



collana

Didattiche, Tecnologie e Media Education

Frontiere per la sostenibilità

Gabriele Biagini, Elena Gabbi

L'educazione attraverso i dati

**Dall'alfabetizzazione critica
all'educational data mining**





Didattiche, tecnologie e media education

Frontiere per la sostenibilità

Co-directors:

MARINA DE ROSSI, MONICA FEDELI

Scientific Committee

Laura Bierema (University of Georgia)
Joellen Coryell (Texas State University)
Clara Bolante e Simeon-Fayomi (Obafemi Awolowo University, Ile-Ife)
Gerald Knezeck e Rhonda Christensen (North Texas University)
Anneke Smits (Windesheim University)
Edward W. Taylor (Penn State University)
Regina Egetenmeyer (University of Wuerzburg)
Loretta Fabbri (Università di Siena)
Luciano Galliani (professore emerito, Università di Padova)
Salvatore Colazzo (Università del Salento)
Ada Manfreda (Università di Roma Tre)
Mario Giampaolo (Università di Siena)
Alessandra Romano (Università di Siena)
Francesca Bracci (Università di Firenze)
Daniela Frison (Università di Firenze)
Concetta Tino (Università di Padova)
Simone Visentin (Università di Padova)
Elisabetta Ghedin (Università di Padova)
Marina Santi (Università di Padova)
Chiara Panciroli (Università di Bologna)
Corrado Petrucco (Università di Padova)
Maria Ranieri (Università di Firenze)

Young Scientists Committee – Young Scientists are a varied group of people.

They are typically early career PhD scientists in academia or researchers:

Marica Liotino, Tommaso Reato, Ruoyi Qui, Taiwo Isaac Olatunji, Ottavia Trevisan

Anna Macaudo, Alessandro Soriani, Daniele Agostini, Francesco Fabbro.

**The volumes of this series have undergone
a double-blind refereeing procedure**

Gabriele Biagini, Elena Gabbi

L'educazione attraverso i dati

Dall'alfabetizzazione critica
all'educational data mining



Il volume è frutto della ricerca svolta presso il Dipartimento di Formazione, Lingue, Intercultura, Letterature e Psicologia dell'Università degli Studi di Firenze e beneficia, per la pubblicazione, di fondi di Ateneo.

Il volume è stato ideato congiuntamente dai due autori che ne sono pertanto co-primi autori. Per la redazione dei capitoli, Gabriele Biagini ha scritto i capitoli 1 e 4, mentre Elena Gabbi ha elaborato i capitoli 2 e 3.



Quest'opera è assoggettata alla disciplina *Creative Commons attribution 4.0 International Licence* (CC BY-NC-ND 4.0) che impone l'attribuzione della paternità dell'opera, proibisce di alterarla, trasformarla o usarla per produrre un'altra opera, e ne esclude l'uso per ricavarne un profitto commerciale.

ISBN volume 979-12-5568-154-0

2024 © by Pensa MultiMedia®

73100 Lecce • Via Arturo Maria Caprioli, 8 • Tel. 0832.230435

www.pensamultimedia.it

Indice

Prefazione	9
di <i>Maria Ranieri</i>	

Capitolo I.

I big data nell'educazione

di <i>Gabriele Biagini</i>	23
1.1 Dal dato ai big data educativi	23
1.1.1 Cos'è un dato?	23
1.1.2 La quantificazione nella storia	28
1.1.3 Gli usi dei dati digitali: implicazioni e criticità	33
1.2 I big data	38
1.2.1 Le origini dei big data: digitalizzazione e datificazione	38
1.2.2 Big data: definizioni e caratteristiche	42
1.2.3 La filiera dei big data	50
1.3 Usi e applicazioni educative dei big data	56
1.3.1 Lo sviluppo dei big data nel campo educativo	56
1.3.2. Le fonti dei big data nell'educazione	60
1.3.3. Le applicazioni dei big data nell'educazione	65

Capitolo II.

Origine, definizioni e fondamenti teorici della *data literacy*

di <i>Elena Gabbi</i>	83
2.1 Definire la <i>data literacy</i> : dagli approcci tradizionali alle prospettive contemporanee	83
2.1.1 Origini ed evoluzione del concetto	83
2.1.2 Inquadramento teorico e dimensioni chiave della <i>data literacy</i>	87
2.2 Elementi di congruenza e discrepanza tra <i>data literacy</i> e altre forme di alfabetizzazione	97

2.2.1	Tecnologie e pratiche: in dialogo con digital, information e <i>data visualization literacy</i>	97
2.2.2	Dimensioni critiche: l'educazione ai media e il ruolo del pensiero computazionale	102
2.2.3	Dati come evidenze: il rapporto con le competenze statistiche e di valutazione	108
2.2.4	La necessità di una prospettiva interdisciplinare	112
2.3	Il ruolo della <i>data literacy</i> nel contesto educativo attuale	115
2.3.1	Un panorama normativo e socioculturale in evoluzione	115
2.3.2	Uno strumento a favore dei cittadini nella società dei dati	121

Capitolo III.

Didattica della *data literacy* ed esperienze di ricerca nel panorama educativo

di *Elena Gabbi*

		147
3.1	Portare l'alfabetizzazione ai dati in classe: formare cittadini consapevoli nell'era dei dati	147
3.1.1	Sviluppo del curriculum e contesti disciplinari	147
3.1.2	Sfide e limitazioni in relazione agli ordini scolastici	152
3.1.3	Progettazione didattica e valutazione	156
3.2	Strumenti e risorse per conoscere, visualizzare, interpretare e creare con i dati	163
3.2.1	Data literacy: dimensione conoscitiva di base	163
3.2.2	Data literacy: dimensione visuale	170
3.2.3	Data literacy: dimensione critico-etica	176
3.2.4	Data literacy: dimensione generativa e creativa	183
3.3	Oltre l'aula scolastica: la <i>data literacy</i> per gli educatori	189
3.3.1	Sviluppare le competenze sui dati nella formazione degli insegnanti	189
3.3.2	Altri attori rilevanti nella prospettiva dell'educazione ai dati	194

Capitolo IV.

Pratiche e tecniche di Educational Data Mining: metodi ed esempi

di *Gabriele Biagini*

	209
4.1 Elementi di Educational Data Mining	209
4.1.1 L'Educational Data Mining: una definizione	209
4.1.2 Le origini dell'EDM	214
4.1.3 Le fasi del processo dell'EDM	216
4.2 Metodologie, tecniche e applicazioni dell'EDM	223
4.2.1 I metodi predittivi	224
4.2.2 I metodi di structure discovery	232
4.2.3 I metodi di relationship mining	238
4.2.4 Il text mining	243
4.3 Questioni aperte nell'uso dell'EDM	245
4.3.1 Questioni economico-sociali: la proprietà dei dati come fonte di vantaggio competitivo	246
4.3.2 Questioni metodologiche: le implicazioni pedagogiche dei sistemi predittivi	249
4.3.3 Questioni tecnico-organizzative: dall'acquisizione dei dati al cambiamento organizzativo	252
4.3.4 Questioni etico-giuridiche: privacy e uso responsabile dei dati	253

Ringraziamenti

Desideriamo esprimere la nostra più sincera gratitudine al Dipartimento di Formazione, Lingue, Intercultura, Letterature e Psicologia dell'Università degli Studi di Firenze e, in particolare, alla professoressa Maria Ranieri. Il suo prezioso accompagnamento nella stesura di questo volume, l'accurata revisione e la supervisione scientifica sono stati fondamentali per il successo di questo progetto. Inoltre, la prefazione da lei scritta ha dato ulteriore valore al nostro lavoro, offrendo una chiave di lettura illuminante e stimolante per il lettore. Senza il suo costante supporto, incoraggiamento e impegno, non sarebbe stato possibile raggiungere questo traguardo. La sua dedizione alla ricerca e all'educazione ci ispirano profondamente e siamo molto grati per l'opportunità di poter collaborare con lei. Grazie di cuore per aver creduto in noi e per averci guidato in questo percorso.

Prefazione

Maria Ranieri

Il tema dei dati nei contesti educativi non è certamente nuovo. Il Ventesimo Secolo ha visto anche il mondo dell'educazione coinvolto in una profonda riflessione di carattere epistemologico sulla funzione dei dati per una migliore comprensione dei fenomeni educativi e un impegno consistente nello sviluppo e messa a punto di strategie di ricerca di tipo empirico o basate su dati, riconducibili al dominio della Pedagogia Sperimentale (La Rosa, 2017). Da Aldo Visalberghi e Luigi Calonghi a Mauro Laeng, Benedetto Vertecchi e Egle Becchi – tanto per citare alcuni dei grandi autori della pedagogia italiana –, si è fatta strada nel nostro Paese, a partire dagli anni Cinquanta/Sessanta, l'idea che sia possibile anche in ambito educativo una ricerca di tipo scientifico, fondata su ipotesi e successive verifiche e alimentata da una riflessività – tutt'altro che ingenua – capace di dialogare criticamente con *ciò che è dato*. Rigore e miglioramento della pratica educativa, o meglio didattica, sono le ragioni che troviamo costantemente richiamate nei classici della Pedagogia Sperimentale, allo scopo di chiarire il senso e il significato di una pedagogia che fa ricorso non solo alle idee ma anche alle parole e ai numeri, intesi come istanze rappresentative del mondo reale e funzionali ad una maggiore conoscenza dei fatti educativi. La didattica, l'insegnamento, non solo non sono terreno di improvvisazione in quanto possiedono le loro tecniche, ma devono maturare un proprio sapere in grado di render conto dell'efficacia delle tecniche, orientando la pratica verso le migliori soluzioni. La costruzione di questo sapere si basa sull'osservazione sistematica dell'esperienza, su una postura

euristica tesa alla comprensione profonda dei problemi educativi, superando l'annosa contrapposizione tra qualità e quantità. Come sottolinea Vertecchi, in un'opera degli anni Novanta dal titolo *Quantità e qualità nella ricerca didattica* (Vertecchi, 1995), la contrapposizione tra quantità e qualità nell'affrontare i problemi della scuola sottintende pregiudizi che ostacolano una piena comprensione della realtà scolastica; la questione va invece vista in termini diversi: qualità e quantità sono momenti diversi di uno stesso processo finalizzato all'interpretazione dei fenomeni educativi, dove il dato, la quantità è ciò su cui si esercita l'interpretazione e ciò senza di cui non ci sarebbe niente da interpretare. E, tuttavia, la centralità del dato non va intesa in senso riduzionistico: il dato non è in sé la verità del fenomeno educativo, sarebbe questo un approccio superficiale. Al contrario, è necessario contestualizzare: non è, infatti, sufficiente raccogliere dati numerici e statistici, ma occorre comprendere il contesto in cui tali dati sono stati raccolti per interpretarli correttamente, evitando di trarre conclusioni errate o fuorvianti. Dal Laeng a Vertecchi, tutti sottolineano come la raccolta dei dati in educazione deve sempre essere accompagnata da una profonda comprensione del contesto in cui tali dati vengono ottenuti, altrimenti rischiano di essere vuoti e privi di significato. I dati vanno usati criticamente, tenendo conto dei limiti delle tecniche e delle possibili distorsioni. Solo così possono contribuire al miglioramento dei sistemi educativi.

Dall'inizio del Nuovo Millennio, dopo un periodo di crisi degli approcci – semplificando – positivisti allo studio dei fenomeni educativi, si assiste ad una nuova ondata di entusiasmo verso i dati educativi attraverso il movimento dell'Evidence-based Education (EBE). Ispirato a visioni analoghe adottate in campo medico e basate sull'idea di fondare i processi decisionali sulle evidenze, l'EBE affonda le sue origini in Inghilterra intorno agli anni Novanta, quando in occasione di una ormai celebre conferenza H. D. Hargreaves prese la parola per muovere una critica radicale alle mancanze dell'indagine accademica in ambito edu-

cativo, auspicando una ricerca in grado di rispondere in modo più esplicito e sistematico ai criteri correnti di scientificità (Ranieri, 2007). A tali carenze, l'EBE contrappone una visione della ricerca educativa che fa leva sull'uso sistematico e rigoroso delle migliori evidenze scientifiche disponibili per informare e guidare le pratiche educative e le decisioni politiche. Slavin (1986), uno dei maggiori studiosi rappresentativi di questo approccio, sottolinea la rilevanza di una postura scientifica e basata sull'evidenza per la comprensione dei fenomeni educativi, incoraggiando gli educatori a prendere decisioni informate dai risultati della ricerca empirica piuttosto che dalle intuizioni o tradizioni non verificate. L'EBE ha trovato anche in Italia una sua diffusione (Pellegrini, Vivanet, 2018), soprattutto come sguardo privilegiato per lo studio dei processi educativi e formativi come pure nell'enfasi posta sull'importanza di ottimizzare i risultati della ricerca in prospettiva sia diacronica sia sincronica (Calvani, 2012).

Tuttavia, nelle sue manifestazioni più rigide, l'EBE è stato oggetto di critiche, anche molto dure (Biesta, 2007). Intanto, vale la pena specificare che all'interno del movimento per l'evidenza in educazione si possono individuare differenze nel modo di intendere il rapporto con l'evidenza attraverso posizioni che oscillano da quello che potremmo definire "determinismo quantocentrico" a quella che potremmo invece chiamare "saggezza del dato". Detto grossolanamente, sono riconducibili alla prima posizione coloro che pretendono di *dedurre* le decisioni educative dai dati statistici, intesi come entità oggettive sintetizzabili tramite metanalisi, mentre più sfumati sono gli orientamenti di coloro che parlano non tanto di educazione evidence-based quanto di pratiche e politiche educative *informate* dall'evidenza. Quest'ultimo è l'approccio tipico dell'EPPI Center che ricorre al verbo *informare* piuttosto che basare (in tal senso si suggerisce di parlare di saggezza del dato...), e che – oltre alle sintesi statistiche – contempla anche quelle narrative e meta-etnografiche (Calvani, Ranieri, 2016). Le differenze tra gli approcci rendono le critiche non uniformemente

applicabili a tutti i fattori dell'evidenza in educazione, e tuttavia il richiamo all'influenza del contesto e dell'ambiente, all'importanza del pluralismo metodologico senza privilegiare il dato quantificato o quantificabile, alla discutibilità della natura causale della relazione tra risultati della ricerca e pratica educativa, tutto ciò ci (ri)porta a guardare ai dati in educazione nella duplice prospettiva dei benefici che possono apportare come base per il miglioramento e delle cautele riflessive che sono indispensabili per un uso critico e non superficiale dei dati educativi. Una simile visione appare tanto più necessaria per il momento storico che stiamo attraversando, al punto da rendere opportuno un approfondimento sul tema dei dati educativi nell'era digitale. Infatti, anche se abbiamo aperto questo contributo, osservando che il tema dei dati in educazione non è nuovo, inedita è invece la massa dei dati oggi in circolazione come anche inedito è il valore oggi assegnato ai dati in termini di produzione di ricchezza, e ciò in tutti i settori incluso quello educativo, richiedendo nuove attenzioni che spaziano dal ripensamento delle alfabetizzazioni ad un approccio critico verso l'uso degli ambienti digitali per la formazione. La "svolta digitale", o anche "the digital turn" (Mills, 2010), ha impresso infatti un ritmo senza precedenti alla velocità di generazione degli dati nel Nuovo Millennio, dando addirittura un rinnovato impulso alla ricerca sull'Intelligenza Artificiale (Elliott, 2021) e allo sviluppo di campi come il *machine learning* o *deep learning*, il cui combustibile è costituito proprio dai dati, i dati digitali. L'apprendimento automatico, infatti, lavora sullo sviluppo di algoritmi e modelli che consentono ai computer di imparare dai dati e di migliorare le loro prestazioni su compiti specifici senza essere esplicitamente programmati per eseguirli. Tali tecnologie sono alla base di molte applicazioni moderne, dagli assistenti virtuali ai sistemi di raccomandazioni di prodotti, ai veicoli autonomi e alle diagnosi mediche. Non solo. Queste applicazioni stanno conoscendo una diffusione anche in ambito educativo attraverso soluzioni tecnologiche per il supporto personalizzato degli studenti

grazie ad agenti intelligenti o la raccomandazione di contenuti adatti alle loro esigenze. Il meccanismo che alimenta la macchina si fonda, ancora una volta, sui dati, dati educativi, ovviamente, se parliamo di istruzione e formazione: dati d'ingresso funzionali alla profilazione e alla personalizzazione, dati in uscita che informano di ritorno il sistema per un suo eventuale adattamento nel senso del miglioramento. È ciò che viene definito: *AI feedback loop*, un processo iterativo attraverso cui il sistema di intelligenza artificiale utilizza i dati generati dalle sue azioni o dai suoi output per migliorare ulteriormente le sue prestazioni. Ironia della sorte, sembrerebbe qui valere il noto detto: "*historia magistra vitae!*", che noi umani stentiamo a ricordare, cioè quello che è stato può informare e informa quello che sarà. Ovviamente, le cose non stanno proprio così, ma le immagini forti servono per stimolare una riflessione... in questo caso, ci interessa sottolineare la circolarità di un processo che diventa o aspira a diventare migliorativo, suggerendo analogie che andrebbero approfondite.

Insieme al *machine learning* per la formazione, un altro settore della ricerca educativa che sta conoscendo una significativa espansione sulle tracce dei dati digitali è quello dei *learning analytics*, ossia delle tecniche utilizzate per misurare, raccogliere e analizzare i dati sugli studenti e le attività che li riguardano per comprendere e ottimizzare i processi di apprendimento e le risorse e gli strumenti che li supportano (Fulantelli, Taibi, 2014). Combinando le tecniche di tracciamento dei comportamenti degli utenti negli ambienti digitali con i dati generati dalle interazioni degli studenti nelle piattaforme si generano situazioni di apprendimento personalizzate, finalizzate all'ottimizzazione dei processi di insegnamento e apprendimento. In realtà, le cose sono più complicate e ciò emergerà chiaramente dalla lettura delle pagine che seguono. Tuttavia, è importante osservare come nell'era digitale l'interesse per la personalizzazione della didattica e il miglioramento dei processi di insegnamento e apprendimento negli ambienti elettronici, tipico dell'*educational technology*, si saldi con l'attenzione

prestata ai dati, tipica della ricerca educativa di taglio sperimentale. Le cesure disciplinari reggono sempre meno alla prova della complessità: un tecnologo dell'educazione che voglia oggi confrontarsi consapevolmente con le nuove frontiere dell'intelligenza artificiale, dell'apprendimento automatico o dei learning analytics e relative tecniche come quelle dell'educational data mining, non può ignorare la questione dei dati in educazione. Viceversa, lo sperimentista non può ignorare il funzionamento delle nuove macchine per apprendere che forniscono la base materiale per lo svolgimento dei processi di insegnamento e apprendimento, influenzando – attraverso le sue grammatiche dettate dal design – tanto l'erogazione dei contenuti quanto la gestione delle interazioni, in una parola la didattica dei processi educativi e formativi. Pertanto, la questione dei dati educativi va affrontata da almeno tre prospettive: quella didattica, in quanto parliamo di strumenti che impattano sui processi di apprendimento; quella valutativa, poiché abbiamo a che fare con sistemi in grado di realizzare un monitoraggio continuo dei processi permettendoci di intervenire in funzione (ri)orientativa e formativa; quella euristica, dal momento che questa disponibilità di dati si presta come base per una maggiore conoscenza dei fatti educativi. Il tutto senza dimenticare la lezione dei padri della Pedagogia Sperimentale, secondo cui – come abbiamo osservato in apertura – non è ammissibile un approccio riduzionistico all'uso dei dati in educazione: occorrono capacità critiche di decodifica dei dati sulla base dei contesti, capacità interpretative fondate su quadri concettuali dinamici e complessi, capacità analitiche in grado di proteggerci dalla tentazione di scambiare la correlazione con la causazione, l'interpretazione con la realtà. Queste capacità vanno formate sia negli insegnanti che nei cittadini. Vanno formate negli insegnanti perché, nonostante i richiami della comunità scientifica ad una maggiore alfabetizzazione della docenza sui temi della ricerca educativa – ivi inclusa la questione dei metodi e dei dati –, il mondo della scuola e dei suoi principali attori, in particolare gli insegnanti, ri-

mane ancora distante da quello dell'università e degli approcci scientifici alla comprensione dei problemi educativi. Questa frattura diventa ancora più stridente con l'avanzare delle tecnologie algoritmiche, che stanno massicciamente varcando la soglia delle aule scolastiche ed universitarie senza una diffusa competenza da parte degli attori umani di comprendere e governare ciò che questo comporta. Porte aperte all'algoritmo non solo a scuola o all'università, ma anche a casa, sui posti di lavoro, nei luoghi della cura o del divertimento: in altri termini, la penetrazione di applicazioni che si nutrono di dati, dei nostri dati, non riguarda solo l'educazione, ovviamente, ma tutti gli aspetti della nostra vita. Ciò rende necessaria una alfabetizzazione ai dati della cittadinanza tutta: come argomenta Elena Gabbi, nelle pagine che seguono, la rilevanza di questo tema ha implicazioni di carattere non solo scolastico in termini di conoscenze di base da acquisire, ma anche politico-sociali rispetto ad una alfabetizzazione in grado di favorire una cittadinanza attiva e consapevole. La questione dei dati nella società contemporanea, infatti, sollecita ulteriori punti da attenzionare, che riguardano aspetti più generali di natura sociale, politica ed etica. Alla digitalizzazione delle nostre società, infatti, si accompagnano i fenomeni della piattaformaizzazione e datificazione (Van Dijck, 2014). Su di essi Gabriele Biagini e Elena Gabbi offrono un affondo critico e analitico nel corso di questo volume. In questa sede introduttiva, ci limitiamo a richiamare un paio di definizioni utili per completare queste note preliminari ad un testo che assume questi fenomeni come prospettiva privilegiata per osservare i processi trasformativi in atto nel mondo educativo contemporaneo. Parafrasando una delle definizioni più note del concetto di piattaformaizzazione, possiamo dire che riguarda la crescente penetrazione sul piano infrastrutturale delle piattaforme digitali nei diversi settori economici e della vita quotidiana, come pure la riorganizzazione delle pratiche e dell'immaginario culturale associato alle piattaforme (Poell, Nieborg, van Dijck, 2019). Per quanto riguarda la seconda espres-

sione, datificazione, si tratta di quel fenomeno attraverso cui le piattaforme digitali trasformano in dati, pratiche e processi che hanno storicamente eluso la quantificazione, includendo non solo dati demografici e di profilazione volontariamente rilasciati dagli utenti nei sondaggi, ma soprattutto metadata comportamentali, la cui raccolta avviene attraverso infrastrutture che espandono le piattaforme come le app, i plugin, i sensori attivi e passivi, i trackers (Mejias, Couldry, 2019). Questi fenomeni, come spiega Gabriele Biagini, stanno investendo pesantemente la sfera educativa al punto che si parla di “piattaformizzazione educativa”, per riferirsi alla crescente dipendenza da piattaforme digitali per la gestione e la distribuzione dell’educazione. Tale processo ha indubbiamente conosciuto un’impennata a causa della pandemia da Covid-19 e alla conseguente transizione della didattica scolastica e universitaria dal formato presenziale alla didattica da remoto o a distanza (DAD) (Decuyper, Grimaldi, Landri, 2020). Possiamo addirittura parlare di un prima e dopo la pandemia da Covid-19 rispetto alla diffusione del digitale nella scuola e nelle università, rendendo il tema dei dati educativi (digitali) un’emergenza culturale che non si può trascurare. Ancora una volta ribadiamo che il tema dei dati in educazione non è nuovo, ma qualcosa è cambiato, qualcosa sta cambiando. Viviamo in un mondo data-driven, dove non è più solo la teoria a guidare la ricerca – e la costruzione dei suoi strumenti – e la pratica – ivi comprese le applicazioni tecno-educative –, ma sono i dati, in particolare i dati digitali generati attraverso le piattaforme senza una specifica intenzionalità che ne guidi la produzione, a suggerire modelli, tendenze, elementi per prendere decisioni, anche in ambito educativo. Purtroppo, però, all’indomani dell’esperienza pandemica non è stata effettuata una adeguata riflessione sull’impatto della piattaforma educativa nell’era data-driven, e oggi, ci troviamo impreparati davanti al dilagare degli algoritmi in educazione e alle loro conseguenze sia sul versante pedagogico che socio-politico. Approfittiamo di questo spazio per avanzare alcune consi-

derazioni a partire dai pochi studi realizzati su questa materia e da una serie di evidenze raccolte a seguito della pandemia da Covid-19 e documentate nel rapporto europeo *What did we learn from schooling practices during the COVID-19 lockdown* (Carretero Gomez et al., 2021).

In primo luogo, la stragrande maggioranza delle scuole europee non era pronta dal punto di vista delle infrastrutture digitali alla DAD e questo ha comportato l'adozione di pedagogie digitali dell'emergenza non basate sul vaglio critico e una governance consapevole, ma – tranne qualche eccezione come ad esempio la Francia – sull'improvvisazione di un consumatore (più che di un professionista, l'insegnante) che può scegliere tra una miriadi di opzioni apparentemente gratuite. Questa abbondanza è risultata disorientante da più punti di vista per la scuola, gli insegnanti, gli studenti e le famiglie. Da un lato, il ricorso ad una pluralità di ambienti digitali, a seconda delle disponibilità del momento e delle possibilità dei docenti, ha generato un sovraccarico cognitivo per gli attori coinvolti: studenti, insegnanti e genitori sono stati chiamati a navigare tra diverse piattaforme di apprendimento, generando confusione e frustrazione (Carretero Gomez et al., 2021). Dall'altra, l'uso di applicativi proprietari e piattaforme commerciali ha surrettiziamente generato una sorta di privatizzazione dello spazio pubblico della scuola, laddove se le aule fisiche della scuola dipendono dagli enti pubblici, le classi virtuali rispondono alle piattaforme digitali dei vari colossi dell'high tech (Decuyper et al., 2020). La transizione emergenziale di scuola e università verso il digitale ha generato anche questo effetto tutt'altro che collaterale, senza una discussione pubblica sulle implicazioni di questa trasformazione. Il tempo è oltremodo maturo per questa discussione, ma occorre una maggiore e migliore conoscenza del significato e delle conseguenze della digitalizzazione dei processi educativi. Da questa prospettiva, il presente volume offre un contributo sostanziale.

Veniamo ad una seconda questione: essa riguarda la gestione

dei dati personali degli studenti e la loro sicurezza. Come si legge nel report curato da Carretero Gomez et al. (2021), l'impressione generale è che tale questione sia stata sottovalutata, sia da parte delle scuole che delle famiglie. Alcuni genitori e insegnanti hanno osservato che c'erano altre priorità e che le piattaforme fornite dalle scuole avrebbero dovuto essere sicure. Abbiamo visto sopra che questo non poteva essere dato per scontato e, pertanto, le applicazioni usate in classe non sempre erano sicure rispetto alla protezione dei dati personali degli studenti. Oltre a ciò, pratiche discutibili si sono diffuse tra gli utenti come ad esempio quella di fotografare i propri figli durante le lezioni online per poi condividerle sui social media oppure mostrare gli ambienti domestici senza alcuna attenzione verso la privacy degli abitanti di questi spazi (Adami, Djonov, 2022).

Se alcuni adulti hanno ritenuto non prioritaria la questione della sicurezza dei dati personali, altri insegnanti si sono al contrario lamentati per questo approccio generale superficiale alle questioni di privacy e sicurezza dei minori in rete. Da più parti è emersa l'esigenza di una formazione di competenze digitali in generale e nello specifico di data literacy in materia di privacy, in riferimento al triplice target di insegnanti, studenti e genitori (Carretero Gomez et al., 2021).

Da ultimo, è stato trascurato il modo in cui le piattaforma digitali influiscono sulle pratiche stesse di insegnamento e apprendimento. Come sottolineano Decuypere et al. (2020), le piattaforme non sono strumenti digitali neutrali, ma artefatti connettivi costitutivi di assemblaggi socio-tecnici attivi che impattano sui processi educativi. In particolare, attraverso le loro attività di tracciamento e generazione di dati digitali che vengono analizzati informando le successive attività, le piattaforme per l'educazione sono di fatto "un tipo di organizzazione emergente: molto spesso operano come un nuovo tipo di impresa che ha inventato e adottato un nuovo mezzo di produzione del valore. Questo nuovo mezzo di produzione del valore ruota centralmente attorno alla

misurazione delle attività degli utenti come dati, così come intorno alla successiva capitalizzazione (ad esempio, delle relazioni potenzialmente stabilite) di tali attività e varie forme di ‘scambio’” (*ivi*, p. 6). Ciò che conta non sono i contenuti ma la misurazione delle transazioni: pertanto, come Facebook non produce media così Classroom di Google, per fare un esempio, non crea contenuti educativi, ma opera come un mezzo per facilitare lo scambio di contenuti ed attività educative, realizzando valore attraverso la capitalizzazione quindi dei processi di scambio di dati (*ibidem*). Sono fenomeni complessi che richiedono una approfondita comprensione per poterne governare le direzioni di sviluppo, piuttosto che essere inconsapevolmente traghettati in un sistema che non capiamo.

Abbiamo preso a pretesto l’esperienza della pandemia da Covid-19 e il suo impatto sulla scuola per portare all’attenzione del lettore il mondo dei dati da un’altra prospettiva, una prospettiva critica che richiede di considerare rischi e opportunità senza scorciatoie. Una prospettiva che guarda all’innovazione digitale senza perdere di vista l’umano, non in termini antropocentrici, ma di benefici sociali che tale innovazione si auspica sia in grado di realizzare. Dare una direzione piuttosto che un’altra è una decisione politica che ha pesanti ricadute sulla vita sociale. Formare cittadini e cittadine in grado di affrontare con profondità questi temi è dal nostro punto di vista essenziale.

Il tema dei dati non è nuovo, lo ribadiamo. È caro da tempo alla ricerca educativa. È tempo, però, oggi, di prendere maggiormente in considerazione la funzione dei dati nei sistemi educativi e formare insegnanti e cittadini da questo punto di vista più consapevoli.

Firenze, 12 Luglio 2024

Bibliografia

- Adami, E., Djonov, E. (2022). 12 Everyday acts of social-semiotic inquiry. Insights into emerging practices from the research collective PanMeMic. In Tan S., K. E. M. (eds.) *Discourses, Modes, Media and Meaning in an Era of Pandemic. A Multimodal Discourse Analysis Approach*. London: Routledge.
- Biesta, G. (2007). Why “what works” won’t work: Evidence based practice and the democratic deficit in educational research. *Educational theory*, 57(1), 1-22.
- Calvani, A. (2012). *Per un’istruzione evidence based. Analisi teorico-metodologica internazionale sulle didattiche efficaci e inclusive*. Trento: Erickson.
- Calvani, A., Ranieri, M. (2016). La dimensione metodologico-conoscitiva. In Bonaiuti, G., Calvani, A., & Ranieri, M., *Fondamenti di didattica. Teoria e prassi dei dispositivi formativi* (pp. 145-169). Roma: Carocci.
- Carretero Gomez, S., Napierala, J., Bessios, A., Mägi, E., Pugacewicz, A., Ranieri, M., Triquet, K., Lombaerts, K., Robledo Bottcher, N., Montanari, M., Gonzalez Vazquez, I. (2021). *What did we learn from schooling practices during the COVID-19 lockdown*. Luxembourg: Publications Office of the European Union.
- Decuyper, M., Grimaldi, E., Landri, P. (2021). Introduction: Critical studies of digital education platforms. *Critical Studies in Education*, 62(1), 1-16.
- Elliott, A. (2021). *Contemporary social theory: An introduction*. London: Routledge.
- Fulantelli, G., Taibi, D. (2014). Learning Analytics: opportunità per la scuola. *Italian Journal of Educational Technology*, 22(3), 157-164.
- La Rosa, V. (2017). *Linee evolutive della pedagogia sperimentale in Italia*. Milano: Franco Angeli.
- Mejias, U. A., Couldry, N. (2019). Datafication. *Internet policy review*, 8(4), 1-10.
- Mills, K. A. (2010). A Review of the “Digital Turn” in the New Literacy Studies. *Review of Educational Research*, 80(2), 246–271.
- Pellegrini, M., Vivanet, G. (2018). *Sintesi di ricerca in educazione. Basi teoriche e metodologiche* (pp. 1-184). Roma: Carocci.

- Poell, T., Nieborg, D., Van Dijck, J. (2019). Platformisation. *Internet policy review*, 8(4), 1-13.
- Ranieri, M. (2007). Evidence based education: un dibattito in corso. *JE-LKS. Journal of E-Learning and Knowledge Society*, 3, 147-152.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational researcher*, 15(9), 5-11.
- Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & society*, 12(2), 197-208.
- Vertecchi, B. (1995). *Quantità e qualità nella ricerca didattica*. Firenze: La Nuova Italia.

I.

I big data nell'educazione

Gabriele Biagini

1.1 Dal dato ai big data educativi

1.1.1 *Cos'è un dato?*

Con questo capitolo intendiamo accompagnare l'ingresso del lettore nel mondo dei big data educativi, ma prima di confrontarci con questo tema, è opportuno soffermarsi sull'elemento fondamentale che ne costituisce la base materiale: il dato. Dunque, cos'è un dato? A questa domanda, gli studiosi hanno risposto proponendo una varietà di spiegazioni, riconducibili ad ambiti disciplinari diversi dalla scienza dell'informazione alla statistica, dall'epistemologia alle scienze sociali. In prima approssimazione, possiamo fare riferimento al Nuovo dizionario De Mauro (2014), che alla voce *dato*, riporta: “certo, determinato”, seguito da cinque occorrenze e dall'elenco di ben dodici parolimitiche (ovvero combinazioni di parole percepite dai parlanti come unità lessicali: dato di fatto, dato che, banca dati, centro elaborazione dati, dati strutturati, dato segnaletico, dato sensibile, flusso di dati, lettore di dati, ecc.). Similmente, ma con alcune accentuazioni, nell'Enciclopedia Treccani il vocabolo *dato* è definito come “ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione¹” e, in tal senso, si distingue dall'informazione in

1 Fonte: <https://www.treccani.it/vocabolario/dato/>

quanto “notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere” (Atzeni, Ceri, Fraternali, Paraboschi, Torlone, 2018, p. 2). In termini più articolati, l’interconnessione tra dato (quantità o condizione nota), informazione e conoscenza, può essere rappresentata in forme più complesse come formalizzato nella piramide DIKW (acronimo di Data, Information, Knowledge e Wisdom) (Ackoff, 1989), che illustra la progressione gerarchica dei concetti di dato, informazione, conoscenza e saggezza, in cui ogni livello si basa su quello precedente.

Alla base della piramide troviamo i dati grezzi, risultanti da input, applicazioni e oggetti connessi. Al livello superiore, quello delle informazioni, i dati vengono raggruppati, verificati, resi coerenti e armonizzati, affinché possano essere confrontati. Il livello successivo, quello della conoscenza, rappresenta la comprensione delle informazioni, la loro contestualizzazione supportata dal confronto con i dati esterni. All’apice della piramide troviamo infine il livello della saggezza, che consiste nella capacità di applicare la conoscenza e l’esperienza acquisite per prendere decisioni.

Queste prime schematizzazioni sono utili per comprendere il contesto più ampio all’interno del quale si colloca la nostra riflessione sui dati e per perimetrare al tempo stesso i contorni di questa riflessione. Tuttavia, è importante notare come questa interpretazione oggettivistica del dato quale elemento primo e antecedente a qualsiasi elaborazione sia stata superata da approcci più recenti, che – come vedremo nel prosieguo – ne hanno ampliato i confini incorporando dimensioni della soggettività maturate nel tempo.

Del dato ci è nota l’etimologia: il termine deriva dal latino *datum*, participio passato del verbo dare (Rivière, 2020), mentre i greci, da Platone ad Aristotele, Euripide e altri, utilizzavano la parola *dotos* nell’ambito del linguaggio matematico, per indicare il punto di partenza o la base di una indagine da cui si prendono le mosse successive. Oltre alle origini del termine, conosciamo

anche l'epoca che ne registra il battesimo nella lingua inglese grazie a un importante studio statistico-semanticò ed epistemologico di Rosenberg (2013), da cui si apprende che l'uso del termine – ancorché in senso teologico – è attestato a partire dal XVII secolo. Infatti, la prima occorrenza nell'Oxford English Dictionary risale al 1646 (Borgman, 2020; Rosenberg, 2013).

Anche se per il contesto italiano non disponiamo di uno studio analogo a quello di Rosenberg, sia il "Dizionario etimologico della lingua italiana" (Cortellazzo-Zolli, 2004) sia il "Grande Dizionario della lingua italiana" (Battaglia, 1966), fanno riferimento a Egnazio Danti che usa esplicitamente il termine *dato* in senso scientifico (e non teologico) in una sua opera del 1573.

Questa tendenza si consolida negli anni successivi. Tra i secoli XVII e XVIII, infatti, in piena epoca illuministica, il ricorso alla terminologia del dato è stato funzionale al tentativo di oggettivare la conoscenza, ricorrendo a fatti e prove scientifiche e prendendo le distanze dall'interpretazione teologica incentrata sulle scritte, nella convinzione che i dati avessero un significato intrinseco indipendentemente dal contesto, dal mezzo, dal formato e dalla lingua (Kitchin, 2014a).

Oggi, a distanza di secoli dagli esordi di questo termine, non esiste una sua definizione univoca e condivisa, mentre si fanno strada visioni che si spingono oltre l'assimilazione del concetto di dato all'idea di "un oggetto naturale provvisto di una essenza propria" (Borgman, 2020, p. 52). Rivière (2020, p. 129) addirittura osserva che i dati "[...] non esistono in natura [...]. In altri termini i 'dati non sono dati' [*les données ne sont pas données*]. Occorre costruirli, acquisirli. Necessitano di un lavoro di modellazione a monte, di astrazione, di specificazione dei concetti, quindi dei contesti prima di immaginare di produrre valori"².

2 Traduzione degli autori qui e altrove.

In effetti, quando utilizziamo il termine *dati*, solitamente ci riferiamo a elementi che vengono raccolti, frammenti selezionati da un insieme più vasto di informazioni. Da questo punto di vista, Kitchin (2014a) sostiene che sarebbe più corretto parlare di *capta* anziché di dati, sottolineando che ciò che viene misurato e raccolto rappresenta solo una frazione del panorama informativo complessivo e, quindi, intrinsecamente selettivo e rappresentativo di una realtà limitata. In questa ottica, i dati devono essere considerati come prodotti sociali, come unità di misura estrapolate da fatti o eventi sulla base delle convenzioni con cui sono stati acquisiti. In altre parole, è necessario prestare molta attenzione alla loro individuazione. A tal proposito, Borgman (2018, p. 18) sottolinea come “i dati sono rappresentazioni [...]. Queste rappresentazioni variano in base allo scienziato, alle circostanze, e nel tempo” e ancora: “Essi esistono in un contesto, che ne influenza il significato insieme alla prospettiva dell’osservatore”. Insomma, è cruciale comprendere che non si tratta di materiale grezzo, passivo che, una volta identificato, costituisce la base per la produzione di prove o informazioni (Bowker, 2008). Al contrario, “i dati richiedono la nostra partecipazione, hanno bisogno di noi” (Gitelman, Jackson, 2013, p. 6). Pertanto, non vengono semplicemente forniti, ma vengono attentamente selezionati in base a vari fattori sociali, tra cui il background degli esperti, le deontologie professionali e le pratiche tecnologiche applicate: organizzazioni e istituzioni che li raccolgono, analizzano e diffondono. I contesti, finanziari e di mercato, influenzano le modalità d’uso dei dati, insieme alla disponibilità di strumenti per la loro generazione e ai contesti normativi ed etici che regolano l’intero processo di acquisizione e utilizzo. I dati, quindi, non esistono in un vuoto concettuale e sociale, ma sono plasmati dalle idee, dalle tecniche, dalle tecnologie, dai sistemi, dalle persone e dai contesti in cui si trovano e in cui vengono elaborati (Kitchin, 2014a).

Concludendo, la sintesi più esaustiva, sostiene Borgman (2020, p. 52), “consiste nel dire che i dati sono rappresentazioni

di osservazioni, di oggetti o altre entità finalizzati a rilevare fenomeni a scopo di ricerca”. Il riferimento di Borgman al concetto di entità arricchisce di ulteriori elementi le prospettive sopra delineate, in quanto conferisce importanza a ciò che può “avere una esistenza concreta” (*ibidem*) – come testi, papiri, elementi digitali, segnali – e assumere un significato diverso in relazione al contesto: “Una entità diventa un dato solo se è utilizzata per evidenziare un fenomeno e una medesima entità può essere la manifestazione di fenomeni diversi” (tra gli esempi più calzanti è citato l’uso di vecchie fotografie che, a seconda dei casi, possono servire per illustrare diverse tendenze sociali, mode, relazioni familiari) (*ibidem*). Da qui l’invito ad assumere un punto di vista diverso (quasi provocatorio) per superare la difficoltà di definizioni non esaustive, scegliendo di chiedersi, al posto di “che cosa è un dato”, “quando un termine assume la funzione di dato?”, tenendo però conto anche di altri fattori basilari: ad esempio, “come gli individui, i laboratori e le comunità creano, selezionano e utilizzano i dati” (ivi, p. 39), quali ne siano i fattori di rilievo associati, e quali le domande o i metodi di ricerca, i criteri di valutazione dell’evoluzione dei dati e quale sia l’incidenza dovuta alla disponibilità del numero sempre crescente di dati in formato digitale.

Interrogando la Storia, troviamo diverse risposte che vanno nella stessa prospettiva immaginata da Borgman (2020) e molte di esse permettono di determinare la nascita, la costruzione e l’evoluzione nel tempo del concetto di dato, prima ancora della sua formulazione assiomatica, di cui si trova traccia nei dizionari e che spesso lo vede associato, erroneamente, al concetto di realtà inconfutabile. Il cammino è stato abbastanza lungo e complesso e, come vedremo nel paragrafo successivo, ripercorrerlo serve non tanto a spiegare “cos’è un dato” ma a comprendere “come è nato” e “come si è trasformato”, con particolare riguardo al momento in cui (XVIII secolo) la sua originaria formulazione astratta (al singolare) ha finito per coesistere con il riferimento a entità con-

crete (di natura operativa e funzionale come dati personali, economici, statistici, medici, metereologici, ecc.).

1.1.2 *La quantificazione nella storia*

Se sul piano lessicale il termine *dato* si afferma in età moderna con una varietà di significati, che abbiamo precedentemente esplorato, dal punto di vista storico la necessità di quantificare e raccogliere informazioni è un'esigenza antica. La storia dell'umanità è ricca di esempi che dimostrano come il concetto di dato sia stato fondamentale per lo sviluppo delle civiltà. I greci, ad esempio, avevano già compreso l'importanza della raccolta dei dati per scopi pratici. Come Tucidide racconta, nel 431 a.C., Sparta dichiarò guerra ad Atene. Nel suo resoconto della guerra, lo storico descrive come "le forze plateane assediate, fedeli ad Atene, pianificarono di fuggire scalando il muro che circondava Platea, costruito dalle forze del Peloponneso, guidate dagli spartani. Per fare questo avevano bisogno di sapere quanto era alto il muro in modo da poter realizzare scale di lunghezza adeguata. Gran parte del muro del Peloponneso era stato ricoperto di grezzi ciottoli, ma fu trovata una sezione in cui i mattoni erano ancora chiaramente visibili e ad un gran numero di soldati fu affidato il compito di contare gli strati di questi mattoni esposti" (Holmes, 2017, p 25). In questo brano di Tucidide sulla guerra fra Atene e Sparta, troviamo tracce di quelle entità di cui parla Borgman: sono i mattoni a rappresentare le entità e i dati, e la presenza e la disponibilità di oggetti sono diventate materia di rappresentazione ed espressione per registrare e comunicare. Lo spiega bene Holmes (ivi, pp. 25-26) nel suo lavoro dedicato al significato e alle implicazioni dei big data nella società moderna: "Lavorare a distanza al sicuro dagli attacchi nemici comportava inevitabilmente degli errori, ma come spiega Tucidide, dati i molti conteggi effettuati, il risultato più frequente sarebbe stato quello corretto. I Plateesi

conoscevano la dimensione dei mattoni locali utilizzati e vennero così costruite scale della lunghezza necessaria per scalare il muro. Ciò permise a un manipolo di diverse centinaia di uomini di fuggire e l'episodio può essere considerato l'esempio più impressionante di raccolta e analisi di dati storici. Ma la raccolta, l'archiviazione e l'analisi dei dati sono anteriori addirittura a Tucidide di molti secoli". Holmes (*ibidem*) si serve di questa nota storica per rappresentare il lungo cammino percorso dall'essere umano nella messa a punto delle tecniche di raccolta dei dati, a partire dalle antiche modalità di ricerca, misurazione e registrazione di oggetti. L'autore si spinge fino al Paleolitico superiore, passando poi dai reperti di "ossi misuratori" con tacche in Africa (Swaziland), ai geroglifici scritti su legno e papiro presso gli antichi egizi, alle corde colorate intrecciate delle popolazioni Inca. Il testo di Holmes fa parte di quell'insieme più ampio di lavori che hanno tentato di costruire una genealogia del dato prima dell'uso dei computer. Questi contributi documentano come persino la nascita della celebre Biblioteca di Alessandria (verso il 300 a.C.) abbia costituito una fucina per il calcolo dei dati dell'Impero, censiti da molti dei libri che vi erano depositati. All'epoca, lo scopo consisteva nell'ottenere un vantaggio strategico nei confronti degli altri Imperi (quello romano, ad esempio) in termini di potere e controllo; oggi, le ragioni che alimentano la corsa ai dati sono di natura prevalentemente economica, coinvolgendo i processi di avanzamento della ricerca scientifica e tecnologica. Ad ogni modo, tutte queste tracce di misurazioni che ritroviamo nella storia, anche lontana, evidenziano come la volontà di registrare, censire e comunicare dati rappresenti una preoccupazione tutt'altro che recente, la quale si è a sua volta evoluta fino a raggiungere la forma più nota: quella del censimento. Holmes (*ivi*, p. 27) segnala che "uno dei primi usi ben documentati dei dati è il censimento intrapreso dai Babilonesi nel 3800 a.C.", un'impresa volta a registrare "il numero della popolazione e dei beni di consumo, come latte e miele, al fine di fornire le informazioni necessarie per il

calcolo delle tasse”. È noto che anche gli antichi romani usavano realizzare censimenti ogni cinque anni. Come riporta Tito Livio nella sua opera *Ab Urbe Condita*, il primo fu istituito alla fine del VI secolo a.C., sotto il regno di Servio Tullio, sesto re di Roma. Questo procedimento, che si svolgeva nel Foro Romano ad opera dei censori, era inteso non solo come un conteggio della popolazione, ma anche come una valutazione delle proprietà e delle risorse economiche dei cittadini romani. L'obiettivo era duplice: da un lato, organizzare in modo efficiente le strutture militari della Repubblica, distribuendo gli oneri militari in base alla ricchezza e allo status sociale; dall'altro, garantire una giusta ripartizione delle responsabilità fiscali.

Questo sistema di censimento era incredibilmente avanzato per l'epoca, costituendo uno dei primi esempi di raccolta sistematica di dati su larga scala. La capacità di gestire tali volumi di informazioni, classificarli e utilizzarli per la pianificazione e l'amministrazione dimostra un'intuizione precoce dei principi che sottendono al moderno concetto di dato digitale. Sebbene le tecnologie e le metodologie siano radicalmente cambiate, l'idea di base di raccogliere e analizzare dati per prendere decisioni informate è un filo conduttore che collega l'antica Roma al mondo contemporaneo.

Il censimento ha quindi rappresentato una sorta di prima forma organizzata per registrare, quantificare beni, possedimenti, persone, fino alle soglie delle scoperte della statistica e del digitale nell'era tecnologica che stiamo vivendo. È interessante notare che è proprio l'evoluzione del concetto di censimento (*censere* in latino vuol dire stimare), usato nell'accezione di dedurre, a costituire agli occhi di Mayer-Schönenberger e Cukier (2013) un ulteriore progresso storico e un primo tentativo di approccio statistico ai dati. Un approccio che nasce dall'idea, attribuita a un merciaio inglese (John Graunt), appassionato di ricerche demoscopiche, in occasione della peste di Londra (1665), secondo cui da “un piccolo campione si potessero estrapolare utili conoscenze sulla

popolazione in generale” (ivi, p. 36). Il metodo (statistico) per campionatura venne così adottato, malgrado risultasse impreciso e soggetto a casualità; tuttavia, nelle operazioni di censimento di grandi masse di informazioni finì per essere comunque affiancato al “metodo primitivo di contare ogni singolo elemento” (*ibidem*) con tempi che si rivelarono eccessivamente lunghi, specie nell’espletamento decennale dei censimenti di paesi con crescente popolazione come quella degli Stati Uniti. Alla fine del XIX secolo, si comprese che una svolta sarebbe avvenuta, accorciando i tempi della registrazione manuale di dati sempre più vasti, precisi e vari, derivati dal campo di quella rivoluzione scientifica che tra Sette e Ottocento fu opera di studiosi, matematici, epidemiologi (Holmes, 2017). Secondo Holmes, l’occasione materiale di avanzamento si presentò nel corso dell’espletamento del censimento statunitense del 1890, quando l’inventore Hermann Hollerith (un impiegato dell’US Census Bureau) propose di utilizzare un tabulatore a schede perforate che permise di accorciare notevolmente i tempi di archiviazione e analisi dei dati rilevati nel censimento (8 anni si ridussero a un anno). L’impresa non solo ebbe successo in molti paesi ma, in seguito, Hollerith vendette la sua invenzione all’impresa che getterà le basi della futura IBM e delle sue macchine a schede perforate. Nel 1969 Hollerith ebbe un riconoscimento – la dedica della Hollerith Card Code – per la sua provvidenziale scoperta (*ibidem*). Il successivo avanzamento tecnologico nella raccolta e analisi di dati avverrà grazie all’utilizzo massiccio del computer, che offrirà “la possibilità di analizzare enormi quantitativi di dati su un certo argomento, anziché essere costretti a lavorare su insiemi più ristretti” (Mayer-Schönenberger, Cukier, 2013, p. 33). Quel momento segnò l’avvio di un’accelerazione senza precedenti verso l’impiego del computer nella raccolta e analisi dei dati, sancendo l’inizio di una trasformazione epocale che ha visto il dato digitale infiltrarsi in modo sempre più pervasivo nelle nostre vite quotidiane. Tale processo di digitalizzazione ha, inoltre, rivoluzionato radicalmente il concetto stesso

di dato, trasformandolo da forma statica e relativamente circoscritta nel suo utilizzo a inestimabile risorsa dinamica e in continua evoluzione, caratterizzata da una straordinaria duttilità e adattabilità alle mutevoli esigenze dell'ambiente digitale contemporaneo. Come sottolineato da de Kerckhove (2020), la trasformazione digitale in corso riflette il crescente ruolo del software (cioè, l'insieme dei programmi) nella gestione e nell'analisi dei dati. Infatti, mentre l'evoluzione dell'hardware (insieme dei dispositivi meccanici e elettronici) ha fornito la base per una sempre più evoluta raccolta ed elaborazione dei dati, è lo sviluppo della parte software che ha assunto un ruolo sempre più centrale per dare vita a nuove modalità di interpretazione e utilizzo dei dati. Questo processo tecnologico ha contribuito, però, anche alla trasformazione del concetto di dato, da entità statica a, come abbiamo detto, risorsa dinamica. Si può dire, quindi, che l'evoluzione del significato di dato è strettamente intrecciata con lo sviluppo tecnologico – quasi a divenire un tutt'uno – con la trasformazione digitale che rappresenta l'ultimo capitolo di una lunga storia di progresso culturale.

Nel tempo, il concetto di dato è stato modellato e ridefinito dalle idee, dalle innovazioni tecnologiche e dalle pratiche socio-culturali delle persone coinvolte nella loro elaborazione. Pertanto, una comprensione approfondita di questa evoluzione storica è essenziale per affrontare le sfide e per cogliere le opportunità emergenti nel contesto del mondo digitale contemporaneo.

Come vedremo più dettagliatamente nel prossimo paragrafo, la diffusione dei dispositivi digitali e delle piattaforme online ha generato una quantità senza precedenti di dati, che vengono costantemente raccolti, elaborati e utilizzati per informare decisioni, guidare l'innovazione e plasmare l'esperienza individuale.

1.1.3 *Gli usi dei dati digitali: implicazioni e criticità*

I dati digitali sono rappresentazioni numeriche di informazioni che possono essere create, memorizzate, processate e trasmesse utilizzando tecnologie digitali. Questi dati, a differenza dei dati tradizionali che possono essere raccolti manualmente o archiviati in forma fisica, sono generati e gestiti attraverso dispositivi elettronici, sistemi informatici e reti di comunicazione (Kitchin 2014a). La differenza tra dato e dato digitale risiede principalmente nel modo in cui vengono raccolti, archiviati e utilizzati. Mentre un dato può essere qualsiasi tipo di informazione raccolta da una osservazione o misurazione, un dato digitale è specificamente codificato in formato binario, il linguaggio dei computer, che ne permette l'elaborazione automatizzata, l'analisi avanzata e la trasmissione istantanea attraverso le reti globali (Mayer-Schönenberger, Cukier, 2013).

I dati digitali si sono imposti come pilastri centrali nell'ambito delle operazioni aziendali, governative, dei media, delle città, nonché nel panorama dell'innovazione scientifica influenzando direttamente la vita quotidiana di miliardi di individui in tutto il globo. Una vasta gamma di organizzazioni commerciali, tra cui supermercati, rivenditori online, istituzioni finanziarie, società di intrattenimento, social media e molti altri, si affida alla massiccia raccolta di dati dai propri clienti e utenti per ottimizzare le loro operazioni. Come sottolineato da Beer (2013), nell'ampio panorama delle aziende digitali, colossi come Amazon, Netflix e Spotify si distinguono per la capacità di predire le preferenze dei propri utenti e formulare raccomandazioni personalizzate sulla base di dati acquisiti su larga scala dei loro clienti. Società come Google e Meta raccolgono informazioni dettagliate sui comportamenti degli utenti attraverso ogni loro clic per creare profili ad hoc che, sulla base delle preferenze espresse, guidano l'adattamento e la personalizzazione dei contenuti. Come descritto da van Dijck (2013), questi dati vengono anche sfruttati per fini

pubblicitari mirati, influenzando l'esperienza complessiva degli utenti sui social media. Perfino una iconica destinazione di divertimento come Disney World ha abbracciato la logica dei dati: a ogni visitatore vengono forniti braccialetti personalizzati, noti come MagicBands, che costantemente registrano informazioni riguardanti la loro posizione all'interno del parco, le scelte di ristorazione, i tempi di attesa per le attrazioni e gli spettacoli a cui assistono. Tutti questi dati sono vagliati dagli analisti della Disney nei loro laboratori di ricerca per ottimizzare l'esperienza dei visitatori e migliorare le attrazioni del parco, trasformandolo in una sorta di città sperimentale, dove la sorveglianza e l'utilizzo dei dati sono parte integrante dell'esperienza complessiva dei visitatori, come descritto da Bogost (2014). Questo modello sta recentemente ispirando molti paesi, che cercano di adattare strategie simili su scala nazionale: attraverso i dati biometrici e le informazioni contenute nei documenti di identità, possono monitorare i movimenti dei propri cittadini e mantenere, ad esempio, un registro dettagliato dei loro viaggi. Similmente, quando si accede ad un qualunque servizio pubblico online si generano informazioni di varia natura che vengono catalogate e analizzate.

Da tempo, anche in politica si è iniziato ad adottare in modo massiccio l'analisi dei dati, sia per condurre sondaggi, affidandosi ad agenzie specializzate, sia utilizzando metodi statistici. L'obiettivo è la creazione di gruppi di elettori rappresentativi sulla base delle varie caratteristiche della popolazione nazionale per ottenerne informazioni comportamentali. Pur non avendo pretese di scientificità, queste informazioni ottenute sono comunque considerate attendibili e vengono utilizzate per "organizzare o mobilitare segmenti specifici dell'elettorato" in vista delle elezioni (Chadwick, Stromer-Galley, 2016, p. 284).

Recenti e numerose indagini della magistratura, in tutti i Paesi dell'Occidente, sull'impropria conservazione di dati personali e sul loro potenziale fraudolento utilizzo per fini illeciti hanno evidenziato quanto questo campo sia delicato ed abbia quindi ne-

cessità di essere regolamentato. A livello europeo, l'argomento – insieme ad altri – è stato regolato nel 2018 attraverso la legge sulla protezione dei dati personali (GDPR³) e un Libro bianco della Commissione Europea dal titolo “Un approccio europeo all'eccellenza e alla fiducia”, pubblicato nel 2020 (Commissione europea, 2020). In Italia, è stata creata la figura del Garante della Privacy che ha appunto lo scopo di normare e vigilare su questo aspetto al fine di evitare che soggetti privati o pubblici commettano violazioni in tal senso e l'Autorità per la Garanzia nelle Comunicazioni (AGCOM), che ha, tra le altre finalità, anche il compito di tutelare “i consumi di libertà fondamentali degli utenti”. Il concetto è un po' criptico, ma vuole significare il controllo delle attività e dei comportamenti che possono incidere sui diritti fondamentali di ciascuno, che abbracciano – come è noto – due dimensioni, quella etico-politica e quella giuridica.

Il tema dell'impropria conservazione e utilizzo dei dati evidenzia l'impellente necessità di una regolamentazione rigorosa nel panorama digitale. Questa esigenza è strettamente correlata alla natura fluida e pervasiva dei dati digitali, la quale ha aperto la strada a una sempre più facile manipolazione degli stessi, agevolando la creazione e la diffusione di notizie distorte e, di conseguenza, andando a minare la fiducia pubblica e la stabilità dei processi informativi. In questo contesto, la regolamentazione dell'uso dei dati diventa cruciale per contrastare effetti dannosi sull'attendibilità delle informazioni, dalla falsificazione dei dati personali al furto di identità e alla costruzione di notizie infondate (fake news). Serve inoltre a garantire l'integrità dell'informazione, sempre più spesso condizionata dalla ricerca del consenso a di-

3 Il General Data Protection Regulation o GDPR è una regolazione varata nel 2018 dalla Commissione Europea, “il cui intento è quello di proteggere i cittadini dall'abuso dei loro dati, problematizzarne la compravendita e il riuso e istigare forme più sofisticate di archiviazione e mobilitazione dei dati stessi” (Leonelli, 2018, p. 69).

scapito dei fatti. Così, sebbene l'analisi accurata e professionale dei dati sia certamente preziosa per creare le condizioni di una comprensione più approfondita degli avvenimenti rilevanti, in politica o nel giornalismo, e del loro impatto sul pubblico, una interpretazione inaffidabile di essi può sfociare in situazioni di condizionamento, di alterazione della percezione della realtà, influenzando le persone (ad esempio, per consolidare le convinzioni esistenti, personalizzando le informazioni sulla base della visione del mondo di una particolare categoria di utenti anziché impostare un cambiamento sfidandone gli interessi con diverse prospettive). Come osserva de Kerckhove (2020, p. 82), “la maggior parte delle persone non si preoccupa di verificare la veridicità o addirittura la credibilità delle affermazioni che sente”. La questione è complessa e, come vedremo nel capitolo 2 (par. 2.2.2), trova ampia trattazione nell'ambito della *media and information literacy*; tuttavia, uno dei motivi per questa tendenza a credere nelle informazioni senza controlli preventivi può essere sicuramente ricercato nel fatto che la sempre maggiore pervasività delle informazioni digitali nella società contemporanea ha condotto a una sorta di saturazione cognitiva (il noto problema del *knowledge overload* di Weinberger, 2011). In un contesto in cui siamo costantemente sommersi da una moltitudine di dati provenienti da varie fonti, le persone possono sentirsi sopraffatte e talvolta disinteressate all'autenticità dei dati che incontrano. Questa abbondanza di notizie, spesso presentata senza filtri o controlli sulla veridicità, può generare un atteggiamento di apatia nei confronti della ricerca della verità con il risultato di accontentarsi di ciò che viene proposto senza metterne in discussione l'attendibilità o senza mantenere un atteggiamento critico nei confronti delle fonti.

Abbiamo esteso il nostro campo, tuttavia è essenziale considerare l'esistenza di questi fenomeni legati all'uso sempre più massivo dei dati digitali e, talvolta, alla loro mistificazione. Siamo, in sostanza, nel campo della post-verità – in inglese *post-truth* – che

si ha quando la notizia viene accettata come vera sulla base di sensazioni e emozioni senza alcuna verifica della reale veridicità dei fatti menzionati (Keyes, 2004). In altre parole, i fatti – accertati – sono meno importanti per formare una opinione rispetto a emozioni o convinzioni personali. Per sottolineare la portata di questo fenomeno è interessante menzionare come l'espressione *post-truth* sia stata eletta parola dell'anno nel 2016 dall'“Oxford English Dictionary”.

Vi è anche un ulteriore aspetto rilevante nel contesto dei dati digitali, che conduce ad un risultato analogo: il problema dei dati parziali. Nel mondo digitale, la raccolta e l'analisi dei dati avvengono spesso attraverso algoritmi che selezionano specifici campioni di dati da un vasto panorama informativo. Questa selezione, necessaria per rendere i dati gestibili e analizzabili, può portare a interpretazioni parziali e talvolta fuorvianti della realtà. Ogni studio scientifico necessariamente semplifica la realtà in modo da poterne investigare un aspetto specifico. Quello che preoccupa non è il fatto di ridurre la realtà, a partire dalla sua infinita complessità, ad un campione di dati limitati nella loro portata e rappresentatività. Questa riduzione è una componente essenziale di ogni processo di analisi. Il rischio di mistificazione è generato, invece, dalla tendenza degli utenti a non considerare che stanno utilizzando non è un campione ben bilanciato della realtà, ma piuttosto una selezione di informazioni fatta in parte per limiti pratici e in parte per ragioni concettuali, “facilitando la produzione di interpretazioni che mistificano i fatti invece che aiutare a comprenderli” (Leonelli, 2018, p. 58). Questa osservazione non deve comunque inficiare la validità complessiva dell'utilizzo dei dati, specie in campo tecnologico.

Dall'analisi dei diversi utilizzi e delle implicazioni circa la gestione dei dati digitali alle sfide legate alla veridicità delle informazioni, nel prossimo paragrafo ci soffermeremo sul mondo dei big data che, come vedremo più nel dettaglio, essendo caratterizzati da volumi enormi, velocità di generazione rapida e varietà di

fonti, rappresentano, a partire dalla loro evoluzione storica e tecnologica, un fenomeno chiave nell'ambito della gestione dell'informazione digitale.

1.2 I big data

1.2.1 *Le origini dei big data: digitalizzazione e datificazione*

Dall'analisi delle sfide della post-verità e dell'uso di dati parziali, emerge chiaramente quanto sia cruciale l'accuratezza nella raccolta e nell'interpretazione dei dati per la comprensione della realtà. Si tratta di una preoccupazione fondamentale che ha trovato il suo punto di svolta con la nascita del computer e la conseguente possibilità di digitalizzare e immagazzinare nelle macchine le informazioni analogiche, rendendole di conseguenza più facili da archiviare e, poi, da elaborare e utilizzare. Questo avanzamento tecnologico ha avuto inizio intorno agli anni '90, quando l'uso massiccio del computer – nato come strumento per computare/contare, fanno notare Mayer-Schönenberger e Cukier (2013) – ha permesso di analizzare enormi quantità di dati, rompendo con le limitazioni dei campionamenti più ristretti. Questa trasformazione ha dato il via all'era di quell'eccedenza denominata big data, caratterizzata dalla crisi del metodo per campionamento e dall'emergere della formula di ricerca identitaria ($N = \text{tutti}$). Tale formula indica come l'uso di enormi volumi di dati comporti, come conseguenza, l'abbandono dell'approccio di tipo statistico, basato su criteri selettivi e risposte determinate da domande prestabilite (su dati precisi e rilevati da piccole campionature dei fenomeni), per accogliere la massa dei big data, spesso non strutturati e soggetti a “confusione” e “imprecisione” (ivi, pp. 52-53). Tuttavia, siamo sempre agli albori: la presenza dei computer non significava disponibilità di dati e tanto meno la loro interconnessione, ossia la capacità di combi-

nare e analizzare insieme dataset⁴ provenienti da diverse fonti per ottenere nuove informazioni. I gruppi di dati, in questa prima fase, erano ancora spesso classificati per specifiche finalità e in genere slegati gli uni dagli altri. Di solito, si procedeva per campioni. Nel tempo, sono stati gli sviluppi della matematica insieme alla nascita e all'aumento di potenza dei calcolatori a consentire il passaggio dal procedimento di raccolta manuale e analogica dei dati a quello indicato da Mayer-Schönenberger e Cukier (ivi, p. 109) con il neologismo *datification*, vale a dire “conversione di un fenomeno in forma quantitativa in modo da poterlo tabulare e analizzare”. Tale procedimento – considerato quale punto di arrivo di una lunga storia, legata al desiderio di misurare la realtà e di registrarne i dati – conosce una ulteriore fase di svolta in tempi più recenti grazie alla digitalizzazione, processo attraverso cui si rendono intelligibili al computer, codificandole, determinate informazioni – comprese quelle che erano storicamente sfuggite alla quantificazione – per poterle registrare e archiviare (Van Dijck, 2014). “In poche parole” – spiegano Mayer-Schönenberger e Cukier (2013, p. 115) – “la digitalizzazione mette le ali alla datificazione. Ma non la sostituisce. La mera digitalizzazione [...] di per sé non datizza”. Parallelamente, con l'avvento e l'enorme diffusione delle piattaforme digitali, abbiamo visto come il fenomeno della datificazione abbia riguardato anche la trasformazione in dati di vari aspetti della vita sociale e individuale, rendendoli analizzabili e permettendo di trasformarli in informazioni dotate anche di valore economico (Cavagnero, 2019). Questo fenomeno,

4 Un insieme di dati o dataset (anche data set) è una collezione di dati. La struttura più comune è quella di forma relazionale in cui ogni colonna della tabella rappresenta una particolare variabile, e ogni riga corrisponde ad un determinato membro del dataset in questione. La dimensione del dataset è quindi data dal numero dei membri presenti, che formano le righe, e dal numero delle variabili di cui si compone, che formano le colonne.

noto come piattaforma (platformisation), si riferisce alla crescente centralità delle piattaforme digitali che non solo facilitano la raccolta, l'archiviazione e l'analisi dei dati su larga scala, ma strutturano anche le interazioni economiche e sociali attraverso le loro architetture tecnologiche (Corazza, 2021; Perrotta, 2021; Raffaghelli, 2021). Piattaforme come Google, Facebook e Amazon, fungono da catalizzatori e mediatori di enormi flussi di dati, organizzando e influenzando l'accesso alle informazioni e il valore che ne deriva. Oltre ad accumulare grandi quantità di dati, stabiliscono anche nuove norme per la privacy, la sicurezza e l'economia dei dati, configurando così nuove modalità di interazione sociale e commerciale (Rivoltella, 2024). La piattaforma, insomma, ha ampliato la portata e l'impatto dei big data, introducendo complessità aggiuntive nella gestione e nell'interpretazione dei fenomeni globali (Selwyn, 2020).

Ciò detto, qual è stata l'evoluzione dei big data nel corso del tempo? Si può parlare di big data prima dell'era digitale? Se sì, in che senso? L'avvento dei big data non ha seguito un processo lineare, ma anzi si è sviluppato seguendo le spinte dell'evoluzione tecnologica, secondo un movimento che potremmo definire a ondate (Ambrose, 2015).

La prima ondata dei big data può essere fatta risalire alla metà del XIX secolo, un periodo segnato da un'intensificazione nell'accumulo e nell'analisi dei dati. Questa fu innescata da un bisogno crescente di affrontare e comprendere le complessità emergenti dalla rivoluzione industriale, che trasformò radicalmente le società in termini di urbanizzazione, economia e struttura sociale. Durante questo periodo, la raccolta di dati assunse una nuova urgenza e finalità differenti, diventando un pilastro per le decisioni governative, commerciali e scientifiche. Ovviamente, il senso in cui si può parlare di big data prima della digitalizzazione non è lo stesso di quello odierno. Infatti, questa prima ondata non è caratterizzata tanto dal volume dei dati (se paragonati ai volumi attuali), quanto dal cambiamento avvenuto nel modo di percepire

e utilizzare le informazioni. Tuttavia, come sottolineato da Mackenzie (2012), la moderna dipendenza da sistemi di database e da tecniche di smistamento e classificazione è radicata in queste pratiche storicamente definite e maturate nel contesto della rivoluzione industriale. Tale legame genealogico tra le prime raccolte di dati e i moderni big data evidenzia una continuità nel desiderio umano di catalogare, comprendere e influenzare il mondo attraverso i dati.

Se la prima ondata di big data del XIX secolo si concentra principalmente sulle procedure statistiche tradizionali, la seconda del XXI secolo si distingue, come anticipato, per la sua dipendenza dalle tecnologie digitali avanzate e dalle modalità di gestione dei dati che ne conseguono (Ambrose, 2015). Questa nuova era dei big data si caratterizza per la capacità di raccogliere, archiviare e analizzare non solo dati quantitativi o numerici, ma anche contenuti di natura qualitativa come testi, immagini, video e audio. La rivoluzione digitale ha ampliato significativamente la portata e la varietà dei dati disponibili, grazie all'incremento della larghezza di banda delle telecomunicazioni, alla connessione tra sistemi di archiviazione dati che facilitano l'accumulo massiccio di informazioni e a server farm supportati da capacità computazionali avanzate (Hilbert, 2016; Kitchin, 2014a).

È quindi fondamentale riconoscere che i big data non sono mai veramente grezzi; richiedono un processo di acquisizione e traduzione in formati quantificabili e leggibili dalle macchine, per poi essere elaborati attraverso algoritmi che dipendono da infrastrutture hardware specifiche (Bowker, 2008). Questo processo evidenzia come i big data, nella loro essenza, siano costruiti attraverso una serie di interventi tecnici e interpretativi. Questo ci obbliga a considerare i big data non solo in termini tecnologici ma anche come espressione di specifici modi di pensare e di agire nello sviluppo di infrastrutture e pratiche analitiche (Beer, 2016). La sfida sta nel definire chiaramente i big data, visto che, come osservato da Kitchin e McArdle (2015), molti sistemi descritti

come tali spesso non soddisfano pienamente i criteri stabiliti dalla letteratura accademica. Questa ambiguità ha reso il concetto di big data meno rigido nella sua inquadratura e definizione ontologica, spesso percepito come un'entità amorfa priva di chiarezza concettuale (Williamson, 2017).

In conclusione, la seconda ondata di big data del XXI secolo si connota per la sua enfasi sulle tecnologie digitali e sulle capacità di analisi avanzate, riflettendo un cambiamento significativo nelle modalità di raccolta, archiviazione e interpretazione dei dati rispetto alla precedente era statistica. Questa evoluzione evidenzia l'importanza di considerare i big data non solo come collezioni di informazioni ma come fenomeni complessi che intersecano tecnologia, società e pensiero.

1.2.2 *Big data: definizioni e caratteristiche*

“Ne parlano tutti, è vero, eppure non vi è una definizione univoca dei *big data*. Secondo l'Unione europea i megadati sono grandi quantità di tipi diversi di dati prodotti da varie fonti, fra cui persone, macchine e sensori” (Del Mastro, Nicita, 2019, p. 10). Infatti, nonostante il termine big data sia diventato popolare negli ultimi anni, definire precisamente cosa siano rimane un compito arduo. Rinaldi (2019, p. 9) è stato tentato di rispondere: “le nostre identità digitali”. Altri studiosi (Kitchin, McArdle, 2016) hanno indicato addirittura ventisei possibilità di definizione e si è andati ben oltre (Leonelli, 2018, p. 20). In effetti, il punto che mette tutti d'accordo è la quantità iperbolica: già dieci anni fa Mayer-Schönenberger e Cukier (2013, pp. 19-20) scrivevano che, se i dati “fossero tutti racchiusi in libri cartacei, coprirebbero l'intera superficie degli Stati Uniti 52 volte. Se venissero raccolti su dei CD-ROM e messi uno sull'altro, arriverebbero alla Luna in 5 pile separate”. Comunque sia, è indubbio che i big data non sono semplicemente una realtà, bensì sono strettamente legati allo svi-

luppo della tecnologia informatica (Holmes, 2017). Infatti, il termine big data con tutta evidenza fa riferimento a un concetto di grandi volumi di dati raccolti in maniera statistico-informatica e caratterizzati da elevata velocità e varietà (Laney, 2001). Si tratta di informazioni eterogenee che vengono generate continuamente da una moltitudine di fonti diverse. Questi dati sono distinti per la loro granularità fine, la possibilità di essere combinati con altri set di dati, e la loro flessibilità e scalabilità (Boyd, Crawford, 2013; Kitchin, 2014a, 2014b; Mayer-Schönberger, Cukier, 2013). In merito al concetto di granularità, è interessante notare come la rivoluzione dei big data non sia da attribuirsi esclusivamente al volume dei dati ma anche alla loro granularità che consiste nel loro livello di dettaglio (ridurre in granelli). Più la granularità è fine, più il dato è ridotto a livelli minimi e più cresce la capacità di analisi. Il grado si ritiene raggiunto nel momento in cui il livello di granularità di un sistema è ridotto ai minimi termini. Kitchin e McArdle (2015) sottolineano l'eshaustività e la velocità come le caratteristiche distintive dei sistemi di big data, contrapposte ai metodi tradizionali di raccolta e analisi dei dati che si basavano su campionamenti limitati per portata, varietà, temporalità e dimensione. Ciò che rende i dati particolarmente grandi – per così dire – e utili è la possibilità di incrociare e analizzare dataset per scoprire incroci e pattern difficilmente individuabili attraverso metodi tradizionali. Non a caso quella che Holmes definisce come “esplosione di dati”, si riferisce alle quantità sempre più grandi di dati strutturati, non strutturati e semistrutturati generati minuto dopo minuto grazie allo sviluppo del sistema Web, lanciato dall'informatico inglese Berners-Lee e descritto dallo stesso inventore come “un sistema di informazione globale”⁵ tramite l'accesso libero ad Internet da parte di tutti gli utenti connessi.

5 Executive summary del 6 agosto 1991 di Berners-Lee consultabile su: <https://www.historyofinformation.com/detail.php?id=1032>

Ma da dove provengono questi dati? “Un cellulare in ogni tasca, un computer in ogni zaino e sistemi informativi negli uffici di tutto il mondo” (Mayer-Schönberger, Cukier, 2013, p. 15): sono queste le fonti dei dati (grezzi). Mezzo secolo dopo l’introduzione del computer nella nostra società, i dati hanno cominciato ad accumularsi in misura tale da creare un fenomeno nuovo e particolare: il mondo è stato subissato di dati come mai prima d’ora e queste informazioni crescono sempre più rapidamente. Secondo uno studio condotto da IDC (acronimo di Internet Database Connector), entro il 2025 il mondo genererà 163 zettabyte⁶ di dati (Reinsel, Gantz, Rydning, 2018). È noto che questo accumulo quantitativo di dati, in continua crescita, sta occupando tutti i campi dell’attività umana e, con la rapidità che caratterizza la loro natura, ne sta rivoluzionando gli scenari futuri in virtù della loro “capacità di esplorare, aggregare e relazionare vasti insiemi di dati” (Boyd, Crawford, 2012, citato in Leonelli, 2018, p. 23). Il nostro appare effettivamente come un “universo digitale in espansione” (Mayer-Schönenberger, Cukier, 2013, p. 17), anche se Leonelli parla di una società dell’informazione datocentrica, che pone anche la necessità di interrogarsi “su ciò che costituisce conoscenza affidabile” (Leonelli, 2018, p. 11). Da qui la constatazione che la quantità di informazioni per i big data sia meno rilevante dell’“uso che se ne fa” o, meglio, dell’uso che ne fanno “coloro che sono incaricati di raccogliere, analizzare ed elaborare questi dati per individuare ciò che è veramente importante per l’individuo e la collettività” (Ispas, 2019, p. 41).

Così come la precisa definizione dei big data rimane un compito elusivo che ha visto il susseguirsi di contributi diversi anche per la descrizione delle caratteristiche i contributi degli studiosi sono stati molteplici.

Per quanto riguarda le caratteristiche fondamentali trattate

6 Uno zettabyte è pari a un miliardo di gigabyte.

nella letteratura sui big data, esse risalgono al famoso modello delle 3 V (Volume, Velocità e Varietà) teorizzato da Laney che, nel 2001, ne ha reso ufficiale la formulazione⁷. Tale modello è stato poi arricchito con altre variabili quali: Valore (AGCM, 2018; Del Mastro, Nicita, 2019), Veridicità e Valore (Rinaldi, 2019), Visualizzazione (AGCOM, 2018), Volatilità e Mobilità (Leonelli, 2018). A cosa si riferiscono queste caratteristiche? Vediamolo analiticamente nel seguito:

- Volume: connota “la dimensione dei file usati per archiviare e disseminare i dati, che grazie al potere crescente dei processori elettronici sta aumentando vertiginosamente” (Leonelli, 2018, p. 19). Si pensi ad esempio alle piattaforme di social media, come Meta o X, che gestiscono quotidianamente petabyte⁸ di dati derivanti da post, immagini, video e interazioni degli utenti. Questi dati sono utilizzati non solo per personalizzare l’esperienza dell’utente e per il targeting pubblicitario, ma anche per analisi di tendenze globali e comportamenti di massa.
- Velocità: “è necessariamente connessa al volume: più velocemente vengono generati i dati, più crescono” (Holmes, 2017, p. 41) e concerne sia il ritmo dell’elaborazione elettronica che la trasmissione delle informazioni. Nel mondo finanziario, per esempio, i dati sulle transazioni vengono generati e analizzati in frazioni di secondo. Le piattaforme di trading ad alta fre-

7 L’origine del termine non è di chiara provenienza. In una breve (e anonima) “Storia dei Big Data” si legge: “Il termine Big data viene utilizzato dall’inizio degli anni Novanta. Sebbene non si sappia esattamente chi ha utilizzato il termine per la prima volta, molti attribuiscono a John R. Mashey (all’epoca impiegato presso la Silison Grafics) il merito di aver reso il termine popolare [...]. Il Big data è ormai un ambito di conoscenze ben riconosciute, sia nel mondo universitario che in quello dell’impresa”.

8 Un petabyte è pari a un milione di gigabyte.

quenza utilizzano questi dati per eseguire ordini basati su minimi movimenti di prezzo, il tutto avvenendo in tempi così brevi da poter essere misurati in millisecondi. La capacità di processare rapidamente queste informazioni è cruciale per il successo nel trading ad alta velocità.

- **Varietà:** riguarda la disponibilità dei dati in formati molteplici, tra i quali dati strutturati tradizionali (anagrafe, banche...), semi-strutturati e non strutturati come audio, video, fotografie, e-mail, pagine web e testi. Nell'ambito del commercio al dettaglio, per esempio, le aziende raccolgono una vasta gamma di dati che vanno dalle informazioni strutturate dei sistemi di gestione degli inventari, ai feedback dei clienti semi-strutturati sui social media, fino ai dati non strutturati come le immagini e i video delle recensioni dei prodotti. La gestione e l'analisi di questa varietà di dati sono fondamentali per comprendere le preferenze dei consumatori e per personalizzare le offerte, aumentando così l'efficacia delle campagne di marketing.
- **Veridicità:** determina l'aspetto qualitativo dei dati raccolti e filtrati e valutati in termini di accuratezza, affidabilità, significatività e credibilità. Non sempre i dati reperiti sono veritieri né possono essere verificati, andando a inquinare le nostre analisi. Pertanto, la veridicità costituisce un fattore determinante per conoscere come sono strutturate e come possono essere immagazzinate simili moli di dati.
- **Valore:** è la qualità che assumono i dati una volta elaborati e analizzati. In ambito commerciale questo permette, ad esempio, la loro vendita – cioè, l'attribuzione di un valore monetario – ad aziende che poi li elaborano ulteriormente per i propri fini. In campo educativo a questo elemento è stata attribuita una particolare attenzione. In una interessante “Ricerca internazionale tra studenti universitari” (2019) professori e ricercatori hanno sondato la consapevolezza dell'importanza che gli studenti universitari attribuiscono ai propri dati in rete. Dalle singole valutazioni, tutte degne di nota, emerge la con-

ferma “che per la maggior parte degli studenti universitari intervistati i propri dati non hanno particolare valore”. Da qui l’auspicio di “una educazione al valore del dato che tarda a venire” (Rinaldi, 2019, pp. 9-11).

- Visualizzazione: riguarda grafici e schemi destinati a presentare e a comunicare in maniera visuale i dati per renderli facilmente interpretabili e determinabili mediante “grafici e diagrammi statistici, spazializzazioni, mappe e animazioni” (Kitchin 2014a, p. 106). Possono diventare a loro volta oggetto di “analisi visiva” in vista della determinazione delle scelte che le persone devono compiere (Mackenzie 2015, p. 437). Ad esempio, la visualizzazione dei dati è fondamentale nelle analisi di mercato, dove i grafici sono impiegati per identificare e presentare tendenze di consumo, aiutando le aziende a prendere decisioni informate sulla base delle preferenze dei clienti.
- Volatilità: questa caratteristica va messa in relazione con la stabilità delle informazioni nel tempo. Ovvero con “la capacità dei dati di rimanere affidabili e leggibili nonostante l’evoluzione di nuove tecnologie di archiviazione” (Leonelli, 2018, p. 22).
- Mobilità: Leonelli aggiunge un’ulteriore caratteristica ai big data – molto interessante e al di fuori della lettera V – la Mobilità che descrive come conseguenza della “capacità dei dati di viaggiare attraverso diverse situazioni di analisi e riuso e di essere relazionati con quanti più tipi di dati diversi possibili. Senza mobilità non ci sarebbero big data” (ivi, p. 31).

La figura che segue include, schematizzandole, le sei “V” dei big data (figura 1.1):

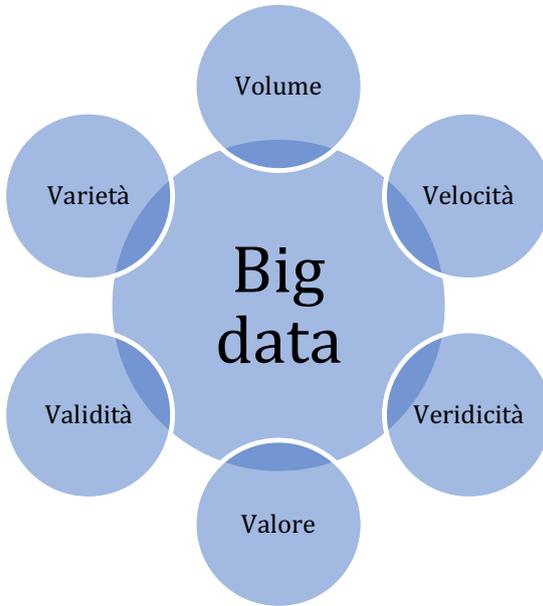


Fig. 1.1: Le sei "V" dei big data

Da tutto questo si deduce che i big data non hanno più molto in comune con le rilevazioni di ordine matematico-statistico in uso prima della diffusione del computer, o nella fase della prima ondata dei big data, sebbene ne siano l'evoluzione diretta e rappresentino ormai una risorsa per l'economia e per nuove conoscenze, tramite "correlazioni di natura (per lo più) probabilistica, di tendenze o modelli" (AGCM, 2018, p. 7). Come è stato notato (Rinaldi, 2019), i big data si servono della massa delle nostre tracce come oggetto. "Come persone, stiamo diventando i nostri stessi dati", diceva Bowker già nel 2013 (Bowker, 2013, p. 168). Sono il nostro doppio virtuale che ci registra, esamina e processa attraverso le transazioni delle nostre carte fedeltà e di credito, dei

nostri pagamenti che tengono il conto degli acquisti, del loro costo e valutano i nostri gusti. Ha informazioni su di noi attraverso i siti web, il Wi-Fi, l'uso di applicazioni dei nostri smartphone, la gestione di oggetti sulla rete, l'utilizzo di social network, per captare informazioni sui nostri comportamenti personali, e orientarsi su quali prodotti e servizi è più probabile che verteranno le future scelte di acquisti, richieste di servizi, letture, tempo libero. In quanto alla loro tipologia (si pensi alla V di Varietà), i big data non sono solo dati numerici – ovvero strutturati (suscettibili di analisi statistica, come quelli di cui ci serviamo per fare un bonifico o per iscriversi a un servizio fornendo dati anagrafici) – bensì includono spesso foto e video dei social media (si pensi a YouTube, Instagram e TikTok), messaggi di testo, e-mail, file di registro, conversazioni telefoniche, trascrizioni di riunioni, video di sorveglianza, di telecamere stradali, strumenti di geolocalizzazione, ricerche su Google, conversazioni di chatbot e molto altro ancora.

Parallelamente alla raccolta di vasti volumi di dati personali e commerciali, l'importanza degli *open data*,⁹ ovvero dati resi liberamente accessibili e utilizzabili da chiunque senza restrizioni, assume un ruolo sempre più centrale (Chignard, 2013). Quali sono allora le differenze e complementarità tra le due locuzioni? Mentre i big data tendono ad essere determinati dalla loro portata, gli *open data* sono definiti dal loro utilizzo (Raffaghelli, 2017). La locuzione *open data* (registrata nel 2013 nel Lessico del XXI secolo della Treccani) fa parte di un insieme di concetti correlati come: *open access* (accesso libero), *open science* (scienza aperta), *open content* (contenuto aperto), *open source* (fonte aperta). Ad esempio,

9 L'*open data handbook* ne dà questa definizione: “I dati aperti sono dati che possono essere liberamente utilizzati e ridistribuiti da chiunque, soggetti eventualmente alla necessità di citarne la fonte e di condividerli con lo stesso tipo di licenza con cui sono stati originariamente rilasciati” (fonte: <https://opendatahandbook.org/>).

Wikipedia è un autentico esempio di *open content*. Gli *open data* rappresentano quindi una risorsa fondamentale per l'innovazione e la trasparenza, permettendo a ricercatori, sviluppatori e cittadini di accedere a informazioni pubbliche che possono essere utilizzate per una varietà di applicazioni, dalla creazione di app di servizi pubblici all'analisi di tendenze economiche e sociali. Questa apertura non solo facilita nuove scoperte e servizi, ma promuove anche una maggiore responsabilità e partecipazione civica, rafforzando il legame tra tecnologia e società.

1.2.3 *La filiera dei big data*

Come vengono raccolti, archiviati, memorizzati e decifrati i big data? Dove sono fisicamente collocati? Una risposta sul tema dell'archiviazione dei dati la propone Leonelli (2018, pp. 33-34) quando, rilevandone l'importanza e sottolineando che gli "archivi non possono essere semplici contenitori (*data dumps*)", ricorda come tanti filosofi del Ventesimo secolo, tra cui Michel Foucault e Jacques Derrida, abbiano riconosciuto che "la chiave per la gestione di un archivio è il modo in cui si ordina". Infatti, come viene osservato, l'assetto della schedatura (classificazione) dei dati, dei modelli e algoritmi di visualizzazione è determinante ai fini dell'interpretazione e del riutilizzo dei dati. Tutto questo nell'ambito dei big data prende la forma di una grande banca dati digitale, ad opera di ingegneri dell'elettronica ed esperti in informatica in grado di creare le tecnologie (*software e hardware*) che determinano il funzionamento delle banche dati che usiamo nella consultazione via Internet. Naturalmente la parte archivistica è affidata ad esperti in formattazione dei dati stessi per consentirne la visualizzazione da parte degli utenti.

In quanto alla raccolta dei dati, esiste un testo di riferimento importante: di raccolta, elaborazione e interpretazione hanno trattato, infatti, gli autori di un prezioso report ministeriale (basato

su sondaggi e audizioni dirette), dove vengono descritti i risultati dell'“Indagine conoscitiva sui big data” (AGCM, 2018), da cui emerge la complessa filiera di quello che viene indicato come piano logico del processo di estrazione di conoscenza dai big data (figura 1.2). Non sarà possibile descrivere estensivamente ogni passaggio di tale processo nelle sue molteplici scansioni e sottocategorie; vale, però, la pena di tentare di riportare alcuni snodi di questo documento visualizzabili all'interno della figura seguente.

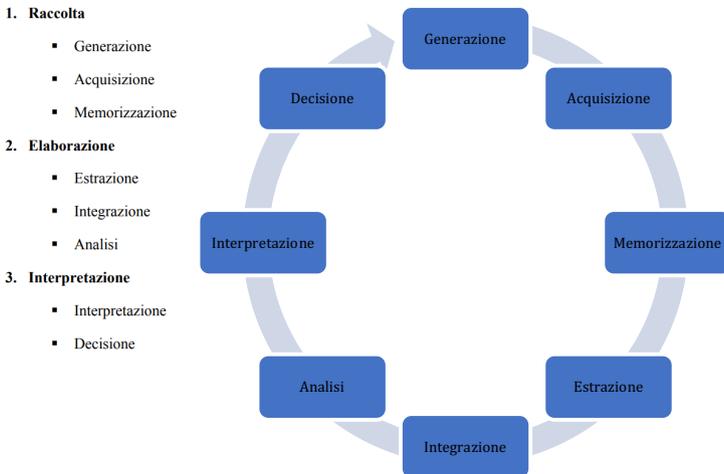


Fig. 1.2: La filiera dei big data (adattato da AGCM, 2018)

A sinistra sono elencati i tre ordini principali di attività (raccolta, elaborazione, interpretazione con lo spaccettamento delle rispettive sottosequenze per ogni attività, otto complessivamente); a destra, il diagramma circolare non presenta tagli di valore proporzionale, bensì frecce che riproducono le attività e le sotto-attività programmatiche, seguendo un movimento rotatorio, consequenziale. Il grafico agisce come modo per semplificare, ri-

durre e lasciare traccia dell'enorme ricerca che non appare, ma sta dietro ad ogni click di Internet, senza distinzione tra dati personali (oggetto del Regolamento generale della protezione dei dati emanato dalla Commissione Europea e adottato nel 2018) e non (come informazioni sul meteo, sui luoghi, sull'economia), e tenendo conto che ogni fenomeno può fornire una enorme ed eterogenea mole di dati.

Esaminiamo ora più dettagliatamente la fase iniziale della filiera, che è rappresentata dal processo di raccolta dei dati. Nella prima sotto-fase, detta di *generazione*, i dati vengono prodotti in un contesto fortemente digitalizzato, dove ogni interazione online tra gli utenti o con dispositivi intelligenti contribuisce alla generazione di dati. Questo coinvolge l'uso dei social media, la navigazione su Internet, le comunicazioni via e-mail, l'interazione con app mobili e l'utilizzo di dispositivi connessi nell'ambito dell'Internet delle Cose (IoT, acronimo di Internet of Things¹⁰). Ad esempio, i dispositivi indossabili (*wearable*) monitorano parametri biologici e attività fisica, mentre i sensori IoT raccolgono dati su vari aspetti dell'ambiente e delle macchine. Questa sotto-fase è caratterizzata dalla diversità e dalla vastità dei dati generati, che riflettono una gamma di comportamenti, preferenze e attività degli utenti, sia online che offline.

Successivamente, nella sotto-fase di *acquisizione*, i dati generati devono essere raccolti per essere elaborati e analizzati. Questo processo implica la raccolta dei dati da varie fonti, come dispositivi mobili, sensori e piattaforme online. L'acquisizione dei dati richiede meccanismi per raccogliere efficacemente le informazioni mantenendo al contempo attenzione alla privacy e alla sicurezza degli utenti. Ciò implica ottenere il consenso per l'uso di dati personali, utilizzare cookie per tracciare le preferenze degli utenti on-

10 L'Internet of Things (IoT) è una rete di oggetti e dispositivi connessi – detti “cose” – dotati di sensori che consentono loro di trasmettere e ricevere dati, da e verso altre “cose”, sistemi e altri strumenti digitali.

line e implementare sistemi di tracciamento che permettano una raccolta dati quanto più precisa. I dati acquisiti possono poi essere analizzati per migliorare i servizi o per personalizzare l'esperienza utente. La sfida in questa fase risiede nel bilanciare l'acquisizione e l'uso dei dati con la necessità di rispettare la privacy degli utenti e le normative sulla protezione dei dati.

Infine, la sotto-fase di *memorizzazione* comporta il trasferimento dei dati raccolti verso sistemi di elaborazione. Dato l'enorme volume di dati generati e acquisiti, è fondamentale disporre di infrastrutture di memorizzazione che siano non solo capienti ma anche sicure e in grado di garantire l'integrità dei dati. Questo richiede l'uso di database avanzati, data warehouse e soluzioni di cloud storage che possano incrementare le proprie prestazioni (in gergo "scalare") in base alle esigenze. La tecnologia ha fatto grandi passi avanti nel ridurre i costi e nel migliorare l'efficienza delle soluzioni di memorizzazione, facilitando l'accesso a sistemi ad alta capacità e a bassa latenza. La gestione efficiente della memorizzazione è cruciale per assicurare che i dati siano non solo conservati in modo sicuro, ma anche facilmente accessibili per l'analisi e l'elaborazione.

Una volta ultimata la fase di *raccolta*, il passaggio dai dati, già organizzati ma ancora grezzi e non strutturati, alla possibilità di fornire informazioni pratiche, utili in materia di "data economy" (Delmastro, Nicita, 2019, p. 14), ci conduce alla seconda fase, quella elaborativa, che comporta tre ulteriori passaggi: quello dell'*estrazione* dei dati dalle fonti, accompagnato da un ulteriore passaggio di raccolta di informazioni, detto di *integrazione*, e seguito dal vero e proprio campo dell'*analisi* dei dati.

La procedura di *estrazione* inizia con l'identificazione e la raccolta dei dati dalle molteplici fonti disponibili, che possono includere database interni, sistemi IoT, piattaforme online e social media. Questo processo comporta il reperimento dei dati, la loro selezione e il successivo caricamento nella memoria di un sistema di elaborazione. L'obiettivo è quello di raccogliere un insieme di

dati il più completo e variegato possibile, in modo da avere una base solida su cui costruire le analisi successive.

Segue l'*integrazione*, durante la quale i dati estratti vengono combinati e armonizzati per creare un dataset unificato e coerente. Questo passaggio è fondamentale per garantire che le informazioni riferite agli stessi elementi o domini applicativi siano integrate in modo da formare una visione olistica. L'integrazione consente di superare le disparità tra diverse fonti di dati, risolvendo eventuali incongruenze e duplicazioni, e facilita l'aggregazione di dati eterogenei in formati strutturati o semi-strutturati. L'obiettivo è di preparare i dati per un'analisi efficace, garantendo che siano puliti, completi e pronti per essere esaminati in profondità.

Infine, la sotto-fase di *analisi* rappresenta il cuore dell'elaborazione dei big data, dove algoritmi avanzati e tecniche di intelligenza artificiale, inclusi quelli di apprendimento automatico¹¹, sono impiegati per esplorare i dati integrati. In questa procedura, si utilizzano metodi statistici, algoritmi di apprendimento automatico per identificare pattern, trend, correlazioni e anomalie nei dati. L'analisi può essere diretta a rispondere a specifiche domande (attraverso l'uso di algoritmi di interrogazione) o a scoprire nuova conoscenza senza una direzione predeterminata, sfruttando la capacità degli algoritmi di apprendimento di estrarre *insight* e generare previsioni basate sui dati. Questo processo non solo fornisce interpretazioni e risposte basate sui dati esistenti, ma apre anche la strada all'identificazione di nuove opportunità, alla ge-

11 L'apprendimento automatico (in inglese *machine learning*) è una branca dell'intelligenza artificiale che utilizza metodi statistici per migliorare la performance di un algoritmo nell'identificare pattern nei dati. Nell'ambito dell'informatica, l'apprendimento automatico è una variante alla programmazione tradizionale nella quale in una macchina si predispone l'abilità di apprendere qualcosa dai dati in maniera autonoma, senza istruzioni esplicite.

nerazione di modelli predittivi e alla presa di decisioni informate basate su analisi complesse (Panciroli, Macaudo, 2020).

La terza e ultima fase prevista dalla filiera dei big data consiste nell'*interpretazione* dei dati, precedentemente raccolti ed elaborati, che anticipa e informa il processo decisionale a conclusione dello schema circolare rappresentato nel grafico. Qui giocano un ruolo rilevante gli algoritmi di intelligenza artificiale che possono analizzare i dati per identificare correlazioni complesse e sconosciute ai ricercatori. Inoltre, gli algoritmi possono continuare a monitorare le decisioni prese per migliorarle in futuro o utilizzarle come base per nuove decisioni in campi diversi. Questo approccio data driven al decision making può essere applicato in diversi ambiti per aumentare l'innovazione e supportare processi decisionali complessi in ambito scientifico, economico e sociale. Tuttavia, i modelli algoritmici attualmente in uso spesso non sono condivisi apertamente dai loro sviluppatori, persino quando producono risultati distorti. Quali prodotti di un contesto socioculturale, infatti, anche gli algoritmi possono incorporare pregiudizi (*bias*) in modo così radicato da renderli spesso impercettibili (O' Neil, 2016).

Anche il settore dell'istruzione partecipa a questa incessante tendenza di raccolta, elaborazione ed interpretazione di dati che si depositano come tracce digitali nell'"infosfera" che ci circonda (Floridi, 2017, p. 44). Studenti, insegnanti, scuole e università producono ogni giorno una mole di dati senza precedenti, utilizzando piattaforme e-learning, applicazioni educative, sistemi di gestione dell'apprendimento (learning management system o LMS), social media, sensori IoT. Si stima che il solo mercato edutech genererà 1,5 zettabyte di dati, dei 163 totali, entro il 2025 (Liu, Sun, Wang, Wu, 2020). Attraverso una combinazione di strumenti di analytics, intelligenza artificiale e visualizzazione dei dati, è infatti possibile estrarre spunti preziosi da questo oceano di informazioni nell'intento di produrre cambiamenti non solo quantitativi bensì qualitativi. Nel paragrafo seguente, passeremo

al vaglio le potenzialità e i limiti legati all'utilizzo dei big data in ambito educativo. Verranno quindi proposte applicazioni concrete finalizzate a potenziare la didattica, ottimizzare le risorse, guidare il processo decisionale e creare modelli predittivi a supporto di studenti e docenti. Esaminando casi di studio ed esempi reali, dimostreremo come i big data possano influire sul mondo dell'educazione, aprendo la strada a sistemi formativi più efficaci, inclusivi e centrati sul discente. Ciò non toglie che continueremo a mantenere uno sguardo critico su tali pratiche, tenendo conto che la "rivoluzione dei big data" può comportare anche dei rischi, soprattutto in tema di privacy e di atteggiamenti fideistici nei confronti dei numeri, qualora dovesse prevalere, invece di scenari ottimistici, il lato oscuro rappresentato da situazioni distopiche simili a quelle descritte da Orwell nel suo romanzo "1984" (Mayer-Schönenberger, Cukier, 2013, pp. 206-220). Allo stesso modo, dobbiamo affrontare le preoccupazioni riguardanti le possibili distorsioni che minacciano la ricerca scientifica (Leonelli, 2018) e, perfino, i paventati timori di una dittatura dei dati (Kaiser, 2019).

1.3 Usi e applicazioni educative dei big data

1.3.1 *Lo sviluppo dei big data nel campo educativo*

La raccolta, il calcolo e la comunicazione di dati su larga scala nell'ambito dell'educazione hanno una lunga storia (Williamson, 2017). Già nel corso del XIX secolo, sia in contesti europei che americani, le grandi esposizioni universali rappresentavano un palcoscenico comparativo emblematico tanto per i settori produttivi e le economie nazionali quanto per un raffronto tra arti, culture, idee, mode e sistemi educativi: in sintesi, un confronto tra civiltà. Questi eventi erano occasioni per narrare i progressi nazionali in termini di evoluzione scientifica e tecnologica, e of-

frire i propri successi culturali, artistici e educativi come esempi per altre realtà, trasformandosi in efficaci strumenti di politica.

Nell'era post-bellica, l'emergere di un tessuto internazionale di ricerca ha ulteriormente integrato l'uso dei dati nella gestione dell'educazione. La complessità della raccolta di informazioni, che si avvaleva di reti di varia natura, richiedeva una serie di competenze tecniche per il trattamento e l'analisi dei dati. In questo contesto, l'educazione, intesa come istituzione sociale, ha perseguito la quantificazione e la divulgazione dei propri numeri come strumenti di chiarimento e persuasione, un trend oggi accelerato dalle nuove tecnologie digitali e dalla crescente tendenza alla datificazione (Grosvenor, Roberts, 2013; Lawn, 2013; Sobe, 2013).

Nell'era contemporanea, l'enfasi sulla raccolta di dati nell'ambito educativo ha portato alla concezione dell'educazione come dominio "governato dai numeri", caratterizzato da incessanti valutazioni e confronti anche a livello internazionale (Grek, 2009; Lawn, Grek, 2012; Ozga, Dahler-Larsen, Segerholm, Simola, 2011). Questa tendenza ha trasformato i dati numerici in strumenti politici, con organizzazioni internazionali che esercitano un'influenza sempre maggiore sulla definizione delle politiche educative a livello mondiale. La necessità di una complessa rete di dati emerge per coordinare l'ampio spettro di attori coinvolti nella governance dell'educazione, sottolineando l'importanza dei dati per il controllo del sistema (Ozga, 2009). In tal senso, l'OCSE si distingue come uno degli attori più influenti, attraverso la gestione di programmi come PISA e PIAAC e studi sulla prima infanzia che valutano le competenze dei bambini di 5 anni (Moss, Mangez, Ozga, 2016). Queste indagini hanno un impatto significativo sulle politiche educative nazionali, generando dati comparativi che spesso guidano verso un allineamento dei curricula ai requisiti dei test (Morris, 2016). Tuttavia, c'è il rischio di un ciclo chiuso di miglioramenti apparenti, in cui le politiche si concentrano esclusivamente sulle prestazioni nei test a scapito di una vera e propria innovazione educativa. Il risultato di questa com-

petizione globale tra i sistemi educativi nazionali, innescata dai programmi di valutazione sopra citati, ha portato all'emergere di gerarchie internazionali, trasformando i sistemi educativi di paesi come Singapore e Finlandia in modelli da seguire, oggetto dell'ammirazione e dell'imitazione da parte di altre nazioni (Rizvi, Lingard, 2010). Queste ampie operazioni di raccolta e analisi dati mirano a creare una cosiddetta conoscenza di governo, un corpus informativo strategico in ambito educativo volto a guidare le decisioni politiche (Fenwick, Mangez, Ozga, 2014). Tuttavia, l'introduzione dei big data e delle tecnologie avanzate, come l'intelligenza artificiale e l'apprendimento automatico, ha ulteriormente ampliato l'uso dei dati nel contesto educativo. Questi strumenti, ormai parte integrante non solo della nostra vita quotidiana attraverso social media, smartphone e web, ma anche dei sistemi educativi, offrono nuove opportunità per monitorare e migliorare i processi di apprendimento. Lungi dall'essere elementi puramente tecnici, questi avanzamenti sono saldamente ancorati al tessuto sociale ed economico, evidenziando come il campo educativo stia diventando un terreno fecondo per l'applicazione e l'evoluzione del concetto di big data, che emergono come un concetto sociale potente, il cui impatto va ben oltre il semplice accumulo di vasti volumi di informazioni. I big data sono inestricabilmente legati a programmi software, algoritmi e analisi complesse necessarie per la loro raccolta e gestione, le quali richiedono competenze specializzate. Ci sono persone dietro i big data: non solo data scientist o data analyst, ma sviluppatori di software e progettisti di algoritmi, nonché attori politici, scientifici ed economici che cercano di sviluppare e utilizzare sistemi di big data per scopi diversi. E i big data riguardano anche le persone che li compongono, le cui vite sono registrate sia individualmente che su vasta scala come aggregati di popolazione. I big data, in altre parole, sono allo stesso tempo tecnici e sociali: tecnici in quanto sono il prodotto di programmi e processi software; sociali perché prodotti e utilizzati da operatori che lavorano in contesti

organizzativi specifici e sono generati dalla vita quotidiana di persone in tutto il mondo. In quanto fonte di conoscenza, i big data hanno anche il potere di modificare cosa sappiamo della società, delle persone e delle istituzioni che la compongono. Come si è visto, una migliore conoscenza basata sui dati potrebbe quindi essere utilizzata per catalizzare innovazioni negli ambiti più diversi: dagli affari all'intrattenimento, dal governo ai servizi pubblici. Anche il campo dell'istruzione è emerso come un luogo chiave per la produzione di visioni del futuro basate sui dati. Si tratta, insomma, della prosecuzione e di un'espansione della tradizione storica di utilizzo dei dati nell'educazione, proiettata verso nuove frontiere che promettono di migliorare ulteriormente il modo in cui concepiamo e implementiamo le pratiche educative.

Le applicazioni e i servizi che elaborano i big data possono aiutare gli studenti nelle scuole e nelle università, fornendo feedback sui loro progressi rilevati e consigli su cosa fare per migliorare. Possono aiutare gli insegnanti a rivedere e valutare i corsi, poiché possono monitorare il coinvolgimento e i risultati degli studenti e consentire ai dirigenti scolastici e universitari di rivedere e valutare allo stesso tempo le prestazioni istituzionali e del personale (Ranieri, Cuomo, Biagini, 2024).

Per i sostenitori dell'utilizzo dei big data in ambito educativo, questi potrebbero essere utili come *insight* da utilizzare nella progettazione di nuovi corsi, nuove risorse, nuove politiche e nuove pratiche. Potrebbero anche diventare disponibili applicativi per l'apprendimento basati sull'intelligenza artificiale in grado di fungere da assistenti digitali sia per insegnanti che per studenti, capaci di fornire in tempo reale feedback per migliorare le attività educative o forse anche per automatizzarle. I big data nell'istruzione sono tanto una risorsa da sfruttare per le possibilità future quanto una realtà tecnica emergente. Infatti, immaginare il futuro dell'istruzione con i big data sta catalizzando sviluppi reali destinati ad avere un impatto sui processi educativi in tutto il mondo (Williamson, 2017).

1.3.2 *Le fonti dei big data nell'educazione*

I big data in ambito educativo provengono anch'essi da una molteplicità di fonti, interne ed esterne agli istituti di istruzione, che consentono di raccogliere e analizzare grandi volumi di dati su studenti, docenti, processi formativi e risultati di apprendimento (Daniel, 2015). Una prima fonte è costituita dai dati amministrativi. Essi costituiscono la colonna portante della raccolta di dati in ambito educativo, fornendo informazioni essenziali su iscrizioni, voti, presenze e dati demografici degli studenti (età, genere, ecc.), sui loro percorsi scolastici pregressi (voti, corsi seguiti, titoli acquisiti), oltre che informazioni sul background familiare (status socioeconomico). Questi dati vengono raccolti dai sistemi di gestione scolastica e universitaria, permettendo un'analisi longitudinale del progresso degli studenti e della loro performance accademica. L'analisi di queste informazioni può rivelare tendenze importanti nella frequenza e nel rendimento, contribuendo alla progettazione di interventi mirati e alla distribuzione efficace delle risorse; possono, inoltre, servire per predirne le possibilità di successo accademico (Agasisti, Bowers, 2017). Tuttavia, vale la pena sottolineare come l'utilizzo di questi dati per prevedere il successo accademico sollevi questioni etiche significative: il rischio di etichettare prematuramente gli studenti può influenzare negativamente le loro opportunità e l'equità del sistema educativo (per una trattazione più esaustiva delle questioni aperte si rimanda al capitolo 4).

Una ulteriore fonte è rappresentata dai dati sull'apprendimento online. L'ascesa dell'e-learning e dei MOOC (Massive Online Open Courses) ha portato alla generazione di quantità massicce di dati, registrando ogni interazione degli studenti con i materiali didattici online. Questi dati, che includono, tra le altre cose, i click per l'accesso alle risorse, i tempi di permanenza su specifiche risorse, le risposte ai quiz e la partecipazione ai forum di discussione, offrono una panoramica dettagliata sull'engage-

ment e sul comportamento degli studenti. La loro analisi permette di personalizzare l'esperienza educativa, adattando i contenuti e i percorsi didattici alle esigenze individuali degli studenti (Williamson, 2017).

Altra fonte è quella dei social media e dei forum online che rappresentano una fonte di dati in rapida espansione, in grado di riflettere l'apprendimento informale e le interazioni sociali degli studenti. L'analisi dei post, dei commenti e delle discussioni su queste piattaforme può offrire intuizioni preziose sulle percezioni, le esperienze e le esigenze degli studenti (Gao, Luo, Zhang, 2012; Junco, Elavsky, Heiberger, 2013). Questi dati possono anche evidenziare le dinamiche delle comunità di apprendimento online, fornendo agli educatori informazioni utili per supportare e facilitare l'apprendimento collaborativo (Greenhow, Askari, 2017).

Poi vi sono i sensori e dispositivi indossabili la cui introduzione negli ambienti educativi apre nuove prospettive per la raccolta di dati. Questi dispositivi possono monitorare vari aspetti fisici e cognitivi degli studenti, come i livelli di attenzione e il battito cardiaco, offrendo nuove dimensioni dei dati sull'apprendimento e sul benessere degli studenti. Analizzando queste informazioni, sarà possibile comprendere meglio come gli ambienti fisici e le attività didattiche influenzino l'apprendimento e l'*engagement* degli studenti (Aljohani, Davis, 2016).

Importanti sono anche i dati provenienti da valutazioni formative e sommative che, insieme al feedback raccolto dagli studenti, risultano fondamentali per valutare l'efficacia dell'insegnamento e dell'apprendimento. Tali dati permettono ai docenti di identificare lacune nella comprensione e nelle competenze degli studenti, orientando la personalizzazione degli interventi didattici. Inoltre, il feedback degli studenti sui corsi e sui materiali didattici può guidare le revisioni dei curricula e il miglioramento delle pratiche pedagogiche (Lang, Siemens, Wise, Gašević, 2017).

Similmente, i dati dei questionari, tradizionalmente utilizzati per valutare atteggiamenti, opinioni e comportamenti degli stu-

denti, sono usati per la loro capacità di strutturazione e ampiezza di copertura, benché siano limitati da potenziali *bias* di risposta e dalla rigidità del loro design (Kays, Gathercoal, Buhrow, 2012).

Anche i dati delle applicazioni mobili emergono come fonti di *insight* comportamentali e preferenze degli utenti, captati attraverso dispositivi quali smartphone e smartwatch che le persone portano costantemente con sé (Couper, 2013). Questi dati possono permettere un elevato grado di personalizzazione e contestualizzazione dell'apprendimento, nonostante le preoccupazioni per la privacy e la dipendenza dall'utilizzo attivo dei dispositivi.

Infine, i dati di geolocalizzazione: essi si rivelano strumenti molto utili per analizzare i pattern spaziali e il movimento fisico, essenziali per studiare gli spazi di apprendimento e l'effetto della mobilità studentesca sull'esperienza educativa, con lo studio StudentLife del Dartmouth College che rappresenta un esempio pionieristico (Facer, 2014).

La tabella che segue schematizza caratteristiche, punti di forza, limitazioni e applicazioni delle principali fonti di dati educativi (Tabella 1.1).

Fonte di dati	Caratteristiche principali	Punti di forza	Limitazioni	Applicazioni
Dati amministrativi	Informazioni su iscrizioni, voti, presenze, dati demografici, percorsi scolastici e background familiare.	Forniscono una base per analisi longitudinali del progresso e della performance accademica.	Richiedono sistemi di gestione dati robusti e possono presentare sfide nella protezione dei dati personali.	Predire il successo accademico, progettare interventi mirati, distribuzione delle risorse

<p>Dati sull'apprendimento online</p>	<p>Dati generati dalle interazioni degli studenti con materiali didattici online, inclusi click, tempo di permanenza, risposte a quiz, partecipazione a forum.</p>	<p>Permettono la personalizzazione dell'esperienza educativa in base alle esigenze degli studenti.</p>	<p>Dipendono dalla partecipazione attiva degli studenti e possono sollevare questioni di privacy e accesso ai dati.</p>	<p>Personalizzazione dell'apprendimento, miglioramento dell'engagement e del comportamento di apprendimento degli studenti</p>
<p>Social media e forum online</p>	<p>Dati generati da post, commenti e discussioni degli studenti su piattaforme sociali e forum online, riflettenti l'apprendimento informale e le interazioni sociali.</p>	<p>Offrono intuizioni preziose sulle percezioni, esperienze e bisogni degli studenti.</p>	<p>Possono presentare sfide legate alla privacy e alla rappresentatività dei dati.</p>	<p>Studio delle dinamiche delle comunità di apprendimento online, facilitazione dell'apprendimento collaborativo</p>
<p>Sensori e dispositivi indossabili</p>	<p>Monitoraggio di aspetti fisici e cognitivi tramite sensori e dispositivi indossabili, inclusi livelli di attenzione, battito cardiaco e mobilità.</p>	<p>Offrono nuove dimensioni di dati sull'apprendimento e sul benessere degli studenti.</p>	<p>Richiedono il consenso esplicito per la raccolta di dati e possono presentare sfide tecniche e etiche.</p>	<p>Comprendere l'influenza degli ambienti fisici e delle attività didattiche sull'apprendimento e l'engagement</p>
<p>Assessment e feedback</p>	<p>Dati da valutazioni formative e sommative e feedback degli studenti su corsi e materiali didattici.</p>	<p>Consentono di identificare lacune di comprensione e competenze, orientando la personalizzazione.</p>	<p>La qualità dei dati dipende dalla progettazione delle valutazioni e dalla sincerità del feedback degli studenti.</p>	<p>Guidare le revisioni dei curricula, miglioramento delle pratiche pedagogiche</p>

Dati di indagine	Sondaggi quantitativi e qualitativi, spesso attraverso questionari.	Strutturati, possono raggiungere grandi campioni.	Possono soffrire di <i>bias</i> di risposta, limitati dalla progettazione del sondaggio.	Valutare atteggiamenti, opinioni, e comportamenti degli studenti.
Dati delle applicazioni mobili	Dati raccolti da smartphone e smartwatch, inclusi l'uso delle app e i dati comportamentali.	Elevata personalizzazione e contestualizzazione.	Rischi per la privacy, dipendono dall'uso costante del dispositivo da parte degli utenti.	Ricerca su apprendimento mobile, monitoraggio del comportamento e della salute degli studenti.
Dati di geocalizzazione	Dati relativi alla posizione fisica raccolti tramite dispositivi mobili.	Permettono l'analisi di pattern spaziali.	Possono richiedere il consenso esplicito per questioni di privacy, precisione variabile.	Esplorare spazi di apprendimento, studiare l'impatto della mobilità degli studenti sulla loro esperienza educativa.

Tab. 1.1: Sintesi delle fonti dei big data in ambito educativo

Complessivamente, l'utilizzo strategico e consapevole di queste diverse fonti di dati apre nuove porte nella ricerca educativa, consentendo approfondimenti più dettagliati e una comprensione più sottile delle dinamiche di apprendimento. Ogni fonte, con i suoi punti di forza e limitazioni, richiede un'attenta valutazione per massimizzare il suo potenziale contributo agli studi educativi, sottolineando l'importanza di metodi di raccolta di dati etici e metodologicamente solidi.

1.3.3 *Le applicazioni dei big data nell'educazione*

I big data generati nel contesto dei sistemi formativi moderni trovano applicazione in una moltitudine di ambiti, apportando – sotto determinate condizioni - benefici per studenti, docenti, ricercatori e istituzioni educative (Boyd, Crawford, 2013; Gabbi, 2023; Long, Siemens, 2011; Wagner, Ice, 2012; Williamson, 2017).

Con l'ampio volume di informazioni in ambito educativo, che include iscrizioni, record accademici e disciplinari, gli istituti di istruzione superiore dispongono dei dataset necessari per effettuare analisi mirate. L'uso dei big data e delle analitiche in questi contesti può avere un ruolo trasformativo, modificando i processi esistenti di amministrazione, insegnamento, apprendimento e lavoro accademico (Baer, Campbell, 2011), contribuendo ai risultati delle azioni e delle politiche applicate e partecipando positivamente alle sfide contemporanee che l'istruzione superiore sta affrontando. Il contributo chiave dei big data in ambito educativo dipende fondamentalmente dall'applicazione di tre approcci al dato: l'analisi descrittiva, quella predittiva e quella prescrittiva (*descriptive, predictive, prescriptive analytics*) (Rinaldi, 2018).

L'analisi descrittiva mira a descrivere e analizzare i dati storici raccolti sugli studenti, l'insegnamento, la ricerca, le politiche e altri processi amministrativi. Utilizza i dati aggregati degli studenti per identificare pattern e tendenze, consentendo di rispondere alla domanda "cosa è successo?". Si avvale di strumenti statistici per riepilogare grandi volumi di dati in modo comprensibile, attraverso l'uso di report, dashboard e visualizzazioni. Questo tipo di analisi è fondamentale per comprendere il contesto attuale o storico di un'organizzazione o di un fenomeno, ma non fornisce indicazioni dirette sulle azioni future.

Fanno parte di questo utilizzo le seguenti applicazioni in campo educativo:

- Ottimizzazione della gestione delle lezioni e delle risorse: a livello istituzionale, i big data generati da sistemi informativi, piattaforme digitali, sensori IoT ed altre fonti consentono di migliorare i processi chiave come la pianificazione delle lezioni, l’allocazione delle risorse, la progettazione di spazi e orari, il reclutamento di studenti e docenti (Daniel, 2015). L’analisi dei dati può rivelare inefficienze e guidare l’ottimizzazione di questi processi organizzativi e gestionali. Ad esempio, la pianificazione delle lezioni può tenere conto di correlazioni tra specifici docenti, materie e classi di studenti, in modo da massimizzare risultati e soddisfazione. Risorse come aule e laboratori possono essere allocate in base all’analisi dei trend di utilizzo. Gli spazi fisici possono essere riprogettati per favorire modalità di apprendimento più efficaci emerse dai dati.
- Valutazione: per valutare l’efficacia della didattica le analisi possono abbracciare una varietà di dati, dai risultati degli studenti ai feedback raccolti attraverso questionari, alle interazioni online e alle revisioni su piattaforme pubbliche (Ferguson et al., 2015). I big data integrano utilmente le valutazioni più tradizionali, fornendo informazioni imparziali difficilmente ottenibili altrimenti.
- Valutazione dei contenuti digitali per la didattica online: in ambito di e-learning, i dati raccolti sulle interazioni degli studenti con i materiali e le risorse didattiche digitali disponibili sulla piattaforma (es. tempi di fruizione, sezioni consultate, risultati nei test di autovalutazione, ecc.) risultano fondamentali per capire dove questi contenuti funzionano bene e dove invece generano difficoltà o sono poco efficaci (Sergis, Sampson, 2017). Queste informazioni guidano il miglioramento e la riprogettazione dei contenuti digitali secondo i feedback restituiti dagli utenti (Shayaa et al., 2018).
- Ottimizzazione del curriculum: a livello di istituto e di sistema scolastico, l’analisi dei big data può rivelare preziose informazioni per ottimizzare la progettazione del curriculum, alline-

andoli maggiormente alle necessità degli studenti e del mercato del lavoro (Cumbley, Church, 2013). Ad esempio, correlando dati sui percorsi formativi con dati occupazionali è possibile identificare eventuali mismatch tra competenze sviluppate e competenze richieste dalle professioni, permettendo un adeguamento dei piani di studio.

L'analisi predittiva, invece, utilizza i dati per modellare e prevedere eventi o comportamenti futuri. Può fornire alle istituzioni sollecitazioni per decisioni informate e spunti d'azione basati sui dati, creando degli *alert* che giungano all'attenzione dei *decision maker*. Essa si basa su tecniche di statistica, *machine learning* e modellazione per stimare la probabilità di scenari futuri, rispondendo alla domanda "cosa potrebbe accadere in futuro?". Questo tipo di analisi si concentra sulla previsione di tendenze con un certo grado di probabilità, ad esempio, individuare studenti che mostrano comportamenti a rischio all'inizio del semestre che potrebbero portare all'abbandono o al fallimento del percorso. L'analisi predittiva è particolarmente utile per la pianificazione e l'ottimizzazione strategica, permettendo alle istituzioni di anticipare cambiamenti e adattarsi proattivamente. Fanno parte dell'analisi predittiva le seguenti applicazioni:

- Predizione dei risultati: tramite tecniche di *machine learning* applicate ai dati raccolti sugli studenti nel corso del tempo, è possibile costruire modelli predittivi in grado di prevedere variabili come il superamento o meno di esami (Asif, Merceron, Ali, Haider, 2017). Queste previsioni consentono di identificare per tempo gruppi di studenti potenzialmente a rischio ed attivare tempestivamente misure correttive o di supporto, cercando di prevenire situazioni critiche.
- Sistema di allerta precoce: altra faccia della predizione dei risultati, uno degli impieghi più diffusi e promettenti dei big data in educazione sono i sistemi di *early warning*, che spesso

sfruttano l'intelligenza artificiale per elaborare i dati raccolti su ogni studente durante il suo percorso e costruire modelli predittivi specifici in grado di calcolare il rischio di abbandono scolastico o insuccesso accademico (Arnold, Pistilli, 2012; Johar, Kew, Tasir, Koh, 2023). I sistemi di *early warning* sfruttano algoritmi di *machine learning* per elaborare dati quali voti, presenze, attività online, risultati in itinere, e identificare precocemente studenti a rischio di abbandono o insuccesso formativo. Ciò consente di attivare per tempo interventi di mentorship e supporto personalizzato.

L'analisi prescrittiva va oltre la predizione, fornendo raccomandazioni su come agire per influenzare positivamente gli eventi o gli esiti futuri. Essa combina i risultati analitici sia dei modelli descrittivi che predittivi per determinare linee guida utili alla risoluzione delle eventuali problematiche emerse. Inoltre, incorpora tecniche avanzate di ottimizzazione e simulazione per esplorare una varietà di possibili azioni e prevedere l'impatto di ciascuna, rispondendo alla domanda "cosa dovremmo fare?". L'analisi prescrittiva consente ai decisori di valutare gli sviluppi futuri delle loro attività considerando le opportunità (e i problemi connessi), oltre ad indicare la migliore strada da intraprendere per sfruttare quella previsione in modo tempestivo (Basu, 2013).

Fanno parte dell'analisi prescrittiva le seguenti applicazioni:

- Personalizzazione dell'apprendimento: la disponibilità di una grande mole di dati su profilo, interessi, stili di apprendimento, lacune e progressi del singolo studente consente di creare percorsi formativi e contenuti didattici personalizzati, tagliati su misura per i bisogni specifici di ciascun discente (Johnson et al., 2016). Ad esempio, le piattaforme formative come Khan Academy raccolgono in maniera continuativa dati quali: livello di partenza dello studente, tempo impiegato per terminare un'attività, risultati nei quiz, sezioni di un corso che

generano maggiori difficoltà, argomenti di interesse ecc. Questi dati consentono alla piattaforma di comprendere lacune ed esigenze specifiche dello studente. Di conseguenza, il sistema può suggerire attività di approfondimento su argomenti che risultano più complessi per quel discente, fornire esempi aggiuntivi laddove servono, proporre nuovi contenuti in linea con gli interessi dimostrati. In questo modo, la raccolta e l'analisi continua dei dati di apprendimento permette di plasmare un percorso formativo su misura per i bisogni di quel determinato studente, garantendo un'esperienza personalizzata molto più efficace rispetto ad un curriculum standard uguale per tutti. Un curriculum personalizzato che tenga conto delle inclinazioni dello studente si dimostra molto più efficace in termini di coinvolgimento e motivazione rispetto ad un apprendimento standardizzato (Institute for the Future, 2019).

- Miglioramento dei risultati di apprendimento: attraverso tecniche di learning analytics¹² (LA), i dati generati dagli studenti durante il loro percorso di studi possono essere raccolti e analizzati in tempo reale per identificare lacune nella preparazione o difficoltà di apprendimento (Papamitsiou, Economides, 2014). I docenti hanno quindi la possibilità di rivedere le strategie didattiche e mettere in atto interventi mirati laddove necessario, andando anche a supportare individualmente gli studenti che manifestano particolari esigenze. Diversi studi suggeriscono che questo monitoraggio continuo dei progressi di apprendimento, se accompagnato da adeguate letture pedagogico-didattiche dei dati e da interventi didatticamente fondati, può condurre a risultati accademici migliori, tempi di completamento ridotti e tassi di abbandono più bassi nei

12 I learning analytics sono la misurazione, raccolta, analisi e reportistica dei dati relativi agli studenti e ai loro rispettivi contesti, allo scopo di comprendere e ottimizzare l'apprendimento e gli ambienti in cui questo avviene.

- percorsi formativi online (Xing, Chen, Stein, Marcinkowski, 2020).
- Miglioramento dei contenuti formativi digitali: nell'ambito della formazione online, le tecniche di LA (si veda il capitolo 4 per un dettaglio sull'educational data mining e i learning analytics) permettono di tracciare nel dettaglio le interazioni degli studenti con i materiali e le attività didattiche disponibili sulle piattaforme di e-learning (Garavaglia, 2021; Sergis, Sampson, 2017). Questi dati evidenziano eventuali difficoltà riscontrate in specifiche sezioni di contenuti, aree di maggiore interesse, tempi di fruizione e così via. Sulla base di queste informazioni è possibile migliorare i contenuti digitali, rendendoli più chiari, interattivi ed efficaci. L'analisi dei dati di interazione degli studenti con i contenuti formativi digitali disponibili sulla piattaforma e-learning può rivelarsi fondamentale per identificare sezioni o componenti di un corso online che non funzionano in modo ottimale. Ad esempio, frequenti picchi di abbandono in corrispondenza di specifici moduli o unità didattiche potrebbero indicare che quella parte del corso risulta poco coinvolgente, di difficile comprensione o non allineata agli obiettivi formativi. Oppure, scarsi risultati medi in alcuni quiz o esercizi interattivi potrebbero segnalare domande mal formulate o attività non efficaci. Sulla base di queste informazioni emerse dall'analisi dei dati, il team di sviluppo può intervenire miratamente per riprogettare le sezioni "problematiche" del corso e-learning, in modo da migliorarne coinvolgimento ed efficacia didattica.
 - Supporto alla ricerca educativa: la massa di dati generati dal sistema scolastico rappresenta una fonte preziosa per ricerche sull'efficacia di nuove metodologie didattiche, approcci pedagogici innovativi, strumenti e tecnologie educative (Picciano, 2012). I big data consentono di condurre studi su campioni molto ampi e approfondire fenomeni difficilmente investigabili su scala ridotta. Ad esempio, l'abbondante quantità di dati

generati dai MOOC permette di esaminare le dinamiche di apprendimento e interazione su una scala senza precedenti. Queste ricerche possono offrire l'opportunità di esplorare e comprendere meglio i modelli formativi, guidandone il miglioramento continuo. I big data possono quindi rappresentare una risorsa preziosa anche per la ricerca educativa, fornendo accesso a campioni di analisi molto ampi e consentendo approfondimenti altrimenti difficilmente realizzabili. Eccone alcuni esempi:

- Analizzando i dati di migliaia di MOOC sulla piattaforma edX, ricercatori del MIT hanno identificato strategie didattiche e modalità di coinvolgimento degli studenti che massimizzano il completamento di questi corsi online (Gardner, Brooks, 2018).
 - Attraverso i dati di utilizzo di milioni di risorse educative aperte (OER), studiosi della Open University UK hanno determinato tipologie e formati di contenuti digitali che ottimizzano l'apprendimento (Farrow, Pitt, de los Arcos, Perryman, Weller, McAndrew, 2016).
 - Incrociando dati occupazionali con i percorsi formativi di laureati, istituti come la Purdue University negli USA hanno valutato l'efficacia dei diversi curricula nel preparare gli studenti al mercato del lavoro.
- Tutoraggio virtuale: infine, sofisticate applicazioni di intelligenza artificiale come i chatbot vengono addestrate su una grande mole di dati per fornire tutoraggio e supporto personalizzato agli studenti, rispondendo a domande, fornendo spiegazioni, feedback, suggerimenti sui percorsi formativi online (es. il chatbot Jill Watson della Georgia Tech University).

Concludendo, i big data rappresentano indubbiamente una risorsa preziosa per abilitare una nuova generazione di servizi a supporto dell'apprendimento, promettendo il passaggio da un approccio educativo generalista a uno che riconosce e valorizza le

peculiarità di ogni studente. Attraverso la possibilità di fungere come strumenti per la personalizzazione e l'adattamento dei percorsi formativi, i big data sembrano offrire le basi per un apprendimento più inclusivo ed efficace, in grado di rispondere dinamicamente alle esigenze individuali e di promuovere l'engagement e la motivazione degli studenti.

Tuttavia, l'entusiasmo per le potenzialità offerte dai big data non deve oscurare la necessità di affrontare criticamente gli aspetti etici e le sfide poste dalla loro implementazione nel settore educativo. Come evidenziato da Selwyn (2020), l'incremento della piattaformaizzazione e della datificazione nell'educazione porta con sé rischi significativi legati alla privacy, alla sicurezza dei dati e all'autonomia degli individui. La piattaformaizzazione educativa, ovvero la crescente dipendenza da piattaforme digitali per la gestione e la distribuzione dell'educazione, può portare a una centralizzazione del controllo sui dati educativi, spesso in mani di entità private che possono non avere come priorità gli interessi educativi o l'equità (Perrotta, 2021). Inoltre, la datificazione, o trasformazione di ogni aspetto dell'esperienza educativa in dati quantificabili, rischia di ridurre l'apprendimento a mere metriche, ignorando le dimensioni qualitative, emotive e relazionali che sono fondamentali nell'educazione.

La raccolta, l'analisi e l'utilizzo dei big data devono quindi essere governati da principi etici rigorosi e da normative adeguate che tutelino i diritti e la privacy degli studenti. È fondamentale garantire che le innovazioni guidate dai dati siano implementate in modo responsabile, promuovendo la trasparenza nei processi decisionali basati sui dati, assicurando la giustizia nell'accesso alle opportunità educative e proteggendo l'equità per evitare che le disparità esistenti vengano amplificate dalla tecnologia (Ranieri, 2019).

Per affrontare in modo efficace questi temi e garantire un uso consapevole dei dati, è indispensabile partire dalle competenze di base, promuovendo una data literacy diffusa tra studenti e edu-

catori. Questo sarà il focus dei prossimi due capitoli, dove esploreremo le strategie e gli strumenti necessari per sviluppare una solida competenza nell'uso dei dati a supporto dell'apprendimento.

Riferimenti bibliografici

- Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16(1), 3-9.
- Agasisti, T., Bowers, A.J. (2017). Data Analytics and Decision-Making in Education: Towards the Educational Data Scientist as a Key Actor in Schools and Higher Education Institutions. In G. Johnes, J. Johnes, T. Agasisti, & L. López-Torres (Eds.), *Handbook of Contemporary Education Economics* (pp. 184-210). Cheltenham, UK: Edward Elgar Publishing.
- Aljohani, N.R., Davis, H.C. (2012). Significance of Learning Analytics in Enhancing the Mobile and Pervasive Learning Environments. *2012 Sixth International Conference on Next Generation Mobile Applications, Services and Technologies*, 70-74.
- Ambrose, M. (2015). Lessons from the avalanche of numbers: big data in historical perspective. *I/S: A Journal of Law and Policy for the Information Society*, 11(2), 201-77.
- Arnold, K.E., Pistilli, M.D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270).
- Asif, R., Merceron, A., Ali, S.A., Haider, N.G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177-194.
- Atzeni, P., Ceri, S., Fraternali, P., Paraboschi, S., Torlone, R. (2018). *Basi di dati*. Milano: McGraw-Hill.
- Autorità per le Garanzie nelle Comunicazioni (AGCOM) & Autorità Garante della Concorrenza e del Mercato (AGCM). (2018). *Indagine conoscitiva sui big data*.
- Baer, L., Campbell, J. (2012). From metrics to analytics, reporting to action: Analytics' role in changing the learning environment. In

- Game changers: Education and information technologies* (pp. 53-65). EDUCAUSE.
- Baker, R.S., Siemens, G. (2013). Educational data mining and learning analytics. Retrieved June 7, 2024, from www.columbia.edu/~rsb2162/BakerSiemensHandbook2013.pdf
- Basu, A. (2013). Five pillars of prescriptive analytics success. *Analytics*, March/April Issue, 8-12.
- Battaglia, S. (1966). *Grande dizionario della lingua italiana* (Vols. 1-21). Torino: UTET.
- Becker, S.A., Cummins, M., Davis, A., Freeman, A., Glesinger Hall, C., Ananthanarayanan, V. (2017). *NMC Horizon Report: 2017 Higher Education Edition*. Austin, Texas: The New Media Consortium. Retrieved June 7, 2024, from <https://www.learntechlib.org/p/174879/>
- Beer, D. (2013). *Popular Culture and New Media: The politics of circulation*. London: Palgrave Macmillan.
- Beer, D. (2016). How should we do the history of big data? *Big Data & Society*, 3(1). <http://dx.doi.org/10.1177/2053951716646135>
- Blikstein, P., Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220-238.
- Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. Cambridge: MIT Press.
- Borgman, C. L. (2020). *Qu'est-ce que le travail scientifique des données? Big data, little data, no data* (C. Matoussowsky, Trad.). OpenEdition Press.
- Bowker, G. C. (2008). *Memory Practices in the Sciences*. London: MIT Press.
- Bowker, G. C. (2013). Data flakes: an afterword to 'Raw Data' is an Oxymoron. In L. Gitelman (Ed.), *'Raw Data' is an Oxymoron* (pp. 167-72). London: MIT Press.
- Bowker, G. C., Star, S. L. (1999). *Sorting Things Out: Classification and its consequences*. Cambridge: MIT Press.
- Boyd, D., Crawford, K. (2013). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society*, 15(5), 662-679.

- Cavagnero, S. M. (2019). Dati digitali e azioni sociali. In R. Grimaldi, S. M. Cavagnero, M. A. Gallina (eds.), *Big data e tracce digitali: Una ricerca internazionale tra studenti universitari*. Roma: Aracne.
- Chadwick, A., Stromer-Galley, J. (2016). Digital media, power, and democracy in parties and election campaigns: party decline or party renewal? *The International Journal of Press/Politics*, 21(3), 283-293.
- Chignard, S. (2013). *A Brief History of Open Data*. Retrieved June 7, 2024, from <http://www.paristechreview.com/2013/03/29/brief-history-open-data/>
- Cope, B., Kalantzis, M. (2015). Interpreting evidence-of-learning: educational research in the era of big data. *Open Review of Educational Research*, 2(1), 218-239.
- Cope, B., Kalantzis, M. (2016). Big data comes to school: implications for learning, assessment and research. *AERA Open*, 2(2), 1-19.
- Corazza, P. (2021). *L'intelligenza collettiva al tempo delle piattaforme digitali. Il modello del formicaio: implicazioni pedagogiche e alternative possibili* [Dissertation thesis, Alma Mater Studiorum Università di Bologna]. Dottorato di ricerca in Scienze pedagogiche, 33 Ciclo.
- Cortellazzo, M., Zolli, P. (2004). *Dizionario etimologico della lingua italiana*. Torino: Paravia.
- Couper, M. P. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. *Survey Research Methods*, 7(3), 145-156.
- Crawford, K. (2014). The anxieties of big data. *The New Inquiry*, 30 May. Retrieved June 7, 2024, from <http://thenewinquiry.com/essays/the-anxieties-of-big-data/>
- Cumbley, R., Church, P. (2013). Is “big data” creepy? *Computer Law & Security Review*, 29, 601-609.
- Daniel, B. K. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904-920.
- de Kerckhove, D. (2022). Algoritmo, big data e il sistema legale. In U. Ruffolo, G. Riccio, & A. F. Uricchio (Eds.), *L'intelligenza artificiale tra etica e diritti* (pp. 73-91). Bari: Cacucci.
- De Mauro, T. (Ed.). (2000). *Grande Dizionario Italiano dell'uso* (Vols. I-VI). Torino: UTET.
- Delmastro, M., Nicita, A. (2019). *Big data. Come stanno cambiando il nostro mondo*. Bologna: Il Mulino.

- Derrida, J. (1995). *Mal d'archive: une impression freudienne*. Paris: éditions Galilée.
- Facer, K. (2011). *Learning Futures: education, technology and social change*. Abingdon: Routledge.
- Facer, K. (2014). What is Space for? Towards a Politics and a Language for the Human in Education. *Technology, Pedagogy and Education*, 23(1), 121-126.
- Farrow, R., Pitt, R., de los Arcos, B., Perryman, L.-A., Weller, M., McAndrew, P. (2015). Impact of OER use on teaching and learning: Data from OER Research Hub (2013-2014). *British Journal of Educational Technology*, 46(5), 972-976.
- Fenwick, T., Mangez, E., Ozga, J. (Eds.). (2014). *Governing Knowledge: Comparison, knowledge-based technologies and expertise in the regulation of education*. London: Routledge.
- Ferguson, R., Clow, D., Beale, R., Cooper, A. J., Morris, N., Bayne, S., Woodgate, A. (2015). Moving through MOOCS: Pedagogy, learning design and Patterns of Engagement. In *Design for Teaching and Learning in a Networked World: 10th European Conference on Technology Enhanced Learning, EC-TEL 2015, Proceedings* (pp. 70-84). (Lecture Notes in Computer Science; Vol. 9307). Springer-Verlag. Advance online publication. https://doi.org/10.1007/978-3-319-24258-3_6
- Floridi, L. (2017). *La quarta rivoluzione: come l'infosfera sta trasformando il mondo*. Milano: Raffaello Cortina.
- Foucault, M. (1967). *Le parole e le cose*. Milano: Rizzoli.
- Foucault, M. (1969). *L'archéologie du savoir*. Paris: éditions Gallimard.
- Gao, F., Luo, T., Zhang, K. (2012). Tweeting for Learning: A Critical Analysis of Research on Microblogging in Education Published in 2008-2011. *British Journal of Educational Technology*, 43(5), 783-801.
- Gabbi, E. (2023). *Learning Analytics per lo sviluppo professionale: un approccio metodologico allo studio delle comunità di pratica*. Milano: Francoangeli.
- Garavaglia, A. (2021). Analytics e Big Data. In P. C. Rivoltella (Ed.), *Apprendere a distanza: Teorie e metodi* (1^a ed., pp. 399-404). Milano: Raffaello Cortina
- Gardner, J., Brooks, C. (2018). Student success prediction in MOOCS. *User Modeling and User-Adapted Interaction*, 28(2), 127-203.

- Gibson, D. C., Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In B. K. Daniel (Ed.), *Big data and learning analytics: Current theory and practice in higher education* (pp. 29-42). New York, NY: Springer.
- Gitelman, L., Jackson, V. (2013). Introduction. In L. Gitelman (Ed.), *'Raw Data' is an Oxymoron* (pp. 1-14). London: MIT Press.
- Greenhow, C., Askari, E. (2017). Learning and teaching with social network sites: A decade of research in K-12 related education. *Education and Information Technologies*, 22(2), 623-645.
- Grek, S. (2009). Governing by numbers: the PISA 'effect' in Europe. *Journal of Education Policy*, 24(1), 23-37.
- Grek, S. (2016). The life and work of the killer chart: on the art of visually assembling education comparisons. *European Consortium for Political Research*, Prague, September.
- Grimaldi, R., Cavagnero, S. M., Gallina, M. A. (eds.). (2019). *Big data e tracce digitali: Una ricerca internazionale tra studenti universitari*. Roma: Aracne. (Teoria e ricerca in educazione; Vol. 32).
- Grosvenor, I., Roberts, S. (2013). Systems and subjects: ordering, differentiating and institutionalising the modern child. In M. Lawn (Ed.), *The Rise of Data in Education Systems: Collection, Visualization and Use* (pp. 79-96). Oxford: Symposium.
- Hacking, I. (1990). *The Taming of Chance*. Cambridge: Cambridge University Press.
- Hilbert, M. (2016). Big data for development: a review of promises and challenges. *Development Policy Review*, 34(1), 135-74.
- Holmes, D. E. (2017). *Big data: A very short introduction*. Oxford: Oxford University Press.
- Ifenthaler, D., Widanapathirana, C. (2014). Development and Validation of a Learning Analytics Framework: Two Case Studies Using Support Vector Machines. *Tech Know Learn*, 19, 221-240.
- Institute for the Future. (2019). *The Next Era of Human-Machine Partnerships: Emerging Technologies' Impact on Society and Work in 2030*. Institute for the Future.
- Ipsas, C. (2019). Big data: tra necessità e responsabilità. In R. Grimaldi, S. M. Cavagnero, & M. A. Gallina (Eds.), *Big data e tracce digitali: Una ricerca internazionale tra studenti universitari*. Roma: Aracne. (Teoria e ricerca in educazione; Vol. 32).

- Johar, N. A., Kew, S. N., Tasir, Z., Koh, E. (2023). Learning analytics on student engagement to enhance students' learning performance: A systematic review. *Sustainability*, 15(10), 7849.
- Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., Hall, C. (2016). *NMC Horizon Report: 2016 Higher Education Edition*. Austin, TX: The New Media Consortium.
- Junco, R., Elavsky, C. M., Heiberger, G. (2013). Putting twitter to the test: Assessing outcomes for student collaboration, engagement and success. *British Journal of Educational Technology*, 44(2), 273-287.
- Kaiser, B. (2019). *La dittatura dei dati*. Milano: HarperCollins.
- Kays, K., Gathercoal, K., Buhrow, W. (2012). Does Survey Format Influence Self-Disclosure on Sensitive Question Items? *Computers in Human Behavior*, 28(1), 251-256.
- Keyes, R. (2004). *The Post-Truth Era: Dishonesty and Deception in Contemporary Life*. New York: St. Martin's Press.
- Khalil, M., Ebner, M. (2015). Learning Analytics: Principles and Constraints. In *Proceedings of ED-Media 2015 Conference* (pp. 1789-1799). VA: AACE.
- Kitchin, R. (2014a). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin, R. (2014b). Big data, new epistemologies and paradigm shifts. *Big Data and Society*, 1(1). <http://dx.doi.org/10.1177/2053-951714528481>
- Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication and Society*, 20(1), 14-29.
- Kitchin, R., Dodge, M. (2011). *Code/Space: Software and Everyday Life*. London: MIT Press.
- Kitchin, R., Lauriault, T. (2014). Towards critical data studies: charting and unpacking data assemblages and their work. *The Programmable City Working Paper 2*. Retrieved June 7, 2024, from <http://ssrn.com/abstract=2474112>
- Kitchin, R., Lauriault, T., McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, 2(1), 6-28.
- Kitchin, R., McArdle, G. (2015). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data and Society*, 3(1). Retrieved June 7, 2024, from <http://bds.sagepub.com/content/3/1/2053951716631130>

- Knewton. (2021). *Using Data Science to Personalize Learning*. Knewton.
- Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety*. Meta-Group.
- Lang, C., Siemens, G., Wise, A., Gaševi, D. (Eds.). (2017). *Handbook of Learning Analytics*. (1st ed.) Society for Learning Analytics Research.
- Lawn, M. (2013). The rise of data in education. In M. Lawn (Ed.), *The Rise of Data in Education Systems: Collection, Visualization and Use* (pp. 7-10). Oxford: Symposium.
- Lawn, M., Grek, S. (2012). *Europeanizing Education: Governing a New Policy Space*. Oxford: Symposium.
- Leonelli, S. (2019). *La ricerca scientifica nell'era dei big data. Cinque modi in cui i big data danneggiano la scienza, e come salvarla*. Milano: Meltemi.
- Liu, X., Sun, R., Wang, S., Wu, Y.J. (2020). The research landscape of big data: a bibliometric analysis. *Library Hi Tech*, 38(2), 367-384.
- Livio, T. (1982). *Storia di Roma dalla sua fondazione* (M. Scàndola, Trad.). Milano: Rizzoli.
- Long, P., Siemens, G. (2011). Penetrating the fog: analytics in learning and education. *EDUCAUSE Review*, 46(5), 30-40.
- Mackenzie, A. (2006). *Cutting Code: Software and Sociality*. Oxford: Peter Lang.
- Mackenzie, A. (2012). More parts than elements: how databases multiply. *Environment and Planning D: Society and Space*, 30, 335-350.
- Mackenzie, A. (2013). Programming subjects in the regime of anticipation: software studies and subjectivity. *Subjectivity*, 6(4), 391-405.
- Mackenzie, A. (2015). The production of prediction: what does machine learning want? *European Journal of Cultural Studies*, 18(4-5), 429-445.
- Mackenzie, A., Vurdubakis, T. (2011). Codes and codings in crisis: performativity, signification and excess. *Theory, Culture and Society*, 28(6), 3-23.
- Matcha, W., Gaševi, D., Uzir, N. A. A., Jovanovi, J., Pardo, A. (2019). Analytics of learning strategies: associations with academic performance and feedback. In C. Brooks, R. Ferguson, U. Hoppe (Eds.), *Proceedings of the 9th International Conference on Learning Analytics*

- and Knowledge (LAK'19): Learning Analytics to Promote Inclusion and Success* (pp. 461-470). New York: ACM.
- Matcha, W., Uzir, N. A., Gašević, D., Pardo, A. (2020). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies*, 13(2), 226-245. <https://doi.org/10.1109/TLT.2019.2916802>
- Mayer-Schönberger, V., Cukier, K. (2013). *Big Data: A Revolution That Will Change How We Live, Work and Think*. London: John Murray.
- Mayer-Schönberger, V., Cukier, K. (2014). *Learning from Big Data: The Future of Education*. New York: Houghton Mifflin Harcourt.
- Moss, P., Dahlberg, G., Grieshaber, S., Mantovani, S., May, H., Pence, A., Rayna, S., Blue Swadener, B., Vandebroek, M. (2016). The Organisation for Economic Co-operation and Development's International Early Learning Study: Opening for Debate and Contestation. *Contemporary Issues in Early Childhood*, 17(3), 343-351.
- O'Neil, C. (2017). *Armi di distruzione matematica*. Milano: Bompiani.
- Ozga, J. (2009). Governing education through data in England: from regulation to self-evaluation. *Journal of Education Policy*, 24(2), 149-163.
- Ozga, J. (2016). Trust in numbers? Digital education governance and the inspection process. *European Educational Research Journal*, 15(1), 69-81.
- Ozga, J., Dahler-Larsen, P., Segerholm, C., Simola, H. (Eds.). (2011). *Fabricating Quality in Education: Data and Governance in Europe*. London: Routledge.
- Panciroli, C., Macaudo, A. (2020). Ambienti integrati per la didattica. In *Colloquia* (pp. 61-67). Padova: Padova University Press.
- Papamitsiou, Z., Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49-64.
- Perrotta, C., Gulson, K. N., Williamson, B., Witzemberger, K. (2021). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. *Critical Studies in Education*, 62(1), 97-113.
- Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, 16(3), 9-20.

- Pistilli, M. D., Arnold, K. E., Bethune, M. (2012). Signals: Using Academic Analytics to Promote Student Success. *Educause. Review Online*, 10(8).
- Raffaghelli, J. E. (2017). Data Literacy in the Big and Open Data Society: An Educational Challenge. *Training & Teaching: European Journal of Research on Education and Teaching*, 25(3), 279-304.
- Raielli, R. (a cura di). (2017). *Progressi dell'informazione e progresso delle conoscenze. Granularità, interoperabilità e integrazione dei dati*. Roma: AIB.
- Ranieri, M., Cuomo, S., & Biagini, G. (2024). *Scuola e intelligenza artificiale. Percorsi di alfabetizzazione critica*. Carocci.
- Ranieri, M. (2019). La Media Literacy Education tra vecchie e nuove sfide in un mondo iperconnesso. In M. Gui (ed.), *Benessere digitale a scuola e a casa. Un percorso di educazione ai media nella connessione permanente* (pp. 9-35). Firenze: Mondadori Università.
- Reinsel, D., Gantz, J., Rydning, J. (2018). *The Digitization of the World from Edge to Core*. An IDC White Paper (#US44413318). Retrieved June 7, 2024, from <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>
- Rinaldi, F. M., Grimaldi, R., Cavagnero, S. M., Gallina, M. A. (a cura di). (2019). *Big Data e tracce digitali: Una ricerca internazionale tra studenti universitari*. Roma: Aracne. (Teoria e ricerca in educazione; Vol. 32).
- Rivière, P. (2022). Qu'est-ce qu'une donnée? Impact des données externes sur la statistique publique. *Courrier des statistiques*, N5.
- Rivoltella, P. C. (2024). Intervento e competenze per le professioni educative nella condizione postmediale. In *COO. DE Cooperative Digital Education* (pp. 33-39). Trento: Erickson.
- Rizvi, F., Lingard, B. (2010). *Globalizing Education Policy*. Abingdon: Routledge.
- Roncaglia, G. (2020). *L'età della frammentazione*. Roma: Laterza.
- Roncaglia, G. (2023). *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*. Roma-Bari: Laterza.
- Rosenberg, D. (2013). Data before the fact. In L. Gitelman (Ed.), *'Raw Data' is an Oxymoron* (pp. 15-40). London: MIT Press.
- Selwyn, N., Gašević, D. (2020). The datafication of higher education: Discussing the promises and problems. *Teaching in Higher Education*, 25(4), 527-540.

- Selwyn, N., Hillman, T., Eynon, R., Ferreira, G., Knox, J., Macgilchrist, F., Sancho-Gil, J. M. (2020). What's next for Ed-Tech? Critical hopes and concerns for the 2020s. *Learning, Media and Technology*, 45(1), 1-6.
- Sergis, S., Sampson, D. G. (2017). Teaching and Learning Analytics to Support Teacher Inquiry: A Systematic Literature Review. In A. Peña-Ayala (Ed.), *Learning Analytics: Fundamentals, Applications, and Trends* (pp. 279-304). Cham: Springer.
- Shayaa, S., Jaafar, N.I., Bahri, S., Sulaiman, A., Seuk Wai, P., Wai Chung, Y., Piprani, A.Z., & Al-garadi, M.A. (2018). Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges. *IEEE Access*, 6, 37807-37827.
- Sobe, N. (2013). Educational data at late nineteenth- and early twentieth-century international expositions: 'accomplished results' and 'instruments and apparatuses'. In M. Lawn (Ed.), *The Rise of Data in Education Systems: Collection, Visualization and Use* (pp. 41-56). Oxford: Symposium.
- University of Louisville. (2021). *A Tale of Three Models: What Contributes to Student Persistence*. Louisville: University of Louisville. Retrieved June 7, 2024, from <https://louisville.edu/oapa/institutional-research-and-planning/docs/student-retention-predictive-models>
- Van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. Oxford: Oxford University Press.
- Van Dijck, J. (2014). Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. *Surveillance and Society*, 12(2), 197-208.
- Wagner, E., Ice, P. (2012). Data changes everything: delivering on the promise of learning analytics in higher education. *Educause. Review*, July/August, 33-42.
- Weinberger, D. (2011). *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. New York: Basic Books.
- Williamson, B. (2017). *Big Data in Education: The Digital Future of Learning, Policy, and Practice*. London: SAGE Publications Ltd.
- Xing, W., Chen, X., Stein, J., Marcinkowski, M. (2016). Temporal prediction of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, 58, 119-129.

II. Origine, definizioni e fondamenti teorici della *data literacy*

Elena Gabbi

2.1 Definire la *data literacy*: dagli approcci tradizionali alle prospettive contemporanee

2.1.1 *Origini ed evoluzione del concetto*

La formula *data literacy* rappresenta un neologismo, il cui significato non è ancora del tutto chiaro e univocamente condiviso. Tuttavia, per iniziare la disamina delle varie connotazioni che può ricevere, si può dire che, in prima approssimazione, la traduzione italiana di questa espressione rimanda alle conoscenze e abilità fondamentali per la comprensione e l'uso dei dati digitali oppure al concetto di alfabetizzazione ai dati.

Come è stato illustrato nel capitolo 1, l'origine della parola *dato* (quantità o condizione nota) non aiuta a circoscrivere un concetto di per sé polisemico e complesso. Nell'Enciclopedia Treccani¹ la sua definizione è associata all'ambito informatico, quale minima parte di informazione abbinata ad un codice. Nella stessa fonte bibliografica, il dato digitale è definito come un'informazione assimilata a una quantità finita e discreta, a differenza del dato analogico che coincide con un valore stabilito tra infinite possibilità. Permettendoci un'ardita semplificazione, è possibile

1 Fonte: <https://www.treccani.it/enciclopedia/ricerca/dato/?search=dato>

affermare che i dati digitali che rappresentano il prezioso capitale insito nelle nostre azioni online, nonché il supporto quotidiano delle nostre funzioni cognitive quali memoria e comunicazione e infine lo strumento di informazione a servizio dei sistemi di potere non sono altro che rappresentazioni imperfette di informazioni, infinitamente più ricche e inaccessibili. Ciononostante, questa semplificazione consente di cogliere, attraverso i dati, caratteristiche e funzioni analizzabili, gestibili e manipolabili dalla mente umana.

Abbiamo discusso in precedenza come, nel panorama socio-politico odierno, la proliferazione dei dati digitali abbia suscitato il desiderio di sfruttarli a fini vantaggiosi e di lucro, dando il via allo sviluppo di sistemi capaci di superare i limiti computazionali finora esistenti. Non si tratta soltanto di contenuti digitalizzati e raccolti da esperti, bensì di multiformi tracce di comportamenti digitali e informazioni personali. L'avvento della rivoluzione digitale ha radicalmente trasformato le dinamiche in gioco, ridefinendo le abitudini e alterando gli equilibri di potere tra coloro che hanno accesso e possiedono dati, da un lato, e coloro che non possono permetterselo, dall'altro (Kitchin, 2014). Questo cambiamento ha portato a un ampliamento significativo del divario esistente tra gli esperti, impegnati nella manipolazione e nella interpretazione avanzata dei dati, e gli inesperti, che rischiano di sentirsi sempre più emarginati da un panorama informativo in rapida evoluzione. Nonostante il crescente interesse nei confronti delle applicazioni dei big data, dell'analisi dati e della loro visualizzazione, la ricerca sulla componente umana e la sua interazione con tali tecniche è ancora limitata, specialmente per quanto riguarda i dati più complessi (Wolff, Gooch, Caverio Montaner, Rashid, Kortuem, 2016).

Interrogarci sui presupposti e le competenze utili per entrare in dialogo con questi temi è non solo importante, ma anche urgente per il mondo educativo ai fini della cittadinanza consapevole. Quali sono le competenze necessarie alle persone per

imparare e risolvere problemi con i dati? Quali nuove abilità sono indispensabili per progettare, interpretare e criticare le analisi dei dati e le relative visualizzazioni complesse?

La nozione di *data literacy*, o “alfabetizzazione ai dati”, descrive in modo generico le abilità legate all’uso quotidiano dei dati per risolvere problemi reali. Tale competenza è considerata sempre più cruciale, in un contesto che vede l’incremento delle interazioni quotidiane con i dati e le decisioni che coinvolgono l’uso dei dati, anche personali. Lo studio pionieristico di Habermann, Burton e Frender (1998, p. 237) ha utilizzato il termine *data literacy* per riferirsi alla sfida educativa, per studenti e docenti, generata dall’emergere di “una vasta quantità di materiale da esplorare, ma in una varietà di formati poco familiari”, dovuto all’avvento del web. Gli autori hanno delineato una sequenza per l’insegnamento di capacità e abilità critiche nell’uso dei dati, che riflette la gerarchia Data Information Knowledge Wisdom (DIKW), distinguendo quindi tra dati, informazioni, conoscenza e saggezza (cfr. capitolo 1, par. 1.1.1). Attraverso tale premessa, l’approccio di Habermann e colleghi opera una netta distinzione tra il contenuto e la sua presentazione, associando al primo i dati grezzi inclusi nel sistema digitale e assimilando le informazioni alla rappresentazione elaborata degli stessi dati. Il focus di questa concezione della *data literacy* è pertanto costituito dalla lettura competente dei dati da parte dello studente, che deve elaborarli come informazioni e poi trasformarli in conoscenza. In seguito, si evidenzierà il limite di questo approccio, che considera il dato e l’infrastruttura come oggetti neutri (Borgman, 2017; Kitchin, 2014), sulla scia di convinzioni molto diffuse nel mondo industriale e commerciale sintetizzate dal motto “*letting the data speak for itself*” (lasciar parlare i dati). Tuttavia, partire da questa precoce proposta di definizione rappresenta un punto di partenza per evidenziare come la portata dell’alfabetizzazione ai dati sia andata costantemente espandendosi, anche in assenza di una univoca concettualizzazione.

Il dibattito europeo sulle competenze digitali ha di recente incluso la *data literacy* nel suo quadro DigComp 2.1 (Carretero, Vuorikari, Punie, 2017), collegando però la sua funzione alle prime fasi dell'alfabetizzazione digitale in relazione alla gestione delle informazioni (cercare, recuperare, interpretare) e meno all'elaborazione tecnica o creativa dei dati. La *data literacy* è stata concettualizzata dalle Nazioni Unite come il catalizzatore che abilita gli agenti del cambiamento a progredire verso "il futuro che desideriamo" (Gray, Gerlitz, Bounegru, 2018, p. 2), definendo graficamente il costrutto come l'intersezione tra *statistical literacy*, *information literacy* e le competenze tecniche per lavorare con i dati. I limiti di tale definizione possono essere ricondotti alla mancanza di una prospettiva umanistica e civico-sociale che metta al centro l'obiettivo e il ruolo dell'alfabetizzazione alla cittadinanza. Per gli organismi sovranazionali e governativi si auspica, infatti, la concettualizzazione di una competenza più complessa che permetta ai cittadini e alle società di sfidare le attuali strutture e dinamiche di potere, sottese a controllare i processi e i risultati socioeconomici della società, per perseguire gli Obiettivi di Sviluppo Sostenibile (Data Revolution Group, 2014a; Letouzé, 2016).

La mancanza di accordo tra le definizioni per la *data literacy* rischia di creare ostacoli rilevanti nell'insegnamento e nella valutazione di questa competenza, soprattutto considerando l'evoluzione continua della complessità e del volume dei dati. Questa lacuna teorica rende difficile promuovere in modo efficace le competenze necessarie per affrontare dati sempre più vasti e complessi. Allo stesso tempo, la valutazione dell'efficacia dell'insegnamento diventa un compito veramente arduo, se non si definisce una chiara cornice concettuale. Affrontare una simile sfida richiede pertanto una revisione delle definizioni esistenti per garantire che siano adeguate alle mutevoli esigenze del contesto attuale e futuro della *data literacy*.

2.1.2 *Inquadramento teorico e dimensioni chiave della data literacy*

Nell'analisi della letteratura del settore, possiamo considerare le diverse definizioni come punti collocati in un continuum che va dalle competenze tecnico-statistiche fino a quelle civiche, per poi procedere a disambiguare termini semanticamente affini e a tracciare una mappa delle connessioni tra i concetti.

Nel contesto dell'insegnamento di capacità e abilità critiche nell'uso dei dati, in riferimento agli studenti Deahl (2014) definisce la *data literacy* come:

La capacità di comprendere, trovare, raccogliere, interpretare, visualizzare e sostenere argomentazioni utilizzando dati quantitativi e qualitativi (p. 41).

Un aspetto di questa definizione che la distingue da altre è che tale concezione considera una componente fondamentale della *data literacy* la capacità di pensare in modo trasversale ai metodi di analisi dei dati, in una prospettiva più completa e ben fondata. Diversamente da approcci che tendono a essere determinati da confini metodologici, privilegiando l'analisi quantitativa dei dati o l'indagine qualitativa, esaminare dati prodotti da entrambi i metodi permette di collocare le interpretazioni dei dati numerici all'interno del contesto delle narrazioni umane. Consentire alle persone di interagire con entrambi i tipi di ricerca offre la possibilità di esplorare un problema da diverse prospettive. Deahl (2014) sviluppa ulteriormente questa idea delineando un insieme di abilità più specifiche per gli studenti che hanno acquisito un adeguato livello di *data literacy*. Queste competenze spaziano dalla comprensione generale dei dati alla verifica dei dati e alla loro presentazione per sostenere argomentazioni e raccontare storie, considerando le ampie implicazioni culturali, sociali e politiche dei dati raccolti.

Diversi autori sostengono che la *data literacy* debba essere una

competenza situata e ancorata ai problemi del mondo reale, perché il contesto svolge un ruolo cruciale nell'interpretazione e nell'applicazione dei dati. Inoltre, la situazione in cui i dati sono generati, raccolti e utilizzati influisce molto sulla loro rilevanza, affidabilità e interpretazione. Vahey e colleghi (2006) affermano che gli studenti devono essere consapevoli del contesto in cui si trovano i dati per essere in grado di generare argomenti appropriati e di eseguire le analisi per risolvere i problemi. Diversamente dalla capacità di agire con altre nozioni matematiche, indipendenti dalle circostanze e dalle conoscenze di altri insegnamenti, l'abilità di lavorare con i dati richiama diverse competenze disciplinari, richiedendo connessioni trasversali nel curriculum scolastico. Tali autori propongono i seguenti elementi chiave per una definizione esaustiva di *data literacy*:

- formulare e rispondere a domande utilizzando i dati come parte di un pensiero basato sulle evidenze;
- utilizzare dati, strumenti e rappresentazioni appropriate per supportare tale pensiero;
- interpretare le informazioni;
- saper sviluppare e valutare inferenze e spiegazioni basate sui dati;
- utilizzare i dati come base di partenza e valutare inferenze e spiegazioni basate sui dati;
- utilizzare i dati per risolvere problemi reali e comunicare le loro soluzioni.

La professoressa Wolff e i colleghi (2016) della Open University dal Regno Unito sottolineano che le competenze di *data literacy* da promuovere in ambito scolastico dovrebbero essere acquisite attraverso dati e compiti del mondo reale, utilizzando un approccio *inquiry-based*. In tale approccio, il processo di apprendimento si avvia mediante la presentazione di un problema che richiede una soluzione ed è strutturato in modo tale che gli

studenti siano sollecitati ad acquisire nuove conoscenze prima di poter procedere con la risoluzione. Gli studenti possono avvalersi di questo strumento sia in relazione a set di dati più piccoli che a quelli più grandi e complessi. La definizione degli autori è la seguente:

La *data literacy* è la capacità di porre e rispondere a domande del mondo reale a partire da insiemi di dati grandi e piccoli attraverso un processo di indagine, tenendo conto dell'uso etico dei dati. Si basa su competenze pratiche e creative di base, con la possibilità di ampliare la conoscenza di competenze specialistiche nella gestione dei dati in base agli obiettivi. Queste includono le capacità di selezionare, pulire, analizzare, visualizzare, criticare e interpretare i dati, nonché di comunicare storie a partire dai dati e di utilizzare i dati come parte di un processo di progettazione (p. 23).

Le diverse fasi sono parte di un ciclo iterativo, in cui le conclusioni possono sollecitare ulteriori domande e analisi, spesso di complessità crescente man mano che il problema viene affrontato e risolto. L'enfasi, qui, è posta sulla capacità di formulare e rispondere a domande legate al mondo reale, facendo affidamento su insiemi di dati di varie dimensioni attraverso un processo continuo di indagine.

Il report canadese dell'esperta statistica Ridsdale e colleghi (2015), a seguito di una mappatura sistematica della letteratura esistente, propone questa definizione sintetica di *data literacy*: “la capacità di raccogliere, gestire, valutare e applicare i dati in modo critico” (p. 3). Il collegamento tra pensiero critico e *data literacy* è pertanto un ulteriore tassello da aggiungere per sviluppare programmi per cittadini informati e consapevoli nell'era dei dati. Nello specifico, il pensiero critico in relazione alla *data literacy* può comprendere la valutazione dei dati, dei loro risultati e la riflessione sulle fonti e sul contesto, oltre che la consapevolezza delle

implicazioni etiche, inclusi problemi di *privacy*, *bias* e discriminazione. Un'ulteriore revisione delle concettualizzazioni esistenti sulla *data literacy* in senso critico è stata operata dalla ricercatrice tedesca Sander (2023), proponendo una categorizzazione degli approcci suddivisi in tre categorie:

- competenze pratico-strumentali
- riflessione critica attraverso l'uso di dati
- educazione critica alla datificazione.

L'analisi della letteratura dell'autrice ha confermato la forte predominanza della componente pratico-strumentale quale prima categoria di studi, ma ha anche individuato in un limitato numero di ricerche un orientamento alternativo che incoraggia un esercizio critico e creativo sui dati e sui dataset. Nella seconda categoria, i contributi mirano a favorire una visione critica durante l'utilizzo dei dati, integrando gli obiettivi di conoscenza e manipolazione dei dati con una prospettiva critica. La riflessione, in questo caso, si concentra spesso sulla valutazione del contenuto dei media digitali o sull'uso personale dei dati da parte degli studenti, riconoscendo, ad esempio, il processo e le circostanze della generazione dei dati, oltre alle motivazioni per la loro raccolta. Tuttavia, Sander (2023) evidenzia che il limite di questa prospettiva risiede nella possibilità che l'onere e la responsabilità ricadano prevalentemente sugli individui, considerando che l'agency attuale dei cittadini è ancora limitata. Sebbene ancora sporadici nella letteratura, gli approcci alla *data literacy* della terza categoria avanzano in questa direzione, concentrandosi sulla promozione di una comprensione più ampia e sulla riflessione circa le implicazioni strutturali e sistemiche della *datafication*, che si stanno radicando in profondità nelle società odierne (cfr. capitolo 1), e su come opporsi efficacemente.

Nella prospettiva sociotecnica, i sistemi tecnologici sono il risultato dell'interazione tra forze sociali e caratteristiche strutturali

(Huysman, Wulf, 2006; Raffaghelli, Stewart, 2020). Pertanto, progettare, implementare e utilizzare le tecnologie informatiche e digitali, comprese le pratiche afferenti ai big data, non è un'operazione solo di natura tecnico-informatica. Le strutture e le procedure rappresentano il prodotto finale delle dinamiche e delle negoziazioni tra diversi fattori: le esigenze di tipo tecnologico, gli elementi umani e di interazione con le persone e le condizioni derivanti dai diversi contesti politici e socioculturali. Le definizioni di *data literacy*, che aspirano a una comprensione approfondita della portata della *datafication*, includono anche elementi che favoriscono la comprensione delle varie sfumature dell'ingerenza che possono assumere la politica economica e le ideologie nel corso di questo processo (Pangrazio, Sefton-Green, 2020).

Includere nelle definizioni di *data literacy* anche le dimensioni di analisi, gestione e lettura dei dati in relazione ai big data fa sorgere alcune considerazioni rilevanti a proposito delle connessioni tra contesto, cognizione e struttura tecnico-informatica (D'Ignazio, Bhargava, 2015). Per molti individui usare i dati in modo competente potrebbe non essere possibile a causa della complessità tecnica; inoltre, spesso, non è possibile risalire ai dati se questi sono stati raccolti con strumenti di *data mining* (cfr. capitolo 4) e non sono pertanto disponibili. Infine, se manca il rispetto dei principi di trasparenza al momento in cui vengono raccolti, anche la capacità di lettura dei dati risulta compromessa. Per costruire una competenza che permetta di lavorare sulle implicazioni dei big data anche per i soggetti non esperti, D'Ignazio e Bhargava (2015) propongono di aggiungere le seguenti competenze alle definizioni precedenti di *data literacy*:

- identificare quando e dove i dati vengono raccolti in modo passivo attraverso le azioni e le interazioni degli utenti;
- comprendere quali sono le manipolazioni degli algoritmi eseguite su grandi dataset per individuare i modelli d'uso degli utenti;

- valutare l’impatto etico, sia effettivo che potenziale, delle decisioni *data-driven* per gli individui e la società.

Questo approccio ha l’obiettivo di permettere alle persone di “*speak data*” (parlare il linguaggio dei dati) (ivi, p. 4), anche di fronte alle sfide della datificazione, diventando consapevoli rispetto alle orme digitali che lasciano le azioni online, al funzionamento e alle limitazioni dei modelli previsionali e di raccomandazione e all’adozione di una prospettiva etica e socialmente responsabile. Tale visione è diametralmente opposta all’intento di “*letting the data speak for itself*” e implica la capacità di ragionare e intervenire sulle motivazioni e le finalità degli interventi data-based.

Perseguendo l’obiettivo di avvicinare i cittadini alle trasformazioni in corso nella società dei dati, D’Ignazio (2017, p. 7) identifica in seguito come *creative data literacy* “la capacità di leggere, lavorare, analizzare e argomentare con i dati come parte di un processo più ampio di indagine sul mondo”. Tale contributo, che richiama esplicitamente la pedagogia popolare di Freire (2002), vuole distanziarsi da quei programmi didattici che assimilano la *data literacy* alla conoscenza matematico-statistica, per evidenziare che gli studenti hanno bisogno di percorsi formativi di stampo non tecnico per perseguire il fine ultimo dell’alfabetizzazione, ovvero l’emancipazione. Tale prospettiva permette di accennare alla complessità della definizione del concetto di *literacy* e delle diverse accezioni che può ricoprire nei contesti dell’educazione alla cittadinanza e di trasformazione della società, temi su cui torneremo in seguito nel corso di questo capitolo.

Secondo Pangrazio e Selwyn (2019), è necessario introdurre il termine *personal data literacies* per riferirsi alla comprensione dei dati quali elementi situati socialmente e che dipendono dal contesto. Tale prospettiva sottolinea la necessità per gli individui di sviluppare strategie e tattiche per navigare le complessità dei dati digitali personali, per poter affrontare gli attuali problemi di

ineguaglianze, privacy, *bias* e disinformazione. Pertanto, le *personal data literacies* comprendono le competenze e la consapevolezza necessarie per navigare le complessità dei dati digitali personali e le più ampie implicazioni sociali dell'uso e dell'abuso dei dati, suddivise in cinque ambiti:

- *Identificazione*: la capacità di riconoscere e comprendere i diversi tipi di dati personali.
- *Comprensione*: la capacità di interpretare e valutare in modo critico i dati personali, compresi i concetti di privacy e sicurezza.
- *Riflessività*: la consapevolezza delle implicazioni sociali, etiche e politiche legate all'uso dei dati personali.
- *Utilizzo*: la capacità di applicare in modo efficace i dati personali in contesti specifici (es. per prendere decisioni informate).
- *Tattica*: lo sviluppo di strategie e abilità pratiche per gestire, proteggere e negoziare l'uso dei propri dati personali.

Gray, Gerlitz e Bounegru (2018) suggeriscono di orientarsi verso un approccio che problematizza le infrastrutture sottostanti ai dati che influenzano e organizzano concretamente le relazioni tra persone, oggetti, prospettive e tecnologie. Il concetto di *data infrastructure literacy* si riferisce infatti alla capacità di comprendere in modo critico e navigare attraverso le strutture fondamentali che gestiscono e producono dati. La premessa a questa definizione è la considerazione che i dataset non sono mezzi neutrali per registrare informazioni, ma “plasmano il mondo stesso in conformità a visioni, valori e culture diverse, trasformandolo in un territorio navigabile attraverso i dati” (ivi, p. 3). Le infrastrutture, infatti, possono anche esercitare un'influenza normativa, modellando la generazione e la gestione della priorità assegnata alle informazioni. Questo aspetto richiama la capacità di sviluppare una consapevolezza critica delle dinamiche e delle possibilità intrinseche nella *datification*, superando una visione di

alfabetizzazione tecnica limitata alla mera manipolazione di dataset. La *data infrastructure literacy* prende pertanto in esame le situazioni in cui i dati vengono riutilizzati in modi creativi, non previsti dall'intenzione originale di chi li ha creati, include la capacità di raccogliere dati alternativi oppure, in alcuni casi, la scelta di non produrre ulteriori informazioni, con l'obiettivo di creare spazi vuoti di "non-datificazione" rispetto ad alcuni argomenti o temi sensibili.

Un ultimo contributo rilevante è il concetto di *literacy in the age of data* proposta da Bhargava, Deahl, Letouzé, Noonan, Sangokoya e Shoup (2015) – alla lettera, alfabetizzazione nell'era dei dati – che si differenzia dalla tradizionale *data literacy*, in quanto si sposta dall'acquisizione di competenze tecniche specifiche per l'analisi dei dati, a una visione più ampia e inclusiva che si concentra sulla capacità di comprendere, utilizzare e comunicare i dati in modo efficace e significativo per la vita personale e della propria comunità. I dati non sono più concepiti come risorse da sfruttare per alimentare o agevolare i sistemi di potere esistenti, ma come "un ecosistema da plasmare per sfidare proprio questi sistemi di potere" (Letouzé, 2016, para 3). Questo slittamento di significato implica una maggiore attenzione alla comprensione del contesto in cui i dati sono prodotti e utilizzati, nonché alla capacità di valutare criticamente le fonti di dati e le implicazioni etiche e sociali dell'uso dei dati. Inoltre, la *literacy in the age of data* richiede una maggiore attenzione alla partecipazione e all'empowerment dei gruppi e delle comunità, in modo che i diversi attori sociali possano utilizzare i dati per affrontare le sfide che affrontano e per promuovere la trasformazione dello status quo. Ciò richiede la progettazione di processi di coinvolgimento delle comunità che siano inclusivi, partecipativi e rispettosi delle diverse prospettive e conoscenze. In sintesi, la *literacy in the age of data* si concentra sulla capacità di comprendere e utilizzare i dati in modo significativo e contestualizzato, nonché sulla promozione dell'empowerment delle comunità attraverso l'uso dei dati. Questo cam-

biamento di prospettiva ha importanti implicazioni per la progettazione e l'implementazione di programmi di *data literacy* e per la promozione di una società più inclusiva e partecipativa nell'era dei dati.

Possiamo collocare le concettualizzazioni di *data literacy* su uno spettro che va dalla comprensione pratico-strumentale fino alla componente critico-riflessiva, anche se è importante notare che la maggior parte delle definizioni si basano su tradizioni di alfabetizzazione informativa e statistica, si riferiscono ad integrazioni nel curriculum scolastico oppure riguardano soltanto la capacità di compiere operazioni con i dati.

A conclusione di questo percorso che ha preso in considerazione numerosi contributi, talvolta inconciliabili, per circoscrivere il campo semantico e operativo della *data literacy*, è essenziale riconoscere che la sua definizione è un ambito tuttora in evoluzione, arricchito da varie prospettive che riflettono la complessità e la diversità delle competenze coinvolte. La figura 2.1 ha lo scopo di rappresentare una panoramica sintetica degli elementi chiave associati alla *data literacy*, evidenziando le diverse prospettive adottate dal mondo scientifico, le competenze fondamentali che gli individui dovrebbero sviluppare e i temi educativi cruciali a cui risponde il concetto.

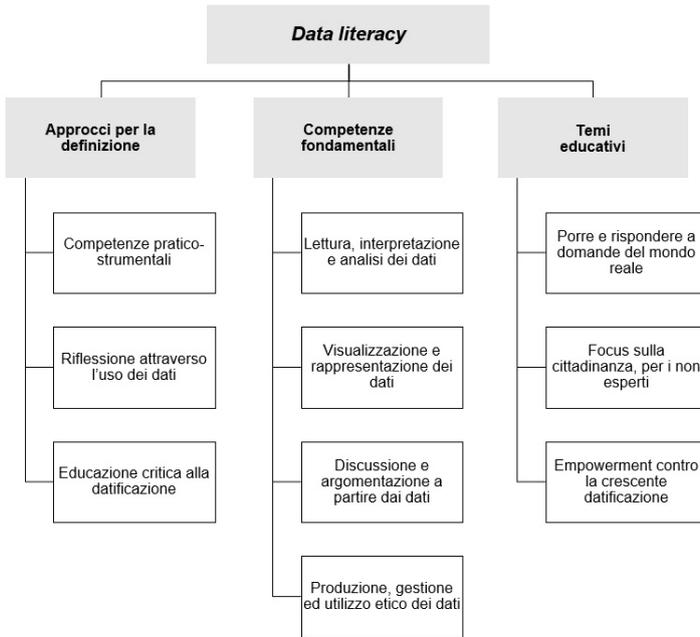


Fig. 2.1: Sintesi dei concetti legati alla definizione di *data literacy*: approcci, dimensioni principali e temi educativi sollecitati

Le differenze di approccio delineate sottolineano la varietà di obiettivi a cui possono riferirsi e le implicazioni per l'applicazione in ambito educativo, attraverso la realizzazione di curricula anche molto diversificati tra loro e destinati a svariati target (Sander, 2023). Diversamente dall'accezione di competenza tecnica e strumentale, l'approccio riflessivo alla *data literacy* si concentra sulla valutazione consapevole del contenuto delle rappresentazioni e sull'uso dei dati personali, ponendo attenzione al processo e alle circostanze della generazione dei dati. L'approccio critico va oltre la semplice riflessione individuale, sottolineando le implicazioni strutturali e sistemiche della *datafication* nella società contemporanea, nonché gli strumenti per agire in tale contesto.

Le varie componenti della competenza possono essere sintetizzate in quattro aree che costituiscono i pilastri fondamentali in comune alle definizioni più accreditate (D'Ignazio, 2017; Deahl, 2014; Wolff et al., 2016): lettura e analisi dei dati, visualizzazione e rappresentazione, discussione sui dati e loro generazione e gestione. Infine, parlare di *data literacy* permette al settore educativo di rispondere ad alcuni temi sollecitati dal dibattito contemporaneo sull'ingresso dei big data e delle loro implicazioni per la società, a partire dalla gestione quotidiana consapevole dei dati per rispondere ai propri bisogni di conoscenza e approfondimento, all'interpretazione e alla scoperta delle recenti evoluzioni tecnologiche da parte di chi non ha familiarità con la *data science*, fino a forme più complesse di collaborazione e attivismo per contribuire consapevolmente allo sviluppo della realtà in cui vivono.

2.2 Elementi di congruenza e discrepanza tra *data literacy* e altre forme di alfabetizzazione

2.2.1 *Tecnologie e pratiche: in dialogo con digital, information e data visualization literacy*

Per delineare più precisamente l'approccio teorico della *data literacy* e i suoi obiettivi, possiamo posizionarla all'interno di un contesto più ampio, mettendola in relazione alle altre forme di alfabetizzazione e alle discipline da cui hanno origine. La *data literacy* è infatti collegata ad altri tipi di alfabetizzazione, quali la *media literacy*, la *digital literacy*, l'*information literacy*, la *statistical literacy* e il pensiero computazionale, che sono stati ampiamente esplorati e analizzati. Le distinzioni tra le varie categorie sono oggetto di dibattito e, senza voler esaurire la discussione su questi argomenti, ci proponiamo di evidenziare le aree di sovrapposizione e le visioni comuni tra le diverse concezioni in relazione alla *data literacy*.

A questo punto diventa opportuno chiedersi: in che modo la *data literacy* si differenzia da questi ambiti che la precedono? E quali possibili sovrapposizioni concettuali si evidenziano?

La visione che incontra un maggior accordo nella comunità scientifica è quella che considera l'alfabetizzazione ai dati nello stesso bacino concettuale della *digital literacy*, l'*information literacy* e della *media literacy* (Markham, 2020). Nel corso del paragrafo, analizzeremo le similitudini e le caratteristiche uniche della *data literacy* rispetto a *digital*, *information* e *data visualization literacy*, posponendo la discussione relativa alla *media literacy* che pure viene talvolta nominata in associazione con le competenze sui dati.

Con una portata ampia, la *digital literacy* si concentra sui media digitali e sulle tecnologie interattive in rete e può far riferimento a competenze performative e strumentali, così come a competenze critiche (Calvani, Fini, Ranieri, 2010; Hobbs, 2011). Esistono connessioni tra la *digital literacy* e la *data literacy*, in particolare in relazione alla comprensione consapevole di come avvalersi in modo critico della struttura e delle caratteristiche dei media digitali e delle risorse tecnologiche. Come anticipato, nel panorama istituzionale europeo, la *data literacy* è concepita come una componente della più vasta e complessa *digital literacy*, o competenza digitale², ed è associata all'*information literacy* come capacità di estrarre informazioni e manipolare dati. Tale gerarchia teorica presuppone una diversa complessità concettuale tra dato, quale unità "grezza" e minimale di significato, e informazione ela-

- 2 La traduzione di *digital literacy* in italiano come "competenza digitale" solleva questioni complesse e articolate. Il dibattito scientifico ha problematizzato non solo la terminologia, ma anche le diverse concezioni di *literacy* in termini di abilità strumentali o come pratica sociale (Cappello, 2017; Ranieri, 2018). Tuttavia, l'uso del termine *competenza digitale* porta inevitabilmente a sottolineare il collegamento tra *data literacy* e *digital literacy*, richiamando la collocazione della prima all'interno del framework europeo DigComp, come discusso di seguito in questo paragrafo.

borata nei contesti digitali³; inoltre, è la concezione più diffusa nei modelli e nei documenti istituzionali dell'Unione Europea, sebbene i più recenti aggiornamenti includano anche elementi che suggeriscono un cambio di prospettiva.

Su tutti, il quadro europeo per le competenze digitali dei cittadini (Vuorikari, Kluzer, Punie, 2022), che abbiamo precedentemente menzionato, identifica le componenti chiave in cinque aree: informazioni e dati, comunicazione e collaborazione, creazione di contenuti, sicurezza e problem solving. La prima area “*Information and data literacy*” comprende le seguenti competenze:

- articolare i bisogni informativi, localizzare e recuperare dati, informazioni e contenuti digitali;
- giudicare la rilevanza della fonte e del suo contenuto;
- archiviare, gestire e organizzare dati, informazioni e contenuti digitali.

Le competenze digitali individuate dal quadro Europeo per quest'area si concentrano pertanto su tre aspetti cruciali che abbinano gli elementi digitali e i dati, inclusi quelli personali e connessi ai servizi della rete e delle app. Tra gli esempi di conoscenze, abilità e atteggiamenti per il cittadino digitalmente competente, aggiunti in questa versione aggiornata del modello originale del 2013, si annoverano a titolo esemplificativo (pp. 9-11):

- è consapevole che i contenuti online disponibili per gli utenti senza alcun costo monetario sono spesso pagati dalla pubblicità o dalla vendita dei dati dell'utente;

3 Tale concezione deriva dal percorso tracciato nella “Raccomandazione del Consiglio sulle competenze chiave per l'apprendimento permanente” del 2018 in relazione alla priorità sociale ed educativa riconosciuta alla competenza digitale.

- sa valutare i vantaggi e gli svantaggi dell'uso di motori di ricerca guidati dall'intelligenza artificiale (es. se da un lato possono aiutare gli utenti a trovare le informazioni desiderate, dall'altro possono compromettere la privacy e i dati personali o sottoporre l'utente a interessi commerciali);
- è consapevole dei potenziali *bias* delle informazioni causati da vari fattori (es. dataset, algoritmi, scelte editoriali, censura, limiti personali);
- sa che esistono archivi di dati aperti dove chiunque può ottenere dati per supportare alcune attività di problem solving (es. i cittadini possono utilizzare i dati aperti per generare mappe tematiche o altri contenuti digitali).

In questa concezione, tali competenze consentono una partecipazione consapevole e tutelata nella società contemporanea, riflettendo la necessità di una formazione completa nell'ambito delle competenze digitali anche per comprendere l'importanza dell'etica e della privacy associata alle tecnologie digitali. Il legame tra *data* e *digital literacy* sottolinea l'importanza di una comprensione critica nell'utilizzo delle infrastrutture e nella valutazione delle caratteristiche delle risorse digitali in un contesto sempre più interconnesso e personalizzato. Tuttavia, questa prospettiva tende a enfatizzare le dimensioni di base e trasversali, non addentrandosi in discussioni approfondite sull'impatto etico-sociali delle tecnologie digitali e più specificamente della datificazione⁴.

4 Sono stati evidenziati dalla studiosa italiana Ranieri (2019a) i limiti di una mancata concettualizzazione degli aspetti etico-sociali del Dig-Comp. Sebbene siano stati incorporati elementi relativi alla sicurezza (Area 4), tali riferimenti appaiono insufficienti a coprire appieno l'ampiezza della responsabilità delle azioni in rete. Tale responsabilità è da considerare sia da un punto di vista individuale, in termini di consapevolezza e gestione delle conseguenze per i propri comportamenti online, sia collettivo, riguardante la gestione delle disparità e delle iniquità derivanti dal *digital divide*. Questi aspetti sono invece inclusi nel modello

Come illustrato in precedenza, il complesso di competenze connesse con *information* e *data literacy* presenta elementi comuni, tra cui la capacità di gestire l'informazione, valutare criticamente le fonti e utilizzare i dati secondo principi etici. Secondo Calzada Prado e Marzal (2013), la *data literacy* risulta pertinente soprattutto nei contesti scientifici e di ricerca, mentre l'*information literacy* si applica in maniera più generica a diversi contesti. Tali autori definiscono l'alfabetizzazione ai dati come "la componente dell'*information literacy* che consente agli individui di accedere ai dati, interpretarli, valutarli criticamente, gestirli, trattarli e utilizzarli eticamente" (p. 126). Da questo punto di vista, l'*information literacy* e la *data literacy* farebbero parte di un continuum, un processo graduale di educazione scientifico-investigativa che inizia a scuola, si perfeziona e si specializza nell'istruzione superiore e fa parte del bagaglio di competenze degli individui per tutta la vita.

La rappresentazione visiva costituisce un'occasione per migliorare l'accessibilità e l'interattività dei risultati derivanti dall'analisi dei dati, traducendoli in un registro visuale immediato e coinvolgente (torneremo su questo argomento all'interno del capitolo 3). Il processo di visualizzazione dei dati, *data visualisation* o *dataviz*, sintetizza informazioni qualitative o quantitative attraverso rappresentazioni grafiche, facilitando l'esplorazione, l'interpretazione e la comunicazione dei dati (Azzam, Evergreen, Germuth, Kistler, 2013). In tale ambito, la *data visualization literacy* è stata definita come "la capacità e l'abilità di leggere e interpretare i dati rappresentati visivamente e di estrarre informazioni dalle visualizzazioni dei dati" (Lee, Kim, Kwon, 2016, p. 552). Sebbene l'uso del linguaggio visuale abbia radici storiche antiche e sia stato a lungo utilizzato per riassumere e descrivere dati attraverso grafici stati-

concettuale di competenza digitale messo a punto da Calvani, Fini e Ranieri (2010).

stici, diagrammi, mappe e animazioni, di recente si registra un forte interesse nella *data visualization*, soprattutto in relazione ai big data (Williamson, 2017). Una considerazione importante da premettere è che le visualizzazioni dei dati non sono indipendenti dalla loro provenienza, dalla persona che le sta osservando e dal contesto in cui sono progettate e collocate. In virtù della non neutralità dei dati, la *data visualization* può diventare una forma di argomentazione retorica in forma visiva, amplificando la funzione narrativa dei dati e utilizzandoli per creare prospettive e generare spiegazioni sul mondo, al fine di convincere gli altri che tali prodotti siano la fedele rappresentazione della realtà (Kitchin, Lauriault, McArdle, 2015). La sociologa Pinney (2020, p. 226) afferma che “tralasciare questi fattori socioculturali limita la possibilità di comprendere come le competenze e la consapevolezza necessarie per interpretare le *data visualization* influiscano sulle relazioni di potere e sulla partecipazione in una società basata sui dati”. La *data literacy* associata al tema della visualizzazione può pertanto fornire un quadro di riferimento utile sia come strumento pratico e condiviso per l’interpretazione delle *dataviz*, sia come promozione del cambiamento socioculturale.

2.2.2 *Dimensioni critiche: l’educazione ai media e il ruolo del pensiero computazionale*

La *media literacy* viene spesso utilizzata per descrivere la capacità degli adulti di agire come cittadini in società satura di media (Ranieri, 2019b; Rivoltella, 2020). Secondo l’UNESCO (2013), il concetto di *Media and Information Literacy* (MIL) riunisce una serie più ampia di competenze di base, dall’*information, digital e media literacy* fino alle tecnologie dell’informazione e della comunicazione per la ricerca e la generazione di contenuti:

un insieme di competenze che consente ai cittadini di accedere, recuperare, comprendere, valutare e utilizzare, creare e condividere informazioni e contenuti mediali in tutti i formati, utilizzando vari strumenti, in modo critico, etico ed efficace, al fine di partecipare e impegnarsi in attività personali, professionali e sociali (p. 29).

Nella definizione curricolare della *media literacy* in molti paesi, l'obiettivo è consolidare le capacità creative e comunicative dei giovani (Buckingham, 2020) tanto quanto offrire una prospettiva di ampio respiro che possa rivolgersi ad un vasto numero di soggetti, anche esperti in ambiti specialistici quali "consumatori, creativi, ricercatori, educatori, *policy makers* e mondo degli affari" (Hobbs, 2021, p. 23). Oggetti di ricerca della *media literacy education* includono i media testuali, come giornali, riviste e libri di testo, con attenzione al linguaggio e ai messaggi sottostanti; i media audiovisivi, come film e video su YouTube, implicando l'analisi degli elementi visivi e sonori; i media digitali e social media quali contenuti online e sui siti di social networking, con un'enfasi sui pregiudizi e la presentazione delle informazioni; i prodotti multimediali che combinano diversi media in forme quali siti web e app.

Le pratiche mediali contemporanee non si limitano alla sfera d'azione dell'individuo, in quanto consumatore anche se "critico", ma sono intrinsecamente collegate in rete e richiedono competenze proattive legate alle proprie esperienze sociali e comunicative (Bateman, 2021; Cappello, 2021). Il framework per l'analisi dei prodotti mediali sviluppato da Buckingham coinvolge l'analisi di concetti chiave quali la rappresentazione, il linguaggio, il pubblico e la produzione (Buckingham, 2006, 2020). È possibile osservare come ci siano delle connessioni tra la *media* e la *data literacy*, in relazione all'attenzione offerta al modo in cui l'esperienza è mediata dalle scelte di rappresentazione dei contenuti e del linguaggio espressivo utilizzato e a come le istituzioni che desiderano il potere e il profitto possano agire da produttori e rivolgersi ad

un'audience a cui veicolare il proprio messaggio. Un esempio di questa possibile combinazione di approcci è presente nell'analisi mediaeducativa di un *data storytelling* problematico, un prodotto realizzato e condiviso in rete dai fautori di una teoria del complotto (Fabbro, Gabbi, 2021). Utilizzando il quadro concettuale della *media literacy* di Buckingham (2006), sono state esaminate le rappresentazioni veicolate dal *data storytelling* in esame, alcune pratiche di produzione che lo caratterizzano, oltre che le funzioni retoriche dei linguaggi impiegati e il modo in cui esso potrebbe rivolgersi a diversi pubblici. Tale analisi è stata condotta nel contesto di un workshop rivolto a un gruppo ristretto di educatori socio-pedagogici in formazione (Fabbro, Gabbi, 2024). I risultati emersi suggeriscono che i partecipanti considerano la combinazione di *data* e *media literacy* come una strategia pedagogica efficace per guidare diversi gruppi di utenti (adolescenti, pubblico in generale e gli stessi educatori) nella valutazione critica della (dis)informazione mediatica. Inoltre, questa integrazione promuove il pensiero critico dei partecipanti riguardo alla mediatizzazione e alla datificazione delle teorie del complotto.

È importante riconoscere che la *data literacy* non coincide con il campo delle competenze mediali e digitali, ma si collega con gli elementi di ciascuna di esse per articolare l'insieme di capacità necessarie per poter sviluppare *agency* in un mondo datificato (Pangrazio, Sefton-Green, 2020). Possiamo affermare che i principi della *media literacy* hanno mantenuto la loro rilevanza con l'avvento dei social network. Tuttavia, per comprendere fenomeni come lo sfruttamento economico dei dati digitali e le dinamiche delle *echo chambers*, aree circoscritte della rete in cui le credenze e le opinioni personali sono amplificate e replicate mentre le idee divergenti faticano a trovare spazio, sono necessarie conoscenze legate al funzionamento degli algoritmi e a temi affini al campo dell'alfabetizzazione ai dati. Alcuni autori si riferiscono alla *data literacy* come parte di una concezione aggiornata della *media literacy*, in cui le tracce informative digitali rappresentate dai dati

sono viste come qualcosa di situato e necessariamente compreso in relazione ad altre pratiche sociali (Dencik, Hintz, Redden, Treré, 2019). In Italia, Rivoltella (2022) sottolinea la necessità di una prospettiva più ampia rispetto alla semplice idea di controllo dell'uso dei dispositivi, sostenendo che la *media education* contemporanea non può esimersi dall'affrontare il tema dei dati, attraverso l'adozione di nuovi strumenti per l'analisi critica e per l'esercizio della cittadinanza attiva.

Sintetizzando, la *data literacy* si accosta, e in parte si sovrappone, all'ambito degli studi di *media literacy education* soprattutto per due motivi (Van Audenhove, Van den Broeck, Mariën, 2020): da un lato, alcuni processi di *datafication* sono chiaramente situati nel campo dei media, includendo, a titolo esemplificativo e non esaustivo, motori di ricerca, sistemi di raccomandazione, personalizzazione delle notizie e la profilazione degli utenti; dall'altro lato, alcune di queste evoluzioni si trovano in settori al di fuori dei media digitali (come le *smart cities*, l'ambito sanitario o l'economia digitale) e pertanto non possono essere del tutto inglobate nello stesso contesto. Gli algoritmi, l'intelligenza artificiale, il *machine learning* e le piattaforme di social networking sono tecnologie e applicazioni che si basano sui dati, i quali rappresentano un elemento fondamentale per la loro definizione e la loro comprensione. Non sorprende, quindi, che vi sia una crescente domanda di nuove forme di alfabetizzazione più specifiche: l'*algorithmic literacy* (Koenig, 2020), la *digital platform literacy* (Ha, Kim, 2023), l'*Artificial Intelligence literacy* (Ranieri, Cuomo, Biagini, 2024) o la già citata *data infrastructure literacy* (Gray et al., 2018). Pertanto, alla *data literacy*, a differenza di altri ambiti di competenza, è riconosciuto il ruolo di navigare attraverso le complesse intersezioni dei processi di *datafication*, che si estendono dai media digitali ad ambiti infrastrutturali della società. Ciò fornisce una base essenziale per comprendere criticamente il mondo moderno, senza addentrarsi nelle specificità di uno tra i molteplici sviluppi tecnologici possibili.

Alcuni autori hanno associato alla *data literacy* il pensiero computazionale, un processo cognitivo che permette di risolvere problemi di varia natura attraverso l'uso di metodi e strumenti scelti in base a una strategia predeterminata (Wing, 2006). Sebbene la programmazione informatica sia una pratica principalmente riscontrata tra i professionisti, le recenti discussioni sulla democratizzazione della *data science* invitano anche i non esperti a imparare competenze computazionali che supportano l'analisi e la risposta a domande utilizzando i dati.

Berikan e Özdemir (2020) hanno evidenziato che “risolvere problemi con i dataset” rappresenta una componente fondamentale per l'implementazione efficace del pensiero computazionale. Dapprima, è necessario individuare un problema per il quale l'utilizzo della capacità di calcolo risulti appropriato e agire con le competenze di *data literacy* (raccogliere, analizzare e rappresentare i dati), trasversali a tutti i processi di natura scientifica. In seguito, gli individui devono adottare un approccio algoritmico e sviluppare astrazioni consecutive per formulare una strategia di soluzione del problema che possa essere eseguita con l'ausilio di un elaboratore elettronico. Infine, ad un livello più avanzato, vengono integrate le competenze tecniche per automatizzare il processo tramite software appropriati.

Il termine *computational data literacy*, coniato da Chen e colleghi (2023), si riferisce agli aspetti della *data literacy* collegati alle abilità di programmazione degli studenti combinati alla capacità di risolvere problemi attingendo ai concetti fondamentali dell'informatica. Nell'intervento didattico sperimentale, illustrato nel loro contributo, gli autori hanno associato le competenze di programmazione ad alcuni compiti chiave, per ordinare, filtrare, pulire, unire e creare visualizzazioni significative di ampi e complessi *dataset* in modo efficiente e affidabile. Nel loro lavoro, hanno descritto diverse pratiche della *computational data literacy*, tra cui svolgere operazioni su interi set di dati, elaborare dati e utilizzare i risultati, sperimentare con dati e codice, verificare incrociando

fonti diverse, operare *debug* e aggiornare la presentazione dei dati ai risultati ottenuti in corso d'opera. È interessante notare che il punto di convergenza con la *data literacy* è invece costituito dalle considerazioni conclusive espresse dai partecipanti durante i workshop:

- i grandi dataset sono incomprensibili senza un supporto computazionale automatizzato;
- i dati sono plasmati dalle decisioni umane;
- i dati possono essere incompleti e “disordinati”;
- gli *outlier* (i valori anomali) possono influenzare la visualizzazione e l'analisi;
- i dati possono essere utilizzati per rispondere ad una serie di domande.

L'apporto della *data literacy* al pensiero computazionale si può manifestare, pertanto, attraverso la prospettiva critica sulla natura e i processi di elaborazione dei dati a partire dalla soluzione di una domanda scientifica, nondimeno il pensiero computazionale offre un contesto di applicazione e lo sviluppo di un progetto a partire dai dati, ad esempio un algoritmo, con un risultato concreto ed efficace (Berikan, Özdemir, 2020). Il pensiero computazionale, arricchito delle capacità e abilità critiche nell'uso dei dati, consentirebbe agli individui di suddividere problemi complessi in componenti più semplici e realizzare soluzioni attraverso processi di automazione che tengano conto dei limiti stessi dei dati di partenza, migliorando in tal modo la loro capacità di manipolare ed analizzare dati strutturati in modo efficiente ed equo.

2.2.3 *Dati come evidenze: il rapporto con le competenze statistiche e di valutazione*

Proseguendo l'analisi delle potenziali connessioni e sovrapposizioni con altri costrutti, Schield (2004) sostiene che la *data literacy* costituisce una componente cruciale della *statistical literacy*, proponendola sia da una prospettiva critica che disciplinare. Da un lato, i dati rappresentano le unità fondamentali con cui operano le statistiche e le informazioni, dall'altro lato, partendo dalle competenze di base si può giungere a una comprensione più avanzata e complessa. A differenza dell'ambito della *data literacy*, le definizioni di competenza statistica sono piuttosto coerenti. Tra le più accreditate, le definizioni di Wallman (1993) e Carlson, Fosmire, Miller e Nelson (2014) presentano notevoli convergenze. Il primo propone che la *statistical literacy* sia considerata come la capacità di comprendere e valutare criticamente i risultati statistici che permeano la nostra vita quotidiana e di apprezzare il contributo del pensiero statistico nelle decisioni professionali e personali, mentre, nel secondo contributo, si considera la *statistical literacy* equivalente alla capacità di leggere e interpretare le informazioni statistiche incontrate nella fruizione quotidiana dei media, dai grafici alle tabelle, fino alle affermazioni, ai risultati dei sondaggi e degli studi.

Watson e Callingham (2003) hanno costruito una scala unidimensionale di *statistical literacy*, descrivendo la seguente gerarchia di sei livelli, caratterizzata da processi cognitivi sempre più complessi in cui sono coinvolti sia le procedure matematico-statistiche, sia la comprensione del contesto in cui si ricavano le informazioni:

- *Idiosincratico*: attenzione frammentaria alle informazioni del contesto, uso tautologico della terminologia e abilità matematiche di base associate al conteggio e alla lettura dei valori nelle tabelle.

- *Informale*: impiego discorsivo o informale delle informazioni del contesto, che spesso riflette convinzioni intuitive e non statistiche, uso di singoli elementi di una terminologia e di un contesto più complesso e calcoli semplici rispetto a tabelle, grafici e probabilità.
- *Incostante*: impiego selettivo delle informazioni del contesto, spesso sotto forma di supporto alle argomentazioni, riconoscimento appropriato delle conclusioni ma senza giustificazioni e uso qualitativo piuttosto che quantitativo delle idee statistiche.
- *Coerente non critico*: impiego appropriato ma non critico delle informazioni del contesto, uso versatile della terminologia e con abilità statistiche associate alla conoscenza di media, probabilità semplici e caratteristiche dei grafici.
- *Critico*: attenzione alle informazioni del contesto sia in ambiti conosciuti che non noti ma che implicano l'uso appropriato della terminologia, l'interpretazione qualitativa dei fenomeni casuali e l'analisi delle variazioni.
- *Critico-matematico*: atteggiamento critico e problematizzante nei confronti del contesto e delle informazioni, ragionamento basato su relazioni matematiche e proporzioni, consapevolezza dei margini di incertezza nel fare stime e previsioni e interpretazione degli aspetti più sottili del linguaggio tecnico.

Nel loro insieme, tali definizioni indicano una comprensione condivisa della *statistical literacy* come la capacità di affrontare e valutare criticamente i compiti e le sfide in ambito matematico-statistico incontrate nella vita quotidiana. Tuttavia, è importante notare che queste stesse competenze compaiono, in qualche forma, in molte definizioni di *data literacy*, senza costituire l'intera definizione. Gould (2017), pur riconoscendo il valore della *data literacy* nel dotare l'individuo di competenze e conoscenze statistiche necessarie per partecipare a una società che sempre più spesso raccoglie dati su di lui per fare previsioni sui suoi modelli

di consumo e di comportamento, suggerisce di includere il concetto di *data literacy* all'interno della *statistical literacy*. In questa prospettiva, l'agentività dei cittadini è connessa alla competenza tecnica: l'arricchimento delle nozioni statistiche con le capacità e abilità critiche nell'uso dei dati potrebbe far crescere il numero di cittadini capaci di accedere e analizzare dati pubblici o derivati dai dispositivi personali per rispondere alle proprie domande, conferendo loro una voce più potente ed efficace nel dibattito democratico. I diversi contributi suggeriscono pertanto due posizioni contrastanti: la *statistical literacy* potrebbe far parte dell'insieme più ampio di competenze che costituiscono la *data literacy* o, al contrario, può rappresentare una sua evoluzione in senso professionalizzante.

La promozione delle competenze nell'uso dei dati non implica automaticamente né inevitabilmente la comprensione degli stessi, nonostante questa correlazione sia talvolta considerata implicita. Seymoens, Van Audenhove, Van den Broeck e Mariën (2020) suggeriscono di differenziare le componenti cognitive fondamentali per la *comprensione* dei dati, da un lato, e per il loro *utilizzo*, dall'altro, in quanto tali abilità rappresenterebbero due aree distinte di apprendimento. Nel loro studio hanno infatti osservato questa separazione di aree di competenza, indipendenti tra loro, nel contesto dell'esperienza di formazione e ricerca con studenti di diverse età e adulti in Belgio. I cittadini potrebbero essere ben consapevoli del ruolo dei dati e degli algoritmi nella società, ma non avere le competenze numeriche e statistiche per utilizzare i dati in modo autonomo e, al contrario, gli studenti possono essere abili nella programmazione di algoritmi o nella gestione dei dati, senza però comprenderne le conseguenze sociali. L'area di competenza associata all'utilizzo dei dati è prevalentemente orientata alla pratica, mentre quella legata alla comprensione dei dati è in maggior misura concentrata sulla riflessione critica e consapevole del loro ruolo nella società e nella vita personale. Quest'ultimo aspetto è spesso trascurato o, a dir poco meno sviluppato, in par-

ticolare nei contesti disciplinari per la formazione in statistica, analisi dei dati o programmazione informatica, mentre non dovrebbe essere sottovalutata anche per gli esperti e i professionisti la comprensione concettuale e critica dei limiti dei dati e delle loro implicazioni etiche.

Un ulteriore campo di potenziale sovrapposizione della *data literacy* è individuabile nell'*assessment literacy*, la competenza valutativa, anche se le differenze tra le due non sono state chiaramente esplicitate nella letteratura scientifica di riferimento (Mandinach, Gummer, 2013). La distinzione tra i due ambiti è importante poiché, mentre l'*assessment literacy* si focalizza sulla comprensione e l'interpretazione dei risultati delle valutazioni, la *data literacy* è un concetto più ampio che comprende la capacità di lavorare con diversi tipi di dati, inclusi quelli relativi a percezioni, opinioni e comportamenti. Le valutazioni – intese come dati raccolti dalle misurazioni di apprendimento degli studenti a livello di classe, scuola e nazione – costituiscono solo una tra le tipologie di dati che possono entrare nel processo decisionale in educazione, anche se la loro comprensione è senz'altro una componente importante nella presa di decisioni data-based. Sebbene le proprietà intrinseche dei dati e dei relativi sistemi di raccolta ed elaborazione possano modellare le modalità di impiego dei dati nei contesti educativi, l'utilizzo delle informazioni ottenute da valutazioni su ampia scala sul rendimento degli studenti rimane condizionato dalle competenze degli utenti, soprattutto da parte di insegnanti e dirigenti (Freddano, Pastore, 2018). L'*assessment literacy* riguarda, quindi, il processo di analisi e interpretazione dei dati nell'individuazione di eventuali criticità nelle prestazioni della scuola e nella definizione di azioni correttive. Di conseguenza, mentre le competenze relative alla comprensione e alla gestione dei dati riguardano tutti i cittadini, indipendentemente dalle conoscenze pregresse, per favorire una partecipazione consapevole nella società datificata, le competenze e le abilità valutative rappresentano una competenza specifica del contesto educativo, poi-

ché coinvolgono la raccolta e l'interpretazione di informazioni relative ai processi di apprendimento e insegnamento.

2.2.4 *La necessità di una prospettiva interdisciplinare*

Le definizioni concettuali delineate in questo contesto, volte a delimitare l'ambito di studio tra le diverse forme di alfabetizzazione, non devono essere interpretate come confini netti tra competenze indipendenti. Nella figura 2.2 viene proposta un'indicazione sintetica degli elementi comuni alle diverse *literacy* in relazione al processo di alfabetizzazione ai dati. Per ciascuna delle diverse componenti è stato individuato l'elemento che li accomuna alla *data literacy* e che è assimilabile ai temi chiave discussi in questo capitolo.

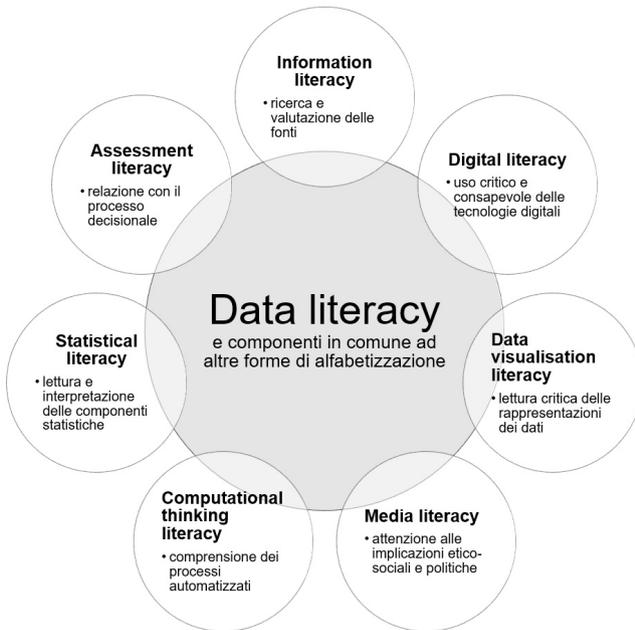


Fig. 2.2: *Sovrapposizioni concettuali tra data e altre tipologie di literacy*

La stretta relazione tra i diversi tipi di alfabetizzazione è evidente in particolare nel contesto sociorelazionale dei cittadini nell'era della rivoluzione digitale. Le competenze nell'ambito della *data literacy*, come pure le altre forme di alfabetizzazione discusse in precedenza con i loro evidenti connotati sociali, culturali e politici, sono intrinsecamente complesse e coinvolgono una serie di pratiche che si combinano in varie forme che spesso si adattano a specifici contesti sociali. Una concezione dinamica e complessa dell'influenza reciproca tra competenze era già emersa a partire dagli anni Novanta (Ranieri, 2018), dal gruppo di esperti multidisciplinare noto come The New London Group, che identificò una serie di elementi sociali e culturali fondamentali che compongono il paradigma della *multiliteracy* (Cazden et al., 1996). Questa nozione supera la semplice alfabetizzazione tradizionale, enfatizzando la capacità di combinare linguaggi e altre forme di espressione, di negoziare significati che provengono da media e tecnologie della comunicazione in evoluzione e di comprendere criticamente il panorama di prospettive della società contemporanea.

Lo studio di Yates e Carmi (2022) ha contribuito, in particolare, alla comprensione della stretta relazione tra *digital*, *media* e *data literacy*, sottolineando che il livello delle diverse alfabetizzazioni dei cittadini è dipendente dal formato tecnologico e spesso è fortemente influenzato dalle reti sociali, come anche dal background individuale, dallo status socioeconomico, dal livello di istruzione e dall'età. Gli autori hanno coniato il termine *networks of literacy* per riferirsi “ai modi in cui le persone sono coinvolte con gli altri, dove e con quali media” al fine di acquisire conoscenze, abilità e competenze per adattarsi e operare attraverso i media e i dispositivi tecnologici (ivi, p. 192). La ricerca ha rivelato, infatti, che le persone dipendono in modo rilevante dalla propria cerchia locale, sia per la verifica dei dati e dei contenuti digitali che per l'aiuto reciproco nelle attività quotidiane. Questi *networks of literacy*, caratterizzati dalla contaminazione di varie

competenze e conoscenze tra i membri di un gruppo, risultano fondamentali e costituiscono la base per navigare i contenuti dei media, acquisire competenze, verificare informazioni e sostenere una partecipazione attiva alla comunità. Il contributo di questa riflessione ci permette di estendere la concezione di *literacy*, su cui torneremo più avanti, al di là del contesto meramente “digitale” o “mediale”, comprendendo aspetti sia online che offline, e sottolinea l’importanza delle istituzioni educative, governative e socioculturali in questo processo.

Nell’arduo compito di chiarire le ambiguità e le contaminazioni nel campo della *data literacy*, emerge la fitta rete che collega il termine con nozioni di ambiti disciplinari estremamente eterogenei. Nel quadro articolato per definire le competenze che la caratterizzano, si fondono dimensioni di matrice cognitiva, di consapevolezza civica e critica e di abilità pratiche e situate nel contesto digitale. L’etnografa digitale Markham (2020) sintetizza:

“qualunque sia il nome che diamo alla *data literacy* e se questa sia o meno marcatamente diversa dalla *media literacy*, dalla *multiliteracy* o dalla *digital literacy*, è senz’altro più complessa del semplice apprendimento dell’accesso, dell’uso, dell’analisi o della creazione di dati attraverso le statistiche o della comprensione delle visualizzazioni e dei calcoli sui dati” (p. 203).

La nozione recente di *data literacy* deve riconoscere le radici del proprio sviluppo nel lungo dibattito che l’ha preceduta, pur affermando la propria specificità. Tale sviluppo ha rivelato tappe discontinue in cui la discussione è profondamente intrecciata con le evoluzioni storiche delle varie forme di alfabetizzazione e con il mutare del contesto disciplinare di riferimento. L’affermazione della sua specificità è, quindi, il risultato di un processo di continuo dialogo e affinamento concettuale all’interno di un panorama più ampio di competenze legate alla cittadinanza digitale nell’era della datificazione.

2.3 Il ruolo della *data literacy* nel contesto educativo attuale

2.3.1 *Un panorama normativo e socioculturale in evoluzione*

Nella Raccomandazione del Consiglio Europeo relativa alle competenze chiave per l'apprendimento permanente (2018) si riconosce l'importanza di sviluppare abilità quali "la capacità di risoluzione di problemi, il pensiero critico, la capacità di cooperare, la creatività, il pensiero computazionale, l'autoregolazione" e si aggiorna il concetto di competenza digitale. Come anticipato, in questa sede, viene riconosciuto un ruolo, seppur marginale, alla *data literacy* all'interno della definizione più ampia di competenza digitale:

Le persone dovrebbero assumere un approccio critico nei confronti della validità, dell'affidabilità e dell'impatto delle informazioni e dei dati resi disponibili con strumenti digitali ed essere consapevoli dei principi etici e legali chiamati in causa con l'utilizzo delle tecnologie digitali. [...] Le persone dovrebbero essere in grado di gestire e proteggere informazioni, contenuti, dati e identità digitali, oltre a riconoscere software, dispositivi, intelligenza artificiale o robot e interagire efficacemente con essi (2018, para 4).

La Raccomandazione del Consiglio Europeo delinea un quadro che amplia la nozione di competenza digitale, introducendo la necessità di specifiche competenze di lettura e analisi dei dati per affrontare le sfide e le opportunità della società contemporanea. Un'accresciuta consapevolezza sull'importanza delle sfide insite nella rivoluzione digitale, insieme alla necessità di colmare le disparità, si riflette attualmente in Europa sia a livello normativo che attraverso iniziative strategiche nell'ambito educativo.

La "Strategia europea in materia di dati" delineata dalla Commissione Europea (2020) si configura come un quadro strategico

mirato a plasmare l'approccio del continente europeo nei confronti dei dati nell'era digitale, nell'ambito delle priorità delineate per il periodo 2019-2024. Essa si propone di guidare l'Europa verso una posizione di rilevanza nel contesto digitale, fornendo un approccio coerente e innovativo alla gestione, all'accesso e all'utilizzo dei dati con l'obiettivo di sfruttare appieno il potenziale dei dati per sostenere l'innovazione, la competitività e la crescita economica, nonché per affrontare sfide sociali e ambientali. Attraverso un approccio inclusivo e collaborativo, la Commissione Europea mira a coinvolgere tutti gli attori interessati, compresi cittadini, imprese e settore pubblico, per plasmare un futuro digitale sostenibile e centrato sui valori europei. All'interno della citata Strategia, l'"European Data Act" (Commissione Europea, 2023) costituisce una legislazione entrata in vigore l'11 gennaio 2024 volta a istituire un quadro normativo unificato per la gestione dei dati a livello europeo. L'obiettivo principale è superare le barriere e agevolare la condivisione dei dati tra settori e paesi diversi. L'*European Data Act* si focalizza sull'apertura e la trasparenza dei dati per consolidare una governance dei dati equa ed efficiente. Questo processo mira a promuovere l'apertura e la trasparenza nelle decisioni, assicurando contemporaneamente elevati standard di sicurezza e rispetto della privacy in modo etico e conforme ai principi fondamentali dell'Unione Europea, inclusi quelli relativi alla tutela dei diritti individuali.

In tale contesto politico e normativo, è stato recentemente avviato un bando di finanziamento per la sperimentazione di politiche in materia di istruzione e formazione nell'ambito del Programma Erasmus+ (2023) per sostenere un'istruzione digitale inclusiva e di alta qualità. In tale avviso è stata collocata una terza priorità "Strategie di alfabetizzazione ai dati nell'istruzione primaria e secondaria", in aggiunta alla prima, che era incentrata a promuovere l'equilibrio di genere negli ambiti di studio e professionali legati alle TIC, e alla seconda, che si propone di promuovere le iniziative sul benessere digitale, già presenti nel Piano

d'azione per l'istruzione digitale 2021-2027. Il Programma Erasmus+ è chiamato a valutare e finanziare azioni per:

1. identificare le strategie esistenti per promuovere la *data literacy*, sia come costruito autonomo che come parte delle più ampie strategie di digitalizzazione delle istituzioni nell'istruzione primaria e secondaria;
2. mappare le strategie di *data literacy* nei curricula nazionali e identificare, in particolare, in che misura e in che modo la *data literacy* è affrontata nei curricula dei diversi Stati membri e come i curricula vengono modificati per riflettere tale necessità.

Gli esiti di tali iniziative si concretizzano nella formulazione di pratiche avanzate per incorporare l'alfabetizzazione ai dati all'interno della strategia generale di istruzione digitale degli Stati membri. Particolare attenzione dovrebbe essere rivolta a come gli insegnanti sono supportati nello sviluppare competenze di *data literacy* per l'insegnamento, sia per migliorare l'insegnamento in sé sia quale capacità di sviluppare tale competenza per i loro studenti. Inoltre, vengono proposte raccomandazioni a livello europeo e nazionale, considerando i fattori di successo e le sfide nel processo di sviluppo della data literacy nei programmi educativi della scuola primaria e secondaria.

Nonostante il processo di riflessione istituzionale in atto a livello europeo, abbiamo già sottolineato come il concetto di *data literacy* nel contesto dei quadri di competenze europei, quali il DigComp 2.2, rimanga associato soprattutto all'information e alla media literacy e orientato alle funzioni pratico-strumentali a livello individuale. Per superare tale limite, è necessario inserire tali competenze composite nel contesto più ampio della "cittadinanza digitale", enfatizzando il benessere digitale, lo sviluppo del pensiero critico e promuovendo un uso consapevole e attivo delle tecnologie digitali. La scuola italiana può affrontare questo tema

mediante l'insegnamento trasversale dell'educazione civica, introdotto dalla Legge 92/2019. Tale legge identifica la cittadinanza digitale come uno dei tre pilastri concettuali dell'educazione civica, insieme allo sviluppo sostenibile e allo studio della Costituzione, definendola come "la capacità di un individuo di utilizzare consapevolmente e responsabilmente i mezzi di comunicazione virtuali". Per quanto riguarda le "Linee guida per l'insegnamento dell'educazione civica", inviate nel 2020 in allegato alla già citata Legge 92/2019, il legislatore riconosce e specifica anche l'importanza di abilità come "analizzare, confrontare e valutare criticamente la credibilità e l'affidabilità delle fonti di dati, informazioni e contenuti digitali", recependo l'uso e la comprensione dei dati in funzione dello sviluppo dell'*information literacy*. Si ribadisce l'importanza del contesto scolastico per facilitare lo sviluppo di competenze tali da permettere una piena e consapevole partecipazione alla comunità, a partire dall'infanzia:

L'approccio e l'approfondimento di questi temi dovrà iniziare fin dal primo ciclo di istruzione: con opportune e diversificate strategie, infatti, tutte le età hanno il diritto e la necessità di esserne correttamente informate. Non è più solo una questione di conoscenza e di utilizzo degli strumenti tecnologici, ma del tipo di approccio agli stessi; per questa ragione, affrontare l'educazione alla cittadinanza digitale non può che essere un impegno professionale che coinvolge tutti i docenti contitolari della classe e del Consiglio di classe (MIUR, 2020, p. 2).

Il ruolo specifico del primo ciclo d'istruzione è quello di favorire lo sviluppo delle competenze di base, promuovendo la comprensione dei linguaggi e dei codici che costituiscono il fondamento della nostra cultura, con una prospettiva aperta alle altre culture presenti nel contesto sociale e all'utilizzo consapevole dei nuovi media. Attraverso le "Indicazioni nazionali per il curriculum della scuola dell'infanzia e del primo ciclo d'istruzione"

(MIUR, 2012), è possibile intravedere riferimenti impliciti alla *data literacy*, concepita in una prospettiva ampia e collegata alle competenze civiche. Nel “Profilo delle competenze al termine del primo ciclo di istruzione”, infatti, in riferimento all’ambito matematico-scientifico, si prevede che uno studente sviluppi la capacità di “analizzare dati e fatti della realtà e di verificare l’attendibilità delle analisi quantitative e statistiche proposte da altri”. Parallelamente, nell’ambito tecnologico-digitale, si sottolinea l’importanza dell’uso consapevole dei mezzi di comunicazione “per ricercare e analizzare dati ed informazioni, per distinguere informazioni attendibili da quelle che necessitano di approfondimento, di controllo e di verifica” (MIM, 2024). I due elementi, logico-razionale e critico-tecnologico, idealmente si dovrebbero integrare per preparare gli studenti che completano la scuola primaria ad affrontare problemi e situazioni reali, con la consapevolezza dei limiti delle strutture digitali e della natura dei contenuti che esplorano. Questa prospettiva risulta coerente con lo sviluppo di competenze nell’utilizzo dei dati all’interno di un pensiero *evidence-based*, che mira a dotare gli individui della capacità di distinguere argomentazioni valide ed erranee fondate sui dati e di formularne di corrette (Calvani, 2013; Ridsdale et al., 2015; Vahey, Yarnall, Patton, Zalles, Swan, 2006). La competenza nei confronti degli strumenti matematici e statistici è, quindi, connessa alla loro funzione interpretativa del contesto, nonché alla responsabilità circa l’efficacia del metodo d’indagine per la comprensione dei fenomeni complessi (MIUR, 2018; Raffaghelli, 2019).

Un recente rapporto dell’Osservatorio Italiano sui Media Digitali, centro nazionale affiliato all’Osservatorio Europeo sui Media Digitali, ha riscontrato un notevole interesse per la tematica della privacy e dei dati tra gli studenti e gli insegnanti italiani (IDMO, 2023). Ciò suggerisce che si tratta di una questione sensibile e che la consapevolezza politica su questo tema è in crescita.

Malgrado questi segnali, la *data literacy* non ha ancora ricevuto

un'adeguata formalizzazione e legittimazione a livello curricolare. Ferrari e Pasta (2023) hanno sintetizzato l'approccio italiano all'educazione digitale, in una visione in cui la *data literacy* è ricondotta a componente dell'*information literacy*. Da un lato, in una concezione delle competenze digitali come *literacies* dinamiche applicate al contesto, superando la mera competenza tecnica, lo Stato italiano adotta un approccio trasversale alle discipline anziché curricolare, enfatizzando l'alfabetizzazione basata sull'educazione al pensiero critico e alla responsabilità digitale. Dall'altro lato, il Ministero dell'Istruzione incoraggia ogni istituzione scolastica, indipendentemente dall'ordine e dalla fascia di età degli studenti, a contribuire alla creazione di curriculum per l'educazione digitale, che in un secondo tempo potranno essere integrati nel sistema educativo nazionale, promuovendo un coinvolgimento attivo delle scuole ed evitando di fornire soluzioni preconfezionate.

La necessità di raccogliere la sfida di implementare percorsi istituzionali che permettano lo sviluppo di competenze legate ai dati emerge anche nel rapporto del gruppo di lavoro sul big data del MIUR (2016), che fornisce un'analisi sulle opportunità e le sfide che i big data presentano per il sistema formativo italiano. Questo documento propone azioni specifiche a diversi livelli, a partire dalla Scuola secondaria, per sviluppare competenze propedeutiche ai big data tra gli studenti italiani, in relazione a svariati ambiti quali pensiero computazionale, Internet of Things (IoT) e i diritti in rete, e successivamente fino agli istituti di formazione superiore che sono chiamati a sviluppare percorsi specialistici legati alla *data science*.

L'ampia disponibilità di dati online, se da un lato offre opportunità di trasparenza e partecipazione democratica, dall'altro presenta sfide legate alla finalità e alle modalità con cui è condotta l'alfabetizzazione. Dalla metà degli anni 2000 in avanti, i governi a livello globale hanno avviato un'ampia pubblicazione online di dati, contrassegnando l'inizio del movimento mondiale verso gli

open data, i dati aperti (Chignard, 2013, cfr. capitolo 1, par. 1.2.2). L'obiettivo primario del movimento *open data* consiste nella promozione dell'accessibilità e della disponibilità dei dati al pubblico. Tale iniziativa sollecita governi, organizzazioni e istituzioni a condividere i propri dati in un formato agevolmente accessibile, utilizzabile e comprensibile, con l'intento di stimolare una maggiore trasparenza, collaborazione e innovazione. Questo è stato interpretato come il primo passo in un processo di trasparenza a sostegno dei governi democratici: le iniziative sui dati aperti promettono una società più trasparente, una democrazia partecipativa più profonda e la capacità di generare valore dai dati (Huijboom, Broek, 2011). Tuttavia, le gravi disuguaglianze sociali diffuse a livello globale, riflesse direttamente nel livello di istruzione della popolazione, costituiscono un terreno fertile perché continui ad aumentare il numero di individui che non possiedono ancora le competenze per utilizzare questi strumenti. Il movimento sottolinea l'importanza di garantire che l'accesso alle informazioni non sia limitato solo a coloro che possiedono già un elevato livello di *data literacy*. Affrontare le disuguaglianze educative diventa cruciale per realizzare a pieno le promesse di trasparenza e partecipazione democratica associate all'iniziativa degli *open data*. Gli sforzi per promuovere la *data literacy* devono, pertanto, essere guidati da una comprensione più approfondita di ciò che tale promessa implica, rispetto a quanto effettivamente realizza.

2.3.2 *Uno strumento a favore dei cittadini nella società dei dati*

In una visione in cui la *data literacy* viene ricondotta all'opportunità di considerare il dato solamente quale risorsa da cui estrarre valore (economico, tecnologico, sociale, etc.) emerge il rischio di trascurare le questioni relative alla prospettiva mediaeducativa. Una concezione puramente basata sulle competenze

tecniche quali “saper utilizzare o analizzare i dati” non si preoccupa né dell’origine dei dati né dello scopo della loro analisi né degli attori che ne curano la gestione (Gray et al., 2018). Letouze (2016), del gruppo Data Pop Allianz, sostiene che restringere all’ambito tecnico-strumentale la portata della *data literacy* possa “rafforzare e perpetuare, piuttosto che sfidare e cambiare, le strutture e le dinamiche di potere prevalenti” (para 3). Al contrario, una concezione dell’alfabetizzazione ai dati in cui questi sono utilizzati da e per le persone, con lo scopo di raggiungere i propri obiettivi e partecipare pienamente alla comunità e alla società, richiede la valutazione degli strumenti e la trasparenza degli scopi alla luce di norme sociopolitiche condivise, soprattutto di matrice etica e filosofica. Ad esempio, un individuo “alfabetizzato ai dati” potrebbe scegliere di non utilizzare o diffondere quelli raccolti in modo non etico o che potrebbero arrecare danno a qualcuno.

Prima di procedere oltre nel definire il ruolo che può rivestire la *data literacy* in educazione, risulta cruciale definire il concetto di alfabetizzazione, un termine che è oggetto di dibattito in diversi contesti (Cappello, 2017; Ranieri, 2018; Street, 2003). Accostare il termine “literacy” ad un ambito circoscritto, come è avvenuto per la *media* o la *data literacy*, evidenzia, da un lato, l’importanza di attirare su di esso l’attenzione, sottolineando che, così come l’alfabetizzazione tradizionale⁵, anche questa forma di competenza

5 Anche se con il termine *literacy* (alfabetizzazione) si intende comunemente l’insegnamento e l’apprendimento delle abilità minime della comunicazione scritta, questo non può essere ridotto alle competenze di lettura, scrittura e far di conto. L’Organizzazione delle Nazioni Unite per l’Educazione, la Scienza e la Cultura (UNESCO) definisce la *literacy* come un concetto complesso e dinamico: “identificazione, comprensione, interpretazione, creazione e comunicazione in un mondo sempre più digitale, mediato dal testo, ricco di informazioni e in rapida evoluzione” (UNESCO, 2024, para 2), senza menzionare esplicitamente la lettura e la scrittura. In tale visione, si sottolinea un processo continuo

risulta fondamentale. Dall'altro, implica anche considerazioni di portata più ampia: essere completamente alfabetizzati non riguarda solo abilità individuali, ma implica anche una partecipazione sociale e culturale consapevole, che permetta di riconoscere le dinamiche di potere in azione nella società moderna (Buckingham, 2022). In modo analogo, l'UNESCO (2005, 2024) ha modificato la sua prospettiva, passando da un concetto di alfabetizzazione "funzionale", incentrato sulle competenze di base e la sopravvivenza, a una prospettiva socioculturale che enfatizza l'autonomia personale e sociale e che considera la competenza come integrata in contesti specifici che richiedono pratiche diverse, compresa la sfera digitale (Rivoltella, 2020; Wickens, Sandlin, 2007).

Dal punto di vista educativo, l'alfabetizzazione spesso viene concepita con una valenza intrinsecamente positiva: un processo di evoluzione storica e miglioramento delle condizioni umane. Tuttavia, un'analisi più approfondita del ruolo storico della scrittura, dell'alfabetizzazione e degli sforzi per contrastare l'analfabetismo rivela sfumature più complesse, fortemente influenzate dalle circostanze sociopolitiche (Letouzé, 2016).

L'antropologo francese Lévi-Strauss (1955) suggerisce un'interpretazione controversa del processo di alfabetizzazione, quale mezzo per esercitare maggiore controllo sulle masse, piuttosto che come uno strumento di progresso intellettuale per favorire l'emancipazione individuale e la partecipazione alla società (Ricciardi, 2014). Lévi-Strauss sostiene che il movimento per l'istruzione obbligatoria delle masse aveva principalmente una funzione di facilitare il controllo e l'asservimento degli individui da parte di coloro che detenevano il potere. Secondo questa interpretazione, le campagne di alfabetizzazione del XIX secolo sono state

di apprendimento che abilita gli individui a perseguire i propri obiettivi e partecipare in modo attivo alla comunità e alla società.

sviluppate di pari passo ad altre iniziative, quali l'estensione del servizio militare e l'ampliamento delle città a discapito delle campagne durante la Rivoluzione Industriale, con il fine di radicare e ampliare le strutture di potere emergenti, creando individui conformi alle esigenze degli Stati nazionali e dell'economia capitalista in formazione.

L'alfabetizzazione riveste un ruolo centrale nel lavoro educativo promosso dal famoso pedagogista brasiliano Freire (2002), quale strumento indispensabile per la presa di coscienza sulla propria condizione sociale e politica e leva per l'emancipazione collettiva (Fiorucci, Vaccarelli, 2022). Secondo la sua prospettiva, l'alfabetizzazione non è intesa come apprendimento meccanico di procedure per la lettura e la scrittura, ma come processo per sviluppare il pensiero critico per l'interpretazione e la "lettura del mondo" (Freire, 2002). Tale processo di alfabetizzazione critica segue tre fasi sequenziali:

1. *Ricerca*: si studia il contesto in cui vivono le persone al fine di determinare il lessico comune e le parole più importanti per il gruppo di persone che parteciperanno al percorso di alfabetizzazione.
2. *Tematizzazione*: si identificano i temi significativi e si contestualizzano, collegando il processo di alfabetizzazione alla realtà quotidiana e sociale dei futuri alfabetizzati.
3. *Problematizzazione*: si analizzano in modo critico i temi identificati, incoraggiando i partecipanti a esplorare le cause e gli effetti delle situazioni problematiche presenti nella loro vita.

Una fase aggiuntiva, chiamata "sistematizzazione delle esperienze", è stata incorporata più tardi, in seguito alla diffusione e all'implementazione del modello di Freire in diverse nazioni (Tygel, Kirsch, 2016). Questa fase implica il processo collettivo di costruzione della conoscenza mediante l'esplorazione dei temi

che emergono dall'esperienza vissuta dalla comunità coinvolta nel percorso di alfabetizzazione.

Un parallelismo può essere tracciato tra il concetto di analfabetismo e le lacune contemporanee emergenti dalla società sempre più orientata ai dati, che si aggiungono alle criticità sia sociali che tecniche derivanti dal *digital divide* (Pasta, Rivoltella, 2022; Ranieri, 2019b). Vengono così chiamati “oppressi digitali” coloro che mancano degli strumenti culturali necessari per sfruttare le opportunità offerte dalla rete, trovandosi così esclusi dai processi di emancipazione culturale e personale (Ravanelli, Ceretti, 2017). Inoltre, è stato precedentemente evidenziato anche il divario che può esistere tra l'utilizzo dei dati e la loro comprensione. Quando si adotta questa prospettiva, oltre alle strategie pratiche e alle competenze operative, diventa essenziale integrare elementi di riflessione critica e consapevolezza riguardo al ruolo dei dati nella società e nella vita personale (Seymoens et al., 2020).

Tygel e Kirsch (2016) hanno proposto un'analisi della *data literacy* basata sul metodo di Freire, in cui si evidenzia la relazione tra il metodo dell'alfabetizzazione critica e la *data literacy* (Tabella 2.1).

Fase	Alfabetizzazione critica tradizionale	Alfabetizzazione critica ai dati	Risultato
<i>Ricerca</i>	Comprensione del contesto educativo e scoperta dei temi socialmente rilevanti in una specifica realtà		Indagine sull'universo lessicale
<i>Tematizzazione</i>	Codifica e decodifica delle parole e comprensione del loro significato sociale	Codifica dei temi in dati esistenti (o non esistenti) e decodifica per comprenderne la natura	Temi e parole codificati in immagini, film o dati
<i>Problematizzazione</i>	Ricerca delle contraddizioni che circondano i temi decodificati e demistificazione della realtà	Scoperta e analisi della non neutralità dei dati: quali aspetti sono espressi dai dati e quali sono nascosti?	Prospettiva critica sui temi

<i>Sistematizzazione</i>	Organizzazione, interpretazione e presentazione dell'esperienza vissuta	Organizzare e interpretare la realtà attraverso i dati e comunicare le scoperte	Prodotti di comunicazione
--------------------------	---	---	---------------------------

Tab. 2.1: Elementi in comune tra alfabetizzazione freiriana e data literacy
(adattato da Tygel, Kirsch, 2016)

Gli autori sottolineano, pertanto, il carattere emancipatorio della *data literacy*, quale strumento necessario all'interpretazione del contesto dei dati, alla discussione dei concetti di senso comune (come la presunta neutralità o assertività dei dati) e nella costruzione di nuove prospettive per l'uso e la comunicazione attraverso i dati, in una direzione che supporti la partecipazione civica e l'empowerment delle persone.

Oltre all'ambito pedagogico, il termine *data literacy* è gradualmente entrato anche nei riferimenti di esperti, rappresentanti politici e sostenitori che iniziano a considerare cosa sarebbe necessario per consentire ai cittadini di fare un migliore uso della vasta quantità di dati a loro disposizione. Un rapporto delle Nazioni Unite ha sottolineato l'impellente necessità di una *global data literacy* al fine di innescare una "rivoluzione dei dati" per il raggiungimento degli Obiettivi di Sviluppo Sostenibile, volendo cogliere "l'opportunità di migliorare i dati essenziali per il processo decisionale, la responsabilità e la risoluzione delle sfide di sviluppo" (Data Revolution Group, 2014a, p. 5), finalità ribadita anche da Letouzé (2016). Nel contesto delle discussioni pubbliche in corso sulla cosiddetta "rivoluzione dei dati", Bhargava e colleghi (2015) del gruppo Data Pop Allianz hanno individuato tre temi emergenti: il controvalore economico, il fattore di empowerment e la partecipazione civica nella società digitale.

In primo luogo, per gli enti privati e il mondo industriale la *data literacy* dei lavoratori aumenta l'impatto economico dei big data, dei small data e degli *open data*: "Le competenze legate ai dati sono ormai indispensabili in quasi ogni posizione all'interno di

qualsiasi organizzazione. Tuttavia, data la loro scarsa diffusione in molte aziende, risulta complesso individuare da quale punto iniziare per potenziare queste abilità. Sugeriamo di focalizzarsi sui concetti fondamentali” (Harvard Business Review, 2020, para 4).

Considerato che le imprese mirano a capitalizzare il potenziale valore commerciale generato dai dati, i dipendenti con competenze in *data science* sono diventati molto preziosi nell’economia odierna. Le imprese hanno iniziato a investire in formazioni della loro forza lavoro finalizzate a sviluppare competenze per creare prodotti e servizi innovativi basati sui dati, per aumentare i propri margini di profitto.

In secondo luogo, gli operatori per lo sviluppo socioculturale si riferiscono alla *data literacy* come fattore di empowerment che consente alla popolazione locale di comprendere e risolvere problemi del proprio territorio. Lavorare con dati molto dettagliati o circoscritti dal punto di vista geografico, diversamente dalle statistiche spesso aggregate a un livello più generale, può aumentare l’efficienza e la resilienza degli attori e delle comunità locali nel risolvere i problemi che affrontano, sviluppando una visione più critica ed estendendo il raggio d’azione non solo ai dati esistenti, ma anche a quelli finora mancanti. L’Agenzia delle Nazioni Unite per i Rifugiati (UNHCR) collabora con un vasto network di stakeholder, che includono governi, sistemi statistici a livello nazionale, regionale e internazionale, agenzie delle Nazioni Unite, organizzazioni della società civile, istituzioni accademiche e altri ancora per trasformare il modo in cui vengono raccolti e utilizzati i dati, massimizzandone il potenziale. Tra gli altri, l’UNHCR insieme a Microdata Library ha sviluppato lo strumento aperto “Refugee Data Finder”⁶ per semplificare la consultazione delle statistiche sui rifugiati degli ultimi 23 anni, suddivise per paese

6 Lo strumento *Refugee Data Finder* di UNHCR è disponibile all’indirizzo <https://www.unhcr.org/refugee-statistics/>, oltre che attraverso una versione app per iOS e Android.

di asilo e di origine grazie all'ausilio di mappe interattive, report e altre pubblicazioni.

Infine, da un punto di vista politico e di partecipazione civica, lo sviluppo della *data literacy* dei cittadini può sollecitare i governi ad essere responsabili e trasparenti, migliorando l'accesso degli *open government data* e contribuendo a colmare un divario digitale sempre crescente. Gli attivisti per i dati aperti hanno persuaso molti governi e organizzazioni *non profit* a rendere i propri dati accessibili, portando a numerosi esempi positivi di utilizzo dei dati per generare cambiamenti concreti nella vita delle persone. Il focus del dibattito ora si sposta dal “se” rendere pubblici i dati al “come” renderli più aperti, standardizzati e significativi (Data Revolution Group, 2014b). La produzione di dati disaggregati basati su genere, etnia, disabilità o età è ritenuta di valore inestimabile per sostenere la responsabilità e informare politiche fondate sul “progetto umano” (Floridi, 2020), per realizzare un esempio di società equa e sostenibile. Inoltre, la gestione dell'ecosistema dati può essere configurata in modo da riflettere i valori culturali e i principi delle comunità, assicurando un maggiore controllo e sovranità decisionale riguardo all'utilizzo e all'interpretazione dei dati. Il processo di riappropriazione, noto come “decolonizzazione”, si attua quando le comunità locali sostituiscono le norme esterne con sistemi e prospettive proprie, influenzando così la definizione, la raccolta e l'utilizzo dei dati in maniera congruente con la propria identità e le proprie finalità (Carroll, Rodriguez-Lonebear, Martinez, 2019). Infine, la *data literacy* può aiutare le porzioni svantaggiate della società civile a individuare le violazioni dei diritti, alimentare le azioni di *data journalism*⁷ e

7 Il *data journalism* è un processo mediante il quale l'analisi e la presentazione dei dati vengono impiegate per informare e coinvolgere il pubblico (Royal, Blasingame, 2015). Sebbene affondi le sue radici nell'uso dei big e open data, i prodotti derivati da questa pratica offrono un coinvolgimento maggiore delle semplici rappresentazioni dei dati, at-

stimolare il coinvolgimento dei cittadini negli sforzi di trasparenza e anticorruzione.

In precedenza, è stato illustrato come numerosi preconcetti o concetti che affondano le radici nel senso comune considerati “verità” sono formulati sulla base dei dati e che acquisire una comprensione approfondita su come questi dati siano stati generati permette uno sguardo critico su tali rappresentazioni (Maggino, 2017). Un esempio è costituito dal concetto di Prodotto Interno Lordo (PIL), generalmente utilizzato per fare confronti sul “valore economico” di diversi paesi. Sebbene il PIL stesso sia considerato la misura più importante dell’economia di un paese, esso non tiene conto della distribuzione del reddito o delle conseguenze ambientali dello sviluppo economico (Tygel, Kirsch, 2016).

Comprendere in che modo la conoscenza viene prodotta, valutata e utilizzata rientra nell’ambito dell’epistemologia. La riflessione epistemologica è rilevante anche nel contesto della *data literacy* perché ci aiuta a capire come gli esseri umani attribuiscono ai dati un certo tipo di valore conoscitivo e a valutare la loro validità e affidabilità. Il contributo di Špiranec, Kos e George (2019) definisce tre livelli nella critica ad una concezione epistemologicamente scorretta dei dati: dati come *autorità*, dati come *riduzione* e dati come *ideologia*. Nel primo caso, i dati sono trattati come una forma superiore di evidenza poiché intrinsecamente deterrebbero tutte le risposte, a prescindere da come siano stati generati. Ne deriva una concezione dei dati, qualora mediati da rappresentazioni grafiche di fenomeni complessi, che conduce a visioni riduttive e semplificate del mondo, accettando tali pro-

traverso la personalizzazione e il contributo degli utenti e sfruttando le tecniche avanzate di sviluppo web e programmazione. Per conoscere alcuni tra i progetti più significativi è possibile consultare il premio annuale Sigma Awards, il riconoscimento che celebra i contributi più efficaci nel campo del *data journalism* internazionale (<https://www.sigmaawards.org/>).

spettive senza indagare cosa sia stato incluso o escluso nel processo scientifico che ha generato la sintesi. Infine, i dati, a causa della loro intrinseca non neutralità (Borgman, 2017; Kitchin, 2014), potrebbero essere interpretati come veicoli di un'ideologia. Di conseguenza, anche le affermazioni basate sui dati, dotate di una "seducente precisione e oggettività" (Tygel, Kirsch, 2016, p. 16), spesso possono nascondere intenzioni, credenze e valori legati a ciò che si desidera dimostrare.

La considerazione del valore epistemologico attribuito ai dati getta luce sulla complessità e sulla pervasività delle ideologie nel contesto contemporaneo, influenzando non solo la produzione e l'interpretazione dei dati, ma anche la comprensione più ampia del nostro mondo. Il termine *dataism*, utilizzato dalla studiosa olandese van Dijck (2014), si riferisce a un'ideologia che implica una diffusa fiducia nell'oggettività della quantificazione e nel potenziale di tracciamento di tutti i (meta)dati, nonché la fiducia negli agenti istituzionali che raccolgono, interpretano e condividono i (meta)dati raccolti dai social media e delle proprie piattaforme. Con le parole di Davies (2022, p. 148), possiamo pensare al *dataism* come una "sovrastima delle potenzialità dei dati, suggerendo che questi abbiano intrinsecamente proprietà trasformative: se si producono e si rendono accessibili più dati e si insegna a più persone ad analizzarli, si creeranno maggiori opportunità di realizzare ambizioni".

È un'ideologia (o perfino una religione, nella versione enfatizzata del filosofo e futurologo Harari, 2016) radicata in affermazioni ontologiche ed epistemologiche problematiche ed è caratterizzata dalla fiducia nell'oggettività della quantificazione e nel potenziale di tracciamento di tutti i dati. Questa ideologia è associata alla graduale normalizzazione della datificazione, come nuovo paradigma nella scienza e nella società, e ha conosciuto una grande diffusione, perché sono ormai assai numerose le persone che, nelle proprie azioni quotidiane, affidano i loro dati personali a varie enti istituzionali e non, tra cui governi, aziende e

università, spesso ingenuamente o inconsapevolmente. I potenziali rischi per gli individui includono la perdita della privacy personale, la manipolazione e lo sfruttamento dei dati personali, gli effetti limitanti sulla libertà di espressione dovuti alla sorveglianza digitale costante e le controversie etiche legate alla raccolta e uso dei dati personali (Zuboff, 2019). La datificazione, inoltre, non impatta su tutti i cittadini in maniera uniforme: genere, etnia, livello socioeconomico, religione sono fattori che influenzano molto la portata delle sue implicazioni. Nel contesto dell'etica dei big data e dei sistemi automatizzati, la politologa Eubanks (2018) ha sollevato preoccupazioni significative riguardo alle possibili ingiustizie derivanti dall'impiego dei dati e degli algoritmi, evidenziando il rischio intrinseco di discriminazione nei confronti delle minoranze attraverso il processo di automazione decisionale. Particolare enfasi è posta sulla carenza di controllo esterno in questi procedimenti, oltre che sul problema dell'opacità degli algoritmi, descritti come "black box" in virtù della mancanza di trasparenza nel processo che guida le decisioni.

In contrapposizione a queste criticità e per definire percorsi etici attraverso un mondo che si sta rapidamente datificando, è stato sviluppato il concetto di *data justice* quale "equità nel modo in cui le persone sono rese visibili, rappresentate e trattate a seguito della produzione di dati digitali" (Taylor, 2017, p.1). La ricercatrice olandese Taylor propone tre pilastri per la concezione della giustizia in materia di dati: a) *(in)visibility*, b) *(dis)engagement with technology*, c) *non-discrimination* (figura 2.3). Questi pilastri integrano diritti e libertà sia in positivo che in negativo, sfidando così le basi delle attuali normative sulla protezione dei dati e la convinzione molto diffusa, ma anche certamente ideologica, che la visibilità prodotta dai dati sia un vincolo inevitabile del contratto sociale della realtà contemporanea.



Fig. 2.3: Framework della data justice (adattato da Taylor, 2017)

Negli ultimi anni, si è osservato un notevole aumento di iniziative volte a influenzare le pratiche associate ai big data e a innescare una trasformazione dal basso degli elementi più controversi (Raffaghelli, 2023). Il concetto di *data activism*, o attivismo dei dati, identifica una vasta gamma di strategie che comprendono sia azioni proattive di coinvolgimento che tattiche di resistenza nei confronti della massiccia raccolta di dati (Milan, Gutiérrez, 2015). Tra i numerosi esempi di *data activism*, si citano l'attivismo legato al movimento per i dati aperti che si concentra sull'accesso aperto e trasparente ai dati (Baack, 2015), promuovendo la condivisione e l'utilizzo dei dati in modo che siano liberamente accessibili a tutti, e il *data feminism*, che si concentra sull'analisi critica e sull'utilizzo dei dati per opporsi alle disuguaglianze e per promuovere l'equità di genere (D'Ignazio, Klein, 2020).

Anche l'ambito educativo rischia di sostenere la diffusione dell'ideologia del *dataism*, quando non si interroga sulle implicazioni etiche dei processi di raccolta e utilizzo dei dati e su come promuovere una cittadinanza digitale consapevole. Il ricercatore finlandese Mertala (2020), attraverso una revisione della letteratura scientifica sulla *datafication* in educazione, ha rilevato la presenza di due forme interconnesse di approcci taciti e dannosi. Per farlo, ha adottato il concetto di *hidden curriculum* (o "curriculum nascosto"), un'espressione che, come noto, si riferisce ai contenuti implicitamente appresi ma non consapevolmente compresi come tali (Kentli, 2009), quale strumento euristico per esaminare le pratiche quotidiane legate ai dati nell'istruzione scolastica. Innanzitutto, l'autore mostra come nelle pratiche quotidiane associate ai dati nel settore educativo viene trasmessa agli studenti un'immagine dei dati in termini di "autorità cognitiva". Tale concetto, elaborato da Wilson (1983), si riferisce a una fonte di informazione secondaria, non derivata dall'esperienza, che le persone ritengono credibile e legittima. Ad esempio, in campo educativo, i dati sugli esiti di apprendimento degli studenti possono essere utilizzati come misura diretta delle prestazioni degli insegnanti, producendo una semplificazione del processo valutativo. Questo approccio pedagogico involontario può spingere gli studenti a sopravvalutare l'accuratezza dei dati, utilizzati per scopi diversi da quelli per cui sono stati raccolti. Ciò può portare a una fiducia eccessiva nei sistemi data-based, poiché gli studenti potrebbero non essere consapevoli delle limitazioni dei dati o delle possibili distorsioni nella loro interpretazione al di fuori del contesto originale di raccolta. In secondo luogo, a scuola si assiste spesso ad una normalizzazione delle onnipresenti abitudini di raccolta dei dati digitali, conseguenza anche della graduale piattaformaizzazione dell'ambito educativo (cfr. capitolo 1, par. 1.2.1). Questa pratica si traduce nel fatto che, all'aumentare dell'informatizzazione delle pratiche scolastiche attraverso le infrastrutture e i portali digitali, la *datafication* e la sorveglianza dei dati appaiono

sempre più naturali e accettabili agli occhi degli studenti, riducendo così la loro capacità di agire e di contestare tale abitudine. Queste due forme di apprendimento implicito e fuorviante in ambito scolastico possono produrre, secondo Mertala (2020), un'(an)alfabetizzazione sui dati, caratterizzata da una comprensione acritica e unidimensionale degli stessi e del loro utilizzo da parte degli studenti.

È stato illustrato come in educazione si possano adottare strategie diverse, e talvolta discutibili, nei confronti del progressivo avvento dei dati nei propri territori e in relazione agli studenti. Per questo diventa lecito interrogarsi sul ruolo che possono giocare la scuola e le istituzioni in questo scenario, per arginare le pratiche che non salvaguardano l'interesse degli allievi e dei cittadini. In che modo il settore educativo può quindi combattere la diffusione del *dataism* e le conseguenze della progressiva datificazione e piattaformaizzazione dell'educazione?

L'istruzione svolge un ruolo fondamentale nel promuovere la consapevolezza dei diritti e delle responsabilità dei cittadini all'interno di una democrazia, tuttavia l'introduzione di programmi volti a sviluppare la *data literacy* degli studenti non può essere considerata la panacea alle criticità attuali.

Pangrazio e Sefton-Green (2020) hanno delineato le possibili strategie di risposta alla *datafication*, al fine di formulare interventi che considerino la complessità dei diversi fattori interconnessi: prospettiva normativa, strategica ed educativa. Una prima risposta è di tipo legale-politico e ruota attorno al tema dei diritti, con una prospettiva più indirizzata a limitare i danni che ad affrontare il problema nel suo complesso. In questo ambito, che comprende sia i diritti umani che i diritti relativi alla proprietà, spesso il linguaggio delle normative offre una strategia euristica che tenta di giustificare, razionalizzare e legittimare la relazione tra l'individuo e la società, cercando di rimediare a posteriori alle questioni sulla privacy e alle asimmetrie strutturali esistenti. Inoltre, alcune risposte normative tendono a contrapporsi alla motivazione fon-

damentale dell'uso della rete e dei social media, ossia condividere e connettersi con gli altri. La seconda risposta è di tipo tattico-tecnologico e consiste nella progettazione di una serie di soluzioni tecniche per avviare processi alternativi alla *datafication*, contrastando il costante flusso dei dati immesso nell'economia digitale e creando piattaforme e algoritmi più etici. Spesso, le soluzioni tecniche emergono in risposta alla mancanza di regolamentazione e si basano sull'idea che una maggiore trasparenza sul trattamento dei dati motiverà i cittadini a utilizzare una serie di strumenti a propria tutela (es. per anonimizzare e crittografare le informazioni personali e migliorare il "consenso informato"). Tali strategie, da un lato, spostano la responsabilità della protezione dei dati sull'individuo e talvolta compromettono l'esperienza su Internet (Acar, Eubank, Englehardt, Juarez, Narayanan, Diaz, 2014); dall'altro, richiedono una manutenzione continua e un livello sofisticato di competenza tecnica. Infine, la risposta educativa si configura come l'ultima strategia per affrontare le complesse sfide connesse alla *datafication*, poiché anche l'affermazione dei diritti individuali o l'utilizzo di software per la tutela dei dati personali richiedono un adeguato livello di consapevolezza e comprensione delle dinamiche dell'economia dei dati. Anche per questo motivo, l'educazione più spesso intesa come intervento per favorire la *data literacy* si è affermata come la risposta prevalente. In questo contesto, per rispondere ai nuovi bisogni emersi l'ambito educativo può promuovere una concezione di equità e promozione della *data literacy* come parte di un dibattito più ampio sulla *data justice*, integrando considerazioni tecniche, etiche e sociopolitiche in modo strategico lungo tutto il curriculum (Raffaghelli, 2023). La pratica educativa può quindi sostenere sia gli studenti che gli educatori attraverso un percorso di partecipazione collettiva al dibattito nell'era dei dati, che consenta dapprima di padroneggiare strumenti critici di lettura e interpretazione, fino a giungere a riconoscere e avviare azioni civiche di *data activism*.

Al fine di esplorare la relazione tra dati, potere e contesto, il

quadro teorico della *data citizenship*, la cittadinanza dei dati, sviluppato da Carmi, Yates, Lockley e Pawlucz (2020), sottolinea la necessità che i cittadini possano esercitare la capacità di agire in modo critico e attivo relativamente ai dati. Il framework evidenzia l'importanza dell'agency dei cittadini, in un momento storico di normalizzazione della datificazione e delle decisioni basate sugli algoritmi nella società contemporanea (figura 2.4). A tal fine gli autori hanno delineato tre livelli progressivi di partecipazione significativa alla propria comunità: saper pensare, saper agire e poter partecipare attraverso i dati.



Fig. 2.4: Framework della data citizenship (elaborazione da Carmi et al., 2020)

Nella prima area di competenza della *data citizenship*, il *data thinking* (pensare i dati), si sottolinea l'importanza delle capacità di pensiero critico dei cittadini in relazione ai dati nella loro vita quotidiana. In questo livello sono incluse pertanto le iniziative per lo sviluppo di capacità e abilità critiche nell'uso dei dati. Il secondo livello, *data doing* (agire con i dati), è un concetto che si configura come un insieme di competenze specifiche per affrontare le sfide collegate all'utilizzo delle informazioni, come la gestione critica ed etica dei dati, inclusi i processi come la loro

creazione e la cancellazione. La terza area del framework, *data participation* (partecipare con/attraverso i dati), pone l'accento sull'importanza del coinvolgimento proattivo dei cittadini nella società e il valore della dimensione di promozione dell'alfabetizzazione. Tale area amplia e integra i primi due ambiti per permettere ai cittadini informati di utilizzare le loro competenze per partecipare attivamente alla comunità, cercando opportunità per esercitare i propri diritti e contribuire a modellare le loro esperienze collettive in relazione ai dati, anche attraverso la diffusione delle iniziative volte a favorire lo sviluppo della *data literacy* per diversi target. In questo ultimo livello, si collocano infine le iniziative per la *data justice* e del *data activism*, a favore di una rappresentazione equa e non distorta della società, e il movimento per l'*open data movement*, all'interno del quale si incentivano pratiche di utilizzo trasparente e democratico dei dati pubblici.

In conclusione, nonostante la *data literacy* emerga come uno strumento fondamentale per l'emancipazione e la partecipazione civica dell'era digitale, non le è stato ancora riconosciuto un ruolo formale all'interno delle policy e dei curricula in Italia, malgrado alcuni primi segnali di attenzione. Il diffondersi di posizioni ideologiche sulla presunta autorità dei dati e i problemi connessi alla datificazione progressiva della società rappresentano sfide significative ed urgenti che il mondo civile ha iniziato ad affrontare. A fianco di strumenti normativi e soluzioni tecniche, è possibile intravedere una possibile risposta attraverso l'educazione, con la costruzione di una cittadinanza consapevole e proattiva nei confronti dei dati, che possa offrire strumenti essenziali per affrontare in modo informato le sfide emergenti.

Riferimenti bibliografici

Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., Diaz, C. (2014, November). The web never forgets: Persistent tracking

- mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (pp. 674-689). New York: ACM.
- Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, 16(1), 3-9.
- Azzam, T., Evergreen, S., Germuth, A. A., Kistler, S. J. (2013). Data visualization and evaluation. *New Directions for Evaluation*, 2013(139), 7-32.
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2), 2053951715594634.
- Bateman, J. A. (2021). What are digital media? *Discourse, Context & Media*, 41, 100502.
- Berikan, B., Özdemir, S. (2020). Investigating “problem-solving with datasets” as an implementation of computational thinking: A literature review. *Journal of Educational Computing Research*, 58(2), 502-534.
- Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., Shoup, N. (2015). *Beyond data literacy: Reinventing community engagement and empowerment in the age of data*. Working Paper for Discussion.
- Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. Cambridge: MIT press.
- Buckingham, D. (2006). *Media education. Alfabetizzazione, apprendimento e cultura contemporanea*. Trento: Erickson.
- Buckingham, D. (2020). *Un manifesto per la media education*. Milano: Mondadori Università.
- Buckingham, D. (2022, maggio). *The uselessness of literacies*. Retrieved June 7, 2024, from <https://davidbuckingham.net/2022/05/16/the-uselessness-of-literacies/>
- Calvani, A. (2013). Evidence Based (Informed?) Education: neopositivismo ingenuo o opportunità epistemologica? *Form@ re-Open Journal per la formazione in rete*, 13(2), 91-101.
- Calvani, A., Fini, A., Ranieri, M. (2010). *La competenza digitale nella scuola: modelli e strumenti per valutarla e svilupparla*. Trento: Erickson.
- Calzada Prado, J., Marzal, M. Á. (2013). Incorporating data literacy

- into information literacy programs: Core competencies and contents. *Libri*, 63(2), 123-134.
- Cappello, G. (2017). Literacy, media literacy and social change. Where do we go from now? *Italian Journal of Sociology of Education*, 9(1), 31-44.
- Cappello, G. (2021). Bridging the gaps: Literacy, media literacy education, and critical digital social work. *J. Soc. & Soc. Welfare*, 48, 99-116.
- Carlson, J., Fosmire, M., Miller, C. C., Nelson, M. S. (2014). Determining data information literacy needs. *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, 23, 11-33.
- Carmi, E., Yates, S. J., Lockley, E., Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, 9(2), 1-22.
- Carretero, S., Vuorikari, R., Punie, Y. (2017). DigComp 2.1. *The Digital Competence Framework for Citizens. With eight proficiency levels and examples of use*. Bruxelles: Publications Office of the European Union.
- Carroll, S. R., Rodriguez-Lonebear, D., Martinez, A. (2019). Indigenous data governance: strategies from United States native nations. *Data Science Journal*, 18(31).
- Cazden, C., Cope, B., Fairclough, N., Gee, J., Kalantzis, M., Kress, G., ... Nakata, M. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard educational review*, 66(1), 60-92.
- Cheng, R., Dangol, A., Ello, F. M. T., Wang, L., Dasgupta, S. (2023, June). Concepts, practices, and perspectives for developing computational data literacy: Insights from workshops with a new data programming system. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 100-111). New York: ACM.
- Chignard, S. (2013). *A Brief History of Open Data*. Retrieved June 7, 2024, from <http://www.paristechreview.com/2013/03/29/brief-history-open-data/>
- Commissione Europea (2020). *Strategia europea in materia di dati*. Retrieved June 7, 2024, from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_it

- Commissione Europea (2023). *European Data Act*. Retrieved June 7, 2024, from <https://www.eu-data-act.com/>
- D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1), 6-18.
- D'Ignazio, C., Bhargava, R. (2015, September). Approaches to building big data literacy. In *Proceedings of the Bloomberg data for good exchange conference* (Vol. 6). Bloomberg.
- D'Ignazio, C., Klein, L. F. (2020). *Data feminism*. Cambridge: MIT press.
- Data Revolution Group (2014a). *A World That Counts: Mobilising the Data Revolution for Sustainable Development*. New York: United Nations Secretary-General's Independent Expert Advisory Group.
- Data Revolution Group (2014b). *Accessible Data: open data, accountability and data literacy*. United Nations Secretary-General's Independent Expert Advisory Group.
- Davies, H. C. (2022). Rescuing Data Literacy from Dataism. In *Data justice and the right to the city* (pp. 146-164). Edinburgh: Edinburgh University Press.
- Deahl, E. (2014). *Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments*. M.S. Thesis, Massachusetts Institute of Technology.
- Dencik, L., Hintz, A., Redden, J., Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. *Information, Communication & Society*, 22(7), 873-881.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.
- Fabbro, F., Gabbi, E. (2021). Alfabetizzazione critica ai dati e alle teorie del complotto. Un approccio media educativo. In P. Lucisano (Ed.), *Ricerca e Didattica per promuovere intelligenza comprensione e partecipazione, Atti del X Convegno SIRD 9-10 aprile 2021 – II TOMO* (pp. 68-83). Lecce: Pensa MultiMedia.
- Fabbro, F., Gabbi, E. (2024). Addressing conspiracy theories through Media and Data Literacy Education. An exploratory case study. *Journal of Media Literacy Education*, 16(1), 19-36.
- Ferrari, S., Pasta, S. (2023). Information Literacy: l'istanza educativa centrale per la declinazione delle competenze digitali. *Annali online della Didattica e della Formazione Docente*, 15(26), 170-192.

- Fiorucci, M., Vaccarelli, A. (Eds.). (2022). *Pedagogia e politica, in occasione dei 100 anni dalla nascita di Paulo Freire*. Lecce: Pensa Multimedia.
- Floridi, L. (2020). *Il verde e il blu: Idee ingenuie per migliorare la politica*. Milano: Raffaello Cortina.
- Freddano, M., Pastore, S. (Eds.). (2018). *Per una valutazione delle scuole oltre l'adempimento: Riflessioni e pratiche sui processi valutativi*. Milano: FrancoAngeli.
- Freire, P. (2002). *La pedagogia degli oppressi*. Ed. it. a cura di Linda Bimbi, Torino: EGA.
- Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, 16(1), 22-25.
- Grey, J., Gerlitz, C., Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, 5(2), 2053951718786316.
- Ha, S., Kim, S. (2023). Developing a conceptual framework for digital platform literacy. *Telecommunications Policy*, 102682.
- Habermann, T., Burton, N., Frender, K. (1998). Information arcology and data explorations: scientific content for multiple learning styles and environments. *Journal of Science Education and Technology*, 7, 235-247.
- Harari, Y. N. (2016). *Big data, Google and the end of free will*. Financial Times. Retrieved June 7, 2024, from <https://www.ft.com/content/50bb4830-6a4c-11e6-ae5b-a7cc5dd5a28c>
- Harvard Business Review (2020, Febbraio). *Boost Your Team's Data Literacy*. Retrieved June 7, 2024, from <https://hbr.org/2020/02/boost-your-teams-data-literacy>
- Hobbs, R. (2011). *Digital and media literacy: Connecting culture and classroom*. Apple Valley: Corwin Press.
- Hobbs, R. (2021). *Media literacy in action: Questioning the media*. Lanham: Rowman & Littlefield Publishers.
- Huijboom, N., Broek, T. Van Den. (2011). Open data: an international comparison of strategies. *European Journal of ePractice*, 1(12), 1-13.
- Huysman, M., Wulf, V. (2006). IT to support knowledge sharing in communities, towards a social capital analysis. *Journal of Information Technology*, 21(1), 40-51.
- IDMO. (2023). *Digital Media Literacy Gaps and Needs*. Italian Digital Media Observatory project. Retrieved June 7, 2024, from https://www.idmo.it/wp-content/uploads/2023/12/IDMO-Digital-Media-Literacy-Gaps-and-Needs-IT_finale_compressed.pdf

- Kentli, F. D. (2009). Comparison of hidden curriculum theories. *European Journal of educational studies*, 1(2), 83-88.
- Kitchin R. (2014), *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin, R., Lauriault, T. P., McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, 2(1), 6-28.
- Koenig, A. (2020). The algorithms know me and I know them: using student journals to uncover algorithmic literacy awareness. *Computers and Composition*, 58, 102611.
- Lee, S., Kim, S. H., Kwon, B. C. (2016). Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics*, 23(1), 551-560.
- Legge 20 agosto 2019, n. 92 - Introduzione dell'insegnamento scolastico dell'educazione civica. Gazzetta Ufficiale Serie Generale n.195 del 21-08-2019. Retrieved June 7, 2024, from <https://www.gazzettaufficiale.it/eli/id/2019/08/21/19G00105/sg>
- Letouzé, E. (2016). *Should 'data literacy' be promoted?*. UN World Data Forum. Retrieved June 7, 2024, from <https://unstats.un.org/unsd/undataforum/blog/should-data-literacy-be-promoted/>
- Lévi-Strauss, C. (1955). *Tristi tropici*. Milano: Il Saggiatore.
- Maggino, F. (Ed.). (2017). *Complexity in society: From indicators construction to their synthesis* (Vol. 70). Cham, Switzerland: Springer.
- Mandinach, E. B., Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30-37.
- Markham, A. N. (2020). Taking data literacy to the streets: Critical pedagogy in the public sphere. *Qualitative Inquiry*, 26(2), 227-237.
- Mayer-Schönberger, V., Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
- Mertala, P. (2020). Data (il)literacy education as a hidden curriculum of the datafication of education. *Journal of Media Literacy Education*, 12(3), 30-42. <https://doi.org/10.23860/JMLE-2020-12-3-4>
- Milan, S., Velden, L. V. D. (2016). The alternative epistemologies of data activism. *Digital culture & society*, 2(2), 57-74.
- MIM. (2024). *Schema di decreto di adozione dei modelli di certificazione*

- delle competenze* [Allegato B - Certificazione delle competenze al termine del primo ciclo di istruzione]. Decreto ministeriale n. 14/2024.
- MIUR. (2012). *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione*. Retrieved June 7, 2024, from https://www.miur.gov.it/documents/20182/51310/DM+254_2012.pdf
- MIUR. (2016). *Big Data*. Rapporto del gruppo di lavoro. Retrieved June 7, 2024, from <http://www.miur.gov.it/-/big-datarapporto-del-gruppo-di-lavoro-del-miur>
- MIUR. (2018). *Indicazioni Nazionali e Nuovi Scenari*. Comitato Nazionale per le Indicazioni Nazionali il curricolo della scuola dell'infanzia e del primo ciclo di istruzione. Retrieved June 7, 2024, from <http://www.miur.gov.it/-/indicazioni-nazionali-di-infanzia-e-primo-ciclo-piu-attenzione-alle-competenze-di-cittadinanza>
- MIUR. (2020). *Linee guida per l'insegnamento dell'educazione civica*. Decreto ministeriale n. 35/2020.
- Pangrazio, L., Sefton-Green, J. (2020). The social utility of 'data literacy'. *Learning, Media and Technology*, 45(2), 208-220.
- Pasta, S., Rivoltella, P. C. (2022). Superare la "povertà educativa digitale". Ipotesi di un nuovo costrutto per la cittadinanza digitale. In *La formazione degli insegnanti: problemi, prospettive e proposte per una scuola di qualità e aperta a tutti e tutte* (pp. 600-604). Lecce: Pensa MultiMedia.
- Pinney, L. (2020). Is literacy what we need in an unequal data society?. In M. Engebretsen, H. Kennedy (Eds.), *Data visualization in society*, Amsterdam: Amsterdam University Press.
- Programma Erasmus+ (2023). Call for proposals. European policy experimentations (ERASMUS-EDU-2024-POL-EXP). Retrieved June 7, 2024, from https://ec.europa.eu/info/funding-tendersopportunities/docs/2021-2027/erasmus/wp-call/2024/call-fiche_erasmus-edu-2024-pol-exp_en.pdf
- Raffaghelli, J. E. (2019). Alfabetizzare ai dati attraverso l'uso di Open Data. *Idee in form@zione*, 8(7), 101-128.
- Raffaghelli, J. E. (2023). Pathways for Social Justice in the Datafied Society: Reconsidering the educational response. *Media Education*, 14(1), 5-14.
- Raffaghelli, J. E., Stewart, B. (2020). Centering complexity in 'educa-

- tors' data literacy' to support future practices in faculty development: A systematic review of the literature. *Teaching in Higher Education*, 25(4), 435-455.
- Ranieri, M. (2018). La "svolta digitale" negli studi sulle nuove literacy ed implicazioni educative. In M. Ranieri (Ed.), *Teoria e pratica delle new media literacies nella scuola* (pp. 1-316). Roma: Aracne.
- Ranieri, M. (2019a). Le competenze digitali per la formazione dei cittadini. In P. C. Rivoltella, P. G. Rossi (Eds.), *Tecnologie per l'educazione* (pp. 227-238). Torino: Pearson Italia.
- Ranieri, M. (2019b). La Media Literacy Education tra vecchie e nuove sfide in un mondo iperconnesso. In M. Gui (Ed.), *Benessere digitale a scuola e a casa. Un percorso di educazione ai media nella connessione permanente* (pp. 9-35). Firenze: Mondadori Università.
- Ranieri, M., Cuomo, S., Biagini, G. (2024). *Scuola e intelligenza artificiale. Percorsi di alfabetizzazione critica*. Roma: Carocci.
- Ravanelli, F., Ceretti, F. (2017). Pedagogia degli oppressi digitali. Freire, il digitale e l'utopia pedagogica. *Sapere pedagogico e Pratiche educative*, 2017(1), 157-170.
- Ricciardi, M. (2014). *La comunicazione: maestri e paradigmi*. Bari: Gius. Laterza & Figli Spa.
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., ... Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report*. Dalhousie University.
- Rivoltella, P. C. (2020). *Nuovi alfabeti. Educazione e culture nella società post-mediatale*. Brescia: Scholé-Morcelliana.
- Rivoltella, P. C. (2022). Critical Thinking in the Data Age. New Challenges. In *Humane Robotics. A multidisciplinary Approach toward the Development of Humane-centered Technologies* (pp. 327-342). Milano: Vita e Pensiero.
- Royal, C., Blasingame, D. (2015). Data journalism: An explication. *International Symposium in Online Journalism*, 5(1), 24-46.
- Sander, I. (2023). Critical datafication literacy—a framework for educating about datafication. *Information and Learning Sciences*, 125(3/4), 270-292.
- Schild, M. (2004). Information literacy, statistical literacy and data literacy. *IASSIST Quarterly*, 28(2/3), 6-11.
- Seymoens, T., Van Audenhove, L., Van den Broeck, W., Mariën, I.

- (2020). Data literacy on the road: Setting up a large-scale data literacy initiative in the DataBuzz project. *Journal of Media Literacy Education*, 12(3), 102-119.
- Špiranec, S., Kos, D. George, M. (2019). Searching for critical dimensions in data literacy. In *Proceedings of CoLIS, the Tenth International Conference on Conceptions of Library and Information Science, Ljubljana, Slovenia, June 16-19, 2019. Information Research*, 24(4), paper colis1922.
- Street, B. (2003). What's 'new' in new literacy studies? Critical approaches to literacy in theory and practice. *Current issues in comparative education*, 5(2), 77-91.
- Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2), 205-3951717736335.
- Tygel, A.F., Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics*, 12 (3), 108-121.
- UN. (2020, November). *Data Literacy – The Human Element in Data*. United Nations Data Blog. Retrieved June 7, 2024, from <https://unite.un.org/blog/data-literacy-%E2%80%93-human-element-data>
- UNESCO. (2005). *Education for All Global Monitoring Report: Literacy for life*, UNESCO Publishing, Paris. Retrieved June 7, 2024, from <https://www.unesco.org/gem-report/en/literacy-life>
- UNESCO. (2024, gennaio). *What you need to know about literacy*. Retrieved June 7, 2024, from <https://www.unesco.org/en/literacy/need-know>
- Vahey, P., Yarnall, L., Patton, C., Zalles, D., Swan, K. (2006). Mathematizing middle school: Results from a cross-disciplinary study of data literacy. In *Annual Meeting of the American Educational Research Association, San Francisco, CA* (pp. 1-15). Washington: AERA.
- Van Audenhove, L., Van den Broeck, W., Mariën, I. (2020). Data literacy and education: Introduction and the challenges for our field. *Journal of Media Literacy Education*, 12(3), 1-5.
- Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & society*, 12(2), 197-208.
- Vuorikari, R., Kluzer, S., Punie, Y., (2022). *DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge*,

- skills and attitudes*. Bruxelles: Publications Office of the European Union.
- Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1-8.
- Watson, J. M. Callingham, R. (2003). Statistical literacy: A complex hierarchical, construct. *Statistics Education Research Journal*, 2(2), 3-46.
- Wickens, C. M., Sandlin, J. A. (2007). Literacy for what? Literacy for whom? The politics of literacy education and neocolonialism in UNESCO-and World Bank-sponsored literacy programs. *Adult Education Quarterly*, 57(4), 275-292.
- Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. Newbury Park: Sage Publications Ltd.
- Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Westport: Greenwood Press.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.
- Wolff, A., Gooch, D., Cavero Montaner, J.J, Rashid, U., Kortuem, G., (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, 12(3), 9-26.
- Yates, S., Carmi, E. (2022). Citizens' networks of digital and data literacy. In *The Palgrave handbook of media misinformation* (pp. 191-205). Cham: Springer International Publishing.
- Zuboff, S. (2019). *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*. Roma: Luiss University Press.

III.

Didattica della *data literacy* ed esperienze di ricerca nel panorama educativo

Elena Gabbi

3.1 Portare l'alfabetizzazione ai dati in classe: formare cittadini consapevoli nell'era dei dati

3.1.1 *Sviluppo del curriculum e contesti disciplinari*

“La questione dei dati non riguarda solo la matematica: è una questione di cultura” (Buckingham, 2022, para 21). L'importanza della *data literacy* nel contesto educativo va ben oltre la mera dimensione matematica: la gestione e la comprensione dei dati non sono soltanto abilità numeriche, ma si radicano profondamente nella cultura contemporanea. Nel corso dei primi due capitoli abbiamo osservato come la datificazione, sebbene spesso associata all'era digitale, affondi le sue radici in una storia che precede l'avvento delle tecnologie digitali. L'impiego sempre più diffuso dei dati per rappresentare caratteristiche personali e comportamenti rientra in un lungo processo di sviluppo dei sistemi di registrazione, risalente almeno a un secolo fa. Tali sistemi, che mirano a misurare, classificare e monitorare segmenti di popolazione su larga scala, portano inevitabilmente con sé valori e presupposti ideologici. L'attuale proliferare dell'uso dei dati, dunque, modella in maniera significativa la nostra percezione della realtà, di noi stessi e degli altri, e ci rende partecipi (o succubi) di un meccanismo economico e sociale difficile da decostruire. Pertanto affrontare la questione dei dati in educazione significa indagare la complessità delle relazioni tra individui, società e informazioni.

Nel presente capitolo, esamineremo non solo il modo in cui la didattica della *data literacy* si inserisce nei contesti disciplinari e nei curricula, ma anche come essa affronta sfide e limitazioni in riferimento ai diversi ordini scolastici. Esploreremo, inoltre, le strategie didattiche e le pratiche di valutazione che possono essere adottate per promuovere una formazione completa e consapevole in questo ambito cruciale per una cittadinanza digitale in continua evoluzione. In seguito, presenteremo i principali temi riguardanti l'insegnamento della *data literacy*, utilizzando schede didattiche, finalizzate alla promozione delle competenze fondamentali per comprendere, visualizzare, interpretare e creare con i dati. Inoltre, considereremo come tali competenze possano essere integrate nella formazione degli insegnanti e coinvolgere altri stakeholder rilevanti.

Nella letteratura, si riscontra ampia convergenza sull'idea secondo cui le basi per sviluppare una società alfabetizzata ai dati debbano essere poste a livello scolastico; nonostante ciò, l'importanza della *data literacy* stenta ad essere adeguatamente riconosciuta a scuola o all'università attraverso una sua integrazione nelle pratiche didattiche correnti (Raffaghelli, 2023; Wolff, Gooch, Caverio Montaner, Rashid, Kortuem, 2019). L'inserimento di percorsi finalizzati all'acquisizione di *data literacy* nei diversi contesti educativi richiede un impegno coordinato tra istituzioni, docenti ed esperti di alfabetizzazione ai media e ai dati. I programmi formativi possono essere progettati per rispondere alle esigenze e agli interessi degli studenti e possono coinvolgere gli insegnanti stessi, fornendo loro opportunità di formazione e risorse per maturare competenze sempre più avanzate (Van Audenhove, Van den Broeck, Mariën, 2020).

Una formazione in tal senso può essere realizzata in modo efficace sia a livello di curriculum generale tramite percorsi *ad hoc* di una certa durata (Martin, Leger-Hornby, 2012), che attraverso attività formative più brevi dedicate ad aspetti specifici (De Marco, 2023; Qin, D'Ignazio, 2010; Teal et al., 2015). La prima

tipologia di approcci si caratterizza per una struttura ben definita, con obiettivi legati alla progressiva acquisizione di competenze e ad un monitoraggio continuo per fornire assistenza specifica quando necessario (Ridsdale et al., 2015). Un esempio per la tipologia curricolare è costituito dal percorso ideato da Martin e Leger-Hornby (2012), che hanno progettato e realizzato un curriculum per un corso universitario presso l'University of Massachusetts Medical School. Questo corso, articolato in sette moduli e quattro studi di caso, copre aspetti chiave della gestione dei dati di ricerca, tra cui tipologie, formati, fasi e considerazioni etiche. Il programma rappresenta un modello adattabile ai diversi curricula, offrendo un'organizzazione strutturata dei contenuti da trattare al fine di acquisire competenze specifiche nell'ambito della gestione dei dati. Tuttavia, a fronte degli ostacoli incontrati dalle istituzioni nell'ampliare la gamma dei corsi, anche l'integrazione della *data literacy* all'interno di percorsi esistenti relativi a *literacy* affini costituisce un approccio praticabile e sostenibile. Zilinski, Scherer e Maybee (2013) hanno sperimentato l'introduzione di tali competenze nelle classi universitarie degli studenti STEM (discipline scientifico-tecnologiche), integrata all'interno di curricula disciplinari già istituiti. Connettendo la *data literacy* all'*information literacy*, gli autori hanno sviluppato nel proprio intervento didattico una *data credibility checklist*, cioè un elenco che comprende gli elementi fondamentali per valutare la qualità di un insieme di dati.

Diversamente dagli approcci curricolari, la seconda tipologia di interventi si caratterizza per attività formative brevi e intense, come ad esempio un workshop, e sono spesso mirate a competenze specifiche, livelli e/o settori. Simili approcci risultano efficaci soprattutto per mitigare il divario esistente nelle competenze o colmare la distanza tra conoscenza e pratica, attraverso opportunità di apprendimento situate e interattive, sia in contesti formali che informali. Teal e colleghi (2015) hanno specificato le caratteristiche chiave per la progettazione di tali interventi brevi:

- essere specifici del settore disciplinare, per ottenere una comprensione più immediata e facilitare l'applicazione al proprio ambito, utilizzando esempi concreti e realistici
- comprendere la spiegazione delle diverse fasi del ciclo di vita dei dati per un determinato set di dati o problema, per consentire la riproducibilità dei risultati e offrire un modello di un possibile flusso di lavoro
- essere progettati per persone senza precedenti esperienze in materia di dati, per consentire a tutti i partecipanti di imparare secondo il proprio ritmo, e basarsi sulle pratiche e conoscenze pregresse
- essere traducibili e applicabili ad un diverso ambito, poiché la flessibilità dei principi del ciclo di vita dei dati permette di trasportare tali metodi e risorse a qualunque dominio conoscitivo.

Venendo al contesto disciplinare in cui inserire interventi di alfabetizzazione ai dati, un'analisi comparativa degli standard di contenuto nelle discipline degli studi sociali, della matematica, delle scienze e dell'ambito umanistico ha evidenziato una convergenza su tre ampie categorie di competenze presenti in modo trasversale in questi ambiti disciplinari (Vahey, Yarnall, Patton, Zalles, Swan, 2012):

- la capacità di formulare e rispondere a domande basandosi sui dati;
- l'abilità nell'uso di dati, di strumenti e relative rappresentazioni;
- la capacità di sviluppare e valutare inferenze e spiegazioni basate sui dati.

La maggior parte degli interventi in materia di *data literacy* mostra una mancanza di specificità disciplinare (Ridsdale et al., 2015). Tra le diverse iniziative, si osserva che la maggior enfasi è posta sull'alfabetizzazione ai dati per gli studenti STEM, con al-

cuni riferimenti integrativi alle scienze sociali, all'economia e al settore medico (Ghodoosi, West, Li, Torrisi-Steele, Dey, 2023). Nonostante un curriculum di matematica includa spesso indicazioni per l'insegnamento della gestione dei dati e dell'analisi statistica, Wolff e colleghi (2016) suggeriscono che sarebbe più efficace insegnare la *data literacy* come materia interdisciplinare, incorporandola in ambiti come scienze e geografia. In questa visione, i concetti e i metodi appresi verrebbero contestualizzati in modo più efficace, portando a competenze maggiormente trasferibili in altri ambiti. L'approccio interdisciplinare può aiutare gli studenti a capire come i dati siano rilevanti in molteplici contesti, rendendoli consapevoli dell'applicazione pratica di queste competenze in diverse materie, tra cui matematica, scienze, studi sociali e informatica (Van Audenhove et al., 2020).

Un efficace esempio di curriculum interdisciplinare è offerto dal lavoro di Vahey e collaboratori (2012) nell'ambito del progetto "Thinking with Data". Questo intervento didattico, realizzato negli Stati Uniti, ha coinvolto quattro materie di una classe di scuola media: studi sociali, matematica, scienze e lingua inglese. Le fasi del lavoro hanno incluso la progettazione dei moduli, la co-progettazione con gli insegnanti e il continuo perfezionamento dei materiali fino alla versione finale per l'implementazione. L'obiettivo del progetto è stato quello di utilizzare l'insegnamento della matematica per preparare gli studenti alla vita al di fuori del contesto disciplinare, riconoscendo l'importanza trasversale di sviluppare competenze di alfabetizzazione ai dati. Come caso di studio, è stato selezionato il tema della distribuzione equa dell'acqua nel bacino idrografico del Tigri/Eufrate condiviso da Siria, Iraq e Turchia. Alla fine del progetto, ci si attendeva che gli studenti determinassero se l'attuale assegnazione dell'acqua tra i tre stati fosse equa, utilizzando come riferimento la Convenzione delle Nazioni Unite sul diritto all'uso delle acque internazionali non navigabili.

La *data literacy* è stata utilizzata come ponte tra le discipline e come strumento per aiutare gli studenti a comprendere meglio i

dati e le affermazioni a cui sono esposti quotidianamente, sia a scuola che nel quotidiano. Il modello è stato applicato attraverso quattro moduli consequenziali:

1. l'insegnamento degli studi sociali per il contesto storico-politico e la preparazione di base;
2. l'insegnamento della matematica per il ragionamento e il calcolo delle proporzioni;
3. l'insegnamento delle scienze per applicare i concetti matematici alla comprensione della distribuzione dell'acqua;
4. l'insegnamento linguistico per consolidare l'apprendimento degli studenti che devono comunicare le proprie soluzioni ai problemi di allocazione dell'acqua, avvalendosi di argomentazioni basate sui dati.

Tra i vantaggi osservati nell'implementazione del progetto "Thinking with Data" vi sono l'incremento della *data literacy* attraverso l'approccio interdisciplinare, il miglioramento della comprensione dei concetti matematici e la capacità di interpretare i dati in contesti reali, nonché un innalzamento della qualità degli elaborati degli studenti che hanno articolato argomentazioni efficaci e data-based.

3.1.2 *Sfide e limitazioni in relazione agli ordini scolastici*

La capacità di raccogliere, analizzare e formulare inferenze accurate a partire dai dati non dovrebbe essere riservata soltanto all'istruzione secondaria e post-secondaria, come tradizionalmente è avvenuto. Per affrontare la crescente necessità di alfabetizzazione ai dati per la cittadinanza è importante integrare questo tipo di competenza già nell'istruzione primaria (Vahey et al., 2012). L'introduzione precoce di tali concetti consente infatti agli individui di acquisire competenze fondamentali per comprendere come i

dati possano essere utilizzati per dare significato alla loro vita quotidiana (Rubin, 2020). La promozione delle competenze di alfabetizzazione ai dati nei bambini più piccoli presenta notevoli sfide, tra cui la questione della maturità cognitiva e metacognitiva necessaria per affrontare nozioni complesse come l'analisi causale e il ragionamento inferenziale (Cottone, Yoon, Shim, Coulter, Carman, 2023). Alcune ricerche suggeriscono che gli studenti della scuola primaria potrebbero non essere ancora pronti, dal punto di vista cognitivo e metacognitivo, ad assimilare appieno tali capacità di ragionamento (Seel, 2017).

Rendere l'apprendimento dei concetti legati ai dati più accessibile e adattarlo al livello di competenza degli studenti della scuola primaria, aiuta a promuovere una base solida per lo sviluppo futuro delle loro abilità di *data literacy*. Per garantire l'efficacia, è essenziale ripetere la proposta educativa e collegare i concetti fondamentali a un'attività pratica. Inoltre, nel contesto dell'insegnamento della *data literacy* ai bambini della scuola primaria, è fondamentale adottare un linguaggio e un approccio didattico appropriati, aiutando gli studenti a trasformare i dati disponibili in conoscenza significativa (Barzilai, Chinn, 2018). Ad esempio, quando si lavora con gli studenti della scuola primaria, utilizzare il termine "cercare la storia" (*story-finding*) anziché "analizzare i dati" (*data analysis*) risulterebbe più efficace e coerente, poiché l'analisi dei dati è spesso associata a conoscenze avanzate e tecniche di programmazione e utilizzo di software, mentre con i più piccoli l'obiettivo è di formare competenze di base (Bhargava et al., 2015).

Nel contesto dell'istruzione secondaria, la ricerca e gli interventi di alfabetizzazione ai dati si sono concentrati soprattutto sullo sviluppo delle competenze degli studenti nell'analisi e nell'interpretazione di dati quantitativi strutturati. Gli educatori integrano spesso competenze statistiche e matematiche nei percorsi formativi, sebbene vi sia una limitata attenzione alla concettualizzazione e all'utilizzo dei dati da parte degli studenti (Wolff et

al., 2017). Sebbene queste competenze tecniche siano cruciali, da sole non sono sufficienti a stimolare gli studenti ad affrontare questioni rilevanti per loro e a utilizzare i dati per risolvere problemi nella propria comunità. Spesso la visione che gli studenti della scuola secondaria hanno dei dati è limitata alla loro esperienza di lavoro in contesti scolastici come sondaggi, dati di esperimenti o grafici e tabelle (Gebre, 2018). Le concezioni degli studenti su un determinato argomento sono il risultato di un processo cognitivo che affonda le proprie radici nelle conoscenze ed esperienze pregresse. Quando agli studenti viene richiesto di spiegare un concetto, le loro risposte sono spesso l'esito di una combinazione di ricordi e frammenti di conoscenza, assemblati al momento per soddisfare le esigenze specifiche della domanda. Pertanto, se gli studenti hanno familiarità principalmente con dati strutturati da esperimenti di laboratorio o fogli di lavoro, è probabile che le loro spiegazioni sui dati siano influenzate da tali esperienze. Nel panorama dell'istruzione secondaria, vi è anche una carenza di attenzione circa il ruolo dei giovani adulti come fonte di dati e la conseguente necessità di aumentare la consapevolezza degli studenti riguardo alla loro relazione con i dati e alle relative implicazioni. Questo aspetto comprende il contributo significativo dei giovani adulti attraverso la generazione di una vasta quantità di dati personali su piattaforme di social media (Carenzio, Rivoltella, Ferrari, 2021). Tuttavia, non è del tutto chiaro in che misura gli studenti siano consapevoli del proprio ruolo e delle implicazioni associate (Pangrazio, Selwyn, 2018).

L'educazione alla *data literacy* viene sempre più offerta a livelli post-laurea, fornendo agli studenti le competenze e gli strumenti per affrontare la "big data question" (Ridsdale et al., 2015). Tuttavia, insegnare la *data literacy* a questo livello influisce solo su un numero limitato di studenti, a differenza dell'ambito universitario. Inoltre, è spesso difficile introdurre conoscenze di base nel post-laurea. Qin e D'Ignazio (2010) hanno riscontrato nei feedback dei loro studenti che la mancanza di conoscenze di base

rende difficile padroneggiare il gergo e gli esercizi di *data literacy*, in modo particolare quando ci sono livelli di competenza che variano sensibilmente tra i componenti del gruppo. In termini di utilità, inoltre, le competenze apprese nei corsi post-laurea sono molto specifiche e incentrate sulla disciplina. E, ancora, insegnare l'educazione alla *data literacy* all'inizio dei corsi di laurea potrebbe introdurre buone pratiche e migliorare il rendimento generale degli studenti.

Una prospettiva orientata al *lifelong learning* riconosce infine la necessità per le persone di un'esperienza di apprendimento continua e a lungo termine, ricercando un contesto in grado di fornire sostegno, indirizzare in caso di difficoltà e monitorare i progressi compiuti. Tuttavia, secondo Carmi e Yates (2020), i programmi di alfabetizzazione ai dati spesso presentano una durata limitata, concentrata in interventi specifici, di solito della durata di poche settimane o mesi. Una possibile soluzione potrebbe consistere nel coinvolgimento delle comunità locali supportate da biblioteche e finanziamenti specifici, contribuendo così a costruire una rete di alfabetizzazione a lungo termine e sostenibile. Un altro aspetto da considerare nell'insegnamento di abilità critiche per la lettura e l'uso dei dati riguarda l'importanza di comunicare in modo chiaro ed esplicito fin dall'inizio i vantaggi legati al loro utilizzo e alle competenze ad essi correlate (Ridsdale et al., 2015). Tale aspetto risulta particolarmente rilevante per gli studenti in età adulta, i quali potrebbero essere più inclini ad investire il loro tempo limitato e le risorse disponibili, se individuano l'opportunità di apportare un contributo significativo alla loro comunità o ad altri contesti rilevanti e di progredire attraverso un processo graduale di acquisizione di autonomia. Tra le iniziative rivolte alla cittadinanza, il progetto "Data Murals" (Bhargava, Kadouaki, Bhargava, Castro, D'Ignazio, 2016) ha affiancato una comunità per collaborare alla creazione di un'opera d'arte - un murale - che riflettesse le loro indagini sui dati, utilizzando dati pertinenti al loro quartiere. Inoltre, Anslow, Brosz, Maurer e Boyes (2016)

hanno esaminato il potenziale dei *datathon*¹ nel promuovere l'alfabetizzazione ai dati, riunendo studenti e cittadini per affrontare sfide e risolvere problemi attraverso l'analisi dei dati.

3.1.3 *Progettazione didattica e valutazione*

La revisione della letteratura di Ridsdale e colleghi (2015) ha permesso di identificare diverse *best practices* per l'insegnamento della *data literacy*, alcune delle quali si discostano dalle strategie didattiche tradizionali, ma risultano coerenti con le pratiche di insegnamento già in atto nelle istituzioni post-secondarie: applicazioni pratico-operative, strategie modulari o di sequenziamento e metodologie problem-based.

È essenziale sottolineare che, come per lo sviluppo di altre competenze, la pratica costante è cruciale per acquisire competenze nella *data literacy*, poiché la comprensione delle dinamiche sottostanti al processo è di primaria importanza. Affrontare errori può risultare frustrante, ma rappresenta una componente essenziale per stimolare il pensiero critico e la risoluzione dei problemi. L'apprendimento *learning by doing* (Dewey, 1993), realizzato attraverso workshop e laboratori, fornisce agli studenti l'esperienza pratica necessaria per acquisire una piena comprensione di competenze tecniche specifiche; in tal senso, è molto importante offrire agli studenti la possibilità di esplorare i processi e i metodi

1 Un *datathon* (neologismo composto dalle parole *data* e *marathon*) è una competizione di idee in cui individui o gruppi si riuniscono per sviluppare progetti o applicazioni utilizzando dati aperti pubblicati da istituzioni. Si tratta di una sfida che mira a dimostrare il valore dei dati aperti, a evidenziarne il potenziale e a promuoverne l'utilizzo come risorsa gratuita a disposizione del pubblico più ampio. Ad esempio, EU Datathon è un'iniziativa dell'Unione Europea rivolta ai cittadini che dal 2017 ogni anno promuove l'utilizzo degli *open data*.

in autonomia, commettendo errori che, in seguito, possano essere corretti per migliorare la comprensione.

Adottare una strategia di sequenziamento prevede che gli studenti si confrontino dapprima con attività di apprendimento semplici e progrediscano in un secondo tempo verso attività più complesse, con la possibilità di raggiungere gli obiettivi di apprendimento in maniera graduale e sistematica (Van Merriënboer, Kirschner, Kester, 2003). Cominciare, quindi, con compiti di minore complessità e avanzare verso compiti più articolati consente agli studenti di maturare fiducia nelle proprie abilità. Un processo di apprendimento basato su moduli consente di costruire sulle competenze acquisite in precedenza, promuovendo un approccio centrato sullo svolgimento dell'attività piuttosto che sulla mera memorizzazione o sull'adesione ad una procedura prestabilita. Questa modalità didattica facilita l'apprendimento di concetti inizialmente sconosciuti.

L'apprendimento basato su progetti, quando implementato secondo l'approccio sequenziale, risulta particolarmente efficace per sviluppare la *data literacy* (Ridsdale et al., 2015). Progetti che incorporano una vasta gamma di metodi d'indagine e offrono contesti d'applicazione concreti rafforzano il legame tra teoria e pratica (Kokotsaki, Menzies, Wiggins, 2016), consentendo ai docenti di valutare le competenze degli studenti in contesti situati, anziché basarsi su una valutazione puramente astratta e contenutistica. Inoltre, è opportuno che i progetti includano dati ed esempi del mondo reale, pertinenti agli interessi degli studenti e contestualizzati in un ambito di interesse per loro, evitando di lavorare su dati per il solo scopo di utilizzarli. I progetti dovrebbero pertanto fornire agli studenti l'opportunità di superare le proprie aspettative e pregiudizi nell'ambito dei dati. Sviluppare competenze di base in classe è importante, ma incoraggiare gli studenti a migliorare queste competenze attraverso sperimentazioni a casa è essenziale per raggiungere un adeguato livello di competenza (Cobo Romani, 2009). Un maggiore coinvolgimento può infatti

innescare dinamiche di innovazione, migliorare il processo di apprendimento e accrescere la probabilità di proseguire il percorso in ottica *lifelong learning*.

L'integrazione della *data literacy* nell'ambito scolastico richiede un approccio olistico e ben strutturato. I principi di design per attività di *data literacy* sviluppati da Wolff, Wermelinger e Petre (2019) delineano un quadro che enfatizza gli elementi di indagine proattiva, l'ampliamento graduale del set di dati, il contesto reale e familiare, lo sviluppo di competenze trasversali e di base, l'approccio ibrido tra creazione e analisi e la raccolta dei dati in prima persona (tabella 3.1).

Principi di design per attività didattiche sulla <i>data literacy</i>	
Inquiry-based	Seguire un processo di esplorazione e ricerca per supportare l'analisi dei dati, stimolando gli studenti a formulare domande e a individuare azioni per risolvere problemi e capire i fenomeni. Condurre gli studenti prima in un'esplorazione guidata dall'insegnante, alla quale segue una fase d'indagine aperta ed indipendente, quando gli studenti hanno sviluppato maggior familiarità con i dati e il metodo.
Espansione	Iniziare da una porzione rappresentativa di una piccola parte del set di dati ed espanderla in seguito, anziché iniziare con un set di dati completo e di grandi dimensioni e concentrarsi da subito su di esso. Questo mira ad aiutare gli studenti a collegare più facilmente le domande di esplorazione e ricerca sui dati.
Contesto	Insegnare attraverso un contesto familiare e comprensibile per gli studenti, utilizzando dati provenienti dal proprio ambiente, sia locali che correlati ad essi in qualche altro modo.
Competenze fondamentali	Concentrarsi sullo sviluppo delle competenze di base e trasversali piuttosto che delle abilità pratiche, ad esempio come saper porre "buone" domande di tipo scientifico a partire dai dati.
STEAM	Adottare un approccio educativo che integra le discipline di STEM (Scienza, Tecnologia, Ingegneria e Matematica) con l'Arte ² , lavorando in modo collaborativo su attività creative insieme a quelle pratiche.

Raccolta dei dati	Far interagire gli studenti con i dati che hanno raccolto in modo autonomo. Quando gli studenti analizzano un set di dati esterno e prestabilito, dovrebbero essere previste attività aggiuntive che li supportino nella comprensione di cosa significhi raccogliere quel tipo di dati.
-------------------	---

Tab. 3.1: Elementi chiave per la progettazione di interventi volti a sviluppare la *data literacy* per gli studenti (adattato da Wolff et al., 2019)

Questi principi mirano a creare un intervento didattico coinvolgente e efficace, fornendo agli studenti le competenze necessarie per comprendere, analizzare e interpretare i dati in modo critico. Particolarmente interessanti sono i principi di scalabilità e di raccolta dei dati. Da un lato, il primo principio ha lo scopo di fornire agli studenti l'opportunità di orientarsi all'interno dei dati, prima di esplorarli, ad esempio, attraverso il tempo, lo spazio e/o qualche altra dimensione, diminuendo quindi la complessità iniziale dello stimolo e permettendo di approfondire l'esplorazione per gradi. Dall'altro, il principio relativo alla raccolta di dati grezzi mira ad aiutare gli studenti a interagire con grandi set di dati, familiarizzando con le difficoltà e i limiti del processo di trasformazione della realtà in unità informative. Tale principio ha il fine di contestualizzare e interpretare i dati già esistenti e limitare gli effetti di una visione ideologica e idealizzata del dato, quale "verità oggettiva" e con caratteristiche di presunta "neutralità" (van Dijck, 2014).

Data l'enfasi sul contesto reale nelle competenze di *data literacy*, i ricercatori hanno suggerito di insegnare la *data literacy* come una disciplina trasversale attraverso l'apprendimento basato

- 2 Rispetto all'approccio STEM (Scienza, Tecnologia, Ingegneria, Matematica), l'aggiunta dell'Arte mira a incorporare la creatività e l'espressione artistica nel processo di apprendimento basato su queste discipline. Ne sono un esempio la *creative data literacy* e le proposte di attività didattiche di D'Ignazio (2017).

sui progetti e utilizzando dati autentici (Kjelvik, Schultheis, 2019). La progettazione di un intervento didattico coinvolgente ed efficace, basato su principi e metodologie attive, mira a dotare gli studenti delle competenze essenziali per affrontare in modo critico la comprensione, l'analisi e l'interpretazione dei dati (Raf-faghelli, 2018a). Valutare l'efficacia dell'apprendimento basato su tali buone pratiche e principi di design risulta tuttavia una sfida significativa, a causa della complessità delle dimensioni da considerare all'interno di un evento didattico (es. design, condizioni di erogazione, strategie didattiche, target). Tale valutazione comporta inoltre un rischio di soggettività e richiede un considerevole impegno temporale, soprattutto quando coinvolge un ampio numero di studenti.

Lo studio sperimentale di Reeves e Chiang (2019) ha analizzato l'efficacia di un breve intervento online di *data literacy*, interattivo e strutturato, che ha impegnato i partecipanti a porre e rispondere a diversi tipi di domande attraverso l'uso di dati di valutazione esterni e standardizzati presentati in tabelle, grafici e punteggi. Lo studio ha analizzato l'impatto dell'intervento sull'implementazione a scuola delle pratiche di utilizzo dei dati, nonché l'autoefficacia e l'ansia legate al processo decisionale data-based. Inoltre, ha rilevato le credenze sulla valutazione, così come le percezioni dei partecipanti, insegnanti in servizio e in formazione, sull'impatto dell'intervento didattico.

In modo parallelo, anche la valutazione delle competenze di *data literacy* rappresenta una sfida, in particolare a causa della mancanza di una definizione universalmente condivisa e delle competenze associate nella letteratura, come è stato illustrato nel capitolo 2 (par. 2.2).

La revisione sistematica condotta da Cui, Chen, Lutsyk, Leighton e Cutumisu (2023) ha categorizzato le valutazioni esistenti della *data literacy* secondo due principali orientamenti: rilevazioni eterodirette e approcci autoriflessivi. Le prime, più comuni in letteratura, si distinguono in tre forme: test tradizionali con do-

mande a scelta multipla e a risposta aperta, valutazioni basate su giochi digitali e osservazioni dei partecipanti. In questa tipologia, i ricercatori assumono un ruolo predominante nell'analisi delle competenze di *data literacy* dei partecipanti, sviluppando e convalidando gli strumenti di valutazione. Test, rubriche di valutazione e scenari realistici sono gli strumenti più utilizzati, tuttavia non mancano metodi meno convenzionali. Ad esempio, Chin, Blair e Schwartz (2016) hanno progettato un *assessment game-based* per studenti di biologia delle scuole superiori attraverso un esercizio di *data storytelling*: tramite il gioco Storylet gli studenti hanno interpretato il ruolo di redattori junior, leggendo storie diverse e selezionano le foto, i grafici e i titoli appropriati. Diversamente, gli approcci autoriflessivi, meno frequenti in letteratura, valutano le percezioni individuali delle competenze di *data literacy* attraverso sondaggi, questionari, interviste semi strutturate e con la metodologia *think-aloud*. Quest'ultima prevede che i partecipanti valutino le proprie abilità di *data literacy* in modo autonomo o verbalizzino quello che pensano durante l'esecuzione del compito. Tali strumenti sono stati impiegati con l'obiettivo di misurare la capacità di prendere decisioni basate sui dati, le credenze e gli atteggiamenti e le prassi in uso. Per esempio, Reeves e Honig (2015) hanno indagato l'efficacia di un intervento didattico sui futuri insegnanti, analizzando le diverse pratiche di utilizzo dei dati degli studenti, mediante item su scala Likert e domande aperte. Queste pratiche comprendono attività come "fornire un feedback specifico sui punti di forza e di debolezza dei singoli studenti" e "supportare le decisioni sull'insegnamento basandosi sui dati della valutazione" (*ivi*, p. 95). L'intervento didattico sulla *data literacy* ha coinvolto i partecipanti nel confronto tra le valutazioni tradizionali delle prestazioni in classe e le decisioni basate sui dati. Dalle rilevazioni qualitative sono emersi come aspetti rilevanti nell'esperienza dei partecipanti l'uso di strumenti predisposti per l'analisi e il mentoring da parte dei docenti.

Infine, sono rari gli studi che si sono concentrati sullo sviluppo

e sulla validazione di uno strumento di misurazione della *data literacy* (Cui et al., 2023). Tra questi, Larasati e Yunanta (2020) hanno sviluppato e validato uno strumento con 12 item a scelta multipla, nell'ambito disciplinare della fisica per le scuole superiori. Gli autori hanno testato la validità di costruito attraverso l'analisi degli esperti, oltre all'affidabilità e alla validità interna tramite analisi psicometriche, utilizzando sia la teoria classica dei test sia la teoria della risposta agli item (IRT). Diversamente, Wu, Xu e Zhan (2021) hanno elaborato un Cognitive Diagnostic Assessment (CDA) volto a valutare la *data literacy* degli insegnanti. Questo tipo di valutazione viene utilizzata per progettare test mirati a determinare se i partecipanti hanno acquisito un particolare insieme di abilità e a fornire informazioni sui loro punti di forza e di debolezza cognitiva (Leighton, Gierl, 2007). Gli autori hanno validato uno strumento composto da 19 item, mediante l'analisi delle risposte di 531 insegnanti in servizio cinesi, identificando otto distinte abilità, ciascuna indagata da almeno un item del test.

Concludendo, la valutazione delle competenze di *data literacy* può beneficiare dall'integrazione di approcci valutativi diversificati, sia ancorati ad una situazione autentica e a strumenti di rilevazione esterni, che orientati alla riflessione e all'introspezione (Cui et al, 2023). Ciò consente di ottenere una visione più completa delle competenze di *data literacy* degli studenti e dei futuri cittadini, permettendo agli educatori di adattare e perfezionare le loro strategie didattiche per far fronte alle trasformazioni della società contemporanea.

3.2 Strumenti e risorse per conoscere, visualizzare, interpretare e creare con i dati

3.2.1 *Data literacy: dimensione conoscitiva di base*

Il concetto di *data literacy* ha attraversato, nel tempo, diverse fasi di definizione. Come illustrato nel capitolo 2, in principio è stato radicato in una prospettiva tecnica che si basava sulla comprensione di concetti statistici di base e sulla gestione dei dataset; in seguito, si è trasformato in un contesto più ampio legato alla cittadinanza digitale e alla comprensione dello sfruttamento dei dati in senso economico e infrastrutturale. Questa evoluzione riflette la crescente consapevolezza che la *data literacy* non è una competenza riservata agli esperti, ma piuttosto una capacità fondamentale per tutti, in particolare per i futuri cittadini di un mondo a poco a poco sempre più datificato (Pangrazio, Sefton-Green, 2020). È necessario pertanto includere alcuni concetti fondamentali per comprendere il “linguaggio” e la “grammatica” dei dati, offrendo una base minima di conoscenze per comunicare efficacemente con essi (Tygel, Kirsch, 2016).

I concetti principali includono le definizioni e le distinzioni tra dato, dato digitale, dati aggregati, dataset e database, che sono stati trattati in altre sezioni del volume (capitolo 1, par. 1.1 e capitolo 2, par. 2.1). Nel contesto della comprensione dei dati, è cruciale anche distinguere tra dati qualitativi e quantitativi (Deahl, 2014). Con una prima sommaria definizione, i dati qualitativi rappresentano informazioni che non possono essere espresse in forma numerica, ma piuttosto attraverso parole, immagini o altre forme narrative. Al contrario, i dati quantitativi sono informazioni che possono essere quantificate e misurate numericamente, come numeri e statistiche. Oltre a questa distinzione, i dati possono essere categorizzati come primari o secondari (Hox, Boeijs, 2005). I dati primari sono raccolti attraverso attività in prima persona e sono ottenuti in tempo reale. Questo tipo di

dati richiede un livello elevato di controllo sulla procedura di raccolta, che può comportare tempo, costi e competenze specifiche. Gli esempi di dati primari includono interviste, questionari, osservazioni sul campo e esperimenti condotti dagli individui o dai ricercatori, senza intermediari. D'altra parte, i dati secondari sono stati raccolti in passato da altre persone, organizzazioni o istituzioni e comprendono varie fonti di informazioni pubbliche e *open data*. Questi dati offrono un livello più basso di controllo sulla procedura di raccolta e non richiedono la stessa quantità di tempo e risorse. Esempi di dati secondari includono vari tipi di documenti, diari, statistiche ufficiali e risultati di ricerche pregresse.

Nella sua forma più semplice, un dato digitale è rappresentato da un valore immagazzinato in una posizione specifica, come all'interno di un database o all'interno di una piattaforma di tracciamento. La psicologia sociale Livingstone e colleghi (2019) hanno delineato una distinzione tra tipologie di dati fondata sulla loro origine: dati forniti consapevolmente dagli utenti, tracce di dati lasciate in modo involontario durante le interazioni con la tecnologia (come i dati monitorati dai cookie) e dati inferiti, ovvero nuove informazioni ottenute attraverso la connessione e l'analisi di diverse fonti di dati. La crescente disponibilità e registrazione di tali informazioni in formato digitale può essere interpretata sia come un'opportunità che come una sfida. Da un lato, la vasta mole di dati raccolti da attività di ricerca e da servizi pubblici offre nuove possibilità di collaborazione e condivisione, come evidenziato ad esempio dalla comunità scientifica di Zenodo che ha raccolto dati sull'epidemia di COVID-19 (Ioannidis, 2020). Tuttavia, è importante considerare anche le implicazioni etiche, legali e di privacy legate alla raccolta, alla condivisione e all'uso di tali dati (cfr. capitolo 4, par. 4.3.4).

I dati pubblici derivano dalle attività di servizio sostenute dal settore pubblico e sono spesso accessibili attraverso portali governativi o altre piattaforme online. Allo stesso tempo, l'*open data* costituisce un sottoinsieme importante di questa categoria, com-

prendendo dati resi disponibili per l'accesso pubblico e l'uso gratuito (Chignard, 2013). Mentre alcuni dati aperti sono di natura pubblica, altri possono essere proprietari, creati da istituzioni o entità private, ma comunque caratterizzati dalla possibilità di essere utilizzati e distribuiti liberamente da chiunque, con l'unica condizione che venga attribuita la paternità e che siano condivisi apertamente e senza costi aggiuntivi. Il grado di apertura dei dati può essere valutato in cinque livelli progressivi, come indicato dall'Open Data Institute (Orrù, 2020). Il primo livello si verifica quando i dati sono accessibili al cittadino e possono essere visualizzati, stampati o copiati sotto licenza aperta Creative Commons³. Per raggiungere il secondo livello, i dati devono essere strutturati tramite applicazioni o software proprietari, come Microsoft Excel. Il terzo livello si raggiunge quando i dati sono disponibili per il download in formato CSV (*comma-separated value*), che è aperto e non proprietario. Il quarto livello viene attribuito ai dati accessibili tramite formato URI (*uniform resource identifier*), garantendo tracciabilità universale e univoca e consentendo la fruibilità condivisa in ogni ambiente web connesso. Infine, la più alta valutazione è assegnata ai *linked open data* che sono collegati ad altre serie di dati esterni, facilitando l'interoperabilità simultanea di dati correlati e dialogando con lo stesso schema strutturato.

La comprensione di queste prime distinzioni di base è fondamentale per interpretare correttamente e utilizzare in modo effi-

- 3 Creative Commons è un'organizzazione non profit e una rete internazionale che si dedica ad ampliare l'accesso all'istruzione e al patrimonio creativo disponibile per la condivisione e l'utilizzo nel pubblico dominio. L'organizzazione fornisce una gamma di licenze copyright e strumenti del dominio pubblico che danno a ogni persona e organizzazione del mondo un modo gratuito, semplice e standardizzato per concedere in licenza le opere creative e accademiche, garantire l'attribuzione corretta e consentire ad altri di copiare, distribuire e utilizzare quelle opere.

cace i dati in diversi contesti di studio e di ricerca. La trasformazione precisa e accurata dell'informazione in dati costituisce, infatti, un pilastro fondamentale nel campo scientifico e di ricerca. In questo contesto, gli esperti del settore si confrontano sui limiti e le metodologie più idonee per ripulire i dati, estrarre conoscenza e attribuire significato ai dati raccolti attraverso l'analisi e lo studio del loro oggetto di ricerca (Wickham, 2014). È pertanto essenziale interrogarsi se le conclusioni siano supportate da dati raccolti in modo appropriato, utilizzati in analisi condotte in modo sistematico e coerenti con gli obiettivi della ricerca. Questa dimensione interviene sia in relazione alle competenze di base, quale consapevolezza del processo di raccolta e analisi, sia per i livelli più avanzati di applicazione, come nel caso del rigore richiesto alla metodologia della ricerca scientifica.

A questo proposito, possiamo segnalare due esperienze che ci permettono di descrivere il processo di guida e supporto alla comprensione e manipolazione di base dei dati all'interno del contesto educativo. La prima è "A scuola di OpenCoesione" (ASOC)⁴, un percorso educativo che mira a promuovere e sviluppare la cittadinanza attiva e consapevole nelle scuole secondarie di primo e secondo grado in Italia attraverso la ricerca e il monitoraggio civico dei finanziamenti della politica di coesione europea e nazionale. Gli studenti utilizzano i dati aperti per monitorare e analizzare l'uso dei fondi pubblici delle politiche di coesione europee e nazionali. Il programma insegna agli studenti come utilizzare gli *open data* e il *data journalism* per comprendere e comunicare l'impatto di queste politiche sulle loro comunità locali. L'aspetto distintivo di questa iniziativa risiede nel suo approccio didattico che promuove l'apprendimento attraverso attività di attività progettuali a scuola basate su compiti autentici, dove i dati aperti vengono elaborati ed utilizzati come risorse edu-

4 Fonte: <https://www.ascuoladiopencoazione.it/>

cative aperte. La seconda applicazione è il sito “DataBasic.io”⁵ che racchiude una serie di strumenti web intuitivi progettati per principianti che introducono i concetti fondamentali per lavorare con i dati. I quattro strumenti sviluppati semplificano il processo di manipolazione dei dati in modo semplice e piacevole, consentendo agli studenti di apprendere come individuare storie interessanti a partire da dati familiari:

- *WordCounter*: analizza il testo e indica le parole e frasi più comuni (introduzione all’analisi quantitativa del testo).
- *SameDiff*: confronta due o più documenti di testo e mostra quanto sono simili o diversi (introduzione all’analisi qualitativa).
- *WTFcsv*: spiega cosa può esserci di inconsueto in un file .csv (introduzione all’analisi quantitativa).
- *ConnectTheDots*: rivela come i dati sono connessi analizzando le loro relazioni (introduzione all’analisi delle reti).

I ricercatori, nonché sviluppatori delle applicazioni, hanno testato l’uso degli strumenti con gli studenti, utilizzando *WordCounter* per confrontare e analizzare i testi di canzoni, chiedendo di creare una presentazione visiva della propria storia da condividere con i compagni e utilizzando il risultato come punto di partenza per una discussione sui concetti di data mining e analisi testuale (Bhargava, D’Ignazio, 2015).

Oltre alla familiarizzazione con i concetti di base dell’analisi, è importante riconoscere che i dati raccolti e non ancora elaborati, i cosiddetti dati “grezzi”, non sono intrinsecamente obiettivi e possono presentare limitazioni (Borgman, 2017; Kitchin, 2014). Guidare gli studenti attraverso il processo di raccolta dei dati, categorizzazione e creazione di standard o regole di classificazione

5 Fonte: <https://databasic.io/en/>

li aiuta a capire come gli obiettivi dell'indagine, gli interessi e le procedure contribuiscono alla creazione dei dataset (Wolff et al., 2019). Inoltre, aiuta gli studenti a mettere in atto le capacità di pensiero critico di cui avranno bisogno per porre domande su altri dataset in futuro. Un esercizio a tal fine è la creazione di una biografia dei dati (*data biography*), che rappresenta un'analisi completa dell'origine, della nascita e della vita di qualsiasi dataset (Krause, 2019). È importante riconoscere che utilizzare un insieme di dati senza dedicare il tempo necessario a comprendere a pieno il loro significato non è solo non etico, ma può anche portare a interpretazioni errate e conseguenze indesiderate. Ne consegue che coltivare uno scetticismo nei confronti dei dati grezzi non implica sposare un approccio relativistico o cinico, ma piuttosto adottare uno sguardo consapevole sul modo in cui i dati sono raccolti, ripuliti e analizzati (D'Ignazio, 2017).

Concludendo questa sezione sulla dimensione conoscitiva di base della *data literacy*, proponiamo la seguente scheda didattica esemplificativa di un'attività realizzabile a scuola per supportare la comprensione della natura dei dati e lo sviluppo di competenze basilari nel trattamento delle informazioni.

Scheda didattica 1

Introduzione al processo “disordinato” di creazione e categorizzazione dei dati

Obiettivo: Introdurre ai partecipanti il concetto di raccolta dati attraverso un'attività pratica e interattiva. L'attività si rivolge a studenti di scuola secondaria di primo e secondo grado. Con adattamenti riguardo alla discussione guidata dall'insegnante può essere proposta anche ad alunni della scuola primaria.

Procedura:

1. *Introduzione:* Spiegare ai partecipanti che la raccolta dati è un processo fondamentale per ottenere informazioni utili da analizzare e interpretare. Presentare l'attività di simulazione di raccolta dei dati che sarà svolta.
2. *Creazione della tabella:* Disegnare una tabella semplice sulla lavagna con due colonne: “Nome” e “Colore della maglietta”.
3. *Raccolta dati:* Passare tra i partecipanti chiedendo loro di fornire il loro nome e il colore della maglietta che indossano. Registrare le risposte iniziali sulla lavagna senza commentarle.
4. *Discussione:* Avviare una discussione guidata sulla raccolta dati: se qualcuno indossa una maglietta a righe o con motivi, avvia una conversazione su come rappresentare questo dato nella tabella. Discutere se sia necessario permettere più colori per ciascuna maglietta o se sia opportuno aggiungere una colonna “Motivo/disegno” per descrivere meglio i dati.
5. *Conclusione:* Chiedere ai partecipanti di identificare alcune informazioni che non sono state incluse nella tabella. Usare questo momento per illustrare che i dati sono sempre una semplificazione intenzionale di una realtà più complessa. Parlare anche di come conoscere il possibile obiettivo della raccolta e dell'utilizzo futuro dei dati possa influenzare la raccolta stessa delle informazioni. Far riflet-

tere i partecipanti sull'importanza di comprendere che la raccolta dati richiede scelte e che i dati raccolti possono influenzare le analisi e le decisioni future.

Esercitazione ispirata all'attività "Data messy" descritta in D'Ignazio (2017)

3.2.2 *Data literacy: dimensione visuale*

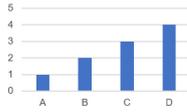
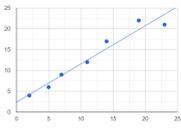
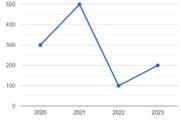
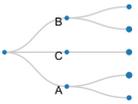
La visualizzazione dei dati è un processo che trasforma i dati complessi in forme grafiche o diagrammi; tuttavia diversi ambiti di ricerca interdisciplinari affrontano tale tema da prospettive diverse (Franconeri et al., 2021). Gli studiosi dell'educazione si concentrano su come superare le sfide nell'interpretazione delle rappresentazioni visive per gli studenti, mentre quelli della comunicazione politica esaminano i fattori che influenzano la percezione e la persuasività delle visualizzazioni. Gli psicologi cognitivi esplorano i limiti della memoria di lavoro nel comprendere le relazioni statistiche complesse attraverso le visualizzazioni, mentre gli esperti di informatica si occupano della progettazione di algoritmi e interfacce utente per la creazione di *dataviz* interattive.

L'uso di metodi visivi consente agli utenti di dare senso ai dati, trasformando le misurazioni quantitative in grafici, tabelle o diagrammi che possono essere interpretati con facilità e utilizzati per supportare argomentazioni e decisioni (Azzam, Evergreen, Germuth, Kistler, 2013). Tale processo aiuta a semplificare la complessità dei dati, consentendo agli utenti di identificare agevolmente modelli, tendenze e relazioni tra le variabili. Le rappresentazioni visive dei dati suggeriscono delle chiavi di lettura per interpretare le informazioni e comunicarle in modo chiaro ed efficace ad altre persone; tuttavia; non si limitano a rappresentare informazioni numeriche in forma visiva ed estetica, bensì svol-

gono anche un ruolo retorico e argomentativo (Kitchin, Lauriault, McArdle, 2015). Infatti, quando i dati vengono trasformati in visualizzazioni grafiche, essi diventano sì più accessibili e comprensibili per l'utente, ma al tempo stesso il modo in cui vengono rappresentati può influenzare la percezione e l'interpretazione stessa delle informazioni: diversi tipi di visualizzazioni possono essere utilizzati per influenzare e persuadere il pubblico in modi diversi.

I pionieri della visualizzazione dei dati, tra cui William Playfair (1759-1823), Charles Joseph Minard (1781-1870), John Snow (1813-1858) e Florence Nightingale (1820-1910), sono stati tra i primi a riconoscere e sfruttare il potenziale della rappresentazione grafica dei dati nei secoli XVIII e XIX (Lupi, 2017). Questa forma di comunicazione visiva ha dimostrato la capacità di veicolare in modo immediato ed efficace, senza l'intermediazione del linguaggio verbale, di grandi quantità di informazioni, sia strutturate che non strutturate, superando molte barriere culturali. Nel corso degli ultimi vent'anni, autori come Edward Tufte e Alberto Cairo hanno contribuito alla rinascita dell'ambito della visualizzazione dei dati, agevolando la trasposizione di tali principi nel contesto dei big data.

La *data visualization* è un processo che si concentra sulla rappresentazione grafica di dati numerici o categorici, mentre grafici, diagrammi e tabelle sono strumenti specifici all'interno di questo processo. Esistono varie forme di rappresentazioni dei dati, ciascuna delle quali è ottimizzata per rispondere a tipi specifici di domande o per descrivere determinati tipi di informazioni in modo più efficace. Ad esempio, alcuni grafici sono ideali per mostrare tendenze nel tempo, mentre altri sono più adatti per confrontare valori tra diverse categorie (vedi tabella 3.2). La scelta del tipo di grafico dipende quindi dalle domande specifiche che si vogliono porre sui dati e dall'obiettivo della visualizzazione stessa (Fisher, Meyer, 2017).

Domanda guida	Obiettivo	Tipologie di grafici	Esempi
“Come si distribuisce una misura?”	Comprendere la distribuzione di una o due serie di dati.	Istogrammi, distribuzioni	
“In che modo i gruppi si differenziano l'uno dall'altro?”	Confrontare i valori quando questi sono suddivisi per categorie o gruppi.	Grafici a barre, grafici a torta	
“Esiste una relazione tra le misure?”	Osservare le relazioni tra gli oggetti e i raggruppamenti.	Scatterplot, heatmap	
“Come varia una misura nel corso del tempo?”	Rappresentare i dati in serie temporali.	Grafici lineari	
“Come sono in relazione gli oggetti?”	Tracciare i rapporti di associazione o gerarchia tra gli elementi.	Visualizzazioni di reti, visualizzazioni ad albero	
“Dove si trovano gli oggetti misurati?”	Collocare i dati in relazione a punti di riferimento conosciuti.	Mappe	
“Che cosa contiene questo testo?”	Rappresentare il conteggio dei temi e delle frequenze di parole o frasi.	Word cloud	

Tab. 3.2: Diversi grafici rispondono a differenti tipi di domande sui dati

La visualizzazione dei dati, soprattutto con dataset di dimensioni ridotte, è il risultato di un processo decisionale che coinvolge diversi attori, dai committenti che richiedono la creazione della visualizzazione ai professionisti incaricati della sua realizzazione, e persino agli intermediari responsabili della raccolta e dell'elaborazione dei dati. L'uso sempre più diffuso degli strumenti di visualizzazione dei dati, noti anche come *dataviz*, è mirato a rendere comprensibili dataset e modelli complessi che diversamente risulterebbero ostici da interpretare, permettendo di rivelare pattern, strutture e interconnessioni altrimenti nascoste tra i dati. Nell'ambito della visualizzazione dei big data, l'analisi visuale si basa sull'interazione sinergica tra esseri umani e algoritmi: mentre gli algoritmi assistono nell'organizzazione e nella presentazione dei dati in modo significativo, è l'utente che guida il processo decisionale e interpreta i risultati (Williamson, 2017).

Le ricerche esistenti indicano che sia i bambini che gli adulti incontrano difficoltà nell'identificare e interpretare varie tipologie di visualizzazioni dei dati e che tale comprensione è influenzata dalla loro familiarità con le convenzioni grafiche presenti all'interno e tra le diverse visualizzazioni dei dati (Börner, Maltese, Balliet, Heimlich, 2016). Shreiner (2017) suggerisce che gli insegnanti possano svolgere un ruolo fondamentale nell'aiutare gli studenti a comprendere che le visualizzazioni dei dati sono strumenti utili per rispondere alle domande fondamentali nel processo di ricerca. Inoltre, possono guidare gli studenti nell'acquisire familiarità con le convenzioni grafiche, temporali e cartografiche comuni, fornendo loro un approccio strutturato per analizzare in modo più approfondito specifiche visualizzazioni dei dati. È essenziale che tale analisi includa la valutazione critica del contesto, la capacità di stabilire connessioni tra la visualizzazione dei dati e il suo contesto, oltre alla valutazione dell'efficacia della visualizzazione stessa nel rispondere alle domande di ricerca e nel rappresentare accuratamente la realtà.

Tufte, designer pioniere dell'ambito delle infografiche, sostiene

che le visualizzazioni dei dati siano “idee complesse comunicate con chiarezza, precisione ed efficienza” (Tufte, 2001, p. 51). Le percezioni sulla realtà, il senso comune e il modo in cui avviene la scoperta e l’acquisizione di intuizioni influenzano direttamente le concezioni di bellezza nella visualizzazione dei dati. Al giorno d’oggi la distinzione tra visualizzazione e arte si fa sempre più sfumata, poiché le nuove rappresentazioni sono concepite per catturare l’attenzione del fruitore attraverso un’estetica accattivante, per poi offrire una ricostruzione dei dati su cui sono basate. Non sempre le competenze acquisite tramite l’analisi di dati di piccole dimensioni, comunicati attraverso grafici e diagrammi piuttosto convenzionali, sono facilmente trasferibili e applicabili nell’ambito di dati di natura più complessa (Wolff et al., 2016). Se è vero che saper decostruire le *data visualisation* permette di accedere ad informazioni inedite, l’introduzione di forme molto complesse all’interno dei materiali didattici per promuovere la *data literacy* in classe richiede un approccio graduale e adattato al contesto di applicazione.

“Information is beautiful”⁶ è un progetto del giornalista McCandless (2011) che, sfruttando gli strumenti di visualizzazione delle informazioni, propone un modo per osservare e per comprendere le relazioni tra i fatti, il loro contesto e la rielaborazione mediatica a cui vengono sottoposti. La visualizzazione interattiva dei dati permette di comprendere, ordinare e rendere più efficace la comunicazione. Oltre ad essere molto belle e curate nel design, tutte le *dataviz* sono basate su fatti e dati reperibili nel sito: le informazioni sono aggiornate regolarmente, riviste e revisionate. Per concludere la sezione sulla componente visuale della *data literacy*, viene proposta una scheda didattica, a partire da questo caso di studio, che illustra un’attività pratica per favorire la comprensione dei concetti legati alla rappresentazione dei dati.

6 Fonte: <https://informationisbeautiful.net/>

Scheda didattica 2.

Introduzione alla lettura e interpretazione delle *dataviz*

Obiettivo: Introdurre gli studenti all'efficacia delle *dataviz* e costruire un linguaggio condiviso per pensare criticamente al design delle visualizzazioni. L'attività si rivolge a studenti di scuola secondaria di secondo grado. Con adattamenti riguardo alla terminologia e all'uso di esempi familiari di *dataviz* tratte dai libri scolastici può essere proposta anche ad alunni della scuola primaria.

Procedura:

1. *Introduzione:* Vedere con i partecipanti il video TEDex di David McCandless “La bellezza della visualizzazione delle informazioni”⁷. Durante la visione, i partecipanti sono incoraggiati a prendere appunti sui concetti chiave presentati e sulle visualizzazioni di dati mostrate.

2. *Discussione:* Dopo la visione del video, guidare una discussione sui concetti presentati (es. rappresentazione delle informazioni, potenzialità creativa del dato, scoperta di nuovi *insight*). I partecipanti saranno stimolati a condividere le loro osservazioni sul modo in cui le *dataviz* hanno contribuito a trasmettere informazioni in modo chiaro ed efficace.

3. *Analisi:* Chiedere ai partecipanti di condividere esempi di visualizzazioni di dati che hanno incontrato in contesti quotidiani o scolastici e di discuterne secondo alcune domande guida:

- A chi è rivolta questa *dataviz*? Con quale proposito è stata progettata?
- Quali tipi di dati vengono presentati?
- Come sono stati visualizzati e organizzati graficamente i dati?

7 Fonte: https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization

d. Ci sono considerazioni etiche o preoccupazioni sull'affidabilità dei dati presentati in questa *dataviz*?

4. *Sintesi e riflessione*: Alla fine della discussione e dell'analisi, invitare i partecipanti a riflettere su quanto appreso dalla visione del video e dalla discussione. Chiedere loro di individuare i concetti chiave che hanno acquisito e di valutare come potrebbero applicarli in situazioni reali.

3.2.3 *Data literacy: dimensione critico-etica*

Per delineare la dimensione critico-etica della *data literacy* è necessario riferirsi ad alcune idee fondamentali come: 1) la *data justice*, che mette in discussione l'equità nella gestione e nell'accesso ai dati (Raffaghelli, 2023), 2) la natura rappresentativa dei dati, in quanto riflettono uno specifico modo di guardare e parlare del mondo (Borgman, 2017) e 3) l'importanza dell'agency delle persone nel contesto d'uso dei dati (Pangrazio, Selwyn, 2023). Come osservano Tygel e Kirsch (2016, p. 119), “i dati da soli non cambiano il mondo, ma crediamo che le persone che comprendono in modo critico la realtà attraverso i dati abbiano strumenti migliori per farlo”.

Nel primo capitolo è stato sottolineato che i dati vanno considerati come prodotti sociali, il cui significato è fortemente influenzato dal contesto e dalla prospettiva dell'osservatore. Le applicazioni tecnologiche, il cui funzionamento si basa *su* e si nutre *di* dati, possono spesso essere intrinsecamente discriminanti, sia perché sono sviluppate su insiemi di dati e presupposti di design influenzati dalle disuguaglianze e discriminazioni presenti nella società, sia perché vengono implementate in contesti sociali caratterizzati da ingiustizie e discriminazioni strutturali (Corazza, 2022; Pangrazio, Selwyn, 2023). Ne consegue che se si assume che “nessuna società possa mai essere completamente equa, allora è difficile credere che ci possa mai essere un'intelligenza artificiale,

un riconoscimento facciale, un algoritmo o altre tecnologie data-based che siano equi” (*ivi*, p. 24). In altre parole, non è opportuno attribuire la responsabilità degli esiti discriminatori agli algoritmi stessi o ai software di intelligenza artificiale: piuttosto, in una società iniqua, le applicazioni tecnologiche tenderanno a riflettere tale ingiustizia (Ranieri, Cuomo, Biagini, 2024). Il rischio, sia per gli individui che per la società nel suo complesso, è che le decisioni basate su dati possano riconfermare le disuguaglianze esistenti legate alla razza, al genere, all’etnia e allo status socioeconomico (O’Neil, 2017). Inoltre, spesso i risultati ottenuti attraverso l’analisi dei big data non vengono integrati in un’indagine più ampia che consideri le dimensioni umane e sociali che possono essere cruciali per una comprensione completa dei fenomeni studiati, quali le interpretazioni, le esperienze, gli atteggiamenti, i valori morali, i dilemmi etici, le contraddizioni e le implicazioni sociologiche (Fuchs, 2017).

In un’epoca dominata dal digitale, le banche dati possono evolversi in nuove forme di narrazione, come nel caso delle collezioni digitalizzate del patrimonio culturale (Panciroli, Macaudo, 2019). Nondimeno, è importante considerare anche che la rappresentazione dei dati, seppur talvolta distorta, tende a definire ciò che è considerato rilevante, mentre ciò che non si adatta a questo modello rischia di essere trascurato. Le mancanze nei dati e la scarsità delle informazioni disponibili non solo riflettono le priorità culturali e sociali, ma anche i pregiudizi impliciti e i temi verso cui la società è indifferente. Un’efficace rappresentazione visiva su questo argomento è l’installazione artistica denominata “Library of Missing Datasets”⁸: si tratta di uno schedario riempito di cartelle vuote che simboleggiano gli argomenti esclusi o trascurati nella società contemporanea (Onuhoa, 2016). I “dataset mancanti” (*missing datasets*) evidenziano i problemi sociali esclusi dalla

8 Fonte: <https://mimionuoha.com/the-library-of-missing-datasets>

datificazione su cui sarebbero necessarie indagini per supportare la ricerca, la trasparenza e l'azione civica, ad esempio il numero di persone transgender uccise o ferite in reati d'odio o statistiche sulla povertà e sull'occupazione che includano le persone in carcere. Spiega l'artista e ricercatrice Mimi Onuhoa che le motivazioni alla base di questa assenza possono essere diverse. A volte coloro che hanno accesso alle risorse per la raccolta dati sono anche coloro che possono influenzarne l'esclusione o la rimozione, come ad esempio in casi di corruzione o per autodifesa dell'organizzazione stessa. Inoltre, alcuni fenomeni sono più difficili da quantificare o categorizzare, come le emozioni o gli atteggiamenti, e di conseguenza vengono trascurati in favore di dati più facilmente gestibili. Infatti, essi potrebbero richiedere un lavoro considerevole per essere documentati, e la mancanza di incentivi o di percezione dei benefici può portare alla loro esclusione dalla raccolta dati. Infine, alcuni gruppi potrebbero trarre vantaggio dal rimanere al di fuori della raccolta dati per proteggere la loro privacy o la loro sicurezza. La soluzione ai problemi rappresentati dai dati mancanti, pertanto, può esulare dalla semplice raccolta di maggiori informazioni, poiché tali problemi complessi possono richiedere approcci che ricadono oltre il campo d'azione della mera documentazione.

In risposta ai problemi menzionati di rappresentazione con i dati, a fianco dei movimenti nati per promuovere la *data justice* già menzionati nel capitolo 2 – quali il *data feminism* e l'*open data movement* – si è fatta strada anche la prospettiva del *data for good* (dati per il bene comune), gettando le basi per una pratica della scienza dei dati prosociale, con una propria etica, regole e casi di studio esemplari. Questo concetto, diffuso tra sviluppatori tecnologici, leader aziendali e responsabili politici, si riferisce all'utilizzo dei dati digitali per fini socialmente utili, come l'istruzione, la salute e lo sviluppo della comunità e implica l'idea che i dati possano essere impiegati per promuovere obiettivi positivi a favore dell'intera società. Quale forma di partecipazione e comunica-

zione dei dati, il *data journalism* è una nuova forma di giornalismo incentivata dal movimento dei dati aperti, in cui le storie sono presentate o integrate attraverso grafici o visualizzazioni di dataset analizzati (Bhargava et al., 2015). È possibile implementare un'applicazione didattica del *data journalism* nell'ambito dell'istruzione: gli insegnanti possono assegnare progetti in cui gli studenti devono raccogliere dati, condurre analisi e creare visualizzazioni per sviluppare storie basate sui dati. Con un approccio fondato sul project-based learning, Gebre (2018) ha elaborato un processo che guida gli studenti nella selezione di argomenti rilevanti per sé stessi o per le loro comunità, incoraggiandoli a formulare domande cruciali in varie fasi del loro progetto:

- rilevanza dell'argomento: "Perché scelgo questo argomento e in che modo è legato a me e alla mia comunità?"
- fonte e credibilità dei dati: "Da dove provengono i dati e quanto sono completi e affidabili?"
- organizzazione e spiegazione: "Quali sono le prove alla base della mia argomentazione e quali implicazioni emergono dai risultati?"
- rappresentazione e comunicazione: "Che cosa ho appreso dai dati e come posso trasmetterlo in modo efficace?"

L'autore ha osservato come le competenze pregresse degli studenti sui dati siano state acquisite soprattutto attraverso esperienze con sondaggi ed esperimenti di laboratorio, mentre la natura situata e autentica dell'esercizio può supportare una visione critica e complessa dell'uso dei dati nei contesti autentici di risoluzione dei problemi. L'attività di comprendere un argomento da diverse prospettive e di costruire una rappresentazione coerente ha richiesto l'utilizzo di diversi tipi di dati e strumenti di rappresentazione, nonché la valutazione dell'adeguatezza e dell'efficacia della sua comunicazione.

Le iniziative finalizzate alla promozione delle competenze cri-

tiche sui dati devono trovare strategie efficaci per superare alcuni ostacoli lungo il percorso. Ad esempio, le prospettive sociopolitiche e critiche non emergono in modo spontaneo durante le attività e il loro sviluppo deve essere supportato in modo coerente dagli insegnanti. Oltre a questo, la discussione delle disuguaglianze sociali e sistemiche può richiedere agli educatori uno sforzo in termini di sensibilità e di prospettiva inclusiva che permetta di inquadrare i temi senza arrecare danni ai soggetti più vulnerabili (Louie, 2022). Pertanto, i progettisti di formazione e gli insegnanti/educatori devono prepararsi adeguatamente ad affrontare le molteplici difficoltà che possono ostacolare lo sviluppo della consapevolezza sociopolitica, sia del mondo circostante che dei dati stessi. In questo contesto, seguire percorsi dedicati e specifici, oltre ad essere affiancati da soggetti esperti, può rivelarsi fondamentale per proteggere dall'insorgere di tali problematiche in aula. Come per tutte le innovazioni delle pratiche didattiche, la formazione è una componente essenziale, ma in questo contesto gli aspetti critico-etici meritano particolare attenzione, come avviene per gli interventi di *media education* (Rivoltella, 2019; Rivoltella, Rossi, 2019). La discussione di questi temi, talvolta controversi e divisivi, richiede un'ampia conoscenza delle questioni in gioco e anche una certa familiarità con le circostanze in cui tali questioni possono emergere (Ranieri, Fabbro, Nardi, 2019).

Esemplificativo di un approccio educativo volto a contrastare le distorsioni che influenzano la percezione dei fatti, “Dollar Street”⁹, all'interno del sito Gapminder, è la rappresentazione fotografica della vita quotidiana in centinaia di case di diversi livelli di reddito in tutto il mondo, raccontate tramite l'analogia con una strada immaginaria in cui i più poveri vivono ai numeri civici più bassi e i più ricchi ai numeri più alti. La visualizzazione inte-

9 Fonte: <https://www.gapminder.org/dollar-street>

rattiva dello strumento permette di esplorare e confrontare gli aspetti comuni della vita quotidiana, attraverso le culture e i continenti, con il fine di sfatare gli stereotipi e offrire una visione più autentica, ma non per questo oggettiva. Questa rappresentazione intende infatti sollevare questioni critiche, stimolare riflessioni e mettere in evidenza le implicazioni più profonde dei dati presentati. La scheda didattica illustra un possibile utilizzo di questo strumento con l'obiettivo di favorire lo sviluppo di competenze critiche all'uso dei dati e alle relative rappresentazioni.

Scheda didattica 3

Decostruzione delle rappresentazioni attraverso i dati

Obiettivo: Esplorare le rappresentazioni offerte da una *dataviz* interattiva e realizzare un confronto sulle vite delle persone descritte dai dati. L'attività si rivolge a studenti di scuola primaria e secondaria di primo grado. Con adattamenti riguardo alla complessità del confronto tra famiglie e delle connessioni interdisciplinari può essere proposta anche ad alunni della scuola secondaria di secondo grado.

Procedura:

1. *Esplorazione:* Assegnare ai partecipanti il compito di visitare “Dollar Street”, familiarizzando con la struttura della *dataviz* ed eseguendo il *quick tour*.
2. *Analisi:* Chiedere di selezionare tre diverse case, caratterizzate da un reddito simile ma geograficamente distanti, e di esaminare con attenzione le fotografie e le informazioni di ciascun nucleo familiare. I partecipanti descriveranno brevemente le vite delle famiglie visitate, includendo informazioni sul loro stile di vita, le loro occupazioni, le loro abitudini quotidiane e altro ancora.
3. *Confronto:* In seguito, chiedere agli studenti di comparare le vite di queste famiglie, confrontando le fotografie di almeno 5 elementi diversi (es. cibo, abbigliamento, mobili, mezzi di trasporto, infrastrutture domestiche). I partecipanti utilizzeranno le osservazioni realizzate durante la loro visita virtuale per evidenziare le differenze e le somiglianze tra le famiglie. Infine, prepareranno un resoconto finale, operando collegamenti con temi come globalizzazione culturale ed economica e/o disuguaglianza sociale.
4. *Discussione:* Riflettere sul tipo di messaggio che i dati veicolano, considerando sia i dati numerici come il reddito, che visuali come le fotografie. Discutere su quale sia

la finalità del progetto “Dollar Street”, su quale storia venga raccontata attraverso la rappresentazione visiva dei dati e sull’efficacia di tale rappresentazione.

3.2.4 *Data literacy: dimensione generativa e creativa*

Il *data storytelling* rappresenta il metodo che integra la visualizzazione dei dati in una narrazione che comunica un messaggio, selezionando e organizzando le intuizioni emerse dall’analisi dei dati (Bennato, 2018). Una *dataviz* infatti può essere concepita come una forma di narrazione lineare, diretta dall’autore, oppure interattiva, coinvolgendo il pubblico nell’esplorazione attraverso verifiche, nuove domande e spiegazioni alternative (Lee, Riche, Isenberg, Carpendale, 2015). In alcuni casi, gli autori hanno voluto creare un’esperienza intensa attraverso i dati, dando priorità all’impatto emotivo piuttosto che fornire un’analisi puramente obiettiva dei dati. In una “data-visualization emozionale”¹⁰ l’obiettivo è quello di trasmettere un messaggio che coinvolga emotivamente il pubblico (Cairo, 2013).

Nel discutere il processo di trasformazione dei dati in storie, Lee e colleghi (2015) descrivono le *visual data stories* attraverso alcuni elementi chiave. Innanzitutto, le storie di dati visivi sono costituite da un insieme di fatti specifici che sono corroborati da dati empirici. In secondo luogo, la maggior parte di tali fatti è presentata attraverso rappresentazioni grafiche al fine di trasmettere uno o più messaggi, di solito arricchite da annotazioni o nar-

10 Un esempio efficace di data-visualization emozionale è rappresentato dal progetto U.S. Gun Deaths di Perisopic, dove viene visualizzato con enfasi drammatica il numero di persone morte prematuramente perché uccise da un’arma da fuoco negli Stati Uniti (fonte: <https://guns.perisopic.com/>).

razioni atte a chiarire ed enfatizzare aspetti rilevanti. Infine, la struttura e l'organizzazione di queste porzioni narrative sono progettate per sostenere l'obiettivo comunicativo generale degli autori, ad esempio informare, intrattenere o persuadere il pubblico mediante argomentazioni convincenti.

La ricercatrice Echeverria e colleghi (2017) hanno elaborato una serie di regole per la realizzazione di un *data storytelling* visivo e hanno associato a ciascuna regola l'insieme di azioni corrispondenti che possono essere eseguite sulle visualizzazioni (tabella 3.3). Un esempio di riprogettazione della visualizzazione tradizionale è illustrato dalla figura 3.1.

Regole	Elementi di storytelling con i dati
Definire con chiarezza l'obiettivo	Identificare finalità e pubblico
Rimuovere elementi che non aggiungono valore informativo	Rimuovere etichette dei dati, marcatori, griglie, legende, segni di spunta ed etichette degli assi
Utilizzare sapientemente le narrazioni	Aggiungere le etichette testuali sui punti critici dei dati Oscurare le aree meno rilevanti per concentrare le informazioni
Attirare l'attenzione	Evidenziare con i colori Sottolineare i dati più importanti Rendere le linee più spesse
Chiamare all'azione	Aggiungere un titolo descrittivo che fornisca una panoramica immediata dei dati.

Tab. 3.3: Principi per la progettazione di un *data storytelling* (adattato da Echeverria, Martinez-Maldonado, Buckingham Shum, 2017)

La figura 3.1 mette a confronto il tipico output di uno strumento di visualizzazione utilizzato per tracciare il numero di visite per modulo in una classe online durante una settimana, con la stessa visualizzazione arricchita da elementi di *data storytelling*.

Confrontando entrambi i grafici, possiamo osservare che l'informazione più rilevante, ossia l'aumento delle visite per il Modulo 4, risulta più evidente nell'esempio di destra, senza che l'utente si perda nell'esplorazione della *dataviz*.



Fig. 3.1: A sinistra, un grafico tradizionale e, a destra, la visualizzazione degli stessi dati progettata secondo i principi del *data storytelling* (Echeverria et al., 2018)

Similmente alle altre dimensioni della *data literacy*, le competenze relative alla creazione e comunicazione di messaggi attraverso i dati non sono riservate agli esperti, ma possono essere apprese e applicate anche dal pubblico generale (D'Ignazio, 2017). Abbiamo già assistito a forme di partecipazione dal basso nella raccolta e creazione di dati, come nel caso dell'attivismo dei dati, che utilizza tecniche e strumenti di *data storytelling* per comunicare messaggi sociali rilevanti per la comunità.

Il concetto di *data humanism* promuove un approccio individualizzato nell'acquisizione, analisi e rappresentazione dei dati, evidenziando l'importanza della soggettività e del contesto nella comprensione degli eventi e dei mutamenti sociali, specialmente quando i dati riguardano gli individui (Lupi, 2017):

Disegnare con i dati [...] aiuta a rivelare nuove possibili analisi da eseguire: invece di lasciarci sopraffare dalle dimensioni di un dataset e da milioni di numeri, ci concentriamo solo sulla loro natura, sulla loro organizzazione, e così facendo spesso si aprono nuove opportunità che scaturiscono da questo punto di vista (para 3).

A questo riguardo, si possono segnalare due esperienze che consentono di delineare i benefici di un approccio creativo (e non tecnologico) alla *data literacy* in un'ottica critica. La prima è stata realizzata dall'informatica Sturdee e colleghi (2022), i quali hanno condotto uno studio finalizzato a esplorare le potenzialità espressive della creazione con i dati, al fine di approfondire la comprensione delle necessità e delle modalità con cui le persone si avvicinano a tale attività. Attraverso un metodo pratico e artistico per la manipolazione dei dati (*data painting*), i partecipanti sono stati coinvolti in modo personale nella creazione delle visualizzazioni e hanno dimostrato di poter analizzare i dati in maniera più approfondita rispetto a quanto avrebbero fatto tramite l'uso convenzionale di software per la creazione di grafici (figura 3.2).



Fig. 3.2: Alcune opere dei partecipanti alla sessione di *data painting* (Sturdee, Knudsen, Carpendale, 2022)

Inoltre, situando la *data literacy* nel contesto della gestione dei dati personali, Acker e Bowler (2017) hanno realizzato un'attività creativa (*data silhouette*) per i giovani, concepita come uno strumento per inquadrare le competenze necessarie per creare, raccogliere, gestire e curare set di dati nella vita quotidiana. È stato chiesto ai partecipanti di creare una visualizzazione disegnata a mano sui dati personali tracciati da hotspot WIFI, dispositivi e social media e di connetterli graficamente attraverso la creazione di un circuito. Un tale approccio, coinvolgendo i partecipanti in modo più diretto ed emotivo, può stimolare a pensare e a esplorare i dati in modo non convenzionale e autonomo.

“Dear Data”¹¹ rappresenta un progetto di *data visualisation* nel quale, per un anno intero, le designer Lupi e Posavec (2016) hanno impiegato il disegno tradizionale per elaborare e comunicare i dati relativi alle loro vite quotidiane. Il progetto con oltre 50 cartoline illustrate ha coperto una vasta gamma di argomenti, dalle abitudini personali alle interazioni sociali e all'ambiente urbano, con la finalità di esplorare il potenziale dei dati come mezzo di narrazione, espressione personale e comprensione del mondo circostante. Attingendo da tale esempio, la scheda didattica di questa sezione è un esercizio in cui gli studenti hanno l'opportunità di diventare autori e designer dei propri dati e di apprendere che è possibile controllare il proprio *data storytelling* decidendo quali dati raccogliere, condividere o evidenziare.

11 Fonte: <https://www.dear-data.com/theproject>

Scheda didattica 4

Raccolta, documentazione e racconto con i dati

Obiettivo: Coinvolgere i partecipanti nell'esplorazione e nella rappresentazione creativa dei propri dati, riconoscendosi un ruolo attivo nel definire e produrre una storia data-based. L'attività si rivolge a studenti di scuola secondaria di primo e secondo grado.

Procedura:

1. *Raccolta dei dati:* Iniziare chiedendo ai partecipanti di raccogliere e categorizzare i dati che riguardano le app presenti sui loro telefoni cellulari. In una tabella far annotare individualmente il nome delle app, la categoria a cui appartengono (es. social media, giochi, produttività, salute, ecc.) e la frequenza con cui vengono utilizzate (es. mai utilizzata, usata una o due volte, ogni tanto, con regolarità, ogni giorno).

2. *Creazione della visualizzazione:* Supportare i partecipanti alla produzione di una legenda, associando una rappresentazione visiva a ciascuna caratteristica delle app. Ogni app è rappresentata da un cerchio, il genere di app è descritto dal colore (es. rosso per i social network) e il riempimento della forma circolare individua la frequenza d'uso (da vuoto per lo scarso tempo di utilizzo fino a pieno per l'uso quotidiano). Chiedere ai partecipanti di realizzare su un foglio il loro *data storytelling*, in cui disegneranno i simboli che rappresentano le app in ordine di frequenza d'uso, dalla meno alla più utilizzata. Ad esempio, un utente molto affezionato ad Instagram, lo inserirà in fondo alla lista come un cerchio pieno (ovvero massima frequenza d'uso) di colore rosso (ovvero social media) e così via.

3. *Personalizzazione:* Poiché le app sul telefono offrono una prospettiva sull'identità e sulle abitudini digitali, chiedere ai partecipanti quali elementi vorrebbero che emergessero e quali fossero esclusi dal *data storytelling*. In-

vitare a personalizzare la propria rappresentazione, aggiungendo commenti e inventando altri criteri visivi, come evidenziare le app preferite o quelle che si sono pentiti di aver installato.

4. *Discussione*: Riflettere in gruppo sulle diverse tendenze emerse, sul significato dei dati rappresentati e sui messaggi che si desidera comunicare attraverso la loro visualizzazione. È possibile confrontare le visualizzazioni, individuare somiglianze e differenze e riflettere sulle implicazioni dei dati raccolti sulla propria vita digitale.

Esercitazione ispirata all'attività "Il mio telefono" illustrata in Lupi e Posavec (2018)

3.3 Oltre l'aula scolastica: la *data literacy* per gli educatori

3.3.1 *Sviluppare le competenze sui dati nella formazione degli insegnanti*

La *data literacy* per l'insegnamento rappresenta un sottoinsieme dell'alfabetizzazione ai dati e consiste nella raccolta, l'analisi e l'interpretazione di una vasta gamma di dati degli studenti, inclusi quelli relativi alla valutazione, al clima scolastico, ai comportamenti, nonché l'uso di dati prodotti in tempo reale e misure protratte nel tempo (Gummer, Mandinach 2015). Tale competenza è cruciale per guidare le decisioni didattico-pedagogiche, integrando la comprensione dei dati con i contenuti disciplinari, la struttura del curriculum, la progettazione e la valutazione dei processi di apprendimento degli studenti (Ndukwe, Daniel, 2020). Nel capitolo 4 saranno approfonditi alcuni strumenti per migliorare l'apprendimento e l'insegnamento attraverso l'uso dei dati, compresi i big data, mentre in questa sede ci concentreremo sulle abilità di base e sul loro ruolo nella formazione degli insegnanti.

Le ricercatrici statunitensi Mandinach e Gummer (2016)

hanno individuato tre ragioni principali a sostegno dello sviluppo della *data literacy* per gli insegnanti. Innanzitutto, si evidenzia l'importanza che i docenti acquisiscano competenze nell'utilizzo dei dati, specialmente in considerazione dell'interesse crescente a livello politico per l'adozione di pratiche educative evidence-based (Calvani, 2011). In secondo luogo, emerge una crescente confusione tra *data literacy* e *assessment literacy*, fenomeno che ostacola la diffusione globale della prima nel settore dell'istruzione, sebbene sia necessario comprendere che i dati non si limitano solo alle valutazioni. Infine, per quanto riguarda la preparazione degli insegnanti, introdurre l'uso dei dati una volta che sono già inseriti nel lavoro potrebbe risultare tardivo, mentre una precoce esposizione alla *data literacy* a livello di istruzione universitaria potrebbe offrire molteplici vantaggi. A maggior ragione, dal momento che nei programmi formativi per sviluppare le competenze legate ai dati degli studenti in modo efficace queste ultime dovrebbero essere collegate ed integrate con i contenuti curricolari e le metodologie didattiche sperimentate dai futuri insegnanti.

A questo proposito, il concetto di *pedagogical data literacy* connette la dimensione di instructional design con la comprensione critica dei dati, quali prove ed evidenze per favorire processi educativi data-based più partecipativi e trasparenti (Raffaghelli, 2018b). In altre parole, la competenza relativa ai dati per insegnare si concentra sul loro utilizzo per informare e migliorare le pratiche educative in modo più completo e efficace, con la considerazione del contesto e degli altri attori coinvolti nel processo (tabella 3.4). Il tema dell'uso di sistemi automatizzati per l'elaborazione dei dati ai fini di migliorare l'insegnamento sarà affrontato in dettaglio nel capitolo 4.

III. Didattica della *data literacy* ed esperienze di ricerca nel panorama educativo

Fasi del processo	Competenze operative
Identificare i problemi	Concentrarsi su un problema della pratica educativa, comprendere il contesto del problema relativo allo studente e all'istituzione scolastica, coinvolgere gli stakeholder (colleghi, studenti, famiglie), discutere le questioni relative alla privacy.
Utilizzare i dati	Identificare le fonti dei dati e relativi scopi, proprietà e qualità (accuratezza, adeguatezza e completezza); capire come vengono generati i dati e come possono essere estratti; utilizzare misure/fonti diversificate (quantitative-qualitative) di dati; comprendere come organizzare, analizzare, gestire, aggregare e disaggregare i dati. Valutare l'uso dei dati per ciascuna sequenza della pratica didattica: progettazione, sviluppo, implementazione e valutazione.
Trasformare i dati in informazioni	Considerare l'impatto e le conseguenze del processo di indagine e delle decisioni basate sui dati. Comprendere come i dati possono essere visualizzati, rappresentati e condivisi; generare collegamenti ipotetici con l'istruzione; testare ipotesi; valutare modelli, causalità e tendenze; sintetizzare dati diversificati; trarre inferenze; riassumere e spiegare i dati.
Trasformare le informazioni in decisioni	Determinare le fasi successive del processo didattico; monitorare le prestazioni degli studenti; diagnosticare i bisogni e le esigenze degli studenti; apportare aggiustamenti al percorso didattico; comprendere il contesto in cui si inserisce la loro decisione (contenuto disciplinare, curriculum, etc.).
Valutare i risultati	Riesaminare la domanda o il problema iniziale; confrontare le prestazioni prima e dopo la decisione; monitorare i cambiamenti nelle pratiche in classe e nelle prestazioni degli studenti. Considerare la natura iterativa dei cicli decisionali basati sui dati nel processo di insegnamento e apprendimento.

Tab. 3.4: Elementi chiave per una pedagogical data literacy (rielaborazione da Mandinach, Gummer, 2016 e Raffaghelli, 2018b)

Gli insegnanti necessitano di diverse esperienze per sviluppare le abilità legate all'uso quotidiano dei dati nel corso delle loro carriere, dalla preparazione pre-servizio fino allo sviluppo continuo delle competenze professionali (Ridsdale et al., 2015). Un esem-

pio di iniziativa di base rivolta ai docenti è il progetto Erasmus+ “Data Literate”, finalizzato a potenziare le capacità e abilità critiche nell’uso dei dati per i docenti delle scuole secondarie, a cui offre un MOOC gratuito e disponibile a tutti¹².

Dopo aver esaminato la letteratura sullo sviluppo professionale dei docenti, Henderson e Corry (2021) hanno formulato alcune raccomandazioni riguardanti i programmi di *data literacy*: implementare programmi presso gli istituti di istruzione superiore incentrati sulle competenze; promuovere opportunità collaborative per i partecipanti; modellare e incoraggiare l’uso dei dati provenienti sia da fonti quantitative che qualitative; indagare sul ruolo della tecnologia e dei big data nel processo di alfabetizzazione ai dati. Tra gli interventi finalizzati a migliorare l’uso dei dati tra gli insegnanti in servizio, risalta anche il ruolo delle comunità di apprendimento professionale e del coaching con gli esperti (Marsh, Bertrand, Huguet, 2015). Questi interventi svolgono un ruolo significativo nello sviluppo della *data literacy* dei docenti, influenzando il modo in cui apprendono e applicano tali competenze. L’efficacia del coaching emerge come un elemento distintivo, con un impatto più duraturo nel determinare cambiamenti nell’insegnamento, mentre nelle comunità di apprendimento professionale gli insegnanti tendono a condividere strategie specifiche e pratiche d’uso, operando principalmente come centro di diffusione e scambio di idee.

- 12 I sei moduli del corso online riguardano: 1) introduzione ai concetti di dato e opinione, sottolineando la non neutralità dei dati e l’importanza del fact-checking; 2) l’esplorazione di formati e licenze per la gestione dei dati, inclusa la rappresentazione visiva; 3) l’analisi e la pulizia dei dati e le statistiche di base rilevanti per gli insegnanti; 4) la presentazione di strumenti e strategie di comunicazione visiva dei dati; 5) la riflessione sulla relazione tra dati e insegnamento, mostrando come la *data literacy* possa supportare l’apprendimento; 6) la progettazione guidata di un piano di *digital data literacy* per gli studenti, attraverso l’esempio di una buona pratica scolastica (fonte: <https://www.dataliterateproject.eu/>).

L'incremento del ruolo della tecnologia nei processi decisionali basati sui dati è una tendenza evidente, soprattutto con l'avvento di sistemi di intelligenza artificiale e *machine learning* che forniscono raccomandazioni basate sui dati degli studenti. Tali implicazioni saranno discusse nel capitolo 4 (par. 4.3), tuttavia in questa sede è opportuno premettere che la *data literacy* è considerata un prerequisito per l'applicazione e l'uso dei learning analytics e dell'educational data mining nella pratica educativa, considerando queste ultime come forme sofisticate di decisioni basate sui dati. La discussione attuale verte sul livello di competenza nell'analisi dei dati necessario per adottare strumenti di learning analytics e sulle responsabilità per la loro adozione nei contesti educativi, dal momento che con la crescente complessità nell'uso dei dati, sussistono rischi legati ai *bias* e ai problemi etici (Gabbi, 2023; Mandinach, Abrams, 2022). Le competenze di *data literacy* diventano quindi fondamentali non solo per comprendere le raccomandazioni generate da tali sistemi, ma anche per favorire la fiducia degli educatori nell'adozione di tali tecnologie.

La sfida è capire come sfruttare a pieno il valore dei learning analytics e assicurarsi che gli insegnanti non solo siano in grado di lavorare con dati complessi, ma soprattutto di utilizzare tali *insight* per informare i diversi momenti delle proprie pratiche educative: supporto alla progettazione didattica, identificazione dei bisogni degli studenti con relativa azione d'intervento a loro beneficio e riflessione sulla pratica di insegnamento (Gabbi, 2021). Trasformare i dati dei contesti educativi in azioni concrete rappresenta, infatti, da un lato, l'essenza della funzione della *pedagogical data literacy* per i docenti, dall'altro, una finalità prioritaria dei learning analytics, secondo cui i dati dovrebbero essere estratti ed analizzati solo qualora comportino un reale beneficio per i soggetti che li cedono (Clow, 2012).

Infine, le disposizioni, le credenze e gli atteggiamenti degli insegnanti possono influenzare l'utilizzo efficace dei dati (Schil-

dkamp, Poortman, 2015). Bocala e Boudett (2015) hanno sottolineato il ruolo positivo di alcune abitudini mentali, sia per gli insegnanti in formazione che per quelli già in servizio, quali la creazione di mindset incentrati alla collaborazione, l'uso delle evidenze nella presa di decisioni e l'adesione a una sequenza ciclica del processo di decision-making basato sui dati. Analogamente, Mandinach e Gummer (2016) hanno identificato diversi fattori che supportano questo processo, tra cui:

- la riflessione critica ed etica sull'uso dei dati;
- la convinzione che tutti gli studenti siano capaci di apprendere;
- la consapevolezza che gli insegnanti abbiano un impatto significativo sul rendimento degli studenti.

Le autrici hanno altresì sottolineato l'importanza della capacità di comunicare in modo efficace con studenti, genitori e altre parti interessate, adattandosi alle loro esigenze. Tuttavia, le disposizioni, qualora implicite e non consapevoli, possono anche avere conseguenze negative, per esempio nel caso dell'adesione a pratiche d'uso riduttive e dettate dall'ideologia del *dataism* che sovrastima le potenzialità dei dati (Ali, 2023) o riguardo la diffusione di idee distorte in relazione alle decisioni data-based (Mandinach, Schildkamp, 2021). Date le potenziali implicazioni negative per gli studenti derivanti da interpretazioni errate causate da una mancanza di competenza nell'uso e nella lettura dei dati (Cui, Zhang, 2022), è fondamentale lavorare sullo sviluppo delle competenze degli insegnanti in questo ambito.

3.3.2 *Altri attori rilevanti nella prospettiva dell'educazione ai dati*

L'alfabetizzazione ai dati riveste un ruolo cruciale nel trasformare i dati in informazioni utili per migliorare la qualità dell'educa-

zione, tuttavia studenti e personale docente non sono i soli a poter supportare tale trasformazione in un sistema complesso. Una prospettiva critica è auspicata ad esempio nell'ambito socioeducativo: gli educatori che spesso esercitano la loro professione in contesti sociali svantaggiati possono interrogarsi sul ruolo della conoscenza quantitativa e basata su dati nella loro pratica professionale (Ranieri, Biagini, Cuomo, Gabbi, 2023). Anche i genitori e il pubblico in generale condividono la preoccupazione per la crescente enfasi e dipendenza dai dati e hanno sollevato interrogativi in merito ai tipi di dati raccolti sugli studenti, al loro utilizzo, alla sicurezza e alla gestione dell'accesso alle informazioni personali (Mandinach, Parton, Gummer, Anderson, 2015). Il coinvolgimento di tutte le parti interessate, compresi i dirigenti, il personale non docente e le autorità istituzionali, rappresenta un passaggio cruciale nell'adozione di un approccio efficace alla *data literacy* e all'uso dei dati in campo scolastico (Fornasari, 2017). In aggiunta, settori come la ricerca scientifica e l'ambito bibliotecario sono particolarmente sollecitati su questioni relative alla crescente pervasività dei dati nella società e nella cultura contemporanea.

La cultura dell'uso dei dati nella scuola esercita un'influenza significativa sul modo in cui gli educatori concepiscono e utilizzano i dati e i dirigenti scolastici rivestono un ruolo cruciale nel plasmare tale percezione condivisa (Henderson, Corry, 2021). Un leader potrebbe influenzare negativamente l'approccio e la percezione circa i dati scolastici qualora mettesse in atto tattiche coercitive o di controllo, quali l'uso di tali informazioni per incolpare o svalutare il docente, limitando così la fiducia degli insegnanti nell'esprire i propri dubbi e nell'assumere rischi ponderati che potrebbero portare benefici agli studenti (Schildkamp, Poortman, 2015).

In relazione al livello di trasformazione locale, i processi decisionali basati sui dati nelle istituzioni scolastiche possono essere ottimizzati attraverso una collaborazione sinergica tra insegnanti

e dirigenti, sfruttando una vasta gamma di fonti di dati per informare e migliorare le pratiche educative. Secondo la professoressa Schildkamp e collaboratori (2019), per i dirigenti scolastici che intendono istituire un *data team* efficace all'interno del proprio istituto, sono essenziali cinque elementi chiave:

- chiarire e comunicare in modo esplicito la prospettiva, le norme e gli obiettivi dell'utilizzo dei dati;
- offrire supporto personalizzato al gruppo, adattato alle esigenze e alle competenze individuali;
- mantenere uno stimolo intellettuale attraverso la condivisione della conoscenza, la garanzia di autonomia e la leadership condivisa;
- creare e promuovere un clima aperto e trasparente riguardo all'utilizzo dei dati nell'ambiente scolastico;
- sviluppare una rete di scambio delle conoscenze per facilitare la condivisione e la diffusione delle migliori pratiche e delle scoperte nel campo dell'utilizzo dei dati educativi.

In virtù dello stretto legame concettuale tra *data literacy* e *information literacy* (cfr. capitolo 2, par. 2.2), anche i bibliotecari possono svolgere una funzione cruciale nel supportare gli studenti ad acquisire competenze relative alla comprensione e all'analisi dei dati (Calzada Prado, Marzal, 2013). Tale ruolo si estende oltre l'ambiente scolastico, fungendo da agente promotore di connessioni interdisciplinari che possono facilitare l'integrazione di concetti e competenze provenienti da diverse aree di studio. In questo contesto, tra le mansioni del bibliotecario rientra la promozione di un cambiamento culturale orientato verso una ricerca condivisa, coinvolgendo sia i produttori sia i riutilizzatori dei dati (Orrù, 2020). Questo richiede un impegno per facilitare il trasferimento tecnologico necessario per la scienza aperta, tramite una collaborazione sinergica a tutti i livelli operativi, finanziari e decisionali dell'istituzione. Inoltre, il personale delle biblioteche

può supportare piattaforme di scambio per informazioni pratiche, risposte a domande frequenti e soluzioni replicabili, anche attraverso gruppi di discussione per il mutuo sostegno, come per esempio l'iniziativa internazionale Research Data Alliance¹³. Le ricercatrici statunitensi Fontichiaro e Oehrli (2016) hanno identificato sei temi prioritari per i bibliotecari delle scuole superiori statunitensi:

- comprensione dei dati e delle statistiche, inclusi i termini tecnici
- creazione e interpretazione di visualizzazioni dei dati
- costruzione e critica di argomentazioni basate sui dati
- comprensione delle promesse e dei pericoli legati ai big data
- gestione dei dati personali e del *digital footprint*, la traccia di informazioni digitali che viene lasciata attraverso le attività online
- comportamento etico nell'uso, nella raccolta e nella rappresentazione dei dati.

In Italia, il forte legame tra settore bibliotecario e *open science* ha permesso la nascita di occasioni formative sulla gestione dei dati, in particolare relativi alla ricerca scientifica (Tammaro, 2017). Secondo i principi dell'*open science*, i dati di ricerca sono considerati risorse comuni e pubbliche che richiedono un'adeguata gestione per essere condivise e riusate. La frequente richiesta

13 La Research Data Alliance (RDA), con oltre 14.500 membri provenienti da 151 paesi, offre uno spazio gratuito e aperto a tutti per sviluppare e adottare infrastrutture che promuovano la condivisione dei dati e la ricerca data-based. Fondata nel 2013, la RDA facilita la condivisione e il riutilizzo dei dati tra diverse tecnologie, discipline e nazioni. La sua missione è costruire legami sociali e tecnici per ridurre le barriere alla condivisione dei dati, coinvolgendo scienziati, ricercatori ed esperti in gruppi di lavoro e comunità di pratica (fonte: <https://www.rd-alliance.org/>).

di fornire un Data Management Plan (DMP) da parte dei programmi di finanziamento dell'Unione Europea ha portato maggiore attenzione su questa tematica nell'ottica di ottemperare alle limitazioni della normativa europea General Data Protection Regulation (GDPR) e di rendere i dati della ricerca aderenti al modello FAIR: cioè reperibili (*findable*), accessibili (*accessible*), interoperabili (*interoperable*) e riutilizzabili (*re-usable*) (Wilkinson et al., 2016). Il piano di gestione dei dati dovrebbe comprendere informazioni dettagliate riguardanti le diverse fasi del ciclo di vita dei dati nella ricerca (Commissione Europea, 2021). In primo luogo, è essenziale chiarire come i dati di ricerca saranno gestiti durante il corso del progetto e dopo la sua conclusione. Ciò include procedure per la raccolta, l'elaborazione e la generazione dei dati, nonché piani per il trattamento dei dati una volta terminato il progetto. Inoltre, il DMP dovrebbe specificare quali dati verranno raccolti, processati e/o generati durante lo svolgimento della ricerca, insieme alla metodologia e agli standard che saranno applicati per garantire la qualità e l'affidabilità dei dati stessi. Una componente importante del DMP riguarda la condivisione dei dati e l'accesso aperto: è necessario indicare se e in che modo i dati saranno condivisi con altri ricercatori o resi accessibili al pubblico. Infine, il piano dovrebbe includere strategie per la cura e la conservazione dei dati nel lungo termine, garantendo la loro disponibilità e accessibilità anche dopo la conclusione del progetto di ricerca. All'interno del Piano nazionale per la ricerca aperta (MUR, 2022) si promuove, grazie alle nuove tecniche di generazione automatizzata, l'impegno alla produzione di dati FAIR-by-design, che rispettino quindi tali principi fin dalla produzione, associando l'uso corretto dei metadati e delle appropriate informazioni contestuali che ne facilitino la ricerca e il riuso. Il Piano prevede anche un rafforzamento delle competenze attraverso azioni formative volte a "avviare la formazione di base per tutti i ricercatori e tecnologi sulla gestione dei dati FAIR" ed "estendere a tutti i curriculum formativi un'alfabetizzazione al-

l'utilizzo dei dati della ricerca" (*ivi*, p. 9). È quindi considerato una priorità che sia i docenti e i ricercatori che gli studenti comprendano l'importanza della gestione dei dati lungo l'intero ciclo della ricerca (Rivoltella, 2022; Viganò, 2020).

In conclusione, per avviare una riflessione circa l'importanza della *data literacy* nell'istruzione, è necessario discutere delle finalità e delle strategie per promuovere una formazione critica in questo campo. Con l'obiettivo di formare cittadini consapevoli nell'era dei dati, la ricerca e le esperienze dimostrano la coesistenza di diversi approcci nel contesto scolastico, che implicano considerazioni sullo sviluppo del curriculum, sulle sfide legate agli ordini scolastici e sulle pratiche di progettazione didattica e di valutazione. A tal fine, sono disponibili una serie di strumenti, esperienze e risorse per conoscere, visualizzare, interpretare e creare con i dati, affrontando diverse dimensioni della *data literacy*, dalla conoscitiva e visuale alla critico-etica fino alla dimensione creativa. Infine, è rilevante evidenziare il ruolo degli educatori, in considerazione della necessità di promuovere lo sviluppo delle competenze relative ai dati nel percorso formativo degli insegnanti, nonché del coinvolgimento di altri attori rilevanti nell'ambito scolastico. Tale partecipazione è essenziale per delineare soluzioni operative che siano coerenti con i principi pedagogici delle istituzioni educative e di ricerca, che considerino gli individui come soggetti attivi nella società dell'informazione e che promuovano modelli di agentività per i futuri cittadini.

Riferimenti bibliografici

- Acker, A., Bowler, L. (2017, July). What is your Data Silhouette? Raising teen awareness of their data traces in social media. In *Proceedings of the 8th International Conference on Social Media & Society* (pp. 1-5). New York: ACM.
- Ali, M. M. (2023). Datafication, Teachers' Dispositions and English

- Language Teaching in Bangladesh: A Bourdieuan Analysis. *Tesol Quarterly*, 57(4), 1282-1308.
- Anslow, C., Brosz, J., Maurer, F., Boyes, M. (2016, February). Data-thons: An experience report of data hackathons for data science education. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (pp. 615-620). New York: ACM.
- Azzam, T., Evergreen, S., Germuth, A. A., Kistler, S. J. (2013). Data visualization and evaluation. *New Directions for Evaluation*, 2013(139), 7-32.
- Barzilai, S., Chinn, C. A. (2017). On the goals of epistemic education: Promoting apt epistemic performance. *Journal of the Learning Sciences*, 27(3), 353-389.
- Bennato, D. (2018). Osservare tanti dati. Il ruolo della visualizzazione dei Big Data nella ricerca valutativa. *RIV. Rassegna Italiana di Valutazione*, 21(68), 63-83.
- Bhargava, R., D'Ignazio, C. (2015, April). Designing tools and activities for data literacy learners. In *Workshop on data literacy, Webscience*. Short paper, Oxford, UK.
- Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., Shoup, N. (2015). *Beyond data literacy: Reinventing community engagement and empowerment in the age of data*. DATA-POP Alliance White Paper Series. Retrieved June 7, 2024, from <https://datapopalliance.org/wp-content/uploads/2015/11/Big-Data-Official-Statistics-FINAL.pdf>
- Bhargava, R., Kadouaki, R., Bhargava, E., Castro, G., D'Ignazio, C. (2016). Data murals: Using the arts to build data literacy. *The Journal of Community Informatics*, 12(3), 197-216.
- Bocala, C., Boudett, K. P. (2015). Teaching educators habits of mind for using data wisely. *Teachers College Record*, 117(4), 1-20.
- Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. Cambridge: MIT press.
- Börner, K., Maltese, A. V., Balliet, R. N., Heimlich, J. (2016). Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15, 198-213. <https://doi.org/10.1177/147387-1615594652>
- Buckingham, D. (2022, aprile). *Who needs data literacy?*. Retrieved June 7, 2024, from <https://davidbuckingham.net/2022/04/20/who-needs-data-literacy/>

- Cairo, A. (2013). *L'arte funzionale: Infografica e visualizzazione delle informazioni*. Milano: Pearson Italia Spa.
- Calvani, A. (2011). «Decision Making» nell'istruzione. «Evidence Based Education» e conoscenze sfidanti. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal)*, 2(3), 77-99.
- Calvani, A., Fini, A., Ranieri, M. (2010). *La competenza digitale nella scuola. Modelli e strumenti per valutarla e svilupparla*. Trento: Erickson.
- Calzada Prado, J., Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, 63(2), 123-134.
- Carenzio, A., Rivoltella, P. C., Ferrari, S. (2021). A Media Diet Today: A Framework and Tool to Question Media Uses. In *Media Education at the Top* (pp. 1-14). Cambridge: Cambridge Scholars Publishing.
- Carmi, E., Yates, S. J. (2020). What do digital inclusion and data literacy mean today?. *Internet Policy Review*, 9(2). <https://doi.org/10.14763/2020.2.1474>
- Chignard, S. (2013). *A Brief History of Open Data*. Retrieved June 7, 2024, from <http://www.paristechreview.com/2013/03/29/brief-history-open-data/>
- Chin, D. B., Blair, K. P., Schwartz, D. L. (2016). Got game? A choice-based learning assessment of data literacy and visualization skills. *Technology, Knowledge and Learning*, 21, 195-210.
- Clow, D. (2012, April). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134-138). New York: ACM.
- Cobo Romani, J. C. (2009). *Strategies to promote the development of e-competencies in the next generation of professionals: European and international trends*. SKOPE Publications. Retrieved June 7, 2024, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1904871
- Commissione Europea (2021). *Horizon Europe. Data Management Plan Template*. Retrieved June 7, 2024, from <https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf>
- Corazza, P. (2022). *L'intelligenza collettiva al tempo delle piattaforme digitali. Il modello del formicaio: implicazioni pedagogiche e alternative possibili*. Roma: Armando.
- Cottone, A. M., Yoon, S. A., Shim, J., Coulter, B., Carman, S. (2023).

- Evaluating the apt epistemic processes of data literacy in elementary school students. *Instructional Science*, 51(1), 1-37.
- Cui, Y., Zhang, H. (2022). Integrating teacher data literacy with TPACK: A self-report study based on a novel framework for teachers' professional development. *Frontiers in Psychology*, 13, 966575.
- Cui, Y., Chen, F., Lutsyk, A., Leighton, J. P., Cutumisu, M. (2023). Data literacy assessments: A systematic literature review. *Assessment in Education: Principles, Policy & Practice*, 30(1), 76-96.
- D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1), 6-18.
- De Marco, E. L. (2023). La cultura del dato e la teacher data literacy: un caso d'uso del linguaggio di programmazione Python. *Lifelong Lifewide Learning*, 19(42), 570-579.
- Deahl, E. (2014). *Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments*. M.S. Thesis, Massachusetts Institute of Technology.
- Dewey, J. (1993). *Esperienza e educazione* (1938). Venezia: La Nuova Italia.
- Echeverria, V., Martinez-Maldonado, R., Buckingham Shum, S. (2017, November). Towards data storytelling to support teaching and learning. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (pp. 347-351). New York: ACM.
- Echeverria, V., Martinez-Maldonado, R., Buckingham Shum, S., Chiluiza, K., Granda, R., Conati, C. (2018). Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *Journal of Learning Analytics*, 5(3), 73-97.
- Fisher, D., Meyer, M. (2017). *Making data visual: a practical guide to using visualization for insight*. Sebastopol: O'Reilly Media, Inc.
- Fontichiaro, K., Oehrli, J. A. (2016). Why data literacy matters. *Knowledge Quest*, 44(5), 21-27.
- Fornasari, A. (2017). Social Privacy. Informare, comunicare, educare ai tempi del web 3.0. *Mondo Digitale*, 16(71), 1-13.
- Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., Hullman, J. (2021). The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest*, 22(3), 110-161. <https://doi.org/10.1177/15291006211051956>

- Fuchs, C. (2017). From digital positivism and administrative big data analytics towards critical digital and social media research!. *European Journal of communication*, 32(1), 37-49.
- Gabbi, E. (2021). La sfida dei Learning analytics per gli insegnanti tra ambienti digitali di apprendimento e competenze professionali. *Formazione & Insegnamento*, 19(2), 193-205.
- Gabbi, E. (2023). About or with Teachers? A Systematic Review of Learning Analytics Interventions to Support Teacher Professional Development, *QWERTY*, 18(2), 88-109.
- Gebre, E. H. (2018). Young adults' understanding and use of data: Insights for fostering secondary school students' data literacy. *Canadian Journal of Science, Mathematics and Technology Education*, 18(4), 330-341.
- Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G., Dey, S. (2023). A systematic literature review of data literacy education. *Journal of Business & Finance Librarianship*, 28(2), 112-127.
- Gummer, E. S., Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, 117(4), 1-22.
- Henderson, J., Corry, M. (2021). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning*, 14(2), 232-244.
- Hox, J. J., Boeije, H. R. (2005). Data collection, primary vs. secondary. *Encyclopedia of social measurement*, 1(1), 593-599.
- Ioannidis, A. (2020, April). *Fighting the Coronavirus Disease (COVID-19)*. Zenodo blog. Retrieved June 7, 2024, from <https://blog.zenodo.org/2020/04/02/fighting-the-coronavirus-disease-covid-19/>
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin, R., Lauriault, T. P., McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, 2(1), 6-28.
- Kjelvik, M. K., Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE—Life Sciences Education*, 18(2), es2.
- Kokotsaki, D., Menzies, V., Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving schools*, 19(3), 267-277.
- Kolb, D. A. (1984). *Experiential learning experience as the source of learning and development*. Hoboken: Prentice Hall.

- Krause, H. (2019). *An Introduction to the Data Biography*. Data Biography, Data Collection & Sourcing. Retrieved June 7, 2024, from <https://weallcount.com/2019/01/21/an-introduction-to-the-data-biography/>
- Larasati, P. E., Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. In *Journal of Physics: Conference Series* (Vol. 1440, No. 1, p. 012020). London: IOP Publishing.
- Lee, B., Riche, N. H., Isenberg, P., Carpendale, S. (2015). More than telling a story: Transforming data into visually shared stories. *IEEE computer graphics and applications*, 35(5), 84-90.
- Leighton, J., Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Livingstone, S., Stoilova, M., Nandagiri, R. (2019). *Children's data and privacy online: Growing up in a digital age. An evidence review*. London: London School of Economics and Political Science.
- Louie, J. (2022, September). Critical data literacy: Creating a more just world with data. In *Workshop on Foundations of Data Science for Students in Grades K-12*. Woodside: Valhalla Foundation.
- Lupi, G. (2017). *Data Humanism: The Revolutionary Future of Data Visualization*. PrintMag. Retrieved June 7, 2024, from <https://www.printmag.com/article/data-humanism-future-of-data-visualization/>
- Lupi, G., Posavec, S. (2016). *Dear data*. San Francisco: Chronicle books.
- Lupi, G., Posavec, S. (2018). *Osserva, raccogli, disegna!: un diario visivo: scopri i pattern nella tua vita quotidiana*. Mantova: Corraini.
- Mandinach, E. B., Abrams, L. M. (2022). Data literacy and learning analytics. In *The handbook of learning analytics* (pp. 196-204). New York: SOLAR, Society for Learning Analytics and Research.
- Mandinach, E. B., Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, 60, 366-376.
- Mandinach, E. B., Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature. *Studies in Educational Evaluation*, 69, 100842.
- Mandinach, E. B., Parton, B. M., Gummer, E. S., Anderson, R. (2015).

- Ethical and appropriate data use requires data literacy. *Phi Delta Kappan*, 96(5), 25-28.
- Marsh, J. A., Bertrand, M., Huguet, A. (2015). Using data to alter instructional practice: The mediating role of coaches and professional learning communities. *Teachers College Record*, 117(4), 1-40.
- Martin, E., Leger-Hornby, T. (2012). *Framework for a data management curriculum: course plans for data management instruction to undergraduate and graduate students in science, health science, and engineering programs*. Retrieved June 7, 2024, from https://library.umassmed.edu/pdfs/data_management_frameworks.pdf
- McCandless, D. (2011). *Information is beautiful. Capire il mondo al primo sguardo*. Milano: BUR Rizzoli.
- MUR. (2022). *Piano nazionale per la ricerca aperta*. PNR 2021-2027. Ministero dell'Università e della Ricerca. Retrieved June 7, 2024, from https://www.mur.gov.it/sites/default/files/2022-06/Piano_Nazionale_per_la_Scienza_Aperta.pdf
- Ndukwe, I. G., Daniel, B. K. (2020). Teaching analytics, value and tools for teacher data literacy: A systematic and tripartite approach. *International Journal of Educational Technology in Higher Education*, 17, 1-31.
- O'Neil, C. (2017). *Armi di distruzione matematica*. Milano: Bompiani.
- Onuoha, M. (2016, 10 Febbraio). *The Point of Collection*. Data & Society: Points. Retrieved June 7, 2024, from <https://points.datasociety.net/the-point-of-collection-8ee44ad7c2fa#.y0xtfxi2p>
- Orrù, D. (2020). Open data steward: bibliotecari e alfabetizzazione ai dati aperti. *AIB studi*, 60(2), 311-323.
- Panciroli, C., Macaudo, A. (2019). Spazi digitali per educare al Patrimonio: il MOdE, Museo Officina dell'Educazione. In *Studi avanzati di educazione museale. Lezioni* (pp. 49-62). Napoli: Edizioni Scientifiche Italiane.
- Pangrazio, L., Sefton-Green, J. (2020). The social utility of 'data literacy'. *Learning, Media and Technology*, 45(2), 208-220.
- Pangrazio, L., Selwyn, N. (2023). *Critical Data Literacies: Rethinking Data and Everyday Life*. Cambridge: MIT Press.
- Qin, J., D'Ignazio, C. (2010). Lessons learned from a two year experience in science data literacy education. *International Association of Scientific and Technological University Librarians, 31st Annual Conference*, Paper 5. Retrieved June 7, 2024, from <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>

- Raffaghelli, J. E. (2018a). Oltre il “far di conto” nell’era digitale. La frontiera della data literacy. In M. Ranieri (Ed.) *Teoria e pratica delle new media literacies*, (p. 99-133). Roma: Aracne.
- Raffaghelli, J. E. (2018b). Educators’ Data Literacy. Supporting critical perspectives in the context of a “datafied” education. In M. Ranieri, L. Menichetti, M. Kashny-Borges (Eds.) *Teacher education & training on ict between Europe and Latin America* (pp. 91-109). Roma: Aracne.
- Raffaghelli, J. E. (2023). Pathways for Social Justice in the Datafied Society: Reconsidering the educational response. *Media Education*, 14(1), 5-14.
- Ranieri, M., Biagini, G., Cuomo, S., Gabbi, E. (2023). “At the tip of data...”: Developing data literacy in educators’ professional development. *Italian Journal of Educational Technology (IJET)*, 31(2), 23-35.
- Ranieri, M., Cuomo, S., Biagini, G. (2024). *Scuola e intelligenza artificiale. Percorsi di alfabetizzazione critica*. Roma: Carocci.
- Ranieri, M., Fabbro, F., Nardi, A. (2019). *La media education nella scuola multiculturale. Teorie, pratiche, strumenti*. Pisa: ETS.
- Reeves, T. D., Chiang, J. L. (2019). Effects of an asynchronous online data literacy intervention on pre-service and in-service educators’ beliefs, self-efficacy, and practices. *Computers & Education*, 136, 13-33.
- Reeves, T. D., Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, 50, 90-101.
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., ... Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report*. Nova Scotia: Dalhousie University.
- Rivoltella, P. C. (2019). Media education. In P. C. Rivoltella, P. G. Rossi (Eds.), *Tecnologie per l’educazione* (pp. 127-138). Torino: Pearson Italia.
- Rivoltella, P. C. (2022). Critical Thinking in the Data Age. New Challenges. In *Humane Robotics. A multidisciplinary Approach toward the Development of Humane-centered Technologies* (pp. 327-342). Milano: Vita e Pensiero.
- Rivoltella, P. C., Rossi, P. G. (2019). *Il corpo e la macchina. Tecnologia, cultura, educazione*. Brescia: Scholé.

- Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know? *Journal of the Learning Sciences*, 29(1), 154-164.
- Schildkamp, K., Poortman, C. (2015). Factors influencing the functioning of data teams. *Teachers college record*, 117(4), 1-42.
- Schildkamp, K., Poortman, C. L., Ebbeler, J., Pieters, J. M. (2019). How school leaders can build effective data teams: Five building blocks for a new wave of data-informed decision making. *Journal of educational change*, 20(3), 283-325.
- Seel, N. M. (2017). Model-based learning: A synthesis of theory and research. *Educational Technology Research and Development*, 65, 931-966.
- Shreiner, T. L. (2017). Data Literacy for Social Studies: Examining the Role of Data Visualizations in K–12 Textbooks, *Theory & Research in Social Education*, 46(2), 194-231. <https://doi.org/10.1080/00933104.2017.1400483>
- Sturdee, M., Knudsen, S., Carpendale, S. (2022). Data-painting: Expressive free-form visualisation. In D. Lockton, S. Lenzi, P. Hekkert, A. Oak, J. Sádaba, P. Lloyd (Eds.), *DRS2022: Bilbao, 25 June - 3 July, Bilbao, Spain*. DRS Digital Library.
- Tamarro, A. M. (2017). Data literacy: formare docenti e studenti alla gestione dei dati di ricerca. *Biblioteche oggi*, 35, 19-25.
- Teal, T., Cranston, K., Lapp, H., White, E., Wilson, G., Ram, K., Pawlik, A. (2015). Data carpentry: Workshops to increase data literacy for researchers. *International Journal of Digital Curation*, 10(1), 135-143.
- Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2). Cheshire: Graphics press.
- Tygel, A.F. Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics*, 12(3), 108-121.
- Vahey, P., Rafanan, K., Patton, C., Swan, K., van't Hooft, M., Kratcoski, A., Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. *Educational Studies in Mathematics*, 81(2), 179-205. <https://doi.org/10.1007/s10649-012-9392-z>
- Van Audenhove, L., Van den Broeck, W., Mariën, I. (2020). Data literacy and education: Introduction and the challenges for our field. *Journal of Media Literacy Education*, 12(3), 1-5.

- Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & society*, 12(2), 197-208.
- Van Merriënboer, J. J., Kirschner, P. A., Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational psychologist*, 38(1), 5-13.
- Viganò, R. (2020). La validità della ricerca educativa tra criteri scientifici, contesti di pratica, responsabilità politica. *Pedagogia Oggi*, 18(1), 323-334.
- Wickham, H. (2014). Tidy data. *Journal of statistical software*, 59, 1-23.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.
- Wolff, A., Gooch, D., Cavero Montaner, J.J, Rashid, U., , Kortuem, G., (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, 12(3), 9-26.
- Wolff, A., Wermelinger, M., Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies*, 129, 41-54.
- Wu, X., Xu, T., Zhang, Y. (2023). Research on the data analysis knowledge assessment of pre-service teachers from China based on cognitive diagnostic assessment. *Current Psychology*, 42(6), 4885-4899.
- Zilinski, L., Scherer, D., Maybee, C. (2013). *Librarians as Consultants: The Convergence of Information Literacy, Data Literacy and Scholarly Communication in Undergraduate Research*. New Albany: Indiana University Libraries Information Literacy Colloquium.

IV.

Pratiche e tecniche di Educational Data Mining: metodi ed esempi

Gabriele Biagini

4.1 Elementi di Educational Data Mining

4.1.1 *L'Educational Data Mining: una definizione*

A quale scopo vengono usati i dati in ambito educativo? Come ottimizzare i benefici dei dati nel contesto della formazione? Quali metodi e quali tecniche possono essere utilizzate? Se nel capitolo 1 abbiamo parlato degli educational big data e nei capitoli 2 e 3 di *educational data literacy*, in questo capitolo conclusivo affronteremo i quesiti sopra menzionati, approfondendo quell'area di ricerca nota a livello internazionale con la formula Educational Data Mining (EDM).

Si tratta di un campo di studi oggetto di crescente interesse, che integra due componenti fondamentali: gli *educational data* e il (data) *mining* (DM). Come abbiamo già osservato (cfr. capitolo 1), i dati educativi si riferiscono a qualsiasi tipo di informazione raccolta nei contesti formativi, includendo, fra le altre cose, dati sul rendimento degli studenti, interazioni in classe e risorse didattiche utilizzate. Il termine (data) mining, che tradotto dall'inglese significa estrazione, si riferisce all'uso di tecniche e strumenti statistici e computazionali per scoprire schemi e modelli nascosti nei grandi set di dati (Nguyen, Gardner, Sheridan, 2020; Romero, Ventura, 2020). Mettendo insieme le due espressioni, possiamo dire che l'EDM si occupa di sviluppare, ricercare e applicare metodologie informatiche mirate all'identificazione di schemi e

modelli entro le ampie raccolte di dati relative al mondo della formazione. Per una sua definizione, gli studiosi Romero e Ventura (2013) fanno riferimento al sito web della comunità internazionale dell'EDM, che lo inquadra come “una disciplina emergente, focalizzata sullo sviluppo di strumenti per l'esplorazione di dati specifici provenienti dai contesti educativi e sull'applicazione di tali strumenti per un migliore inquadramento degli studenti e degli ambienti in cui si sviluppa il processo di apprendimento” (*Educational data mining consortium*¹, 2022). In effetti, come evidenziano alcuni ricercatori, l'obiettivo dell'EDM consiste nell'analisi di dati provenienti da ambienti formativi, al fine di affrontare e risolvere interrogativi rilevanti nel campo della ricerca educativa (Baker, Yacef, 2009; Scheuer, McLaren, 2012). In questa prospettiva, l'EDM si avvale delle tecniche di data mining per il trattamento di dataset specifici che provengono da diversi contesti istruttivi, ponendo così le basi per una risposta articolata a quesiti di rilievo nel campo dell'educazione (Ray, Saeed, 2018).

Numerosi studi (Aldowah, Al-Samarraie, Fauzy, 2019; Ampadu, 2023; Okewu, Adewole, Damasevicius, Maskeliunas, Misra, 2021; Romero, Ventura, 2020; Safitri, Setiadi, Suryani, 2022) hanno proposto di categorizzare gli obiettivi dell'EDM, basandosi sulla prospettiva dell'utente finale (il discente), nonché sulle sfide da affrontare. Questi obiettivi comprendono la previsione dei comportamenti degli studenti tramite il miglioramento dei modelli di classificazione, la scoperta o il perfezionamento dei modelli conoscitivi, l'impiego di sistemi di apprendimento per fornire un supporto pedagogico più efficace, insieme allo sviluppo di dati empirici per sostenere o definire teorie e framework pedagogici, al fine di identificare le componenti essenziali dell'apprendimento e promuovere la creazione di ambienti formativi più efficaci.

1 Fonte: <https://www.educationaldatamining.org>

Alla luce di questi obiettivi, possiamo osservare come l'EDM si estenda ben oltre le semplici interazioni degli studenti con i sistemi educativi digitali, il comportamento di navigazione o le risposte ai quiz, abbracciando anche dati collaborativi (es. chat testuali tra studenti), dati amministrativi (come quelli relativi a scuole o insegnanti), dati demografici e informazioni sulla motivazione e sugli stati emotivi degli studenti (Anjum, Badugu, 2020). Caratteristiche tipiche di questi dati sono la presenza di vari livelli gerarchici (es. materia, compito, domanda), contesto specifico (un determinato studente in una specifica classe oggetto di una data domanda in un momento preciso), dettagli minuziosi e aspetti longitudinali (dati raccolti su lunghi periodi di tempo).

L'EDM si configura, inoltre, come un'area interdisciplinare che spazia dal recupero di informazioni ai sistemi di raccomandazione, dall'analisi visiva dei dati fino all'analisi delle reti sociali, la psicopedagogia, la psicologia cognitiva e la psicometria (Baker, 2010; Ghorpade, Chaudhari, Patil, 2020; Govindarajan, 2021; Pal, Pal, 2013). In questo contesto, l'EDM si interseca con l'informatica, le teorie dell'educazione e la statistica, dando vita a sottocampi correlati come l'educazione basata sul computer (nei vari progetti di scuola digitale), il data mining (DM), l'apprendimento automatico (o *machine learning*) e i learning analytics (LA) (Long, Siemens 2011; Siemens, 2013). Tuttavia, come illustrato nel capitolo 3 (par. 3.3.1), mentre i LA si focalizzano sul processo decisionale basato sui dati, integrando dimensioni tecniche e socio-pedagogiche, l'EDM privilegia tecniche come la classificazione, il clustering e la modellazione. Occorre, però, precisare che esiste una somiglianza sostanziale tra il data mining educativo e i LA, in quanto entrambi perseguono l'obiettivo di incentivare un processo di miglioramento dell'istruzione (in particolare, attraverso l'identificazione dei problemi e il procedimento analitico che ne consegue) grazie all'uso dei big data. Per converso, i due campi derivano da contesti disciplinari differenti ed entrambi hanno priorità diverse. Le ricerche di Siemens e Baker (2012) aiu-

tano a comprendere le differenze che intercorrono tra i due ambiti, evidenziando diversità nell’approccio di ricerca, nell’analisi, nella personalizzazione e nei metodi utilizzati. L’EDM, come viene spesso precisato, è radicato nella modellazione e nei sistemi predittivi, si focalizza sull’estrazione (allo scavo secondo l’etimologia della formula) e sull’esplorazione dettagliata dei dati, sull’individuazione di schemi specifici attraverso tecniche quali classificazione, clustering, modeling, e si concentra sulla riduzione e sull’analisi dei singoli aspetti e delle loro interrelazioni. Questo approccio enfatizza la scoperta automatizzata e l’adattamento, pur riconoscendo il valore insostituibile del contributo umano nel raggiungimento degli obiettivi educativi. Al contrario, i LA si basano su una varietà di fonti come questionari, web semantico e learning management system (LMS) utilizzati per l’e-learning, puntando a un’integrazione dell’analisi dei dati con un’interpretazione umana per offrire una visione complessiva (olistica) del processo educativo. Metodi quali l’analisi statistica, la visualizzazione dei dati, l’analisi dei social network e l’analisi del testo sono frequentemente impiegati nei LA per fornire *insight* profondi che supportino sia gli studenti che i docenti, stimolando una personalizzazione dell’apprendimento basata sui dati. Curiosamente, la convergenza tra EDM e LA è sottolineata dalla comune sfida dell’*overload* informativo, evidenziata da ricercatori come Naisbitt, che già nel 1984, nel suo libro “Megatrends”, osservava che “stiamo affogando nell’informazione ma morendo di fame di conoscenza” (Naisbitt, 1984, p. 6), o da altri autori quali Papamitziou ed Economides (2014), che hanno notato come l’enorme volume di dati prodotti possa comportare un “sovraccarico di informazioni”² e renderne problematica l’analisi. Un argomento simile è stato avanzato da Raffaghelli, Cucchiara e Persico (2015)

2 Il Dizionario Treccani registra i termini “infobesità” e “infodemia”, rispettivamente: [https://www.treccani.it/vocabolario/infodemia_\(Neologismi\)](https://www.treccani.it/vocabolario/infodemia_(Neologismi))/ https://www.treccani.it/enciclopedia/infobesita_%28altro%29/

in relazione all'analisi dell'apprendimento per i MOOC (Massive Open Online Course – corso online aperto e di massa), sulla base di una revisione di sessanta articoli di ricerca (pubblicati tra il 2008 e il 2014), da cui risulta che le tecniche di analisi dell'apprendimento comprendono strumenti per la raccolta dei mega-dati allo stesso modo dei DM. Nel tempo, insomma, la frontiera tra i due campi sta diventando sempre più sfumata. Lo schema seguente (figura 4.1) fornisce una rappresentazione di sintesi dei principali obiettivi dell'EDM, delle sue applicazioni e dei suoi strumenti.

Obiettivi dell'EDM	Es. usi/applicazioni	Es. strumenti/tecniche
Previsione del comportamento degli studenti	Predizione del successo/in-successo accademico	Classificazione, Clustering
Perfezionamento dei modelli conoscitivi	Supporto pedagogico più efficace	Modellazione, Sistemi di raccomandazione
Ottimizzazione dei percorsi di apprendimento	Personalizzazione dell'apprendimento	<i>Machine Learning</i> , Sistemi adattivi
Generazione di dati empirici	Sostenere/definire teorie e framework pedagogici	Data Visualization (Data-Viz), Analisi statistica
Individuazione delle componenti essenziali dell'apprendimento	Identificazione di strutture relazionali e dinamiche sociali	Sistemi predittivi, content/thematic analysis
Promozione di ambienti formativi efficaci	Analisi di dati amministrativi, demografici ecc.	Analisi delle reti sociali, Analisi demografica
Sviluppo, ricerca e applicazione di metodologie informatiche	Identificazione di schemi e modelli nel comportamento degli studenti	Data reduction, Analisi di pattern
Analisi e risoluzione di questioni rilevanti nel campo della ricerca educativa	Miglioramento delle strategie didattiche basate su evidenza	Modellazione di simulazione, Analisi qualitativa
Analisi dettagliata dei dati educativi	Esplorazione di grandi dataset per rivelare nuovi <i>insight</i>	Data mining, Analisi di pattern complessi

Fig. 4.1: Obiettivi, usi e strumenti dell'EDM

4.1.2 *Le origini dell'EDM*

Si può parlare di una storia dell'EDM? Come abbiamo ricordato, il DM in generale è una recente disciplina informatica che consente di estrarre in maniera automatica conoscenza da un vasto insieme di dati digitali. Roiger e Geatz (2004, p. 24), nella loro "Introduzione al Data Mining", parlano di "una strategia di apprendimento di natura induttiva che costruisce modelli per identificare pattern nascosti di significato" e spiegano, inoltre, che "un modello creato da un algoritmo è una generalizzazione concettuale dei dati" e che "[tale] generalizzazione può assumere la forma di un albero, di una rete, di un'equazione di un insieme di regole".

Di solito, i manuali di DM si limitano a indicare il significato metaforico della formula (scavo dei dati), ad elencarne i presupposti scientifici (matematica, statistica, informatica, linguistica computazionale e scienze umane), accompagnando le indicazioni strumentali con un dato storico che situa la nascita del DM intorno al 1990. L'anno di nascita viene fatto coincidere con il potenziamento delle tecnologie digitali e dell'informazione – tra le quali l'esplosione dei big data – e con lo sviluppo esponenziale degli ambiti di applicazione (soprattutto marketing, finanza, bioinformazione, bioingegneria, sicurezza, sport e giochi). Il campo dell'educazione insieme a quello del text mining sono considerati come ultimi arrivati, ma non di minore importanza, tra i vari contesti possibili d'indagine. La natura induttiva ne sancisce il carattere sperimentale basato su congetture da validare per creare teorie da riapplicare in altri contesti. Banalmente, come per ogni ricerca, occorre formulare giuste domande a un'immensa massa di dati e testi. Ma non si tratta di domande improvvisate o illogiche. Roiger e Geatz (2004) tengono a precisare che, diversamente da altre modalità di interrogazione ("data query" e "sistemi esperti" – ovvero domande rivolte a specialisti), che loro si soffermano a considerare, il data mining "offre le risposte a do-

mande che non avremmo mai pensato di porci” (p. 25). Da qui l’idea di presentarlo come “la strategia più appropriata alla risoluzione di un particolare problema” (pp. 16-17) da preferire ad altre. Naturalmente, il percorso di apprendimento prevede una procedura in steps – presentata come un semplice modello di processo di data mining – che vedremo nel dettaglio più avanti. In pratica, abbiamo davanti la formulazione di un metodo di volta in volta adattabile e riutilizzabile.

Non abbiamo ancora risposto alla domanda iniziale se si possa parlare di una storia dell’EDM. In realtà, pensando più in generale all’idea di scavo dei dati, la risposta è sì, anche se non in senso tradizionale e comunque in modo non codificato (Rockwell, Berendt, 2016). L’EDM, insomma, seppur non possa vantare discendenza diretta, possiede antenati lontani dai nomi di spicco quali Francesco Petrarca (1304-1374), Niccolò Machiavelli (1469-1527) o Lorenzo Valla (1407-1557). L’accostamento – che solo *prima facie* può apparire azzardato e su cui torneremo al termine di questo paragrafo – tra gli antichi umanisti (alle prese con la cura dell’edizione critica dei testi) e i nuovi, impegnati a sperimentare il text mining per studiare le grandi collezioni di testi letterari e storici resi disponibili dai progressi informatici, è stato proposto da Rokwell e Berendt (2016p. 29), nell’evocare antiche “pratiche di interpretazione e analisi testuale”. Tali pratiche si fondano su operazioni editoriali esemplari: dall’“Index Thomisticus”, dello studioso gesuita Roberto Busa (1913-2011), pioniere della linguistica informatica, autore dello spoglio lessicale di tutta l’opera di S. Tommaso d’Aquino (1980, ma avviato alla fine degli anni Quaranta), al “Thesaurus Linguae Graecae” (1971-1985) e al “Trésor de la langue française” (1971-1994), comportanti la catalogazione (informatica) delle opere ma anche lo studio del linguaggio e la ricerca delle concordanze, volte a far emergere significati spesso non evidenti. Sono tutti esempi di applicazioni delle modalità macroanalitiche e cultoromiche di text mining, si- nona praticate, che si tratti di analisi basate su metodi statistico-

informatici (le concordanze) oppure di vere e proprie applicazioni macro-teoriche e concettuali (come il “Distant Reading” di Franco Moretti, 2013).

Gli esempi fin qui menzionati, che danno per acquisita l’intercambiabilità dei ruoli tra l’automa (Gigliozzi, Mordenti, Zampolli, 2000; Savoca, 2000) – il computer – e il testo letterario, sancendo l’incontro tra filologia e informatica (alla base delle attuali *digital humanities*), non sembrano avere stretta attinenza con l’EDM. Eppure, la vicinanza, e quindi l’impulso che l’EDM ha indirettamente ricevuto dalla cosiddetta “Galassia delle Digital Humanities” (Ciotti, 2023; Orlandi, Tommasi, 2023; Previtali, 2023), va ricercata nella propensione dell’EDM per l’indagine dei processi di apprendimento, volta a far emergere significati non evidenti proprio come gli antichi critici erano interessati al “senso riposto dei testi”. Possiamo ipotizzare che, come altri settori affini, non sia destinata a restare un campo sperimentale di confine proprio grazie a questa sua vocazione.

4.1.3 *Le fasi del processo dell’EDM*

La transizione verso l’utilizzo di tecnologie dell’informazione avanzate e l’archiviazione dei big data nelle discipline umanistiche, come evidenziato dai pionieri menzionati da Rockwell e Berendt (2016), hanno preparato il terreno per l’avvento di un nuovo ambito di ricerca che fonde l’antica arte della filologia con le possibilità quasi illimitate offerte dalle moderne tecnologie informatiche. L’introduzione di progetti innovativi – come quelli sopra menzionati realizzati da Busa (1980) e Moretti (2013) – ha dimostrato il potenziale delle metodologie computazionali nell’analisi dei testi, aprendo nuove strade per l’interpretazione e la comprensione del linguaggio umano su scala mai vista prima. Questa evoluzione metodologica ha trovato un fertile campo di applicazione nel settore dell’EDM, dove le tecniche di data mi-

ning e text mining diventano strumenti chiave per esplorare e interpretare i vasti set di dati generati nel contesto educativo. L'EDM, con il suo focus sull'applicazione di metodi e tecniche di data mining per estrarre conoscenze utili dai dati educativi, rappresenta una convergenza tra le metodologie tradizionali e le esigenze attuali dell'ambiente educativo digitalizzato. Il processo dell'EDM incarna questa sintesi di approccio matematico e analisi computazionale applicata ai dati educativi per migliorare l'apprendimento e l'insegnamento e si articola in sei fasi distinte, delineate da Larose e Larose (2015) e Romero e colleghi (2010). Esse abbracciano l'intero arco operativo: in pratica, dalla definizione del problema fino alla valutazione dei modelli generati. Ne consegue che l'EDM si pone come ultimo anello di una lunga catena di sviluppi metodologici, un ponte tra l'esperienza ereditata dalle modalità di ricerca nelle discipline umanistiche e la nuova frontiera dell'analisi dei dati educativi, ancorati alla comprensione e all'interpretazione del linguaggio umano, ora arricchita grazie alle risorse offerte dall'informatica e dalle tecnologie dell'informazione. La figura 4.2 illustra le fasi del processo dell'EDM.



Fig. 4.2: Le fasi del processo dell'Educational Data Mining

Si è anticipato che l'EDM si concentra sull'applicazione di metodi e tecniche in ambito educativo per estrarre informazioni utili dai dati generati da studenti, insegnanti e sistemi formativi digitali. Abbiamo anche detto che il processo dell'EDM consta

tipicamente di sei fasi (Larose, Larose, 2015; Romero, Baker, Pechenizkiy, Ventura, 2010). La raccolta dei dati è il prerequisito del processo dell'EDM (Shruthi, Chaitra, 2016). Come affermano Romero e Ventura (2013, p. 12), questa fase comporta “la raccolta di dati grezzi da diverse fonti”. Nell’ambito educativo, i dati possono essere raccolti attraverso sistemi di gestione dell’apprendimento (LMS), ambienti virtuali di apprendimento, risposte a test standardizzati, sensori in aule smart, interazioni su piattaforme online e una varietà di altre fonti (Daniel, 2015; Williamson, 2017).

I dati raccolti includono comunemente informazioni demografiche e di background degli studenti, risultati di apprendimento, tracciamenti delle attività e interazioni degli studenti, network di relazioni sociali, contenuti di discussioni online, risposte a sondaggi e questionari e altro ancora (Williamson, 2017). L’accumulo dei dati può avvenire attraverso procedure di import automatico o manuale, API (Application Programming Interface) delle piattaforme educative, formulari somministrati agli studenti (Baker, Siemens, 2014).

Una volta riuniti, i dati devono essere selezionati in modo accurato – prima fase – per costruire un dataset rilevante ed omogeneo rispetto agli obiettivi specifici di apprendimento e ricerca (Zaïane, 2002). Criteri per la selezione dei dati includono la completezza, la correttezza, la rappresentatività della popolazione target e l’attrattività in termini di potenziali pattern significativi (Romero, Ventura, 2013). Il risultato è un insieme curato di dati pronto per il pre-processamento, seconda fase e momento critico che prepara i dati grezzi raccolti per l’effettiva analisi di data mining (Ibrahim, Shiba, 2019). Come spiegano Baker e Yacef (2009), attività tipiche di pre-processamento comprendono:

- integrazione: combinare dati da diverse fonti in un unico dataset coerente;

- pulizia: rilevare, correggere o rimuovere dati corrotti, duplicati, inconsistenti;
- normalizzazione: standardizzare attributi in scale comuni per consentire confronti;
- discretizzazione: convertire attributi continui in categorie discrete;
- riduzione della dimensionalità: selezionare un sottoinsieme di caratteristiche rilevanti;
- de-identificazione: rimuovere o oscurare dati sensibili che identifichino individui.

Un pre-processamento efficace è essenziale per produrre dataset di alta qualità che consentano al data mining di generare modelli accurati e utili (Zaiane, 2002); inoltre, l'attenzione e l'abilità con cui questa fase viene eseguita possono fare una grande differenza nel successo complessivo di un progetto di EDM (Baker, Alevan, Corbett, 2008).

La terza fase riguarda la trasformazione dei dati: in pratica, si prende il dataset pre-processato e lo si converte nel formato ottimale per l'applicazione degli algoritmi di data mining scelti (Romero, Ventura, 2013). Tipiche trasformazioni includono (Baker, Siemens, 2014):

- creazione di una matrice features-instances: i dati vengono organizzati in un dataset con istanze (osservazioni) sulle righe e caratteristiche sulle colonne;
- conversione in formati specifici: i dati vengono convertiti in formati come CSV, ARFF, o altri formati richiesti da *tool* di data mining;
- costruzione di dataset di training e test: il dataset viene suddiviso in sottoinsiemi disgiunti per l'addestramento e il test dei modelli;
- creazione di nuove features: possono essere derivate e aggiunte nuove features dalle esistenti per migliorare i modelli;

- codifica di feature categoriche: gli attributi categorici sono convertiti in formato numerico.

Le trasformazioni applicate devono essere appropriate al tipo di tecniche di data mining che verranno utilizzate nella fase successiva (Baker, Siemens, 2014). Questa fase richiede quindi familiarità con i requisiti algoritmici e capacità di manipolazione dei dati.

La fase di data mining, ossia la quarta, consiste nell'applicazione di algoritmi per estrarre modelli e scoprire relazioni nei dati educativi (Romero, Ventura, 2010). Questo è il cuore del processo dell'EDM (Jassim, Abdulwahid, 2021). Tale pratica è finalizzata alla scoperta di tendenze, modelli e relazioni statisticamente significative, nascoste in grandi insiemi di dati. Il data miner appare come una sorta di detective che esplora e analizza grandi quantità di dati per scoprire indizi al fine di prendere decisioni informate ed efficaci. Le principali tecniche di data mining utilizzate in ambito educativo includono (Baker, Siemens, 2014):

- classificazione: costruire modelli predittivi per attribuire gli studenti a categorie come, ad esempio, studenti a rischio di abbandono;
- regressione: elaborare modelli che predicono valori numerici come voti o performance;
- clustering: identificare gruppi di studenti con caratteristiche comuni;
- analisi di reti sociali: analizzare pattern di connessioni e interazioni tra studenti;
- analisi di sequenza: individuare sequenze frequenti di azioni degli studenti;
- text mining: estrarre conoscenza da contenuti testuali come discussioni online;
- scoperta di regole di associazione: identificare relazioni frequenti tra eventi educativi;
- pattern mining: trovare pattern interessanti nei dati.

I risultati del data mining sono modelli e pattern che rivelano nuova conoscenza e informazione sugli ambienti e i processi educativi analizzati (Romero, Ventura, 2010). La selezione delle giuste tecniche deve essere commisurata agli obiettivi e al tipo di dati.

La quinta fase è quella di valutazione dei modelli; essa è importante per determinare la qualità e l'utilità dei modelli prodotti nella fase di data mining (Baker, Siemens, 2014). I modelli vengono valutati rispetto a metriche di performance appropriate al tipo di modello e di problema analizzato. Per i modelli predittivi le metriche comunemente usate includono (Romero, Ventura, 2010):

- Accuratezza (*Accuracy*): percentuale di predizioni corrette complessive (sia positive che negative) rispetto al totale delle predizioni;
- Precisione (*Precision*): la percentuale di predizioni positive corrette rispetto al totale delle predizioni positive (es. un antifurto è molto preciso se ogni volta che suona c'è un ladro in casa; non suona se passa il gatto o un topo);
- Sensibilità (*Recall*): la percentuale di casi positivi reali che sono stati correttamente identificati dal modello (es. un antifurto è sensibile se suona ogni volta che c'è un ladro in casa; potrebbe però suonare anche se passa un gatto o un topo);
- Matrice di confusione (*Confusion Matrix*): una tabella che mostra le predizioni del modello rispetto ai valori reali; permette di visualizzare facilmente i veri positivi, i falsi positivi, i veri negativi e i falsi negativi;
- Curva ROC (*ROC Curve*): un grafico che illustra la capacità di un modello di classificazione di distinguere tra classi al variare della soglia di decisione. Sull'asse delle y si trova il tasso di veri positivi, mentre sull'asse delle x si trova il tasso di falsi positivi. Più l'area sotto la curva ROC è grande, più il modello è capace di classificare correttamente le classi.

Per modelli descrittivi come clustering e pattern mining, la valutazione si basa su (Baker, Siemens, 2014):

- validità del clustering: quanto i cluster catturano la struttura intrinseca dei dati;
- interesse dei pattern: quanto i pattern rivelano nuova conoscenza;
- comprensibilità: quanto facilmente i modelli sono interpretabili.

I modelli con prestazioni inadeguate vengono scartati o migliorati attraverso tecniche come ri-campionamento dei dati e configurazione dei parametri (*parameter tuning*). L'obiettivo consiste nel predisporre modelli affidabili da usare nella pratica educativa.

La sesta fase riguarda la rappresentazione della conoscenza e consiste nell'interpretare e presentare i modelli e i pattern scoperti durante il data mining in modo che forniscano intuizioni significative e agibili (Baker, Siemens, 2014). Questa fase traduce i risultati tecnici del data mining in una forma comprensibile per insegnanti, amministratori e altri stakeholder non tecnici. Queste le attività chiave:

- interpretazione: esaminare i modelli per derivarne inferenze e comprensione degli aspetti educativi modellati; ad esempio, quali fattori influenzano maggiormente il rendimento?
- visualizzazione: rappresentare graficamente modelli e pattern tramite grafici, mappe di calore (*heatmap*), alberi di decisione; le rappresentazioni visive aiutano la comprensione;
- generazione di rapporti: creare rapporti e dashboard che riassumono in linguaggio naturale i risultati e le implicazioni per gli stakeholder;
- storytelling: raccontare i risultati come narrazioni coinvolgenti che evidenziano gli *insight* chiave;

- supporto decisionale: fornire raccomandazioni concrete su come applicare le conoscenze per migliorare l'apprendimento.

Questa fase finale completa l'intero ciclo, trasformando l'output tecnico dell'EDM in conoscenza pedagogica fruibile quale guida dell'azione educativa.

4.2 Metodologie, tecniche e applicazioni dell'EDM

L'EDM impiega un'ampia gamma di metodi e tecniche per estrarre conoscenze dai dati educativi. Questi metodi possono essere suddivisi in diverse categorie, ognuna con il proprio insieme di approcci e obiettivi specifici. Tra i metodi più rilevanti nell'EDM, Baker e Yacef (2009) menzionano gli approcci sotto riportati, che illustreremo nei paragrafi successivi:

- Metodi predittivi:
 - Classificazione
 - Regressione
- Metodi di *structure recovery*:
 - Clustering
 - Analisi delle reti sociali
- Metodi di *relationship mining*:
 - Analisi di sequenza
 - Regole di associazione
- Text mining:
 - Sentiment Analysis
 - Estrazione di concetti
 - Analisi tematica dei contenuti.

4.2.1 *I metodi predittivi*

In generale, nell'ambito dell'EDM, i metodi predittivi giocano un ruolo fondamentale nella comprensione e nel miglioramento dei processi educativi (Ahmad, Ismail, Aziz, 2015; Bakhshinategh, Zaiane, Elatia, Ipperciel, 2018; Lei, Yang, Cai, 2016). Tali approcci si basano sullo sviluppo di modelli capaci di inferire un aspetto specifico dei dati, noto come variabile predetta, attraverso l'analisi di altre variabili, definite di predizione (Alyahyan, Dü tegör, 2020; Dekker, Pechenizkiy, Vleeshouwers, 2009). Ad esempio, in un contesto educativo, un modello predittivo potrebbe essere utilizzato per prevedere il rendimento accademico futuro degli studenti, basandosi su variabili di predizione quali la partecipazione a lezioni, i voti ottenuti in test precedenti e il tempo trascorso su piattaforme di apprendimento online. La predizione può avvalersi sia di classificatori che di "regressori", con aggiunta di diverse sottocategorie, che attingono alla vasta esperienza dei campi dell'intelligenza artificiale e del *machine learning* (Baker, Buckingham Shum, Duval, Stamper, Wiley, 2012; D'Mello, Craig, Graesser, McDaniel, Witherspoon, 2008).

La classificazione

La classificazione è una delle tecniche di data mining supervisionato³ più utilizzate in ambito educativo (Romero, Ventura, 2013). Lo scopo della classificazione consiste nel costruire un modello predittivo, chiamato classificatore, in grado di assegnare istanze non etichettate (es. nuovi studenti) a categorie o classi predefinite, sulla base di un insieme di dati di training contenente

- 3 Il data mining supervisionato (in inglese *data mining supervised*), si riferisce agli algoritmi di apprendimento utilizzati nella classificazione e nella previsione. L'algoritmo apprende dai dati di addestramento che sono stati precedentemente etichettati rendendo l'attività supervisionata.

istanze già correttamente etichettate (Baker, Siemens, 2014). Nell'EDM, la classificazione viene tipicamente applicata per sviluppare modelli predittivi di risultati di apprendimento, come il superamento o meno di un corso, il voto finale o ancora l'abbandono degli studi (Öztürk, 2016; Romero et al., 2010). Ad esempio, un modello di classificazione potrebbe categorizzare gli studenti in due gruppi distinti: "a rischio di abbandono" e "non a rischio di abbandono", sulla base di un set diversificato di attributi quali risultati scolastici pregressi, dati demografici, comportamento in classe, motivazione, tratti cognitivi e molti altri (Pérez, Castellanos, Correal, 2018).

I principali algoritmi di classificazione supervisionata applicati nell'EDM comprendono alberi decisionali, reti neurali, classificatori bayesiani (*naive bayes*), le *support vector machines* e le *random forest* (Baker, Siemens, 2014). Questi algoritmi costruiscono modelli complessi a partire da un insieme di training inclusivo di istanze (es. studenti) già etichettate rispetto alla classe di output (es. abbandono sì/no). Il modello viene poi utilizzato per assegnare etichette di classe a nuove istanze non ancora classificate. La previsione fornita da un'analisi può indicare direttamente l'appartenenza ad una specifica classe (a rischio di abbandono) oppure una probabilità stimata di appartenenza alle possibili classi di output (es. 70% probabilità di abbandono) (Romero & Ventura, 2013). Nel complesso, i modelli predittivi di classificazione forniscono informazioni estremamente utili in ambito educativo, consentendo l'identificazione precoce di situazioni problematiche, permettendo quindi interventi mirati e preventivi da parte di insegnanti ed istituzioni (Abou-Elfetouh, Nabil, Seyam, 2022). Tuttavia, per ottenere efficaci prestazioni predittive, questi modelli richiedono la disponibilità di ampi dataset di training adeguatamente bilanciati rispetto alle classi target e correttamente etichettati (Baker, Siemens, 2014).

Scheda di approfondimento 1

Scenario applicativo sull'uso di tecniche di classificazione per predire il rischio di abbandono

Un'università si propone di ridurre il tasso di abbandono degli studi tra i suoi studenti, facendo ricorso alle tecniche dell'EDM. Lo scopo del modello da costruire consisterà, pertanto, nel tentare di identificare gli studenti a rischio di abbandono; successivamente, in base alla previsione ottenuta, l'Università implementerà programmi di supporto personalizzati, come tutoraggio, consulenza accademica e psicologica e corsi di miglioramento delle abilità di studio.

Supponiamo di avere un dataset con le seguenti caratteristiche per ogni studente:

- Media dei voti (alta, media, bassa)
- Partecipazione alle lezioni (alta, bassa)
- Attività extracurricolari (sì, no)
- Supporto familiare (alto, basso)
- Rischio di abbandono (sì, no)

Passaggi per la costruzione del modello

- Selezione degli algoritmi: si scelgono uno o più algoritmi di classificazione basandosi sulla natura dei dati e sull'obiettivo specifico; nel nostro caso, esamineremo gli alberi decisionali.
- Training del modello: si addestra il modello utilizzando un set di dati di training, che include studenti già etichettati come "a rischio di abbandono" o "non a rischio".
- Validazione e ottimizzazione: si valuta l'accuratezza del modello su un set di dati di test separato e si affina per migliorarne le prestazioni.
- Applicazione del modello: il modello è ora pronto per categorizzare gli studenti, assegnando probabilità di appartenenza alle classi "a rischio" e "non a rischio".

Applicazione

Simuliamo a questo punto un albero decisionale ipotetico sulle nostre variabili al fine di fare un esempio pratico. L'albero decisionale risulterà così composto:

Livello 1 (Radice): Media dei voti

La prima divisione si basa sulla "Media dei voti": è la caratteristica che, ad esempio, potrebbe meglio distinguere gli studenti a rischio da quelli non a rischio.

Se alta, si proceda al Livello 2A. - Se media o bassa, si proceda al Livello 2B.

Livello 2A: attività extracurricolari (per studenti con media alta)

Se sì, la probabilità di abbandono è bassa. - Se no, la probabilità di abbandono è media.

Livello 2B: Partecipazione alle lezioni (per studenti con voti bassi o medi)

Se alta, si proceda al Livello 3A. - Se bassa, si proceda al Livello 3B.

Livello 3A: supporto familiare (per studenti con partecipazione alta)

Se alto, la probabilità di abbandono è media. - Se basso, la probabilità di abbandono è alta.

Livello 3B: supporto familiare (per studenti con partecipazione bassa)

Indipendentemente dal supporto familiare, la probabilità di abbandono è alta.

Per prevedere il rischio di abbandono di uno studente, si inizia dalla radice dell'albero e si segue il percorso che corrisponde alle caratteristiche dello studente fino a raggiungere una foglia, che indica la previsione del modello. Nel nostro caso, ad esempio, uno studente con media dei voti alta, che non partecipa ad attività extracurricolari, sarà

classificato come a rischio medio di abbandono secondo questo albero decisionale.

Questo schema offre un modello interpretabile che può aiutare gli educatori a identificare gli studenti a rischio e a comprendere quali fattori contribuiscono maggiormente al rischio di abbandono. Gli interventi possono quindi essere personalizzati per affrontare specificamente questi fattori, migliorando così le possibilità di permanenza degli studenti. La figura 4.3 rappresenta un albero decisionale costruito sullo scenario sopra descritto.

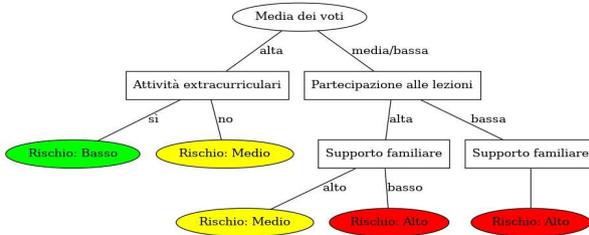


Fig. 4.3: Esempio di albero decisionale costruito sullo scenario applicativo presentato

La regressione

La regressione è una tecnica di data mining supervisionato utilizzata per predire valori numerici continui, come voti, punteggi o performance, invece di semplici etichette di classe (Romero, Ventura, 2013). Nell'EDM i modelli di regressione vengono tipicamente impiegati per prevedere il rendimento scolastico, ad esempio per stimare il voto che uno studente otterrà in un esame o il suo punteggio in una valutazione standardizzata (Baker, Siemens, 2014). A differenza della classificazione, che predice etichette di classe (variabili categoriche), la regressione si concentra sulla previsione di valori continui (variabili numeriche) (Zhu, Li, Hou, Hu, 2021). Per costruire un modello di regressione, gli al-

goritmi analizzano le relazioni tra gli attributi di input (caratteristiche degli studenti) e la variabile numerica da predire nel set di training. Il modello cattura queste relazioni sotto forma di equazione matematica che mappa gli input sulla base dei valori target (Romero et al., 2010). I principali algoritmi di regressione usati nell'EDM sono: regressione lineare, regressione logistica, alberi di regressione, reti neurali e support vector regression (Baker, Siemens, 2014). Questi modelli forniscono in output una stima numerica continua anziché una semplice classificazione. I modelli di regressione consentono previsioni granulari del rendimento e possono rivelare i fattori che influenzano più fortemente le performance numeriche (Romero, Ventura, 2013). Ad esempio, si può determinare l'impatto relativo di precedenti voti scolastici, motivazione, stato socioeconomico e altre caratteristiche sul voto dell'esame finale. Tuttavia, la qualità delle previsioni dipende dalla rappresentatività dei dati di training e dalle assunzioni dell'algoritmo scelto rispetto ai dati (Baker, Siemens, 2014).

Scheda di approfondimento 2

Scenario applicativo sull'uso di tecniche di regressione per prevedere il voto finale di uno studente

Supponiamo di voler sviluppare un modello di regressione per prevedere il voto finale di uno studente in matematica. Gli input (variabili indipendenti) potrebbero includere:

- voti intermedi
- ore di studio settimanali
- partecipazione alle lezioni
- motivazione auto-dichiarata
- supporto familiare

Il modello di regressione analizzerà i dati storici per identificare come queste variabili influenzino il voto finale, risultando in un'equazione che stima il voto in base agli input forniti.

Passaggi per la costruzione del modello

- Sviluppo del modello: utilizzare un dataset di training per addestrare il modello di regressione, testandolo poi con un set di dati separato per valutarne l'accuratezza.
- Valutazione: esaminare gli errori del modello e la sua capacità di generalizzazione su nuovi dati.
- Simuliamo quindi un albero di regressione ipotetico sulle nostre variabili al fine di fare un esempio pratico. L'albero di regressione risulterà così composto:

Nodo Radice

Condizione: voti intermedi ≤ 23.30

Questo è il punto di partenza dell'albero, dove vengono presi in considerazione tutti i dati. La condizione divide il set di dati in base ai voti intermedi degli studenti. Se

uno studente ha voti intermedi minori o uguali a 23.30, segue il ramo a sinistra dell'albero; altrimenti segue il ramo a destra.

Primo Livello di Divisione

Ramo Sinistro: ore di studio settimanali ≤ 6.29

Gli studenti con voti intermedi ≤ 23.30 sono ulteriormente divisi in base al numero di ore di studio settimanali. Se studiano 6.29 ore o meno a settimana, procedono al prossimo nodo sul ramo sinistro.

Ramo Destro: Ore di studio settimanali > 6.29

Gli studenti con voti intermedi ≤ 23.30 che studiano più di 6.29 ore a settimana procedono sul ramo destro.

Secondo Livello di Divisione per gli Studenti con Voti Intermedi ≤ 23.30

Ramo Sinistro (di Sinistra): ore di studio settimanali ≤ 6.00

Questo nodo rappresenta gli studenti che studiano molto poco. Se uno studente studia 6 ore o meno a settimana, raggiunge una "foglia" dell'albero con una previsione specifica del voto finale.

Ramo Destro (di Sinistra): ore di studio settimanali > 6.00

Gli studenti che studiano più di 6 ore ma meno o uguale a 6.29 ore a settimana raggiungono un'altra foglia dell'albero con una diversa previsione del voto finale.

Secondo Livello di Divisione per gli Studenti con Voti Intermedi > 23.30

Ramo Sinistro (di Destra): ore di studio settimanali ≤ 7.07

Gli studenti con voti intermedi maggiori di 23.30 sono divisi basandosi sulle ore di studio settimanali. Quelli che studiano fino a 7.07 ore seguono questo percorso.

Ramo Destro (di Destra): ore di studio settimanali > 7.07
 Gli studenti che studiano più di 7.07 ore a settimana vengono ulteriormente analizzati nei nodi successivi.

Terzo Livello di Divisione e “Foglie”

A questo punto, gli studenti sono stati divisi in gruppi molto specifici basati su voti intermedi, ore di studio settimanali, e altri criteri nei rami successivi (come il supporto familiare per alcuni rami). Ogni foglia dell'albero fornisce una previsione del voto finale basata sulle caratteristiche che definiscono il percorso verso quella foglia. Ogni foglia, quindi, rappresenta il valore previsto (il voto finale in questo caso) per gli studenti che rientrano in quella categoria specifica basata su tutte le condizioni incontrate lungo il percorso dall'inizio dell'albero fino a quella foglia. La figura 4.4 fornisce un esempio visivo di albero di regressione sulle variabili descritte nello scenario.

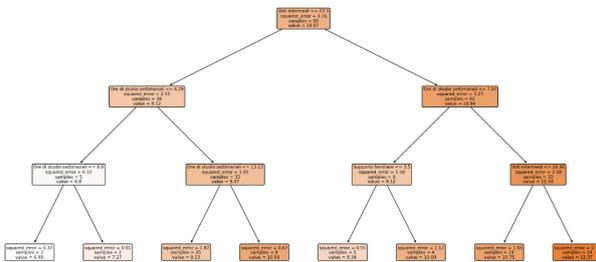


Fig. 4.4: Esempio di albero di regressione costruito sullo scenario applicativo presentato

4.2.2 I metodi di structure discovery

La ricerca di strutture nei dati rappresenta una componente cruciale dell'EDM (Corbett, Anderson, 1994); essa cerca di individuare una struttura nei dati senza un'idea a priori di ciò che si dovrebbe trovare.

Questo processo si distingue per il suo approccio esplorativo: a differenza dei metodi predittivi, che mirano a modellare una variabile specifica identificata a priori, la scoperta di strutture si avventura senza presupposti predefiniti sulla natura o sulla forma delle relazioni che emergono dall'analisi dei dati. In altre parole, mentre nella previsione l'obiettivo è chiaro e diretto verso un target definito, nella ricerca di strutture l'obiettivo è scoprire, senza preconcetti, quali pattern o configurazioni si manifestano spontaneamente all'interno del dataset. Questa differenziazione metodologica riflette un diverso orientamento nell'indagine scientifica all'interno del campo dell'EDM: si passa dalla modellazione di variabili e relazioni specifiche all'esplorazione libera dei dati, per identificare strutture e correlazioni non immediatamente evidenti.

Gli algoritmi di scoperta di strutture servono quindi a identificare organizzazioni intrinseche nei dati, senza presupporre alcuna verità di fondo o idea preconcetta su ciò che dovrebbe emergere (Ahuja, Jha, Maurya, Srivastava, 2019). Questa caratteristica li rende particolarmente adatti all'esplorazione libera dei dati educativi, al fine di rivelare pattern e relazioni precedentemente non osservati (Dawson, 2008; Suthers, Rosen, 2011).

Il clustering

Il clustering è una tecnica di data mining non supervisionato che identifica e raggruppa istanze simili in cluster distinti (Larose, Larose, 2015; Romero, Ventura, 2013). A differenza di classificazione e regressione, non sono definite a priori delle etichette o valori target. Nell'EDM, il clustering viene tipicamente usato per individuare gruppi omogenei di studenti sulla base di caratteristiche e comportamenti condivisi, come risultati scolastici, stili di apprendimento, livelli di motivazione, pattern di studio online e così via (Baker, Siemens, 2014). La scelta dell'algoritmo di clustering appropriato dipende dalla natura dei dati e dagli obiettivi specifici della ricerca. I principali algoritmi di clustering applicati

sono K-means, clustering gerarchico e DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Ester, Kriegel, Sander, Xu, 1996; Ikotun, Abualigah, Abuhaija, Ezugwu, Heming, 2023; Johnson, 1967; Romero et al., 2010). In generale, la caratteristica di questi algoritmi è raggruppare gli studenti sulla base della similarità, massimizzando quella intra-cluster e minimizzando la similarità inter-cluster. I cluster generati forniscono informazioni preziose sulle caratteristiche di porzioni di studenti e consentono di adattare l'insegnamento in base alle esigenze dei gruppi (Baker, Siemens, 2014). Ad esempio, gli studenti possono essere clusterizzati in base ai risultati ottenuti per identificare gruppi con performance alte, medie o basse; oppure, sulla base delle interazioni su forum e contenuti didattici condivisi dai docenti, per distinguere studenti motivati, mediamente partecipativi e non impegnati. I cluster individuati possono essere poi utilizzati per creare classi omogenee o attivare azioni mirate per gruppi con necessità simili (Romero, Ventura, 2007). La clusterizzazione si presta anche al campo dello sviluppo professionale degli insegnanti, ad esempio nello studio delle forme di aggregazioni informali delle community online di insegnanti (Gabbi, 2023). Tuttavia, la qualità dei cluster dipende fortemente dalle caratteristiche considerate e dai parametri algoritmici: è quindi importante applicare tecniche di validazione per garantire cluster significativi e interpretabili (Romero, Ventura, 2013).

Scheda di approfondimento 3

Scenario applicativo sull'uso di tecniche di clustering per identificare gruppi omogenei di studenti

Supponiamo di voler identificare gruppi omogenei di studenti basati sulle loro performance accademiche tramite l'ausilio dell'algoritmo K-means. L'obiettivo è suddividere gli studenti in tre cluster distinti: alte performance, medie performance e basse performance, per poter poi personalizzare l'insegnamento e le risorse a disposizione di ciascun gruppo.

Gli input (variabili indipendenti) potrebbero includere:

- voti finali di 100 studenti in una particolare materia, espressi su una scala da 0 a 10
- numero di ore dedicate allo studio settimanalmente
- livello di partecipazione alle lezioni.

Passaggi per la costruzione del modello

- Selezione del numero di cluster (k): scegliamo $k=3$ per riflettere i nostri gruppi target: alte, medie, e basse performance.
- Inizializzazione dei "centroidi": inizialmente, selezioniamo tre voti a caso dall'insieme dei dati come centroidi iniziali dei cluster.
- Assegnazione degli studenti ai cluster: ogni studente viene assegnato al cluster il cui centroide è il più vicino in termini di differenza di voti; questo si basa sulla distanza euclidea tra il voto dello studente e il valore del centroide.
- Aggiornamento dei Centroidi: dopo aver assegnato tutti gli studenti, il centroide di ogni cluster viene aggiornato, calcolando la media dei voti degli studenti appartenenti a quel cluster. Ripetizione fino alla convergenza: i passaggi 3 e 4 vengono ripetuti fino a quando i centroidi non cambiano significativamente

tra le iterazioni, indicando che l’algoritmo ha raggiunto la convergenza.

Dopo aver applicato l’algoritmo K-means con i passaggi sopra indicati, supponiamo che il clustering risultante sia così composto: gli studenti con voti ≥ 8 formano il cluster “Alte Performance”, quelli con voti tra 5 e 7.9 formano il cluster “Medie Performance”, e quelli con voti < 5 formano il cluster “Basse Performance”. La figura 4.5 rappresenta visivamente degli esempi di cluster costruiti sullo scenario sopra descritto.

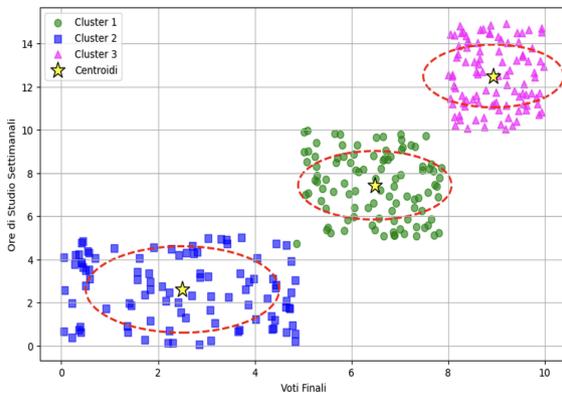


Fig. 4.5: Visualizzazione dei cluster costruiti sullo scenario applicativo presentato

Analisi delle reti sociali

L’ascesa dei social media, tra cui Meta e X (precedentemente noto come Twitter), ha notevolmente stimolato l’interesse verso l’analisi delle reti sociali. L’origine dell’analisi delle reti sociali (Social Network Analysis, SNA) può essere tracciata nei contributi di pionieri della sociologia come Durkheim (1858-1917), il quale postulava che l’obiettivo della sociologia dovesse essere l’esame olistico della società e delle interconnessioni fra gli attori sociali.

In generale, una rete sociale può essere definita da un insieme di nodi (o punti) che possono avere relazioni gli uni con gli altri (Wasserman, Faust, 1994). Questa architettura di nodi può essere costituita da individui, gruppi, organizzazioni, governi, paesi o altre unità sociali distinte di analisi. Tali legami possono variare in numero e intensità, offrendo una mappa complessa delle strutture sociali che la SNA aspira a decodificare mediante tecniche di mappatura e rappresentazione grafica. All'interno della SNA convivono diverse prospettive di ricerca: alcuni studiosi pongono l'accento sulla predizione di dinamiche future in termini di interazioni e relazioni all'interno delle reti; altri mirano alla generalizzazione attraverso l'uso di statistiche inferenziali per esplorare fino a che punto i risultati dell'analisi delle reti sociali siano rappresentativi di popolazioni più ampie; altri ancora guardano in direzione diversa come ad esempio Buch-Hansen (2014, p. 312), il quale ritiene che "è perfettamente possibile applicare tecniche di analisi della rete sociale senza utilizzarle deduttivamente, senza combinarle con la teoria della scelta razionale, senza fare previsioni e generalizzazioni, e senza fare affidamento su modelli formali crudamente riduzionisti". Ciò si allinea con le crescenti richieste di integrare dati qualitativi nell'analisi delle reti, una istanza che acquista particolare rilievo nel contesto dell'EDM. La SNA offre, infatti, una lente attraverso cui è possibile esaminare le intricate reti di relazioni che caratterizzano gli ambienti educativi, identificando come nodi non solo gli individui – studenti, insegnanti – ma anche le istituzioni, i materiali didattici e persino le interazioni virtuali. Questo approccio permette agli esperti dell'EDM di mappare e analizzare le complesse dinamiche sociali e di apprendimento all'interno delle comunità educative, evidenziando i pattern di collaborazione, le reti di supporto tra pari e la distribuzione delle risorse educative. L'impiego della SNA nell'EDM si distacca insomma dalla tradizionale enfasi posta sulle sole metriche quantitative, arricchendo l'analisi con una comprensione profonda delle strutture relazionali e delle dinamiche sociali. Que-

sta metodologia, quindi, si colloca all'interno dell'EDM come un prezioso strumento per esplorare non solo le performance individuali ma anche il modo in cui l'ambiente e la società influiscono sugli apprendimenti, consentendo una visione più olistica e interconnessa dell'educazione. La figura 4.6 rappresenta una SNA semplificata realizzata con il software open source GEPHI⁴.

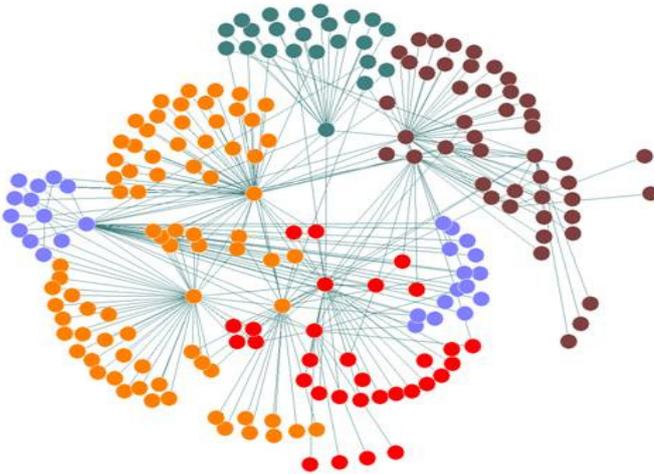


Fig. 4.6: Esempio di SNA realizzata con GEPHI

4.2.3 I metodi di relationship mining

Nel campo dell'EDM, il mining delle relazioni rappresenta un'area focalizzata sulla scoperta di collegamenti tra variabili all'interno di dataset che presentano un ampio numero di fattori (Scheuer, McLaren, 2012). L'obiettivo consiste nell'identificazione delle variabili più fortemente associate a un dato di parti-

4 Fonte: <https://gephi.org/>

colare interesse e nella scoperta di quali relazioni tra due variabili siano le più significative (Algarni, 2016). Fanno parte di questo processo due tipologie di mining delle relazioni comunemente impiegate nell'EDM: le analisi di sequenze e le regole di associazione (Spirtes, Glymour, Scheines, 2001).

Analisi di sequenza

L'analisi di sequenza è una tecnica che identifica pattern e relazioni in serie ordinate di eventi o attività degli studenti (Baker, Siemens, 2014). Nell'EDM, essa viene applicata principalmente per evidenziare successioni frequenti di azioni effettuate dagli studenti in ambienti digitali come piattaforme e-learning e sistemi intelligenti di tutoraggio (Romero, Ventura, 2013; Krpan, Stanokov, 2012). Ad esempio, si possono identificare pattern di navigazione ricorrenti in un corso online.

Gli algoritmi di analisi di sequenza – come la ricerca di serie frequenti – rilevano similarità e differenze tra le successioni di eventi per estrarre pattern significativi (Baker, Siemens, 2014). Possono emergere pattern come “rileggere le slide-fare gli esercizi-rileggere di nuovo le slide”.

I pattern sequenziali forniscono preziose informazioni sul comportamento e le strategie di apprendimento degli studenti in ambienti digitali (Romero et al., 2010). Consentono di riprogettare i percorsi formativi per migliorare l'esperienza o correggere frequenti inefficienze nel rendimento degli studenti.

Tuttavia, data la natura sequenziale e potenzialmente lunga dei dati, sono necessari accurati pre-processamento e filtraggio per identificare sottoinsiemi (*subset*) rilevanti su cui applicare il mining (Baker, Siemens, 2014). Inoltre, la significatività statistica dei pattern va validata per evitare l'overfitting⁵.

5 In statistica e in informatica, si parla di *overfitting* (adattamento eccessivo) quando un modello statistico complesso si adatta eccessivamente ai dati osservati (il campione) perché ha un numero eccessivo di para-

Regole di associazione

La scoperta di regole di associazione è una tecnica che identifica relazioni interessanti tra eventi educativi che si verificano frequentemente insieme (Aleem, Gore, 2020; Baker, Siemens, 2014).

Nell'EDM questa tecnica viene utilizzata per scoprire associazioni tra azioni, risorse e performance degli studenti (Romero, Ventura, 2013). Le regole seguono la formula tipica: “Se A, allora B”; ad esempio, “Se lo studente accede alle video-lezioni, allora ottiene punteggi più alti”.

Gli algoritmi di associazione calcolano le frequenze di co-occorrenza tra eventi ed estraggono le regole che soddisfano soglie minime di supporto e confidenza statistica (Romero et al., 2010). Supporto e confidenza forniscono significatività e forza dell'associazione di due o più elementi.

Ad esempio, una regola di associazione potrebbe rilevare che gli studenti che ripassano gli appunti prima di un esame tendono ad ottenere voti più alti. Questa osservazione suggerisce un'associazione tra due variabili: il ripasso e i voti degli esami. Il “supporto” indica quanto comunemente questi due eventi si verificano insieme nel dataset, mentre la “confidenza” esprime la probabilità che gli studenti ottengano voti più alti, se hanno ripassato. Tuttavia, è necessaria cautela nell'interpretazione, poiché una correlazione riscontrata da una regola di associazione non stabilisce una relazione di causa ed effetto tra gli eventi; riflette piuttosto una tendenza all'interno dei dati esaminati. Per questo motivo, l'interpretazione delle regole di associazione richiede un'analisi attenta da parte di esperti nel campo educativo, i quali devono valutare la rilevanza statistica e il significato pedagogico delle associazioni scoperte, districandosi tra semplici coincidenze e possibili indizi di fenomeni più profondi (Romero, Ventura, 2013).

metri rispetto al numero di osservazioni, perdendo quindi la capacità di generalizzare. Il modello si adatta quindi a caratteristiche che sono specifiche solo del campione, ma che non hanno riscontro nella distribuzione tipica del resto dei casi.

Scheda di approfondimento 4

Scenario applicativo sull' uso di regole di associazione per esplorare le abitudini di studio degli studenti

In una università, un gruppo di ricerca è interessato a esplorare in che modo le abitudini di studio e l'uso delle risorse digitali influenzino le performance degli studenti. L'obiettivo consiste nell'utilizzare la tecnica delle regole di associazione per identificare modelli e correlazioni tra comportamenti di studio specifici e successo accademico, al fine di migliorare le strategie didattiche e le risorse a disposizione degli studenti.

Supponiamo di avere un dataset che include informazioni sugli studenti, come frequenza di accesso alle video-lezioni (Alto, Medio, Basso), partecipazione ai forum di discussione (Sì, No), uso di materiali di sintesi prima degli esami (Sì, No), superamento dell'esame (Sì, No).

L'algoritmo "apriori" (Agrawal, Imieli ski, Swami, 1993) inizia una ricerca degli item singoli nel dataset e calcola il loro supporto⁶ mantenendo solo gli item che soddisfano la soglia minima di supporto decisa. Successivamente, estende la ricerca agli item frequenti, trovati con altri item, creando combinazioni di 2 item, 3 item, ecc., e calcola il loro supporto, ripetendo il processo fino a quando

6 *Supporto*: Misura la frequenza con cui appare un insieme di item nel dataset. Ad esempio, il supporto per {Uso di materiali di sintesi=Sì, Superamento esame=Sì} indica in che percentuale gli studenti che usano le sintesi superano l'esame.

Confidenza: Misura la probabilità che un item appaia in una regola dato che un altro item è presente. Ad esempio, la confidenza della regola {Partecipazione forum=Sì} => {Superamento esame=Sì} indica la probabilità che gli studenti superino l'esame dato che partecipano ai forum.

non ci sono più nuove combinazioni che soddisfino la soglia di supporto impostata.

Potremmo, ad esempio, immaginare che vengano trovate regole con il relativo valore di supporto:

Regola 1: se uno studente accede regolarmente alle videolezioni (A), allora tende ad ottenere voti più alti (B).

Regola 2: se uno studente partecipa attivamente ai forum di discussione (C), allora migliora la sua performance negli esami (D).

Regola 3: se uno studente consulta frequentemente le sintesi dei corsi prima degli esami (E), allora ha maggiori probabilità di superare gli esami al primo tentativo (F). A questo punto, verificati i valori di confidenza e di supporto, si può passare alle interpretazioni dei risultati. Nel nostro caso le regole identificate suggeriscono correlazioni tra l'uso di specifiche risorse educative e le performance accademiche.

A scopo esemplificativo, una delle regole di associazione scoperte potrebbe essere {Uso di materiali di sintesi=Sì, Partecipazione forum=Sì} => {Superamento esame=Sì} con un supporto del 30% e una confidenza del 70%. Questo significa che il 30% degli studenti nel dataset usa le sintesi e partecipa ai forum e, di questi, il 70% supera gli esami. Tuttavia, è cruciale comparare l'interpretazione di queste regole con un'attenta considerazione del contesto educativo e della possibile varietà di fattori influenti, mantenendo un atteggiamento critico circa la distinzione tra correlazione e causalità.

4.2.4 *Il text mining*

Come abbiamo precedentemente illustrato, il text mining indica l'insieme di metodi per estrarre conoscenza da dati testuali non strutturati, tipicamente generati da studenti in ambienti educativi digitali (Bienkowski, Feng, Means, 2012; Melendez-Armenta, Morales-Rosales, Huerta-Pacheco, Rebolledo-Mendez, 2020). Nell'EDM, il text mining viene applicato per analizzare il contenuto di e-mail, discussioni su forum, chat, documenti scritti e altri testi prodotti da studenti in piattaforme online (Romero, Ventura, 2013). Fanno parte del text mining: la sentiment analysis, l'estrazione di concetti e l'analisi tematica dei contenuti.

Sentiment analysis

Questa tecnica è utilizzata per identificare e classificare le opinioni (*sentiment*) espresse in un testo. L'elaborazione computazionale del linguaggio applicata agli elaborati degli studenti consente di rilevare frustrazione, confusione, noia, entusiasmo ed altri stati emotivi attraverso tecniche di sentiment analysis (Lang, Gasevic, Siemens, Wise, 2017). Nell'ambito educativo, tale analisi può essere impiegata per esaminare il feedback degli studenti su corsi, materiali didattici o esperienze di apprendimento. Attraverso l'uso dell'analisi dei *sentiment*, gli educatori possono rilevare tendenze positive o negative nelle reazioni degli studenti, consentendo loro di apportare modifiche mirate per migliorare il processo di insegnamento e apprendimento. Tali tecniche di analisi spaziano dalla semplice conta di parole positive o negative fino all'uso di complessi modelli di classificazione per interpretare il testo (Medhat, Hassan, Korashy, 2014). La figura 3 illustra un esempio di applicazione su un dataset pubblicamente reperibile su Kaggle⁷ composto da tweet relativi all'assistenza clienti.

7 Fonte: <https://www.kaggle.com/datasets/thoughtvector/customer-support-on-twitter>

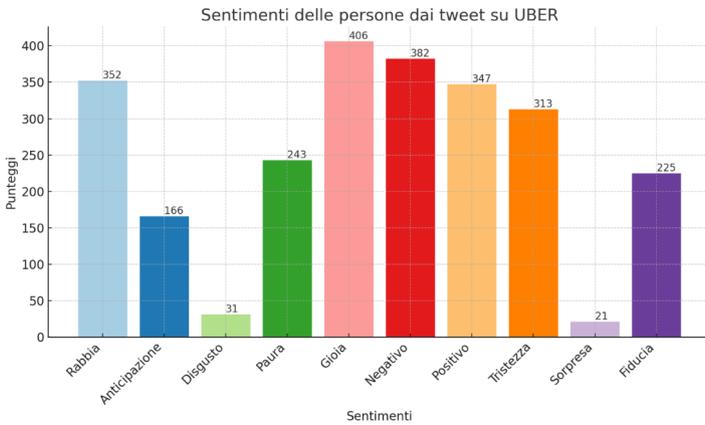


Fig. 4.7: Esempio di sentiment analysis condotta sul dataset “Customer Support on Twitter”

Estrazione di concetti

L'estrazione di concetti si riferisce all'identificazione di entità (come persone, luoghi, concetti) e le relazioni tra queste entità all'interno di un testo (Nadeau, Sekine, 2007). Questa tecnica è particolarmente utile nell'analisi di grandi volumi di testi educativi per scoprire temi comuni o per mappare la struttura del contenuto didattico. L'estrazione di concetti può aiutare gli educatori a identificare quali concetti chiave vengono discussi negli ambienti di apprendimento online e come questi concetti sono interconnessi. L'uso di algoritmi per il riconoscimento di entità (Named Entity Recognition, NER) e l'analisi di co-occorrenza sono esempi di tecniche utilizzate in questo ambito.

Analisi tematica e dei contenuti

La *content analysis*, o analisi del contenuto, è una metodologia di ricerca usata per identificare e quantificare la presenza di certe parole, concetti o categorie all'interno di insiemi di dati testuali. Questa tecnica si basa su un approccio quantitativo per analizzare

la frequenza e le relazioni di tali elementi. Per esempio, l'analisi del contenuto può essere utilizzata per esaminare le descrizioni nei saggi degli studenti e valutare la comprensione di concetti chiave. Oppure, attraverso l'uso di software specializzati, i ricercatori possono processare grandi volumi di testo per identificare modelli e tendenze che emergono negli elaborati degli studenti (Krippendorff, 2018).

La *thematic analysis*, o analisi tematica, si concentra più sulla qualità che sulla quantità del testo analizzato. Questa tecnica si propone di identificare, analizzare e evidenziare i temi (*patterns of meaning*) che emergono all'interno dei dati testuali. A differenza dell'analisi del contenuto, l'analisi tematica non si limita a contare le frequenze, ma piuttosto esplora il modo in cui temi specifici si intersecano e si riferiscono al contesto più ampio (Braun, Clarke, 2006). Nell'ambito educativo, può essere utilizzata per interpretare discussioni di gruppo o risposte agli esami, fornendo approfondimenti sulle percezioni e sulle esperienze degli studenti; inoltre, può essere utile per comprendere come i temi individuati influenzano il processo di apprendimento.

Riassumendo, entrambi gli approcci servono all'individuazione di eventuali schemi nei dati testuali; tuttavia, mentre lo studio dei contenuti si focalizza sul computo della frequenza degli schemi, l'analisi tematica si concentra sul contenuto degli stessi e sulle loro relazioni.

4.3 Questioni aperte nell'uso dell'EDM

Fin qui ci siamo soffermati sui principali usi e le più note metodologie dell'EDM, evidenziando la varietà delle applicazioni e la molteplicità delle tecniche. Il ricorso a questi strumenti può arricchire la nostra comprensione dei fenomeni educativi, consentendoci di migliorare l'esperienza degli studenti attraverso interventi mirati e puntuali. Tuttavia, se ciò è vero in generale, è

altrettanto vero che l'implementazione di queste tecniche genera sfide inedite per il mondo dell'educazione (Boyd, Crawford, 2012; Zoric, 2019). Anzi, alla crescita della complessità degli strumenti utilizzati corrisponde l'aumento della complessità dei problemi risultati dalla loro applicazione. Pertanto, onde evitare l'adozione di approcci superficiali, ingenui o addirittura controproducenti, occorre considerare in modo approfondito alcuni temi, potenzialmente critici, che spaziano dagli aspetti tecnici a quelli organizzativi, economici, politici ed etico-sociali. In particolare, alcuni autori (Baker, Inventado, 2014; Romero, Ventura, 2020; Fischer et al., 2020) hanno provato a mappare le questioni da prendere in considerazione per ottimizzare i benefici dell'EDM e ridurre le possibili criticità. Seguendo e sintetizzando le loro analisi, possiamo distinguere quattro principali questioni: economico-sociali, metodologiche, tecnico-organizzative ed etico-giuridiche.

4.3.1 *Questioni economico-sociali: la proprietà dei dati come fonte di vantaggio competitivo*

Uno dei risvolti critici legati all'implementazione di tecniche dell'EDM nei sistemi istruttivi è riscontrabile nella crescente diffusione, nel mercato dei learning analytics e delle piattaforme di apprendimento adattivo, di servizi erogati da aziende private. Questo fenomeno solleva rilevanti interrogativi in merito alla proprietà e al controllo dei dati prodotti in ambito educativo. Da qui sorge la fondamentale domanda: "Chi detiene i big data?" (Ruppert, 2015). L'industria *educational*, specie in Nord America, ha maturato un progressivo interesse verso la creazione di enormi database contenenti dati educativi: possedendo le infrastrutture tecnologiche e gli strumenti per gestire tali moli di dati, le aziende attive nel settore sono in grado di generare valore economico derivante dalla capacità di analizzare questi dati, comprenderli e fare previsioni basate su di essi (Landri, 2018; Nielsen, 2005).

L'operato di tali aziende nel campo dell'analisi dei dati educativi si fonda su un rigoroso apparato disciplinare legato ai processi di analisi e conoscenza dei dati scientifici. Come evidenziato da Kitchin (2014), questo approccio mira a utilizzare informazioni numeriche per sviluppare nuove pratiche di gestione educativa improntate all'oggettività e a generare approcci innovativi nel settore. Un esempio significativo di questa prospettiva è fornito da Behrens (2013), fondatore del Center for Digital Data, Analytics and Adaptive Learning di Pearson, il quale ha previsto che l'analisi dei dati digitali provenienti dalle interazioni degli studenti con le risorse online e le attività digitali quotidiane sfiderà le attuali pratiche didattiche. Questo perché, secondo Behrens, l'ampia disponibilità di dati digitali potrebbe creare un divario significativo tra l'aumento dei risultati basati sull'analisi dei dati educativi e la base teorica necessaria per interpretarli e integrarli. Nella storia delle tecnologie educative, l'adozione di toni anche enfatici nella rappresentazione del potenziale dei dispositivi tecnici per l'innovazione didattica non è nuova (Ranieri, 2011) e i cicli di entusiasmo, speranza e delusione caratterizzanti questo fenomeno ci ricordano opportuna cautela. Tuttavia, questo divario evidenziato da Behrens potrebbe dare adito alla possibilità di sviluppare nuovi approcci scientifici all'uso dei dati educativi, in grado di generare nuove conoscenze che possano colmarlo con soluzioni in forma di prodotti software per le istituzioni scolastiche e, addirittura, proposte politiche per i governi (Hogan, Lingard, Sellar, 2015).

Tutto ciò suggerisce la presenza di un processo evolutivo in atto indirizzato verso la creazione di un'infrastruttura di dati digitali, dedicata alla produzione di conoscenza e alla generazione di metodi educativi, e guidato da potenti aziende private, in grado di utilizzare il proprio vantaggio tecnologico e organizzativo per adottare pratiche computazionali e algoritmi capaci di superare le capacità umane di analisi (Decuyper, Grimaldi, Landri, 2021; Landri, 2018). Come abbiamo visto nel capitolo 1 in relazione al

tema della *platformisation*, il concetto di “infrastruttura dati” è particolarmente importante, poiché le infrastrutture sono le attrezzature fisiche, materiali e organizzative che sottendono e orchestrano la vita sociale, politica ed economica (Perrotta, Gulson, Williamson, Witzenberger, 2021). Infatti, nuove infrastrutture per la produzione e la comunicazione della conoscenza associate a Internet, ai big data e ai social media sono ormai ampiamente diffuse, portando a “un mondo in cui la conoscenza è perpetuamente in movimento” (Edwards et al., 2013, pp. 6-7). Questi autori osservano che le nuove infrastrutture della conoscenza, oltre ad essere sistemi tecnici, rafforzano o redistribuiscono l'autorità, l'influenza e il potere, privilegiando gli interessi di alcuni e spostando le prospettive di altri, esercitando contemporaneamente effetti sulla forma e sulla possibilità della conoscenza in generale.

Con il crescente interesse dei governi verso la possibilità di avere accesso ai big data a fini della ricerca educativa, organizzazioni private come Pearson, Lytics Lab, Knewton e simili si stanno posizionando come interlocutori sempre più influenti nella produzione di conoscenza. Nel contesto attuale, infatti, avere accesso o, meglio, possedere tali database, rappresenta un prerequisito per la generazione di nuove soluzioni didattiche, mettendo di fatto queste aziende private in una posizione di enorme vantaggio competitivo rispetto ad altri enti. Tuttavia, l'attuale tendenza – intensificata e accelerata anche dalla pandemia da COVID-19 per l'ingresso massiccio delle piattaforme digitali nelle aule scolastiche e universitarie - che vede la ricerca educativa concentrarsi rapidamente nelle mani di aziende e organizzazioni private dotate di ingenti risorse economiche, tecnologiche e organizzative, richiede un'attenta riflessione sulla validità delle nuove soluzioni didattiche generate: infatti, come osservato da Taglietti e colleghi (2021) e Watters (2016), mentre non vi è nessuna garanzia che tali aziende, seppur in possesso di ingenti moli di dati educativi, siano effettivamente in grado di sviluppare le soluzioni migliori per assistere gli studenti nello studio, quasi certamente questo può significare

lo sviluppo di risorse e strumenti educativi oggetto di negoziazione e vendita.

La proprietà dei big data educativi, la formulazione di nuovi approcci nel campo dell'educazione e la loro successiva applicazione all'interno di software brevettati potrebbero delineare uno scenario futuro in cui aziende private, titolari di strumenti di analisi evoluti e di sistemi educativi – ma con obiettivi di mercato – potrebbero, con l'avallo delle autorità governative, assumere un ruolo centrale nel panorama dell'apprendimento e dell'insegnamento (Kerssens, van Dijck, 2022). In questo contesto, lo studio e la comprensione dei processi di apprendimento insieme all'avanzamento della ricerca scientifica in ambito educativo rischiano di diventare una sorta di proprietà intellettuale per aziende e centri di ricerca privati ben finanziati.

4.3.2 *Questioni metodologiche: le implicazioni pedagogiche dei sistemi predittivi*

Dal punto di vista analitico, identificare le metodologie di data mining più adatte per uno specifico caso d'uso educativo è complesso, ma cruciale per la buona riuscita di un progetto di EDM (Bienkowski et al., 2012). Non esiste un approccio unico universalmente valido, ma si tratta di saper selezionare, combinare e mettere a punto tecniche di clustering, classificazione, text mining ecc. adeguate a estrarre *insight* rilevanti, senza incorrere nei rischi già menzionati dell'overfitting (si veda sopra paragrafo 4.2.3).

Inoltre, i risultati degli algoritmi vanno validati per evitare conclusioni sbagliate, e interpretati correttamente alla luce di fattori pedagogici e didattici quasi mai inglobabili nei modelli prefissati. Il rischio di derivare correlazioni spurie è alto: ciò significa che servono particolari competenze multidisciplinari per un'analisi attenta e critica (Alshehri, Alhakami, Alsubait, Baz, 2020).

Nel campo educativo sussiste, altresì, un ulteriore ostacolo le-

gato alla necessità di basare gli algoritmi di *machine learning* su dati già categorizzati (Beer, 2016). Questo aspetto è di particolare rilevanza, in quanto la classificazione iniziale influisce direttamente sul modello di apprendimento degli algoritmi (Mackenzie, 2015). Pertanto, le piattaforme di analisi dei dati producono conoscenza appresa algoritmicamente che può essere usata per modellare attività future, malgrado ciò dipenda dal lavoro classificatorio degli sviluppatori degli algoritmi. Come sottolineano Cope e Kalantzis (2016, p. 11), “i modelli statistici nei dati di apprendimento automatico sono in larga misura generati da pattern già incorporati nei modelli di formazione supervisionati. Nel caso del *machine learning* non supervisionato, i modelli statistici hanno senso solo quando vengono fornite etichette esplicative”. In tal senso, le modalità attraverso le quali i sistemi di apprendimento automatico vengono progettati per imparare dai dati diventano molto importanti per il modo stesso in cui vengono generati approfondimenti dai dati o per il modo in cui gli studenti potrebbero essere coinvolti (Williamson, 2017). Questo dimostra come le promesse matematiche di oggettività delle piattaforme analitiche siano facilmente disattese; in realtà, esse sono profondamente legate ai contesti sociali che influenzano e sono influenzati dai dati stessi (Sellar, 2015). In breve, “tradurre un apprendimento in numeri lo trasforma perché cambia il modo in cui quell’oggetto può essere compreso e come potrebbe essere maneggiato” (*ivi*, p. 172).

Estendendo tali affermazioni all’analisi dei dati educativi, tecniche algoritmiche come l’analisi dei cluster lavorano secondo modalità operative che non scoprono semplicemente modelli nei dati, ma li costruiscono attivamente. Tornando all’esempio relativo alla scheda di approfondimento 3 (*Scenario applicativo sull’uso di tecniche di clustering per identificare gruppi omogenei di studenti*), l’algoritmo di clusterizzazione inizia stabilendo un certo numero di cluster, ognuno con il proprio centro; quindi, prende le osservazioni date e le assegna ai cluster usando la distanza dai centri

come criterio. I centri vengono quindi ricalcolati e il processo ricomincia fino a quando la distanza è minimizzata, in modo che si possano ottenere risultati stabili (Perrotta, Williamson, 2016). Il punto chiave consiste nel fatto che è necessario, per iniziare l'analisi, inserire i dati in un cluster o in un gruppo di altri dati simili, impostando in anticipo alcuni parametri e criteri. L'analisi dei cluster, quindi, non necessariamente scopre modelli nei dati, ma può attivamente costruire una struttura, mentre calcola la distanza tra i punti dati secondo criteri prestabiliti. L'impatto sullo studente deriva dal fatto che tali algoritmi proporranno percorsi di apprendimento personalizzati, suggerendo contenuti ed influenzando il modo in cui essi struttureranno la conoscenza, su cluster forzatamente generati (Williamson, 2017). Detto in altri termini, il modo stesso in cui determinate tecniche dell'EDM funzionano influisce sui processi di insegnamento e apprendimento, senza che alla base vi sia una chiara e consapevole scelta pedagogica sull'impianto educativo e didattico dell'offerta formativa. L'automazione del processo formativo ha pertanto implicazioni pedagogiche rilevanti, che richiedono invece di spostare l'attenzione dalle infrastrutture tecniche a quelle pedagogico-didattiche, attraverso l'adozione di approcci *human centered* e la collaborazione stretta con gli utenti (sia studenti che insegnanti), anche in termini di co-design dei modelli (Gabbi, 2023).

In conclusione, mentre le tecniche dell'EDM offrono opportunità per migliorare i sistemi istruttivi, esse impongono anche una riflessione critica sulla loro implementazione. La necessità di competenze multidisciplinari per l'analisi critica, l'importanza della validazione degli algoritmi e il ruolo cruciale delle scelte umane nel processo di classificazione sottolineano la complessità intrinseca nell'applicazione dell'apprendimento automatico all'educazione.

4.3.3 *Questioni tecnico-organizzative: dall'acquisizione dei dati al cambiamento organizzativo*

Come nel caso dei big data, uno dei punti più critici dell'EDM è rappresentato dalla varietà e qualità dei dati raccolti (Daniel, 2019). I dati educativi, lo si è visto già, provengono da una molteplicità di sistemi eterogenei, spesso mancanti di interoperabilità (ossia della capacità dei sistemi di scambiarsi informazioni tra di essi), aspetto che rende particolarmente complessa la loro integrazione. Inoltre, i dati contengono frequentemente errori, valutazioni sbagliate o incoerenze che ne pregiudicano l'affidabilità ai fini analitici. Operazioni di pulizia dei dati (*data cleaning*) e integrazione richiedono, pertanto, sforzi rilevanti.

Un'ulteriore questione tecnica consiste nell'adattabilità delle analisi su vasta scala. Queste analisi hanno infatti bisogno di tecnologie di big data avanzate per una efficace attuazione (Siemens et al., 2011). Addestrare un modello predittivo computazionale su milioni di record può risultare proibitivo senza l'appoggio di un'adeguata infrastruttura.

Per quanto riguarda gli aspetti organizzativi, l'aumento di iniziative dell'EDM comporta anche un cambiamento gestionale (*change management*) all'interno delle istituzioni formative (Ferguson et al., 2016). Tale cambiamento implica – tra le altre cose – la ridefinizione di ruoli e responsabilità, il coinvolgimento degli insegnanti e l'integrazione di pratiche evidence based nei processi già esistenti. Gli insegnanti sono chiamati ad acquisire competenze di *data literacy* anche per riuscire a superare simili sfide (Mandinach, Gummer, 2016), come è stato precedentemente discusso nel capitolo 3 (par. 3.3.1). Diviene infatti necessario che, sia in ambito universitario che scolastico, i docenti sviluppino competenze specifiche nel campo della gestione dei dati, da intendersi come una vera e propria alfabetizzazione ai dati, per prendere decisioni informate e adattare le pratiche didattiche in base alle esigenze degli studenti. Come sottolineato nel capitolo 2, tale

competenza comprende la capacità di raccogliere, analizzare e interpretare una vasta gamma di dati educativi, inclusi quelli relativi alle valutazioni, al contesto scolastico e ai comportamenti degli studenti, al fine di determinare le azioni educative più appropriate. In tal senso, formazione, comunicazione e incentivi sono fattori chiave per un'adozione ampia e consapevole dell'EDM a livello organizzativo.

4.3.4 *Questioni etico-giuridiche: privacy e uso responsabile dei dati*

Oltre alle questioni relative alla proprietà dei dati e al vantaggio competitivo che questo comporta, i progetti di EDM pongono interrogativi di natura etica, sollevando problemi complessi che rimandano alle considerazioni tradizionalmente affrontate nel contesto dell'Intelligenza Artificiale (IA) (Floridi, 2022; Pancioli, Rivoltella, 2023; Ranieri, Cuomo, Biagini, 2024). Innanzitutto, tali questioni riguardano aspetti cruciali come il consenso informato e la trasparenza nell'utilizzo dei dati degli studenti in qualità di soggetti vulnerabili (Ispas, 2019; Prinsloo, Slade, 2017). È pertanto fondamentale stabilire politiche e regole chiare, ossia un adeguato quadro di policy e governance, al fine di garantire un approccio etico nell'ambito dell'EDM.

Inoltre, questo processo diventa più complesso quando i dati degli studenti vengono raccolti e analizzati tramite sistemi esterni, al di fuori del controllo diretto dei docenti o delle istituzioni scolastiche. In tale contesto, come analizzato nel paragrafo relativo alle questioni economico-sociali, sorge la preoccupazione che la raccolta e l'analisi dei dati da parte di fornitori commerciali con sistemi proprietari possano comportare rischi etici (anche in termini di privacy) e danni potenziali (Miettinen, 2014). Questi rischi relativi alla privacy non sono necessariamente ignorati dagli utenti; tuttavia, il comportamento effettivo spesso non si allinea con le preoccupazioni espresse. Si tratta del così detto *privacy pa-*

radox, ovvero un fenomeno osservato nel comportamento online degli utenti, dove c'è una discordanza significativa tra la preoccupazione espressa per la privacy e le azioni effettive che le persone intraprendono per proteggerla (Gerber, Gerber, Volkamer, 2018). Sebbene molti utenti dichiarino di porre attenzione alla loro privacy online e al modo in cui le loro informazioni personali vengono utilizzate, spesso non agiscono in modo coerente con queste preoccupazioni. Tale comportamento contraddittorio può manifestarsi in varie forme, come continuare a usare servizi di social media nonostante i timori sulla privacy, scaricare app senza leggere le politiche sulla privacy, o accettare acriticamente i termini di servizio e i permessi di app che possono compromettere la propria privacy. A tal proposito, il giurista Solove (2006) ha sviluppato una tassonomia relativa ai potenziali danni alla privacy causati dai sistemi informativi, suddividendoli in quattro categorie. La prima si riferisce alla raccolta dei dati e alle modalità con cui vengono raccolte informazioni dall'individuo oggetto dell'indagine (osservando o registrando le attività o raccogliendo informazioni tramite questionari o sondaggi informativi). Queste pratiche si estendono anche alle attività di analisi e di sorveglianza degli studenti, spesso eseguite in modo inconsapevole o automaticamente da parte di algoritmi specifici.

La seconda categoria si riferisce alla fase di elaborazione delle informazioni, che riguarda l'aggregazione e la combinazione di dati, nonché l'identificazione e il collegamento di informazioni relative a diversi individui, effettuata da parte di coloro che raccolgono i dati – ovvero i titolari dei dati – che li elaborano, combinano e utilizzano. Nel contesto delle piattaforme di analisi dell'apprendimento, tale fase si manifesta nell'aggregazione dei dati degli studenti per creare dettagliati profili derivanti dalle loro attività e interazioni, sollevando di conseguenza preoccupazioni legate alla sicurezza e all'uso improprio dei dati personali raccolti.

Segue la terza categoria, relativa alla diffusione delle informazioni che, in termini di rischi per la privacy, contempla non solo

la violazione della riservatezza, ma anche i rischi legati alla divulgazione di informazioni personali in grado di influenzare il giudizio su un individuo da parte di altri (come l'appropriazione di identità o la diffusione di informazioni false o fuorvianti). La crescita di una vera e propria industria commerciale associata alla raccolta, elaborazione e analisi dai dati degli studenti suggerisce la possibilità di sfruttamento economico di tali informazioni, nonché la vulnerabilità a violazioni dei dati e attacchi informatici.

Infine, nella quarta categoria relativa ai rischi connessi alla violazione della privacy, sono inclusi i possibili danni derivanti dall'interferenza nella vita quotidiana degli individui e nel loro processo decisionale. Molte piattaforme di analisi dell'apprendimento sembrano limitare il coinvolgimento degli studenti nelle decisioni riguardanti l'uso consapevole dei propri dati, delegando la presa di decisione a sistemi informatici automatizzati. Ciò può comportare interferenze nel processo decisionale degli studenti, che rischia sempre più di essere condizionato da meccanismi intrusivi in grado di influenzare le loro scelte sulla base di calcoli derivati da enormi database.

L'interpretazione e l'utilizzo dei risultati dell'EDM devono, quindi, essere integrati all'interno di un quadro etico, che ne garantisca l'uso responsabile, proteggendo la privacy degli studenti e la loro autonomia decisionale (Alamuddin, Brown, Kurzweil, 2016).

Alla luce di tali considerazioni, il concetto di uso responsabile dei dati va interpretato come un approccio consapevole delle questioni relative alla privacy e alla protezione dei dati, e centrato su valori come trasparenza, consapevolezza e autonomia degli studenti (Berendt, Büchler, Rockwell, 2015; Klose, Desai, Gehringer, Song, 2020). Tuttavia, permangono dubbi significativi sull'adeguatezza delle attuali normative sulla privacy e sulla gestione dei dati educativi, specialmente in relazione all'espansione delle pratiche di analisi dei big data e del data mining nel settore

educativo. Tra le principali questioni si annoverano: la capacità delle norme vigenti di proteggere efficacemente le informazioni personali degli studenti dalle potenziali minacce di sicurezza e di uso improprio, l'estensione del consenso informato nell'utilizzo dei dati da parte delle istituzioni educative, la trasparenza nei confronti degli studenti e dei genitori sulle metodologie di raccolta e sull'uso specifico dei dati raccolti, nonché l'equilibrio tra l'utilizzo benefico di tali dati per l'istruzione personalizzata e il rispetto della privacy individuale. Questi dubbi sollevano interrogativi su come il diritto alla riservatezza possa coesistere con le strategie di raccolta e analisi dati sempre più pervasive e sofisticate, in un'epoca dove il valore educativo dei dati è in rapido aumento (Crippa, Girgenti, 2024; Floridi, 2016; Zeide, 2016).

Concludendo, la pratica dell'EDM quale indagine dei processi di apprendimento non deve far pensare a un quadro di successi facilmente raggiungibili. Ogni passo in avanti dovrà continuare faticosamente a mettere in conto aspetti e problemi pedagogico-didattici, etici e sociali in modo del tutto simile a quelli posti dall'uso dell'IA nella formazione (Crippa, Girgenti, 2024; Ranieri et al., 2024). Questi aspetti critici riguardano da vicino sia i learning analytics che l'EDM e devono servire a rendere ogni analista consapevole della necessità di formulare risposte in grado di commisurare l'efficacia della tecnologia con la riflessione critica propria dell'uomo (Crippa, Girgenti, 2024).

In ultima analisi, la *data literacy* e l'EDM sono due facce della stessa medaglia, in grado di offrire possibili risposte alle questioni etiche, gestionali e tecniche sollecitate dal progressivo avvento dei big data nell'ambito educativo. Le due strategie – la prima che lavora in una prospettiva di empowerment dei singoli e delle comunità e la seconda che amplifica le potenzialità esplorative ed esplicative delle dinamiche di apprendimento e insegnamento – sono strettamente interconnesse e si rafforzano reciprocamente. Gli insegnanti che possiedono una solida comprensione dei concetti di *data literacy* sono in grado di trarre beneficio dall'analisi

dei dati per migliorare l'insegnamento e l'apprendimento, ma anche di mettere in luce le eventuali problematiche e distorsioni. Allo stesso tempo, l'EDM può fornire l'occasione per migliorare la *data literacy* degli educatori, offrendo loro accesso a informazioni e risorse avanzate per analizzare i dati in modo critico ed efficace e costituendo un campo d'azione concreto nel quale sviluppare le proprie competenze. In questo contesto, diventa fondamentale non solo continuare a esplorare e a sfruttare le potenzialità offerte dall'EDM e dalla *data literacy*, ma anche rimanere vigili sui rischi associati, lavorando costantemente per garantire che l'evoluzione tecnologica sia sempre guidata da principi etici e inclusivi. Guardando al futuro, la sfida per docenti, sviluppatori e policy maker sarà quella di costruire un ecosistema educativo, in cui la tecnologia serva da leva per un apprendimento significativo, equo e umanizzante, contribuendo a formare cittadini globali consapevoli, critici e proattivi nell'era digitale.

Riferimenti bibliografici

- Abou-Elfetouh, A., Nabil, A., Seyam, M. (2022). Predicting students' academic performance using machine learning techniques: A literature review. *International Journal of Business Intelligence and Data Mining*, 20(4), 456-479.
- Agrawal, R., Imieli ski, T., Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (pp. 207-216).
- Ahmad, F., Ismail, N. H., Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9(129), 6415-6426.
- Ahuja, R., Jha, A., Maurya, R., Srivastava, R. (2019). Analysis of educational data mining. In *Harmony Search and Nature Inspired Optimization Algorithms* (pp. 897-907). Singapore: Springer.
- Alamuddin, R., Brown, J., Kurzweil, M. (2016). *Student Data in the Digital Era: An overview of current practices*. New York: Ithaca S+R.

- Aldowah, H., Al-Samarraie, H., Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13-49.
- Aleem, A., Gore, M. M. (2020). Educational data mining methods: A survey. In *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 182-188). Piscataway, NJ: IEEE.
- Algarni, A. (2016). Data mining in education. *International Journal of Advanced Computer Science and Applications*, 7(6), 456-461.
- Alshehri, E., Alhakami, H., Baz, A., Alsubait, T. (2020). A comparison of EDM tools and techniques. *International Journal of Advanced Computer Science and Applications*, 11(12), 824-831.
- Alyahyan, E., Dü tegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 1-21.
- Ampadu, Y. B. (2023). *Handling Big Data in Education: A Review of Educational Data Mining Techniques for Specific Educational Problems*. IntechOpen. <https://doi.org/10.5772/acrt.17>
- Anjum, N., Badugu, S. (2020). A study of different techniques in educational data mining. In *Advances in Decision Sciences, Image Processing, Security and Computer Vision: International Conference on Emerging Trends in Engineering (ICETE)*, Vol. 2 (pp. 562-571). Cham: Springer.
- Baker, R. S. J. D. (2010). Data mining for education. *International Encyclopedia of Education*, 7(3), 112-118.
- Baker, R. S. J. d., Gowda, S. M., Corbett, A. T. (2011). Automatically detecting a student's preparation for future learning: Help use is key. In *Proceedings of the Fourth International Conference on Educational Data Mining* (pp. 179-188). Eindhoven, The Netherlands.
- Baker, R. S. D., Duval, E., Stamper, J., Wiley, D., Buckingham Shum, S. (2012). Educational data mining meets learning analytics. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 20-20).
- Baker, R. S., Corbett, A. T., Alevan, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. In *International Conference on Intelligent Tutoring Systems* (pp. 406-415). Berlin, Heidelberg: Springer.

- Baker, R. S., Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 253-274). Cambridge: Cambridge University Press.
- Baker, R. S., Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17.
- Bakhshinategh, B., Zaiane, O. R., Elatia, S., Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, 23, 537-553.
- Beer, D. (2016). How should we do the history of Big Data? *Big Data and Society*, 3(1). <http://dx.doi.org/10.1177/2053951716646135>
- Behrens, J. (2013). *Harnessing the currents of the digital ocean*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April.
- Berendt, B., Büchler, M., Rockwell, G. (2015). Is it research or is it spying? Thinking through ethics in Big Data AI and other knowledge sciences. *Künstliche Intelligenz*, 29(2), 223-232.
- Bienkowski, M., Feng, M., Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Office of Educational Technology, US Department of Education. Washington, DC: ERIC.
- Boyd, D., Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662-679.
- Braun, V., Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57-71). American Psychological Association
- Brunet, É. (1989). L'Exploitation des Grands Corpus: Le Bestiaire de la Littérature Française. *Literary and Linguistic Computing*, 4(2), 121-134.
- Buch-Hansen, H. (2014). Social Network Analysis and Critical Realism. *Journal for the Theory of Social Behavior*, 44(3), 306-325.
- Busa, R. (1980). The annals of humanities computing: The Index Thomisticus. *Computers and the Humanities*, 14(2), 83-90.

- Ciotti, F. (2023). *Digital Humanities: Metodi, strumenti, saperi*. Roma: Carocci. (Collana Studi Superiori). ISBN 9788829018437.
- Cope, B., Kalantzis, M. (2016). Big data comes to school: implications for learning, assessment and research. *AERA Open*, 2(2), 1-19.
- Corbett, A. T., Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253-278.
- Crawford, K. (2013). The hidden biases in big data. *Harvard Business Review*. Retrieved June 7, 2024, from <https://hbr.org/2013/04/the-hidden-biases-in-big-data>
- Crippa, M., Girgenti, G. (2024). *Umano, poco umano. Esercizi spirituali contro l'intelligenza artificiale*. Milano: Piemme.
- D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1-2), 45-80.
- Daniel, B. K. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904-920.
- Daniel, B. K. (2019). Big data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*, 50(1), 101-113.
- Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology & Society*, 11(3), 224-238.
- Decuyper, M., Grimaldi, E., Landri, P. (2021). Introduction: Critical studies of digital education platforms. *Critical Studies in Education*, 62(1), 1-16.
- Dekker, G. W., Pechenizkiy, M., Vleeshouwers, J. M. (2009). Predicting students drop out: A case study. In *Proceedings of the 2nd International Conference on Educational Data Mining*, EDM 2009, July 1-3, 2009. Cordoba, Spain (pp. 41-50).
- Durkheim, E. (1955). *Pragmatism and Sociology*. Cambridge: Cambridge University Press.
- Educational Data Mining Consortium. (n.d.). Educational Data Mining. Retrieved June 7, 2024, from <http://www.educationaldatamining.org/>
- Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Bor-

- gman, C. L., Ribes, D., Burton, M., Calvert, S. (2013). *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor: Deep Blue.
- Ester, M., Kriegel, H.P., Sander, J., Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226-231). New York: AAAI Press.
- Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., Vuorikari, R. (2016). Research evidence on the use of learning analytics: Implications for education policy. *Science for Policy Reports*. Seville: Joint Research Centre.
- Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, 44(1), 130-160.
- Floridi, L. (2016). Should we be afraid of AI? *Aeon*, 9 May. Retrieved June 7, 2024, from <https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible>
- Floridi, L. (2022). *Etica dell'Intelligenza artificiale. Sviluppi, opportunità, sfide*. Milano: Cortina editore.
- Gabbi, E. (2021). La sfida dei Learning analytics per gli insegnanti tra ambienti digitali di apprendimento e competenze professionali. *Formazione & Insegnamento*, 19(2), 193-205.
- Gabbi, E. (2023). *Learning Analytics per lo sviluppo professionale: un approccio metodologico allo studio delle comunità di pratica*. Milano: Francoangeli.
- Gerber, N., Gerber, P., Volkamer, M. (2018). Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, 77, 226-261.
- Ghorpade, S. J., Patil, S. S., Chaudhari, R. S. (2020). Educational data mining: Tools and techniques study. *International Journal of Research and Analytical Reviews*, 7, 520-525.
- Gigliozzi, G., Mordenti, R., Zampolli, A. (2000). *La Bella e la Bestia (Italianistica e Informatica)*. Torino: Tirrenia Stampatori.
- Govindarajan, M. (2021). Educational data mining techniques and applications. In *Advancing the Power of Learning Analytics and Big Data in Education* (pp. 234-251). Hershey: IGI Global.

- Grimaldi, R., Cavagnero, S. M., Gallina, M. A. (a cura di). (2019). *Big Data e tracce digitali: Una ricerca internazionale tra studenti universitari*. (Teoria e ricerca in educazione; Vol. 32). Ariccia: Aracne.
- Hogan, A., Sellar, S., Lingard, B. (2015). Network restructuring of global edu-business: the case of Pearson's Efficacy Framework. In W. Au, & J. J. Ferrare (Eds.), *Mapping Corporate Education Reform: Power and Policy Networks in the Neoliberal State* (pp. 43-64). London: Routledge.
- Ibrahim, F. A., Shiba, O. A. (2019). Data mining: WEKA software (an overview). *Journal of Pure and Applied Sciences*, 18(3), 54-58.
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178-210.
- Ispas, C. (2019). Big Data: tra necessità e responsabilità. In R. Grimaldi, S. M. Cavagnero, & M. A. Gallina (a cura di), *Big Data e tracce digitali: Una ricerca internazionale tra studenti universitari*. (Teoria e ricerca in educazione; Vol. 32). Ariccia: Aracne.
- Jassim, M. A., Abdulwahid, S. N. (2021). Data mining preparation: Process, techniques and major issues in data analysis. In *IOP Conference Series: Materials Science and Engineering*, 1090(1), 012053. Bristol: IOP Publishing.
- Johnson, S.C. (1967). Hierarchical Clustering Schemes. *Psychometrika*, 32(3), 241-254.
- Kerssens, N., van Dijck, J. (2022). Governed by edtech? Valuing pedagogical autonomy in a platform society. *Harvard Educational Review*, 92(2), 284-303.
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: Sage.
- Klose, M., Desai, V., Song, Y., Gehringer, E. (2020). EDM and privacy: Ethics and legalities of data collection, usage, and storage. In *Proceedings of the 13th International Conference on Educational Data Mining (EDM)*. Washington: ERIC.
- Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks: Sage Publications.
- Krpan, D., Stankov, U. (2012). Analysis of real-life experience gained from e-learning system. In *Proceedings of the 35th International Convention MIPRO* (pp. 753-757). Opatija: IEEE.

- Landri, P. (2018). *Digital Governance of Education: Technology, Standards and Europeanization of Education*. London: Bloomsbury Publishing.
- Lang, C., Siemens, G., Wise, A., Gasevic, D. (Eds.). (2017). *Handbook of Learning Analytics*. New York: SOLAR, Society for Learning Analytics and Research.
- Larose, D. T., Larose, C. D. (2015). *Data Mining and Predictive Analytics* (1st ed.). Hoboken: John Wiley & Sons, Inc.
- Lei, X. F., Yang, M., Cai, Y. (2016). Educational data mining for decision-making: A framework based on student development theory. In *2nd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2016)* (pp. 628-641). Amsterdam: Atlantis Press.
- Mackenzie, A. (2015). The production of prediction: what does machine learning want? *European Journal of Cultural Studies*, 18(4-5), 429-445.
- Mandinach, E. B., Gummer, E. S. (2016). Every teacher should succeed with data literacy. *Phi Delta Kappan*, 97(8), 43-44.
- Mayer-Schönberger, V., Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.
- Medhat, W., Hassan, A., Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
- Mehra, C., Agrawal, R. (2020). Educational data mining approaches, challenges and goals: A review. *JIMS8I-International Journal of Information, Communication and Computing Technology*, 8(2), 442-447.
- Melendez-Armenta, R., Huerta-Pacheco, N., Morales-Rosales, L., Rebolledo-Mendez, G. (2020). How do students behave when using a tutoring system? Employing data mining to identify behavioral patterns associated to the learning of mathematics. *International Journal of Emerging Technologies in Learning (iJET)*, 15(22), 39-58.
- Miettinen, P. (2014). Interactive Data Mining Considered Harmful (If Done Wrong). KDD 2014 Workshop on Interactive Data Exploration and Analytics (IDEA), 85-87. Retrieved June 7, 2024, from <http://people.mpi-inf.mpg.de/~pmiettinen/papers/miettinen14interactive.pdf>

- Moretti, F. (2005). *La letteratura vista da lontano*. Milano: Einaudi. (Postfazione di A. Piazza).
- Moretti, F. (2007). *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.
- Moretti, F. (2013). *Distant Reading*. London: Verso.
- Nadeau, D., Sekine, S. (2007). A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1), 3-26.
- Naisbitt, J. (1984). *Megatrends*. New York: Warner Books.
- Nauta, L. (2013). Lorenzo Valla. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition). Retrieved June 7, 2024, from <http://plato.stanford.edu/archives/sum2013/entries/lorenzovalla/>
- Nguyen, A., Gardner, L., Sheridan, D. (2020). Data analytics in higher education: An integrated view. *Journal of Information Systems Education*, 31(1), 61-71.
- Nielsen, M. (2015). Who owns big data? *Change: 19 Key Essays on How the Internet Is Changing Our Lives*. Open Mind. Retrieved June 7, 2024, from www.bbvaopenmind.com/en/book/19-key-essays-on-how-internet-is-changing-our-lives/
- Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., Damasevicius, R. (2021). Artificial neural networks for educational data mining in higher education: A systematic literature review. *Applied Artificial Intelligence*, 35(13), 983-1021.
- Öztürk, A. (2016). *Educational data mining: Applications and trends*. Eski ehir: Anadolu University.
- Pal, A. K., Pal, S. (2013). Data mining techniques in EDM for predicting the performance of students. *International Journal of Computer Information Technology*, 2(06), 1110-1116.
- Panciroli, C., Rivoltella, P. C. (2023). *Pedagogia algoritmica. Per una riflessione educativa sull'Intelligenza Artificiale*. Brescia: Scholé-Morcelliana.
- Papamitsiou, Z., Economides, A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society*, 17(4), 49-64.
- Perez, B., Castellanos, C., Correal, D. (2018). Applying data mining techniques to predict student dropout: A case study. In *2018 IEEE*

- 1st Colombian Conference on Applications in Computational Intelligence (Colcaci)* (pp. 1-6). Piscataway, NJ: IEEE.
- Perrotta, C., Williamson, B. (2016). The social life of Learning Analytics: Cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*, 43(1), 3-16.
- Perrotta, C., Gulson, K. N., Williamson, B., & Witzemberger, K. (2021). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. *Critical Studies in Education*, 62(1), 97-113.
- Previtali, G. (2023). *Che cosa sono le digital humanities*. Roma: Carocci.
- Prinsloo, P., Slade, S. (2017). Big data, higher education and learning analytics: Beyond justice, towards an ethics of care. In *Big data and learning analytics in higher education: Current theory and practice* (pp. 109-124). Springer.
- Raffaghelli, J. E., Cuccharia, S., Persico, D. (2015). Methodological approaches in MOOC research: Retracing the myth of Proteus. *British Journal of Educational Technology*, 46(3), 488-509.
- Ranieri, M., Cuomo, S., Biagini, G. (2024). *Scuola e intelligenza artificiale: Percorsi di alfabetizzazione critica* (1a ed., pp. 1-145). Roma: Carocci. (Collana Tascabili Faber).
- Ray, S., Saeed, M. (2018). Applications of educational data mining and learning analytics tools in handling big data in higher education. In *Applications of big data analytics: Trends, issues, and challenges* (pp. 135-160). Cham: Springer.
- Rockwell, G., Berendt, B. (2016). On big data and text mining in the humanities. In *Data mining and learning analytics: Applications in educational research* (pp. 29-40). John Wiley & Sons.
- Roiger, R. J., Geatz, M. W. (2004). *Introduzione al data mining*. Milano: McGraw-Hill.
- Romero, C., Espejo, P.G., Zafra, A., Romero, J.R., Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, 21, 135-146.
- Romero, C., Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146.
- Romero, C., Ventura, S. (2010). Educational data mining: A review of

- the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618.
- Romero, C., Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27. <https://doi.org/10.1002/widm.1075>
- Romero, C., Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355.
- Romero, C., Ventura, S., Pechenizkiy, M., Baker, R. S. (Eds.). (2010). *Handbook of Educational Data Mining*. Boca Raton, FL: CRC Press.
- Ruppert, E. (2015). Who owns big data? *Discover Society*, 30 July. Retrieved June 7, 2024, from <http://discoversociety.org/2015/07/30/who-owns-bigdata/>
- Ruppert, E., Harvey, P., Lury, C., Mackenzie, A., McNally, R., Baker, S. A., Kallianos, Y., Lewis, C. (2015). Socialising big data: from concept to practice. *CRESC Working Paper* no. 138. Retrieved June 7, 2024, from www.cresc.ac.uk/medialibrary/workingpapers/wp13-8.pdf
- Safitri, S. N., Setiadi, H., Suryani, E. (2022). Educational data mining using cluster analysis methods and decision trees based on log mining. *Journal of System and Technology Information (RESTI)*, 6(3), 448-456.
- Savin-Baden, M., Tombs, G. (2017). *Research Methods for Education in the Digital Age*. London: Bloomsbury Publishing.
- Savoca, G. (2000). *Lessicografia letteraria e metodo concordanziale*. Firenze: Olschki.
- Scheuer, O., McLaren, B. M. (2012). Educational data mining. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning*. Springer Science & Business Media.
- Sellar, S. (2015). A feel for numbers: affect, data and education policy. *Critical Studies in Education*, 56(1), 131-146.
- Shruthi, P., Chaitra, B. (2016). Student performance prediction in education sector using data mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(3), 123-126.
- Siemens, G., Baker, R. S. D. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Procee-*

- dings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254).
- Siemens, G. (2013). Learning analytics: the emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400.
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., ... Baker, R. S. J. D. (2011). *Open Learning Analytics: An Integrated & Modularized Platform*. Open University Press.
- Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), 477-564.
- Spirtes, P., Glymour, C., Scheines, R. (2001). *Causation, Prediction, and Search*. Cambridge: The MIT Press.
- Suthers, D. D., Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. In *Proceedings of the First International Conference on Learning Analytics & Knowledge*, Banff, Alberta, February 27-March 1, 2011.
- Taglietti, D., Landri, P., Grimaldi, E. (2021). The big acceleration in digital education in Italy: The COVID-19 pandemic and the blended-school form. *European Educational Research Journal*, 20(4), 423-441.
- Valla, L. (1985). *The Profession of the Religious and the Principal Arguments from the Falsely-Believed and Forged Donation of Constantine* (O. Z. Pugliese, Trans.). Toronto: Centre for Reformation and Renaissance Studies.
- Wasserman, S., Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Watters, A. (2016). Ed-tech patents: prior art and learning theories. *Hack Education*, 12 January. Retrieved June 7, 2024, from <http://hackeducation.com/2016/01/12/patents>
- Williamson, B. (2017). *Big Data in Education: The Digital Future of Learning, Policy and Practice*. London: SAGE.
- Young, J. R. (2016). What clicks from 70,000 courses reveal about student learning. *Chronicle of Higher Education*, 7 September. Retrieved June 7, 2024, from www.chronicle.com/article/What-Clicks-From-70000/237704/
- Zaïane, O. R. (2002). Building a recommender agent for e-learning systems. In *Proceedings of the International Conference on Computers in Education (ICCE 2002)* (pp. 55-59). IEEE.
- Zhu, R., Hu, X., Hou, J., Li, X. (2021). Application of machine learning

ning techniques for predicting the consequences of construction accidents in China. *Process Safety and Environmental Protection*, 145, 293-302.

Zoric, A. B. (2019). Benefits of educational data mining. In *Economic and Social Development: Book of Proceedings*. Varazdin: Varazdin Development and Entrepreneurship Agency.

Finito di stampare
nel mese di GIUGNO 2024 da



per conto di Pensa MultiMedia® • Lecce
www.pensamultimedia.it



collana

Didattiche, Tecnologie e Media Education
Frontiere per la sostenibilità

Co-directors:

Marina De Rossi e Monica Fedeli

Gabriele Biagini. Ha conseguito un master in Data Science and Statistical Learning, attualmente sta svolgendo un dottorato di interesse nazionale in Learning Sciences and Digital Technologies e conduce le sue attività di ricerca sull'intelligenza artificiale per il Laboratorio di Tecnologie dell'Educazione dell'Università degli Studi di Firenze.

Elena Gabbi. PhD in Scienze della Formazione e Psicologia, è attualmente assegnista di ricerca presso l'Università degli Studi di Firenze svolgendo attività di ricerca e insegnamento nell'ambito delle tecnologie dell'educazione, con particolare interesse verso l'alfabetizzazione ai dati e le competenze digitali dei docenti.

La collana si inserisce in una cornice storica e politico-normativa che si orienta verso una trasformazione epocale delle didattiche a tutti i livelli dall'istruzione formale e in tutti i contesti educativi e formativi è necessario ripensare alla "nuova normalità" come sfida irrinunciabile, aperta alla trasformazione e fortemente orientata al paradigma ecologico dell'apprendimento permanente, della sostenibilità e dell'innovazione.

Significa considerare gli obiettivi posti dall'Agenda 2030 declinandoli nei contesti operativi di educazione e formazione in termini progettuali e metodologici volti a favorire accessibilità e inclusione tenendo conto di tratti fondamentali, ma spesso inediti, di condizioni esistenziali della realtà mutati e sempre più rappresentati dagli aggettivi "Volatile, Uncertain, Complex, Ambiguous" (VUCA).

La sfida è l'inversione del processo di apprendimento «Dal saper fare al sapere e al saper essere», come ci ricordava nel 2014 E. Morin in «Insegnare a vivere. Manifesto per cambiare l'educazione». Questa intende essere la nostra prospettiva di lavoro e di ricerca per la formazione iniziale e continua di educatori, formatori e insegnanti affermando il contrasto a logiche deterministiche e di delegittimazione del sapere e della conoscenza a favore di nuove economie.

In tale prospettiva, le istanze relative al tema dell'integrazione fattiva di competenze didattiche e digitali dovranno trovare sempre più ambiti di riflessione scientifica e prassica coinvolgendo studiosi e studiosi, ricercatori e ricercatrici, insegnanti, docenti, formatori e formatrici a ripensare modelli e approcci didattici in chiave propositiva per rendere persone, gruppi e comunità in grado di affrontare situazioni personali e professionali complesse e in continuo cambiamento.





Il volume esplora la contemporanea intersezione tra i dati e l'educazione, un territorio in cui i big data non solo influenzano, ma trasformano attivamente le modalità di insegnamento e apprendimento.

L'opera si apre con un'analisi del ruolo dei big data nella società dell'informazione e dell'influenza che essi esercitano sulle nostre vite individuali e sociali, sollevando questioni di ordine cognitivo, etico e sociale: in che modo sta cambiando il nostro modo di concepire i dati? Quali sono le implicazioni etiche legate alla raccolta e all'utilizzo dei dati in ambito educativo? Quale impatto l'uso dei big data nell'istruzione può avere sull'accesso equo alle opportunità educative? Tali questioni richiedono nuove forme di consapevolezza che vanno promosse tra i cittadini del nuovo millennio per favorire lo sviluppo di *data literacy*. Il libro offre una panoramica sui significati del concetto di *data literacy*, soffermandosi sui suoi fondamenti teorici e fornendo un quadro complessivo di cosa significhi essere alfabetizzati ai dati nel XXI secolo. Vengono così poste le basi per un ripensamento delle componenti di lettura e interpretazione dei dati a favore di una più completa padronanza a favore di cittadinanza digitale, responsabile e critica.

Accanto alla teoria, il volume propone anche riferimenti utili per la pratica, esplorando come la *data literacy* possa essere insegnata e applicata nel panorama educativo, confrontando modelli e prassi didattiche esistenti sull'analisi, la visualizzazione, l'interpretazione e la creazione dei dati. Vengono discusse alcune significative esperienze di ricerca e casi di studio che illustrano l'implementazione e la valutazione di percorsi di alfabetizzazione ai dati nelle istituzioni educative. Il volume si conclude con una disamina delle pratiche e tecniche di educational data mining (EDM). L'obiettivo è far comprendere in che modo l'analisi dei dati possa essere utilizzata per migliorare sia i metodi di insegnamento sia i risultati di apprendimento, presentando metodi specifici ed esempi pratici di applicazioni di EDM.

Attraverso una narrazione che fonde elementi di contesto con esempi operativi, il volume si rivolge a formatori, insegnanti e cittadini. Nell'offrire una visione complessiva e dettagliata della progressiva rilevanza attribuita ai dati e di come tale risalito stia ridefinendo l'ambito educativo, pone una particolare enfasi sulle competenze critiche richieste per navigare in questa nuova era dell'informazione.