International Journal of Tourism Research

WILEY

RESEARCH ARTICLE OPEN ACCESS

A Visitor-Centric Recommender System for Florentine Museums: A Foundation for Sustainable Tourism Management

Stefano Masini[1] | Silvia Bacci[1] | Alessandra Petrucci[1] | Carlo Francini[2] | Bruno Bertaccini[1]

[1]Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence, Florence, Italy | [2]World Heritage and UNESCO Relationship Office, Municipality of Florence, Florence, Italy

Correspondence: Silvia Bacci (silvia.bacci@unifi.it)

Received: 8 December 2024 | **Revised:** 20 February 2026 | **Accepted:** 31 July 2026

Keywords: artificial intelligence | deep learning | FirenzeCard project | neural networks | two-tower model

ABSTRACT

The paper presents the development of a recommender system designed to capture visitors' preferences during museum visits in Florence. The system is trained on data from the entry records of the *FirenzeCard*, the official museum pass of the city offering tourists access to the Florentine museums. The system utilises an ensemble of deep learning models based on a two-tower architecture, in which separate representations of visitors and museums are learned independently and then combined to generate predictions. It incorporates structured and unstructured data and contextual and visitor features. Results show that the integration of new components into the system architecture significantly enhances the recommender system's ability to accurately predict the top-*K* museums a visitor is likely to visit next. This aspect makes it one foundational layer of a larger decision-making framework: a reinforcement learning agent to dynamically manage visitor flows, alleviate crowding in overexposed cultural sites and balance urban resources.

1 | Introduction

Tourism is a powerful and multifaceted driver of economic development, cultural exchange and social interaction. However, in recent years, the positive impacts of tourism have increasingly been offset by its negative externalities, especially in destinations experiencing what has come to be known as overtourism (Coca-Stefaniak et al. 2016; Dodds and Butler 2019; Forster 1964; Koens et al. 2018). This phenomenon occurs when the number of visitors surpasses the hosting capacity of a destination, significantly increasing the risk of environmental degradation, harm to cultural heritage and pressure on public services, ultimately reducing the quality of life for residents. The consequences are multiple and far-reaching: congestion, environmental degradation, commodification of local traditions, rising living costs and growing difficulty in managing small-scale historical and cultural sites while maintaining a balance between residents' daily lives and visitors' needs. The overarching challenge lies in ensuring that the city remains liveable and enjoyable.

Overtourism is not just a matter of numbers, but also of where and when tourism occurs. Tourist flows often concentrate in specific areas and periods, leaving other equally valuable parts of a destination under-visited. This results in uneven distribution of tourist pressure, overexposure of iconic sites and underutilisation of lesser-known attractions. These dynamics are especially problematic in the context of urban cultural tourism. Cities of art, rich in historical landmarks and cultural institutions, are particularly vulnerable to the negative externalities of unbalanced tourism flows. This is especially true for smaller cities with high densities of cultural assets. Florence (Italy) stands as a prominent example. Its historic centre, a UNESCO World Heritage Site since 1982, hosts an exceptional concentration of museums, monuments and artistic treasures within a compact urban layout. Florence attracts millions of visitors each year, and while this spatial compactness contributes to its charm, it also exacerbates the logistical and cultural challenges of accommodating large tourist volumes. Long queues at major museums such as the Uffizi and the Accademia, overcrowded streets and infrastructure bottlenecks are commonly observed during peak seasons. These conditions threaten the sustainability of Florence's tourism model and underscore the urgent need for intelligent, data-driven policies to better manage and redistribute tourist flows across time and space.

To address these challenges, it is essential to equip policymakers with practical tools for managing visitor flows. Literature presents a wide array of proposals (Dodds and Butler 2019; Koens et al. 2018) addressing overtourism in various contexts, including urban and regional settings (Milano et al. 2019; Peeters et al. 2018; WTO et al. 2019), often framed within smart tourism and smart city paradigms (Bastidas-Manzano et al. 2021; Gretzel et al. 2015). Nevertheless, many of the proposed solutions remain theoretical or visionary, often misaligned with local governments' practical and urgent needs (Hollands 2015; Ribes and Baidal 2018). Namely, there is a pressing need for strategies that are not only implementable in the short term but also capable of delivering measurable results (Hospers 2019).

This paper proposes a tool that can be readily implemented in contexts where systems for automatically tracking tourist movements are already in place, primarily museum passes and public transportation cards. When appropriately integrated with Artificial Intelligence systems, such tools can yield valuable management insights without requiring additional user data collection, such as individual profiles, which are often difficult to obtain. Among the many digital innovations in the tourism sector, Recommender Systems (RSs) have emerged as a particularly effective means of enhancing tourist experiences by providing personalised recommendations aligned with users' interests and preferences (Borràs et al. 2014; Hamid et al. 2021; Sarkar et al. 2023). RSs also offer policymakers a 'soft' intervention tool for mitigating overtourism by subtly influencing tourist behaviour. Rather than imposing limits, constraints and restrictions on tourists, such systems gently guide tourists toward lesser-known historical and cultural sites that match their preferences.

This study introduces a museum RS as a first step toward a broader strategic framework for sustainable tourism management. The system's core objective is to anticipate visitor behaviour and preferences by analysing patterns observed in similar users. It accurately predicts which museums a visitor is likely to visit next by analysing their visitation history, patterns observed among similar users and the intrinsic attributes of the museums, such as thematic content, geographic location and overall popularity.

Technically, the system is implemented using a two-tower ensemble learning architecture (Covington et al. 2017), which integrates collaborative and content-based filtering strengths. One tower encodes user-specific information from historical visitation sequences and behavioural data, capturing collaborative signals through user-item co-occurrence and latent preference structures. The other tower models museum-specific attributes, including thematic content, spatial features and metadata, thus incorporating content-based knowledge. Both towers generate embeddings projected into a shared space, where relevance scores are computed using similarity functions. In this way, user and museum similarities are balanced within a unified framework. Additionally, a robust feature engineering pipeline enriches the model with contextual variables such as visitor location, time of visit, visit duration and geographic constraints, enhancing its accuracy and adaptability.

The RS is trained on data collected through the *FirenzeCard* program (https://firenzecard.it/en), a 72-h museum pass providing access to numerous cultural sites in Florence and its surrounding areas. This card, widely adopted by tourists, generates a rich behavioural dataset including visit sequences, timing and frequency. This data offers insights into tourist movement patterns, site prioritisation and evolving preferences. A preliminary analysis of the FirenzeCard dataset is presented in Masini et al. (2025), to which the reader is referred for further methodological details.

The rest of the article is organised as follows. Section 2 reviews the literature on RSs, focusing on their applications and limitations within the tourism context. Section 3 details the data collection process and feature engineering steps and then presents the architecture of the proposed two-tower model. Section 4 discusses the main findings related to the system's implementation. Finally, Section 5 reflects on the theoretical contributions and practical implications, concluding with a discussion of the study's limitations and directions for future research.

2 | Literature Review

This section reviews the literature about RSs, focusing on their use in the tourism context. We first describe the main typologies of RS, classified according to the kind of information used to make recommendations. Then, we focus on the learning systems adopted to process and analyse the data and on integrating contextual features (i.e., geographical location of users and items) to improve the predictive accuracy.

2.1 | Recommender Systems for Tourism

In recent years, the tourism industry has undergone a significant transformation due to advancements in digital technology and the increasing availability of data: nowadays we have plenty of information about travels, reviews on accommodation and catering facilities and many other digital footprints left by tourists. This data often arises from recommendation web systems, which are powerful tools for improving tourist experiences by providing personalised recommendations tailored to individual (declared or inferred) preferences and interests (Ricci et al. 2011).

These tools are generally classified based on two main distinct development techniques (Ricci et al. 2011): collaborative filtering and content-based filtering. For an overview of these techniques in the tourism setting, see Al Fararni et al. (2021), Lucas et al. (2013), and Solano-Barliza et al. (2024), while specific applications are provided by, without pretending to be exhaustive, Abbasi-Moud et al. (2021), Colomo-Palacios et al. (2016), and Nan et al. (2022). Collaborative filtering makes recommendations based on the preferences and behaviours of a community of users to suggest items that similar users have enjoyed (Resnick et al. 1994; Sarwar et al. 2001). The basic principle is that if two users have similar preferences or behaviours, items selected by one user can be recommended to the other. This approach is particularly effective in capturing latent patterns and social proof effects. However, it requires a large amount of user data to generate accurate recommendations and suffers from the cold-start problem when either users or items lack sufficient interaction history. Differently, content-based filtering makes recommendations based on the characteristics of the items (Pazzani and Billsus 2007). The system suggests items similar to ones the user has already chosen or interacted with. This approach can be more robust in cold-start scenarios but may lead to over-specialisation, limiting the system's ability to suggest novel or diverse content. Given that both methods have their strengths and limitations, for a few years now, many modern RSs have been employing a hybrid approach (Burke 2002) that combines both techniques, also integrated with various Data Mining methods, to exploit their advantages and mitigate their weaknesses. In cultural tourism, such as museum visit recommendations, hybrid approaches have proven especially useful due to the heterogeneity of user profiles and the semantic richness of cultural content.

In cultural tourism, where travellers seek immersive experiences in historical landmarks and cultural institutions, recommendation systems have been playing a relevant role for some years, guiding visitors through a plethora of attractions while maximising their enjoyment and engagement. By way of example only, and without any claim to be exhaustive, Lucas et al. (2013) implemented a recommendation system for tourism exploiting an associative classification method, which is a data mining technique that combines concepts from classification and association to enhance recommendation quality. Colomo-Palacios et al. (2016) presented a social and context-aware mobile RS for tourists that performs its recommendations based on data from previous visits while also combining aspects taken from the social environment of the tourist. The system was built to provide efficient pattern detection and rapid adaptability to changes, including relevant recommendations from Destination Management Organisations, which are nonprofit organisations responsible for managing and marketing specific destinations. Abbasi-Moud et al. (2021) presented a recommendation system for tourism that leverages both semantic clustering and sentiment analysis techniques. Semantic clustering was used to group similar tourism-related data. This helps in understanding the semantic context and meaning behind user reviews and descriptions of tourist attractions or services. Sentiment analysis was also incorporated to estimate the emotional tone and sentiment in user-generated content, such as reviews and posts. By combining these two techniques, the proposed system succeeded in more personalised and accurate recommendations, enhancing tourist experience. Nan et al. (2022) discussed the development of a personalised tourism recommendation system mixing data mining techniques with a collaborative filtering strategy. Data mining techniques were employed to analyse large sets of tourism-related data, including user behaviour, preferences and historical data, to identify patterns and trends.

Among the hybrid RS proposals, the two-tower model (Covington et al. 2017; Huang et al. 2013), known also as the dual-encoder model, has been successfully applied in large-scale industrial recommendation scenarios, enabling efficient candidate retrieval and scoring. The two-tower model relies on two subnetworks, one for the users (user tower) and one for the items (item tower). The user tower encodes user features like past behaviour and preferences into a user embedding, while the item tower encodes item features like metadata and content into an item embedding. This architecture has several advantages (Karpukhin et al. 2020; Malkov and Yashunin 2020). Namely, it is efficient in large-scale scenarios (high scalability) as it can use precomputed embeddings and allows for updating a single tower when a new item or user enters (instead of training all over again). Moreover, this architecture is well-suited to scenarios with heterogeneous inputs as it can flexibly incorporate embeddings of different modalities, such as images, texts and behaviours, into the same vector space.

Furthermore, it is particularly useful when the aim is to provide a top-K recommendation (as in our context, where we are interested in recommending the next K museums to visit), as it directly optimises the probability of selection of each item (i.e., the probability of museum visit). Finally, the problem of cold-start is scaled down, as the two-tower model is suitable for recommending a new item even when the history of interactions with users is not available, being sufficient to have metadata (e.g., the description of a museum). For these reasons, an RS based on a two-tower model is particularly suitable for application to the context of museum visit recommendations. To date, the use of two-tower models in the tourism setting appears to be substantially new. To our knowledge, the only work is that recent one of Cui et al. (2024), which demonstrates how multimodal data can be leveraged in a two-tower RS for tourism. In their approach, the user tower inputs the user ID and a brief profile description. In contrast, the item tower incorporates textual descriptions and visual data (e.g., images and videos) of tourist attractions to construct multimodal embeddings.

Our proposal differs substantially from Cui et al. (2024), as it is designed to work with *implicit* preference settings, where detailed user profiles are typically unavailable. Thus, we focus on a situation common to several types of passes around the world (mainly, passes for museums and passes for urban public transport), that is, the absence of any information on the pass holders, such as age, gender and nationality. Instead of relying on predefined personal information, our user representation (user tower) is built upon a richer set of behavioural and contextual features, implicitly provided by the sequence of previously visited museums and the geolocation. On the item side, while we also incorporate textual descriptions of the museums (making our system multimodal), we extend the item representation beyond this. We include structured metadata such as museum typology, accurate description of collections, location and spatial correlations derived from cross-visitation patterns. This design enables our model to generalise more effectively in real-world tourism scenarios where user identities are anonymous, and profiles must be learned from user behaviour alone.

2.2 | Sequence-Based Deep Learning Systems

In addition to the type of information used by the RS that leads to different configurations of the system (e.g., collaborative-based, content-based, hybrid), a further distinguishing feature is represented by the learning algorithms used for data processing to identify underlying patterns and similarities that can be employed to generate recommendations. The comprehensive review by Solano-Barliza et al. (2024) highlights the evolution and application of different types of learning systems across the tourism sector. The authors outline how clustering is the most commonly adopted technique in this setting. However, they also identify a growing trend towards context-aware, location-based and deep learning systems, particularly those capable of integrating real-time environmental data, dynamic visitor profiles and spatial–temporal information. In the context of museum visits, where tourist behaviour tends to be sequential and time-dependent, sequence-based deep learning systems are the natural solution.

Classical sequence-based deep learning approaches include Markov models and probabilistic graphical models, which estimate the probability of visiting the next point of interest based on the historical sequence of visited locations (Kurashima et al. 2013). Another strand of literature focuses on point of interest recommendation (Gavalas et al. 2014), where the goal is to suggest attractions or sites to users based on relevance, proximity and past behaviours. These models typically integrate spatial–temporal data, contextual factors (e.g., time of day, weather) and popularity metrics. However, they struggle with the cold-start problem and lack of robustness in contexts where museum visits are strictly limited by opening hours and capacity limits, as in our case. Moreover, these models are relatively interpretable but limited in their ability to capture the dynamic nature of spatial–temporal sequences.

To address these shortcomings, the recent literature in the tourism setting (e.g., Feng et al. 2015; Liu et al. 2016) has proposed the Recurrent Neural Networks (RNN; Rumelhart et al. 1986), among which, in particular, the Long Short-Term Memory (LSTM; Hochreiter and Schmidhuber 1997) networks and their more efficient version represented by the Gated Recurrent Units (GRU; Cho et al. 2014). These methods allow the system to model temporal dependencies across an entire travel trajectory, often incorporating auxiliary information such as travel times, distances, or user demographics.

In such a context, the work by Zheng et al. (2025) is relevant, who proposed a novel deep learning model for intra-city tourism flow forecasting. The authors leverage a spatial graph neural network combined with the attention mechanism to model the tourist trajectory holistically. Their dataset consists of real-world taxi trajectory data collected from a major Chinese city, which they use to infer tourist movements and visitation patterns across urban attractions. To model intra-city flows, Zheng et al. (2025) construct a directed spatial graph in which nodes correspond to distinct urban areas—typically centred around points of interest—and edges represent the movement of tourists between these areas, inferred from aggregated taxi pick-up and drop-off data. Edge weights reflect transition probabilities derived from empirical taxi flows, allowing the model to account for geographical proximity and empirical mobility patterns. Attention mechanisms are applied over this graph to emphasise high-flow routes and key decision points within a tourist's path.

While Zheng et al. (2025)'s approach provides a strong foundation for flow prediction in data-rich urban contexts, it does not explicitly model user-level preferences or content features of destinations. As a result, their method focuses more on macro-scale mobility patterns rather than fine-grained, user-specific recommendations based on individual interests or visitation histories.

Our system combines personalised modelling with a modular deep learning two-tower architecture that can adapt to micro- and macro-tourism scenarios. We enrich the user tower with a GRU component to account for the dynamic nature of the sequences of visited museums. Compared with the other RNNs, GRU provides stability advantages during the training phase and parsimony (i.e., fewer parameters). Moreover, GRUs are particularly suitable to work with medium-short sequences, as the ones treated in our contribution, where the museum visits concentrate in 72 h (corresponding to the duration of the FirenzeCard pass).

3 | Data and Methodological Structure

In this section, we describe the dataset used in our study and outline the methodological framework of the proposed RS. We begin by presenting the dataset, detailing the available features and the structure of the original data. Next, we illustrate the comprehensive feature engineering and enrichment pipeline we developed to transform the raw data into the structured inputs required by the RS. Finally, we describe the architecture of the proposed two-tower model, following an incremental approach that starts with a baseline configuration and progresses to an enhanced model incorporating contextual features.

3.1 | Data Collection

Data for training and testing the RS are collected as part of the FirenzeCard project (https://firenzecard.it/en). The FirenzeCard is the official 72-h museum pass (with a possible extension of 48 additional hours) for the city of Florence, granting access (often on a priority basis) to 39 museums and other permanent exhibits (collectively referred to as 'museums' from now on) within the city and its surrounding areas. It is worth outlining that almost all the museums of the circuit are included in the UNESCO area and, more precisely, in a circumference of radius 1 km centred in the forum of the ancient Roman city of Florentia (see Figure 1).

At each venue, visitors are required to present their card, which is scanned using an optical reader. In this way, the sequences of museums visited by the cardholders are recorded, with details on *which* museums were visited, *when* and *in what order*. Additional information about the cardholders' profiles (e.g., gender, nationality, ...) is not available. Although this dataset seems quite poor in terms of useful information for the purposes of our research, the visiting sequences and

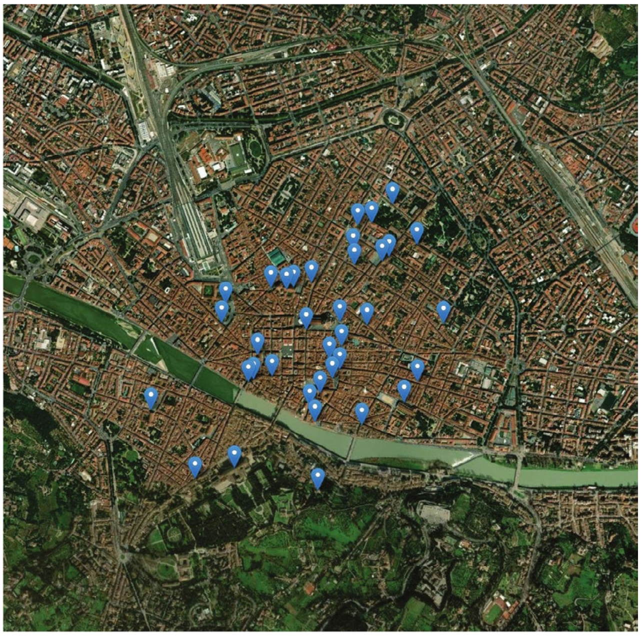

FIGURE 1 | Maps of the city of Florence centred in the ancient Roman forum (the image is a square with sides 2 km); blue markers identify the location of the museums and collections adhering to the FirenzeCard.

the times at which these visits occur implicitly reveal additional information that the implemented two-tower-based RS can learn by itself: among these, the days and specific time slots when the museum is open; the time slots during which it is not ideal to suggest a museum visit, even if the museum is still open; and, most important, visitors' preferences regarding the types of collections they wish to explore and the time they spend on visits, even if these preferences have never been explicitly stated.

The analysis presented here relies on data from FirenzeCards sold in 2018. While the data from 2018 may appear outdated, that year marked an unprecedented peak in tourism for Florence, with approximately 11 million recorded tourist arrivals. The resulting dataset comprises 125,092 sold cards for a total of 846,510 museum visits. Each cardholder visited on average 6.8 museums, with 25% of users visiting at most 4 museums and 75% at most 9 museums. See Masini et al. (2025) for further details on the dataset.

3.2 | Feature Engineering and Data Augmentation

The available data from the FirenzeCard database has been enriched with additional features related to both visitors and museums. Some museum-specific features are readily accessible from public sources and include the geographic location of each museum (latitude and longitude), the type and description of the artworks exhibited and the maximum visitor capacity. In contrast, features related to visitors are not as easily obtainable. Among these, the visitor's position in the city before the next museum visit and the duration of each visit are particularly relevant for training and evaluating the RS. Consequently, these features have been estimated as detailed in the following section.

It is important to note that this estimation approach specifically enables the training and testing of the RS with the available data. In a real-world scenario, where the RS will be integrated into a mobile app, the user's location and visit duration would be automatically captured via the geo-location services of the user's mobile device.

3.2.1 | Assessing the Visitor Location

The effectiveness of an RS whose goal is to suggest the next museum to visit is expected to be inversely proportional to the distance to be covered between the position of the person querying the system and the position of the suggested museum: the greater the distance, the less effective the suggestion will be. Thus, the visitor location becomes a crucial feature to accurately recommend the next museum. As mentioned above, such information is not in the FirenzeCard database, thus we estimate it based on the available data, adding it to the set of visitor features.

Thus, we assume that the visitor location (at the request time) is in a neighbourhood area of the last visited museum. Such an area serves as the continuous sampling space from which a location will be drawn. To improve model interpretability and manage non-linear relationships, we apply discretisation, a technique in which continuous variables (like location) are transformed into categorical ‘bins’ or ‘buckets’. In this work, a 1000×1000 grid on the map of Florence (with 1,000,000 discrete buckets) is used to convert continuous location (i.e., latitude and longitude) samples into discrete buckets that the model can process. Points within a set radius R from the last museum visited are randomly sampled, and each location is assigned to the corresponding bucket in this grid. The radius R varies from 500 to 2000 m with incremental steps of 300 m to observe how different sampling areas affect recommendation accuracy. A larger sampling radius is hypothesised to reduce recommendation effectiveness, a hypothesis tested further in Section 5.2, where we discuss the RS performance. Figure 2 shows an example of drawing from two sampling spaces of different radius centred on Brunelleschi's Dome, as the last museum visited.

3.2.2 | Assessing the Average Visit Duration

Another important feature to improve the effectiveness of the recommendation is the time needed to complete the visit to each visited museum. The goal is to capture visitor behaviour, distinguishing between those who visit, for example, the Uffizi Gallery for a short time, only focusing on its most famous artworks, and those who instead enjoy a more comprehensive visiting experience.

The FirenzeCard database does not contain this information: it only records the time the card is swiped under the optical reader at each entry, and no action is required when leaving the museum. We then estimate each visit time by calculating the time difference between two consecutive entries within the same day. The time taken to complete the visit in the last museum of the day cannot be calculated, but the available data are sufficient to

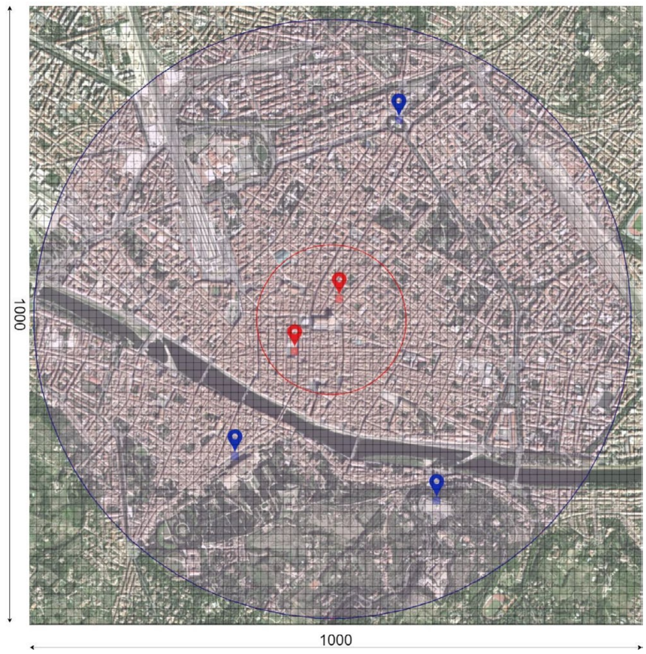

FIGURE 2 | Grid of 1000 × 1000 buckets for discretisation of the map of Florence: 500 m radius sampling space (red circle) and 2000 m radius sampling space (blue circle), centred on Brunelleschi's Dome as the last museum visited. Tick marks identify possible results of the sampling process.

address this limitation. We then compute the average visit time for each museum, adding this information to the set of museum features.

3.2.3 | Museums' Probability of Being Visited and Visitors' Latent Class

Building on the work of Masini et al. (2025), each sequence of museums visited by FirenzeCard holders can be conceptualised as an individual manifestation of an underlying (i.e., latent) trait reflecting the propensity to visit museums. Based on this, cardholders can be differentiated by their varying propensities to visit museums, which are reflected in the length of the visit sequences (i.e., how many museums they visit) and the specific museums included in those sequences.

To operationalise this, each record in the FirenzeCard database is encoded as a sequence of binary (0–1) values, where 1 indicates that the cardholder visited the museum, and 0 otherwise. To link the latent propensity trait to these 0–1 sequences (of length equal to 39, i.e., the number of museums that adhere to the Firenze Card circuit), we adopt a Latent Class Item Response Theory (LC-IRT; Bartolucci 2007; Bartolucci et al. 2015; von Davier 2008; Heinen 1996; Langeheine and Rost 1988; Rost 1990) model, as detailed in Masini et al. (2025), to which we refer for methodological specifics.

This modelling approach yields two outputs that are used to enrich our dataset. First, it provides an estimated probability of being visited for each museum, which is incorporated into the set of museum-specific features. Second, leveraging the latent class formulation, we assume that the visitor population can be segmented into homogeneous groups (i.e., latent classes) sharing a similar propensity to visit museums. For each FirenzeCard holder, we compute the membership probability for each latent class. Based on the maximum a posteriori criterion, each cardholder is then assigned to the most probable latent class and this information is added to the set of visitor-specific features.

3.2.4 | Empowering Information From Visit Sequences

From the recorded entry times, it is straightforward to determine both the weekly closing day and the opening windows on the other days. Moreover, not only which and how many museums are present in a certain sequence, but also the *order* in which museums are visited and the *distance between their ranks* in the sequence provide relevant information about the visitors' preferences. To capture these aspects, we employ a method, inspired by Ranjan et al. (2022), that explicitly incorporates these features into the available dataset.

Given any sequence of visited museums, let s be the vector that identifies the sequence, whose elements are the IDs of visited museums, reported in the order of visit; each sequence s has length $L(s)$, corresponding to the number of visited museums. Let s_l and s_m be the museums at positions (ranks) l and m, respectively, where $l, m = 1, \ldots, L(s)$. For each pair u and v of visited museums in s such that $s_l = u$ and $s_m = v$ (i.e., museum u is the lth visited museum and museum v is the mth visited museum), we compute an association function ψ_uv(s) between museums as

$$\psi_{uv}(s) = \frac{\sum_{(l,m)\in P_{uv(s)}} e^{-hd(l,m)}}{\left|P_{uv(s)}\right|},$$

where $P_{uv(s)}$ is the set of all pairs of position indices of u and v in the sequence s; $d(l, m) = |\, l - m \,|$ is the distance between ranks l and m; and h is an amplification factor accounting for the contribution of distances (the larger h, the smaller the contribution of pair u, v at greater distances).

Hence, for each sequence s we obtain a vector of association measures

$$\psi(s) = \left(\psi_{12}(s), \psi_{13}(s), . \ \ldots, \psi_{uv}(s), \ \ldots\right),$$

whose length is equal to all the possible pairs of museums. This vector is added to the set of visitor features.

3.3 | Methodological Structure

In this section, we provide an overview of the system architecture and the algorithmic approach adopted for the development of the proposed RS. As mentioned above, the system is built upon a two-tower architecture, which has been progressively refined through the integration of additional components.

In a two-tower model, two separate neural networks, or 'towers', are involved in the process. The user tower is responsible for processing the queries (in our case: the FirenzeCard holders or, briefly, the visitors), while the item tower handles the candidate items (in our case: the museums belonging to the FirenzeCard circuit). In the context of RSs, this setup allows the model to encode both the visitor query (e.g., the visiting history contained in the sequences of visited museums or features regarding the current context) and the museum features (e.g., location of museums and other attributes like the total number of visits per museum) separately and then compare them in a joint embedding space. Specifically, embeddings (Liu et al. 2019) are vectors within a multidimensional latent space that are the objective of the training process and are then used to generate museum recommendations.

In the following, we illustrate the compositional nature of the two-tower architecture, showing how the system evolves incrementally: starting from a baseline configuration, we progressively incorporate additional inputs and subsystems to improve the predictive capability. This design is inherently compatible with an ensemble learning paradigm, where multiple sub-models contribute collaboratively to the final prediction.

3.3.1 | Baseline Model

The baseline model is built using only the features available in the original FirenzeCard dataset, consisting of basic information about museums and, crucially, the sequences of museum visits of each cardholder (see Section 3.1 for details on the dataset).

To model the sequential nature of museum visits, we employ an RNN architecture based on GRU, which effectively captures the temporal dependencies and behavioural patterns inherent in the sequences (as outlined in Section 2.2). The architecture of the baseline model is illustrated in Figure 3.

The structure of the visitor tower (Figure 3, left side) is built starting from the sequences of visited museums having an appropriate minimum length (see Section 4.2 for details). Each museum in the input sequence is mapped to a dense vector in a 32-dimensional embedding space. These embeddings are then processed by a GRU network consisting of 32 GRU units, which learn to capture sequential patterns from the ordered museum visits. Then, the output of the GRU network is passed to a stack of two fully connected (dense) layers: the first layer has 32 units with a ReLU activation function, followed by a second layer with 16 linear units. This final 16-dimensional output serves as a visitor embedding, representing user preferences that are not directly observed but implicitly inferred by their visit history. The numbers of units composing the first and second layer (32 and 16, respectively) are hyperparameters empirically selected (see Section 4.2 for details).

The structure of the museum tower (Figure 3, right side) is built, embedding each museum independently into a 16-dimensional

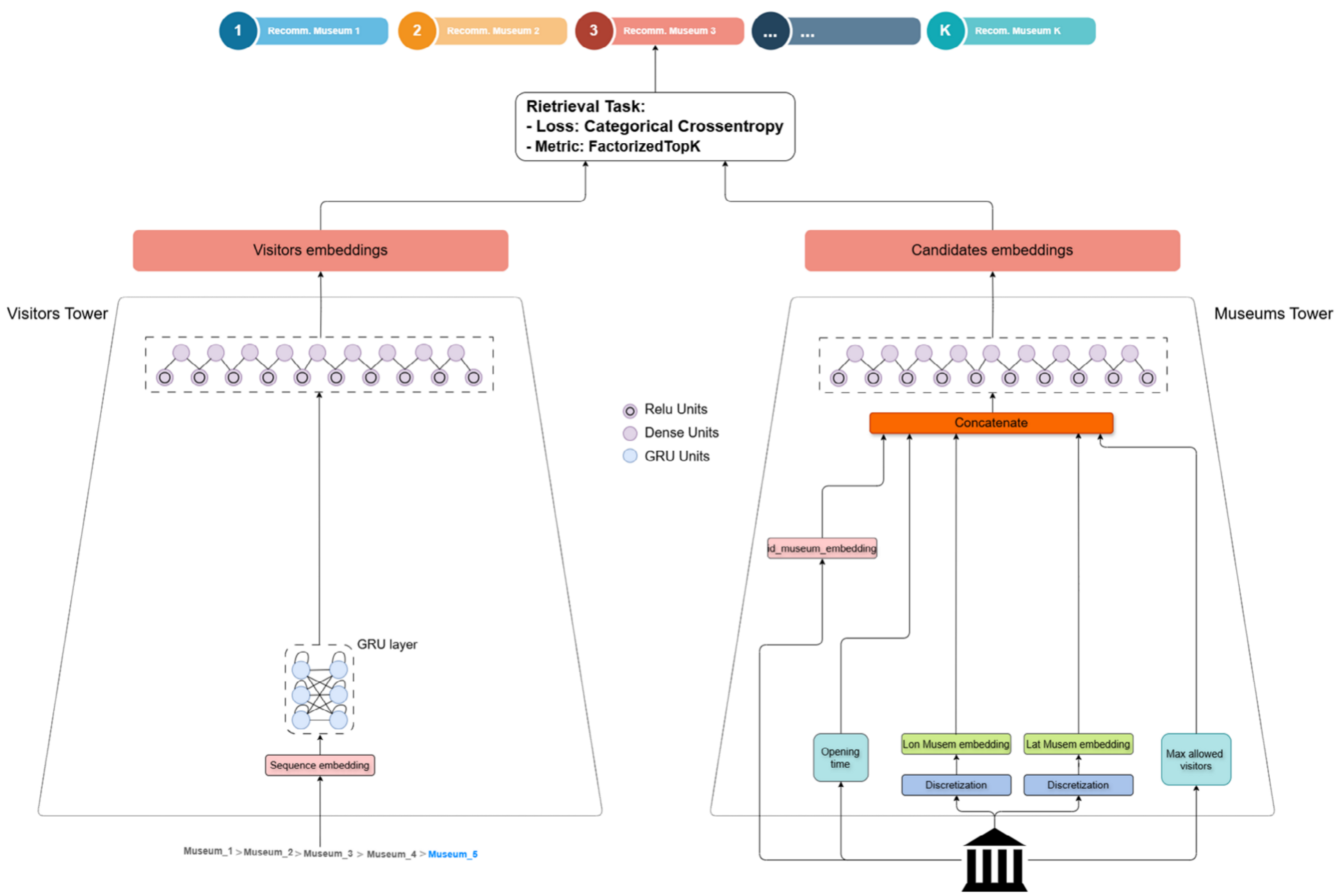

FIGURE 3 | The baseline model with a GRU network on sequence data.

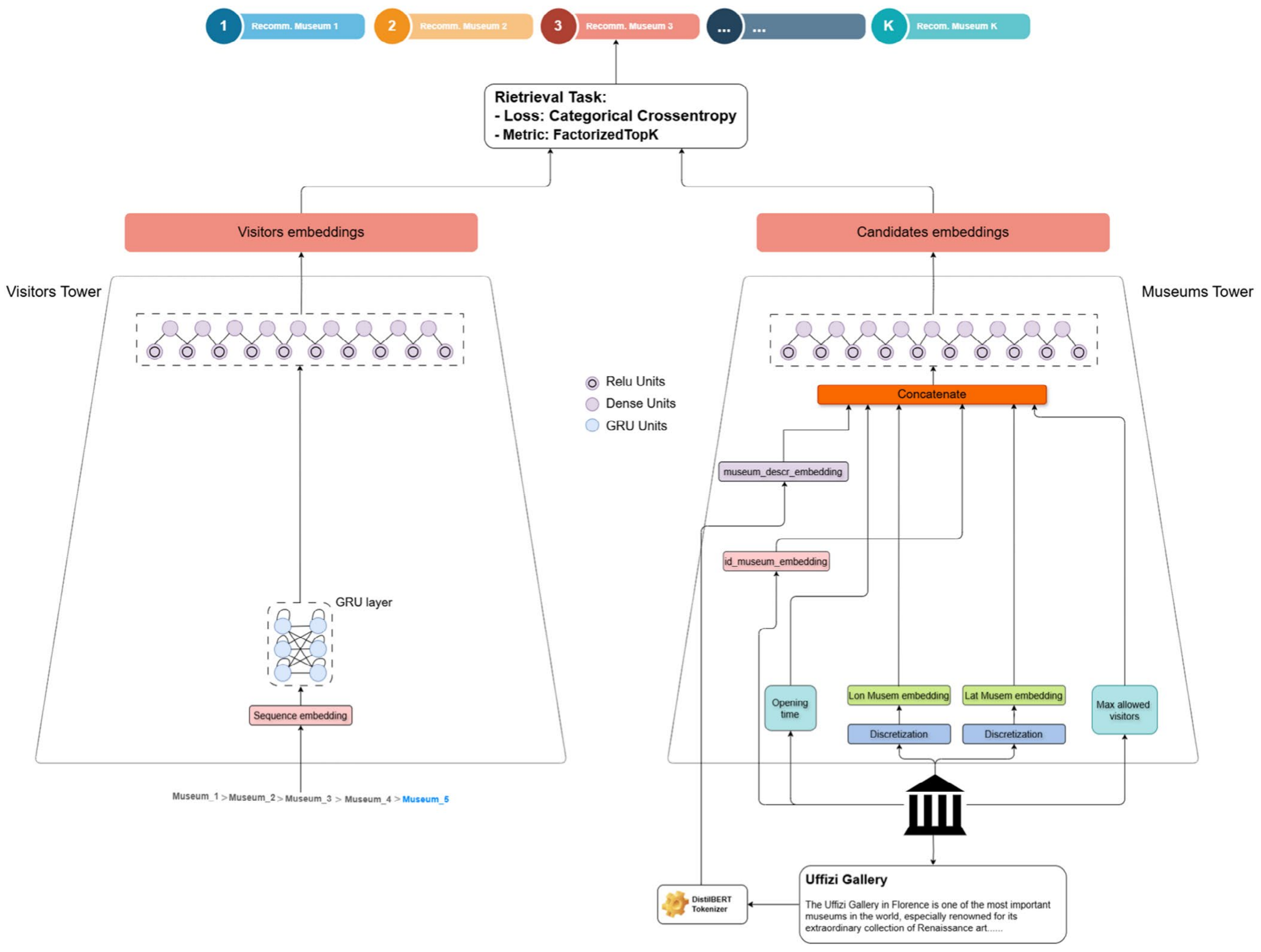

FIGURE 4 | The RS with the museum description with DistilBERT tokeniser.

dense vector space. Moreover, numerical features, such as the opening time, the maximum allowed number of visitors and the geographic coordinates (latitude and longitude), are discretised and incorporated as categorical features. As in the visitor tower, a stack of two fully connected layers, with the same structure: 32 units with ReLU activation followed by 16 linear units, is used to project the input into the same 16-dimensional embedding space. Also in this case, the number of units composing the layers are hyperparameters empirically selected.

3.3.2 | Adding Description of Museum Contents

To improve the predictive performance of the RS, the baseline model described above is enriched by adding the textual description of museums in the museum tower. The textual description of each museum content is encoded using DistilBERT (Sanh et al. 2019) tokeniser library, which is a pre-trained Large Language Model. This procedure projects textual descriptions into a latent space, enabling the computation of similarities between museum contents based on vector distances. In this space, museums with similar descriptions and content will be close, while museums with different descriptions and content will be far apart. The architecture of the two-tower model is updated as illustrated in Figure 4.

3.3.3 | Final Model

The final version of the proposed RS introduces all the features derived from the data augmentation process. The final two-tower RS architecture is illustrated in Figure 5. To the visitor tower, we add the visitor's latent class (Section 3.2.3) and the order and rank of visited museums (synthesised through vector $\Psi(s)$, as detailed in Section 3.2.4); to the museum tower, we add the daily total number of visits per museum, the average visit duration per museum (Section 3.2.2), the probability of museum visit (Section 3.2.3). Moreover, visitor and museum locations are considered in the model through cross-networks, as detailed below.

Treating latitude and longitude as independent numerical inputs fails to capture their combined spatial significance. For example, two museums might share similar latitudes but differ substantially in longitude, meaning they are not geographically close. Accurately modelling spatial proximity requires considering the interaction of both dimensions simultaneously. To address this, our final model leverages cross-networks (Wang et al. 2017) to explicitly learn interactions between latitude and longitude, effectively capturing their joint contribution to spatial proximity. Unlike conventional deep neural layers that implicitly attempt to learn such

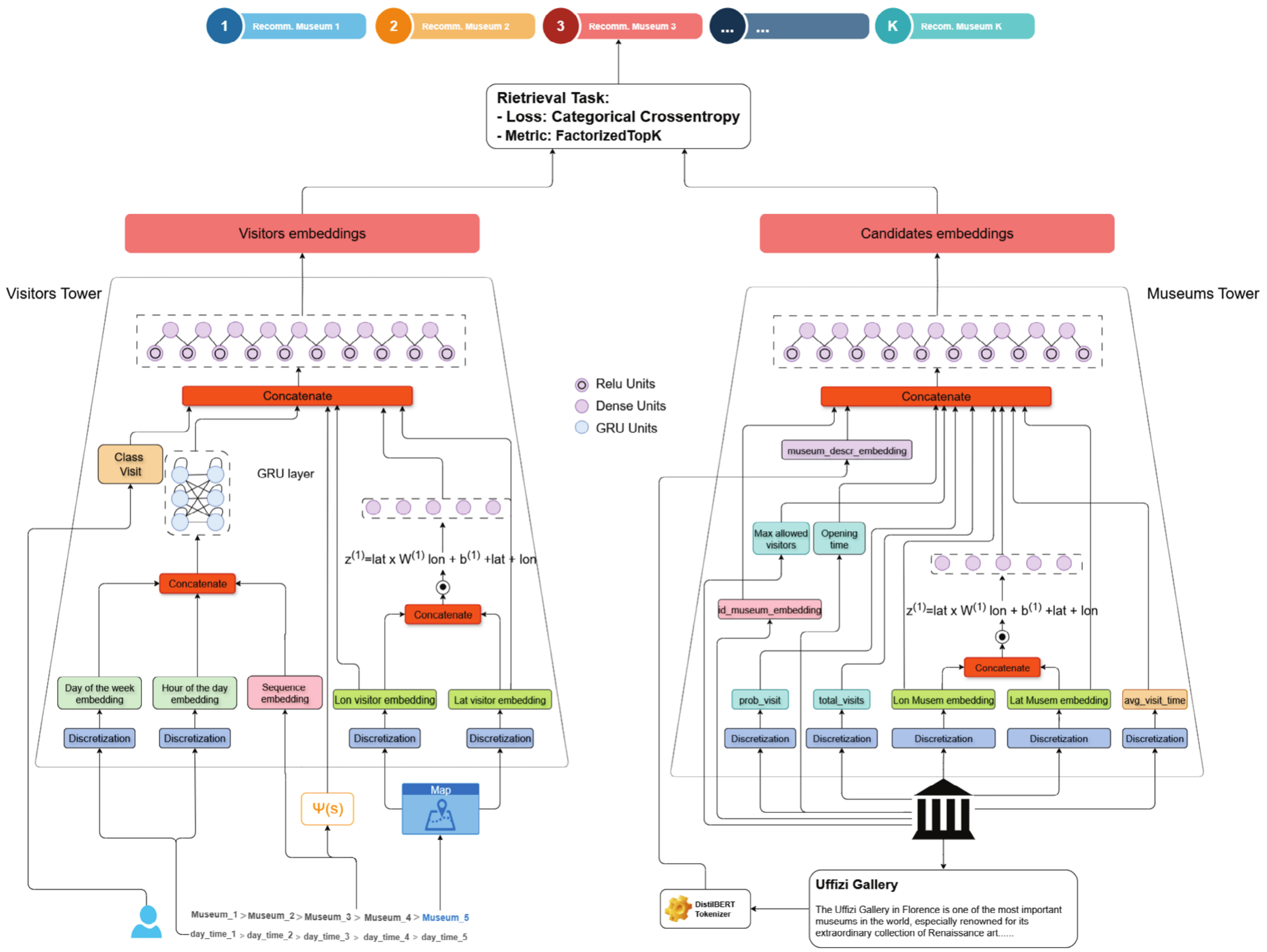

FIGURE 5 | The final two-tower system architecture.

feature interactions, which often require deep and resource-intensive architectures, cross-networks provide a more targeted and parameter-efficient approach by directly modelling polynomial feature interactions from raw inputs. This mechanism allows the system to learn meaningful spatial patterns efficiently without unnecessary model complexity.

Specifically, we apply cross-networks to the spatial features of both users and museums, that is, their latitude and longitude coordinates, ensuring that the model accurately accounts for the spatial dynamics that shape tourist itineraries in dense urban heritage areas. In detail, given two input vectors x and y, the cross-product operation can be represented as

$$z = x \bigodot y,$$

where $\bigodot$ denotes element-wise multiplication.

In cross-networks, this operation is typically extended to higher-order interactions by incorporating multiple layers. The output of each layer is fed into the next layer, allowing the network to capture increasingly complex interactions. Let $x^{(0)}$ and $y^{(0)}$ denote the input features (e.g., latitude and longitude coordinates) and $z^{(l)}$ the output of the lth cross layer. The computation of the cross-network can be expressed recursively as:

$$z^{(l)} = x^{(l-1)} \bigodot W^{(l)} y^{(l-1)} + b^{(l)} + x^{(l-1)} + y^{(l-1)},$$

where $W^{(l)}$ represents the learnable weight matrix for the lth layer, $b^{(l)}$ is the bias term, $x^{(l-1)}$ and $y^{(l-1)}$ are the input features of the previous layer, and $z^{(l)}$ is the output of the lth layer.

By stacking the cross layer with a nonlinear dense ReLU layer with 16 units, the two-tower RS can capture the interactions between latitude and longitude, considering these values not as independent features but as correlated features, and during training, the model can learn to quantify these correlations.

4 | Findings

In what follows, we provide details on the structure of the engineered and augmented dataset, the setting of parameters to train, validate and test the RS and the performance of the proposed system.

TABLE 1 | Training, validation and test sets: Record structure.

Visitorstower						
Sequence	**ClassVisit**	**Lon vis.**	**Lat vis.**	$\psi(s)$	**Day of visit**	**Entry time**
[39, 2, 38, 1, 8]	3	11.259597	43.781967	ψ([39, 2, 38, 1, 8])	[5, 5, 5, 6, 6]	[9:32:15, 11:37:02, ...]
[38, 25, 15, 39, 12]	1	11.245464	43.768742	ψ([38, 25, 15, 39, 12])	[4, 4, 5, 5, 5]	[12:02:27, 15:24:53, ...]
[39, 4, 5, 3, 7]	1	11.239416	43.766966	ψ([39, 4, 5, 3, 7])	[2, 3, 3, 3, 3]	[14:15:41, 09:32:35, ...]
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Museums tower								
Sequence	**nextVisMus**	**Lon Mus.**	**Lat Mus.**	**Avg visit time**	**Total visits**	**Prob visit**	**Description**	**Max numb visitors**
[39, 2, 38, 1, 8]	4	11.254430	43.774932	65.673022	47,453	0.379	[..., ..., ..., ..., ...]	[400, 250, 200, 800, 100]
[38, 25, 15, 39, 12]	37	11.250022	43.765207	214.671528	68,607	0.548	[..., ..., ..., ..., ...]	[200, 200, 100, 400, 300]
[39, 4, 5, 3, 7]	37	11.250022	43.765207	214.671528	68,607	0.548	[..., ..., ..., ..., ...]	[400, 1000, 200, 1500, 30]
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

4.1 | Data Structure

The feature engineering and data augmentation result illustrated in Section 3.2 is a dataset whose structure is shown in Table 1.

Each record represents a FirenzeCard holder to whom visitor and museum features are associated. In particular, on the side of the visitor tower, each FirenzeCard holder is associated with the sequence of visited museums (sequence), the latent class with the highest posterior membership probability (ClassVisit), the assessed position in terms of longitude (Lon vis.) and latitude (Lat vis.), the $\psi(s)$ function, the weekday of visit (Day of visit) and the entry hour (Entry time). On the side of the museum tower, for each cardholder, the Kth museum visited after the first $K-1$ museums (nextVisMus) is identified with its geographical position (Lon Mus. and Lat Mus.), assessed average time of visit (avg visit time), total number of visits recorded by the FirenzeCards (total_visits), probability of being visited estimated through the LC-IRT model (prob_visit), textual description of the artworks exposed (Description) and its maximum capacity (Max numb visitors).

4.2 | Hyperparameter Setting

The construction of the two-tower model requires the definition of certain elements before the training, validation and testing phases, in particular: the minimum length of the sequences to be considered for training, the type of loss function and the number of epochs (see below for details).

The effectiveness of our RS depends on its ability to predict the next museum a tourist will visit based on their history of visits. This capability is closely related to defining effective training and testing datasets. To do this, selecting appropriate sequence lengths between all the available sequences in the FirenzeCard database is crucial. Sequences must be long enough to capture meaningful patterns but not so long as to reduce the dataset size excessively. As outlined in Section 3.1, sequences in the FirenzeCard database show an average of around 7 visits, so we use sequences of at least 6 museums for training and testing the model: the first five museums represent the history of visits, whose information will serve to predict the next museum in the sequence. Thus, after discarding the shorter sequences (i.e., with less than 6 recorded entries), the result is a dataset of 91,650 sequences. We then split such a dataset, allocating 90% of the records for training purposes (i.e., the training dataset consisting of 83,022 sequences), 5% for validation purposes (i.e., the validation dataset consisting of 4583 sequences) and 5% for testing purposes (i.e., the test dataset consisting of 4582 sequences).

The model is trained with a supervised learning approach employing categorical cross-entropy as a loss function. Such a function, suitable for multi-class classification tasks with more than two mutually exclusive classes (Mao et al. 2023), quantifies the dissimilarity between the predicted and the true categorical labels (i.e., the next museum actually visited).

Another element that has to be set for training the model is the number of epochs. An epoch refers to one complete pass of the entire dataset through the neural network. During each epoch, the model processes all data, generates predictions, calculates the loss and updates its parameters accordingly. Multiple epochs may be required to achieve effective learning depending on the model complexity and the dataset size. While more epochs help the model refine its understanding, too many can lead to overfitting, where the model learns the training data too well and fails to generalise to fresh novel data. An early stopping technique is applied to avoid this, which monitors the validation loss and halts training when no improvement is observed for 5 consecutive epochs. In our case, we choose to train the model for a maximum of 40 epochs. This upper limit is set on the basis of empirical observations indicating that performance typically plateaus well before reaching 20 epochs. By doubling this threshold as a precaution, we ensure the model has ample opportunity to improve, while maintaining a safeguard against unnecessary training.

4.3 | Predictive Performance

The model performance is evaluated using the top-K metric (Jagerman et al. 2022), with K representing the size of the recommended set of museums. We choose K values of 1, 3 and 5, meaning the system will suggest one, three, or five potential next museums to visit, respectively. As exposed in Section 3.2, the evaluation of the final model is also conducted by varying the radius R for sampling visitor locations (at the time of the recommendation request) from 500 to 2000 m around the last visited museum, with incremental steps of 300 m.

Table 2 summarises the model performance on the test dataset, evaluated across the defined K values for the top-K metrics and different configurations of the two-tower model: baseline model (Figure 3), baseline model with museum textual descriptions (Figure 4) and final model (Figure 5) with additional features deriving from the engineering process described in Section 3.2. Moreover, we evaluate the performance of the final model for different radius sizes of the sampling space from which the location of the visitor requesting the recommendation is sampled. For each combination of model and K value, Table 2 reports the proportion (probability) with which the recommendation matches the next visited museum in the history of visits (i.e., the one visited after the first five in each sequence).

TABLE 2 | Model performance on the test set, evaluated across the defined K values for the top-K metric and different visitor location sampling radius.

Model	Top-1	Top-3	Top-5
Baseline model	0.11	0.21	0.31
Baseline model with museum description	0.11	0.25	0.35
Final model with 2000 m sampling space radius	0.11	0.25	0.35
Final model with 1700 m sampling space radius	0.13	0.30	0.42
Final model with 1400 m sampling space radius	0.13	0.40	0.54
Final model with 1100 m sampling space radius	0.14	0.43	0.58
Final model with 800 m sampling space radius	0.21	0.51	0.68
Final model with 500 m sampling space radius	0.30	0.68	0.86

As expected, the effectiveness of the recommendation improves as K increases. Furthermore, we observe that feature engineering and feature augmentation activities, along with their proper integration into the two-tower model, positively impact system performance. Table 2 clearly shows a significant improvement in predictive performance when moving from the baseline model to the final model and, for the final model, performance improves further when the sampling radius for estimating the visitor's location is reduced: namely, the probability of correct predictions increases from 0.11 to 0.31 for the baseline model up to 0.30–0.86 for the final model when we consider a 500 m distance around the last visited museum.

As shown in Table 2, the textual descriptions of museum content are not particularly relevant for predictive purposes. A possible explanation for this finding is that the visit sequences provided as input to the system already implicitly capture visitors' preferences. Another factor that could play a significant role in explaining this result is that visit choices are only partially influenced by the museums' content. In contrast, other factors, such as the time available and the need to optimise visit duration imposed by the pass, may have a more substantial impact. This finding is particularly evident when analysing the performance of the final model, which explicitly incorporates the visitor's location when they choose the next museum to visit.

Note that the maximum probability of correctly predicting the next museum visited is 86%, achieved when the location of the FirenzeCard holder (at the time of the hypothetical recommendation request) is within a radius of 500 m from the last visited museum and the system suggests five museums. An 86% success rate is a value considerably greater than the nominal probability that could be calculated if the five recommended museums were selected randomly, which is approximately equal to 14.71%. Remaining within a 500 m radius from the last visited museum, with $K=3$ and $K=1$, the nominal probabilities decrease to approximately 8.82% and 2.94%, respectively. Thus, the designed RS always performs better than chance, and such performance significantly increases as the radius from the last visited museum decreases.

5 | Discussion and Conclusions

5.1 | Discussion

The phenomenon of overtourism has emerged as one of the most pressing negative consequences of contemporary tourism. Overtourism arises when the volume of visitors at a specific tourist destination exceeds its carrying capacity, leading to overcrowding, excessive commercialisation, hostification (i.e., the displacement of residents), rising prices, inadequate infrastructure (e.g., public transport) and the erosion of local traditions. Collectively, these issues contribute to the social, economic and environmental degradation of the tourist sites.

In light of these issues, there is a growing need for agile and non-intrusive tools to support policymakers in managing tourist flows more sustainably. Recommender systems (RSs) represent a promising solution, offering the potential to subtly guide tourist behaviour through personalised suggestions, rather than through restrictive or coercive measures.

This study introduces a hybrid RS grounded in a two-tower neural network architecture, designed to predict and influence museum visitation patterns in Florence. The model was developed and validated using a rich dataset collected in 2018 via the FirenzeCard museum pass, including visit sequences and entry times from over 125,000 tourists. The primary aim of the system is to forecast museum visits by analysing individual visitor histories alongside similar users' preferences, while also accounting for museum-specific features such as thematic content and geographical location. As shown in Section 4.3, the proposed model exhibits a strong predictive performance in forecasting the next top-K museums a tourist will visit, given the museums he/she has already visited.

5.2 | Theoretical Contribution

This study contributes to the tourism literature by introducing a novel application of two-tower neural network architecture, an approach that, to date, has seen limited use within this domain. While two-tower models have been widely adopted in other areas such as e-commerce and media recommendation, their potential in tourism settings, particularly for managing visitor flows and enhancing cultural heritage experiences, remains largely untapped.

Our model offers a genuine Artificial Intelligence framework, instead of more traditional machine learning approaches that rely on explicit user profiles. Indeed, in many tourism contexts, including ours, detailed user data is unavailable due to privacy constraints or the transient nature of tourists. Our system operates instead on implicit preference signals, inferred from observed sequences of museum visits and contextual features such as geolocation. This allows us to generate highly personalised recommendations even in the absence of direct user input or registration, thereby increasing the model's applicability to real-world tourism settings.

At the core of our architecture, the user tower integrates a sequential-based deep learning component, specifically Gated Recurrent Units (GRUs), to capture the time-dependent and ordered nature of tourist behaviour. This design choice addresses key limitations of classical methods such as Markov models, which often suffer from cold-start issues and insufficient robustness in scenarios constrained by opening hours and capacity limits. GRUs provide a stable and parsimonious modelling framework, well suited to the medium-short sequences characteristic of museum visits compressed within a 72-h tourist pass.

On the item side, we adopt a multimodal representation that includes textual museum descriptions alongside structured metadata (e.g., typology and location). Furthermore, we incorporate spatial correlations through cross-networks, which explicitly model the interactions between latitude and longitude. Unlike traditional deep layers that implicitly attempt to learn such spatial patterns, cross-networks provide a more efficient and interpretable mechanism for capturing location-based relationships, enhancing the model's ability to recommend nearby and thematically relevant attractions.

This paper proposes a modular and adaptable RS architecture that is highly replicable across similar tourism contexts. The system's flexibility allows for the seamless integration of additional data sources within the two-tower framework. At the same time, its dynamic nature ensures that recommendations evolve in real time as a visitor's itinerary unfolds. Given that the input data is typically collected in many cultural tourism settings (e.g., through city passes or ticketing systems), our approach offers a scalable foundation for developing sustainable, intelligent, data-driven visitor management tools.

5.3 | Practical Implications

The two-tower recommendation model developed in this study constitutes a foundational component of a broader decision-support system, potentially governed by Reinforcement Learning principles. Currently, the system excels at learning from historical data to predict tourist behaviour with a high degree of accuracy. This retrieval capability is not merely an academic exercise. Still, it has critical practical value when embedded into real-world applications, such as a mobile app, where it can support dynamic, preference-driven itinerary suggestions supervised by the guidance of local policymakers.

For performance evaluation purposes we test the RS using the top-K metric where K represents the size of the recommended set of museums, after a training based on a history of five museums previously visited. However, the underlying two-tower model in our approach is also capable of addressing cold-start scenarios, enabling it to recommend new items even in the absence of prior user interaction history.

Importantly, the model can capture latent patterns and constraints not explicitly encoded in the dataset but inferred from the sequences of observed behaviours. These include factors such as preferred days and times for museum visits, the implicit avoidance of certain time slots despite availability, individual interests and preferences for specific types of exhibitions and typical visit durations. Such features position the model as a sophisticated behavioural predictor, capable of informing real-time decision-making in complex urban tourism environments.

In practical deployment, the RS could be integrated into a mobile app that offers personalised suggestions while respecting system-level constraints. For instance, the app could avoid recommending visits to overcrowded museums at peak times, even if these align with the user's preferences and instead highlight alternative options that match inferred interests while alleviating pressure on the city's most visited attractions. This would enable tourism authorities to re-shape visitor flows sustainably, protecting fragile heritage sites and enhancing the overall tourist experience. In doing so, the RS would serve as a powerful complement to existing initiatives of the Management Plan of the Florence Municipality (De Luca et al. 2020; Liberatore et al. 2023), such as the FirenzeCard project.

5.4 | Limits and Future Research

This study presents a promising step towards intelligent, data-driven tourist flow management, yet some limitations must be acknowledged, suggesting avenues for future research and model refinement.

First, the system has been trained and validated on a dataset with limited details regarding visitor behaviour. Precise information on when each museum visit is completed and the visitor's geolocation at the time of their next decision is not available. Consequently, certain behavioural features had to be estimated, such as the time elapsed between visits or the user's position in the city. However, integrating the RS into a dedicated mobile application offers a clear path forward. A mobile app would allow for the automatic acquisition of real-time geolocation data, access to up-to-date information on museum queues and opening hours and potentially the collection of optional socio-demographic and preference-based information at registration. This would greatly enrich the dataset, transforming the RS into a more adaptive and context-aware system.

Second, the current system functions as a retrieval model: that is, it is designed to identify a set of museums likely to interest a user, but it does not prioritise these suggestions. A natural next step would be to extend the model into a ranking-based framework, capable of producing an ordered list of recommendations according to predefined optimisation criteria. These could include real-time variables such as queue lengths or museum capacity, and user-centric features such as proximity, thematic interest and predicted satisfaction. This passage from a retrieval model to a ranking model necessitates developing and adopting new evaluation metrics tailored to ranking performance.

Third, the two-tower architecture itself offers fertile ground for future exploration. In this study, the user tower is built upon a sequential-based model using GRUs, particularly suitable for the relatively short and dense visit sequences typical in cultural tourism contexts. However, alternative deep learning structures could enhance the model predictive accuracy. Introducing attention mechanisms (Niu et al. 2021) could allow the system to focus more selectively on the most informative parts of a visitor's behavioural sequence, potentially outperforming GRUs in scenarios with more complex or varied user patterns. Future work could empirically assess the impact of such architectural modifications, including comparative analyses between recurrent and attention-based models.

In conclusion, while the current model represents a significant advancement in the application of AI to tourism management, its full potential will be realised through its deployment in live systems and its continuous improvement via user interaction and technical innovation. These developments lay the groundwork for more intelligent, sustainable and user-centred approaches to cultural tourism governance.

Acknowledgements

Authors acknowledge the support provided by the Italian Ministry of University and Research through 'Dipartimenti Eccellenti 2023-2027' project. Authors thank Silfi spa, the company to which the Municipality of Florence has assigned the management of the FirenzeCard project, for having making available the data. Open access publishing facilitated by Universita degli Studi di Firenze, as part of the Wiley - CRUI-CARE agreement.

Funding

The authors have nothing to report.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data used in the article are the property of Silfi Spa, the company to which the Municipality of Florence has assigned the management of the FirenzeCard project.

References

Abbasi-Moud, Z., H. Vahdat-Nejad, and J. Sadri. 2021. "Tourism Recommendation System Based on Semantic Clustering and Sentiment Analysis." *Expert Systems with Applications* 167: 114324. https://doi.org/10.1016/j.eswa.2020.114324.

Al Fararni, K., F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi, and A. Sabri. 2021. "Hybrid Recommender System for Tourism Based on Big Data and AI: A Conceptual Framework." *Big Data Mining and Analytics* 4, no. 1: 47–55. https://doi.org/10.26599/BDMA.2020.9020015.

Bartolucci, F. 2007. "A Class of Multidimensional IRT Models for Testing Unidimensionality and Clustering Items." *Psychometrika* 72: 141–157. https://doi.org/10.1007/s11336-005-1376-9.

Bartolucci, F., S. Bacci, and M. Gnaldi. 2015. *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata*. Chapman & Hall/CRC Press.

Bastidas-Manzano, A. B., J. Sánchez-Fernández, and L. A. Casado-Aranda. 2021. "The Past, Present, and Future of Smart Tourism Destinations: A Bibliometric Analysis." *Journal of Hospitality and Tourism Research* 45: 529–552. https://doi.org/10.1177/1096348020967062.

Borràs, J., A. Moreno, and A. Valls. 2014. "Intelligent Tourism Recommender Systems: A Survey." *Expert Systems with Applications* 41: 7370–7389. https://doi.org/10.1016/j.eswa.2014.06.007.

Burke, R. 2002. "Hybrid Recommender Systems: Survey and Experiments." *User Modeling and User-Adapted Interaction* 12: 331–370. https://doi.org/10.1023/A:1021240730564.

Cho, K., B. van Merrienboer, C. Gulcehre, et al. 2014. "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine

Translation." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, edited by A. Moschitti, B. Pang, and W. Daelemans, 1724–1734. ACL. https://doi.org/10.3115/v1/D14-1179.

Coca-Stefaniak, A., A. Morrison, D. Edwards, et al. 2016. "Editorial." *International Journal of Tourism Cities* 2: 273–280. https://doi.org/10.1108/IJTC-10-2016-0041.

Colomo-Palacios, R., F. J. García-Peñalvo, V. Stantchev, and S. Misra. 2016. "Towards a Social and Context-Aware Mobile Recommendation System for Tourism." *Pervasive and Mobile Computing* 38, no. 2: 505–515. https://doi.org/10.1016/j.pmcj.2016.03.001.

Covington, P., J. Adams, and E. Sargin. 2017. "Deep Neural Networks for YouTube Recommendations." In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys'16)*, 191–198. Association for Computing Machinery. https://doi.org/10.1145/2959100.2959190.

Cui, Y., S. Liang, and Y. Y. Zhang. 2024. "Multimodal Representation Learning for Tourism Recommendation With Two-Tower Architecture." *PLoS One* 19, no. 2: e0299370. https://doi.org/10.1371/journal.pone.0299370.

De Luca, G., A. Shirvani Dastgerdi, C. Francini, and G. Liberatore. 2020. "Sustainable Cultural Heritage Planning and Management of Overtourism in Art Cities: Lessons From Atlas World Heritage." *Sustainability* 12, no. 9: 3929. https://doi.org/10.3390/su12093929.

Dodds, R., and R. Butler. 2019. "The Phenomena of Overtourism: A Review." *International Journal of Tourism Cities* 5, no. 4: 519–528. https://doi.org/10.1108/IJTC-06-2019-0090.

Feng, S., X. Li, Y. Zeng, G. Cong, Y. M. Chee, and Q. Yuan. 2015. "Personalized Ranking Metric Embedding for Next New POI Recommendation." In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, edited by Q. Yang and M. Wooldridge, 965–976. AAAI Press. https://doi.org/10.5555/2832415.2832536.

Forster, J. 1964. "The Sociological Consequences of Tourism." *International Journal of Comparative Sociology* 5: 217–227.

Gavalas, D., C. Konstantopoulos, K. Mastakas, and G. Pantziou. 2014. "Mobile Recommender Systems in Tourism." *Journal of Network and Computer Applications* 39: 319–333. https://doi.org/10.1016/j.jnca.2013.04.006.

Gretzel, U., M. Sigala, Z. Xiang, and C. Koo. 2015. "Smart Tourism: Foundations and Developments." *Electronic Markets* 25, no. 3: 179–188. https://doi.org/10.1007/s12525-015-0196-8.

Hamid, R. A., A. Albahri, J. K. Alwan, et al. 2021. "How Smart Is e-Tourism? A Systematic Review of Smart Tourism Recommendation System Applying Data Management." *Computer Science Review* 39: 100337. https://doi.org/10.1016/j.cosrev.2020.100337.

Heinen, T. 1996. *Latent Class and Discrete Latent Traits Models: Similarities and Differences*. Sage.

Hochreiter, S., and J. Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9: 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hollands, R. G. 2015. "Critical Interventions Into the Corporate Smart City." *Cambridge Journal of Regions, Economy and Society* 8, no. 1: 61–77. https://doi.org/10.1093/cjres/rsu011.

Hospers, G. J. 2019. "Overtourism in European Cities: From Challenges to Coping Strategies." *CESifo Forum* 20, no. 3: 20–24.

Huang, P.-S., X. He, J. Gao, L. Deng, A. Acero, and L. Heck. 2013. "Learning Deep Structured Semantic Models for Web Search Using Clickthrough Data." In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management (CIKM'13)*, 2333–2338. Association for Computing Machinery. https://doi.org/10.1145/2505515.2505665.

Jagerman, R., Z. Qin, X. Wang, M. Bendersky, and M. Najork. 2022. "On Optimizing Top-k Metrics for Neural Ranking Models." In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'22)*, 2303–2307. Association for Computing Machinery. https://doi.org/10.1145/3477495.3531849.

Karpukhin, V., B. Oguz, S. Min, et al. 2020. "Dense Passage Retrieval for Open-Domain Question Answering." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, edited by B. Webber, T. Cohn, Y. He, and Y. Liu, 6769–6781. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.550.

Koens, K., A. Postma, and B. Papp. 2018. "Is Overtourism Overused? Understanding the Impact of Tourism in a City Context." *Sustainability* 10, no. 12: 4384. https://doi.org/10.3390/su10124384.

Kurashima, T., T. Iwata, G. Irie, and K. Fujimura. 2013. "Travel Route Recommendation Using Geotagged Photos." *Knowledge and Information Systems* 37, no. 1: 37–60. https://doi.org/10.1007/s10115-012-0580-z.

Langeheine, R., and J. Rost. 1988. *Latent Trait and Latent Class Models*. Plenum Press. https://doi.org/10.1007/978-1-4757-5644-9.

Liberatore, G., P. Biagioni, C. Ciappei, and C. Francini. 2023. "Dealing With Uncertainty, From Overtourism to Overcapacity: A Decision Support Model for Art Cities: The Case of UNESCO WHCC of Florence." *Current Issues in Tourism* 26: 1067–1081. https://doi.org/10.1080/13683500.2022.2046712.

Liu, Q., S. Wu, L. Wang, and T. Tan. 2016. "Predicting the Next Location: A Recurrent Model With Spatial and Temporal Contexts." In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16)*, 194–200. AAAI Press. https://doi.org/10.1609/aaai.v30i1.9971.

Liu, Y., E. Jun, Q. Li, and J. Heer. 2019. "Latent Space Cartography: Visual Analysis of Vector Space Embeddings." *Computer Graphics Forum* 38: 67–78. https://doi.org/10.1111/cgf.13672.

Lucas, J. P., N. Luz, M. N. Moreno, R. Anacleto, A. Almeida Figueiredo, and C. Martins. 2013. "A Hybrid Recommendation Approach for a Tourism System." *Expert Systems with Applications* 40: 3532–3550. https://doi.org/10.1016/j.eswa.2012.12.061.

Malkov, Y. A., and D. A. Yashunin. 2020. "Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, no. 4: 824–836. https://doi.org/10.1109/TPAMI.2018.2889473.

Mao, A., M. Mohri, and Y. Zhong. 2023. "Cross-Entropy Loss Functions: Theoretical Analysis and Applications." arXiv:2304.07288. https://arxiv.org/abs/2304.07288.

Masini, S., S. Bacci, A. Petrucci, C. Francini, and B. Bertaccini. 2025. "Understanding Intra-Destination Tourist Behaviors as a Tentative Strategy to Mitigate Overtourism in Florence." *Italian Journal of Applied Statistics*. https://doi.org/10.26398/IJAS.16199.

Milano, C., J. M. Cheer, and M. Novelli. 2019. *Overtourism: Excesses, Discontents and Measures in Travel and Tourism*. CABI. https://doi.org/10.1079/9781786399823.0000.

Nan, X., K. Kanato, and X. Wang. 2022. "Design and Implementation of a Personalized Tourism Recommendation System Based on the Data Mining and Collaborative Filtering Algorithm." *Computational Intelligence and Neuroscience* 2022: 1424097. https://doi.org/10.1155/2022/1424097.

Niu, Z., G. Zhong, and H. Yu. 2021. "A Review on the Attention Mechanism of Deep Learning." *Neurocomputing* 452: 48–62. https://doi.org/10.1016/j.neucom.2021.03.091.

Pazzani, M. J., and D. Billsus. 2007. "Content-Based Recommendation Systems." In *The Adaptive Web*, edited by P. Brusilovsky, A. Kobsa, and

W. Nejdl, 325–341. Springer. https://doi.org/10.1007/978-3-540-72079-9_10.

Peeters, P., S. Gössling, J. Klijs, et al. 2018. *Research for TRAN Committee-Overtourism: Impact and Possible Policy Responses*. European Parliament, Policy Department for Structural and Cohesion Policies.

Ranjan, C., S. Ebrahimi, and K. Paynabar. 2022. "Sequence Graph Transform (SGT): A Feature Embedding Function for Sequence Data Mining." *Data Mining and Knowledge Discovery* 36: 668–708. https://doi.org/10.1007/s10618-021-00813-0.

Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. 1994. "GroupLens: An Open Architecture for Collaborative Filtering of Netnews." In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 175–186. Association for Computing Machinery. https://doi.org/10.1145/192844.192905.

Ribes, J. F. P., and J. I. Baidal. 2018. "Smart Sustainability: A New Perspective in the Sustainable Tourism Debate." *Investigaciones Regionales – Journal of Regional Research* 42: 151–170.

Ricci, F., L. Rokach, and B. Shapira. 2011. *Recommender Systems Handbook*. Springer-Verlag. https://doi.org/10.1007/978-0-387-85820-3_1.

Rost, J. 1990. "Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis." *Applied Psychological Measurement* 14, no. 3: 271–282. https://doi.org/10.1177/014662169001400305.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. "Learning Internal Representations by Error Propagation." In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, edited by D. E. Rumelhart and J. L. McClelland, 318–362. MIT Press.

Sanh, V., L. Debut, J. Chaumond, and T. Wolf. 2019. "DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter." ArXiv Abs/1910.01108. https://api.semanticscholar.org/CorpusID:203626972.

Sarkar, J., A. Majumder, C. Panigrahi, S. Roy, and B. Pati. 2023. "Tourism Recommendation System: A Survey and Future Research Directions." *Multimedia Tools and Applications* 82: 8983–9027. https://doi.org/10.1007/s11042-022-12167-w.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2001. "Item-Based Collaborative Filtering Recommendation Algorithms." In *Proceedings of the 10th International Conference on World Wide Web (WWW'01)*, 285–295. Association for Computing Machinery. https://doi.org/10.1145/371920.372071.

Solano-Barliza, A., I. Arregocés-Julio, M. Aaròn-Gonzalvez, et al. 2024. "Recommender Systems Applied to the Tourism Industry: A Literature Review." *Cogent Business & Management* 11, no. 1: 2367088. https://doi.org/10.1080/23311975.2024.2367088.

von Davier, M. 2008. "A General Diagnostic Model Applied to Language Testing Data." *British Journal of Mathematical and Statistical Psychology* 61: 287–307. https://doi.org/10.1002/j.2333-8504.2005.tb01993.x.

Wang, R., B. Fu, G. Fu, and M. Wang. 2017. "Deep & Cross Network for Ad Click Predictions." In *Proceedings of the 2017 ADKDD (ADKDD'17)*, 1–7. Association for Computing Machinery. https://doi.org/10.1145/3124749.3124754.

World Tourism Organization (WTO), Centre of Expertise Leisure, Tourism & Hospitality, NHTV Breda University of Applied Sciences, and NHL Stenden University of Applied Sciences. 2019. *"Overtourism"? – Understanding and Managing Urban Tourism Growth Beyond Perceptions, Volume 2: Case Studies*. UNWTO. https://doi.org/10.18111/9789284420629.

Zheng, W., X. Guo, and J. Li. 2025. "Intra-City Tourism Flow Forecasting: A Novel Deep Learning Model." *International Journal of Tourism Research* 27, no. 2: e70011. https://doi.org/10.1002/jtr.70011.