Intelligent Travel Solutions: Merging User Preferences with Real-Time Contextual Awareness

MSc Artificial Intelligence
Practicum

Shalini Priya
Student ID: 23214333

School of Computing
National College of Ireland

Supervisor: SM RAZA ABIDI

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Shalini Priya…………………………………………………………………………………………
Student ID:	23214333 ……………………………………………………………..……
Programme:	…………………………………………Msc AI **Year:** …2024………………..
Module:	…………………Practicum……………………………………………………………………….………
Supervisor:	……………S M RAZA Abidi…………………………………………………………………….………
Submission Due Date:	…………………………29/01/2024………………………………………………………….………
Project Title:	…… **Intelligent Travel Solutions: Merging User Preferences with Real – Time Contextual Awareness** ……………………………….……….
Word Count:	……………6200……………… **Page Count**……………19……………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	………Shalini Priya……………………………………………………………………………
Date:	…………29/01/2024…………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Intelligent Travel Solutions: Merging User Preferences with Real-Time Contextual Awareness

Shalini Priya
23214333
Artificial Intelligence National
College of Ireland

Abstract

The travel industry increasingly leverages technology for personalized user experiences. This study develops a dynamic recommendation system to address limitations in static models that fail to meet diverse traveler needs. Using geotagged social media data, behavioral patterns, and situational factors, the system significantly enhances recommendation accuracy and relevance. The methodology integrates advanced techniques: LightGCN for collaborative filtering, DBSCAN for clustering, and matrix factorization for filtering user-item interactions. Performance validation is conducted using robust metrics, including Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), ensuring reliability. Findings highlight the system's ability to deliver precise, adaptive travel suggestions, improving decision-making and user satisfaction. The hybrid approach transforms traditional travel planning, making it more engaging and responsive to user preferences. By addressing dynamic traveler needs, the model strengthens customer loyalty and provides a competitive edge in the evolving travel market.

Keywords: Travel Recommendation System, Data Clustering (DBSCAN), Collaborative Filtering (Light GCN), Hybrid Model

1 Introduction

1.1 Background

The tourism industry remains the centerpiece of the global economy, contributing approximately 10-12% of global GDP in 2024, according to recent statistics. This indicates an impressive turnaround and significant evolution since the COVID-19 pandemic. The growth of the travel industry has been powered by greater digital transformation, with modern travelers relying largely on platforms like TripAdvisor, Yelp, and Expedia for trip planning. While these platforms allow easy access to reviews and travel data, they usually fail to provide personalized recommendations based on individual tastes, leaving travelers overwhelmed by the vast amount of information available (Kyrylov et al., 2020).

The surge of big data and social media has transformed the travel and tourism business, creating new prospects for personalized trip suggestions. (Kazandzhieva and Santana, 2019). Social media additionally impacts travel preferences. According to studies, platforms like Twitter, Flickr, and Facebook allow individuals to share their travel experiences, which helps cities' reputations as desirable destinations. According to research, Millennials and Generation Z travelers rely extensively on social media for trip preparation, with content created by users taking preference over traditional advertising. Destinations with a solid presence on the internet attract more tourists, as positive social media reviews have a huge influence on customer decisions. Examining geotagged social media data can help governments and tourism boards promote certain tourist destinations. The travel industry is one of the most data-driven industries in the world, but the sheer volume of available data can be intimidating for individuals. Examining geotagged social media data can help governments and tourism boards promote certain tourist destinations.

The travel industry is one of the most data-driven industries in the world, but the sheer volume of available data can be intimidating for individuals(Marr, 2021). Travelers frequently find it difficult to choose the best place and plan thorough itineraries because they are unfamiliar with available attractions and practical problems. To solve these issues, researchers created a trip recommendation.

Systems that combine user preferences, geotagged data, and contextual elements to deliver personalized recommendations.

1.2 Motivation

The modern traveler wants experiences that are not only enjoyable but also customized to their preferences and situational demands. Despite improvements in technology and data analytics, current travel recommendation systems frequently fall short. These algorithms usually promote well-known landmarks or often-visited locations, ignoring the special and new demands of individual travelers. For example, a visitor looking for off-the-beaten-path sites or specialist activities such as hiking, local cuisine, or cultural activities may find the generic recommendations provided by these systems lacking.

One key disadvantage of the current setup is its static nature. They are unable to adapt dynamically to real-time conditions including weather, crowd density, and local events. This lack of adaptation highlights an important flaw in the capability of current travel recommendations. This lack of adaptation indicates an important flaw in the ability of current travel recommendation models to provide spatially relevant recommendations.

Geotagged data from social media platforms such as Instagram and Flickr provide a transformative opportunity to address these difficulties. (Dom®nech et al., 2020) These platforms provide plenty of metadata, such as location, timestamps, and user-generated content, which may be used to get sights about user behavior, preferences, and trends.

By analyzing such data, algorithms can find unique trends and recommend destinations that are relevant to the user’s interests. Furthermore, incorporating real-time factors like weather, local festivals, and attraction operational status might improve recommendation accuracy and usability. The combination of geotagged data, social media analytics, and smart algorithms has immense potential to transform travel experiences. Instead of providing static, one-size-fits-all recommendations, travel recommendation systems can evolve to be dynamic, personalized, and context-aware. Such an evolution in thinking has the potential to improve travel experiences, simplify trip planning, and increase satisfaction for customers.

By investigating the integration of geotagged data, user behavior analysis, and real-time situational elements, the present research aims to close the current gap in trip recommendation systems and create a strong, versatile model that revolutionizes traveler planning.

1.3 Key Contribution

This project transforms travel planning using AI technologies like Machine Learning, Natural Language Processing, and real-time data analysis. It integrates user preferences with live contextual inputs, including weather, and resource availability, to deliver highly personalized and adaptive travel recommendations through an advanced, dynamic AI-driven framework.

This research advances the field of personalized travel recommendation systems by integrating three critical components for a dynamic, user-centric approach:

- Geotagged Data Integration: Leveraging Flickr metadata and DBSCAN, location clusters are identified and enriched with Yelp business data, delivering context-aware recommendations tailored to user interests.

- User Preferences and Real-Time Data: The system combines historical user behaviors with real-time Yelp reviews and ratings, ensuring recommendations are personalized, relevant, and up-to-date.
- Weather-Aware Recommendations: Incorporating real-time weather data, the model dynamically adjusts suggestions based on environmental conditions, aligning with user preferences and enhancing practicality.

This approach addresses the limitations of static systems, enabling dynamic itinerary adaptation, seamless booking options, and efficient decision-making, transforming traditional travel experiences into personalized, timely, and engaging journeys.

The rest of this article is arranged as follows. **Section 2** describes the past studies and literature review. In **Section 3**, the methodologies, design, and implementation used in this paper are given. In **Section 4,** we present the experiment results, evaluation **Section 5**, Experiments show improved accuracy and user satisfaction in recommendations, and finally, we leave the reader with concluding thoughts, limitations, and future works in **Section 6**.

2 Related Work

Deep research into travel recommendation systems has shown their importance for enhancing user experiences. Numerous articles investigate the integration of recommendation systems and personal assistants, with a focus on personalized, context-aware travel advice. These improvements seek to improvedecision-making by adapting recommendations to individual preferences and behaviors.

2.1 Recommendation System

During the 1990s, research in recommendation models advanced toward predicting product ratings, lay- Ing the groundwork for systems that could suggest relevant goods and services based on user data.As highlighted by (Gentsch, 2018), this evolution integrated social filtering and content-based methods. However, the approach often overlooked user-specific attributes, limiting the potential for deeper personalization in an increasingly data-driven era. According to (Aggarwal, 2016), recommendation systems use collected information to provide personalized choices for goods and services. Amazon’s customized purchasing recommendations, YouTube’s customized video suggestions based on user preferences, and Facebook’s tools for increasing social connections are all notable examples. Information technology has become essential for people and organizations to navigate and extract value from complicated datasetsin the digital age.

2.1.1 Types of Recommendation Systems

Recommendation systems have been classified into four categories: content-based, collaborative-based, knowledge-based, and hybrid (Gupta, 2020). Content-based filtering recommends products by analyzing their properties and connecting them to user profiles via keywords. This approach works at connecting user qualities to items that match their tastes, but it is heavily dependent on a well-maintained knowledgelibrary.

Contrary to that, collaborative filtering predicts user preferences based on the interests and behavior of others who have similar patterns(Khalifeh and A. Al-Mousa, 2021). It is further divided into two types: user-based collaborative filtering, which identifies similar users, and item-based collaborative filtering, which recommends products similar to those a user has previously enjoyed. However, collaborative filtering struggles with limited data and new-user issues.

Knowledge-based algorithms are very good in recommending typically purchased products like pensive goods or real estate. These systems match specifications to user desires, but they can be time-consuming and expensive because of the necessity for a comprehensive knowledge base (Tarus et al., 2018).

Hybrid systems overcome the limits of individual methods by integrating content-based and collaborative filtering approaches, using techniques like integrated, flow, and parallel types to improve prediction accuracy (Walek and Fojtik, 2020). For example, a hybrid approach has been successful in recommending novels based on user ratings and preferences (Khalifeh and A. Al-Mousa, 2021). Similarly, content-based filtering has been used to recommend movies by analyzing genres and obtaining user feedback (Adiy- ansjah et al., 2019), whereas music recommendations have used convolutional recurrent neural networks (CRNN) for feature analysis (Adiyansjah et al., 2019).

Knowledge-based systems have also gained popularity in online learning platforms, which recommend courses and resources based on user preferences (Ali et al., 2022). The enhancements show how versatile and flexible modern recommendation systems are across industries.

2.2 Algorithm used for Recommendation

Deep neural networks (DNNs) have demonstrated excellent performance in speech recognition, computer vision, and natural language processing (He and Chua, 2017). In recommendation systems, multi-layer perceptron (MLP) models (Liu et al., 2016) effectively interpret complicated user-item interactions, whereas recurrent neural networks (RNNs) (Manotumruksa et al., 2017) focus on sequential user feedback. Overfitting in DNN-based collaborative filtering has been addressed using methods such as dropout and regularization (Tan et al., 2016). RNN architectures have recently been expanded to accommodate contextual data, resulting in improved accuracy and relevance of context-aware recommendation systems.

2.2.1 Recommendation based on Text

The research paper offers a robust framework for summarizing tourist reviews, solving the difficulty of getting important details from huge amounts of data on platforms such as TripAdvisor(Avasthi et al., 2023). It uses the unsupervised Attention-based Aspect Extraction (ABAE) technique to extract important details from reviews and combine readability scores, sentiment analysis, and aspect relevance to produce brief, customizable summaries. This method benefits both travelers and service providers by allowing summaries to be adjusted to user preferences, such as gender or area. The experimental validation with a diversified TripAdvisor dataset and crowdsourcing demonstrate the framework's usefulness in improving aspect coverage, readability, and content diversity when compared to existing approaches such as FairSumm and Centroid. While the research excels in scientific rigor, a more in-depth examination of biases in user-generated material and scalability concerns would enhance its practical relevance. Overall, this research significantly advances text summarization and sentiment analysis in the tourism sector.

The effectiveness of the proposed methodology is proved by comparing it to innovative unsupervised methods for summing perspectives. The review was carried out using online methodologies, as well as the Fair Summ and Centroid methods. The next effort will include using tourist reviews from lesser-known destinations and focusing on issues such as culture, language, and heritage.

2.3 Text Summarization

Bhandari et al. (2020) explore text summarization and analyze performance indicators using the TAC and CNN/Daily Mail (CNNDM) datasets. Key findings show that Mover Score succeeds on TAC but performs badly on CNNDM, where ROUGE-2 outperforms. The paper highlights the importance of evaluating metrics across several datasets and applications, pointing out flaws in current evaluations that fail to appropriately compare top systems or individual summaries. The authors suggest collaborative metric development, like the WMT Metrics initiative in machine translation.

In (Alantari et al. (2022)), the authors used three supervised machine learning models to analyze the emotion category of reviews for chosen travel destinations. Three models are used: Naïve Bayes, SVM, and dynamic language models. The study analyzed blog evaluations of seven popular tourism locations in the US and Europe. All three algorithms reach around 80% accuracy in classification.

2.4 Algorithm Used for Travel Recommendation

The tourist recommendation algorithm employs a hierarchical Long Short-Term Memory (LSTM) model. Paper (Shafqat and Byun, 2020) introduced a deep learning-based trip recommendation algorithm that uses data from blogs, Google Maps, and TripAdvisor to offer personalized activities. Latent Dirichlet Allocation (LDA) was used to model topics in tourist blogs and extract sentiments from user reviews. The system used a hierarchical Long Short-Term Memory (LSTM) model in two stages: the first predicted probable next locations based on user travel history and contextual features such as weather and location popularity, and the second refined these predictions using XGBoost to prioritize critical features. This technique achieved 97.2% accuracy by combining historical and contextual data to provide exact, relevant, and personalized travel suggestions.

The paper employs a hybrid method combining the Artificial Bee Colony (ABC) algorithm and the TOPSIS model for tourism recommendations. The ABC algorithm optimizes the selection of tourist destinations by simulating the foraging behavior of bees, while the TOPSIS model evaluates alternatives based on multiple criteria. This approach allows the system to suggest the best tourist spots tailored to user preferences, enhancing decision-making for travelers, Forouzandeh et al. (2022).

2.5 Hybrid Recommendation

A hybrid recommendation system combines multiple recommendation strategies to leverage their complementary strengths. Traditional hybrid systems typically integrate collaborative filtering (CF), which suggests items based on preferences of similar users, and content-based filtering (CBF), which recommends items sharing characteristics with those previously interacted with by the user. Various hybrid systems have been applied across domains; for instance, combining movie ratings with textual data like tags and genres (Yang et al., 2018), or integrating learner profiles with material ratings for educational recommendations (Turnip et al., 2017). However, traditional hybrids often fail to capture the latent user

Preferences. Neural-based approaches address this limitation but typically focus exclusively on either CF or CBF, lacking integration of both.

Many organizations use information technology (IT) to improve corporate operations and compete more efficiently. The Internet's importance in business processes has made online business extremely competitive, with several low-cost online retailers. Customer loyalty is limited online, demanding personalized products and services. (Dwivedi et al., 2023) Traditional marketing often fails online, requiring the use of personalized solutions such as one-on-one marketing. The Penalization Travel Support System uses Reinforcement Learning to analyze user behavior and provide individualized travel recommendations. This system learns from user interactions and improves over time, increasing precision and recall when recommending trips based on user preferences and past data.

3 Methodology

This research paper suggests a Personalized Travel Support System that uses advanced machine learning to provide personalized travel recommendations.

The study adopts a structured Knowledge Discovery in Databases (KDD) methodology, comprising five key stages: data selection, pre-processing, transformation, modeling, and evaluation.

Each stage is further subdivided into detailed steps, forming a framework for developing an effective recommendation model as shown in Fig. 1. This systematic approach ensures the seamless integration of diverse datasets, robust feature engineering, and the creation of highly accurate recommendation systems tailored to individual user preferences and travel needs.

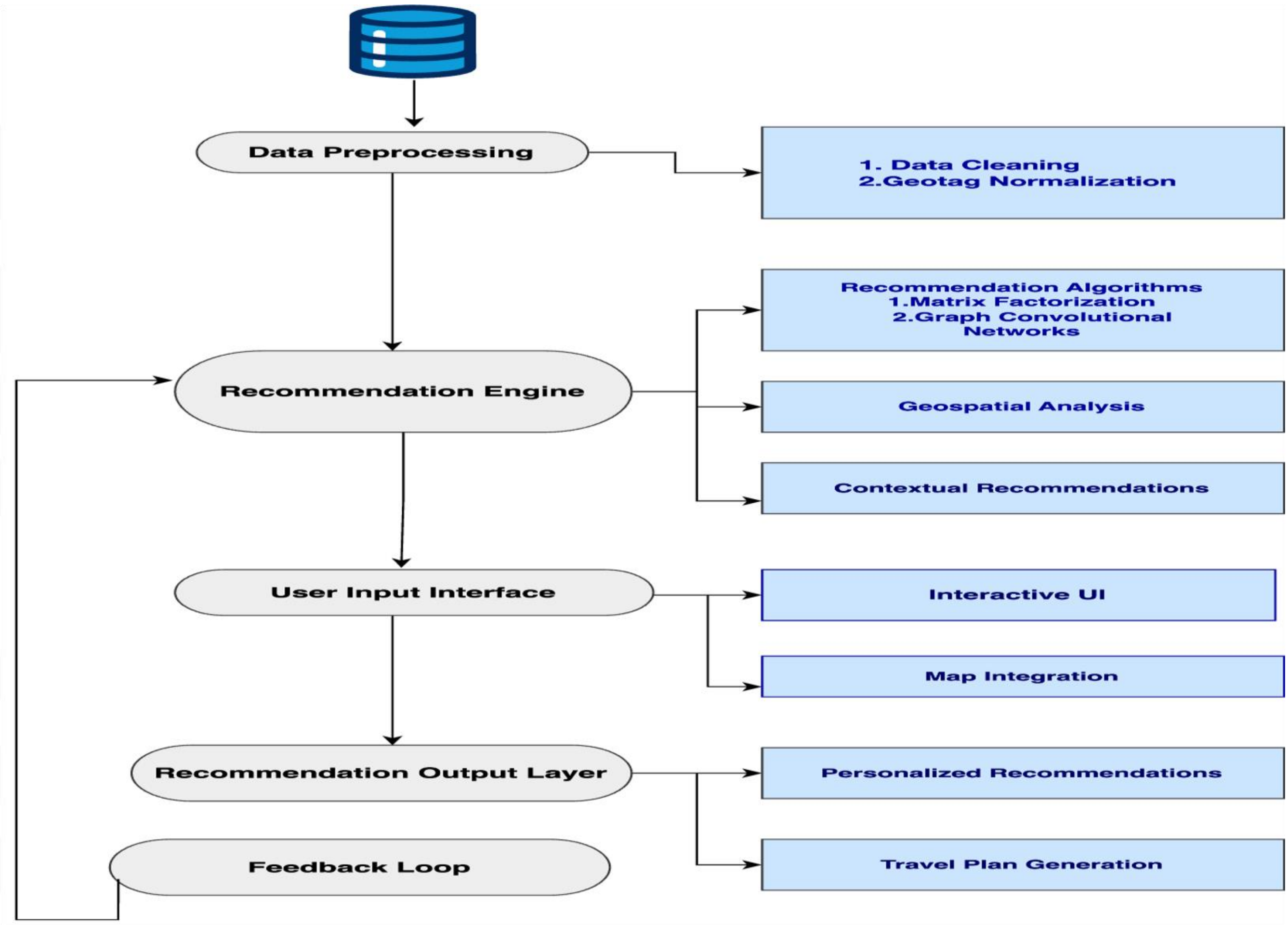


Figure 1: Workflow of Research

Data Sources: Gather Yelp reviews, Flickr geotags, and real-time weather/Google Maps data.

Preprocessing: Clean and format data, extract relevant features, and handle missing values.

Clustering with DBSCAN: Segment Flickr geotags into location clusters based on proximity and timestamps.

Embedding with LightGCN: Use LightGCN for personalized user-item embeddings from Yelp and Flickr data.

Weather Filtering: Integrate real-time weather data to filter out recommendations based on user preferences.

Google Maps Integration: Refine recommendations using proximity and travel time from Google Maps.

Model Optimization: Tune the model to improve recommendation accuracy and relevance.

3.1 Data Source and Exploration

The proposed travel recommendation system combines three data sources, including geotagged social media posts, real-time weather data, and location details from the Google Places API, to deliver personalized travel suggestions.

For this study, each dataset is individually cleaned and processed, then integrated to provide combined recommendations.

Table 1: Datasets Used for Travel Recommendation System

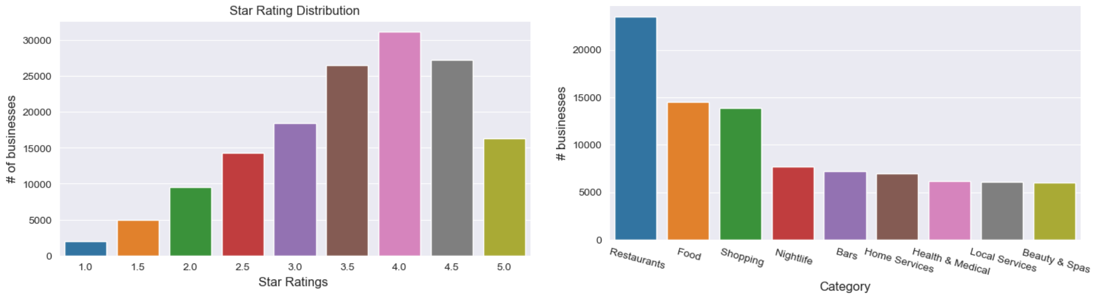
Aspect	Source	Purpose	Data Format	Size Of Dataset	Core Variables
Yelp Academic Dataset	Yelp merechant view platform	Business, review, and user data for NLP and visualization	JSON	7 million reviews, 150,000 businesses, 200,000 images	User reviews, ratings, images, business locations
Kaggle Travel Data- set	Kaggle	Geotagged data for travel recommendation modeling	CSV	20,000 records	Photo ID, user ID, latitude, longitude, timestamp
Weather map	open weather (via API), map	To identify ideal vacation spots based on weather data	JSON, CSV	Varies by APIdata size	Temperature, Humidity, Wind Speed, Cloudiness, City Name

To ensure comprehensive analysis, multiple datasets from various sources were selected:

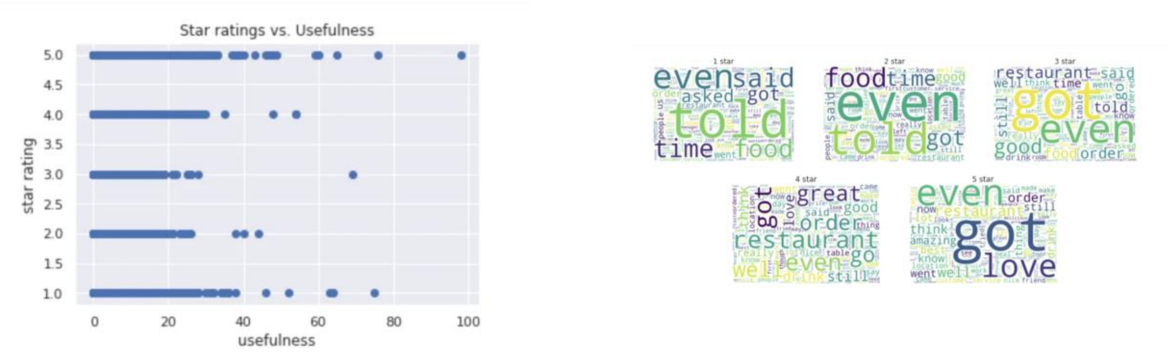
3.2 Data Exploration

Exploratory data analysis (EDA) was conducted to gain insights into data properties, visualize geotagged data distributions, and assess completeness.

- The **Yelp dataset** provides business, review, and user data for academic and personal use. Each business entry includes name, address, location, ratings, review counts, and categories, offering insights into consumer feedback and business scope. Fig. 2 shows most businesses have 4.0-star ratings, while Fig. 3 highlights restaurants as the largest category, reflecting consumer trends.
- **Users and reviews** The dataset, consisting of 100,000 entries from over 8 million reviews, enables analysis of user behavior, review sentiments, star ratings, and business attributes, revealing patterns in feedback and preferences.



(a) Distribution of businesses based on star ratings (b) Distribution of businesses by category
Figure 2: Overview of business distributions: star ratings (left) and categories (right)



(a) Star Rating Vs. Usefulness

(b) Most Popular Reviews word

Figure 3: Yelp Data EDA

- **Social Media Data:** Flickr provided geotagged data, including Data pre-processing prepared the Kaggle dataset with 20,000+ records for modeling.

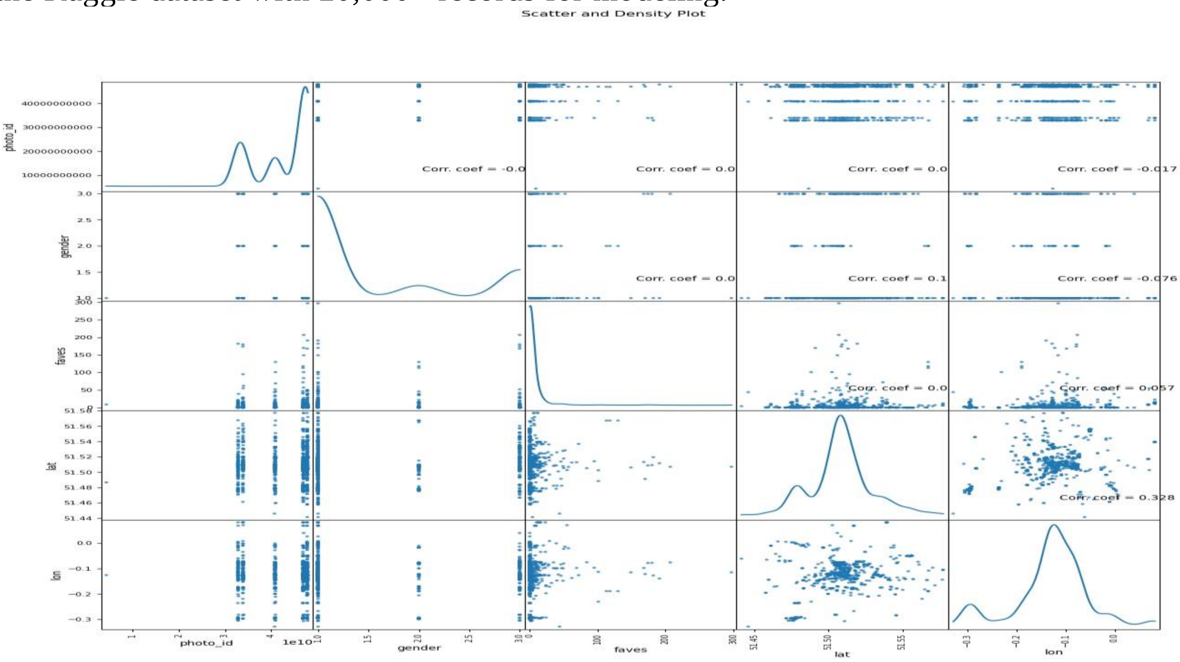


Figure 4. Scatter and Density Plot for Flickr data

From 13 attributes, five key variables were selected: photo ID, user ID, latitude, longitude, and photo timestamp as shown in Table 2. Irrelevant attributes were removed, and data types were optimized for efficient transformations. Latitude, longitude, and date fields were converted for faster operations. The dataset was verified for missing values, ensuring no imputation or removal was required before model construction.

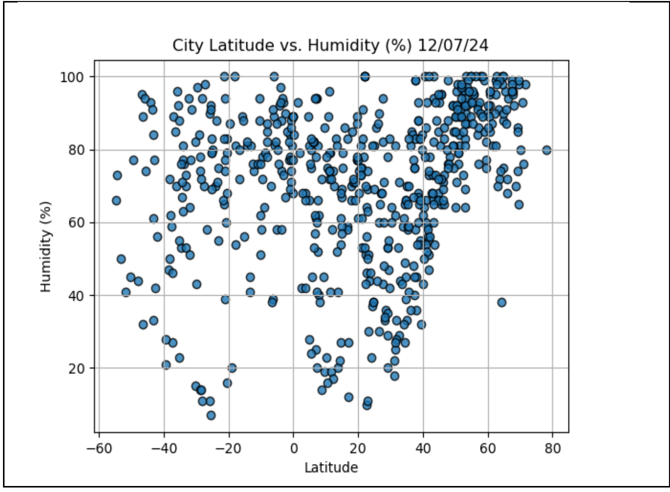
Table 2: Flickr Dataset Attributes Description

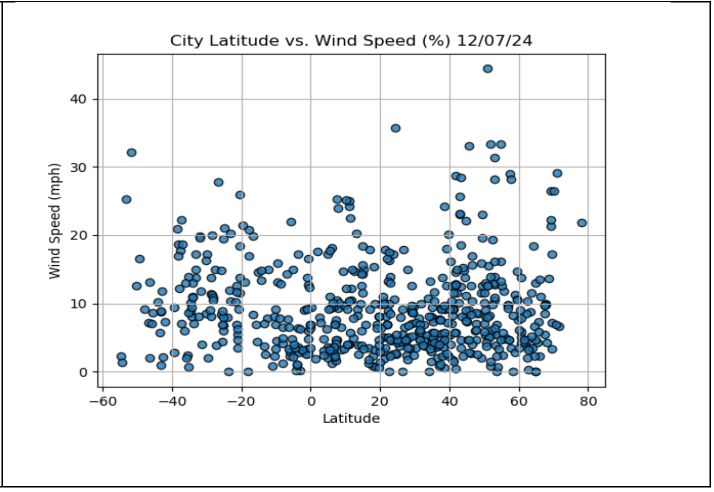
Attribute	Description
photo-id	Photo ID
user id	User ID
let	Latitude
long	Longitude
Taken (timestamp)	Time the photo was taken
location id (cluster label)	Cluster label of the location
cent let (cluster lat)	Latitude of the cluster centroid
cent long (cluster lon)	Longitude of the cluster centroid

- **Weather Data:** Verified for completeness and aligned with geospatial data using city names and coordinates.

The process follows a structured workflow, from data collection and preprocessing to pattern detection, leading to the final recommendation model.

The integration of a weather database enhances the recommendation system by incorporating user-defined parameters and preferences for up to five weather conditions, such as wind, rain, snow, cold, and heat. Users can specify comfort thresholds, enabling dynamic filtering of locations or activities that do not match their weather tolerance. For example, if temperatures exceed a user's preference for walking, the system excludes suggestions requiring extended outdoor exposure. During exploratory data analysis (EDA), the weather database was analyzed to ensure accurate alignment with user-defined thresholds. This integration ensures personalized recommendations based on weather adaptability, proximity, and travel.





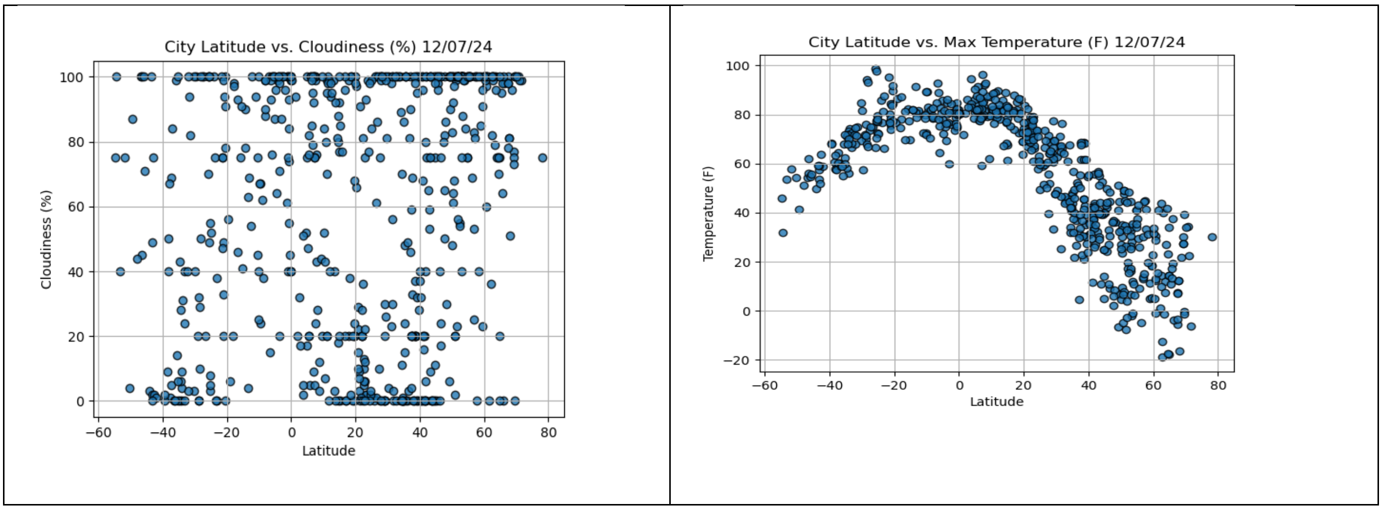


Figure5. EDA of weather data w.r.t tp cloudiness, temp, humidity, and wind

Time, and user preferences.

3.3 Variable Selection

For Yelp data After missing treatment and outlier treatment only keeping

Table 2. Yelp Data Description

Variables	Description
location	Location Id
lat	Latitude
lon	Longitude
user_id	User ID
visit_time	Time of Visit

For Flickr Data: Key variables included latitude, longitude, timestamps, and user IDs. These attributes provided essential geospatial and temporal context for developing the recommendation model by identifying user behavior patterns and popular locations.

Dataset Exploration: Missing values were carefully inspected, and geospatial patterns were visualized to understand the density of data points in different areas. EDA reduced the likelihood of errors in downstream processing and model development.

Table 3: Flickr Dataset Attributes Description

Attribute	Description
photo-id	Photo ID
user id	User ID
lat	Latitude
lon	Longitude
Taken (timestamp)	Time the photo was taken
location id (cluster	Cluster label of the location

label)	-
cent lat (cluster lat)	Latitude of the cluster centroid
cent lon (cluster lon)	Longitude of the cluster centroid

3.4 Pre-Processing

Data Cleaning: Data inconsistencies, missing values, and outliers were addressed. Records with complete latitude and longitude were removed to maintain dataset integrity. Duplicate entries, such as geotagged posts from the same user within three hours, were eliminated to improve data quality.

Splitting of data: The data was split into training (80%) and testing (20%). The recommendations were made based on the users' past travel experiences, and the recommended locations were ranked based on the projected values. The top recommendations to the target users were compared with the actual ratings.

3.5 Clustering:

3.5.1 For Flickr Data

3.5.1.1 Detecting Tourist Hotspots

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was employed to group geotagged data into clusters, effectively identifying tourist hotspots. DBSCAN's capability to detect clusters of arbitrary shapes and handle varying densities made it particularly suitable for this spatial data analysis.

3.5.1.2 Data Preparation and Clustering

Geotagged data was clustered to label travel locations and organize them based on timestamps. To enhance data quality, posts by the same user within a three-hour window were treated as a single visit, following guidelines from prior research. Users with fewer than three geotagged posts were excluded to ensure a robust travel history for collaborative filtering. The final dataset included five attributes:

1. Location ID (cluster label)
2. User ID
3. Latitude
4. Longitude
5. Timestamp

- **Clustering with DBSCAN**:

  DBSCAN was applied to identify travel locations, leveraging its strength in detecting clusters of varying densities while filtering noise. Two key parameters were optimized:

  - **Epsilon (eps)**: Defined the neighborhood radius for clustering.2 Using the elbow method, an optimal value of 0.15 was selected, balancing between missing data points and merging unrelated clusters.
  - **MinPts**: Represented the minimum points required to form a cluster. A value of 10 ensured well-defined clusters without excessive fragmentation.

To compute distances between latitude and longitude coordinates, the **haversine formula** was used.

Haversine formula was employed to calculate great-circle distances between two points on the Earth's surface, specified by their latitude and longitude. This metric is particularly effective for spatial clustering as it accounts for the Earth's spherical shape**.** Formula is

$$d = 2r\ arcsin(\sqrt{sin^2\left(\frac{\phi 2}{2} - \frac{\phi 1}{2}\right) + \cos(\phi 1)\cos(\phi 2)\ sin^2(\lambda 2 - \lambda 1/2)})$$

Where,

r = Earth's radius,

(ϕ1,λ1)(ϕ1,λ1), (ϕ2,λ2)(ϕ2,λ2) = latitude and longitude of two points.

This approach calculates great-circle distances on a sphere, providing accurate spatial clustering.

- **User-Location Rating Matrix**

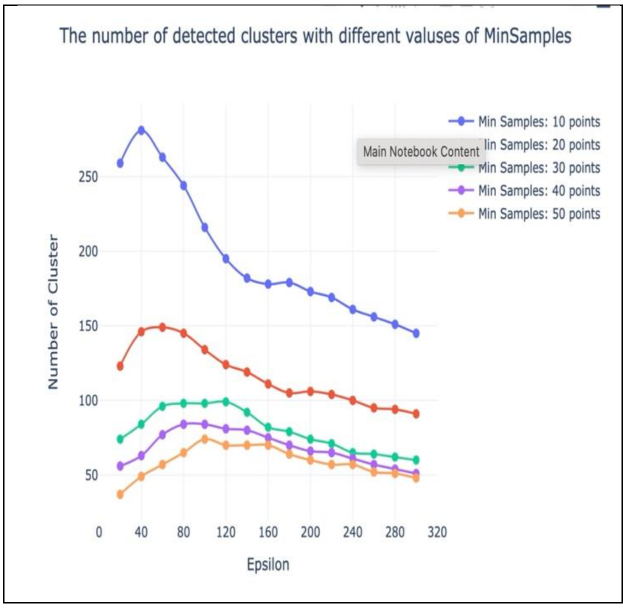


Figure 6. Epsilon

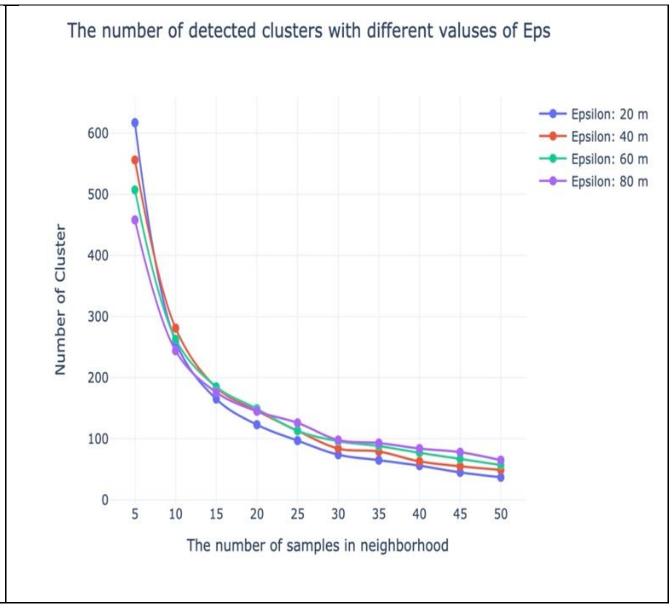
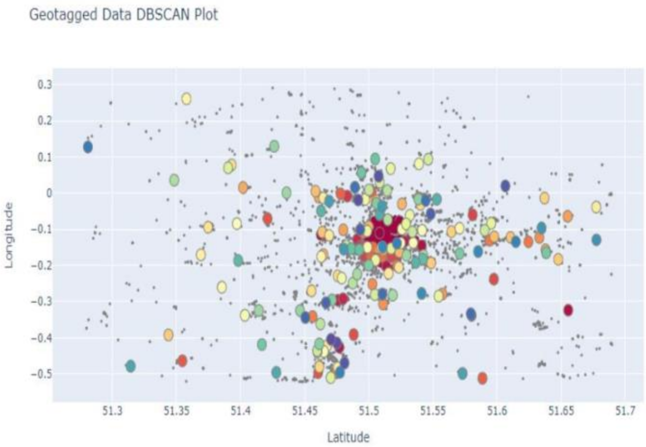


Figure 7. Clusters

In the data transformation phase, a user-location rating matrix was developed from the pre-processed geotagged data. This matrix captured user preferences by counting visit frequencies to specific locations, treating these counts as ratings. To mitigate biases toward heavily visited

Figure 8: DBSCAN Result



3.5.2 Clustering and For Yelp data

To provide personalized recommendations for hotels, restaurants, and attractions from Yelp data, weemploy Two algorithms: Light-GCN for restaurants and Singular Value Decomposition (SVD) for hotels.

And attractions. Each algorithm is used to the unique characteristics of the data subsets to maximize performance and recommendation relevance.

- **Light-GCN for Restaurant Recommendation** Light-GCN models user-restaurant interactions as a bipartite graph, iteratively aggregating neighbor information to create embeddings that capture relationships, enabling effective similarity computation and precise restaurant recommendations.

- **Funk-SVD for Hotel Recommendation** Funk-SVD, a matrix factorization technique, is particularly suited for sparse data often encountered in hotel recommendations. It decomposes the user-hotel rating matrix into two smaller matrices: P (user latent factors) and Q (hotel latent factors). Using Stochastic Gradient Descent (SGD), the algorithm optimizes these matrices to minimize prediction errors. Ratings for unrated hotels are predicted by reconstructing the matrix as the dot product of P and Q, facilitating personalized hotel recommendations.

- **SVD for Attraction Recommendation** For attractions, where data tends to be denser, traditional SVD is employed. The user-attraction rating matrix is factorized into three matrices: U (user factors), S (singular values representing feature importance), and V (attraction factors). By reconstructing the matrix using U × S × V, missing ratings are predicted. Attractions with the highest predicted ratings are recommended, aligning closely with user preferences.
- By combining Light-GCN and SVD, the system delivers recommendations across diverse domains, leveraging advanced machine learning techniques to enhance the travel planning experience.

3.5.3 LightGCN for Collaborative Filtering Recommendations

LightGCN (Graph Convolutional Neural Network) is a streamlined and efficient model designed for collaborative filtering in recommendation systems. Unlike traditional GCNs, LightGCN removes complex matrix transformations and non-linear activation functions, enhancing computational efficiency while preserving recommendation accuracy.

- **Reasons for Choosing LightGCN**

Graph-Based Collaborative Filtering: LightGCN leverages graph structures to model interactions between users and items, capturing complex relationships and high-dimensional collaborative signals, resulting in more precise recommendations.

Simplified Architecture: By reducing structural complexity, LightGCN ensures faster computation, making it suitable for large-scale datasets without sacrificing performance.

Unsupervised Learning: LightGCN operates without requiring labeled data, which is crucial in real-world scenarios where obtaining extensive labels is challenging.

These features make LightGCN a robust, scalable, and practical solution for delivering accurate recommendations in dynamic, data-rich environments.

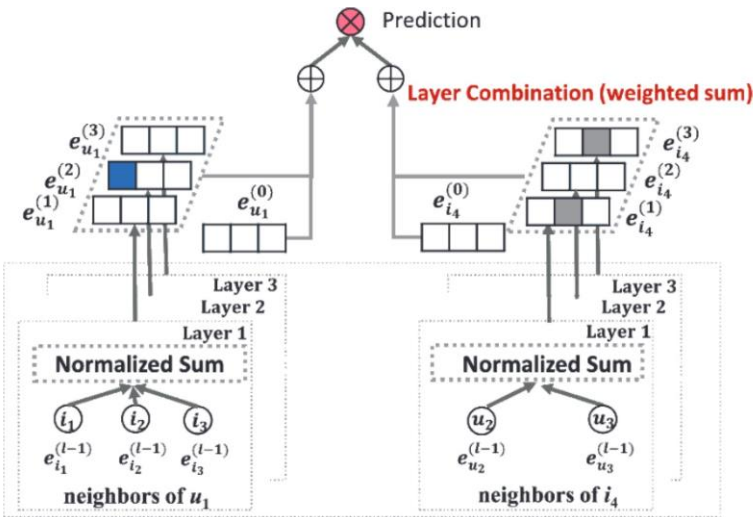


Figure 9: The principle of Light GCN

LightGCN recommends restaurants by iteratively updating node vectors based on interactions, ag-aggregating graph structure information, and calculating similarities to rank items, ensuring personalized recommendations for users.

Integration of Geotagged (Flickr) and Yelp Data Clusters

Yelp data was seamlessly integrated into the system to enrich geotagged clusters with real-time, contextually relevant information. Each cluster, derived from DBSCAN-based geotagging, was augmented with Yelp business data, including ratings, reviews, and business categories. This integration significantly enhanced the value and utility of the clusters.

3.6 Dynamic Ranking of Location

Locations within each cluster were dynamically ranked based on three primary metrics.

1. **Popularity Metrics:** Visitor traffic and engagement derived from Yelp and geotagged data.
2. **User Reviews:** Sentiment analysis and qualitative insights from Yelp reviews.
3. **Average Ratings:** Aggregate Yelp scores for businesses within the cluster.

This ranking mechanism ensured that users received recommendations for top-quality attractions, restaurants, and services in their areas of interest. For example, a cluster identified as a popular tourist spot could highlight nearby restaurants with high ratings or attractions with exceptional reviews.

By combining geospatial clustering with real-time Yelp data, the system provided precise, highly relevant, and user-cantered recommendations. This approach dynamically adjusted to new data, ensuring that recommendations remained current and reflected the latest user trends and preferences.

Dynamic Re-ranking: Recommendations within clusters are dynamically re-ranked based on their suitability under current weather conditions, ensuring the user receives contextually appropriate suggestions.

3.7 Integration with Weather Data

The weather data is seamlessly integrated with Yelp clusters through a modular API framework. Clusters are enriched with a suitability score derived from weather conditions. For instance, during sunny weather, outdoor locations such as trails receive higher scores, while indoor venues are deprioritized. Conversely, during rain or extreme weather, the system suppresses outdoor recommendations and highlights indoor options.

To compute location visit probabilities, we use **TF-IDF**:

$$. w_L^C = TF_l \times IDF_l = N_{c,l}/N_{c,\emptyset} \times \log N_{\theta,\theta} \ / \ log N_{\theta,l}$$

Where

- Nc,l: Visits to location l in context c
- Nc,∅: Visits to all locations in context ccc.
- N∅,∅: Total visits to all locations.
- N∅,l: Total visits to location l.

3.4 Implementation

The final stage of the implementation leveraged **local machines with GPU support** to perform computationally demanding tasks efficiently. Tools and libraries such as **Python**, **NumPy**, **pandas**, **sci-kit-learn**, **LightGCN**, and **SVD** were extensively used to produce the outputs. These outputs included a user-location rating matrix derived from geotagged Flickr data, clusters of tourist hotspots identified using DBSCAN, and embeddings generated using LightGCN for collaborative filtering. The model integrated Yelp reviews, real-time weather data, and Google Maps insights to deliver personalized travel recommendations. Outputs were dynamically re-ranked based on user preferences, proximity, and current weather conditions, ensuring context-aware relevance.

Key challenges during implementation included memory limitations and prolonged runtimes while processing large datasets and training models. These issues were addressed through data batching, pipeline optimization, and GPU acceleration. Geospatial clustering outputs were enriched with Yelp data to provide recommendations based on popularity, sentiment analysis, and ratings, while weather integration enabled suitability scoring to adapt recommendations to changing conditions.

The process delivered a robust recommendation engine capable of generating tailored suggestions for attractions, restaurants, and travel plans, aligning with individual preferences and real-time contexts. This stage demonstrated the integration of advanced algorithms, contextual data, and optimized computational workflows to ensure high accuracy and efficiency.

4 Results and Evaluation

In this study, we integrated three datasets, split in an 80-20 ratio for each data Set. To validate the effectiveness of our approach, we conducted experiments using the Models like

Collaborative filtering with asymmetric cosine similarity (ACOS) identifies user similarities forrating prediction. MF ACOS improves ACOS by using matrix factorization to address similarity.

Matrix sparsity. Popularity Ranking (PR) ranks tourist locations based on their popularity. Models, evaluating their performance with MAPE.

The evaluation of the techniques used in the travel recommendation system can be described as follows:
For the geotag data two methods are used and it is measured based on below

4.1 Geotagged Data: Flickr Results

DBSCAN (Geotag Clustering): Clusters geotagged data to identify travel locations. Evaluated with Silhouette Score and Davies-Bouldin Index.

The results are evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) on both training and test datasets to ensure unbiased ratings. As shown in Fig. The MAE decreased from 1.17 (train) to 1.2 (test), and RMSE slightly decreased from 1.4 (train) to 1.39 (test).

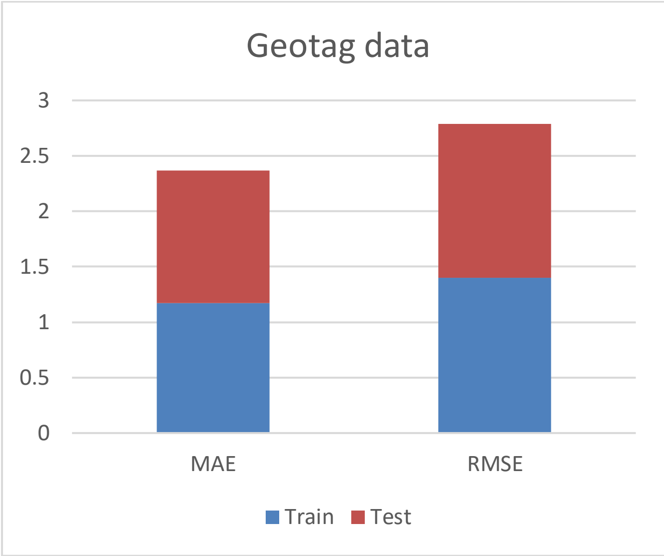


Figure 10. Evaluation Matrix of Flickr data

In the Yelp data each variable like attractions, and hotels clustered separately so

- **Funk-SVD (Hotels)**: For sparse hotel data. Evaluated with RMSE and MAE for prediction accuracy of ratings.
- **SVD (Attractions)**: Collaborative filtering for dense data. Evaluated using Singular Value Retention Ratio, RMSE, and MAE.
- **Light-GCN**: Collaborative filtering for restaurant recommendations. Evaluated with Precision@k, Recall@k, and NDCG for relevance.
- **TF-IDF (Visit Probability)**: Weights location visits based on context. Evaluated using Term Frequency-Weighted Accuracy and IDF distribution.

4.2 Experimental Results of Integrated Data of Yelp and Flickr

However First see the results of this experiment and evaluate the performance of the proposed model using multiple metrics, comparing it against two baseline methods: a popularity-based model and a random model.

The testing dataset, scaled to a range of 1–5, was used to construct a test rating matrix. The proposed matrix factorization model achieved a higher mean average precision (MAP) of 0.83 across recommendations (n = 5, 10, 15), outperforming the baseline models in precision and accuracy.

Table 3 shows the top suggestions for one specific user.

Table 4: Location predictions compared with true values.

location id	True	pred
505088	5	4.583333
594659	0	3.424242
767961	0	3.424242

The training results, as depicted in Table 3, provide insights into the model's performance based on loss, recall, and NDCG (Normalized Discounted Cumulative Gain) across 1000 epochs (seed = 2020):

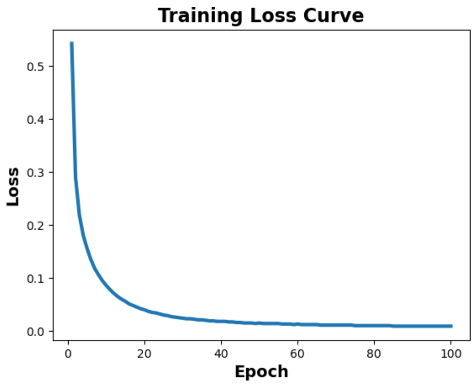
Table 5: Evaluation Metrics for Different Layers of LightGCN

Layer	Recall	NDCG	Precision
Layer 1	0.0765	0.381	0.145
Layer 2	0.0585	0.0475	0.0262
Layer 3	0.0685	0.0575	0.0272
Layer 4	0.0575	0.0495	0.0282

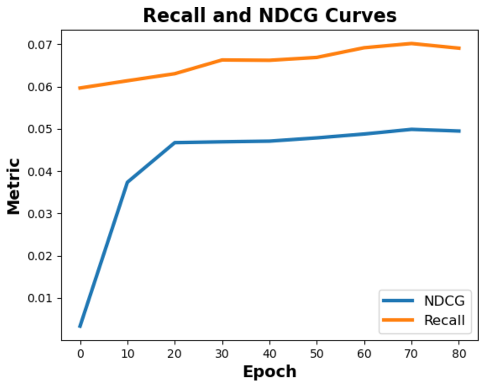
Loss: The model suggests a steady decrease in loss over 1000 epochs, initially at a rather high value and gradually decreasing as the model's parameters are refined. This shows that the training procedure significantly decreases errors and improves prediction accuracy.

Recall: Initially low, but gradually increases with training. For example, at epoch 1, recall is 0.0575 for some layers and reaches 0.0765 at its peak. This indicates the model's increasing capacity to recognize relevant positive samples.

NDCG scores, a measure of ranking quality, demonstrate growth over time. It starts at around 0.0475 for various layers and ends at 0.381 in the most effective layer (Layer 1).



(a) Loss Curve



(b) Recall and NDCG Curve.

Figure 11: Training Evaluation Metrics: Loss and Recall/NDCG Curves

Overall, the model indicates improvement in all specs, showing its ability to learn effectively and make high-quality recommendations. Further benchmarking with baseline models could provide a full Validation of these findings.

The proposed model was evaluated on random and test datasets using MAP@n (n = 5, 10, 15), RMSE, and MAE metrics. On test data, the model achieved MAP values of 0.74, 0.72, and 0.71 for n = 5, 10, and 15, respectively, with RMSE = 1.43 and MAE = 1.27, outperforming random data.

Table 5: Performance Metrics for Model Validation on Random and Test Data

Model - Validated on Data	5	10	15	RMSE	MAE
Random Data	0.723	0.76	0.77	1.23	1.7
Test Data	0.74	0.72	0.71	1.43	1.27

In the final stage, the proposed method was evaluated using two frequently used recommendation system metrics: Mean Average Precision (MAP) and Root Mean Square Error. These metrics offered insight into the model's accuracy and ranking efficacy.

Several experiments have been conducted on travel recommendation systems, drawing inspiration from various methods. In this experiment, we propose that using diverse data sources can enhance results. Our combined model, which incorporates weather data, provides more intuitive and accurate recommendations. When comparing our results with those from existing research, we observed an improvement of approximately 10%-15%. However, for a deeper understanding and more comprehensive analysis, a comparative study would be beneficial.

In addition, the suggested model's performance was assessed with other well-established recommendation systems. This comparison analysis emphasized the proposed method's merits and drawbacks while demonstrating its efficacy in providing personalized and accurate location-based recommendations.

5 Conclusion and Future Works

The experimental design utilized advanced methods like DBSCAN for clustering and Light-GCN for recommendations. However, limits in dataset size and variety may have influenced the portability of the findings. Furthermore, non-random sampling and insufficient parameter optimization may have produced biases. Addressing these issues in future studies may improve the validity of findings.

The experiment highlighted the potential of clustering and recommendation techniques but revealed inconsistencies when compared to prior research. These deviations may indicate either unique insights or flaws in the methodology. Increasing the dataset size, ensuring random sampling and refining parameters could resolve these issues and yield more reliable outcomes.

Despite its limitations, the study contributes significantly to the field of location-based recommendation systems. The integration of Yelp data enriched clusters with real-time information, enhancing the context and quality of recommendations. The user-location rating matrix, combined with Light-GCN, effectively modelled user preferences, offering a strong foundation for personalized travel suggestions.

Future work should incorporate larger datasets, explore advanced parameter tuning methods, and include real-time user feedback to improve recommendation accuracy. While the results are not broadly generalizable, this study provides a valuable methodology and a foundation for further research, with potential applications in enhancing travel experiences and personalized services.

5.2Future Scope

Looking ahead, there are several avenues for further research and development in personalized travel recommendation systems. First, enhancing user engagement through feedback mechanisms can refine recommendations and adapt to changing preferences over time. Second, exploring the integration of emerging technologies, such as augmented reality (AR) and virtual reality (VR), could provide immersive travel planning experiences. Additionally, addressing data privacy and security concerns will be crucial as systems become more reliant on personal data. Finally, expanding the model to include diverse cultural and regional contexts can enhance its applicability across different markets, ensuring that recommendations resonate with a global audience. By pursuing these directions, future research can continue to innovate and improve travel experience for users worldwide.

References

Adiyansjah, Gunawan, A. A. S. and Suhartono, D. (2019), ‘Music recommender system based on genre using convolutional recurrent neural networks’, *Procedia Computer Science* **157**, 99–109. The 4th International Conference on Computer Science and Computational Intelligence (ICCSCI 2019): Enabling Collaboration to Escalate Impact of Research Results for Society.
URL: *https://www.sciencedirect.com/science/article/pii/S1877050919310646*

Aggarwal, C. C. (2016), *An Introduction to Recommender Systems*, Springer International Publishing, Cham, pp. 1–28.
URL: *https://doi.org/10.1007/978-3-319-29659-3*1

Alantari, H. J., Currim, I. S., Deng, Y. and Singh, S. (2022), ‘An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews’, *International Journal of Research in Marketing* **39**(1), 1–19.

Ali, S., Hafeez, Y., Humayun, M., Jamail, N. S. M., Aqib, M. and Nawaz, A. (2022), ‘Enabling re- commendation system architecture in a virtualized environment for e-learning', *Egyptian Informatics Journal* **23**(1), 33–45.
URL: *https://www.sciencedirect.com/science/article/pii/S111086652100030X*

Avasthi, S., Prakash, A., Sanwal, T., Tyagi, M. and Yadav, S. (2023), Tourist reviews summarization and sentiment analysis based on aspects, *in* ‘2023 13th International Conference on Cloud Computing, Data Science Engineering (Confluence)’, pp. 452–456.

Bhandari, M., Gour, P. N., Ashfaq, A., Liu, P. and Neubig, G. (2020), ‘Re-evaluating evaluation in text summarization’, *CoRR* **abs/2010.07100**.
URL: *https://arxiv.org/abs/2010.07100*

Dom®nech, A., Mohino, I. and Moya-Gmez, B. (2020), ‘Using Flickr geotagged photos to estimate visitor trajectories in world heritage cities’, *ISPRS International Journal of Geo-Information* **9**(11).
URL: *https://www.mdpi.com/2220-9964/9/11/646*

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koo-hang, A., Raghavan, V., Ahuja, M. et al. (2023), 'Opinion paper: "so what if ChatGPT wrote it?" multi-disciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy’, *International Journal of Information Management* **71**, 102642.

Forouzandeh, S., Rostami, M. and Berahmand, K. (2022), ‘A hybrid method for recommendation systems based on tourism with an evolutionary algorithm and topics model', *Fuzzy Information and Engineering* **14**(1), 26–50.

Gentsch, P. (2018), *AI in marketing, sales, and service: How marketers without a data science degree can use AI, big data and bots*, springer.

Gupta, S. (2020), 'A literature review on recommendation systems', *Int. Res. J. Eng. Technol* **7**, 3600– 3605.

He, X. and Chua, T.-S. (2017), Neural factorization machines for sparse predictive analytics, *in* 'Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval', pp. 355–364.

Kazandzhieva, V. and Santana, H. (2019), 'E-tourism: Definition, development and conceptual framework', *Tourism: An International Interdisciplinary Journal* **67**(4), 332–350.

Khalifeh, S. and A. Al-Mousa, A. (2021), A book recommender system using collaborative filtering method, *in* 'International Conference on Data Science, E-learning and Information Systems 2021', pp. 131–135.

Kyrylov, Y., Hranovska, V., Boiko, V., Kwilinski, A. and Boiko, L. (2020), 'International tourism development in the context of increasing globalization risks: On the example of Ukraine's integration into the global tourism industry', *Journal of Risk and Financial Management* **13**(12), 303.

Liu, Q., Wu, S., Wang, D., Li, Z. and Wang, L. (2016), Context-aware sequential recommendation, *in* '2016 IEEE 16th International Conference on Data Mining (ICDM)', IEEE, pp. 1053–1058.

Manotumruksa, J., Macdonald, C. and Ounis, I. (2017), A deep recurrent collaborative filtering framework for venue recommendation, *in* 'Proceedings of the 2017 ACM on Conference on Information and Knowledge Management', pp. 1429–1438.

Marr, B. (2021), *Data Strategy: How to profit from a world of big data, analytics, and artificial intelligence*, Kogan Page Publishers.

Shafqat, W. and Byun, Y.-C. (2020),' A context-aware location recommendation system for tourists using hierarchical lstm model', *Sustainability* **12**(10), 4107.

Tan, Y. K., Xu, X. and Liu, Y. (2016), Improved recurrent neural networks for session-based recommendations, *in* 'Proceedings of the 1st workshop on deep learning for recommender systems', pp. 17–22.

Taurus, J. K., Niu, Z. and Mustafa, G. (2018), 'Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning', *Artificial intelligence review* **50**, 21–48.

Turnip, R., Nurjanah, D. and Kusumo, D. S. (2017), Hybrid recommender system for learning material using content-based filtering and collaborative filtering with good learners' rating, *in* '2017 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)', IEEE, pp. 61–66.

Walek, B. and Fojtik, V. (2020), 'A hybrid recommender system for recommending relevant movies using an expert system', *Expert Systems with Applications* **158**, 113452. **URL:** *https://www.sciencedirect.com/science/article/pii/S0957417420302761*

Yang, C., Chen, X., Liu, L., Liu, T. and Geng, S. (2018), A hybrid movie recommendation method based on social similarity and item attributes, *in* 'Advances in Swarm Intelligence: 9th International Conference, ICSI 2018, Shanghai, China, June 17-22, 2018, Proceedings, Part II 9', Springer, pp. 275–

285.