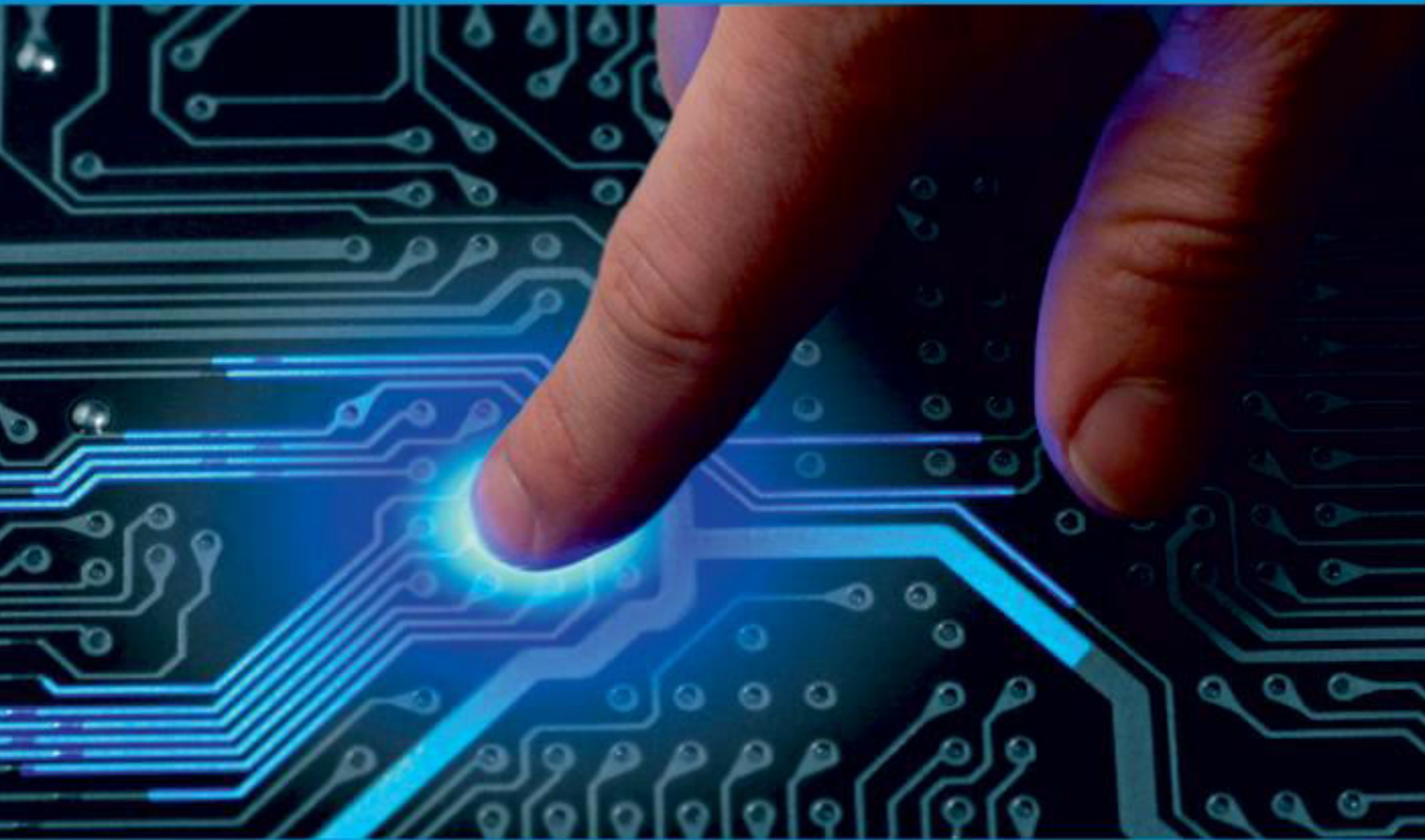
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 2, February 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542|

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002051 |

Developing a User-Friendly Hotel Recommendation System Based on User Reviews

Vishal Paranjape[1], Saurabh Sharma[2], Abhishek Singh[3]

Professor, Department of CSE, Baderia Global Institute of Engineering & Management, Jabalpur, Madhya Pradesh, India[1,2,3]

ABSTRACT: The hospitality sector is a significant global industry, encompassing restaurants, hotels, parks, cruises, and various entertainment services worldwide. Given its role in providing entertainment, relaxation, and tourism opportunities, it holds considerable importance for people everywhere. This project aims to focus specifically on the hotel industry, a key subcategory within hospitality. The hotel industry is dedicated to offering accommodation services to a diverse range of guests. Vacations are popular activities for individuals, couples, and families, typically involving some form of lodging such as hotels. It is crucial for guests to find hotels that align with their specific needs and preferences. Additionally, guests are increasingly discerning about their accommodation choices. Therefore, having a system that recommends hotels based on guests' criteria and standards is essential. This project seeks to develop a system that provides valuable hotel recommendations, helping users and guests find accommodations that meet their expectations and requirements. This research paper aims to develop an innovative and user-centric hotel recommender system that utilizes user reviews to assist travelers in finding hotels that best match their preferences. By integrating advanced natural language processing (NLP) techniques and machine learning algorithms, the proposed system will analyze and categorize hotel reviews, allowing users to filter and rank hotels based on specific criteria such as location, amenities, cleanliness, price, and customer satisfaction.

KEYWORDS: Hotel industry, Hotel recommender system, Natural Language Processing (NLP), Machine Learning, Guest preferences

I. INTRODUCTION

The hospitality industry, a major global sector, encompasses a wide range of services including restaurants, hotels, parks, cruises, and other entertainment services. Within this industry, the hotel sector plays a pivotal role in providing accommodation services to travelers. With the growing trend of vacations and staycations, it is essential for guests to find hotels that align with their specific preferences and needs. This necessitates the development of sophisticated recommender systems that can efficiently guide users in making informed choices about their accommodations. Recommender systems have emerged as a critical tool in various domains, offering personalized suggestions to users based on their preferences and behaviors. As outlined by Aggarwal (2016), these systems employ a variety of algorithms and techniques to filter and prioritize options that are most likely to appeal to the user . Ricci, Rokach, and Shapira (2015) provide a comprehensive overview of the methodologies and applications of recommender systems, highlighting their significance in enhancing user experience and satisfaction . In the context of hotel recommendations, content-based filtering methods, as discussed by Pazzani and Billsus (2007), are particularly relevant. These methods leverage detailed information about the items (hotels) and the user's past preferences to generate recommendations . However, to capture the nuanced preferences of users, it is also important to consider sentiment analysis techniques.
Sentiment analysis and opinion mining, as explored by Liu (2012), play a crucial role in understanding user feedback and reviews . By analyzing the sentiments expressed in hotel reviews, it is possible to gauge the overall user satisfaction and identify specific strengths and weaknesses of different hotels. Pang and Lee (2008) emphasize the importance of sentiment analysis in extracting valuable insights from textual data, which can significantly enhance the accuracy of recommendations . Moreover, comprehensive surveys of recommender systems, such as the one conducted by Bobadilla et al. (2013), provide a thorough examination of the various approaches and their effectiveness in different application areas . These surveys underscore the need for hybrid models that combine multiple techniques to deliver robust and reliable recommendations. Given the increasing importance of personalized recommendations in the hospitality industry, this paper aims to develop a user-friendly hotel recommendation system based on user reviews. By integrating advanced sentiment analysis and machine learning techniques, the proposed system seeks to offer tailored hotel suggestions that meet the specific criteria and standards of users. The following sections will delve into the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542|

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002051 |

methodology, implementation, and evaluation of the proposed system, demonstrating its potential to enhance the decision-making process for travelers.

1.1 Objective of the Study

The primary objective of this project is to create an intuitive recommender system for hotel reviews, enabling users to effortlessly find hotels that meet their specific criteria and access relevant reviews. This system aims to streamline the hotel selection process, offering users personalized suggestions that align with their preferences and needs.

1.2 Significance of the Study

The key beneficiaries of this recommender system include:
1. Guests: This system allows guests to easily access and evaluate hotel reviews, helping them identify hotels that best match their requirements. As a result, they can make more informed decisions regarding their accommodations.
2. Hotel Companies: The system provides hotel companies with easy access to reviews of their establishments, enabling them to devise strategies for improvement based on guest feedback and common criteria. Additionally, it may increase their visibility and help hotel management better understand consumer preferences for accommodations.
3. Online Travel Agencies and Site: This system can enhance the efficiency and effectiveness of online travel agencies and websites in recommending hotel accommodations to users, thereby improving their service offerings.

II. LITERATURE REVIEW

Recommender systems have become an integral part of various online platforms, enhancing user experiences by providing personalized suggestions. These systems leverage a variety of algorithms and techniques to predict user preferences and recommend relevant items. The development and evolution of these systems have been extensively covered in academic literature, providing a robust foundation for their application in different domains, including the hospitality industry.

Aggarwal (2016) offers a comprehensive overview of recommender systems, detailing the fundamental principles, methodologies, and algorithms that underpin these technologies. His work serves as a foundational text, covering collaborative filtering, content-based filtering, and hybrid approaches. Aggarwal's analysis provides a detailed understanding of how these techniques can be applied to develop effective recommender systems, highlighting their strengths and limitations in various contexts.

The Recommender Systems Handbook by Ricci, Rokach, and Shapira (2015) expands on these concepts, presenting a wide array of advanced methods and practical applications. This handbook delves into the nuances of different recommendation approaches, including knowledge-based and context-aware systems, which are particularly relevant for the dynamic and context-dependent nature of hotel recommendations. The comprehensive nature of this work makes it an invaluable resource for understanding the latest advancements and best practices in the field of recommender systems.

Pazzani and Billsus (2007) focus on content-based recommendation systems, a critical aspect of personalized recommendations. Their work, featured in The Adaptive Web, explains how these systems use item features and user profiles to generate recommendations. This approach is especially pertinent to hotel recommendation systems, where user preferences for specific hotel attributes (such as location, amenities, and price) can be leveraged to provide tailored suggestions. Pazzani and Billsus's insights into the implementation and effectiveness of content-based systems offer valuable guidance for developing robust recommendation engines.

Bobadilla et al. (2013) provide an extensive survey of recommender systems, covering the evolution of various techniques and their application across different domains. Their review in Knowledge-Based Systems discusses collaborative filtering, content-based filtering, and hybrid methods, highlighting the importance of incorporating multiple techniques to enhance recommendation accuracy. This survey underscores the need for hybrid models that can adapt to diverse user preferences and improve the overall effectiveness of recommendation systems
.
Sentiment analysis is another crucial component of modern recommender systems, particularly in the context of analyzing user reviews. Liu (2012) presents a thorough examination of sentiment analysis and opinion mining, detailing various methodologies for extracting and interpreting sentiment from textual data. His work emphasizes the importance

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542|

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002051 |

of sentiment analysis in understanding user opinions and enhancing the quality of recommendations. By incorporating sentiment analysis, hotel recommendation systems can provide more nuanced and accurate suggestions based on user feedback.

Overall, these seminal works provide a comprehensive framework for developing a user-friendly hotel recommendation system based on user reviews. By integrating the principles and methodologies outlined in these studies, the proposed system can effectively analyze user preferences and reviews, offering personalized hotel recommendations that cater to individual needs and enhance the user experience.

III. METHODOLOGY

3.1. Data Collection and Preprocessing

- Data Sources: Collect data from various online travel and review platforms such as TripAdvisor, Booking.com, and Yelp. The dataset should include hotel information, user reviews, ratings, and user profiles.
- Data Cleaning: Remove any duplicate entries, handle missing values, and filter out irrelevant information to ensure the quality and consistency of the data.
- Data Transformation: Convert raw data into a structured format. Tokenize and normalize text data, and encode categorical variables.

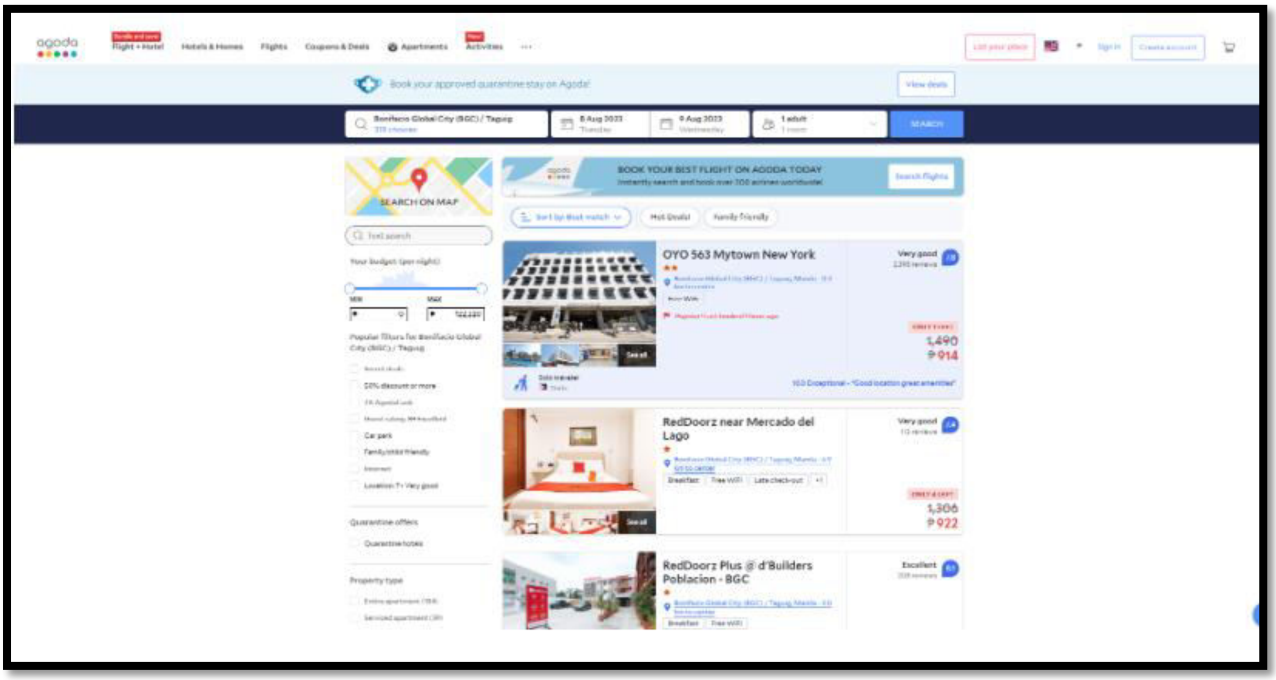


Figure 1: Data is taken from Agoda Website via webscapping

3.2. Exploratory Data Analysis (EDA)

- Descriptive Statistics: Calculate basic statistics (mean, median, mode) for numerical features like ratings and review counts.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542|

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002051 |

	userID	Rating	review_date	hotel_ID
0	Sandip	10.0	January 31 2023	F1 Hotel Manila
1	Duane	10.0	October 21 2022	F1 Hotel Manila
2	Peter	10.0	August 21 2022	F1 Hotel Manila
3	Ma.	10.0	November 05 2022	F1 Hotel Manila
4	Lyra	10.0	January 09 2023	F1 Hotel Manila
5	Desiree	9.6	November 07 2022	F1 Hotel Manila

Figure 2 : Dataset Representation in Data Frame

- Visualization: Create visualizations (histograms, bar charts, word clouds) to understand the distribution of data and identify patterns or trends.

3.3. Feature Engineering

- User Features: Extract features from user profiles such as user ID, review count, average rating given, and user preferences.
- Hotel Features: Extract features from hotel data such as hotel ID, location, amenities, average rating, and price range.
- Text Features: Utilize Natural Language Processing (NLP) techniques to extract features from review text, including sentiment scores, key phrases, and topic modeling.

hotel_ID	Airport transfer	Car park	Free Wi-Fi in all rooms!	Front desk [24-hour]	Fitness center	Swimming pool [outdoor]	Check-in/out [express]	Luggage storage	Valet parking	Hot tub	...	Dishes and dining utensils	Body wash
F1 Hotel Manila	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	NaN	NaN	...	NaN	NaN
Grand Hyatt Manila	1.0	NaN	1.0	1.0	1.0	NaN	1.0	1.0	1.0	1.0	...	NaN	NaN
OYO 563 Mytown New York	NaN	NaN	1.0	1.0	1.0	NaN	NaN	1.0	NaN	NaN	...	NaN	NaN

Figure 3 : Preprocessing & Feature Engineering of Data

3.4 Modeling Considerations

This project will utilize a content-based filtering approach for the recommender system. The primary focus is to suggest hotels to users based on the ratings they have assigned to each hotel. The chosen similarity metric is cosine similarity. Cosine similarity is selected due to its effectiveness in measuring the similarity between two vectors in a multi-dimensional space. In this scenario, the vectors represent the attributes and characteristics of hotels. By calculating cosine similarity, we can assess how closely the features of different hotels match a user's preferences. Additionally, cosine similarity's ability to capture both the direction and magnitude of similarity makes it a robust option for content-based filtering. This metric allows us to identify hotels that best align with a user's preferred features, resulting in more precise and meaningful recommendations.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542|

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002051 |

```
hotel_ID                                                userID
Capital O 804 The Residences at Bonifacio Civic Center  Maria     2
City Park Hotel Manila                                  Maria     1
Espacio Hotel                                           Maria     1
F1 Hotel Manila                                         Maria     1
Lot's Pad nr BGC                                        Maria     1
Muggle Stay Guest House - Bonifacio Global City Taguig  Maria     1
One Uptown BGC across Grand Hyatt Manila Hotel          Maria     1
Rare Unit With Balcony at The Venice Luxury             Maria     1
```

Figure 4 : The user has stayed at some hotels and provided reviews multiple times. To ensure an equal distribution of the ratings, a weighted rating should be calculated.

```
Airport transfer                     10.000000
Doorman                              10.000000
Concierge                            10.000000
Safety deposit boxes                 10.000000
Daily disinfection in common areas   10.000000
BBQ facilities                       10.000000
Kids club                            10.000000
Doctor/nurse on call                 10.000000
Salon                                10.000000
Buzzer/wireless intercom             10.000000
Elevator                             10.000000
Coffee shop                          10.000000
Wheelchair accessible                10.000000
Valet parking                        10.000000
```

Figure 5 : Retrieving Similarity

3.5 Model Evaluation

The model will be evaluated by testing the hit rate for a single user. In this case, we will use Maria, as she has reviewed the most hotels. We will examine how many of the recommendations align with the hotels Maria has actually reviewed.

```
The hit_rate of the top 5 recommendation is 0.4
The hit_rate of the top 10 recommendation is 0.3
The hit_rate of the top 20 recommendation is 0.2
The hit_rate of the top 30 recommendation is 0.2
The hit_rate of the top 40 recommendation is 0.2
The hit_rate of the top 50 recommendation is 0.16
```

Figure 6: Hit ratio for top recommendations

IV. RESULTS & CONCLUSION

From the initial results, we observe that the model successfully recommends hotels based on Maria's interests. The top suggested hotels possess all the features she prefers, indicating that the recommender system effectively matches hotels to user preferences. However, to thoroughly evaluate its accuracy, we will further analyze the model by examining the hit rate for the top N recommendations.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542|

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002051 |

The content-based recommendation system demonstrated promising hit rates across different recommendation scenarios. However, the hit rate decreased as the number of top recommendations increased. The system's accuracy plateaued after the top 10 recommendations, highlighting potential limitations in accurately predicting user preferences. A possible reason for this could be the limited user data, as Maria has only rated 14 hotels. As the number of recommendations increases, there are fewer rated hotels in her profile, leading to a decreasing hit rate. As the number of top recommendations decreases, the system focuses on suggesting a smaller subset of hotels. This allows the system to prioritize and recommend hotels that are more likely to be preferred by the user based on their historical ratings and preferences. Consequently, the recommended hotels are of higher quality and relevance, leading to a higher likelihood of hits. With a smaller "top n," the system is less likely to include hotels that are less relevant to the user's preferences. This reduces the chance of irrelevant suggestions cluttering the recommendation list, thereby improving the overall hit rate. Additionally, users are likely to be more satisfied when they receive a smaller set of high-quality and relevant recommendations, as opposed to a larger set that may include less pertinent options.

REFERENCES

1. Aggarwal, C. C. (2016). Recommender Systems: The Textbook. Springer.
2. Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender Systems Handbook. Springer.
3. Pazzani, M. J., & Billsus, D. (2007). Content-Based Recommendation Systems. In The Adaptive Web (pp. 325-341). Springer.
4. Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender Systems Survey. Knowledge-Based Systems, 46, 109-132.
5. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.
6. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
7. Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, 12(4), 331-370.
8. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. Computer, 42(8), 30-37.
9. Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749.
10. Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender System Application Developments: A Survey. Decision Support Systems, 74, 12-32.
11. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017). Neural Collaborative Filtering. In Proceedings of the 26th International Conference on World Wide Web (pp. 173-182).
12. Zhang, Y., & Yang, Q. (2017). A Survey on Multi-Task Learning. IEEE Transactions on Knowledge and Data Engineering, 29(12), 2319-2339.
13. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
14. McAuley, J., & Leskovec, J. (2013). Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text. In Proceedings of the 7th ACM Conference on Recommender Systems (pp. 165-172).
15. Salakhutdinov, R., Mnih, A., & Hinton, G. (2007). Restricted Boltzmann Machines for Collaborative Filtering. In Proceedings of the 24th International Conference on Machine Learning (pp. 791-798).