Electronics, Communications and Networks
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA241357

Restaurant Recommendation System in Malaysia Using Machine Learning Approach

Nur Idalisa[a] , Muhammad Hazwan Mohd Hazhar[a], Norliana Muslim[b] and Nur Lyana Shahfiqa Albashah[b1]

[a] *College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM), Dungun Campus, Dungun, Terengganu, Malaysia*

[b] *Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia*

ORCiD ID: Nur Lyana Shahfiqa Binti Albashah https://orcid.org/0000-0002-1664-3700

Abstract. People frequently struggle to make decisions when faced with a wider range of possibilities, especially when selecting a dining restaurant. To address this issue, a recommendation system can assist by analyzing user preferences and previous dining experiences to offer personalized suggestions. This research aims to develop a restaurant recommendation system for Malaysian customers using a machine-learning approach. The study focuses on Non-negative Matrix Factorization (NMF), Probability Matrix Factorization (PMF), Principal Component Analysis (PCA), and Singular Value Decomposition (SVD) approaches. Based on an analysis of 2,496 datasets gathered from the TripAdvisor platform, the findings revealed that the SVD method outperformed other approaches, achieving a Root Mean Square Error of 0.1166. This result positions SVD as the most suitable method for developing a restaurant recommendation system. The proposed system features a user-friendly interface built with Streamlit, allowing users to select their location and receive top restaurant suggestions. Additionally, users can view recommendations based on their past dining experiences. The system retrieves all reviews for the selected restaurants and converts them into a Term Frequency-Inverse Document Frequency (TF-IDF) matrix. Cosine similarity is then employed to measure the relevance of review content using the computed TF-IDF. Finally, the system also recommends similar restaurants based on the user's chosen options, enhancing the overall dining experience.

Keywords. Recommendation system, restaurant, machine learning, matrix factorization, cosine similarity

1. Introduction

Due to Malaysia's cultural diversity and varied food landscape, dining establishments in Malaysia serve many types of Malay, Chinese, Indian, and Southeast Asian cuisines. Malaysia's vibrant food culture, encompassing various eating outlets and culinary historical contributions, provides a centering function that promotes the nation's

[1] Corresponding Author: Nur Lyana Shahfiqa Albashah, shahfiqa@utar.edu.my

multicultural history through diversity [1]. Recommender systems (RS) are now common in our everyday lives as the web of things expands dramatically.

They are essential for exploring historical data about user preferences and determining which items might be excellent matches for new observations. In the past decade, several recommendation algorithms have been reported [2-6]. RS also functions as a type of information filtering system that uses analytics and data mining of user behaviors such as activities and preferences to select relevant information from a vast amount of available data [7][8]. This project aims to develop a restaurant RS by implementing machine learning approaches. This enables researchers to generate accurate and personalized recommendations that enhance user experience and satisfaction [9]. The rest of the paper is organized as follows. Section 2 introduces the RS techniques, related work in restaurant RS, and the machine learning approach, including the mathematical theory. In Section 3, the project results will be discussed through data visualization, performance measurement for all ML models, and the system's interface that recommends top dining places based on location and user preferences. Section 4 concludes the research.

2. Related work

2.1 Recommendation system

Table 1 illustrates the application of different restaurant RS that employ machine learning approaches. The current study found that there is a lack of methods utilizing the matrix factorization approach for restaurant RS.

Table 1. Previous research on restaurant RS

Application	Approach	Problem	Result
E-Halal Restaurant Recommender System [10]	K-means clustering.	The presence of "choice overload" can influence a consumer's decision-making. When confronted with extensive options, people often to avoid making a choice or postpone it.	50% believe the system has 51-75% accuracy. 40% agree the system is 75-100% accurate. While only 10% think the system offers 25-50% precision.
Restaurant Recommender System [11]	Semantic approach.	Many RS focus on fixed factors like food quality, prices, and service, often ignoring the changing nature of customer experiences and preferences.	The results show that the proposed system can provide recommendations with an impressive- accuracy of 92.8%.
Restaurant Recommendations with Implicit User Behavior [12]	Image classification and sentiment analysis.	Traditional RS often struggle to accurately capture users' true food preferences on social media, missing the subtleties of their tastes and interests in online culinary choices.	The mode-l correctly identified "1.png" as Curry Me-e, with a confidence of 72.97%. It also classified "10.png" as Spaghetti Bolognese-, with an impressive 99.99% certainty.
A Multi-Criteria Decision Support Model for Restaurant Selection [13]	2-tuple linguistic ordered weighted averaging aggregation, importance weights method.	Internet usage can lengthen the decision-making process because users need to gather and combine information from different sources before deciding.	A list of restaurants catering to various user demands can be compiled by evaluating their frequency in the top 10 across over 340 scenarios created by the proposed model.

Improve Restaurant Recommendation [14]	Combination of knowledge graphs and long-term and short-term interests of users.	Usual collaborative filtering relies heavily on use-r ratings. However, their e-ffectiveness is limited when there are insufficient ratings.	AUC: 0.9320, ACC: 0.8734, F1 score: 0.8773. This performance exceeds 4 other baseline methods.
Menu Recommendation System to Foreign Travelers [15]	Generates tables in the MySQL database, for RS to access the data.	Approximately 1.1% to 10.8% of people worldwide have dietary allergies.	28 out of 30 participants were able to select a meal that they liked.
Location-Based Restaurant Recommendation System [16]	Sentiment analysis and location-based recommendations.	Existing platforms often solely rely on star ratings to evaluate food quality, but this approach may not always accurately capture the full dining experience.	SVM reached 0.98, 60% of reviews were 4-star, indicating positive sentiment and 40% were 5-star, reflecting high satisfaction with the system.
Recommendation System Based on Item and User Similarity [17]	Item similarity and user similarity features.	The increasing number of internet-based businesses are pressing companies to innovate through technology.	The RS based on item similarity outperforms the user similarity approach in terms of the F1-measure metric.
Restaurant Recommendation System in Dhaka City [18]	Weight-based score and cosine similarity matrix.	The quality or the pricing doesn't meet the expectations if compared to other establishments, challenge to identify the best options that suit our preferences.	The cosine similarity method was found to be more accurate in this particular case compared to other algorithms.
Ontology Based Restaurant Recommendation Approach [19]	NLP pipeline, dialog management, action server, and MongoDB.	Consumers are frequently unable to evaluate a restaurant's offerings before physically using them, which can cause doubt and possibly discontent.	The experiment showed that using semantically enriched approaches can achieve high recommendation accuracy in RS

2.2 Matrix Factorization Approach in Recommendation System

Matrix Factorization (MF) is widely used in RS because it effectively handles hidden patterns in data. Common types of MF include Non-negative Matrix Factorization (NMF), Probabilistic Matrix Factorization (PMF), Singular Value Decomposition (SVD), and Principal Component Analysis (PCA). NMF [20] identifies structures that can be weighted to resemble each column (time series) of the original data. This method aims to discover matrices W and H that minimize the difference between V and WH, with $E = \frac{1}{2}||V - WH||^2_{Fro}$ where $||V - WH||^2_{Fro}$ represents the Frobenius norm squared. PMF [21] constructs a ranking by deleting user-rated items and arranging the remaining ones based on reconstruction ratings. It depends on $Loss = \sum_{R_{ij} \neq 0}\left(r_{ij} - U_i^T V_j\right)^2 + \lambda_U ||U||^2_F + \lambda_V ||V||^2_F$ where r_{ij} is the real rating, $U_i^T V_j$ is the approximated rating, λ_U and λ_V are regularized parameters. PCA [22] analyzes high-dimensional data to simplify the user-item rating matrix, uncovering patterns and generating recommendations effectively with $Z = X \cdot P$, where Z is the transformed data, X is the original data, and P is the matrix of principal components (eigenvectors). SVD [23] decomposes a matrix into singular vectors and values, providing a rank-reduced approximation of the original matrix based on $A = USV^T$ where U is an $M \times N$ matrix containing the left singular vectors, S is an $N \times N$ diagonal matrix with the singular values, and V^T is an $N \times N$ matrix with the right singular vectors.

Researchers have examined the effectiveness of MF techniques in different RS. SVD consistently outperforms other MF techniques across various RS applications, including cold-start scenarios [24], movie recommendations [25, 26], and image-based e-commerce recommendations [27]. Studies show that SVD achieves lower root mean square error (RMSE) [25] and better ranking accuracy compared to methods like NMF and k-nearest neighbors (KNN) [24], particularly excelling in sparse data environments and improving ranking performance by up to 32.32% [26]. While NMF tends to underperform in terms of accuracy, PCA-SVD proves highly effective for dimensionality reduction, significantly reducing feature space while retaining 90.01% of the variance in image-based RSs [27]. Based on these findings, the current study will construct models based on NMF, PMF, PCA, and SVD using the secondary dataset from TripAdvisor [28].

3. Results and Discussions

3.1 Data Description

From 2,496 TripAdvisor reviews, several visualizations were created to better understand trends and patterns in customer preferences. Figure 1 shows that most restaurant reviews are concentrated in major cities, with Kuala Lumpur contributing the largest share. In contrast, smaller regions like Penang and Kedah account for only about 1.2% and 1.0% of reviews, suggesting restaurants in urban areas attract more attention, likely due to higher populations, dining options, and tourism. Additionally, Figure 1 highlights the top 10 restaurants based on user ratings, with Pool Bar & Grill, Puree Juice, and Busdriver.Cafe earned perfect 5.0 ratings. Although Pool Bar & Grill received the most reviews (34), the overall number of reviews is moderate, suggesting that high ratings don't always equate to many reviews. Other highly-rated restaurants, such as Flying Monkeys Bar and Pizz, have fewer reviews. This shows that even top-rated restaurants can have varying review counts.

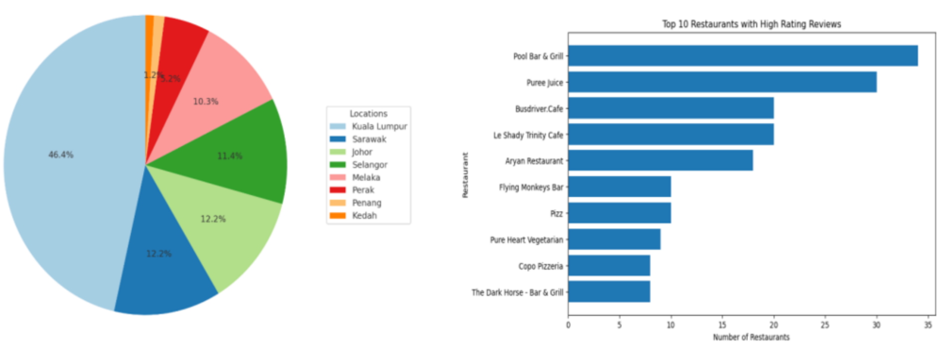

Figure 1. Distribution of restaurant location and rank

3.2 Performance measurement

The chosen machine learning models were developed using a 70% training, 15% testing, and 15% validation data split. This ensured the models were tested on both seen and unseen data, giving a clear measure of how well they generalize to new data. The test and validation sets represented unseen data, allowing evaluation of the model's ability to predict user preferences without overfitting. The Root Mean Square Error (RMSE) was used to measure how accurately the models predicted user preferences for the recommender system. The performance of all methods was evaluated, and Figure 2 shows the test and validation RMSE for different component numbers. For PMF, the test and validation RMSE at 100 components are 3.67 and 3.68, which is underperformed compared to other methods. While the train RMSE for PMF improves as the number of components increases, the validation RMSE fluctuates, indicating inconsistent generalization and no clear improvement trend.

All other three models perform similarly, with a consistent improvement in RMSE as the number of components increases. SVD exhibits the best overall performance, with consistently lower RMSE values for both test and validation sets (0.120 and 0.115 at 100 components). The NMF and PCA also improve but at a slower pace. Since SVD shows superior generalization across components, thus SVD is the most effective method for minimizing prediction error in the RS.

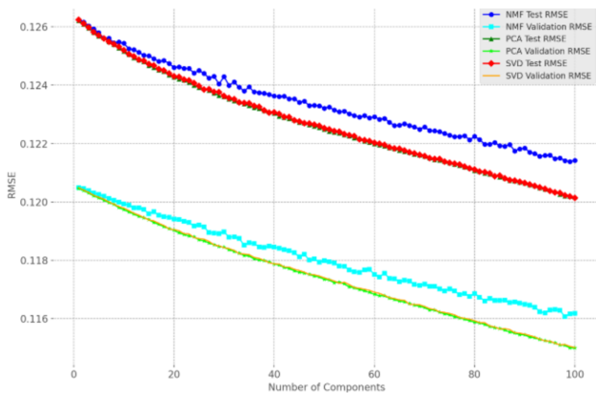

Figure 2. Performance comparison between NMF, PCA and SVD.

3.3 Restaurant Recommendation System

Based on superior performance during evaluation, the SVD and Term Frequency-Inverse Document Frequency (TF-IDF) methods were used to recommend restaurants within a specific location. All reviews for the selected restaurant are combined into a single string and converted into a TF-IDF matrix, which highlights important terms while reducing the impact of common words. Then, the cosine similarity between these reviews and those of other restaurants in the area is calculated using $cosine\ similarity\ (A, B) = \frac{A \cdot B}{\|A\| \|B\|}$. A and B represent the TF-IDF vectors of two documents (in this example, restaurant reviews). The magnitudes of the vectors denoted by $\| A \|$ and $\| B \|$, are Euclidean norms. The restaurants are arranged in descending order according to their similarity ratings. The top five restaurants, excluding the chosen favorite, are then picked as suggestions. Figure 3 shows a sample result when a user picks their preferred eating establishment. Based on the Python codes and Streamlit, the system was successfully developed and recommended five relevant restaurants with dietary preferences.

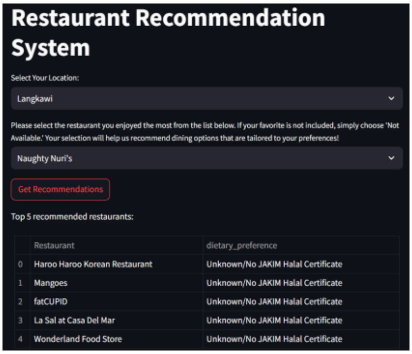

Figure 3. Sample of the recommendations based on selected restaurants.

4. Conclusion and future work

This work has showcased the effectiveness of matrix factorization techniques, in developing a basic dining recommendation system. The data description and visualization using Python set the stage, and the evaluation phase determined that SVD outperformed NMF, PMF, and PCA in terms of RMSE. As a result, the SVD was used to design the recommended dining locations, which were then presented through a simple user interface built with Streamlit. Future research should prioritize using techniques like grid search and Bayesian optimization to comprehensively explore different combinations of hyperparameters, including learning rates, regularization parameters, and the number of latent factors.

Acknowledgment

This scholarly work was related to grant code IPSR/RMC/UTARRF/2023-C2/A01. This academic work benefits from the infrastructural assistance provided by UiTM Terengganu Branch and Universiti Tunku Abdul Rahman.

References

[1] Tryotter. What is Malaysian cuisine? [Internet]. Available from: https://tryotter.com/resource/wiki/what-is-malaysian-cuisine.

[2] Abolghasemi R, Herrera Viedma E, Engelstad P, Djenouri Y, Yazidi A. A graph neural approach for group recommendation system based on pairwise preferences. *Information Fusion*. 2024 Jul 1;107:102343–3.

[3] Gharibi SJ, BagheriFard K, Parvin H, Nejatian S, Yaghoubyan SH. Ontology-based recommender system: a deep learning approach. *The Journal of Supercomputing*. 2024 Feb 7;80(9):12102–22.

[4] Bahrani P, Minaei-Bidgoli B, Parvin H, Mirzarezaee M, Keshavarz A. A hybrid semantic recommender system enriched with an imputation method. *Multimedia Tools and Applications*. 2023 Jul 12;83(6):15985–6018.

[5] Bukhari M, Maqsood M, Aadil F. KGR: A Kernel-Mapping Based Group Recommender System Using Trust Relations. *Neural Processing Letters*. 2024 Jun 19;56(4).

[6] Shrivastava V, Kumar S. Deep Neural Network Empowered Movie Recommender System Using Hesitant Fuzzy Bi Objective Clustering. *Journal of The Institution of Engineers (India): Series B*. 2024 Sep 18.

[7] Zhang Z, Patra BG, Yaseen A, Jie Z, Sabharwal R, Roberts K, et al. Scholarly recommendation systems: a literature survey. *Knowledge and Information Systems*. 2023 Jun 4;65(11):4433–78.

[8] Gwyn S, Caleb J, Christoper A, Guialil JS, Micheal D, Dioses RM. An Enhanced Content-based Filtering Using Maximal Marginal Relevance. *International Journal of Computing Sciences Research*. 2024;8:3070–87.
[9] Sae-Ang A, Chairat S, Tansuebchueasai N, Fumaneeshoat O, Ingviya T, Chaichulee S. Drug recommendation from diagnosis codes: classification vs. collaborative filtering approaches. *International Journal of Environmental Research and Public Health*. 2022 Dec 25;20(1):309.
[10] Mahadi M, Zainuddin N, Bazamah N, Shah A, Naziron NA, Fatimah S, Rum M. E-Halal restaurant recommender system using collaborative filtering algorithm. *Journal of Advanced Research in Computing Applications*. 2018;12:22–34.
[11] Asani E, Vahdat-Nejad H, Sadri J. Restaurant recommender system based on sentiment analysis. *Machine Learning with Applications*. 2021;100114.
[12] Yap WC, Kai KK, Goh CP. Restaurant recommendations with implicit user behavior via image classification and sentiment analysis from social media. 2024.
[13] Shu Z, Carrasco RA, Sánchez-Montañés M, García-Miguel JP. A multi-criteria decision support model for restaurant selection based on users' demand level: The case of Dianping.com. *Information Processing & Management*. 2024;61(3):103650.
[14] Miao L, Li X, Yu D, Ren Y, Huang Y, Cao S. Integrating users' long-term and short-term interests with knowledge graph to improve restaurant recommendation. *Journal of King Saud University - Computer and Information Sciences*. 2023;35(9):101735.
[15] Jiang Q, Wang T, Yamaguchi S. Food allergen database for Japanese restaurants and its application to menu recommendation system to foreign travelers. *Lecture Notes in Electrical Engineering*. 2023.
[16] Khot S, Dhumal R, Singh TP. Location based restaurant recommendation system (RECOSYS) using sentimental analysis. 2023.
[17] Mustafa AA, Budi I. Recommendation system based on item and user similarity on restaurants directory online. 2018.
[18] Ahmed T, Akhter L, Talukder FR, Hasan-Al-Monsur, Rahman H, Sattar A. Restaurant recommendation system in Dhaka city using machine learning approach. *IEEE Xplore*. 2021 Dec 1.
[19] Bandara HMRL, Ranathunga L. Ontology based restaurant recommendation approach. 2023.
[20] Aledavood T, Kivimäki I, Lehmann S, Saramäki J. Quantifying daily rhythms with non-negative matrix factorization applied to mobile phone data. *Scientific Reports*. 2022;12.
[21] Tian B, Gu Y, Liu S. A probability matrix factorization for user behavior perception recommendation model. *2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*. 2023.
[22] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer*. 2009;42(8):30–7.
[23] Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*. 2010;11(1).
[24] Sallam RM, Hussein M, Mousa HM. An enhanced collaborative filtering-based approach for recommender systems. *International Journal of Computer Applications*. 2020;176(41):9–15.
[25] Stodulka J. Collaborative filtering: matrix factorization recommender system. 2019.
[26] Widiyaningtyas T, Ardiansyah MI, Adji TB. Recommendation algorithm using SVD and weight point rank (SVD-WPR). *Big Data and Cognitive Computing*. 2022;6(4):121.
[27] Addagarla SK, Amalanathan A. Probabilistic unsupervised machine learning approach for a similar image recommender system for e-commerce. *Symmetry*. 2020;12(11):1783.
[28] Choon K. Malaysia Restaurant Review Datasets [Internet]. Kaggle; [cited 2024 Sep 24]. Available from: https://www.kaggle.com/datasets/choonkhonng/malaysia-restaurant-review-datasets.