Journal of Informatics and Web Engineering

Vol. 3 No. 3 (October 2024) eISSN: 2821-370X

Performance Evaluation on E-Commerce Recommender System based on KNN, SVD, CoClustering and Ensemble Approaches

Wan-Er Kong[1*], Tong-Ern Tai[2], Palanichamy Naveen[3], Heru Agus Santoso[4]

[1,2,3] Faculty of Computing and Informatics, Multimedia University, 63100 Cyberjaya, Malaysia.

[4]Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia.

**corresponding author: (kong.wan.er@student.mmu.edu.my; ORDIC:0009-0003-4221-1850)*

Abstract - E-commerce recommender systems (RS) nowadays are essential for promoting products. These systems are expected to offer personalized recommendations for users based on the user preference. This can be achieved by employing cutting-edge technology such as artificial intelligence (AI) and machine learning (ML). Tailored recommendations for users can boost user experience in using the application and hence increase income as well as the reputation of a company. The purpose of this study is to investigate popular ML methods for e-commerce recommendation and study the potential of ensemble methods to combine the strengths of individual approaches. These recommendations are derived from a multitude of factors, including users' prior purchases, browsing history, demographic information, and others. To forecast the interests and preferences of users, several techniques are chosen to be investigated in this study, which include Singular Value Decomposition (SVD), k-Nearest Neighbor Baseline (KNN Baseline) and CoClustering. In addition, several evaluation metrics including the fraction of concordant pairs (FCP), mean absolute error (MAE), root mean square error (RMSE) and normalized discounted cumulative gain (NDCG) will be used to assess how well different techniques work. To provide a better understanding, the outcomes produced in this study will be incorporated into a graphical user interface (GUI).

Keywords— Machine Learning, Filtering Technique, E-Commerce System, Recommender System, Collaborative Filtering

Received: 02 April 2024; Accepted: 02 June 2024; Published: 16 October 2024

1. INTRODUCTION

Recommender systems (RS) play an important role in people's everyday lives. Many companies are applying this technology to their website with notable examples such as Netflix, Amazon, and YouTube. This system leverages user data to provide suitable recommendations based on Big Data [1]. RS can help by increasing the company income greatly and easily stand out from competitors if the applied algorithm is well-crafted.

Journal of Informatics and Web Engineering

https://doi.org/10.33093/jiwe.2024.3.3.4

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

People create a tremendous amount of data daily through a variety of platforms in the present era of digitalization [2]. There are many important insights to be found in this enormous data volume using data science concepts. These data can help in forecasting possible trends for businesses, creating business plans, and identifying food preferences. Without the usage of such technology, consumers could find it difficult to sort through the vast amount of information available to them, which would make for a less-than-ideal user experience. This is valid in any field, including education.

RS as a subclass of information filtering systems relies on algorithms to provide consumers with tailored recommendations based on user preferences [3]. The algorithm can identify the most favored product from a wide range of possibilities [4]. The most important reason for building an RS is to help customers make decisions and improve their experience.

The recommendation created is based on a variety of data sources, including demographic information, previous purchases, browsing histories, and more [5]. These recommendations are developed using the data collected from customers. Effective development of the recommendation algorithm has the potential to set a firm apart from competitors and yield significant earnings.

Machine Learning (ML) leverages past data to help computers in building knowledge on their own [6]. ML can create models to forecast user preference using a variety of techniques. The use of this approach is very popular in several domains, including picture identification, ticketing time resolution, and future sales projection.

This paper delivers a thorough knowledge of ML, including the objective of ML and the characteristics of several ML approaches such as advantages and disadvantages, as well as the suitability for use in an e-commerce product RS. One way to discover all this information is by studying several publications and summarizing the conclusion

2. MATERIALS AND METHOD

2.1 Stages of the RS

A fundamental necessity in constructing an effective RS is having ample data to analyze and understand user traits and actions. The volume of data significantly shapes the recommendations provided to users [7]. Insufficient or low-quality data will invariably lead to flawed outcomes, regardless of the system's capabilities. The RS process involves three key phases: information gathering, data analysis, and prediction. Figure 1 illustrates these three stages of the RS.

A range of inputs are being used by the RS, including indirect feedback and response. These inputs help to determine user preferences and interests. Hybrid feedback can be created by integrating both direct and indirect input [8]. A RS cannot achieve its objective without a well-constructed model.

Implicit feedback is one of the approaches for collecting user data. There are various methods for obtaining this data, such as previous purchases, clicked links, and email interactions. Instead of using direct user input, these methods let RS interpret data and generate results by itself. As for explicit feedback, this method requires obtaining input from users. One of the examples of this method is prior rating items. Compared to implicit feedback, this approach needs more effort from users. By integrating both implicit and explicit feedback, the disadvantage of each method can be mitigated. This can generate hybrid feedback which surpasses both types of input [8]. The hybrid feedback is achieved by allowing users to provide ratings and direct comments while incorporating indirect data into the recommendation process. The recommendation process is shown in Figure 1.

The process begins by gathering user data. The data is then examined and put through analysis within the system. Subsequently, the second stage involves using a learning algorithm to comprehend user interests fully. These learning algorithms are quite helpful in identifying the underlying patterns seen in the data [9]. This process concludes at the prediction step. Tailored recommendations are generated at this stage by utilizing the insights obtained from the trained model.

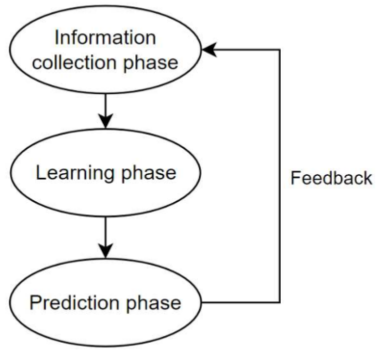


Figure 1. Recommendation Process

2.2. RS Techniques

A RS is expected to employ an accurate and efficient recommendation algorithm to provide each user with suitable suggestions. This emphasizes the importance of understanding the characteristics and applicability of different types of RS. Making good decisions during the algorithm design and development of RS is the most crucial part of the system. The variety of recommendation techniques presented in research publications is illustrated in Figure 2.

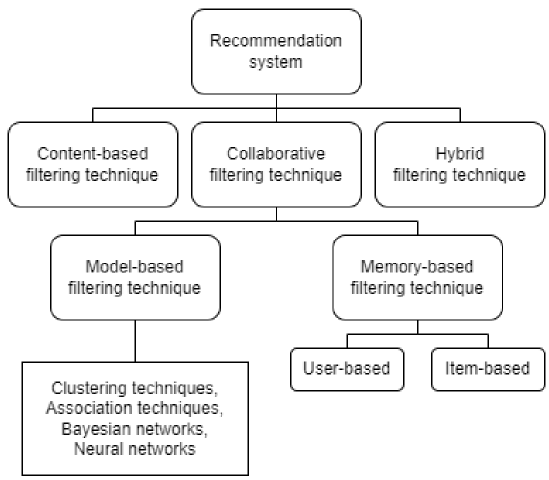


Figure 2. Common Recommendation Approaches

There are three main types of recommendation approaches which are content-based filtering technique, collaborative filtering technique and hybrid filtering technique. Content-based filtering technique analyzes user preferences and item attributes to provide suggestions. As for the collaborative filtering (CF) technique, this approach is derived from the behavior of a big user base. The hybrid filtering method builds from the combination of multiple recommendation techniques on top of one method to increase the model efficacy and accuracy. Below shows the description of each type of recommendation technique:

1) Content-based Filtering Technique: The content-based method employs an algorithm that evaluates the item features to provide users with forecasts. Content analysis is widely used in recommending websites, journals, and others. Each recommendation is customized by utilizing past user data and profiles to discover their preferences. Products are recommended if they have a strong positive link with well-rated content. This method is implemented using two techniques: vector spacing and the classification model [10]. A RS is expected to employ accurate as well as efficient recommendation algorithms to provide each user with suitable suggestions. This emphasizes the importance of understanding the characteristics and applicability of different types

a) *Vector spacing:* The vector spacing approach plays an important role in adjusting the recommendation ranking according to user preferences [11]. The first step of this method is creating vectors for both people and objects and then producing a dot product from these two. The dot product calculations are used to assess an item's relevance. For instance, the algorithm suggests related horror movies to a user who is interested in horror movies. The available movies will be ranked using dots with only the top-ranking movies presented to the user.

b) *Classification method:* The classification method employs decision trees as categorization techniques to produce recommendations [12]. To conclude with a final suggestion, a decision tree constructs a series of related assertions or conditions. When recommending movies to users, for example, the algorithm first determines if the movie fits the user's tastes and is in the horror category before recommending it. It then assesses the movie's rating; if other viewers give it a low rating, the algorithm decides not to suggest it.

2) Collaborative Filtering Technique (CF): CF is one of the most common techniques for recommending media, such as music or video. Especially in situations when metadata is insufficient. This method works by compiling a list of user preferences and then matching users based on similar interests to provide suggestions. These recommendations are based on the comparison of user profiles. The two primary categories of CF are model-based and memory-based techniques.

a) *Memory-based method:* The memory-based technique offers the simplest answer because it doesn't require any models. This method uses prior data to perform nearest neighbor distance computation to find the user groups who share similar tastes and suggest high rating products to them. This strategy can be done in two ways, which are item-based CF and user-based CF. Item-based CF couples objects with comparable ratings across users while user-based CF pairs users with similar ratings on the same items.

b) *Model-based method:* A model-centered approach needs a base model that is capable of precisely corresponding with the results that are attained. In this method, matrix factorization is frequently used to simplify user-item matrices, improving algorithm performance and cutting down on memory consumption and calculation time [13]. Numerous fields require the matrix factorization technique to deal with sparse matrices, for example: recommendations for users with a small number of rated items.

3) Hybrid Filtering Technique: The hybrid filtering technique combines many approaches. There are several types of hybridization in recommendation algorithms, including mixed, switching, weighted, cascade, feature augmentation, and feature combination approaches [11].

a) *Mixed:* This hybrid method involves combining and displaying several ranked lists into a unified one [14]. Every component of this hybrid must be able to provide ranked suggestion lists, which are then combined into a single ranked list by the mixed hybrid's core algorithm. The key concern here revolves around determining the methodology for generating the new rank scores.

b) *Switching:* This strategy chooses only one recommender from a group of potentials. This decision highly depends on unique circumstances encountered. The performance of the components may change based on the situation, thus the selection criterion such as confidence value or external criteria must be specified prior to achieve the optimal performance.

c) *Weighted:* With this hybrid technique, a linear formula is used to aggregate the scores from each component. As such, it requires all components to be able to produce recommendation ratings that are consistent with linear combination. In addition, the components must exhibit constant performance and precision with respect to the product space.

d) *Cascade:* The Cascade Hybridization Method prioritizes efficiency by first generating a broad ranking of candidates and then refining only the necessary items. It reduces required computing power by selectively applying approaches, which is more efficient compared to weighted hybridization methods [15]. This method also allows for alterations in low-priority suggestions without completely overturning higher-level ratings, which helps it withstand noise better.

e) *Feature augmentation:* This method is similar to hybrids of feature combinations, but it has an additional feature in which the contributor creates new features. In comparison to the feature combination approach, it provides more flexibility and adds fewer dimensions.

f) *Feature combination:* This method has two recommender components which are the contributing and the actual recommender. Data modified by the contributing component is used by the recommender itself. By adding characteristics from another source, the contributing component enhances the source of one component

The preceding sections discussed common strategies used in RS across various industries. Nonetheless, each approach possesses distinct characteristics, making it challenging to ascertain the most suitable technique for a specific task. Table 1 is provided to highlight the advantages and disadvantages of each approach.

Table 1. Pros and Cons of RS technique

Techniques	Pros	Cons
Content based Filtering	• The cold start issue is resolved • Can explore user preferences and generate recommendations that may not necessarily appeal to other users.	• Domain expertise is necessary as the system depends on manually created feature representations for items. • Restricted capacity to cope with user's previous interests.
Collaborative Filtering	• Does not require domain knowledge • Can help users explore new interests • Can train a matrix factorization model using only the feedback matrix to a certain level	• Suffering from cold start problems. • Difficult in employing item side features due to data sparsity issues
Hybrid Filtering	• Multiple filtering techniques are combined to overcome issues of each approach	• The model is sometimes too complex and needs strong computation power in order to run the model • May require expensive hardware for application.

Table 1 clearly illustrates that each recommender approach is unique and comes with its own set of advantages and disadvantages. Content-based filtering, which relies on user ratings, is immune to the cold-start problem. Conversely, CF recommends highly rated items to users based on user data. While the hybrid filtering method integrates multiple techniques, potentially leading to greater results, it requires more processing power and could incur higher costs for hardware implementation.

There is no approach that suits all tasks. The most successful approach is not equivalent to the best technique in this domain. It is important to strike a balance between accuracy and budgetary limits

2.3. Literature Review

A study proposed by Bhagampriyal et al. [16] aims to highlight the challenges associated with recommendation methods and introduce improved RS algorithms based on sales data. This study explores common RS algorithms such as association rules, CF, content-based filtering, as well as hybrid model recommendation. The researcher also conducted background research regarding grocery store RS. After completing the background research, a new product RS method is proposed using three algorithms which are the Apriori algorithm, the frequent pattern growth algorithm, and the popular product recommendation algorithm. Among all methods, the Apriori algorithm is the easiest model when it comes to implementation and is commonly utilized for establishing association rules. However, the Apriori algorithm requires a longer time for execution compared to Frequent Pattern (FP) Growth and Popular Product Recommendation methods. FP Growth algorithm is an expanded variant of Apriori algorithm. Consequently, Popular Product Recommendation can provide a quicker recommendation for customers. The primary drawback of the Apriori approach is the high cost of creating candidate sets, especially when working with complex patterns. Similar to this, the FP-growth method's primary drawback is the large number of candidates that are difficult to produce. Future projects involve integrating the FP-Tree with the Apriori data preprocessing method to overcome the limitations of both techniques.

Kumar et al. [17] proposed a three-part RS that can serve new users an excellent experience. The first part of the system is the Product Popularity System which can attract new users to the proposed system. As attracting new users nowadays is quite a challenging task, this method addresses this problem by giving new users access to a carefully selected selection of the most well-liked items on the marketplace. The second part of the system is the Model Based CF System which provides tailored recommendations for users. The third part is the Item-to-item-based RS which stands out as an essential resource for e-commerce ventures launching without prior user-item purchase data. K-Means Clustering algorithm is applied to perform textual clustering analysis to deliver recommendations that are relevant to the context. When a firm launches its e-commerce site without any previous user-item purchase history, K-Means clustering is very important. It makes it possible to group items according to their attributes and descriptions. The RS can display the product cluster after clusters are detected. Recommendations based on the content approach can also benefit from this approach. The evaluation metrics applied in this research project are click-through rates, user engagement, Silhouette Score, mean absolute error (MAE) and root mean square error (RMSE). According to the result, the Popularity-based suggestions have significantly increased user engagement and click-through rates. This shows that CF works well to increase the engagement of customers. In addition, K-Means clustering method has produced systematic product groupings that provide contextually appropriate user suggestions. The future work of this research project involves integrating the Natural Language Processing algorithm into the system to aid in further analysis of product descriptions and reviews.

A research project proposed by Rani et al. [18] uses CF to provide an interface for product RS. A product RS is utilized to provide consumers with recommendations according to prior purchases and search history. The proposed method consists of two methods for product recommendation. The first method uses CF algorithm and cosine similarity to produce recommendations. The system platform enables users to upload a comma-separated values (CSV) file which contains ratings and product details. This information is used to generate a pivot table and compute a cosine similarity matrix. This matrix aids in recognizing similar users according to their preferences. This is important in forming personalized product recommendations. The second method is advanced text analysis using Term Frequency - Inverse Document Frequency (TF-IDF) vectorization and K-means clustering algorithm. Similar to the prior method, the second method also supports users to upload a CSV file. TF-IDF vectorizer is applied to convert the product descriptions into numerical formats while K-means clustering is used to categorize similar products. By combining all three approaches into a single platform, the researchers have successfully built an effective approach that can serve multiple recommendation techniques. Future work planned for this study involves implementing user testing and getting feedback to further enhance the project.

In order to provide individualized recommendations for business clients, a study proposed by Long Li [19] presents a hybrid recommendation strategy that combines friend and product recommendations. The user similarity is computed using a combination of the user location similarity and user interest. The user location similarity refers to the physical location of users and the distance between locations. By involving the purchase history, the proposed method also can forecast what items users are interested in buying at a specific month. An empirical investigation of the suggested approach was carried out by adopting the fast-moving consumer goods e-commerce platform as an example. Offline tests were used to replicate the algorithm's performance. The evaluation metrics involved in this study include precision, recall, cover and recommended product popularity. The results demonstrate that the proposed strategy achieves an improvement in recall rate at 3.91% and an accuracy at 3.74%.

Satheesan et al. [20] proposed a study that develops a product RS that evaluates customers' preferences to suggest the most suitable products. The target users of this system include both existing and new users. For the new users, the system employs two approaches, which are suggesting generally popular items and analyzing the product descriptions for generating recommendations. As for the existing customers, they will receive recommendations through certain recommendation algorithms such as user-based CF, item-based CF, and association rule mining. These algorithms utilize the customer purchase history and priority ratings from fellow shoppers to produce recommendations. To begin, the algorithm identifies customer groups which share similar bought and reviewed products. It then aggregates products favored by customers with similar preferences while excluding bought or reviewed items. A survey is being carried out among individuals who prefer to buy items in supermarkets. This survey aims to collect and analyze their opinions regarding the product RS. Most participants responded that having a product RS is beneficial or potentially valuable to the supermarket. The result shows that the suggested approach successfully enhances customer shopping satisfaction by precisely and effectively suggesting personalized products tailored to their needs.

Cheedella K et al. [21] introduce a data-centric Product RS which is built using the Apache Spark ML libraries. This system takes advantage of the distributed computing ability of Apache Spark to provide tailored recommendations. The proposed method employs the Alternating Least Squares (ALS) and Singular Value Decomposition (SVD) models to build a CF based RS. The proposed system generates a set of forecasts for top ratings using the recommendation algorithm. The result indicates that the proposed method achieved a notable success, attaining the lowest RMSE value for SVD and ALS at 1.098 and 1.247 respectively. Consequently, ALS emerges as the better choice as a product recommender.

A paper proposed by Lourenco et al. [22] introduces a groundwork for future e-commerce studies focusing on new products related to the Covid-19 pandemic. The item recommended is expected to include gloves, hand sanitizers, and face masks. Due to the prolonged duration of this pandemic and the rising demand for these products, numerous vendors are creating innovation in these product categories to meet the consumers' needs. The proposed method includes two commonly used methods in RS, which are item-based CF and association rule mining. The dataset employed in this study is Amazon review data which is related to cell phones and accessories. The Apriori algorithm is used to uncover patterns in transaction history for association rule mining, while the item-based CF utilizes a correlation matrix to identify similar products. This work represents an intelligent data mining framework scalable for big data applications in the e-commerce field. Furthermore, the evaluation of this proposed method is planned to be conducted after applying the Covid-19 related datasets. Future work of this study includes applying sentimental analysis in the textual information of customer reviews and conducting research in advanced ML techniques to enhance performance.

As many studies extensively explored various RS algorithms, this research focuses on a comparative performance evaluation of three ML techniques and ensemble approaches in the context of e-commerce. Most existing studies only focus on an individual or a limited set of algorithms while the proposed method in this paper has an advantage in providing a comprehensive analysis of multiple algorithms and their ensemble. This study not only highlights the advantages and disadvantages of each technique but also provides insight regarding the potential benefits of ensemble methods in improving recommendation accuracy and user satisfaction.

The literature reveals that hybrid and ensemble methods have the potential to address the limitations of a single algorithm. For instance, Bhagampriyal et al. [16] and Satheesan et al. [20] demonstrate the benefits of integrating multiple techniques in enhancing the model accuracy and user satisfaction. Another study by Kumar et al. [17] and Cheedella et al. [21] shows that the scalability and efficiency in RS play a vital role in maintaining model performance. Algorithms such as K-Means and distributed computing frameworks like Apache Spark contributed significantly to addressing this issue.

Most studies suggest that model-based CF outperforms other techniques in recommending e-commerce items. This is because model-based CF considers the perspectives of both users and items, unlike alternative methods. Despite many researchers promoting different techniques in building a RS, there are still gaps remaining to achieve the optimal result. To illustrate, Kumar et al. [17] and Lourenco et al. [22] suggest a need for further research in adopting NLP technique and sentimental analysis to enhance their RS.

This study aims to show the potential of building an ensemble method to combine the strengths of individual approaches compared to traditional approaches. By examining the efficiency, expandability, and suitability of these techniques in an e-commerce environment, the proposed method tends to aid in the creation of stronger and more reliable recommendation systems.

2.4. Proposed Framework for E-Commerce System

A framework specifically designed for RS in the e-commerce domain is presented in this paper. Many efforts are being made to create flexible frameworks which can be used in a variety of contexts. Figure 3 shows the prototype implementation flowchart.

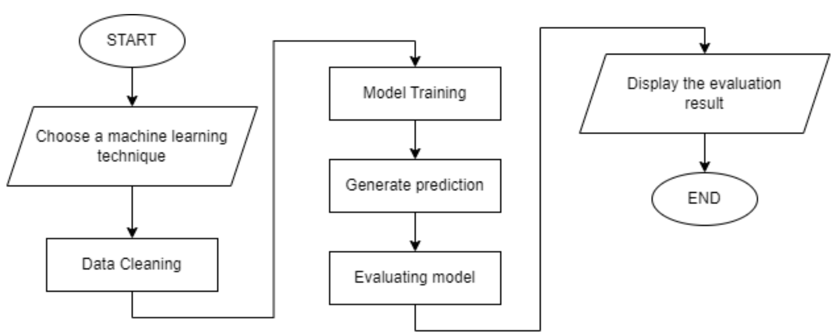


Figure 3. Prototype Implementation Flowchart

During the implementation phase, the prototype is expected to carry out several preliminary activities before generating suggestions. The prototype will first choose a ML model for the application. Data cleansing will then take part to ensure the data is flawless and prepared for further processing. Next, the prototype will conduct model training using the ML model chosen in the previous steps. The model evaluation phase will also be included in the implementation phase to determine the best ML model in terms of accuracy and effectiveness. The assessment results from the prototype will then be presented and compared with various ML techniques in the following phase. In addition, a graphical user interface is integrated into the system to improve the visualizing experience.

1) Deciding the ML Techniques: This research study offers four choices for users to decide what type of ML to use, which include k-Nearest Neighbor Baseline (KNN Baseline), SVD, CoClustering, and an ensemble method of all three techniques. These methods find extensive application in the e-commerce recommendation field. KNN Baseline can maintain precision and accuracy by leveraging local minima of the target function alongside a baseline to identify unknown functions. SVD enables dimensionality reduction, breaking down data into smaller segments for easier interpretation and analysis. CoClustering provides insights into the relationship between rows and columns, particularly useful in high-dimensional datasets.

2) Dataset: The dataset utilized is referred to as the Amazon Dataset. This is a public dataset available on the Kaggle website which contains 21 columns and 17075 rows of data record. Table 2 shows the explanation for each column.

Table 2. Column Explanation in the Dataset

No	Column	Explanation
1	id	The ID of the product
2	name	The name of the product
3	asins	The special code of the product
4	brand	The brand of the product
5	categories	The categories which the product belong to
6	keys	The product keys
7	manufacturer	The manufacturer of the product
8	reviews.date	The date which creates review
9	reviews.dateAdded	The date which add review
10	reviews.dateSeen	The date which seen review
11	reviews.didPurchase	The date which customer purchase the item
12	reviews.doRecommend	The intent of user to recommend the rated item to other users
13	reviews.id	The id of review
14	reviews.numHelpful	The number of helpful comments from other users of review
15	reviews.rating	The rating of reviewed item
16	reviews.sourceURLs	The source url of the review
17	reviews.text	The review text
18	reviews.title	The review title
19	reviews.userCity	The city of the customer which create review
20	reviews.userProvince	The user province of the review
21	reviews.username	The username of user which create the review

3) Data Cleaning: The dataset cleaning process starts with handling missing data present in the dataset. Next, a few columns which are irrelevant to observations such as asins, brand, keys, manufacturer and the url of the source data. The data transformation is also conducted. For example, the date of review is transformed to DateTime format. Some textual content is also cleaned, such as removing the additional spaces in the product name. To maintain the model accuracy, only the users who rated more than 5 times are kept in the dataset.

4) Recommender Engine: When the dataset is cleaned, the user-item matrix is ready to be integrated into the RS. This study enables users to efficiently construct a rating-based RS using a common library named Surprise. Moreover, this module offers several pre-existing functionalities tailored for regression tasks, including a train-test split function and various evaluation metrics for accuracy.

3. RESULTS AND DISCUSSION

The prototype dashboard is displayed by default. The users encounter a selection of 10 recommended items on this page. Each recommendation provides essential information regarding the item such as the item name, product ID as well as the average rating from reviewers. Figure 4 illustrates a sample dashboard of the prototype.

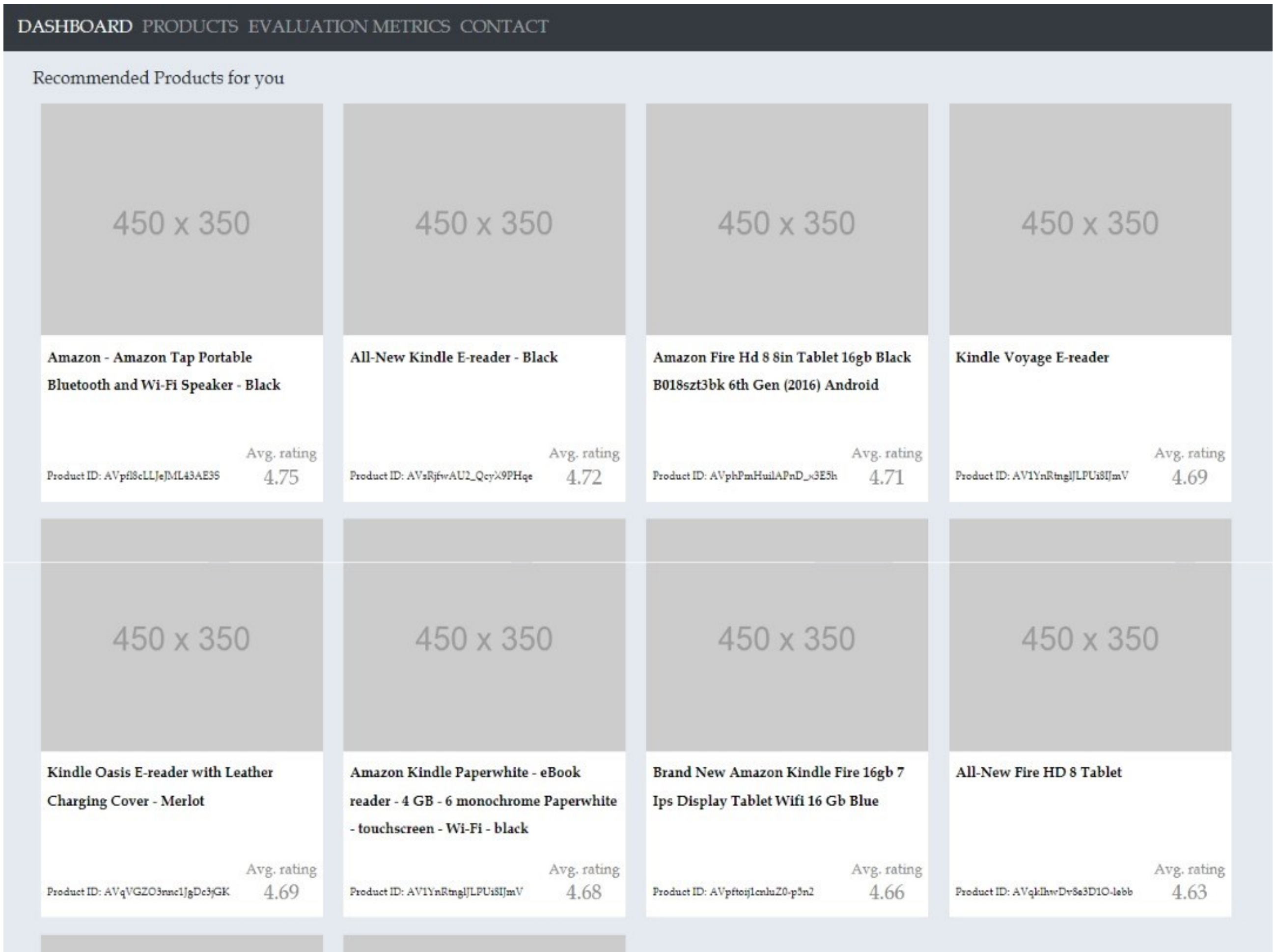


Figure 4. Prototype Dashboard

In this research project, the performance of the model is assessed using evaluation metrics such as MAE, RMSE, fraction of concordant pairs (FCP), and normalized discounted cumulative gain (NDCG). Smaller values of MAE and RMSE indicate higher accuracy, while a higher value of FCP and NDCG suggests better model performance ranking.

According to the results, the ensemble method yields the best performance, with RMSE at 0.557, MAE at 0.388, FCP at 0.465 and NDCG at 0.991. The following highest performance is the SVD model, with RMSE at 0.72, MAE at 0.556, FCP at 0.456 and NDCG at 0.995. The third best performance model is CoClustering, with RMSE at 0.847, MAE at 0.581, FCP at 0.497 and NDCG at 0.992. The lowest performance technique achieved is the KNN Baseline, with RMSE at 0.847, MAE at 0.591, FCP at 0.465 and NDCG at 0.992. Figure 5 and Table 3 demonstrate the results from the prototype.

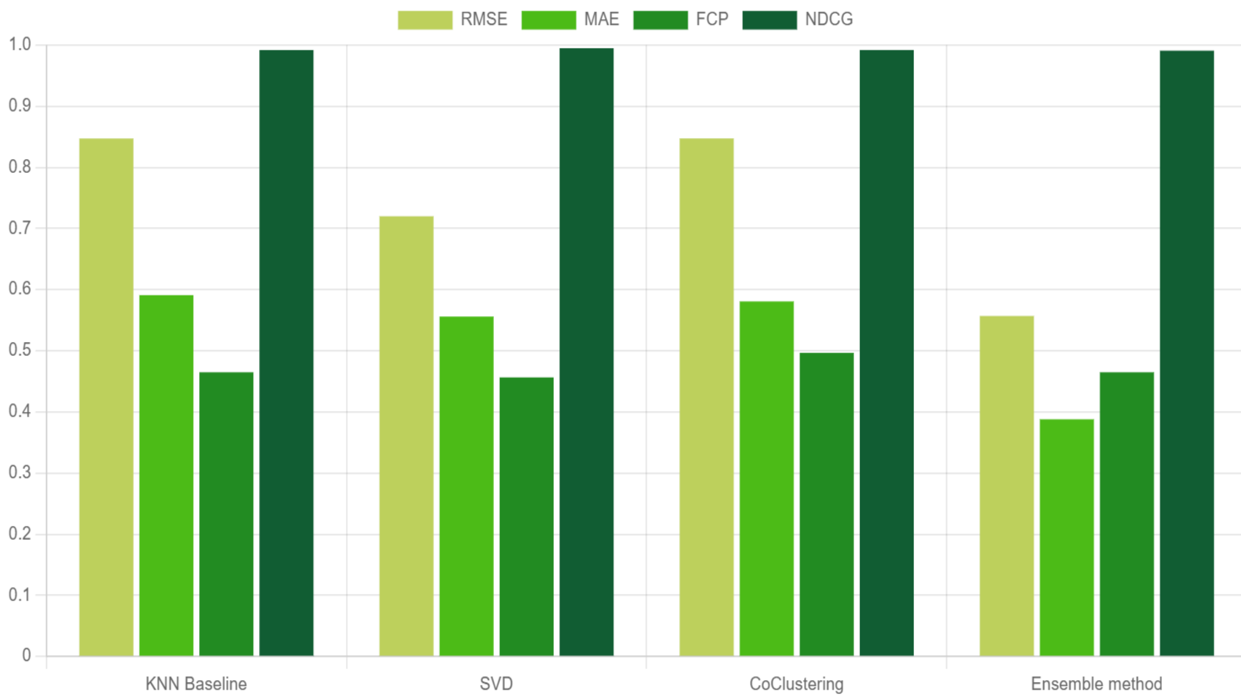


Figure 5. Evaluation Result

Table 3. Evaluation Result

No	ML	RMSE	MAE	FCP	NDCG@100
1	KNN Baseline	0.847	0.591	0.465	0.992
2	SVD	0.72	0.556	0.456	0.995
3	CoClustering	0.847	0.581	0.497	0.992
4	Ensemble method	0.557	0.388	0.465	0.991

The duration of model training is also recorded, revealing the differences in time among various methods. The KNN Baseline model demonstrates the quickest training time, which is 0.009 minutes, outperforming all other approaches. The second-best model is the SVD model, taking 0.069 minutes to fit. The third best model is the CoClustering model, which needs 0.467 minutes for model fitting. Conversely, the ensemble method technique exhibits the lengthiest training time, totaling 0.558 minutes. Figure 6 and Table 4 illustrate the result of model fitting time.

The time required for model training is crucial in selecting the optimal model. Among the methods considered, the KNN Baseline model stands out for its shortest fitting time, followed closely by the SVD model and CoClustering model. The ensemble method requires the longest time to fit. This is due to the KNN Baseline can store the entire dataset in memory, hence it can perform quick lookup operations during forecasting. As the dataset size grows, the time gap in computational demands between each method becomes wider.

The result reveals that the SVD model outperforms other models which include KNN Baseline, CoClustering, and the Ensemble method. In terms of model accuracy, the ensemble method delivers the best result. Although this method has a higher accuracy in predicting the customer preference, this technique requires a much longer model fitting time compared to SVD. This indicates that the ensemble method requires a strong computational resource and cannot adapt to customer changing preference in the real-time system. SVD as a simpler model can be easily implemented, understood, and maintained compared to the ensemble method which combines multiple algorithms. As for the ensemble method, the need for a larger computational resource will lead to a higher requirement in hardware for the company.

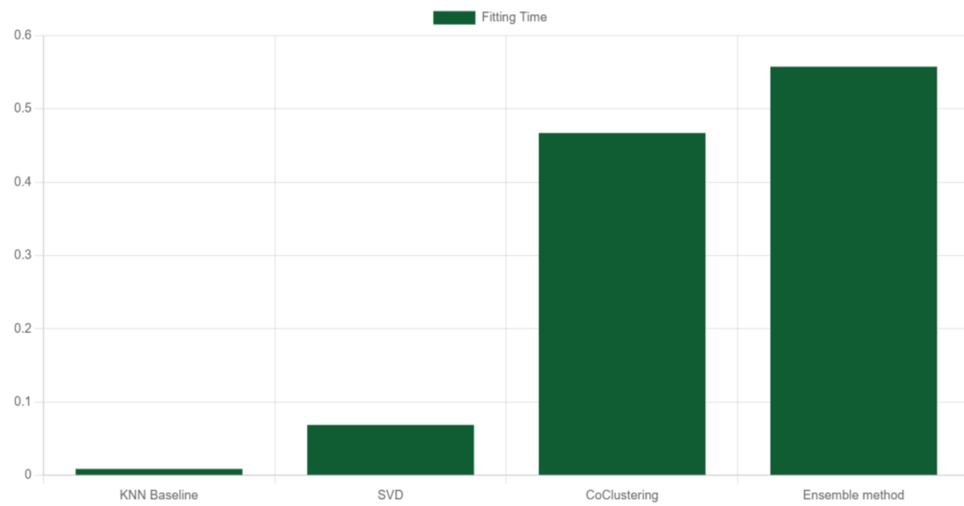


Figure 6. Model Fitting Time

Table 4. Model Fitting Time

No	ML Technique	Model Fitting time
1	KNN Baseline	0.009
2	SVD	0.069
3	CoClustering	0.467
4	Ensemble method	0.558

As for the model fitting time, the result shows that KNN Baseline is the quickest model. Despite the shortest model fitting time, this model delivers the lowest accuracy among all techniques. This may be due to KNN functionality as it only focuses on the direct similarities among users and not being able to detect the complex relationships between users and items. On the other hand, SVD can fulfill this task with ease and is able to discover the underlying factors that influence user preferences. SVD is able to decompose the user-item interaction matrix into latent factors, hence being able to detect deeper patterns and relationships that KNN Baseline would have missed. Besides that, SVD is very effective when dealing with sparse datasets which are common in RS. There are many users who contribute less ratings and comments, which leads to high sparsity in dataset. SVD has the ability to reduce the dimensionality of the data and detect the most significant latent factors that can explain the interaction between users and items. These results can contribute to a better generalization and higher accuracy of recommendation.

In simple words, SVD demonstrates a relatively high accuracy in rating prediction. This model also exhibits moderate training time compared to the quickest technique, which is KNN Baseline. In addition, SVD technique requires less memory during the training process compared to other methods, which helps this model to further stand out among other methods.

4. CONCLUSION

The primary aim of this research is to obtain the fundamental concept of techniques employed in RS. By conducting an extensive literature review, diverse recommendation approaches are examined and deliberated upon. Subsequently, a prototype is developed and integrated into the CF approach, in conjunction with various ML techniques such as KNN Baseline, SVD, and CoClustering. These methodologies are then assessed using metrics like RMSE, MAE, FCP, and NDCG. The results indicate that the SVD model demonstrates the most effective performance as this technique establishes a comparatively high rating prediction accuracy as well as a reasonable training time when compared to other methods. This method not only evaluates the traditional and advanced RS algorithms but also explores the potential of ensemble methods by combining the strengths of individual approaches. This has addressed the performance, scalability, and applicability of these methods under e-commerce domain. As a result, this research has substantially enhanced the understanding within the RS field.

Future work will include enhancing the data visualization of the prototype. While the present investigation focuses solely on constructing e-commerce product recommendations, the depth of data visualization remains limited. Future improvements should strive to integrate more data visualization techniques. Subsequent efforts could explore more about the SVD method with other approaches to enhance the efficacy of models.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for the careful review of our manuscript.

FUNDING STATEMENT

This work is supported by the funding of TM R&D from the Telekom Malaysia, Malaysia (MMUE/240068).

AUTHOR CONTRIBUTIONS

Wan-Er Kong: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Tong-Ern Tai: Project Administration, Writing – Review & Editing;
Palanichamy Naveen: Project Administration, Supervision, Writing – Review & Editing;
Heru Agus Santoso: Conceptualization, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests was disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

REFERENCES

[1] C. Campbell, S. Sands, C. Ferraro, H.-Y. (Jody) Tsao, and A. Mavrommatis, "From data to action: How marketers can leverage AI," *Bus Horiz*, vol. 63, no. 2, pp. 227–243, 2020. doi: 10.1016/j.bushor.2019.12.002.

[2] D. Ferreira, S. Silva, A. Abelha, and J. Machado, "Recommendation System Using Autoencoders," *Applied Sciences*, vol. 10, no. 16, 2020, doi: 10.3390/app10165510.

[3] S. S. Kundu, D. Sarkar, P. Jana, and D. K. Kole, "Personalization in Education Using Recommendation System: An Overview," pp. 85–111, 2021. doi: 10.1007/978-981-15-8744-3_5.

[4] L. Quijano-Sánchez, I. Cantador, M. E. Cortés-Cediel, and O. Gil, "Recommender systems for smart cities," *Inf Syst*, vol. 92, 2020, doi: 10.1016/j.is.2020.101545.

[5] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities," *Applied Sciences*, vol. 10, no. 21, 2020, doi: 10.3390/app10217748.

[6] S. Ben Atitallah, M. Driss, W. Boulila, and H. Ben Ghezala, "Leveraging Deep Learning and IoT big data analytics to support the smart cities development: Review and future directions," *Computer Science Review*, vol. 38, 100303, 2020. doi: 10.1016/j.cosrev.2020.100303.

[7] Z. Xia, A. Sun, J. Xu, Y. Peng, R. Ma, and M. Cheng, "Contemporary Recommendation Systems on Big Data and Their Applications: A Survey," *arXiv,* 2022.

[8] M. Cakir, Ş. G. Oguducu, and R. Tugay, "A Deep Hybrid Model for Recommendation Systems," *arXiv*, pp. 321–335, 2019. doi: 10.1007/978-3-030-35166-3_23.

[9] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput Sci, vol. 2, no. 3, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.

[10] Y. Afoudi, M. Lazaar, and M. Al Achhab, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network," *Simul Model Pract Theory*, vol. 113, p. 102375, 2021, doi: 10.1016/j.simpat.2021.102375.

[11] P. Bahrani, B. Minaei-Bidgoli, H. Parvin, M. Mirzarezaee, and A. Keshavarz, "A hybrid semantic recommender system enriched with an imputation method," *Multimed Tools Appl*, vol. 83, no. 6, pp. 15985–16018, 2023, doi: 10.1007/s11042-023-15258-4.

[12] J. Latrech, Z. Kodia, and N. Ben Azzouna, "CoDFi-DL: a hybrid recommender system combining enhanced collaborative and demographic filtering based on deep learning," *J Supercomput*, vol. 80, no. 1, pp. 1160–1182, 2024, doi: 10.1007/s11227-023-05519-2.

[13] L. Chen, R. Xiong, and Y. Ji, "Application of SVM model based on collaborative filtering hybrid algorithm in e-commerce recommendation," *International Journal of Computers and Applications*, vol. 46, no. 5, pp. 292–300, 2024, doi: 10.1080/1206212X.2024.2309809.

[14] J. Singh, K. Steele, and L. Singh, "Combining the Best of Online and Face-to-Face Learning: Hybrid and Blended Learning Approach for COVID-19, Post Vaccine, & Post-Pandemic World," *Journal of Educational Technology Systems*, vol. 50, no. 2, pp. 140–171, 2021, doi: 10.1177/00472395211047865.

[15] A. S. Lampropoulos, D. N. Sotiropoulos, and G. A. Tsihrintzis, "Evaluation of a Cascade Hybrid Recommendation as a Combination of One-Class Classification and Collaborative Filtering," in *International Conference on Tools with Artificial Intelligence*, IEEE, pp. 674–681, 2012. doi: 10.1109/ICTAI.2012.96.

[16] B. M, G. R, J. Johnson, J. F. Lilian, and S. P, "Recommendation Systems for Supermarket," *International Conference on Innovative Practices in Technology and Management (ICIPTM)*, IEEE, pp. 1–6, 2023. doi: 10.1109/ICIPTM57143.2023.10117637.

[17] M. R. Kumar, S. Vishnu, G. Roshen, D. N. Kumar, P. Revathi, and D. R. L. Baster, "Product Recommendation Using Collaborative Filtering and K-Means Clustering," *IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, IEEE, pp. 1722–1728, 2024. doi: 10.1109/IC2PCT60090.2024.10486625.

[18] G. A. Rani, B. U. Sri, S. S. Deshai, S. N. Bachupally, V. K. R. Patlolla, and P. V Kumar, "E - Commerce Product Recommendation System using Collaborative Filtering," *International Conference on Edge Computing and Applications (ICECAA)*, IEEE, pp. 1521–1525, 2023. doi: 10.1109/ICECAA58104.2023.10212422.

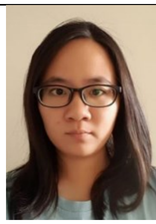
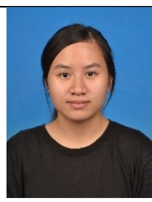
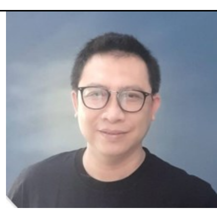
[19] L. Li, "Research on Personalized Recommendation System for E-Commerce Products Based on Collaborative Filtering Algorithm," *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, IEEE, pp. 876–880, 2024. doi: 10.1109/EEBDA60612.2024.10485710.

[20] P. Satheesan, P. S. Haddela, and J. Alosius, "Product Recommendation System for Supermarket," *IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, pp. 930–935, 2020. doi: 10.1109/ICMLA51294.2020.00151.

[21] K. Cheedella, S. Fathimabi, and D. Chinamuttevi, "Amazon Product Recommendation System Using Apache Spark," *International Conference on Cloud Computing, Data Science & Engineering (Confluence),* IEEE, pp. 223–227, 2024. doi: 10.1109/Confluence60223.2024.10463211.

[22] J. Lourenco and A. S. Varde, "Item-Based Collaborative Filtering and Association Rules for a Baseline Recommender in E-Commerce," *IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 4636–4645, 2020. doi: 10.1109/BigData50022.2020.9377807.

BIOGRAPHIES OF AUTHORS

	**Wan-Er Kong** is a master's student at Multimedia University, currently pursuing a Master of Science. Holding a degree in Computer Science with a specialization in Data Science from the same university, Wan Er is passionate about exploring the field of Data Science. Her research focuses on Machine Learning and recommender systems. Driven by a desire to contribute to society, she is dedicated to advancing her knowledge and skills to make meaningful impacts in the world through Data Science. She can be contacted by email: KONG.WAN.ER@student.mmu.edu.my.
	**Tong-Ern Tai** is currently enrolled in the Master of Science program at Multimedia University. She holds a bachelor's degree in computer science with a specialization in Data Science from the same university. She is deeply passionate about Data Science. Her research interests include machine learning, ticket resolution time prediction, and recommender systems. She can be contacted by email: TAI.TONG.ERN@student.mmu.edu.my.
	**Palanichamy Naveen** joined the Faculty of Computing and Informatics, Multimedia University after receiving Ph.D from Curtin University, Malaysia. She received her Bachelor of Engineering (CSE) and Master of Engineering (CSE) from Anna University, India. Her research interest includes Smart Grid, Cloud Computing, Machine Learning, Deep Learning and Recommender system. She is involved in multiple research projects funded by Multimedia University. She can be contacted by email: p.naveen@mmu.edu.my.
	**Heru Agus Santoso** is a faculty member at Dian Nuswantoro University in Semarang, Indonesia. His research interests primarily include ontology, information retrieval, text mining and deep learning. He has published several publications in these areas, showcasing contributions in data and knowledge systems. He can be contacted by heru.agus.santoso@dsn.dinus.ac.id.