International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 6
ISSN: 2088-5334

E-learning Recommender System based on k-Nearest Neighbor (KNN), Singular Value Decomposition (SVD), and CoClustering Approaches

Wan-Er Kong [a], Su-Cheng Haw [a,*], Tong-Ern Tai [a], Palanichamy Naveen [a], Kok-Why Ng [a]

[a] *Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia*
*Corresponding author: *sucheng@mmu.edu.my*

***Abstract*—The advent of the digital age resulted in the development of the e-learning platform as a helpful resource for top-notch e-learning materials. Despite its promise, the entire scope of its potential hasn't been ultimately discovered. E-learning platforms are expected to offer information that meets customers' requirements and interests to draw users and boost profits. These suggestions are created by considering various aspects, including previous browsing habits, purchase history, demographic data, and others. E-learning platforms may improve the quality of the learning process by offering users interesting information tailored to their specific requirements and preferences by utilizing cutting-edge technology. This paper explores Machine Learning (ML) approaches used in e-learning content recommender platforms. Three ML approaches are chosen and applied to predict user interest and need Singular Value Decomposition (SVD), k-nearest Neighbor Baseline (KNNBaseline), and CoClustering. These ML approaches boost user experience by analyzing data trends and patterns of usage to deliver insights into the best customized and appropriate educational materials for every user. The previous user ratings trend is employed to derive the item ratings. Mean absolute error (MAE) and root mean square error (RMSE) are assessed to evaluate the implemented techniques' effectiveness.**

***Keywords*— E-learning; recommendation system; machine learning; filtering technique; collaborative filtering.**

Manuscript received 21 Dec. 2023; revised 16 Apr. 2024; accepted 22 Aug. 2024. Date of publication 31 Dec. 2024.

I. INTRODUCTION

With the rise of websites such as Amazon, YouTube, and Netflix, recommendation algorithms are becoming increasingly common daily. A recommender system (RS) is a program that makes relevant content recommendations to users using Big Data technologies [1]. RS are frequently employed by businesses to predict the preferences and interests of customers according to customer past data [2]. They can point customers to any item or service that grabs their attention.

As the age of digitization grows, people use a variety of platforms to produce a remarkable quantity of data throughout their everyday lives [3]. It is widely recognized that by utilizing data science knowledge, we can mine this immense quantity of data for useful information, for example, determining a particular group of people's favored restaurants, creating business strategy plans, and predicting a company's next big product [4]. In the absence of this technology, users might perceive an abundance of information at their hands to be rather disorganized and confusing, resulting in a less pleasant experience [5]. Every field would experience this; education is no exception. RS plays a critical role in e-learning by helping to mitigate the problem of data overload [6], [7], [8].

A RS is one of the subclasses of information filtering systems that employ the system's algorithm to provide users with customized recommendations based on their interests [9]. Using the information gathered from the user data, the algorithm determines and displays the most favored good or service from the vast range of options. By employing the Machine Learning (ML) technique, people possess the capacity to build a wide range of RS that assist consumers in making choices and offer them a great experience.

Recommendation algorithms are being widely employed in major corporations like Google and Meta. For instance, the popular content-sharing website YouTube uses a RS to assist viewers in finding content likely to be interesting. As a worldwide online retailer, Amazon offers a RS to give customers a selection of items they may be highly interested in. One of the most popular suggestion methods is the randomized advertising that appears when browsing the Internet. A RS is widely implemented to enhance user and platform engagement to achieve a higher advantage.

The generated recommendation could originate from various sources, such as browser history, past purchases, demographic data, and more [10]. These suggestions are created using information gathered from clients. If the recommendation algorithm is developed effectively, this approach can result in immense earnings and set a business apart from rivals.

ML enables computers to generate knowledge autonomously with the aid of historical data [11], [12]. Using various methods, ML builds mathematical models and projections based on historical data or experience. Numerous fields can benefit from applying this technique, such as future sales forecasting, ticketing time resolution, and image recognition [13], [14], [15].

This study covers a comprehensive understanding of ML, including topics such as the objective of ML and the features of various ML techniques, such as their benefits, drawbacks, and applicability in a RS for online courses. This information may be acquired by exploring research articles and summarizing their findings.

II. MATERIALS AND METHOD

A. *Stages of the RS*

One of the basic requirements for building a good RS is sufficient data to investigate and learn about a user's characteristics and behavior. The amount of data could significantly influence the recommendation delivered to users [16]. Low-quality data will only produce flawed output, no matter how good and capable the RS is. The RS comprises three phases: information gathering, data learning, and prediction. Fig. 1 shows the three stages of the RS.

RS relies on various inputs, such as response, which considers user interests in products, and indirect feedback, which infers user preferences by observing user behavior. Direct and indirect feedback can be combined to provide hybrid feedback. A RS cannot produce proper and accurate results without a correctly built user model.

Getting implicit feedback from users is a popular method of gathering user data. This data can be obtained in many ways, including user's past purchases, links clicked, and time spent on websites and emails. Rather than requiring user input, these feedback methods enable the RS to interpret independently and produce results. Nevertheless, in explicit feedback, the system requests users to input, such as purchased item ratings, to create suggestions. This method necessitates more incredible work from the user during feedback compared to implicit input. When both implicit and explicit feedback are combined, the disadvantages of each are removed to provide hybrid feedback, which is superior to both types of input [17]. Allowing users to provide ratings and direct comments while utilizing indirect data as a recommendation element can yield this feedback.

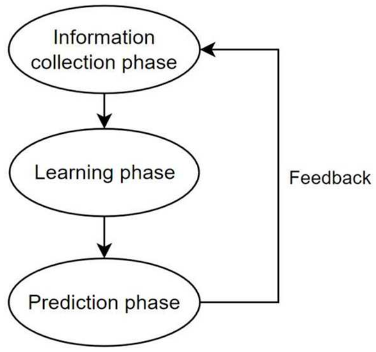

Fig. 1 Process of generating a recommendation

The first phase involves collecting the obtained information from users. The information is studied and further analyzed using the system. Next, the second phase will involve learning the user interest using a learning algorithm. Learning algorithms are methods that help recognize the underlying patterns in data [18]. Lastly, the final stage is the prediction stage. This stage uses the trained learning algorithm to produce recommendations for users.

B. *RS Techniques*

Practical and precise recommendation strategies are necessary for a RS to provide reliable recommendations to every user. This illustrates how important it is to understand the traits and applicability of different RS. Fig. 2 indicates the commonly used recommendation approaches in research papers.

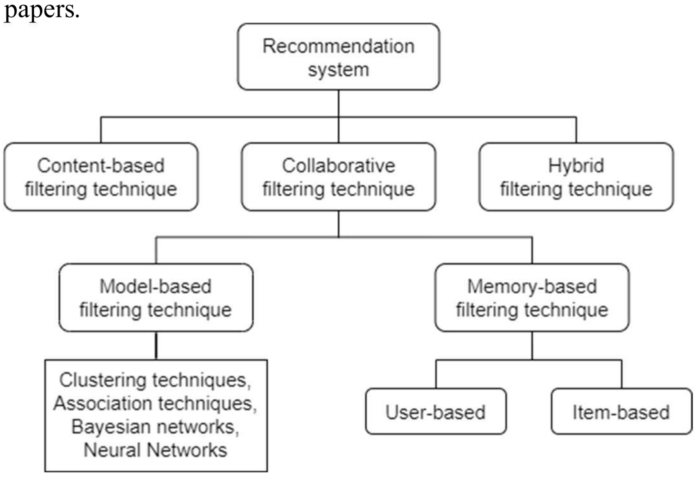

Fig. 2 Common Recommendation Approaches

1) Content-based Filtering Technique: To provide precise predictions to users, the content-based approach employs an algorithm that considers the analysis of item properties. The content-based filtering technique recommends web pages, journals, and other media types. This approach generates recommendations using user profiles and characteristics extracted from past records [19]. Item recommendations that correlate significantly with the highly-rated content are suggested to the user. Two approaches are employed to accomplish this method of work: the classification model and vector spacing.

2) Vector spacing method: As one of the approaches in content-based RS, the vector spacing method organizes ranks of items depending on user input [20]. The steps involved are developing user and item vectors and utilizing dot product computation to ascertain item appropriateness. For instance, if a user favors horror movies, the system suggests similar ones. The ranking of dot products sorts the available movies and recommends the top and highest rank to the user.

3) Classification method: This method uses decision trees as a categorization technique to generate item suggestions [21]. final conclusion will be reached through a series of connected conditions in the decision tree. For example, the algorithm first matches the movie according to the user's preferences and movie genre before recommending it. It then evaluates the movie's rating; if other viewers give it a low rating, the system decides not to suggest it.

4) Collaborative Filtering Technique: Collaborative filtering is commonly used for insufficient metadata content such as videos and audio [22]. To generate recommendations, a database of user preferences is compiled to match users with identical interests. Using the information extracted from user profiles, user recommendations are then generated. There exist two sub-branches of collaborative filtering methods: memory-based and model-based.

5) Memory-based method: The memory-based method is the simplest solution because this method does not involve any models in the technique [23]. This method functions by utilizing past data through a nearest-neighbor distance-measurement approach. It identifies user groups with similar preferences to suggest high-rating items they favor. This method can be split into two approaches: user-based and item-based collaborative filtering. The user-based approach matches users with identical ratings on the same items, while the item-based approach matches items with identical ratings on the users.

6) Model-based method: A model-based approach presupposes the existence of an underlying model that precisely corresponds to the estimated results. This technique commonly uses matrix factorization to simplify user-item matrices, and boost algorithm efficiency while decreasing memory consumption and calculation time [24]. Matrix factorization has applications in many fields, such as RS and image recognition. This method works especially well for sparse matrices, which are prevalent in recommendation scenarios when consumers rate only a restricted quantity of items.

7) Hybrid Filtering Technique: Hybrid filtering techniques combine different methods. Here are several types of hybridization recommendation technique combinations: cascade, feature augmentation, feature combination, meta-level, mixed, switching, and weighted [20].

8) Cascade hybrid: A primary model is used by a hierarchical RS in the Cascade Hybrid framework to generate the main output, while a secondary model is used to handle minor difficulties arising from the original result [20]. Since sparse datasets are common, the secondary suggestion project helps resolve problems resulting from missing data or comparable scores.

9) Feature augmentation strategy: The feature augmentation strategy evaluates and classifies user and object profiles to increase accuracy of recommendation [20]. The generated ratings or categories inform the suggestion procedure, which provides anticipated results. The hybrid feature augmentation enhances system efficiency without changing the fundamental recommendation model.

10) Feature combination strategy: The feature combination technique improves the original profile dataset by introducing a virtual recommendation model by using feature engineering [20]. This strategy eases incorporating some aspects from other recommendation models into the primary model.

11) Meta-level hybrid: The meta-level hybrid method requires a secondary model to enhance the primary recommendation model by offering an improved dataset [25]. A fully trained model from the secondary model creates the dataset in place of the original dataset. This new dataset is then adopted to refine the main recommendation model.

12) Mixed hybrid: The mixed hybrid method generates many candidate datasets according to the attributes and profiles of users [20]. Candidate datasets are then fed into the recommendation model to generate a set of recommendations. Using this strategy, a partial dataset is fitted to the model to improve performance.

13) Switching hybrid: The hybrid method is unique because it can switch between several RS. The recommendation model incorporates an extra layer to provide this switching feature. It helps the system choose the best model for the given circumstances and easily integrates it into the system.

14) Weighted hybrid: The weighted hybrid method enables the recommendation process to employ numerous models simultaneously. It integrates results from many models with a consistent weight that is the same for training and evaluation [26].

Earlier sections covered prevalent RS strategies employed across industries. However, each approach has unique characteristics, making it difficult to determine the best technique for a given task. Table I outlines the pros and cons of each approach to illustrate the distinctions.

Table I makes it evident that every recommender approach is distinct and has pros and cons of its own. Since content-based filtering relies on user ratings, the cold-start problem does not affect it. In contrast, collaborative filtering relies on user data to recommend highly rated things to users. As the hybrid filtering method incorporates many techniques, it may be more accurate but needs more processing power, and hardware implementation might be costly.

TABLE I
ADVANTAGES AND DISADVANTAGES OF RS TECHNIQUES

Techniques	Advantage	Disadvantage
Content-based Filtering	• The cold start issue is resolved since recommendations are personalized for each user and do not rely on the data of other users • Scalability for huge user bases is made easier.	• Domain expertise is required as the system relies on manually crafted item feature representations. • Limited by hand-engineered components

Techniques	Advantage	Disadvantage
	• Can identify a user's preference and produce recommendations that perhaps other users won't find interesting.	• Limited ability to utilize the prior interests of users.
Collaborative Filtering	• Does not need domain knowledge since the embeddings are dynamically learnt • Can assist users in finding new interests • To some extent, the system may use the feedback matrix alone to train a matrix factorization model. Since no particular contextual attributes are required, the system could be employed as a candidate generator.	• Have cold start issues. • Hard to apply side features for items due to data sparsity problems
Hybrid Filtering	• Integrates the pros of multiple filtering techniques • It can give the highest accuracy in recommendations for users.	• The model is too complex when employed in simple datasets. • It may require expensive hardware for application.

It is concluded that the most successful approach differs from the best technique in this domain. Each approach has pros and cons, and no one approach can suit all tasks. Striking a balance between accuracy and budgetary limits is essential.

C. Literature Review

According to research conducted by Tarus et al. [27] developed a hybrid recommendation method that combines CF algorithms, sequential pattern mining (SPM), and context awareness. This hybrid combination helps learners find relevant learning resources. The suggested hybrid recommendation algorithm uses CF to provide recommendations based on contextual data. Context awareness is also adopted by considering the contextual information about learners. In addition, a Generalized Sequential Pattern (GSP) algorithm is used in data mining in weblogs to discover the learners' sequential access patterns. The recall, precision, and F1 measure values are compared to determine the accuracy of the given model. Research findings indicate that the suggested recommendation approach performs better in terms of precision and the quality of suggestions. In addition, by using contextual data and learner sequential access patterns to provide predictions without overlapping learner ratings, the proposed hybrid technique will be able to resolve data sparsity issues.

In a different study, Aher and Lobo [28] introduce a Course RS to suggest courses for students based on their choice groups. This study uses data gathered from Moodle, an open-source learning platform. The RS proposed uses a combination of clustering techniques and association rule algorithms. In precise, Simple K-means and Apriori were used in this approach. A comparison was made between the outcomes generated by the Apriori algorithm and existing algorithms, including results within the Weka data mining tool and other combinations of clustering and association rule algorithms. Integration of the algorithm using Simple K-means clustering and the Apriori association rule algorithm attained better results than the Apriori algorithm alone and other combinations of clustering and association rule algorithms. In summary, the proposed course RS effectively assists students in identifying optimal course combinations based on their preferences, which leads to this approach being beneficial to distance learning students.

Thanh-Nhan et al. [29] employed various approaches in generating course RS that can predict students' learning performance and further aid students in forming a better study strategy. Four methods are employed in this research project: baseline predictors, KNN CF, standard matrix factorization in encoding the latent factors, and biased matrix factorization. These techniques are important in overcoming user and item bias problems. A dataset of student grades from Can Tho University is employed in this study. After performing the prediction, the data is put in a grading matrix and transferred to a web application to form recommendations further. In this research, the RMSE serves as a measure of model accuracy. In conclusion, the biased matrix factorization algorithm exhibits the smallest RMSE, at 0.831, indicating high precision and is thus selected for implementation in the RS framework. Future work will enhance the models by improving the model's forecast accuracy.

Obeidat et al. [30] introduced a CF method for online courses RS, which proposes courses to students based on their interests. This approach utilizes data mining methods to find underlying patterns among courses. The study demonstrates that grouping students into similar clusters based on their course preferences significantly improves the quality of association rules compared to rules generated using the entire dataset. The Apriori algorithm is employed twice: first with the complete dataset and then with the clusters of student course preferences. Evaluation metrics such as coverage, support, and confidence are used in this research project. The results show that rules developed on clusters exhibit higher coverage. In addition, the SPADE method is applied to course sequences to assess the influence of course dependency on recommendations, with results consistent with those obtained using the Apriori algorithm.

Zhao and Pan [31] introduced an improved model for recommending online e-learning courses. The model proposed generates customized recommendations by integrating collaborative filtering algorithms with user behavior data and article analysis. To address the cold start issue, diverse data collection methods aligning with specific interest points are employed. The model computes the user's preference degree and item similarity to enhance the recommendation efficiency. The model performance evaluation is conducted using recall and precision metrics. The proposed method outperforms traditional collaborative filtering models, both user and item-based, in terms of precision and recall, which indicates that a simple ensemble of ML algorithms can generate effective recommendations.

Another study by Ali et al. [32] introduced a RS that integrates CF and clustering techniques to offer e-learning content suggestions tailored to users' interests, thereby enhancing user experience. This method organizes learners into similar groups to extract relevant data effectively. Initially, students are grouped based on their characteristics, performance, and interaction with the system. Next, the system selects learning materials from these clusters that haven't been explored previously. Finally, it generates a list of relevant e-learning courses using two sorting methods and recommends them to users. Moreover, this approach addresses the "cold start" problem often encountered in collaborative RS by assigning each student to a cluster, allowing additional item recommendations. The overall results indicate that model-based CF outperforms other recommendation system techniques. This may be due to the technique's unique ability to consider users' or objects' perspectives.

D. Proposed Framework for e-Learning

Even if attempts are being made to create frameworks that may be used in various fields, there is still a long way to go before they are widely used. This study presents a framework for RS in the field of e-learning. However, there are still many problems, particularly with technology, customization, and user acceptability. Fig. 3 displays the prototype implementation flowchart.

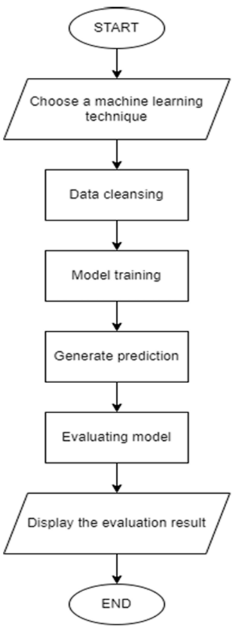

Fig. 3 Prototype Implementation Flowchart

During the implementation phase, the prototype is expected to perform a few background tasks before recommending items to users: data cleaning, model training, and model evaluation. Based on the evaluation results of the recommendation task, only the best ML will be chosen. The following stage will involve displaying and comparing the evaluation results obtained in the prototype using various ML algorithms. A graphical user interface will be constructed to provide a better visualization experience.

1) Choose ML techniques: This research project involved three ML approaches, which include the k-nearest Neighbor Baseline (KNNBaseline), Singular Value Decomposition (SVD), and CoClustering. The KNNBaseline algorithm preserves precision and accuracy by using the local minimum of the target function to identify an unknown function in conjunction with a baseline. The SVD algorithm can reduce dimensionality, which divides data into smaller portions for interpretation and analysis. The CoClustering algorithm offers perspectives on the relationship between rows and columns if the dataset is high dimensional. These approaches are widely used in the e-learning recommendation domain.

2) Dataset: The dataset chosen in this study is Course Reviews on Coursera. This is a public dataset that was also used in another study by Chan et al. [33]. This dataset consists of two tables: courses and reviews. The courses table has 4 attributes and 623 data records, while the reviews table has 5 attributes and 1.45m data records. Table II and Table III explain the details of the dataset.

TABLE II
ATTRIBUTE EXPLANATION IN COURSE TABLE

No	Attribute	Explanation
1	name	Name of the course
2	institution	Course institution
3	course_url	The url to reach the corresponding course
4	course_id	The identifier of each course

TABLE III
ATTRIBUTE EXPLANATION IN REVIEWS TABLE

No	Attribute	Explanation
1	reviews	The textual comments given by reviewers
2	reviewers	The identifier of each reviewer
3	date_reviews	The date that user gives a review
4	rating	Course rating is given by reviewers
5	course_id	The identifier of each course

3) Data Cleansing: Data cleansing plays an important role in ensuring promising result [34]. The process of cleaning the dataset begins with data sampling. Due to the large dataset size, a random sampling method is employed to reduce processing times and high memory demands. This step is done by selecting a representative subset of data from the original dataset. The optimal sample size is selected by conducting simulations to assess model performance using evaluation metrics such as MAE and RMSE with various sample sizes. Fig. 4 and Fig. 5 show the model performance using different sample sizes.

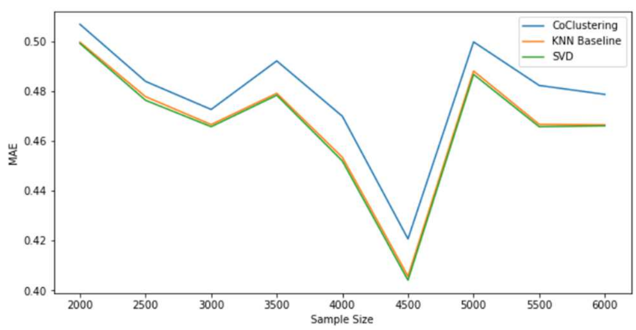

Fig. 4 Data Sampling MAE

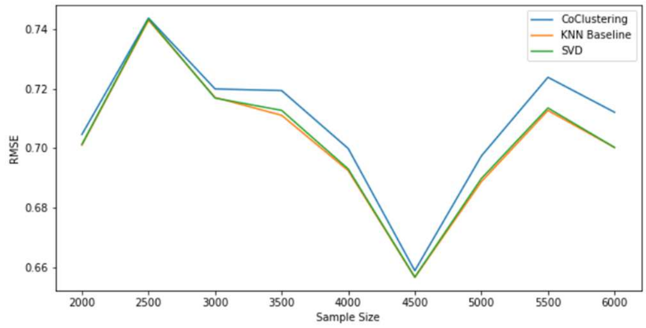

Fig. 5 Data Sampling RMSE

The datasets are sampled from 2000 to 6000. They are then used to train a model to analyze the trend of evaluation metrics using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The analysis reveals that the optimal dataset size is 4500. The next step involves cleaning up the 'reviewers' column in the dataset by removing unnecessary textual content not needed for the prototype. Fig. 6 illustrates the 'reviewers' column data after the cleaning process.

```
In [139]: df2['reviewers'] = df2['reviewers'].str[3:].copy()
          df2['reviewers']

Out[139]: 0                   Ewa M
          1                Dustin H
          2              Felipe M S
          3            Yelitzia R C
          4               Mohamed A
                             ...
          4495        Khaled A F M B
          4496            María X M A
          4497                Eric D S
          4498              Joakim A I
          4499                 Peter O
          Name: reviewers, Length: 4500, dtype: object
```

Fig. 6 Cleaning textual data

The missing values are then checked and removed from this dataset. Fig. 7 shows that each column (4500 rows of data) is complete and has no missing values.

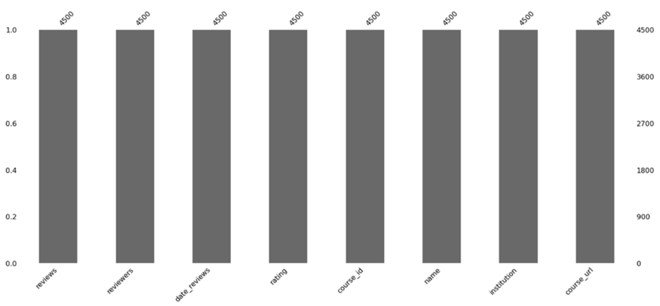

Fig. 7 Missing values in the dataset

4) *Recommender Engine:* The dataset preparation is finished at this point. The user-item matrix is now prepared for incorporation into the RS. By implementing the Surprise Python module, this research allows users to quickly and easily build a rating-based RS without starting from scratch. Additionally, the library contains some pre-built features explicitly designed for regression tasks, such as a train-test split function and numerous assessment criteria for accuracy.

III. Results and Discussion

The prototype shows the dashboard page as the default view. On this page, users will see a selection of 10 recommended courses. Each recommendation will include key details such as the course title, the institution offering the course, and the average rating given by reviewers. Fig. 8 shows the prototype's actual dashboard.

Fig. 8 Prototype Dashboard

Users can check for more detailed information by clicking the intended course. For example, the top-recommended course is chosen for demonstration. Fig. 9 shows more detailed information about the chosen course, including ID, name, institution, link in Coursera, and the average rating from reviewers.

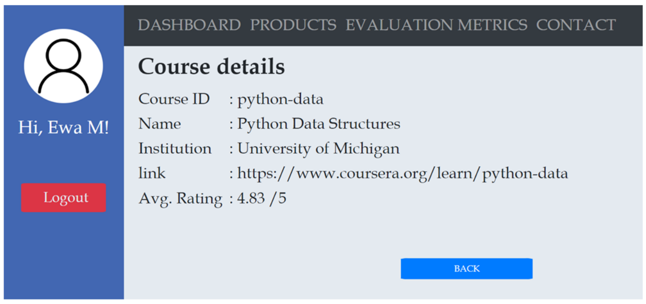

Fig. 9 Prototype Dashboard Course Details

The positive or negative emotions in user responses can be observed by using the course average rating provided on this course details page. The average rating ranges from 0 to 5, with 0 being the lowest possible rating and 5 being the highest. A higher average rating indicates that the course received more positive responses from users.

The evaluation metrics employed for comparing the model performance in this research project include MAE, RMSE, and Fraction of Concordant Pairs (FCP) metrics. Regarding MAE and RMSE, greater accuracy is indicated by smaller values. Conversely, for FCP, a higher value shows a better ranking for the model performance. According to the results, SVD exhibits the highest accuracy, with RMSE and MAE scores of 0.673 and 0.465, respectively. The second most accurate model is the KNN Baseline, with RMSE at 0.673 and MAE at 0.467. The CoClustering demonstrates the lowest accuracy, with RMSE at 0.686 and MAE at 0.48. Regarding FCP, CoClustering achieves the highest value of 0.916, followed by KNN Baseline at 0.485 and SVD at 0.426. As for FCP, CoClustering contributes to the highest value at 0.916, following by the KNN Baseline at 0.485. The model that achieves the lowest value is in terms of FCP is SVD at 0.426. Fig. 10 and Table IV demonstrate the results obtained from the prototype.

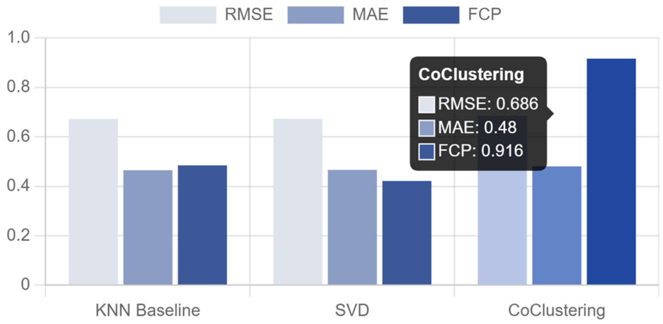

Fig. 10 Prototype Evaluation Results

TABLE IV
PROTOTYPE EVALUATION RESULTS

No	ML Technique	RMSE	MAE	FCP
1	KNN Baseline	0.673	0.467	0.485
2	SVD	0.673	0.465	0.426
3	CoClustering	0.686	0.480	0.916

The average time taken for model training is also recorded. The KNN Baseline model requires the shortest fitting time, just 0.018 minutes in precision, surpassing all other methods. Following closely behind is the SVD model, with a fitting time of 0.063 minutes. In contrast, the CoClustering model took the longest to fit, clocking in at 0.511 minutes. Fig. 11 and Table V show the average model fitting time recorded from the prototype.

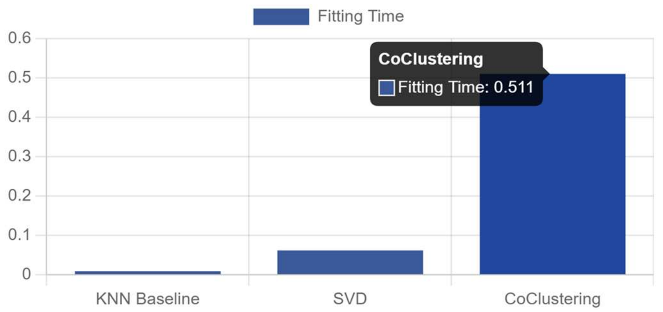

Fig. 11 Average Model Fitting Time

TABLE V
PROTOTYPE MODEL AVERAGE FITTING TIME

No	ML Technique	Average fitting time
1	KNN Baseline	0.018
2	SVD	0.063
3	CoClustering	0.511

The optimal model is the SVD model, followed by the KNN Baseline and CoClustering models. This model has the highest accuracy in predicting the rating regarding RMSE and MAE. As for the FCP, an RS focuses more on accuracy than item ranking. Hence, it is acceptable to choose a model that does not have the best item ranking. In addition, this model scores an average computation time in model training compared with other methods, with less memory required during model training.

The time required in the process of model training time is essential in deciding the best model. The KNN Baseline model has the shortest fitting time compared to the remaining methods. Following closely is the SVD model, with the CoClustering model taking the longest time to fit. The KNN Baseline model excelled due to its capacity to retain the entire dataset in memory, simplifying lookup operations during forecasting. As the dataset size increases, the time difference between the computational requirements of each method becomes longer.

Using a memory-based CF for real-world industrial applications is not advised since this approach necessitates precomputation before inserting data into the database, which will inevitably cause the application to load poorly. The long-term growth of the database would only increase the time needed for data processing. Therefore, a model-based CF, such as the SVD approach, would be the most advised because it produces data with high accuracy and a short processing time.

IV. CONCLUSION

This paper studies and discusses multiple recommendation approaches. A prototype is developed to explore the CF approach alongside various ML techniques such as KNN Baseline, SVD, and CoClustering. These methods are then evaluated using metrics including RMSE, MAE, and FCP, with the findings indicating that SVD performs the most effectively. Future work will involve improving the data visualization features within the prototype. Future enhancement should incorporate more exciting data visualization. In addition, subsequent work could explore integrating SVD methods with other approaches to boost the effectiveness of models.

ACKNOWLEDGMENT

This work is supported by Telekom Malaysia's funding of TM R and D (MMUE/240068).

REFERENCES

[1] Z. Wang, A. Maalla, and M. Liang, "Research on E-Commerce Personalized Recommendation System based on Big Data Technology," in *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, IEEE, Dec. 2021, pp. 909–913. doi:10.1109/ICIBA52610.2021.9687955.

[2] K. Sharma, D. Parashar, K. BasavaRaju, D. K. Singh, A. Kazi, and G. Kumar, "The Role of Big Data Analytics in Enhancing Customer Interactions and Loyalty," in *2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC)*, IEEE, Dec. 2023, pp. 308–311. doi: 10.1109/PEEIC59336.2023.10451643.

[3] M. M. Hussain, S. Akbar, S. A. Hassan, M. W. Aziz, and F. Urooj, "Prediction of Student's Academic Performance through Data Mining Approach," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 241–251, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.16.

[4] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Syst Appl*, vol. 166, p. 114060, Mar. 2021, doi:10.1016/j.eswa.2020.114060.

[5] U. Vellappan, L. Liyen, and S. Y. Lim, "Engaging Learning Experience: Enhancing Productivity Software Lessons with Screencast Videos," *Journal of Informatics and Web Engineering*, vol. 2, no. 2, pp. 189–200, Sep. 2023, doi: 10.33093/jiwe.2023.2.2.14.

[6] A. and S. A. and T. H. Aberbach Hicham and Jeghal, "E-learning Recommendation Systems: A Literature Review," *Lecture Notes in Networks and Systems*, vol 454. Springer, 2022. doi: 10.1007/978-3-031-01942-5_36

[7] C. Susaie, C.-K. Tan, and P.-Y. Goh, "Learning Experience with LearnwithEmma," *Journal of Informatics and Web Engineering*, vol. 1, no. 2, 2022, doi: 10.33093/jiwe.2022.1.2.3.

[8] Z. Yujiao, L. W. Ang, S. Shaomin, and S. Palaniappan, "Dropout Prediction Model for College Students in MOOCs Based on Weighted Multi-feature and SVM," *Journal of Informatics and Web Engineering*, vol. 2, no. 2, 2023, doi: 10.33093/jiwe.2023.2.2.3.

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

[10] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative Filtering for Implicit Feedback Datasets," in *2008 Eighth IEEE International Conference on Data Mining*, IEEE, Dec. 2008, pp. 263–272. doi:10.1109/ICDM.2008.22.

[11] E. Blasch *et al.*, "Machine Learning/Artificial Intelligence for Sensor Data Fusion–Opportunities and Challenges," *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 7, pp. 80–93, 2021, doi:10.1109/MAES.2020.3049030.

[12] N. Zamri, N. Palanichamy, and S.-C. Haw, "College Course Recommender System based on Sentiment Analysis," *Int J Adv Sci Eng Inf Technol*, vol. 13, no. 5, p. 1984, Oct. 2023, doi:10.18517/ijaseit.13.5.19032.

[13] I. Kumar, J. Rawat, N. Mohd, and S. Husain, "Opportunities of Artificial Intelligence and Machine Learning in the Food Industry," *J Food Qual*, vol. 2021, pp. 1–10, Jul. 2021, doi:10.1155/2021/4535567.

[14] S. Fuchs, C. Drieschner, and H. Wittges, "Improving Support Ticket Systems Using Machine Learning: A Literature Review," 2022. doi:10.24251/HICSS.2022.238.

[15] P. Rani, S. Kotwal, J. Manhas, V. Sharma, and S. Sharma, "Machine Learning and Deep Learning Based Computational Approaches in Automatic Microorganisms Image Recognition: Methodologies, Challenges, and Developments," *Archives of Computational Methods in Engineering*, vol. 29, no. 3, pp. 1801–1837, May 2022, doi:10.1007/s11831-021-09639-x.

[16] M. Lee and S. Oh, "An Information Recommendation Technique Based on Influence and Activeness of Users in Social Networks," *Applied Sciences*, vol. 11, no. 6, p. 2530, Mar. 2021, doi:10.3390/app11062530.

[17] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl Based Syst*, vol. 226, p. 107134, 2021, doi:10.1016/j.knosys.2021.107134.

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

[19] Ms. T. Sharad Phalle and Prof. S. Bhushan, "Content Based Filtering And Collaborative Filtering: A Comparative Study," *Journal of Advanced Zoology*, pp. 96–100, Mar. 2024, doi:10.53555/jaz.v45iS4.4158.

[20] 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, 2024, doi: 10.1007/s11042-023-15258-4.

[21] 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, Jan. 2024, doi: 10.1007/s11227-023-05519-2.

[22] G. Behera and N. Nain, "Handling data sparsity via item metadata embedding into deep collaborative recommender system," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 9953–9963, Nov. 2022, doi: 10.1016/j.jksuci.2021.12.021.

[23] J. Wang, A. Kamran, F. Shahzad, and N. A. Syed, "Enhancing group recommender systems: A fusion of social tagging and collaborative filtering for cohesive recommendations," *Syst Res Behav Sci*, Feb. 2024, doi: 10.1002/sres.3000.

[24] 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*, pp. 1–9, Feb. 2024, doi:10.1080/1206212X.2024.2309809.

[25] E. Adeoye, H. K, R. E, and K. P, "Hybrid Recommendation Systems," *SSRN Electronic Journal*, 2024, doi: 10.2139/ssrn.4712941.

[26] S. Suriati, M. Dwiastuti and T. Tulus, "Weighted hybrid technique for recommender system", *Journal of Physics: Conference Series*, vol. 930, 2017, doi: 10.1088/1742-6596/930/1/012050

[27] J. K. Tarus, Z. Niu, and D. Kalui, "A hybrid recommender system for e-learning based on context awareness and sequential pattern mining," *Soft comput*, vol. 22, no. 8, pp. 2449–2461, Apr. 2018, doi:10.1007/s00500-017-2720-6.

[28] S. B. Aher and L. M. R. J. Lobo, "Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data," *Knowl Based Syst*, vol. 51, pp. 1–14, Oct. 2013, doi: 10.1016/j.knosys.2013.04.015.

[29] Huynh-Ly Thanh-Nhan, Huu-Hoa Nguyen, and N. Thai-Nghe, "Methods for building course recommendation systems," in *2016 Eighth International Conference on Knowledge and Systems Engineering (KSE)*, IEEE, Oct. 2016, pp. 163–168. doi:10.1109/KSE.2016.7758047.

[30] R. Obeidat, R. Duwairi, and A. Al-Aiad, "A Collaborative Recommendation System for Online Courses Recommendations," in *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, IEEE, Aug. 2019, pp. 49–54. doi: 10.1109/Deep-ML.2019.00018.

[31] L. Zhao and Z. Pan, "Research on Online Course Recommendation Model Based on Improved Collaborative Filtering Algorithm," in *2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, IEEE, Apr. 2021, pp. 437–440. doi:10.1109/ICCCBDA51879.2021.9442575.

[32] H. Aoulad Ali, C. Mohamed, B. Abdelhamid, and T. El Alami, "A Course Recommendation System for Moocs Based On Online Learning," in *2021 XI International Conference on Virtual Campus (JICV)*, 2021, pp. 1–3. doi: 10.1109/JICV53222.2021.9600379.

[33] H. Y. Chan, R. Rajamohan, K. H. Gan, and N.-H. Samsudin, "Text Analytics on Course Reviews from Coursera Platform," in *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, IEEE, 2021, pp. 1–6. doi:10.1109/IICAIET51634.2021.9573868.

[34] Z. Y. Poo, C. Y. Ting, Y. P. Loh, and K. I. Ghauth, "Multi-Label Classification with Deep Learning for Retail Recommendation," *Journal of Informatics and Web Engineering*, vol. 2, no. 2, pp. 218–232, 2023, doi: 10.33093/jiwe.2023.2.2.16.