Conference Paper

Systematic Review of Machine Learning in Recommendation Systems Over the Last Decade

Weiner, F., Teh, P.L., Cheng, CB.

This is a paper presented at the Intelligent Computing, SAI 2024.

Recommended citation:

Weiner, F., Teh, P.L., Cheng, CB. (2024), 'Systematic Review of Machine Learning in Recommendation Systems Over the Last Decade', In Proc: Arai, K. (eds) Intelligent Computing. SAI 2024. Lecture Notes in Networks and Systems, vol 1016. Springer, Cham. doi: 10.1007/978-3-031-62281-6_5.

The final version of this paper is available here: **https://link.springer.com/chapter/10.1007/978-3-031-62281-6_5**

To cite this paper:

Weiner, F., Teh, P.L., Cheng, CB. (2024). Systematic Review of Machine Learning in Recommendation Systems Over the Last Decade. In: Arai, K. (eds) Intelligent Computing. SAI 2024. Lecture Notes in Networks and Systems, vol 1016. Springer, Cham. https://doi.org/10.1007/978-3-031-62281-6_5

Systematic Review of Machine Learning in Recommendation Systems over the Last Decade

Felix Weiner[1]; Phoey Lee Teh[2] and Chi-Bin Cheng[3]

[1,2] Faculty of Art, Computing and Engineering, Wrexham University, UK.
[2] Department of Information Management, Tamkang University, New Taipei City, 251301, Taiwan.
[1,2]phoey.teh@glyndwr.ac.uk; [3]cbcheng@mail.tku.edu.tw

Abstract. This study presents a comprehensive overview of the approaches employed in recommendation systems over the last decade. The review primarily draws from two categories of filtering techniques: content-based filtering and collaborative filtering methods. We have reviewed and tabulated approximately forty articles that have been published. Major findings include: 1) collaborative filtering is more often used than content-based filtering, 70% to 23%, the rest is hybrid methods of these two; 2) more than half of the machine learning approaches adopted are supervised learning; however, 3) algorithm-wise, K-means the unsupervised learning algorithm emerged as the most frequently adopted approach in recommendation systems. Also notably, cosine similarity stands out as the prevalent measurement technique.

Keywords: Content-based filtering, Collaborative filtering, Recommendation Systems, K-Mean Clustering techniques.

1 Introduction

Recommendation systems aim to suggest or recommend additional products or content to consumers based on their past purchasing behaviors and preferences. Machine learning is the frequently used approach for constructing recommendation systems, as their capability of analyzing large sets of data. With the increasing popularities of online shopping and online streaming contents, there has been growing interest among computer scientists and data scientists to develop the best approaches and processes for recommendation systems.

To achieve optimal performance, recommendation systems must be able to effectively handle large and complex datasets, as well as incorporate a range of different data sources and algorithms. In recent years, machine learning has played an increasingly significant role in developing more sophisticated and accurate recommendation systems, including approaches like collaborative filtering, content-based filtering, and hybrid systems that combine multiple approaches.

2 Literature review

Table 1 summarizes the approaches used in the research and project works related to recommendation systems between the years 2012 to 2023.

Table 1. Summary of approaches used in the past decade

Year	Categories	Approach & Evaluation Method	Type
2013 [1]	Item Categories	Random-Walk Based Algorithm	Unsupervised
2013 [2]	TV Programs	K-means and K Nearest Neighbor	Mixed
2013 [3]	Product reviews	Latent Class Regression model, (LCRM), K- Means and K - Nearest Neighbor	Mixed
2013 [4]	Item Rating	Pearson Correlation Similarity and Mean Absolute Error	Mixed
2014 [5]	User Preferences	Cosine Similarity and Mean Absolute Error	Unsupervised
2014 [6]	Movie Categories	Matrix Factorization and Mean Absolute Error	Supervised
2014 [7]	Product and User	Aprioria Algorithm, Jaccard Similarity	Unsupervised
2014 [8]	Book Reviews	Opinion Mining and Naive Bayes	Unsupervised
2015 [9]	User Rating	Fuzzy Multi Criteria Decision Making Approach	Supervised
2015 [10]	Product and User	Modified Cosine Similarity	Unsupervised

2015 [11]	Book features	Opinion Mining	Unsupervised
2015 [12]	Item Rating	Cosine Similarity, Jaccard Similarity and Mean Absolute Error	Unsupervised
2016 [13]	Item Rating	Mean Absolute Error	Semi-Supervised
2016 [14]	User Rating	K-mean Algorithm and Gradient Descent Approach	Unsupervised
2016 [15]	Suggest Haircut with Face contour	Support Vector	Semi-Supervised
2016 [16]	Website	K-mean Algorithm	Unsupervised
2017 [17]	Movie list	Support Vector Machine and Generic Algorithm	Unsupervised
2017 [18]	TV Stream	Logic Regression and Stochastic Gradient Descent	Reinforcement
2017 [19]	IoT infrastructure	Bayesian Network, Naive Bayes, Decision Tree, Random Forest	Reinforcement
2017 [20]	Item Categories	Pearson Correlation Similarity, Mean Absolut Error and Root Mean Square Error	Semi-Supervised
2018 [21]	Audio sample through log	Convolutional Neural Network (CNN) Approach	Unsupervised
2018 [22]	User Preferences	Jaccard Similarity and K-means	Unsupervised
2018 [23]	Music & Audio	K Nearest Neighbor, Naïve Bayas, Decision Tree, Support Vector Machine and Random Forest	Reinforcement
2018 [24]	Product reviews	Cosine Similarity, Modified Cosine Similarity and Pearson Correlation Similarity	Unsupervised
2019 [25]	User Rating	Modified Cosine Similarity	Supervised
2019 [26]	Query related code segments	Hamming Distance, Cosine Similarity and Euclidean Distance	Supervised
2019 [27]	Online Course	K-means Algorithm, Apriori and SPADE Algorithm were compared	Unsupervised
2019 [28]	User Personalization	Matrix Factorization, Pearson Correlation Similarity and Mean Absolute Error	Unsupervised
2020 [29]	Products	Modified Cosine Similarity and Clustering	Unsupervised
2020 [30]	Item Categories	Singular Value Decomposition (SVD), Matrix, Factorization and Cosine Similarity	Supervised
2020 [31]	Loan Prediction	Logistic Regression and Random Forest	Supervised
2020 [32]	User Behavior	K-means, FunkSVD Algorithm Cosine Similarity, Pearson Correlation Coefficient and Euclidean Distance	Unsupervised
2021 [33]	Topic Modelling	Latent Dirichlet Allocation (LDA) and Jensen-Shannon, Distance	Semi-Supervised

2021 [34]	Product reviews	Singular Value Decomposition and K-means Algorithm	Unsupervised
2021 [35]	Restaurant	Weight Score and Cosine Similarity	Supervised
2021 [36]	Product reviews	K-Means	Unsupervised
2021 [37]	Real time item	Contextual Cluster, Euclidean Distance	Reinforcement
2022 [38]	Tweet Texts	Latent Dirichlet Allocation (LDA), Hierarchical Dirichlet Process (HDP), Laten Semantic Indexing (LSI), Non-Negative Matrix Factorization (NMF) and Structural Topic Model (STM)	Unsupervised
2022 [39]	Coffee Shop	Modified Cosine Similarity and Mean Absolute Error	Supervised
2022 [40]	Leather Product	Modified Cosine Similarity and Mean Absolute Error	Unsupervised

2.1 Approach overview

The field of recommendation systems has seen the development of two major approaches: collaborative filtering and content-based filtering. Collaborative filtering is a technique that recommends products or content based on the preferences and behaviors of other users with similar tastes. This approach can identify similarities in user behavior and make recommendations based on those similarities [5,22,27,33]. Content-based filtering, on the other hand, uses the characteristics and attributes of products to make recommendations. This approach can suggest equivalent products based on shared features or themes.

To achieve even more effective recommendation systems, a hybrid approach can be utilized. By combining both techniques, a system can take advantage of the strengths of each to provide even more personalized recommendations. For example, a hybrid system could use collaborative filtering to identify groups of users with similar interests, and then use content-based filtering to recommend specific products to those groups. Research has demonstrated the efficacy of collaborative filtering in recommendation systems, as evidenced by its frequent usage in commercial and research applications. However, both approaches have their own strengths and limitations, and the best approach will depend on the specific context and goals of the system. As illustrated in fig. 1, Collaborative Filtering approach was the most widely used.

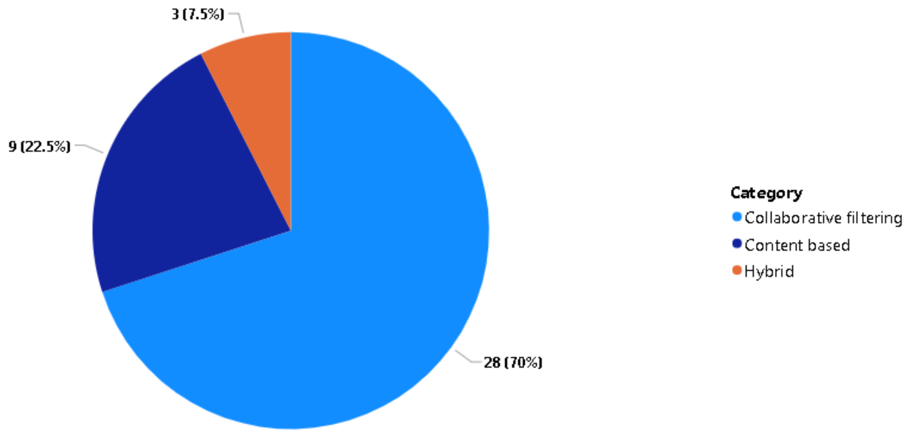

Fig. 1. Approach Overview

2.2 Type of Machine Learning

Recommendation systems can be classified into several types based on the way data is interpreted and learned. These categories include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning utilizes a prepared dataset that has knowledge of several areas to make recommendations. Conversely, unsupervised learning does not involve preset datasets, and the computer must use several methods to discover underlying patterns [20,31]. Semi-supervised learning is a hybrid approach that includes some given labelling, similar to supervised learning, but significantly less. The focus is on analyzing datasets and discovering connections [31]. Reinforcement learning is another type of machine learning, where the goal is to accomplish a specific objective by carrying out a sequence of events that take the previous one into consideration. Multiple criteria are considered throughout the process.

The distribution of these four machine learning types across reviews is illustrated in Fig. 2.

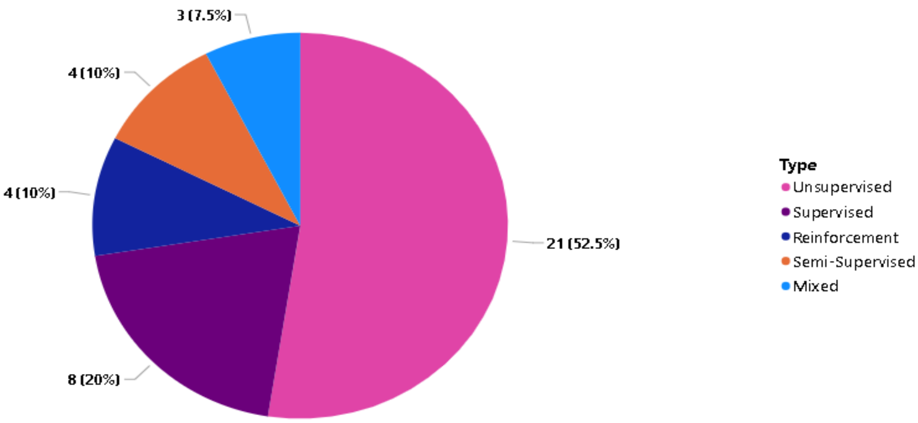

Fig. 2. Type of Machine Learning used

3 Results

K-means is the most often used algorithm, accounting for 30% of all articles. The distribution of algorithms among the papers examined (see Fig. 3.).

A variety of techniques are used to achieve different objectives in data analysis. Among them, the K-means algorithm is commonly used to categorize individuals into groups. The ability to divide users into smaller, more manageable groups can facilitate comparisons between them and speed up algorithms by reducing computation requirements.

To accomplish this, the K-means algorithm selects the same number of vector points as the desired number of clusters. These points can be either data vectors or entirely random points. The algorithm then determines the distance between each data vector and the cluster points, assigning each vector to the closest cluster. Next, each cluster's center is calculated by averaging all vectors within that cluster. The distances between each vector and the cluster center are reevaluated, potentially leading to some vectors shifting to a different cluster based on proximity. If any vectors change their cluster assignment, the center of the affected cluster is recalculated by averaging all of the vectors now included in that cluster. This process of measuring distances and reevaluating centers continues until no further vector changes its cluster assignment, at which point the clusters are established [1,2.27,32].

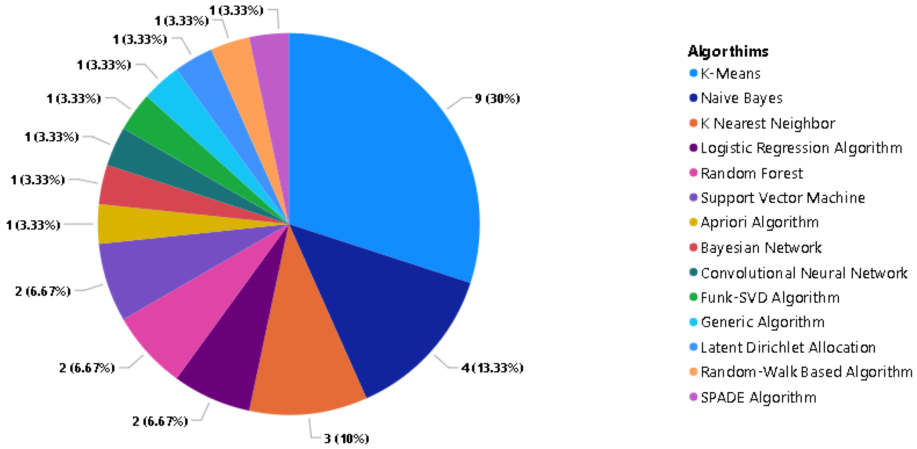

Fig. 3. Algorithm overview

There are many methods available to measure the distances and similarities between vectors. By utilizing techniques such as K-means clustering and cosine similarity, researchers can more effectively analyze large and complex datasets, facilitating a wide range of applications across multiple fields. Cosine similarity is the approach that is most frequently employed across all articles.

In order to assess the degree of similarity between multiple users, a variety of algorithms have been developed. The two most common algorithms are cosine and Pearson correlation similarities. To facilitate this process, the data must first be transformed into a numerical format, which results in a series of vectors. By comparing these vectors and determining their similarity, it becomes possible to compare the different users and assess their level of similarity [20,25,32,40].

Algorithms can be customized to provide benefits or overcome drawbacks, depending on the situation. For instance, you can calculate the distance between two vectors using cosine similarity, but you can also refine the calculation by including the vector's average in all the numbers. This improved cosine similarity can enhance the accuracy of the algorithm and prevent rating issues if ratings are used in an approach [10,25].

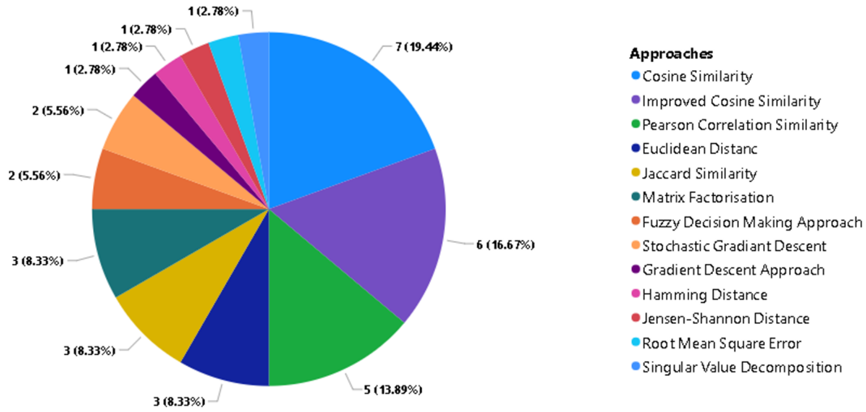

Fig. 4. Measurements approaches

The addition of a new item or user to the system poses a challenge known as the "cold start" problem. This issue arises from the system's attempt to compare objects or user profiles that lack adequate data. For instance, a new item may belong to a novel category that has not been established, or it may lack any ratings. Similarly, the profile information available for new users may be limited, thereby constraining the data available for the algorithms to compare user profiles or objects. A solution to mitigate this issue is to combine collaborative filtering with content-based filtering. One approach is to suggest new goods, while another involves creating recommendations for users based on the profiles of trustworthy individuals [4,30,37].

4 Conclusion

In summary, the K-means algorithm emerges as the most frequently employed technique, with mean absolute error being the predominant evaluation metric across all approaches. Originally developed for signal processing as a clustering method of vector quantization, K-means has transitioned into various applications within recommendation systems across diverse industries, even encompassing the analysis of TV programs. Further research should expand upon this review, delving into a comprehensive exploration of the implementation of each algorithm and approach.

5 References

1. Zhang, L., Xu, J., Li, C.: A random-walk based recommendation algorithm considering item categories, Neurocomputing, vol. 120, pp. 391–396 (2013).
2. Chang, J. H., Lai, C. F., Wang, M. S., Wu, T. Y..: A cloud-based intelligent TV program recommendation system, Computers & Electrical Engineering, vol. 39, pp. 2379–2399 (2013).

3. Chen, L., Wang, F.: Preference-based clustering reviews for augmenting e-commerce recommendation, Knowledge-Based Systems, vol. 50, pp. 44–59 (2013).
4. Chen, C. C., Wan, Y. H., Chung, M. C., Sun, Y. C.: An effective recommendation method for cold start new users using trust and distrust networks, Information Sciences, vol. 224, pp. 19–36 (2013).
5. Zhang, J., Peng, Q., Sun, S., Liu, S.: Collaborative filtering recommendation algorithm based on user preference derived from item domain features," Physica A: Statistical Mechanics and its Applications, vol. 396, pp. 66–76 (2014).
6. Krasnoshchok, O. and Lamo,Y.: Extended Content-boosted Matrix Factorization Algorithm for Recommender Systems, Procedia Computer Science, vol. 35, pp. 417–426 (2014).
7. Choi, Y.K. and Kim, S.K.: An auxiliary recommendation system for repetitively purchasing items in E-commerce, in 2014 International Conference on Big Data and Smart Computing (BIGCOMP), pp. 96–98, IEEE (2014).
8. Tewari, S., Ansari, T. S., and Barman, A. G.: Opinion based book recommendation using Naive Bayes classifier, in 2014 International Conference on Contemporary Computing and Informatics (IC3I), pp. 139–144, IEEE (2014).
9. Parveen, H., Ashraf, M., and Parveen, R.: Improving the performance of multi-criteria recommendation system using fuzzy integrated meta heuristic, in International Conference on Computing, Communication & Automation, pp. 304–308, IEEE (2015).
10. Pujahari and Padmanabhan, V.: Group Recommender Systems: Combining User-User and Item-Item Collaborative Filtering Techniques, in 2015 International Conference on Information Technology (ICIT), pp. 148–152, IEEE (2015).
11. Priyanka, K., Tewari, A. S., and Barman, A. G.: Personalised book recommendation system based on opinion mining technique, in 2015 Global Conference on Communication Technologies (GCCT), pp. 285–289, IEEE (2015).
12. Zhong, Z., Sun, Y., Wang, Y., Zhu, P., Gao, Y., Lv, H., and Zhu, X.: An improved collaborative filtering recommendation algorithm not based on item rating, in 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), pp. 230–233, IEEE (2015).
13. Savadekar, B.B. and Patil, M.E.: Improved recommendation system with review analysis, in 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), pp. 79–82, IEEE (2016).
14. Garg, V. and Tiwari R.: Hybrid massive open online course (MOOC) recommendation system using machine learning, in International Conference on Recent Trends in Engineering, Science & Technology - (ICRTEST 2016), pp. 11 (5.) –11 (5.), Institution of Engineering and Technology (2016).
15. Sunhem, W., Pasupa, K. and P. Jansiripitikul,: Hairstyle recommendation system for women, in 2016 Fifth ICT International Student Project Conference (ICT-ISPC), pp. 166–169, IEEE (2016).
16. Guo, H., Chen, L., Chen, S. and Mi, J.: A personalized recommendation technology for Ecommerce website, in 2016 International Conference on Machine Learning and Cybernetics (ICMLC), pp. 761–765, IEEE (2016).
17. Agrawal, S. and Jain, P.: An improved approach for movie recommendation system, in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 336–342, IEEE (2017).
18. Sunny, B.K., Janardhanan P, A. B. Francis, and R. Murali: Implementation of a self-adaptive real time recommendation system using spark machine learning libraries, in 2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), pp. 1–7, IEEE (2017).

19. Siryani, B.J., Tanju, B., and Eveleigh, T. J.,:A Machine Learning Decision-Support System Improves the Internet of Things' Smart Meter Operations, IEEE Internet of Things Journal, vol. 4, pp. 1056–1066 (2017).
20. Jiang, S.Q., Y.-F. Liu, Zhou, Y.X. and Zhi, H. H.,: Collaborative Recommendation Algorithm Based on Semi-Supervised Learning, in 2017 International Conference on Network and Information Systems for Computers (ICNISC), pp. 189–194, IEEE (2017).
21. Chang, S.H., Abdul, J. A. Chen, and Liao, H.Y.,: A personalized music recommendation system using convolutional neural networks approach, in 2018 IEEE International Conference on Applied System Invention (ICASI), pp. 47–49, IEEE (2018).
22. Li, P., G. Zhang, L. Chao, and Z. Xie, Personalized Recommendation System for Offline Shopping, in 2018 International Conference on Audio, Language and Image Processing (ICALIP), pp. 445–449, IEEE (2018).
23. Elbir, A., Bilal Cam H., Emre Iyican, M., Ozturk, B. and Aydin, N.: Music Genre Classification and Recommendation by Using Machine Learning Techniques, in 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1–5, IEEE (2018).
24. Surekha, S.: A Personalized Recommendation System using Memory-Based Collaborative Filtering Algorithm, in 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), pp. 1–6, IEEE (2018).
25. Liu, F. and Guo, W.W.: Research on House Recommendation Model Based on Cosine Similarity in Deep Learning Mode in Grid Environment, in 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), pp. 121–124, IEEE (2019).
26. Zhou, Y., Cui S. and Wang, Y.: Development of Machine Learning Based Recommendation Engine for Embedded Programmer, in 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), pp. 634–637, IEEE (2019).
27. Obeidat, R., Duwairi, R. and Al-Aiad, A.: A Collaborative Recommendation System for Online Courses Recommendations, in 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), pp. 49–54, IEEE (2019).
28. Li, W., Zhou, X., Shimizu, S., Xin, M., Jiang, J., Gao, H. and Jin, Q.: Personalization Recommendation Algorithm Based on Trust Correlation Degree and Matrix Factorization, IEEE Access, vol. 7, pp. 45451–45459 (2019).
29. Zhou, J., Wan, F. and Jing, R.: Model and Implementation of E-commerce Recommendation System Based on User Clustering, in 2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI), pp. 197–200, IEEE (2020).
30. Wairegi, S., Mwangi, W. and Rimiru, R.: A Framework for Items Recommendation System Using Hybrid Approach, (Kampala), IEEE (2020).
31. Gupta, A., Pant, V., Kumar, S., and Bansal, P.K.: Bank Loan Prediction System using Machine Learning, in 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), pp. 423–426, IEEE (2020).
32. Guo, S. and Li, C.: Hybrid Recommendation Algorithm based on User Behavior, in 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), pp. 2242–2246, IEEE (2020).
33. Bagul, D.V. and Barve, S.: A novel content-based recommendation approach based on LDA topic modeling for literature recommendation, in 2021 6th International Conference on Inventive Computation Technologies (ICICT), pp. 954–961, IEEE (2021).
34. Khatter, H., Arif, S., Singh, U., Mathur, S. and Jain, S.: Product Recommendation System for E-Commerce using Collaborative Filtering and Textual Clustering, in 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 612–618, IEEE (2021).

35. Ahmed, T., Akhter, L., Talukder, F.R., Hasan-Al-Monsur, Rahman, H and Sattar, A.: Restaurant Recommendation System in Dhaka City using Machine Learning Approach, in 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), pp. 59–63, IEEE (2021).
36. Santhosh N.M., Cheriyan, J., and Sindhu, M.: An Intelligent Exploratory Approach for Product Recommendation Using Collaborative Filtering, in 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), pp. 232–237, IEEE (2021).
37. Kabra, A., Agarwal, A., and Parihar, A.S: Potent Real-Time Recommendations Using Multimodel Contextual Reinforcement Learning, IEEE Transactions on Computational Social Systems, vol. 9, pp. 581–593 (2022).
38. Teh, P.L., Piao, S., Almansour, M., Ong, H.F., Ahad, A.: Analysis of Popular Social Media Topics Regarding Plastic Pollution. Sustainability.14(3):1709.https://doi.org/10.3390/su14031709. (2022).
39. Astri,R., Kamal, A., and Sura, S. B.: “Coffee Shop Recommendation System Using an Item Based Collaborative Filtering Approach, in 2022 International Symposium on Information Technology and Digital Innovation (ISITDI), pp. 65–67, IEEE (2022).
40. Ridwan, A. M., Ramdania, D. R, Maylawati, D. S., Wahana, A., Mulyana, E. and Al-Amin, M. I.: Leather Product Recommendation System using Collaborative Filtering Algorithm, in 2022 8th International Conference on Wireless and Telematics (ICWT), pp. 1–5, IEEE (2022).