

Article

Effects of Personalized Recommendation System Based on User Learning Behavior in English-Chinese Translation

Xiaofang Yan^{1,*}

¹ Foreign Language Department, Lyu Liang University, Lyuliang 033000, Shanxi, China

* **Correspondence:** yanxiaofang2023@163.com

Abstract: Users may receive personalised information services and decision support from personalised recommendations. In this paper, a hybrid algorithm-based personalised recommendation approach for learning English is proposed. The user model is created by merging user interest tags, and the Person Rank algorithm is then recommended based on user information. Second, the question-and-answer model is created once the question-and-answer data has been labelled, and the Problem Rank algorithm is suggested in accordance with the question-and-answer data. Then, the approach of tag-based recommendation, comparable user recommendation, and multi-dimensional sliding window are used to construct the recommendation algorithm model. The experimental findings demonstrate that, following the model's training with the gradient descent technique, the recommendation accuracy is steady at around 0.78, the suggested information can accommodate users who are learning English, and the personalised recommendation effect is enhanced.

Keywords: Learning personalization, Hybrid algorithm, Recommendation, Bayesian network

1. Introduction

There are currently numerous studies undertaken both domestically and internationally that characterise participants' learning behaviours, and there are certain restrictions on the choice of study subjects and samples. However, numerous studies on online learners' behaviours have been carried out with the assistance of researchers [1]. In on-line English-Chinese translation courses, the amount of user learning behaviour and reflective learning was investigated [2]. The relationship between various levels of prior knowledge and students' online learning behaviour has been investigated in this research. It has been investigated how learners' online learning behaviours relate to various levels of prior knowledge [3]. Some academics presented their own theories regarding analysis methodologies and the creation of analysis models [4].

The majority of the research on learning analysis technology is based on literature research questionnaires and is still in the early stages of investigation. The domestic online learning analysis technology primarily researches theories, concepts, and influencing elements. There aren't many empirical research on data analysis using on-line English-Chinese translation courses platforms, the data mining and data gathering are shallow, and there isn't any good visual analysis of the data, which is bad for finding hidden information [5]. Therefore, the relationship between online learning behavior and students' offline performance, as well as effective learning guidance methods, deserve special attention. This paper focuses on the following three research issues.

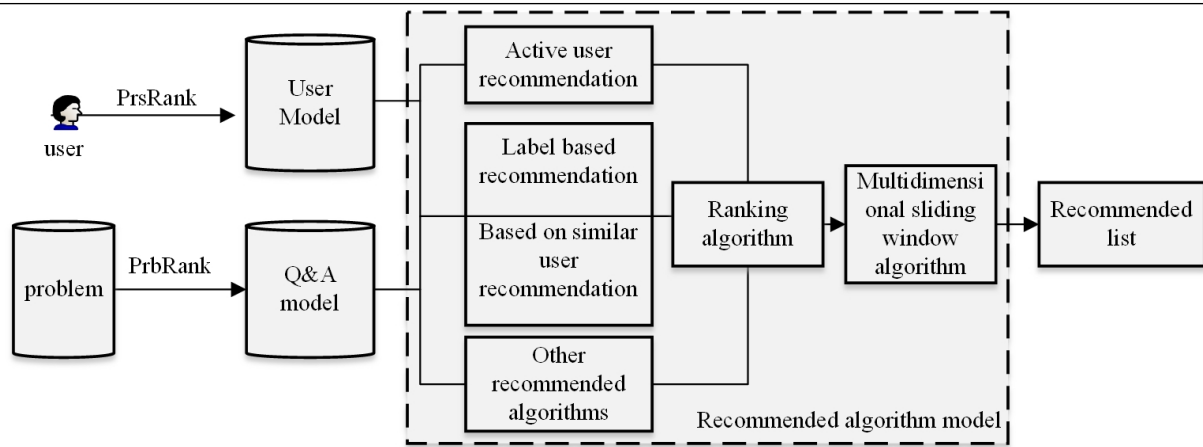


Figure 1. Personalized Recommendation Model

If we utilise online learning behaviour to forecast how well a person will perform offline, we can objectively assess a person's learning effect. Analyse the regularity of your online learning behaviour and whether it is a factor in your performance [6, 7]. Determine how different learner personalities affect their online learning behaviours, then suggest suitable online learning strategies for each personality type.

2. Personalized Recommendation Model Based on Hybrid Algorithm

2.1. Recommended Model Architecture

Figure 1 illustrates the three main components of the personalised on-line English-Chinese translation courses recommendation model for learning English: the user model, the question-answer model, and the recommendation algorithm model. Person Rank (PrsRank), an iterative method, is suggested to develop a user model with its interest tag in accordance with user input for English language learners [8, 9]. A Problem Rank (PrbRank) algorithm is suggested based on the question-and-answer data, and the question-and-answer model is built after labelling the question-and-answer data. Then, the ranking algorithm is constructed by using tag-based recommendation, similar users and other recommendation methods, and the recommendation algorithm model is constructed by combining multi-dimensional sliding windows in stages to finally obtain the recommendation information list.

2.2. Data Acquisition and Analysis

Build a user model by analysing the on-line English-Chinese translation courses learning webpage and user characteristics, collecting user IDs, follower and followed-by counts, questions answered, consent counts, thanks received, attention tags, and other data. The question-answering model is created by gathering data on the question's time, the number of liked questions, the number of answers, the question's title, its tag, etc. We eventually collected 26000 English learning users, 4.61 million following links, and 720000 question data after processing noise data [10].

As an online question and answer platform, English learning can reflect user activity in question and answer. The analysis of the experimental data set shows that most users do not participate in question-and-answer interaction, that is, the overall user activity of English learning is low and shows a long tail effect, as shown in Table 1.

The attention relationships between users on the English learning platform, which has a social component, make up the points and edges of the graph's structure. According to Table 1, 16.24% of users are in charge of resolving issues related to English language acquisition, and only a small portion of answers are approved. Up to 83.77% of users have not responded to queries. As a result, the study map's active user group can increase overall English learning user activity [11, 12].

Didn't answer any questions	21913	83.77%
Agreed and concerned	2373	7.12%
Agreed, no concern	432	1.66%
No approval and concern	864	3.65%
No approval or concern	1863	3.83%

Table 1. User Q&A Report of Zhihu

Data	Net10k	Net50k
Number of points	1859	379
Average path length	2.12	1.86
radius	3	3
diameter	6	5

Table 2. Index Data of Strong Connected Graph

For active users, take users with more than 10000 (Net10k, 1885 people) and more than 50000 (Net50k, 395 people) to quantify the characteristics of active groups in English learning. The graph density indicates the tightness of the graph structure, which can reflect the degree of association between users. The calculated density of Net10k and Net50k graphs are 0.064 and 0.195, respectively, indicating that users with more approval are more closely concerned.

A strongly linked graph can be used to represent the relationship between English learning users, and the strongly connected components can be used to effectively represent the interaction between groups. The findings demonstrate how a close-knit social circle is formed when active users from many fields congregate in one place. Therefore, increasing the influence of active users and widening their social circle will help to increase the overall activity of English learning users. The overall user activity can be significantly increased by promoting active users to inactive users by determining the average path length, radius, and diameter of the strongly connected graph, as shown in Table 2.

2.3. Construction of User Model

2.3.1. Construction of User Model

To calculate the English learning user ranking (PrsRank), After the principal component analysis, the number of users following m_1 , the number of people following m_2 , the number of responses m_3 , the number of approval received m_4 , and the number of thanks received m_5 are introduced as the influencing factors of the PrsRank function, where w_i represents the weight of each influencing factor. As shown in Eq. (1).

$$\text{PrsRank}(m_i) = \sum_{i=1}^5 w_i \times m_i. \quad (1)$$

2.3.2. Iterative PrsRank Function

The PageRank ranking algorithm is introduced in combination with the strong connectivity among English learning users, that is, the user's own ranking is determined by the follower's ranking. English learning users have relatively complex concerns [13]. The strong connected graph contains some isolated points (points with an outgoing chain of 0), that is, user u_0 does not pay attention to other users. Therefore, the damping coefficient q is added, which is generally 0.85.

$$\text{PageRank}(u_i) = \frac{1-q}{N} + q \times \sum_{u_j} \frac{\text{PageRank}(u_j)}{L(u_j)}. \quad (2)$$

User behavior	field	Behavior weight	Behavior description
Points exchange picture book	user exchange	1	The user exchanged the picture book
Downloading picture books	user downloa	0.8	The user downloaded the picture book
Collection	user collect	0.7	Users collected picture books
share	user share	0.6	The user shared the picture book
Phonetic Follow up	user voice read	0.5	The user followed the picture book
Read carefully	user read c	0.9	Users are interested in picture books
Normal reading	user read n	0.6	Medium user interest in picture books
Fast reading	user read f	0.3	Low user interest in picture books

Table 3. Weighting Table of User Behavior

The initial user ranking is obtained from Eq. (1), and the final user ranking is obtained after the iteration of Eq. (2). PageRank (u_j) is the updated ranking value of user u_j , and $L(u_j)$ is the number of people followed by user u_j .

2.4. Final Recommended Treatment

The ranking of users and questions is dynamic due to variables like the quantity of liked messages, the quantity of forwarded messages, and the passage of time. Every piece of recommendation information, both label-based and comparable user-based recommendations, is fundamentally unique. In order to preserve the ideal value within the interval, multi-dimensional sliding windows are employed for suggestion information. Determine users' satisfaction with different recommendation methods sat_i to dynamically adjust different categories of recommendation information rec_i , to further modify the recommendation list. As shown in Eq. (3)

$$\text{recommendation} = \sum sat_i \times rec_i. \quad (3)$$

3. English Text Preprocessing Process

Although there are minor variations, the preprocessing approach for English text is largely the same as that for Chinese text. First, although Chinese text mining preprocessing involves word segmentation, English text mining preprocessing typically does not (unless there are unique requirements). Second, since the majority of English texts are encoded using UTF-8, most processing does not need to take the issue of encoding conversion into account. The issue of Unicode encoding for Chinese texts needs to be solved. In actuality, English text preparation has its own unique characteristics, and then there is the spelling issue. Spelling checks, such as "Hello Word" errors, should typically be part of the preprocessing of English text because they cannot be left for repair after analysis. In order to prepare for pretreatment, this must be fixed. Lemmatization and stem extraction are the more crucial concepts. This step is necessary since English has multiple tenses, as well as the singular, plural, and singular and plural forms of the same word. As an illustration, the words "countries" and "country" and "wolf" and "wolves" are the same [14].

The user's behaviour with picture books in the picture book reading recommendation system shows the user's varied tastes for various kinds of picture books. In the form of logs, the data corresponding to user behaviour will be kept in the database or system file in the background of the system. It is required to retrieve pertinent logs, identify the user's interest in various picture books in light of the user's varied behaviour, and then provide appropriate recommendations in order to monitor the user's behaviour within the system. Table 3 displays the user behaviour weight for the picture book reading recommendation system.

Overall, show behaviour and implicit behaviour can be used to categorise user behaviour in the

Relative reading speed	Corresponding reading status
$\text{Rate}(t) \leq 70\%$	Fast reading
$70\% < \text{Rate}(t) < 140\%$	Normal reading
$\text{Rate}(t) \geq 140\%$	Read carefully

Table 4. User Reading Status

picture book reading recommendation system. When reading picture books, users' display behaviour, such as point exchange for picture books, collection, sharing, voice following, and other behaviours, can be seen as a clear expression of their reading preferences; The implicit behaviour of users, such as reading behaviour [15], can be characterised as behaviour in which users do not overtly and directly indicate their reading preferences. The implicit user behaviour is typically not properly analysed in the majority of the historical personalised recommendation systems, which prevents accurate interest mining. In fact, if users' implicit behaviour can be broken down in depth, it will be quite helpful in identifying their true reading preferences. The user's interest in picture books cannot be effectively assessed in the picture book reading suggestion system based solely on the user's simple reading behaviour, but it may be determined from the user's reading pace of picture books. When a person reads picture books, their various reading moods reflect the variety of reading interests they have [16].

The user's reading status can be roughly divided into fast reading, normal reading and serious reading according to the user's reading speed. Because everyone's cognitive level and reading habits are different, and the reading speed of users will also vary greatly, it is necessary to consider the average reading speed of users and the average length of picture books. The relative reading speed $\text{Rate}(t)$ of picture books is defined as follows, as shown in Eq. (4):

$$\text{Rate}(t) = 0, t < t_0 \quad \text{or} \quad t > t_1; \text{Rate}(t) = \frac{txS}{L}, t_0 < t < t_1, \quad (4)$$

own in Figure 2, and the competition uses a Client/Server architecture, with the Server and Client communicating via the UDP/IP protocol, where, t_0 represents the minimum reading time of the picture book. If the user's real reading time is less than t_0 , it is considered that the user has not read the picture book carefully or has mis operated during reading; t_1 represents the maximum reading time of the picture book. If the user's real reading time is greater than t_1 , it is considered that the user has not been in the system all the time, but has left the system, to reduce the error caused by analyzing the user's reading interest. V represents the average reading speed of the user; P represents the length of the picture book. By analyzing the user's relative reading speed, the user's reading status can be obtained as shown in Table 4:

4. Performance Evaluation

This section will assess the picture book text classification model, the process for creating picture book recommendations, and the overall effectiveness of the picture book recommendation system from the viewpoint of recommendation performance. It will also analyse the rationality and efficacy of the picture book reading recommendation system. It will also evaluate the conventional recommendation algorithm based on association rules in addition to the Ebbinghaus Forgetting Curve-based recommendation algorithm suggested in this research. Get the experimental results and accompanying analysis by experimenting with and contrasting the various CF recommendation algorithms based on users [17].

4.1. Recommended System Metric

The performance evaluation indicators in the general recommendation system mainly include three types, namely, accuracy, recall and F-Measure. These three indicators are described in detail below:

User reading preferences	System recommendation	System not recommended
Like reading	true positives(tp)	false positives(fp)
Don't like reading	false negatives(fn)	true negatives(tn)

Table 5. Reading Preferences of Picture Book Users

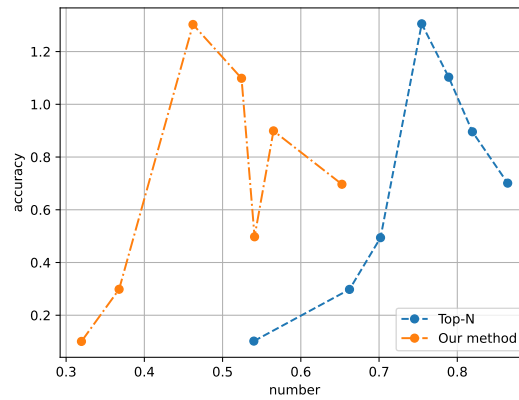


Figure 2. Comparison of Accuracy Rate of Recommended Picture Book Generation Methods

The accuracy rate is typically used in the recommendation system to measure the recommended user’s interest in the content in the recommendation list because the picture book content in the user’s reading recommendation list generated by the system does not entirely meet the user’s reading preferences. Using the picture book recommendation system in this study as an example, as shown in Table 5 the picture book recommendation results have four potential outcomes if the picture book users have never used it or contributed it to the system: 1, does the picture book simply match the user’s reading habits while the system recommends it? 2, The system suggests it but the picture book does not match the user’s reading preferences; 3, The system does not recommend it and the picture book simply matches the user’s reading preferences; and 4, The system does not recommend it and the picture book simply does not match the user’s reading preferences. Table 5 displays the reader’s preferred picture books [18].

4.2. Drawing Book Recommendation Generation Evaluation

The user-based CF recommendation algorithm matches the recommendation results according to their own optimal recommendation proportion and the proportion of interest in the category of picture books. The generation of picture book recommendations is based on the strongest association rules of picture book reading recommendation obtained by using the association rule-based recommendation algorithm. For recommendations, a few picture books that most closely match the user’s reading preferences are chosen. The experimental findings are given in Figure 2. The entire recommendation process involves calculating the proportion of interest in various picture books, the calculation of recommendation algorithms based on association rules, redundant deletion, and filtering of picture books, among other steps [19].

As observed in Figure 2, the suggested generation method of picture books put out in this research is more expressive in terms of recommendation accuracy than the conventional Top-N method.

4.3. Picture Book Recommendation System

This section primarily tests and evaluates the overall effectiveness of the entire picture book reading recommendation system. The traditional recommendation method and the recommended method designed in this paper are selected for analysis and comparison in this experiment because the Ebbing-

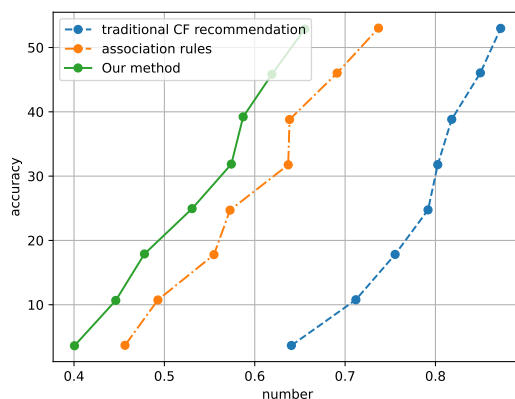


Figure 3. F-measure Comparison Chart of Different Picture Book Recommendation Methods

haus forgetting curve-based recommendation scheme adopted by this system is improved based on the traditional CF recommendation algorithm based on association rules and users. Traditional recommendation methods are mainly traditional association rule-based and user-based CF recommendation algorithms [20]. Figure 3 displays the F-Measure comparison results.

The F-Measure of each suggested strategy will gradually rise as the number of picture books rises, as seen in Figure 3. However, compared to the two conventional recommendation methods discussed above, the F-Measure of the recommendation system provided in this study is significantly greater. The recommendation approach suggested in this study takes advantage of some of its benefits while avoiding some of its drawbacks.

4.4. Relationship Prediction Effects

This work develops a manual annotation experiment, specifically examining the three dimensions of relevance, diversity, and surprise, to confirm the efficacy of the proposed algorithm (CDM) in the job of English learning complementarity discovery. Complementary product discovery based on simultaneous buy frequency (norm prob) and complementary product discovery based on simultaneous purchase condition probability (cooccur) are the two comparison approaches, respectively. In the experiment, we first compute the complementarities between English learning using three ways (including CDM), and then we select the top 150 English learning with the highest complementarities from each approach, combining them into a list of 450 English learning pairs. After the sample was randomised, we gathered five participants to respond to four questions for each English-learning pair in order to assess the complementarities. A score of 1 indicates agreement, while a score of 0 indicates disagreement [21, 22].

The issues are as follows: (1) I might purchase English Learning A and B at the same time; (2) I frequently use or consume English Learning A and B at the same time; (3) English Learning A and B are very different from one another; and (4) The collocation of English Learning A and English Learning B was unexpected, but I thought it was quite appropriate. The Kappa coefficient for the final experimental annotation result was ≥ 0.3 , indicating that the experimental findings were very reliable and could be applied to further study. Figure 4 displays the comparative findings for the three approaches' scores on the four questions.

The results in Figure 4 demonstrate that, in comparison to the method based on norm prob and cooccur, the complementary discovery algorithm CDM introduced in this paper has greatly improved the relevance, diversity, and surprise of English learning. The significant improvement of the approach in questions 1 and 2 demonstrates that the suggested method is capable of finding complements with functional complements that are more in line with the concept of economics as well as complements

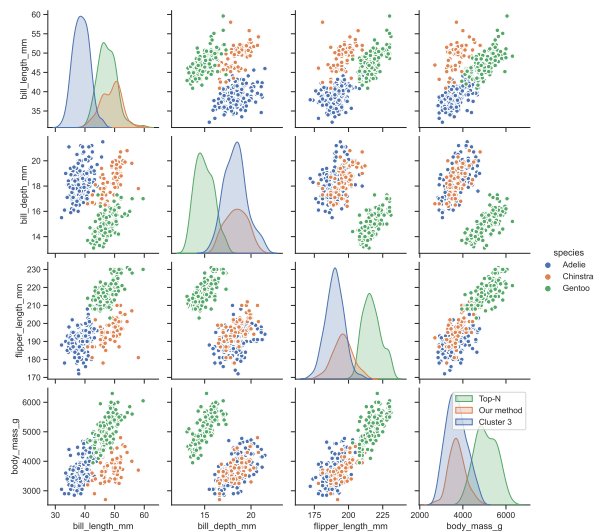


Figure 4. Results of Manual Annotation Experiment

with simultaneous purchase characteristics. This indicates that the method in this paper has a higher probability of finding different kinds of complements (such as different brands/types) and is closer to the actual situation of complements [23].

In terms of English learning diversity, CDM has achieved nearly 50% improvement in question 3 (compared with the second highest method). Additionally, CDM's performance on question 4 demonstrates that the approach used in this paper can find more varied complements, and the addition of English instruction also gives users a feeling of being new and reasonable. For instance, CDM can locate (hotpot components, mutton), (Coca Cola, Sprite) based on the frequency of spot buy, and (Cappuccino, Latte) based on the probability of purchase conditions. These three approaches are utilised to find the top 10 complements. There is no doubt that the English-learning pairs discovered using this method are more in line with the complements of users in the hotpot scenario. The English learning pairings found in the other two techniques, however, are in a different category and are more similar to the link between substitutes.

5. Conclusions

A new area of research focuses on how to employ data mining technologies to examine the accumulated huge data to offer services for teaching decision-making and learning optimisation. This study explores the relationship between learner personality traits and learning effectiveness and realises personalised learning technique recommendations. It also analyses the behaviour features of online learning. The actual entropy value that corresponds to each learner's online learning behaviour is defined and computed in order to assess the regularity of each learner's behaviour and examine the connection between regularity and final score.

Funding

This work was supported by 1) Program for the First-Class Courses Construction of Higher Learning Institutions of Shanxi(K2020298, K2021380); 2) Program for Teaching Reform and Innovation of Higher Learning Institutions of Shanxi in 2022(J20221133).

Conflict of Interest

The author declares no conflict of interests.

References

1. Sollini, M., Bartoli, F., Marciano, A., Zanca, R., Slart, R.H. and Erba, P.A., 2020. Artificial intelligence and hybrid imaging: the best match for personalized medicine in oncology. *European Journal of Hybrid Imaging*, 4, pp.1-22.
2. Zhang, Q., Yang, D., Fang, P., Liu, N. and Zhang, L., 2020. Develop academic question recommender based on Bayesian network for personalizing student's practice. *International Journal of Emerging Technologies in Learning (iJET)*, 15(18), pp.4-19.
3. Chen, S., Huang, L., Lei, Z. and Wang, S., 2020. Research on personalized recommendation hybrid algorithm for interactive experience equipment. *Computational Intelligence*, 36(3), pp.1348-1373.
4. Vedavathi, N. and Anil Kumar, K.M., 2021. An efficient e-learning recommendation system for user preferences using hybrid optimization algorithm. *Soft Computing*, 25(14), pp.9377-9388.
5. Wang, J. and Liu, S., 2019. A novel discrete particle swarm optimization algorithm for solving Bayesian network structures learning problem. *International Journal of Computer Mathematics*, 96(12), pp.2423-2440.
6. Baidada, M., Mansouri, K. and Poirier, F., 2022. Hybrid filtering recommendation system in an educational context: experiment in higher education in Morocco. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, 17(1), pp.1-17.
7. Joy, J. and Pillai, R.V.G., 2022. Review and classification of content recommenders in E-learning environment. *Journal of King Saud University-Computer and Information Sciences*, 34(9), pp.7670-7685.
8. Mikić, V., Ilić, M., Kopanja, L. and Vesin, B., 2022. Personalisation methods in e-learning-A literature review. *Computer Applications in Engineering Education*, 30(6), pp.1931-1958.
9. Ghasemi, F., Ghasemi, A. and Kalatpour, O., 2022. Prediction of human error probability during the hydrocarbon road tanker loading operation using a hybrid technique of fuzzy sets, Bayesian network and CREAM. *International Journal of Occupational Safety and Ergonomics*, 28(3), pp.1342-1352.
10. Yuan, Q., 2020. RETRACTED: Network education recommendation and teaching resource sharing based on improved neural network. *Journal of Intelligent & Fuzzy Systems*, 39(4), pp.5511-5520.
11. Chen, J., Zhang, R., Dong, X., Lin, L., Zhu, Y., He, J., Christiani, D.C., Wei, Y. and Chen, F., 2019. shinyBN: an online application for interactive Bayesian network inference and visualization. *BMC Bioinformatics*, 20, pp.1-5.
12. Pirasteh, P., Bouguelia, M.R. and Santosh, K.C., 2021. Personalized recommendation: an enhanced hybrid collaborative filtering. *Advances in Computational Intelligence*, 1(4), pp.1-8.
13. Kotios, D., Makridis, G., Fatouros, G. and Kyriazis, D., 2022. Deep learning enhancing banking services: a hybrid transaction classification and cash flow prediction approach. *Journal of Big Data*, 9(1), p.100.
14. Lytvyn, V., Vysotska, V., Shatskykh, V., Kohut, I., Petruchenko, O., Dzyubyk, L., Bobrivetc, V., Panasyuk, V., Sachenko, S. and Komar, M., 2019. Design of a recommendation system based on Collaborative Filtering and machine learning considering personal needs of the user. *East European Journal of Advanced Technologies*, (4 (2)), pp.6-28.
15. Gil-Begue, S., Bielza, C. and Larrañaga, P., 2021. Multi-dimensional Bayesian network classifiers: A survey. *Artificial Intelligence Review*, 54(1), pp.519-559.

16. Xing, Y., Lv, C. and Cao, D., 2019. Personalized vehicle trajectory prediction based on joint time-series modeling for connected vehicles. *IEEE Transactions on Vehicular Technology*, 69(2), pp.1341-1352.
17. Du, G., Liang, X., Ouyang, X. and Wang, C., 2021. Risk prediction of hypertension complications based on the intelligent algorithm optimized Bayesian network. *Journal of Combinatorial Optimization*, 42, pp.966-987.
18. Sharma, S., Rana, V. and Malhotra, M., 2022. Automatic recommendation system based on hybrid filtering algorithm. *Education and Information Technologies*, 27(2), pp.1523-1538.
19. García-Gutierrez, F., Díaz-Álvarez, J., Matias-Guiu, J.A., Pytel, V., Matías-Guiu, J., Cabrera-Martín, M.N. and Ayala, J.L., 2022. GA-MADRID: Design and validation of a machine learning tool for the diagnosis of Alzheimer's disease and frontotemporal dementia using genetic algorithms. *Medical & Biological Engineering & Computing*, 60(9), pp.2737-2756.
20. Khanal, S.S., Prasad, P.W.C., Alsadoon, A. and Maag, A., 2020. A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies*, 25(4), pp.2635-2664.
21. Liu, S., Wang, B., Yang, L.T. and Philip, S.Y., 2021. HNF: Hybrid neural filtering based on centrality-aware random walk for personalized recommendation. *IEEE Transactions on Network Science and Engineering*, 9(3), pp.1056-1066.
22. Shen, H., Liu, T. and Zhang, Y., 2020. Discovery of learning path based on bayesian network association rule algorithm. *International Journal of Distance Education Technologies (IJDET)*, 18(1), pp.65-82.
23. Zhou, J., Zhang, D., Ren, W. and Zhang, W., 2022. Auto color correction of underwater images utilizing depth information. *IEEE Geoscience and Remote Sensing Letters*, 19, pp.1-5.



©2024 the Author(s), licensee Combinatorial Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)