Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Personalized Course Recommendation System in E-learning Platforms Using Machine Learning-Based Content Filtering

Rashmi Prakashpant Bijwe[1], Dr. Anjali B. Raut[2]

[1]*Assistant Professor, Department of Computer Science and Engineering, H.V.P.M's College of Engineering and Technology, Amravati, India*
[2]*I/C Principal, Professor & HOD, Department of Computer Science and Engineering, H.V.P.M's College of Engineering and Technology, Amravati, India*

With the rapid expansion of e-learning educational platforms, there is a pressing need to develop personalized course recommendation systems to keep up with student demand and improve the quality of online education. A personalized system for course recommendations built on top of content filtering techniques based on Machine Learning (ML) is introduced in this paper. This method employs the descriptive attributes of courses, such as titles, ratings, summaries, and skills, to provide personalized recommendations for users based on the similarity of course content. The characteristics from textual data are extracted using TF-IDF vectorization with cosine similarity and then ML techniques are applied to estimate course relevance. The efficacy of the proposed system is evaluated utilizing a Coursera benchmark dataset, achieving high accuracy and minimal mean square error rate in forecasting user preferences and providing substantial recommendations. The results also show that content-based filtering can improve e-learning system's user engagement and satisfaction.

Keywords: Course Recommendation, Regressor Model, Machine Learning, Personalized Learning.

1. Introduction

The rapid development of these platforms has significantly influenced the evolution of online learning environments, particularly in the aftermath of the COVID-19 epidemic. Millions of students have benefited from online educational learning platforms, which provide a diverse range of courses on topics ranging from programming to philosophy. The remarkable growth of course catalogs has made it challenging for students to choose the best courses that fit their needs, interests, and skill levels. To rectify these issues and boost user engagement and satisfaction, a personalized recommendation system is necessary to guide students to the best classes for their needs [1]. Traditional recommendation systems in e-learning environments have mostly employed collaborative filtering or hybrid approaches. Collaborative filtering can predict user preferences using past interactions, such as course ratings or descriptions. While this approach has many advantages, it frequently encounters the cold-start problem when suggesting new users or courses without enough interaction data. Furthermore, collaborative methods may face difficulties due to data sparsity when user interaction data is inadequate or fragmented [2].

In contrast, content-based filtering utilizes the intrinsic attributes of the courses, such as titles, descriptions, skills, and ratings, to provide recommendations. Since this approach is not dependent on data derived from user interactions, it is especially useful for less popular courses or new users. Assessment and comparison of course content based on similarities may be accomplished through the use of content-based filtering, which can produce extremely relevant suggestions [3-4]. This paper develops an ML-based content filtering system for individualized course recommendations based on textual features extracted from NLP. The findings demonstrate the potential of content-based filtering to supplement or replace traditional collaborative approaches in environments with little user interaction data. Delivering a strong, scalable, and effective recommendation system that addresses the challenges of modern e-learning platforms is the goal of this paper. The objective is to improve the learning process by offering tailored assistance and direction to each learner. The primary contributions of this research are:

1. Developed a content-based recommendation framework that leverages course metadata for personalized recommendations using ML techniques.

2. Employed TF-IDF for feature extraction and integrating ML algorithms for improved relevance prediction on a real-world Coursera dataset to demonstrate its practical utility.

This paper is structured as follows for the following portions. The second section provides a literature assessment of important research on recommendation systems. The details of the proposed strategy are described in Section 3. Section 4 details the experimental outcomes. A prospective viewpoint on the future is presented in Section 6, which serves as the paper's conclusion.

2. LITERATURE REVIEW

In recent years, there have been several studies conducted on the development of personalized recommendation systems for e-learning. This section highlights the research gaps that the proposed system seeks to address and takes a look at major developments and methods in recommendation systems, with a focus on content-based filtering. MetaCDR for a cold-start recommendation is introduced which includes both meta-optimization and domain knowledge [5]. Intelligent E-Learning platform presented that recommends the most advantageous courses for each student [6]. A resilient and flexible M-learning model is suggested that dynamically investigates learning characteristics, their associated weights, and relationships for M-learners [7]. A hybrid recommendation system framework was developed for course selection [8]. ML-based recommender model for personalized course learning items proposed which integrates the student's real ratings with their learning patterns [9]. An educational recommendation system may determine a person's unique interests in educational materials and then propose resources that are a good fit for those interests [10]. CPT CPT-based model for analogous students was developed to propose university-level courses [11]. Pertinent courses via social media are selected by employing a clustering approach through the establishment of learner profiles and deep flamingo search reinforcement learning [12]. Intelligent systems developed based on neural networks that yielded optimal results [13]. A proficient E-learning recommender system assists users in selecting a course according to their requirements [14]. DL-based course recommender system has been suggested to comprehensively collect course attribute information and high-level user behaviors [15].

Collaborative filtering with the educational domain knowledge graph presented by collecting semantic information about the course, and building a course knowledge graph to supplement the grade prediction process [16]. ML-based course enrollment recommender system was developed to provide individualized advice to students planning to attend classes in the forthcoming semester [17]. Course recommendation systems are developed User-based attributes with profile and rating information [18]. Learners may access a central hub where they can browse courses offered by different providers and get personalized course recommendations based on their actions [19]. An educational domain knowledge graph and collaborative filtering are integrated with a grade prediction model to gather course semantic information and generate a course knowledge graph [20]. The efficacy of ML algorithms in predicting request-based websites has been analyzed, revealing that the Random Forest method has the greatest prediction accuracy at 98.98% [21]. Using course characteristics, academic records, and evaluations of course learning outcomes, a variety of ML algorithms are suggested for use in reinforcement learning to predict suitable behavior [22]. A smart e-learning architecture utilizing Markov Decision Process (MDP) and Reinforcement Learning (RL) can improve the learning experience for individual students by offering a tailored and efficient learning trajectory [23]. RS is suggested employing the collaborative filtering approach for e-learning course suggestions with SVD, KNN, and NCF models [24]. Widely utilized ML models in e-learning content recommendation platforms to precisely predict client interests and preferences [25].

3. PROPOSED METHODOLOGY

The proposed personalized course recommendation system utilizes a machine learning-driven content filtering method that exploits course information to produce customized suggestions. This section delineates the dataset utilized, the preprocessing procedures, the feature extraction techniques, and the algorithms applied to execute the recommendation system, shown in Figure 1.

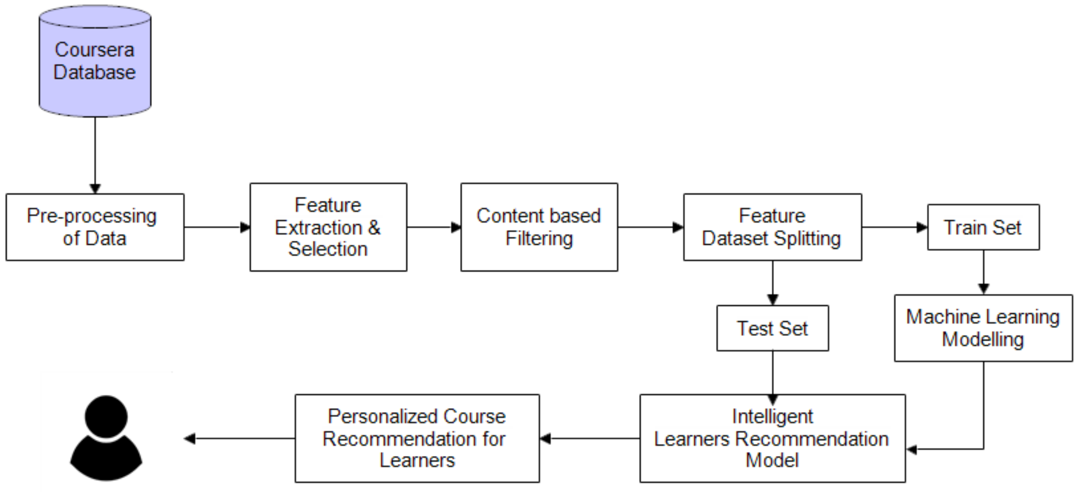

Figure 1: Course Recommendation System

Coursera Dataset

The system is developed and assessed utilizing a benchmark dataset [6] obtained from Coursera, an online educational platform. The abundance of both paid and free online resources might overwhelm students seeking to acquire new abilities. With this data, recommender systems can be built that use students' skill sets and reported difficulty levels to recommend courses. The Course Link is further supplied, facilitating quick access through the Recommender System. The collection comprises organized information for courses, encompassing:

- Course Name: Name of the Course.
- Difficulty Level: It has 3 values: Advanced, Beginner and Intermediate.
- URL: The Course URL
- Course Description: The Description of the Course.
- University: Course is offered by a university or an industry partner.
- Skills: Skill Tags associated with the Course extracted through NLP processes.
- Course Rating: Assessment using a 5-point scale with a 0.1-step minimum.

Data Preprocessing

To ensure meaningful analysis, the dataset undergoes a series of preprocessing steps. In text cleaning, special characters, numbers, and punctuation from the textual attributes (title, description, tags) are removed. All text to lowercase converted to ensure uniformity. Tokenization is a method for decomposing text into its component words. The goal of stop-word removal is to rid a text of overused words that do not add anything to the meaning itself. Words are reduced to their base forms using libraries like NLTK or SpaCy, improving generalization. Missing values are imputed in metadata fields with placeholders to avoid computation errors. At last, combine title, description, and tags into a single combined features column for a holistic representation of course content.

Feature Extraction

To represent the textual data numerically, the following techniques are applied:

1. Term Frequency-Inverse Document Frequency (TF-IDF) Vectorization:

It is used to convert text data into numerical feature vectors. Also, it emphasizes important words in a document while reducing the influence of frequently occurring but less meaningful terms (e.g., "course"). Each course is represented as a high-dimensional vector in the feature space, defined in following equations.

TF – IDF(t,d) = TF(t,d) x IDF(t)

$$IDF(t) = \log\left(\frac{N}{1+n_t}\right)$$

Where, N - Aggregate document count, nt - Number of documents containing term t.

2. Cosine Similarity:

The pairwise cosine similarity between courses are computed based on their TF-IDF vectors. Courses with higher cosine similarity are considered more relevant to each other.

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\|\|B\|}$$

Content-Based Filtering

The core of the recommendation system relies on the similarity between courses. To find courses that are similar to the ones a user has engaged with, TF-IDF vectors were utilized to calculate pairwise cosine similarity. The content-based recommendation system uses the extracted features to generate recommendations. The key steps include:

Step1: Query Course Identification: For a given course that the user has interacted with, extract its feature vector using TF-IDF.

Step 2: Similarity Computation: Calculate cosine similarity between the query course vector and all other course vectors in the dataset.

Step 3: Ranking: Rank the courses based on their similarity scores, with the highest-ranking courses being the most relevant.

Step 4: Top-N Recommendations: Select the top-N courses as recommendations for the user.

Machine Learning Model

To improve customization, four ML models—Linear Regression, Random Forest Regression, and Decision Tree Regression—are trained, verified, and tested to identify course ratings. Forecasted categories were subsequently employed to filter and rank analogous courses, guaranteeing consistency with user preferences. A supervised learning model is employed to categorize courses and improve suggestions for the recommendation system.

The TF-IDF characteristics and information, such as category, serve as input for the classification model. A regression classifier model developed to forecast the category of a course based on its textual characteristics. Hyperparameter tweaking is done by grid search and cross-validation to enhance model performance. Refine the recommendations to correspond with the anticipated category of user interest, guaranteeing tailored congruence.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed methodology for the course recommender system utilizing machine learning techniques is executed in the PyCharm Integrated Development Environment using the Python programming language. The standard benchmark dataset utilized for constructing the course recommender system was Coursera's dataset from Kaggle [6]. The operating system is Windows 11, with a configuration of Intel i5, 16GB RAM, and a 4GB GPU. The performance of the proposed models was evaluated using standard benchmark metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) Score.

There was an 80-20% ratio used to train and test data points in the dataset. For optimal accuracy, the ML model's hyperparameters and TF-IDF matrix dimensions were fine-tuned. Figure 1 shows the histogram data distribution of "difficulty level" from a dataset that has six different categories with a count of course types in Figure 2.

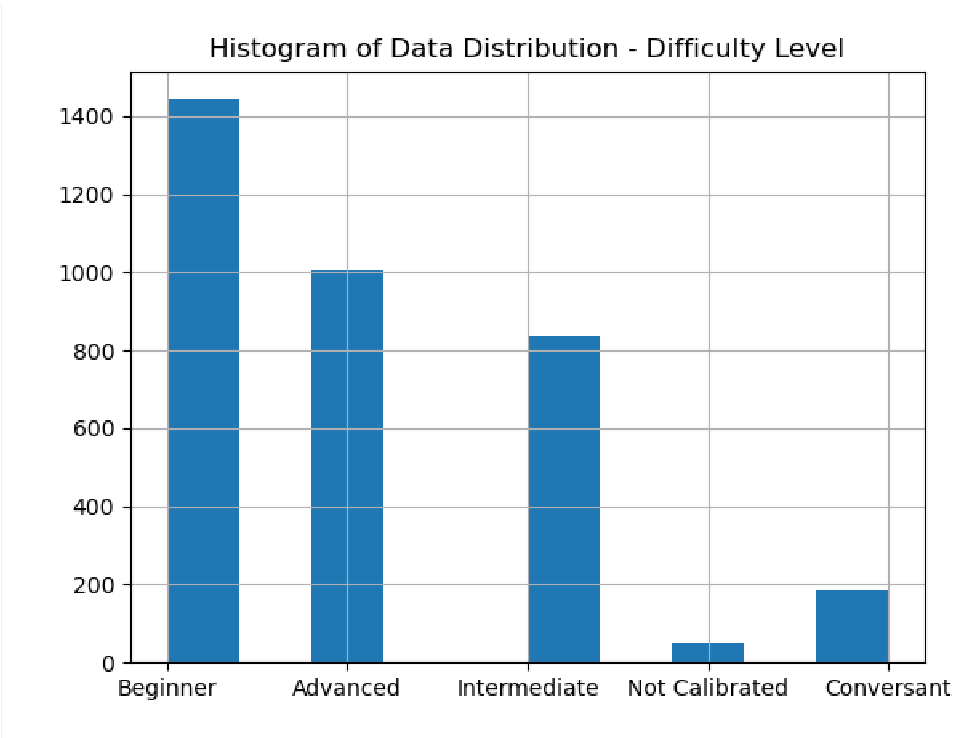

Figure 2: Histogram plot of feature attribute

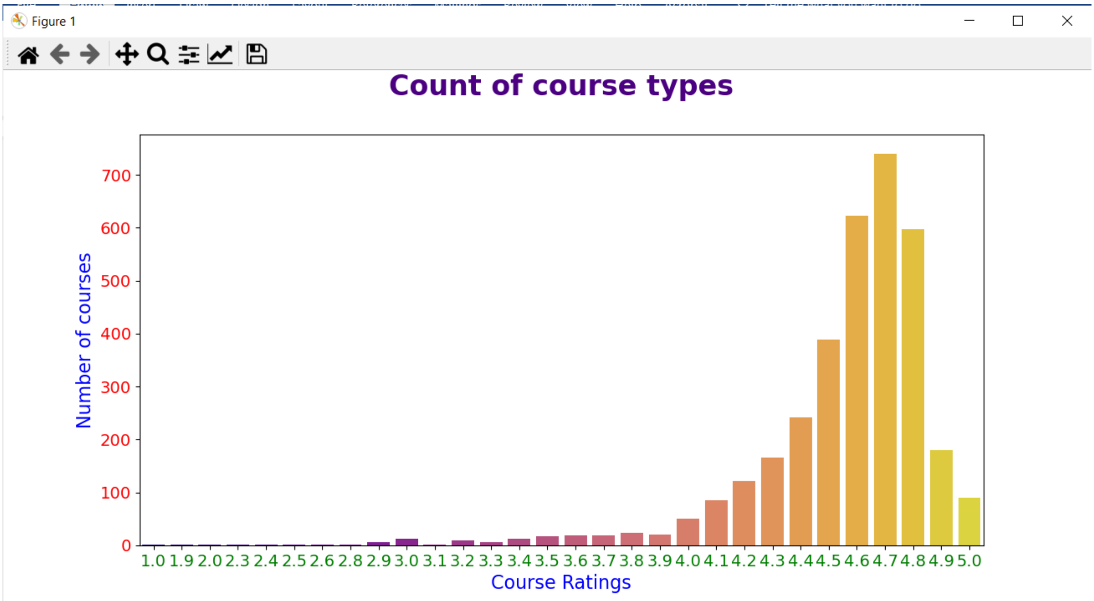

Figure 3: Count plot of course types as per ratings

Figures 4 and 5 show the query course name tested and the top "N" recommended courses to uses as per their preferences. Here, "N" is assumed as 5. There are two types of courses are given as input. Recommended results show that as per the query, courses recommended are matched.

```
Course Name:  Silicon Thin Film Solar Cells
Courses to recommend to user:
-----------------------------------------------------------------
Physics of silicon solar cells
Introduction to solar cells
Organic Solar Cells - Theory and Practice
Solar Energy Systems Overview
Photovoltaic solar energy
-----------------------------------------------------------------
```

Figure 4: Course Recommendations

```
Course Name:  Python Programming Essentials
Courses to recommend to user:
-----------------------------------------------------------------
Python Data Representations
Python Data Analysis
Python Basics
Programming for Everybody (Getting Started with Python)
Python Functions Files and Dictionaries
-----------------------------------------------------------------
```

Figure 5: Course Recommendations

Table 1 discusses the performance evaluation while course rating prediction from different ML-based regressor models using MSE, MAR, and R2 scores. It shows that minimal MSE is obtained which indicates that the actual rating and predicted ratings are nearly equal matched.

Table 1: Performance Evaluation of Different Models for Course Rating Predictions

Regressor Model	MSE	MAE	R2 Score
Linear Regression	0.0524	0.1766	0.2414
Random Forest	0.0366	0.1493	0.1319
Decision Tree	0.0720	0.1998	0.70539

The system demonstrated robust performance in identifying relevant courses, with qualitative feedback indicating improved user satisfaction. The proposed system effectively recommends personalized courses by analysing course metadata and aligning suggestions with user preferences. The content-based approach effectively utilizes course metadata to recommend relevant courses to users. The use of TF-IDF vectorization ensures that the system focuses on significant terms in course descriptions, while cosine similarity provides an efficient metric for matching user profiles with courses. However, the system's reliance on textual data may lead to limited diversity in recommendations.

5. CONCLUSION

This paper introduced a tailored course recommendation system utilizing machine learning-based content filtering on the Coursera dataset. The system attains high accuracy in aligning suggestions with user preferences by employing course metadata and descriptions, therefore addressing the scalability and diversity concerns of e-learning platforms. The suggested method shows considerable promise for improving user engagement and educational results in online learning environments. Future developments of this research may investigate hybrid methodologies using sophisticated NLP models and real-time implementation to further improve recommendation precision and user happiness.

References

[1] H. Oubalahcen, L. Tamym, and M. L. D. E. Ouadghiri, “The Use of AI in E-Learning Recommender Systems: A Comprehensive Survey,” Procedia Computer Science, vol. 224, pp. 437–442, Jan. 2023, doi: 10.1016/j.procs.2023.09.061.

[2] D. G. M., R. H. Goudar, A. A. Kulkarni, V. N. Rathod and G. S. Hukkeri, "A Digital Recommendation System for Personalized Learning to Enhance Online Education: A Review," in IEEE Access, vol. 12, pp. 34019-34041, 2024, doi: 10.1109/ACCESS.2024.3369901.

[3] I. Khan, X. Zhang, M. Rehman and R. Ali, "A Literature Survey and Empirical Study of Meta-Learning for Classifier Selection," in IEEE Access, vol. 8, pp. 10262-10281, 2020, doi: 10.1109/ACCESS.2020.2964726.

[4] M. Murtaza, Y. Ahmed, J. A. Shamsi, F. Sherwani and M. Usman, "AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions," in IEEE Access, vol. 10, pp. 81323-81342, 2022, doi: 10.1109/ACCESS.2022.3193938.

[5] R. Guan, H. Pang, F. Giunchiglia, Y. Liang and X. Feng, "Cross-Domain Meta-Learner for Cold-Start Recommendation," in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 8, pp. 7829-7843, 1 Aug. 2023, doi: 10.1109/TKDE.2022.3208005.

[6] S. Amin, M. I. Uddin, W. K. Mashwani, A. A. Alarood, A. Alzahrani and A. O. Alzahrani, "Developing a Personalized E-Learning and MOOC Recommender System in IoT-Enabled Smart Education," in IEEE Access, vol. 11, pp. 136437-136455, 2023, doi: 10.1109/ACCESS.2023.3336676.

[7] M. Adnan, A. Habib, J. Ashraf, B. Shah and G. Ali, "Improving M-Learners’ Performance Through Deep Learning Techniques by Leveraging Features Weights," in IEEE Access, vol. 8, pp. 131088-131106, 2020, doi: 10.1109/ACCESS.2020.3007727.

[8] S. S, B. P. Sankaralingam, A. Gurusamy, S. Sehar, and D. P. Bavirisetti, “Personalization-based deep hybrid E-learning model for online course recommendation system,” PeerJ Computer Science, vol. 9, p. e1670, Nov. 2023, doi: 10.7717/peerj-cs.1670.

[9] S. M. Nafea, F. Siewe and Y. He, "On Recommendation of Learning Objects Using Felder-Silverman Learning Style Model," in IEEE Access, vol. 7, pp. 163034-163048, 2019, doi: 10.1109/ACCESS.2019.2935417.

[10] H. A. Yazdi, S. J. S. Mahdavi, and H. A. Yazdi, “Dynamic educational recommender system based on Improved LSTM neural network,” Scientific Reports, vol. 14, no. 1, Feb. 2024, doi: 10.1038/s41598-024-54729-y.

[11] G. George and A. M. Lal, "PERKC: Personalized kNN With CPT for Course Recommendations in Higher Education," in IEEE Transactions on Learning Technologies, vol. 17, pp. 885-892, 2024, doi: 10.1109/TLT.2023.3346645.

[12] N. Vedavathi and R. Suhas Bharadwaj, “Deep Flamingo Search and Reinforcement Learning Based Recommendation System for E-Learning Platform using Social Media,” Procedia Computer Science, vol. 215, pp. 192–201, 2022, doi: https://doi.org/10.1016/j.procs.2022.12.022.

[13] T. Ahajjam, M. Moutaib, H. Aissa, M. Azrour, Y. Farhaoui and M. Fattah, "Predicting Students' Final Performance Using Artificial Neural Networks," in Big Data Mining and Analytics, vol. 5, no. 4, pp. 294-301, December 2022, doi: 10.26599/BDMA.2021.9020030.

[14] M. Chakkaravarthy and G. S. Kumar, “A Quality of Experience-based Recommender System for E-learning Resources,” International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 5s, pp. 01–04, May 2023, doi: https://doi.org/10.17762/ijritcc.v11i5s.6590.

[15] Q. Li and J. Kim, “A Deep Learning-Based Course Recommender System for Sustainable Development in education,” Applied Sciences, vol. 11, no. 19, p. 8993, Sep. 2021, doi: 10.3390/app11198993.

[16] Y. Zhang, V. Y. Mariano and R. P. Bringula, "Prediction of Students’ Grade by Combining

Educational Knowledge Graph and Collaborative Filtering," in IEEE Access, vol. 12, pp. 68382-68392, 2024, doi: 10.1109/ACCESS.2024.3390675.
[17] X. Wang, L. Cui, M. Bangash, M. Bilal, L. Rosales, and W. Chaudhry, “A Machine Learning-based Course Enrollment Recommender System,” Jan. 2022, doi: https://doi.org/10.5220/0011109100003182.
[18] P. V. Kulkarni, S. Rai, R. Sachdeo, and R. Kale, “Deep learning-based Educational User Profile and User Rating Recommendation System for E-Learning,” Journal of Information Systems and Telecommunication (JIST), vol. 11, no. 43, pp. 185–195, Aug. 2023, doi: 10.61186/jist.27448.11.43.185.
[19] K. Ramneet, G. Deepali, and M. Mani, “Learner-Centric Hybrid Filtering-Based Recommender System for massive open online courses,” International Journal of Performability Engineering, vol. 19, no. 5, p. 324, Jan. 2023, doi: 10.23940/ijpe.23.05.p4.324333.
[20] Y. Zhang, V. Y. Mariano and R. P. Bringula, "Prediction of Students’ Grade by Combining Educational Knowledge Graph and Collaborative Filtering," in IEEE Access, vol. 12, pp. 68382-68392, 2024, doi: 10.1109/ACCESS.2024.3390675.
[21] G. S. Hukkeri and R. H. Goudar, "Machine Learning-Based Personalized Recommendation System for E-Learners," 2022 Third International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2022, pp. 1-6, doi: 10.1109/ICSTCEE56972.2022.10100069.
[22] N. Yanes, A. M. Mostafa, M. Ezz, and S. N. Almuayqil, “A Machine Learning-Based Recommender System for Improving Students Learning Experiences,” IEEE Access, vol. 8, pp. 201218–201235, 2020, doi: https://doi.org/10.1109/access.2020.3036336.
[23] S. Amin, M. I. Uddin, A. A. Alarood, W. K. Mashwani, A. Alzahrani and A. O. Alzahrani, "Smart E-Learning Framework for Personalized Adaptive Learning and Sequential Path Recommendations Using Reinforcement Learning," in IEEE Access, vol. 11, pp. 89769-89790, 2023, doi: 10.1109/ACCESS.2023.3305584.
[24] K. K. Jena et al., “E-Learning Course Recommender System Using Collaborative Filtering Models,” Electronics, vol. 12, no. 1, p. 157, Jan. 2023, doi: https://doi.org/10.3390/electronics12010157.
[25] K. Wan-Er, et al. “An E-Learning Recommendation System Framework”. International Journal on Advanced Science, Engineering and Information Technology, vol. 14, no. 1, Feb. 2024, pp. 10-19, doi:10.18517/ijaseit.14.1.19043.