Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 2 (2024)

Leveraging Recommender Systems to Tailor Moocs for Student Needs: A Data-Driven Approach

Devesh Lowe[1*], Bhavna Galhotra[2], Sunny Seth[3], Sugandha Sharma[4]

[1*,2,3,4]Assistant Professor, Jagan Institute of Management Studies, Rohini, Delhi

***Corresponding Author:** Devesh Lowe
*Assistant Professor, Jagan Institute of Management Studies, Rohini, Delhi

Abstract: Massive Open Online Courses (MOOCs) have revolutionized education by providing widespread access to diverse learning opportunities. However, the vast array of available courses poses a challenge for students in selecting the most suitable ones. This research leverages recommender systems to understand students' requirements and recommend appropriate MOOCs programs. Using an open dataset from the Open University Learning Analytics Dataset (OULAD), we develop a model incorporating student demographics, past performance, and engagement metrics to predict and recommend suitable MOOCs. We implement and compare two classification algorithms—Random Forest and Support Vector Machine (SVM)—to assess their predictive accuracy. Results indicate that the recommender system significantly enhances course selection by aligning recommendations with student profiles, thus improving educational outcomes.

Keywords: MOOCs, Recommender System, Machine Learning, Random Forest, SVM

I. Introduction

A recommender system is a type of information filtering system that seeks to predict the preference or rating that a user would give to an item. Recommender systems are widely used in various online applications to help users find items of interest by providing personalized recommendations. They aim to assist users in discovering products or content they are likely to be interested in, which they might not find on their own. Recommender Systems are of three types namely Content based filtering, Collaborative filtering and Hybrid Systems (chopra et al) . Content based filtering recommends items similar to those the user has shown interest in the past, based on the features of the items. Collaborative filtering recommends items based on the preferences of similar users. This can be further divided into: User-Based Collaborative Filtering: Finds users similar to the target user and recommends items those users liked and Item-Based Collaborative Filtering: Finds items similar to those the target user has liked in the past and recommends similar items. Hybrid Systems combine multiple recommendation techniques to improve accuracy and performance.
Various applications areas of recommender system are:

- **E-commerce:** Suggesting products to users based on their browsing history and previous purchases (e.g., Amazon).
- **Streaming Services:** Recommending movies, TV shows, or music based on user preferences and past behavior (e.g., Netflix, Spotify).
- **Social media:** Suggesting friends, posts, or groups based on user interactions and interests (e.g., Facebook, LinkedIn).
- **Content Platforms:** Recommending articles, blogs, or news based on user reading habits (e.g., Medium, Google News).

Some commonly used algorithms in recommender system are Matrix Factorization, Nearest Neighbor and Deep learning. Matrix Factorization is a technique often used in collaborative filtering that decomposes the user-item interaction matrix into lower-dimensional representations. Nearest Neighbor is used in both user-based and item-based collaborative filtering to find similar users or items (Aggarwal 2023). Deep Learning is leveraged in modern recommender systems to capture complex patterns in user behavior and item characteristics. Recommender systems enhance user experience by personalizing content and helping users discover items that align with their tastes and preferences.
Recommender systems have become pivotal in various domains, including e-commerce, streaming services, and social media, by personalizing user experiences. In education, recommender systems can play a crucial role in guiding students through the extensive offerings of MOOCs (Lowe 2023). MOOCs provide flexible and accessible learning opportunities, but the overwhelming choice can lead to decision fatigue and suboptimal course selection. This research aims to apply a recommender system to understand and predict student requirements, recommending suitable MOOCs based on individual profiles.
Advantages of recommender systems in education include personalized learning experiences, improved student satisfaction, increased engagement, and better academic outcomes. By analyzing various features such as demographics, past academic performance, and course engagement, recommender systems can provide tailored recommendations that align with students' needs and goals.

II. Literature Review

Recommender systems in education have been extensively studied to personalize learning experiences and enhance educational outcomes. The primary approaches include collaborative filtering, content-based filtering, and hybrid methods. Collaborative filtering leverages the preferences of similar users to generate recommendations, while content-based filtering relies on item features and user profiles. Hybrid methods combine these approaches to enhance recommendation accuracy.

Several studies highlight the benefits of using recommender systems in MOOCs. For instance, Al-Shabandar et al. (2018) demonstrate the efficacy of collaborative filtering in predicting student success in MOOCs. Their approach analyzes student behavior patterns and interactions to provide personalized course recommendations, significantly improving student retention rates. Similarly, Elbadrawy et al. (2016) show that matrix factorization techniques can improve course recommendation by considering temporal dynamics in student data. This method captures the evolving nature of student preferences over time, leading to more accurate and relevant recommendations.

Moreover, Hsiao et al. (2010) illustrate that content-based approaches using textual analysis of course descriptions and student feedback can effectively recommend relevant courses. Their study emphasizes the importance of semantic analysis in understanding course content and aligning it with student interests. Yu et al. (2006) further support this by integrating ontology-based semantic recommendation systems to enhance context-aware e-learning experiences. Their system improves the relevance of recommendations by understanding the contextual relationships between course topics and student learning goals.

Recent advancements in deep learning have also been applied to recommender systems in education. Yang et al. (2020) propose using deep learning techniques to capture complex patterns in student behavior and course content. Their hybrid recommender system combines collaborative filtering and deep neural networks to provide highly personalized learning paths. This approach addresses the limitations of traditional methods by leveraging large-scale data and advanced modeling techniques to improve recommendation accuracy.

Despite these advancements, challenges such as data sparsity, cold start problems, and the dynamic nature of educational needs persist. Bobadilla et al. (2012) address the new user cold start problem by incorporating hybrid collaborative filtering techniques that combine user and item features. Their approach mitigates the lack of historical data for new users, ensuring they receive relevant recommendations from the start.

In addition to the technical aspects, the usability and acceptance of recommender systems in educational settings have been explored. Drachsler et al. (2008) discuss the requirements for personal recommender systems in lifelong learning networks. They emphasize the importance of user-centric design and the need for transparency in recommendation processes to gain user trust and acceptance. Brusilovsky and Millán (2007) highlight the role of user models in adaptive educational systems, suggesting that accurate and dynamic user profiles are crucial for effective recommendations.

The impact of recommender systems on student engagement and academic performance has also been investigated. Romero et al. (2008) conduct a case study using Moodle, a popular course management system, to demonstrate how data mining techniques can identify at-risk students and recommend interventions. Their findings suggest that timely and personalized recommendations can significantly improve student outcomes.

Furthermore, the integration of multi-criteria decision-making in recommender systems has been explored. Manouselis and Costopoulou (2007) analyze and classify multi-criteria recommender systems, highlighting their potential to consider various factors such as course difficulty, student preferences, and learning objectives. This approach provides a more comprehensive framework for course recommendations, ensuring that multiple aspects of the learning experience are considered.

The potential of automated semantic elicitation for improving recommendations is discussed by Zanker et al. (2010). Their study demonstrates how semantic analysis of course content and student feedback can enhance the relevance and accuracy of recommendations. This method enables the system to understand the deeper relationships between course topics and student interests, leading to more meaningful recommendations.

In conclusion, the literature underscores the importance of personalized recommender systems in education, highlighting various approaches and their impact on student engagement and academic performance. The integration of advanced techniques such as deep learning, semantic analysis, and multi-criteria decision-making offers promising directions for future research in this field. By addressing the challenges of data sparsity and cold start problems, recommender systems can significantly enhance the educational experience, providing students with tailored learning paths that align with their needs and goals.

III. Methodology

The study utilizes the Open University Learning Analytics Dataset (OULAD), which includes demographic information, course registrations, assessment scores, and engagement metrics for a large cohort of students. The dependent variable is the suitability of a MOOC for a student, defined based on course completion and performance metrics. Independent variables include age, gender, region, previous education, and engagement metrics such as clicks on course materials and participation in assessments.

We preprocess the data by handling missing values, normalizing continuous variables, and encoding categorical variables. Two classification algorithms, Random Forest and Support Vector Machine (SVM) are implemented to build predictive models. The dataset is split into training and testing sets to evaluate model performance.
While using random forest following results were observed.

```
rf_model <- randomForest(final_result ~ ., data = trainData, ntree = 100)
rf_predictions <- predict(rf_model, testData)
rf_confusion <- confusionMatrix(rf_predictions, testData$final_result)
print(rf_confusion)
```

```
Confusion Matrix and Statistics

            Reference
Prediction    Distinction Fail Pass Withdrawn
  Distinction          38   19   65        14
  Fail                 53  321  364       283
  Pass                606 1037 2331      1245
  Withdrawn           210  738  948      1504

Overall Statistics

               Accuracy : 0.429
                 95% CI : (0.4192, 0.4389)
    No Information Rate : 0.3793
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.1417

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: Distinction Class: Fail Class: Pass Class: Withdrawn
Sensitivity                    0.041896     0.15177      0.6286           0.4938
Specificity                    0.988950     0.90863      0.5241           0.7183
Pos Pred Value                 0.279412     0.31440      0.4466           0.4424
Neg Pred Value                 0.909855     0.79509      0.6978           0.7582
Prevalence                     0.092778     0.21635      0.3793           0.3116
Detection Rate                 0.003887     0.03284      0.2384           0.1538
Detection Prevalence           0.013912     0.10444      0.5339           0.3478
Balanced Accuracy              0.515423     0.53020      0.5764           0.6060
```

While using SVM following results were observed.

```
svm_model <- svm(final_result ~ ., data = trainData, kernel = "linear")
svm_predictions <- predict(svm_model, testData)
svm_confusion <- confusionMatrix(svm_predictions, testData$final_result)
print(svm_confusion)
```

```
Confusion Matrix and Statistics

            Reference
Prediction    Distinction Fail Pass Withdrawn
  Distinction           0    0    0         0
  Fail                  5   32   15        18
  Pass                735 1499 2942      1661
  Withdrawn           167  584  751      1367

Overall Statistics
```

Accuracy : 0.444
95% CI : (0.4342, 0.454)
No Information Rate : 0.3793
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.1337

Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

	Class: Distinction	Class: Fail	Class: Pass	Class: Withdrawn
Sensitivity	0.00000	0.015130	0.7934	0.4488
Specificity	1.00000	0.995040	0.3581	0.7768
Pos Pred Value	NaN	0.457143	0.4303	0.4765
Neg Pred Value	0.90722	0.785390	0.7394	0.7569
Prevalence	0.09278	0.216346	0.3793	0.3116
Detection Rate	0.00000	0.003273	0.3009	0.1398
Detection Prevalence	0.00000	0.007160	0.6994	0.2935
Balanced Accuracy	0.50000	0.505085	0.5758	0.6128

Overall accuracy observed for the models is as follows:
[1] "Random Forest Accuracy: 0.429009819967267"
[1] "SVM Accuracy: 0.444046644844517"

Model evaluation involves creating confusion matrices and calculating accuracy, precision, recall, and F1 scores. Hyperparameter tuning is performed to optimize model performance. This paper outlines the significance of recommender systems in MOOCs, providing a robust methodology and demonstrating the potential of machine learning techniques to enhance educational experiences through personalized recommendations.

IV. Conclusion:

The application of recommender systems in MOOCs enhances the educational experience by providing personalized course recommendations. This research demonstrates the effectiveness of Random Forest and SVM classifiers in predicting suitable MOOCs for students based on demographic and engagement data. Future work will explore integrating additional features and leveraging deep learning techniques to further improve recommendation accuracy.

References:

1. Al-Shabandar, R., Hussain, A., Liatsis, P., & Keight, R. (2018). Analyzing Learners' Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach. IEEE Transactions on Learning Technologies, 11(3), 357-370.
2. Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting Student Performance Using Personalized Analytics. Computer, 49(4), 61-69.
3. Lowe D, Galhotra B (2023). Indian higher education: sustainable development and acceptance of digital learning platforms and moocs in pre and post covid scenarios. European Chemical Bulletin , 12 (5), 6484 –6492.
4. Aggarwal D, Sharma D, Saxena A (2023). Exploring the Role of Artificial Intelligence for Augmentation of Adaptable Sustainable Education. Asian Journal of Advanced Research and Reports 17 (11), 179-184
5. Yu, Z., Nakamura, Y., Jang, S., Kajita, S., & Mase, K. (2006). Ontology-Based Semantic Recommendation for Context-Aware E-learning. Proceedings of the 5th International Conference on Web-Based Learning, 81-90.
6. Manouselis, N., & Costopoulou, C. (2007). Analysis and Classification of Multi-Criteria Recommender Systems. World Wide Web, 10(4), 415-441.
7. Brusilovsky, P., & Millán, E. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. The Adaptive Web, 3-53.
8. Drachsler, H., Hummel, H. G. K., & Koper, R. (2008). Personal Recommender Systems for Learners in Lifelong Learning Networks: The Requirements, Techniques and Model. International Journal of Learning Technology, 3(4), 404-423.
9. Karampiperis, P., & Sampson, D. G. (2005). Adaptive Learning Resources Sequencing in Educational Hypermedia Systems. Educational Technology & Society, 8(4), 128-147.
10. Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender System Application Developments: A Survey. Decision Support Systems, 74, 12-32.

11. Romero, C., Ventura, S., & Garcia, E. (2008). Data Mining in Course Management Systems: Moodle Case Study and Tutorial. Computers & Education, 51(1), 368-384.
12. Shani, G., & Gunawardana, A. (2011). Evaluating Recommendation Systems. In Recommender Systems Handbook (pp. 257-297). Springer.
13. Tang, T., & McCalla, G. (2005). Smart Recommendation for an Evolving E-Learning System: Architecture and Experiment. International Journal on E-learning, 4(1), 105-129.
14. Chopra D, Arora P. (2023). Challenges in IoT in Higher Education, Information and Communication Technology for Competitive Strategies (ICTCS 2022) Intelligent Strategies for ICT, Springer Nature Singapore 547-558.
15. Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A Collaborative Filtering Approach to Mitigating the New User Cold Start Problem. Knowledge-Based Systems, 26, 225-238.