International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I10P106

Volume 72 Issue 10, 31-35, October 2024

Original Article

Machine Learning Recommender System for an Enhanced Students Course Selection

Theresa C. Uzoma[1], Charles O. Ikerionwu[2], Mathew E. Nwanga[1], Chukwuemeka Etus[1]

[1]*Department of Information Technology, Federal University of Technology Owerri, Nigeria.*
[2]*Department of Software Engineering, Federal University of Technology Owerri, Nigeria.*

[2]*Corresponding Author : charles.ikerionwu@futo.edu.ng*

Received: 21 August 2024 Revised: 25 September 2024 Accepted: 10 October 2024 Published: 23 October 2024

Abstract - *The challenges students face in choosing their course of choice recently have been overwhelming. This necessitated a machine learning-driven model for students' course selection. In this work, a machine learning recommender system that accepts the student's O-level as input, analyses the student's result and then recommends undergraduate courses to the student based on the student's academic performance is proposed. This research classified the course data into four datasets (Agric2, chemistry, Building/PMT, and biology). A total of 4943 instances was obtained, such that each dataset is a class containing courses with the exact requirements. Random forest classifier and decision tree classifier were used to implement each of the datasets. At the evaluation, the decision tree classifier gave an accuracy of 98%, 99.2%, 98.5% and 99.0% on the different datasets, while the random forest classifier gave 98.9%, 98.9%, 99.2% and 98.8%, respectively. This development has resolved the challenge of selecting the best-fit courses, as the model now accurately recommends courses to students based on their previous academic performance. As the model recommends courses that align with students' abilities and goals, it can help reduce the risk of course failure and improve overall academic success and retention rates.*

Keywords - *Collaborative filtering, Data normalization, Decision tree classifier, Random forest classifier, Recommender system.*

1. Introduction

Information overload has hindered the quick navigation of information in people's daily lives. [1]. Consequently, society frequently relies on approaches that can help them resolve this problem and make choices. Finding information and making judgments about it becomes more challenging as the amount of information on the Internet grows. This has led to a rise in scholars developing a new technique to help individuals handle this problem more quickly and effectively. Various scholars have presented the phenomenon of the Recommender System (RS). Recommender Systems are software applications explicitly created to propose to users the subsequent activity to partake in based on a range of variables, including preferences and the users' experience [2][3]. In other words, these programs or systems aid people in selecting the things they value. Typically, all recommender systems aim to provide users with insightful recommendations for products that may be of interest. For example, saving a customer the stress of scrolling through the whole product on an e-commerce site by recommending a particular product or brand to the customer because the customer has previously purchased a similar item or item from the same brand [4]. RS is a set of tools for information retrieval [5]. Currently, undergraduate students in their early years/levels have academic challenges and failures in tertiary institutions. This is because they choose these courses without considering their academic background and intellectual capability, which leads to difficulty in coping with the intellectual requirements for their obliviously chosen course of study [6]. Indeed, the wrong choice of course is mainly caused by peer pressure, family and society pressure. Consequently, these deluges of pressures lead the student to a course he or she does not have an interest in and the intellectual capacity to carry on.

In the existing system, students randomly, choose their undergraduate fields of study, gain admission and forge ahead to study the course. Research has shown that many students find it challenging to succeed in their chosen areas of study [1] because of the wrong choice of course. This affects low grades, exam failure, malpractice, and frustration, which further escalates to dropping out of school and obtaining a certificate with no knowledge.

To solve this problem, this research is introducing a machine learning based Rs to enhance student decision-making in course selection, such that a student's previous academic performance is considered before recommending a course. Some of the specific objectives to achieve the aim of this research are data simulation and evaluation of machine learning algorithms to choose the best fit for the model.

2. Literature Review

There are techniques in Rs, these techniques include collaborative, content, context and hybrid filtering techniques. These techniques differ in their pattern of recommendation. Due to increasing demand to meet user needs, machine learning algorithms are applied in the implementation of RS models. Over the years, several studies on recommender systems, techniques for recommender systems, and the application of machine learning to RS have been conducted. A selection of these works is reviewed and presented subsequently. [7] Developed open-source software called recommenderlab. The study used the R programming language, a prominent software environment for statistical computing and data analysis. Open-source programming language R was utilized for the platform since it made it simpler for the researcher to integrate or execute algorithms written in various programming languages, including R, Python, Java, and C/C++.

Currently, this package is used as software support for recommender system research, encompassing extensive evaluation and comparison of algorithms and fast prototyping of algorithms. However, this software allows the implementation of just a few algorithms. [8] applied collaborative and content-based filtering to create a hybrid book recommendation model. While content-based filtering was employed to produce suggestions, collaborative filtering was utilized to gather feedback from user profiles. The researchers used TF-IDF weighting with the Vector Space Model. In [9], Felipe Leite and his colleagues conducted a systematic literature review on education for teaching and learning in 2023. They examined 16 of 756 primary papers published between 2015 and 2020 that satisfied specific criteria. The study's findings indicated that the hybrid method is the most efficient means of generating recommendations. However, the conclusion that the hybrid approach was most efficient for recommendations was not clear because the number of reviewed papers was not enough to generalize the conclusion.

Subsequently, [10] modelled an undergraduate course recommendation system. This system is an enhanced undergraduate course recommender system developed using agile methodology. The authors further personalized the recommendation system by using RIASEC personality for testing. This model was not implemented with machine learning. Similarly, [11] created a learning material recommendation using the Knowledge Graph, where more recommendations are derived from the relationship between users and their preferred products. The data source for this research was not specified. The use of context format for heterogeneous graphs to model students and courses such that the model performance is not dependent on interaction results like grades was demonstrated in [12] dependency on grades, which is user input, was not applied. [13] developed a job and course recommendation model. The authors used collaborative filtering and Naïve Bayes. The authors did not explore other algorithms to ascertain the best performer. Consequently, the work in [14] carried out a literature review on the current state of recommendation systems for academic choices in higher education. It concluded that the hybrid technique is the best recommender technique. The research by [15] developed an enhanced educational recommender system with large language models and knowledge graphs. This model was named AIREG. It is a knowledge-enhanced conversational Rs using LLMS and GAN. The data used in this work is from The ESCO (European Classification of Skills Competency and Qualification). Deep learning (LSTM, MLP, GRU and Bilstm) were explored by [16] where the researchers developed an educational recommendations system based on user interest.

The major input is user data although the researcher did not use sufficient user data. From the literature review, course recommendation is an area of interest as researchers constantly work to eradicate students' challenges in selecting courses. This study proposes a course recommendation model that recommends courses to students based on their previous academic performance. The model was developed using machine learning algorithms. The remainder of the paper is organized as follows. Section 2 has reviewed some of the contributions made in this research area so far. The research methodology is presented in section 3, while section 4 presents the results and discussions of the study. Section 5 further provides the conclusion of the work and open issues.

3. Material and Methods

This paper applies machine learning algorithms to develop the recommender model that takes in students' O-level results and choice courses as input, analyzing the input data to output course recommendations based on the student's performance. The data required for this research were in two forms: Course requirements and West African Examination Council (WAEC) results. The first was collected from tertiary institutions, while the latter was generated through simulation. This data was further divided into dependent and independent data. The WAEC grading system and their score range are as follows: A1 = 75- 100, B2 = 70-74, B3=65- 69, C4 = 60-64, C5= 55-59, C6 =50-54, D7=45- 49, E8= 40-44, F9 =0-39. Courses were grouped so that courses with the exact requirements were in one group/class. This divides the courses into four Classes/datasets: Agric2, Biology, Chemistry, and Building/PMT. The number of instances for each dataset is presented in Table 1.

Table 1. Data size

S/N	Dataset	Instances	Features
1	Agric2	934	9340
2	Biology	1000	10,000
3	Chemistry	1000	10,000
4	Building/PMT	2000	20,000
Total		4,934	49,340

The chemistry dataset is a class of courses with the same academic requirements and includes courses with the same admission requirements. For example, the Geology, Agricultural Engineering, Mathematics, and Statistics departments are some of the courses in the chemistry dataset, and they all require at least a credit pass in the same WAEC results. The courses in the chemistry dataset are 23. The instances refer to the total number of results, which is 1000, while the features include the total number of subjects taken by each student, which is a maximum of 10. This implies mathematically in Equation 1 that.

$$Features = Instances * 10 \quad (1)$$

3.1. Research Design

In the design and execution of the machine learning based recommender system, three phases were adopted. These phases are Data preprocessing, Model Selection and Model Evaluation. During the data preprocessing phase, the Python docx library was used to read, manipulate and modify the data. The data was further formatted into a csv form, parse_doc function was used to identify data details, after which it was converted to a data frame. Dataframe cleaning instruction was applied to instruct direct cleaning functions to row or column. The data was equally normalized to handle missing values. Normalization is performed on the X (independent data) data as follows. Grade A score ranges from 75-100, meaning grade A can be any number between 75 and 100. A “Random” function generates numbers within this range, and then a preprocessing technique called MinMax Scaler is applied to reduce the number range such that the range falls between 0-1.

This is because ML models work better with a lesser number scale. MinMax scaler is calculated mathematically with the formula in Equation 2.

$$MinMax\ Scaler = \frac{X - Min}{Max - Min} \quad (2)$$

Where;
X is that grade in focus.
Min is the minimum value for the grade in focus.
Max is the maximum value for the grade in focus.

The next phase is model selection. At this phase, the data is split into two sets: the training set and the validation set. This splitting was done in the ratio 80:20. The training set is used to train the model, while the validation set is used to test the model’s performance when it encounters new data. For this model development, two classifier algorithms (Decision tree classifier and Random Forest Classifier) were implemented on the four datasets listed in Table 1, and the best-performing algorithm for each model was selected. Pipelining was implemented to integrate these models.

3.2. Proposed System

The recommender system is shown in Figure 1. The model was designed so that users can select a choice course and input their O-level results. The model starts by analyzing the students’ results and then further classifies the students’ results into one of the course datasets. If the student’s results fit into his/her choice course, then it’s recommended; if not, the courses in the course dataset (students’ result fit) are recommended.

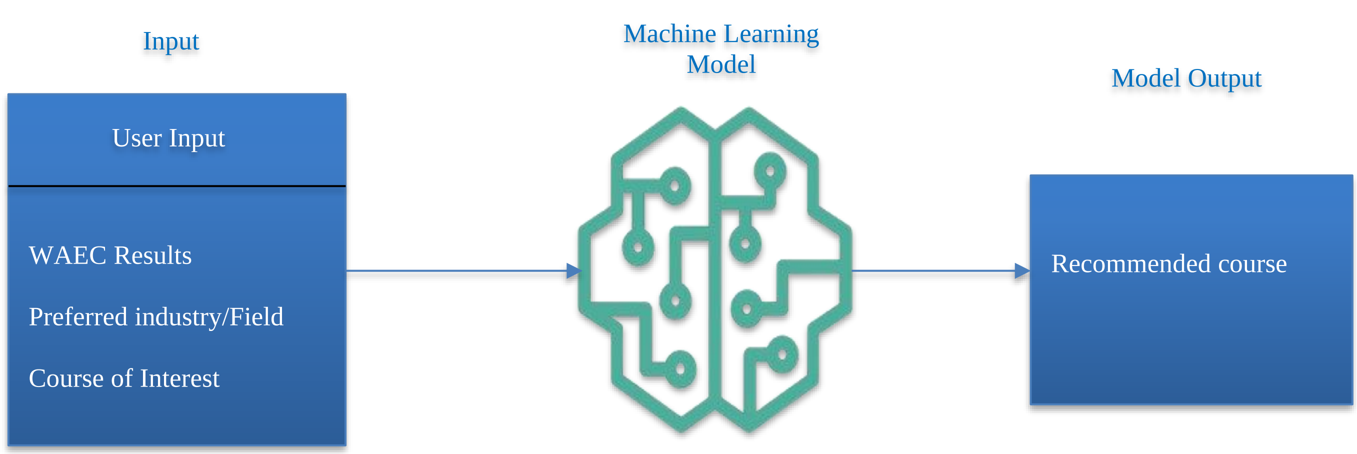

Fig. 1 Overview of the proposed recommender system

As shown in Figure 1, the user inputs, which are the course of interest (choice course) and WAEC results, are inputted into the machine learning model and are processed as the user input. The outputs are the recommended courses based on the student’s performance. For instance, a student requests a choice course in the Optometry department and scores a credit grade in English, Mathematics and Biology. The model checks if the student’s result meets the admission requirement for the optometry department. If the result is satisfactory, Optometry is approved and recommended. However, in a situation, where the student failed Biology and passed Mathematics and English, the recommender model considers other courses that require Mathematics and English with little or no knowledge of Biology.

4. Results and Discussion

Fig. 2 Confusion matrix for decision tree on Building/PMT dataset

The result of this model is stated and calculated using the confusion matrix in Figure 2 and demonstrated in Table 2. Accuracy is calculated using Equation 3, concerning the values obtained from the confusion matrix and contained in Table 3.

$$Accuracy = \frac{(TP+TN)}{TP+FP+FN+TN} \quad (3)$$

Table 3 illustrates each model's performance. The algorithm with the highest accuracy is chosen for the model. For example, in the Chemistry dataset, the decision tree classifier gave 99.2% accuracy while the random forest classifier gave 98.9% accuracy, which is why the decision tree classifier was chosen.

Table 2. Confusion matrix formula

TP	**FP**
FN	TN

Table 3. Result summary

Dataset	Decision Tree Classifier accuracy	Random Forest Classifier accuracy	Algorithm Implemented
Agric2	98.0%	98.9%	Random Forest Classifier
Chemistry	99.2%	98.9%	Decision Tree classifier
Building/PMT	98.5%	99.2%	Random Forest Classifier
Biology	99.0%	98.8%	Decision Tree classifier

5. Conclusion

This study aims to address the issue of students choosing incorrect courses in tertiary institutions. This became imperative because of the increasing number of students who either fail or struggle to enrol in the courses they eventually do. They made their choices without considering their academic standing or the intellectual demands of the courses they selected. To solve this issue, the study created a recommendation algorithm that categorizes and suggests courses to students based on their interests and past academic achievement(s). With this development, students can easily navigate through their undergraduate courses because these courses are chosen with respect and consideration of their intellectual capability. Thus reducing students' failure rate and presumed difficulty in the early years of their studies.

Future Work

This study, having achieved its aim by introducing two algorithms that personalize course recommendation by considering students' previous academic performance, points to several future research. Researchers should explore the possibility of incorporating new courses and changes that will be made to the academic curriculum.

Data of existing courses should equally be updated as curriculum and admission requirements are reviewed.

Conflicts of Interest

This paper is an extract from an M.Sc. thesis and, as such has no conflict of interest or published with any other journal or authority.

References

[1] Nacim Yanes et al., "A Machine Learning-Based Recommender System for Improving Students Learning Experiences," *IEEE Access*, vol. 8, pp. 201218-201235, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Gricela Andrade-Ruiz et al., "Emerging Perspectives on the Application of Recommender Systems in Smart Cities," *Electronics*, vol. 13, no. 7, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Dhruval Patel, Foram Patel, and Uttam Chauhan, "Recommendation Systems: Types, Applications, and Challenges," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 851-868, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] N.D. Lynn, and A.W.R. Emanuel, "A Review on Recommender Systems for Course Selection in Higher Education," *IOP Conference Series: Materials Science and Engineering*, vol. 1098, no. 3, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kareem S. Elattar, and Khaled M. Fouad, "A Survey on Recommender Systems Challenges and Solutions," *2nd International Mobile, Intelligent, and Ubiquitous Computing Conference*, Cairo, Egypt, pp. 296-301, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ádám Kocsis, and Gyöngyvér Molnár, "Factors Influencing Academic Performance and Dropout Rates in Higher Education," *Oxford Review of Education*, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Michael Hahsler, "Recommenderlab: An R Framework for Developing and Testing Recommendation Algorithms," *arXiv*, pp. 1-41, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Sunny Sharma, Vijay Rana, and Manisha Malhotra, "Automatic Recommendation System Based on Hybrid Filtering Algorithm," *Education and Information Technologies*, vol. 27, no. 2, pp. 1523-1538, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Felipe Leite da Silva et al., "A Systematic Literature Review on Educational Recommender Systems for Teaching and Learning: Research Trends, Limitations and Opportunities," *Education and Information Technologies*, vol. 28, no. 3, pp. 3289-3328, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Abdulrazak Yahya Saleh, and Mark Cornalius Julian, "Development of Undergraduate Students Course Recommender System," *3rd International Conference on Emerging Smart Technologies and Applications*, Taiz, Yemen, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Christos Troussas, and Akrivi Krouska, "Path-Based Recommender System for Learning Activities Using Knowledge Graphs," *Information*, vol. 14, no. 1, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Zhengyang Wu, Qingyu Liang, and Zehui Zhan, "Course Recommendation Based on Enhancement of Meta-Path Embedding in Heterogeneous Graph," *Applied Sciences*, vol. 13, no. 4, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] B. Derick Prince et al., "Job and Course Recommendation System using Collaborative Filtering and Naive Bayes Algorithms," *2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation*, Coimbatore, India, pp. 1-4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nabila Kamal et al., "Recommender System in Academic Choices of Higher Education: A Systematic Review," *IEEE Access*, vol. 12, pp. 35475-35501, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Fatemeh Fathi, "AIREG: Enhanced Educational Recommender System with Large Language Models and Knowledge Graphs," *European Semantic Web Conference ESWC*, pp. 1-11, 2024. [Google Scholar] [Publisher Link]

[16] Hadis Ahmadian Yazdi, Seyyed Javad Seyyed Mahdavi, and Hooman Ahmadian Yazdi, "Dynamic Educational Recommender System Based on Improved LSTM Neural Network," *Scientific Reports*, vol. 14, no. 1, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]