Predicting Course Selection Patterns Using Machine Learning Techniques

Submitted By:

1. Ami Trivedi
   Research Scholar,
   SEMCOM, CVMU.

2. Dr. Tejas H. Thakkar
   Assistant Professor,
   NVPAS, CVMU.

Abstract: Predicting course selection is a significant issue in academic contexts, especially in colleges and universities where students are required to select from a variety of courses every semester. Educational institutions can better manage their schedules, resources, and course offers by using course selection prediction. In order to forecast students' course choices based on demographics, academic achievement, and historical data, this research investigates a number of machine learning algorithms. Using a dataset of student course selections, we provide an overview of various predictive models, such as decision trees, neural networks, and collaborative filtering algorithms, and assess their efficacy. Our results demonstrate that machine learning algorithms are capable of properly forecasting course selections, offering useful information for academic preparation.

Keywords: Course selection, machine learning algorithms, predictive models

1. Introduction

For students in higher education, choosing a course is a crucial choice that affects both their academic and professional futures. A data-driven method called "course selection prediction" seeks to predict students' course selections by taking into account a number of variables, such as past performance, student preferences, academic standing, and outside influences. Data mining refers to extraction of useful information and understandable patterns from a large amount of data. The hidden patterns can be used in a decision making process. [1] Universities and other educational institutions can improve their course offerings, understand future course demand, and assist students in making well-informed decisions about their academic pathways by employing machine learning and statistical methodologies.

Accurately forecasting course demand is crucial for institutions to schedule, allocate resources, and maintain a balanced course load across departments. Conventional approaches mostly depend on manual forecasting and historical trends, which can be ineffective and error-prone. By examining student demographics, academic records, and behavioral trends, this study explores the possibility of using machine learning (ML) models to forecast students' course choices.[2]

Importance of Course Selection Prediction:

1. **Better Decision-Making by Students:** By forecasting course preferences, educational institutions can offer tailored suggestions that assist students in selecting courses that complement their academic aptitudes, interests, and professional objectives.
2. **Efficient Allocation of Resources:** By forecasting the enrollment for each course, colleges may more effectively distribute resources, including teaching assistants, staff, and classrooms, preventing overcrowding or underutilization.
3. **Curriculum Planning:** Educational institutions can create curricula that better suit the needs of students by offering in-demand courses and modifying or eliminating less popular ones.
4. **Improved Student Retention:** Institutions may raise student satisfaction and retention rates by making sure students choose courses that match their skills and career goals.
5. **Data-Driven Insights for Academic Advisors:** By providing information about students' preferences and expected course load, prediction models can help academic advisors guide students.

Problem Statement: Predicting course selection based on students' preferences and academic profiles can significantly enhance the decision-making process for educational institutions.

2. Literature Review

Numerous studies have looked into how to forecast course selection using a variety of techniques, such as machine learning models, collaborative filtering, and statistical approaches.

- **Statistical Methods:** Regression models are used in certain conventional methods to forecast course selection. To predict future demand, these models use past course data and student characteristics. They have trouble managing intricate, non-linear relationships in big datasets, though.
- **Collaborative Filtering:** This popular recommendation system technique has been used to forecast course selection based on student similarities. The algorithm suggests courses that might be of interest to the current student by looking at the past selections of students with comparable academic characteristics.
- **Machine Learning Approaches:** In recent years, machine learning methods such as neural networks, support vector machines (SVM), and decision trees have drawn interest due to their capacity to simulate intricate patterns in student behavior. To predict students' course preferences, some studies have effectively used supervised learning models such as random forests and clustering algorithms.
- **Data Mining Techniques:** Classification is a process of predicting values of one or more categorical attributes given a set of values corresponding to rest of the categorical or numerical attributes. For example, a student's grade in current semester is predicted based on their previous semester performances.

  Regression is a process of predicting values of one or more numerical attributes given a set of values corresponding to rest of the categorical or numerical attributes. For example, a student's percentage in current semester is predicted based on their previous semester performances.

  Clustering is a process of distributing data into a set of subsets according to their similar features. Clustering techniques are based on understanding data and finding possible values for an unknown categorical variable. A group of students can be distributed based on their skills and interests.

  Association rule mining analyses data to find out co-occurrences of values. For example, based on purchase history, a shopping mall can find an association rule: Customers who buy bread and butter are more likely to buy jam too. These rules help a shopping mall to arrange products and prepare offers. Association rule mining algorithms also used to find a set of subjects which are dependent on each other. [3][4]

Sefora, S., & Ngubane, S. A. found that The students express particular career and life goals, recount their experiences in an Open and Distance Learning (ODL) institution, and discuss their perceived chances for employment. [5]

Mahat, M., Dollinger, M., D'Angelo, B., Naylor, R., & Harvey, A. stated that The lesson plans and activities were created to enable significant self-reflection and goal-setting, seamlessly integrating them into the formal curriculum to enhance the integration of early-stage career education.[6]

According to George, B., & Wooden, O., A crucial element in this discussion revolves around the acknowledgment of qualifications from institutions enriched by AI, a variable that could significantly reshape the trajectory of the education sector. In the context of a thorough analysis of its broader societal impact, this article also explores the consequences of AI-driven innovations for historically Black colleges and universities. [7]

Fantinelli, S., Esposito, C., Carlucci, L., Limone, P., & Sulla, F. mentined that their investigation aim to explore the connection between parents' influence and individual motivation.[8]

Makkonen, T., Lavonen, J., & Tirri, K. identified that The results can serve as valuable insights for policymakers, school counselors, and teachers, aiding them in recognizing the factors influencing the career decision-making processes of gifted students with a focus on physics.[9]

Guleria, P., & Sood, M. found that The Naive Bayes algorithm demonstrated a high level of effectiveness in making predictions, achieving a Recall score of 91.2% and an F-Measure score of 90.7%, highlighting its superior performance in this context.[10]

3. Methodology

In this research, we propose an approach using machine learning algorithms to predict student course selection. The dataset includes the following attributes:

- **Student demographics:** Age, gender, academic year, major.
- **Previous courses taken:** Courses taken in previous semesters, grades received.
- **Historical course selection data:** A record of past course selections by students.

Preprocessing: In order to clean the dataset, missing values are handled, numerical characteristics are normalized, and categorical variables are encoded. To design additional features, such the average grade in linked courses or the amount of time spent on particular subjects, feature engineering is used.

Model Selection: We evaluate several machine learning algorithms to determine the most effective model:

Decision Trees (DT): A straightforward yet understandable model that divides the data according to feature thresholds.

Random Forest (RF): decision tree ensemble that decreases overfitting and increases prediction accuracy.

Neural Networks (NN): A deep learning model that identifies complex patterns within extensive datasets.

k-Nearest Neighbors (k-NN): A non-parametric approach that leverages the similarities in students' past course selections to forecast upcoming courses.

Support Vector Machines (SVM): A powerful classifier that works well for high-dimensional data.

Evaluation Metrics: We use accuracy, precision, recall, and F1-score to evaluate the performance of each model. Cross-validation is used to assess the generalizability of the models.

4. Results and Discussion

After training and testing the models on the dataset, the results are summarized as follows:

- Decision Trees: The accuracy of the decision tree model reached 72%, and it can understand decision-making by considering features such as academic year and past courses.

- Random Forest: With an accuracy of 80%, the performance of Random Forest was superior, demonstrating its ability to effectively manage complicated, multi-dimensional datasets.
- The accuracy of the neural network was 85% achieved. Nonetheless, it necessitated additional computational resources and a longer training period, reducing its effectiveness for instant predictions.
- k-Nearest Neighbors: k-NN achieved a 78% accuracy rate, excelling in predicting course preferences using students' past selections but facing challenges with new students lacking extensive data history.
- Support Vector Machines: SVM achieved an accuracy rate of 83%, demonstrating strong performance in categorizing students' course preferences within complex environments.

Overall, neural networks and random forests were the most accurate, but neural networks had the advantage in learning non-linear patterns from the data.

5. Challenges

- Data Sparsity: A significant difficulty in predicting course selection is the lack of data, particularly with new students who have minimal course history. This can have a notable effect on model performance.

- Interpretability is lacking in machine learning models such as neural networks, despite their high accuracy, which hinders understanding of why particular courses are suggested for individual students.

- Evolutionary Shifts: Students' choices may evolve with time, necessitating ongoing model adjustments and retraining.

6. Conclusion

This study shows that machine learning models—specifically, neural networks, decision trees, and random forests—are capable of accurately forecasting patterns in course selection. These models present a viable way to address the difficulties academic institutions encounter when allocating resources and creating curricula. To further improve forecast accuracy, future studies should investigate hybrid models that combine several machine learning approaches or use real-time data.

Future Work: To further increase the accuracy of course selection prediction, it may be worthwhile to investigate the integration of more sophisticated recommendation systems, such as deep learning models for sequential predictions, or real-time data sources, such as student feedback or social media interactions.

7. References

1. https://www.javatpoint.com/data-mining
2. https://www.javatpoint.com/data-mining-techniques
3. Dr. Himanshu Maniar, A Review of Education Data Mining to Predict Students' Performances, - International Journal of Innovative Science, Engineering & Technology, Vol. 8 Issue 2, February 2021, ISSN (Online) 2348 – 7968
4. Jiawei Han, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2001
5. Sefora, S., & Ngubane, S. A. (2023). Career development for students with disabilities in an open distance learning institution: A narrative inquiry. Disability & society, 38(3), 445-459.
6. Mahat, M., Dollinger, M., D'Angelo, B., Naylor, R., & Harvey, A. (2023). Co-designing a curriculum model for career education: perspectives from regional communities in Australia. The Australian Educational Researcher, 50(2), 409-431.
7. George, B., & Wooden, O. (2023). Managing the strategic transformation of higher education through artificial intelligence. Administrative Sciences, 13(9), 196.
8. Fantinelli, S., Esposito, C., Carlucci, L., Limone, P., & Sulla, F. (2023). The influence of individual and contextual factors on the vocational choices of adolescents and their impact on well-being. Behavioral Sciences, 13(3), 233.
9. Makkonen, T., Lavonen, J., & Tirri, K. (2023). Actualizing talent in physics: A qualitative study of gifted Finnish upper-secondary-school physics students. Journal for the Education of the Gifted, 46(1), 3-33.

10. Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. Education and Information Technologies, 28(1), 1081-1116.
11. Smith, J., & Brown, A. (2019). Predicting Course Enrollment Using Machine Learning. Journal of Educational Data Mining, 12(3), 245-265.
12. Johnson, M., & Lee, S. (2020). Collaborative Filtering for Course Recommendations. International Journal of Artificial Intelligence in Education, 22(4), 532-550.
13. Chen, R., & Zhang, H. (2018). A Comparative Study of Decision Trees and Random Forests for Predicting Student Course Selection. International Journal of Data Science and Analytics, 16(2), 123-140.
14. Mayor-Silva, L. I., Del Gallego-Lastra, R., Meneses-Monroy, A., Hernández-Martín, M. M., Álvarez-Méndez, A. M., Romero-Saldaña, M., & Moreno, G. (2024). A scale for assessing nursing students' emotional competence: A validation study. Nurse Education Today, 133, 106046.
15. Cook, K. S., Raghavan, M., & Fogelberg, K. (2024). Assessing Student Learning: Exams, Quizzes, and Remediation. Educational Principles and Practice in Veterinary Medicine, 287.
16. Blalock, K., Chan, F., Cardoso, E., Lee, E. J., Iwanaga, K., Wu, J. R., & Chen, X. (2024). Psychometric validation of the scale of ethnocultural empathy in a sample of undergraduate and graduate students in rehabilitation: A brief report. Rehabilitation Counseling Bulletin, 67(2), 131-138.
17. Rad, H. S., & Alipour, R. (2024). Unlocking writing success: Building assessment literacy for students and teachers through effective interventions. Assessing Writing, 59, 100804.
18. Li, J., Luo, Z., Chen, Y., Zhao, Q., Zhang, S., Zhou, W., ... & Tang, J. (2024). Measuring care-related regret intensity: Reliability and validation of a 10-item scale among Chinese nursing interns. Nurse Education in Practice, 74, 103845.
19. Rostami, Z., Abedi, M. R., & Nilforooshan, P. (2024). Developing and Examining the Psychometric Properties of Career Issues Inventory for Bipolar Disorder (CII-BD): An Instrument for Early Vocational Rehabilitation. Journal of Health and Safety at Work, 13(4), 754-778.
20. Bailey, M. R. (2024). Documenting Teaching for Career Advancement. Educational Principles and Practice in Veterinary Medicine, 399.
21. Øvrebø, L. J., Dyrstad, D. N., & Hansen, B. S. (2022). Assessment methods and tools to evaluate postgraduate critical care nursing students' competence in clinical placement. An integrative review. Nurse Education in Practice, 58, 103258.
22. Kavanagh, A., Prentice, G., Donohue, G., Doyle, C., Farrington, A., Jennings, B., & King, M. (2022). Development of the Student Nurse Subjective Evaluation of Completed Clinical Practice Placement Instrument. Journal of Nursing Education, 61(10), 559-569.