

PREDICTING DROPOUT RATES IN ONLINE LEARNING

Dr. S Sreeja ¹ and A Mohamed Ali ²

[1] Associate Professor, Department of Management Sciences, Hindusthan college of engineering and technology, Coimbatore, Tamil Nadu.

[2] Student, Department of Management Sciences, Hindusthan college of engineering and technology, Coimbatore, Tamil Nadu.

Abstract

Online learning is a part of education these days but a lot of students still drop out. This study is about using machine learning to figure out which students are likely to drop out. The study uses information from Kaggle including a set of data about students who took online courses. The researchers looked at things like how much students participated how they did in their classes and what they did while they were taking the courses to see how these things affect whether students drop out.

The study used a kind of computer program called a Random Forest classification model to predict whether a student would drop out or keep taking the course. Before they made the program, they cleaned up the data made features and looked at the data to make sure it was good and the program would work well. They then checked to see how well the program worked by using tests like accuracy, precision, recall and F1-score. What they found out was that students who did not participate much missed assignments. We're not active were the ones who usually dropped out. The program was good at finding students who were at risk of dropping out which means that teachers can help them early. The study says that machine learning can help keep students in school by allowing teachers to give them help when they need it and make better decisions, about online learning.

Keywords: Student Dropout, Online Learning, Machine Learning, Random Forest, Predictive Analytics, Student Engagement, Learning Analytics, Educational Data Mining, Classification Model, Data Analysis

Introduction

Online learning is really popular these days because students can study at any time and from any place. A lot of people use platforms to learn new things and get better at what they do. But there is a problem with online education. A lot of students do not finish their courses and quit before they are done. Students quit their courses for reasons, such as they are not interested they do not feel like learning they do not manage their time well and they have trouble understanding the course. When students learn online they do not get to talk to their teachers much so they might feel like they are not connected and they lose interest after a while. As a result they do not participate much and eventually they stop learning altogether.

These days online platforms collect a lot of information about students, such as when they log in how time they spend on their courses and how well they do on their assignments. We can use this information to understand how students behave and find patterns. By using computer techniques, we can even predict if a student is likely to quit or keep taking the course. In this study we use a computer model called Random Forest to look at student information and predict if they will quit. The main goal is to find students who are having trouble and help them have a better learning experience. This way online platforms can do something to help students on and reduce the number of students who quit.

Overall this study is about using information and technology to fix the problem of students quitting and make online learning better. Online learning can be very helpful, for students. We want to make sure they can succeed. By using data and technology we can make online learning more effective. Help students achieve their goals.

2. Statement of the Problem

Online learning platforms are really popular. They have a big problem. Many students drop out. A lot of students sign up for courses. Do not finish them. This makes online education less effective.

The main issue is that it is hard to tell which students are likely to drop out on. Even though platforms collect lots of data on students like login activity time spent on courses and how well they do on assignments this data is not used well to predict who will drop out. Many things can lead to students dropping out such as engagement not doing well in their courses and not being active. These factors are complicated. Cannot be easily understood using old methods.

So we need a solution that uses data and machine learning to look at how students behave and predict who will drop out. This will help find students who're at risk early and allow to help them in time which will improve student retention, in online courses. Online learning platforms can work better if they use machine learning to predict student dropout and take action to prevent it. Student dropout is a problem that affects online learning platforms and machine learning can help solve it.

Objectives of the Study

1. To analyze student engagement, performance, and behavioural patterns in online learning.
2. To identify the key factors that influence student dropout.
3. To develop a machine learning model (Random Forest) to predict student dropout.
4. To provide suggestions for reducing dropout and improving student retention in online learning platforms.

3. Scope of the Study

This study is about predicting when students drop out of courses. It uses data to understand what makes students complete or quit their courses. The study uses a dataset from Kaggle. This dataset has student records from different online courses. It looks at things like how students log in how much time they spend on the platform how far they get in their courses how many assignments they submit and how often they are inactive. These things help find patterns that show which students might drop out.

The study does a thing. It makes a machine learning model to predict if a student is active or at risk of dropping out. The model used is called Random Forest. It also. Analyses the data, Then it checks how well the model works.

However the study has some limits. It only uses existing data. Does not ask students questions. It also only looks at courses. So it might not be the same, for all courses. The goal of the study is to help keep students in courses. It wants to give information and a way to predict which students might drop out. This can help make education better.

4. Research Methodology

This study uses a numbers-based approach to analyze and predict when students drop out of courses. The research aims to describe and predict student behaviour and to build a computer model that can forecast dropout. The study uses existing data from Kaggle. The data includes student records with information on how engaged they are how well they perform and their behaviour patterns. Relevant records are chosen from this data for analysis.

A sampling method called sampling is used to choose data that is relevant to the study. The focus is on students in online courses where predicting dropout is important.

The research process involves steps:

- * Data collection from Kaggle
- * Data cleaning and fixing missing values
- * Exploring the data
- * Preparing the data for modelling
- * Building a model using Random Forest
- * Evaluating the model using metrics, like accuracy and precision

The Random Forest algorithm is used to predict student dropout. The data is split into two parts to test the models performance. Accuracy, precision, recall and F1-score are used to evaluate the model. Overall this approach provides a way to analyze student behaviour and build a model that can predict dropout in online courses.

Findings

The study provides several important findings related to student dropout in online learning environments based on data analysis and the Random Forest model.

1. **Student engagement is a key factor:** Students who log in frequently and spend more time on the platform are more likely to complete the course. Low engagement leads to higher dropout.
2. **Academic performance affects retention:** Students who submit assignments regularly and score well in quizzes tend to remain active, while poor performance increases dropout risk.
3. **Inactivity is a strong indicator:** Students who do not log in for a longer period are more likely to drop out. “Days since last login” is one of the most important variables.
4. **Assignments play an important role:** Missing assignments is strongly linked to dropout behaviour.
5. **Model performs well:** The Random Forest model shows good accuracy in predicting whether a student will drop out or continue.
6. **Key influencing factors identified:** Features like progress percentage, inactivity, and assignment completion have the highest impact on prediction.

7. **Class imbalance observed:** Dropout cases are fewer compared to active students, but the model still identifies patterns effectively.

Overall, the findings show that student dropout can be predicted using data analysis, and early identification of at-risk students can help improve retention in online learning platforms.

Suggestions (Recommendations)

Based on the findings of the study, the following recommendations are suggested to reduce student dropout and improve retention in online learning platforms:

1. **Improve student engagement:** Platforms should encourage regular participation through interactive content, quizzes, and discussions to keep students actively involved.
2. **Early identification of at-risk students:** Use machine learning models to detect students with low activity or performance and provide timely support.
3. **Provide personalized support:** Offer customized learning plans, reminders, and guidance based on individual student needs and behaviour.
4. **Regular monitoring of student activity:** Track login frequency, time spent, and assignment completion to identify signs of disengagement.
5. **Enhance course design:** Courses should be simple, engaging, and well-structured with clear instructions and short modules to maintain student interest.
6. **Encourage assignment completion:** Provide deadlines, reminders, and feedback to ensure students complete their tasks on time.
7. **Introduce motivation strategies:** Use rewards, certificates, and progress tracking to motivate students to complete courses.
8. **Implement early warning systems:** Integrate predictive models into platforms to automatically alert instructors about at-risk students.

Overall, these recommendations can help online learning platforms improve student retention and provide a better learning experience.

Conclusion

This study focused on predicting student dropout in online learning environments using a data-driven approach and the Random Forest machine learning model. The results show that student dropout is mainly influenced by factors such as low engagement, poor academic performance, and inactivity. Students who frequently log in, spend more time on the platform, and complete assignments are more likely to finish their courses. The developed model performs effectively in identifying students who are at risk of dropping out. This proves that machine learning can be a useful tool in analysing student behaviour and making accurate predictions. By using such models, online learning platforms can take early actions to support students and reduce dropout rates.

The study also highlights the importance of monitoring student activity and providing timely intervention. Features like progress percentage, days since last login, and assignment completion play a key role in predicting outcomes.

In conclusion, this research demonstrates that combining educational data with machine learning techniques can improve student retention and enhance the overall effectiveness of online learning systems.

References

1. Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. *Learning Analytics*, Springer.
2. Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics*, 40(6), 601–618.
3. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
4. Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC Dropout over Weeks Using Machine Learning Methods. *Proceedings of the EMNLP Workshop*.
5. Huang, S., Fang, N., & Qian, Y. (2020). Predicting Student Dropout in Online Courses Using Machine Learning Techniques. *Journal of Educational Computing Research*.
6. Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30–40.
7. Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. *Computational Intelligence and Neuroscience*.
8. Al-Sha bandar, R., Hussain, A. J., Keight, R., et al. (2019). Machine Learning Approaches to Predict Learning Outcomes in MOOCs. *International Journal of Information Management*.
9. Delen, D. (2010). A Comparative Analysis of Machine Learning Techniques for Student Retention Management. *Decision Support Systems*, 49(4), 498–506.
10. Zhang, L., & Rangwala, H. (2018). Early Identification of At-Risk Students Using Machine Learning Techniques. *IEEE Transactions on Learning Technologies*.