Predictive Analysis Of Student Performance in Online Learning

Kolla Puneeth Sai
Computer Science and Engineering
Dhanalakshmi Srinivasan University
Samayapuram,Trichy
puneeths602@gmail.com

Kanna Jyothi Siva Prasad
Computer Science and Engineering
Dhanalakshmi Srinivasan University
Samayapuram,Trichy
jyothisivaprasad03@gmail.com

Kanyadhri premchand
Computer Science and Engineering
Dhanalakshmi Srinivasan University
Samayapuram,Trichy
kanyadharipremchand@gmail.com

Mrs.Subashri Rajendran
Assistant Professor,SET
Dhanalakshmi Srinivasan University
Samayapuram,Trichy
subashrirajendran.cse@dsuniversity.ac.in

Abstract

In recent years, a number of learning platforms have emerged, and this has resulted in a tremendous growth in data related to learning processes. Analyzing this data helps in improving student learning outcomes. Making early predictions about student performance is a key area of research in data mining in education. Data mining and machine learning techniques are useful in finding patterns in student data related to academic performance and making early predictions of success or failure in academics, as suggested in a study by Shahiri et al. [9]. Keeping this in view, this study presents a predictive analysis model for evaluating student performance in an online learning scenario. Various factors are considered in this model, such as student engagement, attendance, and performance in tests, along with their interactions with learning platforms. Machine learning algorithms are employed in this system to classify students into different performance categories and identify those students who are not performing well in academics[8]. Such a system will help in increasing learning outcomes, reducing dropout rates, and making decisions in the field of education to improve the quality of learning in an online scenario.

Keywords — Educational Data Mining, Student Performance Prediction, Online Learning, Machine Learning, Predictive Analytics, Supervised Learning, Learning Analytics, Academic Performance, Data Analysis, Student Engagement.

I Introduction

In the past few years, online learning has been an integral part of the education system. Due to the development of technology, the use of the internet has become widespread. Educational organizations are now using online learning tools to provide learning materials to students[2]. In online learning, students are free to study at their own pace and from different locations. However, there are various challenges associated with online learning, especially in monitoring students' engagement and forecasting students' performance. One of the major challenges that arise in an online learning environment is the challenge of identifying students who are at risk of poor academic performance[3]. Unlike a traditional classroom setting, the instructor of an online classroom often lacks the ability to interact with the students. As a result, it becomes difficult to track the level of participation, motivation, and progress of the students[4]. Some students often find it difficult to comprehend the content of the course or to stay motivated to learn from the online platforms.

With the rapid development of online education, a great deal of information about students is produced through learning management systems or virtual learning environments. These pieces of information include the frequency of login, the amount of time spent on learning materials, assignments completed by the students, quiz results, and participation in online discussions. These pieces of information can give useful insights into the learning behavior of the students[5]. However, the information about the students is not effectively analyzed in most educational institutions..

Educational Data Mining (EDM) and Learning Analytics have become prominent areas of research, and the focus of these approaches is on the analysis of educational data to improve the process of learning and teaching. These approaches employ data analysis and machine learning to identify patterns among the learning and academic behaviors of the students. According to Shahiri et al. [9], predictive models developed through data mining approaches have the potential to reveal factors that influence the academic performance of the students and to predict the academic performance of the students. Machine learning approaches have the potential to analyze the data and reveal patterns to determine the academic performance of the students.

This is because predicting students' performance can offer several benefits to educators and institutions. For instance, predicting students' performance can help educators to identify students who may need extra guidance and assistance[7]. This can improve students' performance because academic intervention can help students overcome their learning challenges. In addition, predicting students' performance can help institutions reduce dropout rates and improve the quality of education.

Machine learning is one of the significant techniques used in the predictive analysis of the performance of students[8]. Supervised machine learning algorithms like Decision Tree, Logistic Regression, Random Forest, and Neural Networks can be used for the analysis of the data of students. These algorithms can predict the performance of students in the future based on the patterns that have been identified. By incorporating the data of the academic performance of students with their behavior in online educational platforms, the accuracy of the predictive model can be increased. In addition to that, recent studies have highlighted the need to take several factors into account when making predictions about students' performance[10]. It has been noted that students' performance is affected not only by their scores but also by their behavior while learning, their participation in activities, as well as their interaction with what they learn. This ensures that predictions made by the system are more encompassing and accurate.

The main aim of this research is to show how predictive analysis and machine learning techniques can be applied to educational data to improve student performance in an online learning system[12]. By using data-driven techniques, educational institutions can improve their capacity to monitor student performance and improve their overall academic performance..

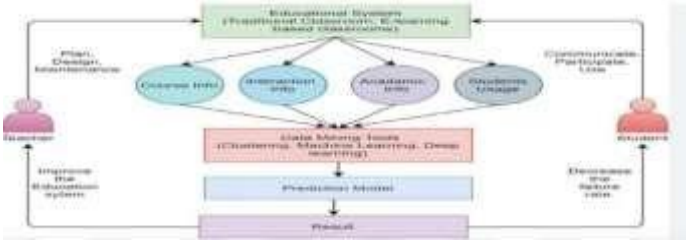

Fig. 1: Architecture of Educational Data Mining System for Student Performance Prediction

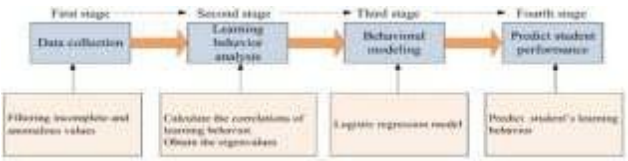

Fig. 2: Workflow of Student Performance Prediction Using Learning Behavior Analysis

II RELATED WORK

Several studies have been carried out to examine data mining and machine learning techniques in predicting student academic performance in educational institutions. One of the key studies in this area was carried out by M. Shahiri, W. Husain, and N. Rashid [9]. In this study, a comprehensive review of data mining techniques in predicting student performance was done. The study highlighted the significance of data analysis in

understanding student learning patterns in order to improve their performance.

According to Shahiri et al. [9], educational institutions are data generators, and this data is generated through learning management systems, online tests, and student activity logs. For example, data generated in this area includes student attendance, assignment performance, test performance, and their interactions with learning materials. The study highlighted the significance of data analysis in this area using data mining techniques in order to identify patterns that affect student performance and success[15].

Some of the machine learning algorithms discussed were the ones most frequently employed for predicting student performance, including the Decision Trees, Neural Networks, Naïve Bayes, Support Vector Machines, and k-Nearest Neighbor algorithms[16]. Among the algorithms discussed, the Decision Trees and Neural Networks were the ones that could provide reliable prediction results due to the ability of the algorithms to deal with complex relationships between variables.

In addition, the study revealed the ability of the predictive models to help the educator identify the students who are at risk of poor academic performance at an early stage[17]. This would enable the educator to provide the students with the necessary academic support, hence improving the overall student success rate and lowering the dropout rates within the educational institutions.

The effectiveness of data mining techniques in analyzing student academic data and creating predictive models for decision-making in education is clearly demonstrated in the research work done by Shahiri et al. [9]. It is clear from their work that it forms a good foundation for creating machine learning systems to help instructors in monitoring student performance in an online learning environment.

III Problem Statement and Objectives

It is difficult for the instructors to keep track of the students' engagement and academic performance in the online learning environment. A huge quantity of data is being created by the students in the online environment[18]. This data is mostly not being used for the prediction of the students' performance. Consequently, students who are underperforming academically are usually identified only after the examination. There is a need to design a predictive system to analyze the students' data to identify the underperforming students at an early stage.

Objectives

The main goal of this research is to build a predictive analysis model that can be used to assess the performance of students in online learning. This research aims to study the data related to the activities of students that are collected from online learning platforms[11]. This research will also focus on finding the important factors that can affect the performance of students. Using the concept of machine learning, this system will be able to classify the students according to their performance levels. This system will also be able to identify the students who are at a higher risk of poor performance.

IV. Proposed Methodology

The proposed methodology is focused on predicting the performance of students in online learning environments using machine learning techniques. In the proposed system, the data collected from online learning environments is analyzed based on the activities performed by students in online learning environments. In the proposed system, initially the data collected is subjected to preprocessing techniques to remove inappropriate values from the collected dataset and to normalize the data for better analysis[8]. Then, features are selected to improve the efficiency of the proposed predictive model In the next stage, supervised machine learning algorithms are used to train the model. For this purpose, the historical data of students is used to train the model. The data is split into training data and testing data to assess the performance of the model.

The model analyzes the patterns of students' behavior and performance to categorize students as high, medium, or low performers. Finally, the performance of the model is evaluated on the basis of accuracy, precision, recall, and F1 score[7]. This approach can be helpful in the early identification of students who are at a high risk of failure. This

can be helpful for the instructor to provide appropriate academic interventions

A. Data Collection

The data used in the current study is obtained from online learning platforms that monitor students' activities during the learning process. The data set used contains information on students' attendance, assignments, quizzes, login frequency, and time spent on learning resources[3]. This information can be used to analyze students' engagement and performance, and then the model can be trained to make predictions using the machine learning algorithm.

B. Data Preprocessing

Data preprocessing is another important step that prepares the collected data set for analysis. During this stage, missing values are identified and removed from the data set to ensure that the data is of high quality[6]. The data set is then normalized to ensure that it can be effectively used to build a model.

C. Splitting of the Dataset

Once the preprocessing is done, the data is divided into two parts: the training data and the testing data. The training data is used to train the machine learning model to learn patterns from the student data, while the testing data is used to test the performance of the trained model[8]. This division is used to measure how well the model can predict the performance of the students on the new or unseen data.

The data is then divided into two sets, which include a training set and a testing set[12]. It is estimated that 80% of the data is used in training, while 20% is used as a test for the model. The test is a measure of the generality of the model, focusing particularly on transactions.

D. Model Testing and Evaluation

Once the model is trained, it is then tested using the testing dataset to assess its performance. The prediction results are compared with the actual values of student performance to determine the precision of the model[13]. Metrics such as accuracy, precision, recall, and F1-score are employed to ascertain the effectiveness of the model in predicting student performance. This helps in assessing the reliability of the predictive system.

Algorithms:

In this research, several supervised machine learning algorithms are used to predict the performance of students in an online learning environment. These algorithms analyze the historical data of students and identify patterns that impact their performance.

Decision Tree is used to classify students based on several attributes like their attendance, scores obtained in quizzes, and assignments completed by students. It constructs a tree that helps make decisions while classifying students.

Logistic Regression:

Logistic Regression is used to predict the probability that a student falls into a specific performance category[14]. It is best used in classification problems.

Random Forest:

Random Forest is an ensemble learning method that aggregates the results from different decision trees to enhance the accuracy of the predictions and prevent overfitting[17].

Neural Networks:

Neural Networks are employed for modeling the intricate relationships existing among the different features of the dataset[19]. They are capable of learning the intricate patterns present in the behavior of students, which aids in the improvement of the accuracy of the predictions.

These algorithms function together for the purpose of analyzing the behavior of students, along with the records, for the prediction of the performance of the students

V. Performance Evaluation Metrics

Performance evaluation metrics are used to measure how accurately the machine learning model can predict the performance of students[18]. These performance evaluation metrics can be useto evaluate the accuracy and reliability of the model.

A. Accuracy

Accuracy measures how accurately the model can make predictions by computing the ratio of the number of correct predictions to the total number

of predictions made by the model.. However[10], the accuracy measure would not be sufficient for understanding the performance of the model for the fraud detection problem, as the frequency of fraudulent transactions is much lower compared to legitimate transactions.

Table I: Accuracy Comparison of Machine Learning Models

Machine Learning Model	Accuracy (%)
Logistic Regression	82%
Decision Tree	85%
Random Forest	90%
Neural Network	88%

B. Precision

Precision refers to the proportion of the actual positive cases that were predicted by the model[18]. Precision tells us how good the model's positive prediction is.

C. Recall

Recall refers to the actual positive cases that were predicted by the model. Recall tells us how good the model's ability to detect relevant cases is[19].D. F1-Score The F1-score is a performance metric that measures the combination of precision and recall. It is a balanced assessment of the performance of the model, especially for datasets that are imbalanced**[20]**. A high F1-score indicates that the model has high precision and recall for the prediction of the performance of the students.

VI. Comparison with Existing Methods

In general, the conventional methods for analyzing the performance of the students mostly involve the evaluation or the application of statistical methods. These conventional methods usually take into consideration a limited number of factors or aspects while evaluating the performance of the students[10]. Moreover, the conventional methods do not usually take into account the various complexities that exist between the different learning behaviors of the students[20]. Hence, the conventional methods do not usually help to effectively predict the performance of the students.

In the proposed system, the machine learning algorithms will be applied to analyze the various aspects of the students' performance, including attendance, submission of assignments, quiz results, login frequency, etc.

VII. Conclusion

This paper has offered a predictive analysis approach to assess the performance of students in virtual learning environments using machine learning techniques. The model can analyze the patterns of students' performance based on their attendance, assignments completed, quizzes completed, and their interaction with course materials[12]. This can assist in the classification of students according to their performance and can identify students who may not perform well.

The experimental results proved that the machine learning approach can accurately predict students' performance. This can assist teachers to offer academic guidance to students who may not perform well. This approach can improve students' performance and reduce the dropout rate among students[20] . In the future, the system can be improved by using real-time data and more advanced the
machine learning techniques to improve the accuracy of the predictions made by the model. machine learning techniques to improve the accuracy of the predictions made by the model.

IX. References

[1]C.Romero.and.S.Ventura
Educational data mining: A review of the state of the art
https://ieeexplore.ieee.org/document/5483663

[2]R.S.Baker,K.Yacef
The state of educational data mining in 2009
https://jedm.educationaldatamining.org/index.php/JEDM/article/view/8

[3] T. Anderson – The theory and practice of online learning
https://www.aupress.ca/books/120146-the-theory-and-practice-of-online-learning/

[4J.L.Rastrollo-Guerreroetal.
Analyzing and predicting students' performance by means,of.machine.learning
https://www.mdpi.com/2076-3417/10/3/1042

[5]S.Kotsiantisetal.
Predicting students' performance in distance learning
https://www.tandfonline.com/doi/abs/10.1080/08839510490442058

[6]J.Xu.and.H.Jaggars
Performance gaps between online face-to-face courses
https://doi.org/10.1353/jhe.2014.0028

[7]M.Hussainetal. Student engagement predictions in online.learning.environments
https://www.sciencedirect.com/science/article/pii/S0747563217305492

[8]G.Siemens,and,R.Baker
Learning analytics and educational data mining
https://dl.acm.org/doi/10.1145/2330601.2330661

[9]M.Shahirietal.
A review on predicting student performance using data mining,techniques
https://www.sciencedirect.com/science/article/pii/S1877050915036182

[10]N.Thai-Ngheetal.
Factorization,techniques,for,predicting,student performance
https://educationaldatamining.org/EDM2011/wp-content/uploads/proc/edm2011_paper_23.pdf

[11] P. Cortez and A. Silva
Using data mining to predict secondary school student performance
https://archive.ics.uci.edu/ml/datasets/Student+Performance

[12] C. Brooks and C. Thompson
Predictive modelling in teaching and learning
https://solaresearch.org/hla-17/hla17-chapter7/

[13] T. Mishra et al.
Students' performance prediction using machine learning algorithms
https://www.ijert.org/students-performance-prediction-using-machine-learning-algorithms

[14] S. Hussain et al.
Educational data mining using WEKA
https://ijeecs.iaescore.com/index.php/IJEECS/article/view/13174

[15] M. Romero et al.
Data mining in course management systems: Moodle case study
https://www.sciencedirect.com/science/article/pii/S0360131507000836

[16] Z. Shou et al.
Predicting student academic performance using multidimensional learning behavior data
https://ieeexplore.ieee.org/document/10412063

[17] N. U. R. Junejo et al.
Multi-category student performance prediction using deep neural networks
https://ieeexplore.ieee.org/

[18] Y. Liu et al.
Fairness-aware machine learning for student performance prediction
https://ieeexplore.ieee.org/document/10434154

[19] Open University Learning Analytics Dataset (OULAD)
https://analyse.kmi.open.ac.uk/open_datase

[20]T.Dietterich
Machine learning for predictive data analysis
https://ieeexplore.ieee.org/document/7080776