International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Student Performance Prediction using Machine Learning

Harshitha[1] , Mrs. C.Mohanapriya[2]

[1] Undergraduate Student [2] Associate Professor Department Of computer Technology, Dr.N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India

https://doi.org/10.55041/ijsmt.v2i3.066

Cite this Article: Harshitha, (2026). Student Performance Prediction using Machine Learning. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.066

Abstract

Education institutions are generating huge volumes of academic data at both student-level and academic performance level. Manual analysis of this data to identify students who might have trouble academically is time consuming for educators and also tends to lack efficiency. If identification of those students are delayed, the negative impact arises from those students having difficulty academically could lead to poor academic outcomes for that student population as a whole.

To address these issues, this research introduces a Student Performance Prediction System using Machine Learning that can support educators in predicting academic outcomes for students and providing timely academic interventions.

The proposed model will analyze data collected from multiple sources of academic data including, but not limited to, attendance, study hours, internal assessment scores, assignment outcomes, and previous academic history; and utilize various Machine Learning algorithms as well as Data Analysis Techniques, to identify trends and relationships between these different data sets and their associated academic outcomes for the student population.

Once all the data is analyzed, the Machine Learning Model will predict how well a given student will likely perform on upcoming examinations. Educators may use these predictions to identify individual students that may need extra academic support based on their predicted performance level. This predictive analysis provides educational institutions with the opportunity to provide early academic intervention strategies, improve teaching strategies, and ultimately enhance the academic performance of their student body.

INTRODUCTION

A recent increase in the amount of data available from schools has provided the opportunity for extensive analysis of student achievement through the use of educational data mining and learning analytics. A thorough analysis of historic patient records provides valuable insight into an institution's student body and can assist in establishing patterns of success and failure based on various factors. Institutions can utilize this knowledge by developing early warning systems that identify students who may be at risk for poor academic outcomes. This information enables institutions to intervene with educational support programs that will help students achieve their full potential.

As technology continues to advance, educational institutions can now use advancements in Machine Learning (ML) and Artificial Intelligence (AI) techniques to study large amounts of data related to education; therefore, they can be accurately predicted how well a student will perform (i.e., their overall achievement). There are many different areas that machine learning algorithms use when predicting how

well a student might perform based on multiple features (or factors) such as attendance, hours spent studying, marks from internal assessments, number of assignments completed and past academic records.

Through ML to develop an intelligent Student Performance Prediction System based on input features of student data. This system will collect data for each student and then will apply ML techniques to detect patterns in the data. Based upon the outcomes of these analyses, the prediction model will indicate how students will likely perform in upcoming exams (i.e., expectations of either performing well, averaging or poorly).

Utilizing an intelligent student performance prediction system will provide educators with greater awareness of their students' progress, which in turn will enable them to identify areas in need of additional support. By providing data-driven insights regarding students, schools will have opportunities to improve their pedagogical practices, encourage greater student engagement and produce increased academic success across their student populations.

Overall, machine learning-based tools for forecasting student success will provide many advantages for educational institutions; these tools can help institutions make proactive decisions and reduce manual effort associated with analyzing data while promoting a more individualized and effective learning experience.

LITERATURE REVIEW

In recent years, the use of data analytics and machine learning in the education sector has increased significantly. Educational institutions generate large volumes of academic data such as attendance records, examination scores, assignment results, and student engagement metrics. Researchers have explored various techniques to analyze this data in order to understand student learning patterns and predict academic performance. These studies aim to assist educators in identifying students who may require additional support and to improve overall learning outcomes.

Romero and Ventura [1] conducted one of the earliest comprehensive studies in the field of **Educational Data Mining (EDM)**. Their research focused on applying data mining techniques to analyze student learning data and identify patterns related to academic performance. The study demonstrated that analyzing historical academic data can help predict student outcomes and support decision-making in educational institutions.

Baker and Yacef [2] further explored the application of machine learning methods in education and highlighted the importance of predictive models in identifying students at risk of academic failure. Their work showed that machine learning algorithms can analyze complex datasets and provide valuable insights that help educators develop targeted intervention strategies for struggling students.

Kotsiantis et al. [3] investigated the effectiveness of various machine learning algorithms in predicting student performance. The study compared different classification techniques such as **Decision Trees, Naïve Bayes, and Artificial Neural Networks**. Their findings indicated that machine learning models can achieve high prediction accuracy when trained with appropriate datasets and relevant features.

Cortez and Silva [4] conducted a study using machine learning techniques to predict student academic results based on factors such as study time, family background, previous grades, and school-related attributes. Their research highlighted that student performance is influenced by multiple academic and non-academic factors, and incorporating these variables into predictive models can improve prediction accuracy.

Recent advancements in machine learning have also encouraged researchers to explore the use of more sophisticated models for performance prediction. Géron [5] discussed how modern machine learning frameworks and tools can be used to build predictive models capable of analyzing large datasets and identifying complex relationships between variables. These techniques have been widely applied in educational analytics to forecast student outcomes and support data-driven educational management.

Han, Kamber, and Pei [6] emphasized the importance of **data preprocessing and feature selection** in building effective predictive models. Their work demonstrated that selecting relevant attributes and cleaning the dataset significantly improves the performance and accuracy of machine learning algorithms.

Despite the progress made in predictive analytics for education, several challenges remain. Student performance is influenced by numerous factors such as

personal motivation, learning environment, teaching quality, and socioeconomic conditions. These factors can make prediction tasks complex and require well-structured datasets and appropriate machine learning models.

Overall, the existing literature highlights the growing importance of **machine learning and data mining techniques in educational systems**. These technologies provide valuable insights into student learning behaviors and enable institutions to identify students who may require academic support. Building upon these studies, the proposed **Student Performance Prediction System using Machine Learning** aims to develop an intelligent model capable of analyzing student data and predicting academic outcomes, thereby helping educators make informed decisions to improve student success.

METHODOLOGY

The Student Performance Prediction System uses machine learning techniques to assess how successful a student will be based on their past performance. Essentially, this process provides a forward-looking prediction of a student's future success or performance based on evaluations of students' academic data records.

The first step in this process includes gathering the right type of academic data from each student, including but not limited to, attendance percentage, number of hours spent studying, scores received on assignments and internal assessments, marks received in previous semesters and all other information that will provide insight into the student's academic performance and learning behaviour. Therefore, without the proper collection of this academic data, the model cannot be created and therefore the prediction will not be reliable nor accurate.

The next step of the process is the preparation of the data sets by performing data preprocessing in order to ensure that the data sets are clean and can be used to represent what is intended to be predicted. There are many challenges involved in preparing the students' data sets for analysis including, but not limited to; null value penalties, inconsistencies among records, duplicate records and irrelevant records. Many methods exist to remove these types of errors; for example, removing duplicates, identifying, and removing irrelevant records.

feature selection process is complete, the next process is feature extraction as shown in the figure below. In this process, features (or attributes) are examined to determine which ones are going to contribute the most towards predicting how well a student will do in their final exam based upon the historical data collected from all students in the same course during the same semester. For instance, the features of attendance, hours studied, on-time completion of assignments, and previous grades will be used in this analysis in order to check how well each feature will predict final performance levels (grades) of all of the students in the course. The use of "good" predictive features allows for more effective machine learning algorithms and reduces the computational overhead of large numbers of attributes being evaluated for making the final predictions.

The next step in the process is the use of machine learning algorithms to train the predictive model using the attributes selected during the feature selection process. The data set is divided into a training and test data set prior to using any machine learning algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), or Logistic Regression for building the predictive model from the training data set. During this training phase, the machine learning system learns about the attributes of the student's record and how they are related to the student's projected final performance level (grade).

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

WORKFLOW

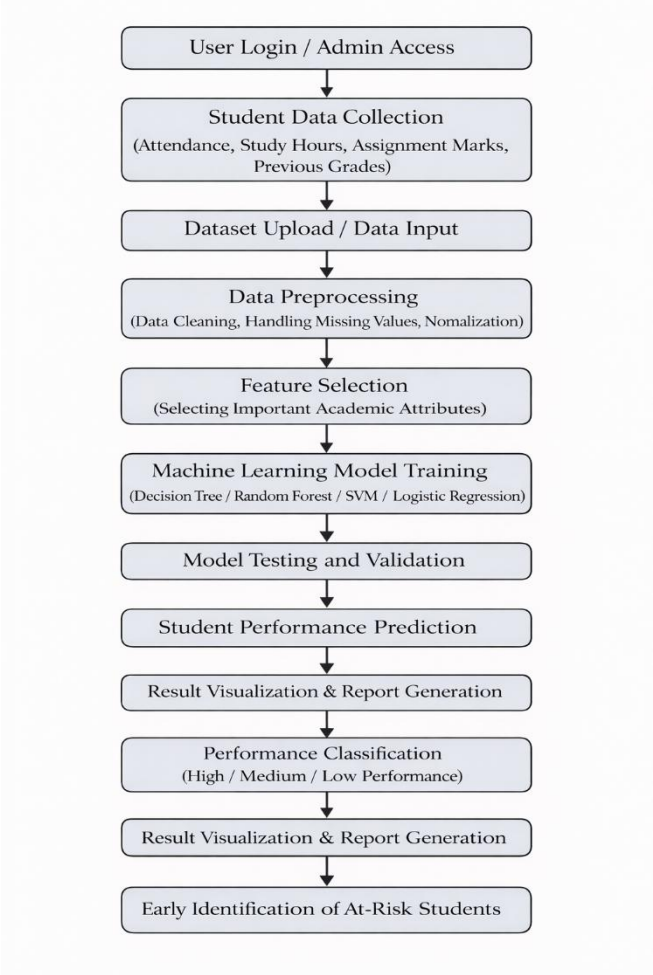

Figure1: Workflow of Proposed Methodology

The design of a prediction system based on student performance using machine learning is described logically in the diagram given in Figure 1. The process starts with the user (administrator or instructor) logging into the system using secure credentials to verify their identity so that only individuals with permission to manage (analyze, etc.) academic data for all students are able to interact with and use the data found within the system.

Once the user has successfully logged into the system, the next step is to collect Student Academics data for storage in a system database. Academic data includes items such as student attendance, hours spent studying, assignment score(s), internal assessment score(s), and previous academic grades that help establish a clear picture of how students learn and progress academically.

Once the Student Academics data has been collected, it will be uploaded as a data set to the system for analysis in order to use this analysis to predict future performance based on past performance. Since this data has been.

CONCLUSION

The Machine Learning-Based Student Academic Performance Prediction System is intended to provide institutions and other entities with a way to assess how well students have performed thus far and to predict what level of academic achievement they will attain in their future educational endeavors. In traditional education systems, educators use their judgment and periodic tests to judge students' progress. If students do not have adequate performance on such evaluations, this may delay identification of these students causing them to fall further behind and have impaired learning. By using machine learning as part of the process of evaluating students' academic progress, this system will give educational institutions the ability to efficiently assess large amounts of data about students, and determine patterns of student performance that are extremely difficult to ascertain via traditional means.

The Machine Learning-Based Academic Performance Prediction System employs machine learning algorithms in order to analyze multiple variables associated with students' Academic Performance, i.e. Attendance, Hours Studied, Assignment Scores, Previous Test Scores. Through the use of data pre-processing, feature selection and model-training techniques, the predicted relationship between these variables and future Academic Performance will be established. The Predictions made by the model can then be used to help identify students that are at risk of failing; and allow educators to develop remediation programs to assist those students.

As a result of the implemented technologies described above, educators will have valuable insight into the learning behaviours of their students as well as their academic achievement. Therefore, by using this Student Academic Performance Prediction System, educational institutions can: have an Automated Process to Predict a Student's Future Achievement; and Implement Interventions; Proactively Identify Students; and Implement Adequate Resources/Support.

An additional significant advantage of using the proposed system is that it enhances the speed and precision with which we analyze large datasets. While

the traditional process is done manually, the proposed automated prediction model calculates education-related data much more efficiently, thus providing consistent output and greatly reducing the number of human errors in evaluating education. In addition, the method will provide educators with structured reports and visual displays related to the prediction and progress of students over time through the use of the machine learning instructors.

The Student Performance Prediction System that utilizes machine learning will show how new methods of analyzing data can help to change traditional educational assessments to a much more efficient and smarter method. By utilizing machine-learning technologies, educational institutions can gain better insight into how students learn, change the way they provide educational assistance to students and grow their students' educational outcomes.

In the future, other types of data, like behavioral data and online learning patterns as well as other psychological factors that affect a student's ability to learn, could be added to the model to improve the operation of the system. Advanced machine-learning technologies, along with access to current and real-time data about education, could also help to create opportunities for...

REFERENCES

1. C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics, vol. 40, no. 6, pp. 601–618, 2010.

2.R. S. J. D. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," Journal of Educational Data Mining, vol. 1, no. 1, pp. 3–17, 2009.

3.S. B. Kotsiantis, C. J. Pierrakeas, and P. E. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, 2004.

4.P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in Proc. Future Business Technology Conference, 2008.

5.H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," Telematics and Informatics, vol. 37, pp. 13–49, 2019.

6.K. Ramaswami and R. Bhaskaran, "A study on feature selection techniques in educational data mining for predicting student performance," International Journal of Computer Applications, vol. 135, no. 7, pp. 1–5, 2016.