

Automated Resume Screening using Machine Learning

Prof. A. S. Phapale¹, Mihir Palatkar², Sanjana Pardeshi³, Prathamesh Pardhi⁴

Department of Information Technology¹⁻⁴

AISSMS Institute of Information Technology, Pune, India

mihir.palatkar@gmail.com, prathameshpardhi6@gmail.com,

sanjanapardeshi65@gmail.com.

Abstract: *As the number of students that attend the current educational institutions continues to increase, assessing and tracking academic performance has become more and more complicated and challenging. The conventional approaches to the analysis of student records are mostly manual based and are therefore time consuming, labor consuming and prone to human error, inconsistency and bias. Consequently, there is an increasing demand to have intelligent systems that are capable of processing huge amounts of educational information and giving reliable information on student performance.*

A Machine Learning-driven Student Performance Prediction System can be used to resolve this problem. Through the use of data-driven methods, such systems can examine various student-related data and can spot anomalies that can be difficult to see otherwise using traditional methods. This will help institutions go beyond the current grade-based assessment and have a more comprehensive picture of student learning behavior.

This paper presents a machine learning-related model to predict student performance based on academic and behavioral features. Some of the key functions that the system entails include attendance records, past academic performance, study habits, classroom attendance, and the level of engagement. All these are processed under data preprocessing methods such as cleaning, normalization, and feature selection to make sure that the input to the prediction models is of high quality.

This is followed by the use of supervised learning algorithms to train predictive models which categorize students into various performance groups. The system will be used to assess whether a student will perform well, average or not and will need any other academic help. The model can be used to support educators by providing more insight into student progress and to identify at-risk students at an early stage with the use of performance scores and categories.

Moreover, automation in the evaluation process will save teachers and administration a lot of manual labor. It improves the accuracy of prediction, provides consistency to the analysis as well as removes the subjective bias. The system is also useful in making decisions based on the data because it delivers actionable insights, which could be applied to develop personalized learning plans and interventions.

On the whole, the suggested solution illustrates that machine learning can change the conventional educational practice into intelligent, efficient, and adaptive systems. Not only does it enhance the precision of performance prediction, but also leads to increased academic performance, optimal use of resources, and higher student achievement rates.

Keywords: Student Performance forecast, ML, Educational Data Mining, Academic Analytics, Classifier Models, Feature Extraction, Intelligent Education System

I. INTRODUCTION

Contemporary educational organizations handle a lot of student-related information in various academic courses, such as grades, attendance, submissions of assignments, and other behavioral indicators. With the amount of such data ever-



increasing, assessing student performance manually has become more challenging and ineffective. Teachers are usually time-limited, and it is not easy to examine the development of every child. Besides, manual methods of evaluation are susceptible to human error, inconsistency, and unconscious bias, which may compromise the objectivity and validity of academic evaluations.

As the digital learning systems and online education systems develop at a rapid pace, a major shift towards data-driven educational practices is observed. There are now large-scale datasets available to institutions that include important information about student behavior and learning patterns. But when there are no smart analytical tools, this data is not utilized. This gives a powerful desire to automated systems that are able to process and interpret student data effectively and correctly.

Machine learning methods provide an effective answer to this issue, as it allows creating predictive models that can be trained using past data. These models are able to discover the latent patterns and associations between different elements affecting student performance. An example of this is where one can study the factors of attendance, past academic performance, study behavior, classroom activity and engagement level on a case by case basis and be able to examine them in combination to determine their effects on learning outcomes. Machine learning models can be used to make a more accurate and comprehensive assessment of multiple variables at the same time, in contrast to traditional methods.

The conventional assessment systems are usually based on the predetermined grading norms and standard tests. Although these techniques are applicable in simple evaluation, they do not tend to provide the complexity of individual learning behaviors. Learning styles may vary among students, as well as their engagement and other extraneous factors that can influence their performance. Rule-based systems have restricted capabilities of adapting to such differences and can also miss out on the students that need extra help or intervention. Moreover, these systems are not flexible and are incapable of changing according to the changing school settings or novel data patterns.

In recent years, machine learning and data analytics have evolved to allow the development of intelligent systems that are able to learn and improve as time goes on. Decision Trees, random forest, support vectors machine (SVM) and Artificial Neural Networks (ANN) are popular algorithms used in classification and prediction in the educational field of data mining. These algorithms are capable of dealing with large data sets, complex relationships among variables and make precise forecasts with minimum human intervention. Consequently, they are very effective in detecting students who risk poor performance and facilitate the prompt correction measures.

The fourth important benefit of machine learning-based systems is that it helps with personalized learning. These systems can be used to offer personalized information about students data that can be used to enhance learning performance. These insights can assist educators to create specific teaching plans, provide extra assistance to those who learn poorly, and improve the general level of education. This is necessary to change the one-size-fits-all model to a model of personalized learning that can enhance the success rates of students in the contemporary education systems.

The main aim of the research is to come up with an intelligent student performance prediction model based on machine learning. The suggested system will examine various student data, derive meaningful features, and make accurate predictions on academic performance. In this way, it will help educators to find the patterns, learn about the needs of students, and establish effective academic planning and intervention strategies.

Finally, machine learning can revolutionize the performance evaluation of students by being embedded into the educational systems. It enhances efficiency, accuracy, as well as fairness, transparency, and making decisions that are data-driven. The study would help in the creation of such smart systems, which would lead to more efficient and responsive learning environments.

II. RELATED WORKS

Student performance prediction is a recent research topic in the field of educational data mining and learning analytics. The use of machine learning methods in advancing academic performance, student engagement, and data-driven decision-making in schools has become a growing area of research interest among researchers. The increasing access to



online learning resources and student information have also contributed to the rapid advancement of predictive models in this field.

Much of the previous studies have focused on the classification-based methods where student data is examined to determine whether or not a student would pass/fail or attain a specific grade. Naive Bayes, Decision Trees, Support Vector Machines (SVM) and random forests are some of the algorithms that have been extensively employed because of their ability to deal with structured data. These models are usually based on aspects like attendance, assignment grades, past academic achievement and involvement rate. Of these, the ensemble approaches such as Random Forest have been more accurate since they use many decision trees to minimize overfitting and maximize generalization.

Besides the classification methods, data preprocessing and feature extraction are also important to enhance the performance of the models. They have used text mining and data analysis techniques to derive significant information using unstructured student data. As an illustration, attendance record, submission rates and performance in exams are evaluated to establish major predictors of academic performance. The features are then utilized in training machine learning models so that they can learn relationships among various variables and predict more accurately.

Clustering and regression models are other approaches that have been examined in the literature based on similarity. Clustering methods help in grouping students with similar traits and as a result, teachers can see performance trends among various groups of students. This aids in recognizing the impact of different factors on academic performance including the behavior of learning, the level of engagement, and study habits. In contrast, regression models serve to forecast continuous variables like final grades or GPA, which gives a more thorough evaluation of student performance.

Deep learning models have been of significant interest in recent research with the development of computational capabilities. Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, and transformer-based models are neural networks that can learn complex and non-linear correlations in data. The models are particularly effective when working with large-scale data as well as with time-series, e.g., student activity logs provided by online learning environments. Deep learning techniques can be much more accurate than traditional machine learning models in terms of prediction accuracy, with enough data.

Although there is great progress in the area of student performance prediction, there are still a number of challenges that persist. The fact that big and quality data are needed to train and accurate models is one of the great constraints. Data can be missing, uneven, or erratic in most learning institutions, which can adversely affect the performance of the models. Indicatively, the dataset that has a greater percentage of high-performing students can result in biased predictions which will decrease the accuracy of the system.

The other issue that is quite significant is the interpretability of complex models, especially deep learning systems. Although these models are highly accurate, they may be hard to interpret by educators on how predictions can be made. Such a lack of transparency may decrease the level of trust in the system and restrict its application in the real world. Also, ethical concerns, including data privacy, fairness, and bias should be properly considered to make sure that predictions do not discriminate against certain categories of learners.

To address these issues, recent studies have been directed at creating efficient, scalable and interpretable machine learning systems. To trade-off between accuracy and interpretability, hybrid models that use a combination of several algorithms are under investigation. Resampling and weighted learning are also being adopted as techniques to deal with unbalanced data to enhance model fairness.

Moreover, an increased focus is on developing adaptive systems that are capable of changing to adapt to changing educational conditions. Such systems will constantly learn based on new information so that predictions will be relevant and accurate as time goes by. By integrating machine learning with real-time data analysis and feedback mechanisms, modern predictive systems aim to provide more personalized and effective learning experiences.

To conclude, current literature shows that machine learning has a large potential in forecasting student performance. Nevertheless, the systems which are not only accurate but also efficient, interpretable, and adaptable are still needed.



The proposed work is built upon such developments by creating a machine learning-based framework that is able to overcome these problems and offer reliable and actionable information to educators

III. OPTIMIZATION, SUSTAINABILITY, AND INTEL- LIGENT RESOURCE MANAGEMENT

Machine learning in predicting student performance has gained relevance in contemporary academic systems where educational institutions must handle extensive data pertaining to students effectively. Conventional assessment techniques that are intensive in terms of manual grading and periodic testing tend to be very time consuming and burdensome to teachers. Furthermore, the approaches are prone to inaccuracies, subjectivity and error. These restrictions show how intelligent, automated systems can be useful in improving the efficiency and accuracy of the academic evaluation processes.

The main aim of optimization in machine learning-driven educational systems is to optimize prediction accuracy with minimal computational power and use of resources. These systems help in decreasing the manual involvement in order to analyze student data, thus facilitating rapid decision making. Machine learning algorithms can find significant patterns within a complicated dataset, discovering the factors of attendance patterns, academic history, the degree of participation, and learning behavior. Such features are then applied to make predictive models which can effectively estimate student performance. The efficiency and reliability of these systems are further improved by optimization methods which include feature selection, hyperparameter optimization and evaluation of the models.

Sustainability in education does not just mean long term success in the academic arena but it also denotes the effective use of the institutional resources. Predictive systems based on machine learning are helpful in being sustainable as they allow anticipating students who are at risk of poor performance early. This enables the educators to carry out timely and specific intervention, including personalized mentoring, extra academic support, or customized learning plans. When the issues of students are addressed in the initial level, institutions will be able to cut down dropout rates, enhance the overall academic performance, and make the learning environment more conducive.

Also, the systems encourage transition towards proactive education management as opposed to reactive. Predictive models enable institutions to foresee possible problems and act proactively, rather than reacting to the poor performance at the time of occurrence. This proactive model improves the academic performance of students as well as minimizes the financial cost of a failure in the academics over time including the need to take a repetitive course or prolong the number of years taken to graduate.

Another important factor of machine learning-based systems is effective resource management. Educational facilities have to manage big data that comes through various channels such as learning management systems, online learning and administrative databases. Machine learning models can be built in such a way that they can handle such data efficiently, even when their workload increases. Scalable architectures and optimal algorithms are used to make sure that data is processed fast and with high precision without the high-computational cost.

Moreover, most of the contemporary systems have feedback that enables the process of continuous learning and improvement. New student information is received, and the models are revised to demonstrate the new tendencies and trends. This dynamic functioning capability makes sure that the system can be pertinent and precise over time, as the educational settings develop. Feedback loops also allow institutions to perfect their teaching strategies depending on the predictive information thus creating a constant improvement in both the teaching and learning processes.

The possibility to facilitate personalized education is another crucial aspect of intelligent resource management. Machine learning systems can make recommendations specific to each student by examining student data separately, which allows them to offer individualized recommendations to students. This will assist the educators in putting their time and effort in the right direction since they will be able to target students that need extra attention and motivate high-performing students to reach their potential.

Although these benefits exist, implementing machine learning in education also has a number of challenges that need to be considered. One of the primary issues is the data privacy, as student data is sensitive and should not be exposed to



unauthorized access. The institutions should make sure that data is gathered, stored and processed in adherence to the ethics and legal requirements.

Another important concern with predictive systems is fairness and bias. In case the training data is biased, the model can give unjust predictions to the disadvantage of some groups of students. Thus, there is a need to develop transparent, interpretable, and discrimination-free models. The reliability and trustworthiness of such systems can be enhanced with the help of techniques like bias detection, fairness assessment, and explainable AI.

Transparency is also relevant to the attainment of acceptance by the educators and stakeholders. Machine learning models must be able to explain clearly how the predictions are made and the users can then comprehend and trust the results. This is especially crucial in the educational context, where the outcomes of student results can be heavily affected by the decisions made on the basis of predictions.

In summary, systems of student performance prediction using machine learning could provide a potent solution to streamlining the educational process, making it sustainable, and managing the resources better. These systems can help promote efficient and effective learning by automating data analysis, facilitating proactive intervention, and assisting with personal learning. Nevertheless, one should take into account ethical issues, including privacy, fairness, and transparency to make sure that such technologies are utilized in a responsible and helpful manner.

IV. PROPOSED ARCHITECTURE FOR DEEP LEARNING-BASED AUTOMATED RESUME SCREENING

The architecture to predict student performance with machine learning is meant to be a high-accuracy, scalable and efficient academic evaluation system. It is organized in a layered style such that each element has a given task and is in co-ordination with other modules. The system is flexible, easy to maintain and adaptable to various learning settings due to this modular design. The general architecture is geared towards converting raw student data into significant forecasts that can guide the educators in decision making.

The system is made up of the following components:

• User Interface Layer

The User Interface (UI) Layer serves as the interface between the users (teachers, administrators or academic coordinators) and the system. It offers a user-friendly and easy-to-use interface, where users can enter student information, see the outcomes of the prediction, and track the overall performance trends.

The UI layer displays insights including the category of student performance, the level of risks, and progress reports through dashboards and visualizations. This layer is user friendly in order to ensure that the system may be well understood and used even by non technical users. It can also be equipped with features such as data entry forms, report generation and graphical representation such as charts and performance summaries.

• Data Collection & Preprocessing Layer

This layer collects raw data that may include academic data bases, learning management system, attendance record and online platforms. Raw data is usually incomplete, inconsistent and noisy thus preprocessing is essential in enhancing the quality of the data.

The preprocessing phase involves a number of steps which involve:

- Removing missing or incorrect values
- Scaling numerical data to normal.
- Removing irrelevant or redundant features.
- Turning categorical data into numbers.
- Running feature extraction and selection.
- This layer will provide high-quality data to the machine learning models because raw data is converted to clean and structured data and this directly enhances prediction quality.



• **Prediction Engine**

The main element of the system is the Prediction Engine where machine learning algorithms are used to process the data received and present predictions. This layer learns the patterns based on previous student history and applies these patterns to forecast a future performance.

An array of machine learning models can be deployed in this layer, such as:

Rule-based classification, decision trees.

Random Forest with better accuracy and minimize overfitting.

• High dimensional data analysis Support Vector Machines (SVM).

• Artificial Neural Networks (ANN) of complex patterns.

Another option is to incorporate deep learning methods in more sophisticated implementations to improve prediction further. The prediction engine classifies students into categories such as high-performing, average, or at-risk, based on their data.

• **Database and Storage Layer**

Database and Storage Layer: This layer is in charge of data storage of any kind of data that is utilized in the system. This contains raw student data, processed data, model parameters, and prediction outcomes.

The structured database guarantees effective data retrieval and scalability so that the system can deal with large datasets without performance problems. It is also conducive to historical data storage, which is fundamental in the process of training and enhancing machine learning models as time goes by.

Moreover, this layer helps to maintain data security and integrity through the application of appropriate storage controls, access controls and backup systems. This will guarantee the protection of sensitive information about students and the availability of the information to be analyzed.

• **Learning Layer and Feedback.**

Feedback and Learning Layer is an important aspect of ensuring the system is adaptive and constantly improving. In contrast to the static system, machine learning-driven models have the advantage of continually updating as new information is provided.

Feedback of prediction results and user interfaces is gathered in this layer. As an illustration, when the actual performance of a student is not in line with the predicted performance, the system is able to retrain and adjust the model accordingly. In the long run, it leads to better prediction accuracy and reliability.

The feedback mechanism also enables the system to change according to changes in patterns of education, changes in curriculum or change in the trend of student behavior. This ensures that the model remains relevant and effective in dynamic learning environments.

• **Overall System Functioning**

Overall, the architecture is organized into a methodical flow wherein student data is gathered, processed, and analyzed and converted into practical predictions. The combination of these layers makes the system run effectively and accurate, as well as scaled.

The synergy of data preprocessing, smart prediction model, and the feedback system provides a strong structure to predict the performance of students. This architecture enhances academic evaluation but also personalized learning and proactive intervention strategies.

V. METHODOLOGY

The main goal of the suggested methodology is to develop and deploy an AI-powered system capable of assessing and forecasting student achievements on the basis of multiple academic and behavioral variables. It is a systematic pipeline



methodology that includes several steps because it is essential to make sure that raw data is converted into valuable predictions in an organized and effective way.

The initial phase is data collection, in which data associated with students is collected through various means including academic databases, learning management systems (LMS) and institutional records. The data gathered usually contains the following attributes; student grades, attendance, submissions of assignments, involvement in classroom activities, and involvement in online learning environments. To increase the accuracy of the predictions, more contextual information, including socio-economic background or study patterns, can also be taken into account, in certain cases.

After collecting the data, it is followed by data preprocessing, which is crucial to enhancing data quality and to guarantee a robust performance of the model. Raw data is frequently full of missing values, inconsistencies, and noise that may adversely affect the results of prediction. Thus, the data is cleaned using preprocessing methods. This involves dealing with missing data, eliminating duplication, correcting inconsistency and standardizing numerical data to a common scale. Encoding techniques are used to convert categorical data into numerical form and irrelevant or redundant features are eliminated to simplify data.

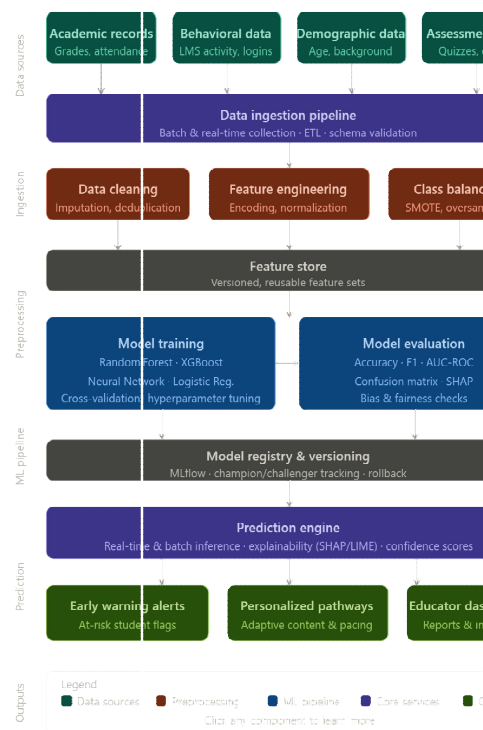


Figure 1: Architecture of Student Performance Prediction using Machine Learning

After preprocessing, feature selection and extraction are implemented in the system, which is key in determining the most relevant factors that affect student performance in the system. Not every feature is equally relevant to prediction accuracy and choosing suitable attributes can be used to improve model efficiency and can also minimise the computational cost. Such techniques include correlation analysis and statistical methods which are applied to identify key features like the consistency in attendance, past academic performance and participation levels.

The second step is the model training during which machine learning algorithms are used on the processed dataset. Depending on the nature of the data, different supervised learning models can be applied, including Decision Trees, Random Forest, Support Vector Machines (SVM) and Artificial Neural Networks (ANN). The models are trained to



have patterns and relationships between the input features and student outcomes. The dataset is usually split into training and testing sets in order to test the performance of the models.

Upon having undergone training, the system is then taken through the prediction phase where the trained models are applied to predict the performance of the students on data that has not been seen. The model categorizes students into various groups based on the acquired patterns, high-performing, average, and at-risk. Such forecasts are useful in terms of monitoring the progress of students and can assist the teachers to determine the ones who might be in need of extra academic assistance.

The machine learning models are evaluated by the use of the evaluation stage. The measurement of how the model can predict student outcomes is in terms of accuracy, precision, recall, and F1-score. By comparing several models, it is possible to choose the best algorithm to be used.

Lastly, a feedback mechanism is also included so that the system is continuously improved. With the advent of new data, the model is modified and retrained to capture the modern trends and patterns. This dynamic learning system is such that the system is correct and pertinent in the long run. The model can enhance predictions and be able to adapt to the changes in student-behavior and educational settings because it keeps learning and updating on the new data.

In general, the suggested methodology offers a detailed outline of how to create an intelligent student performance prediction system. It is a method of data processing, machine learning, and continuous improvement that provides accurate and reliable results.

VI. EXPECTED OUTCOMES AND FUTURE WORK

It is anticipated that the proposed system will considerably improve efficiency and effectiveness of student performance evaluation in schools. The system saves time by automating the process of analyzing the data related to students, thus, eliminating the need to use manual systems to evaluate the data, which may lead to mistakes.

Early identification of at-risk students is one of the key expected outcomes. Anticipating the performance trends will allow the educators to take proactive measures early enough like offering supplementary support, special guidance or remedial courses. This proactive measure assists in enhancing academic success and minimizes the rate of dropouts.

This also encourages objective and unbiased judgement since the predictions are made using objective data and not subjective views. This makes there to be uniformity in decision-making and boost the reliability of academic tests. Moreover, machine learning allows institutions to consider a number of factors at once, giving a more in-depth picture of student performance.

The other significant effect is the optimization of the allocation of resources. Schools are able to channel their energies on those students who need maximum attention thus enhancing the effectiveness of teaching and learning activities. This will result in improved faculty time and institutional resource utilization.

Regarding the future work, there are a number of improvements which can be investigated to further complement the system. A possible way is to incorporate more sophisticated methods of deep learning, including Long Short-Term Memory (LSTM) networks and transformer models, which can be used to detect temporal trends among student data and enhance predictive accuracy.

The other aspect that has evolved is the introduction of real-time data analysis where the system would be used to monitor student performance on a regular basis and give instant feedback. This would facilitate the dynamic decision-making and quick intervention where needed.

The system can also be expanded to incorporate individualized learning suggestions, in which students are given individualized study plans and resources, depending on their projected performance. This would make students more engaged and help them have individual learning experiences.

Lastly, one can look at enhancing the interpretability and fairness of the model used in future research and make sure that predictions are not made in an oblique way. Adding explainable AI methods can assist educators in comprehending the way predictions are made and build the confidence in the system.



VII. CONCLUSION

The growing amount of student data in the past few years has led to a high demand of smart systems that are able to analyze and interpret such data. The old-fashioned evaluation techniques based on the manual evaluation and predetermined grading systems cannot cope with the sophistication of the educational environment today.

A machine learning-enhanced system of student performance prediction can serve as an effective and efficient way to resolve this issue. Using the methods of data analysis, the system can examine several variables, which impact student performance, and make valid forecasts. This can help teachers to understand student behavior more and make effective decisions that will improve academic performance.

The approach suggested incorporates the main elements of data preprocessing, feature extraction, model training, and predictive analysis to provide consistent and reliable results. Involving a feedback mechanism will guarantee that the system will keep on getting better as time goes, adjusting to new data and evolving educational trends.

Furthermore, the system also facilitates a transition to proactive and personalized education, where interventions are informed by predictive insights, as opposed to reactive ones. This does not only increase student success rates but also overall efficiency of learning institutions.

To sum up, machine learning can make the conventional education systems an intelligent, adaptive, and data-driven environment. Mainly because it allows the proper forecasting of the student performance, it is an important tool in enhancing academic performance, efficient resource use, and making informed decisions. The suggested system is one of the major steps towards the fulfillment of these objectives and emphasizes the necessity of the incorporation of modern technologies into education.

REFERENCES

- [1] Prediction of Student's Performance with Learning Coefficients using Regression based Machine Learning Models (2025)
- [2] Predicting Students' Performance Using Machine Learning Techniques (2025)
- [3] Graph-based Ensemble Machine Learning for Student Performance Prediction (2024)
- [4] GritNet: Student Performance Prediction with Deep Learning (2020)
- [5] Machine Learning-Driven Student Performance Prediction for Enhancing Tiered Instruction (2024)
- [6] Student Performance Prediction Using Machine Learning (2025)
- [7] Educational data mining for student performance prediction: feature selection and model evaluation (2024)
- [8] Application of Corrected Methods for High-Resolution XRF Core Scanning Elements in Lake Sediments (2020)
- [9] Predicting students' performance in online courses using multiple data sources (2021)
- [10] Baleen: Robust Multi-Hop Reasoning at Scale via Condensed Retrieval (2022)

