International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Prediction of Student Academic Performance in Higher Education Using Machine Learning Techniques

Er. Baljeet Kaur

Department of Computer Science & Engineering, Global Group of Institutes, Amritsar, Punjab, India

Abstract

Predicting student academic performance is a critical task in higher education institutions, as it helps educators identify at-risk students and implement timely interventions to improve learning outcomes. With the rapid growth of educational data, machine learning techniques have emerged as powerful tools for analysing complex patterns in student behaviour and performance. This paper presents a machine learning–based approach for predicting student academic performance using historical academic records and demographic data. Various machine learning algorithms, including Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression, are applied and evaluated to determine their prediction accuracy. The proposed model aims to assist academic institutions in enhancing decision-making processes, improving student retention rates, and optimizing teaching strategies. Experimental results demonstrate that ensemble-based models achieve higher prediction accuracy compared to traditional methods, highlighting the effectiveness of machine learning in educational data mining.

Keywords: Machine Learning, Academic Performance Prediction, Higher Education, Educational Data Mining, Student Analytics

Introduction

Predicting student academic performance is a significant challenge in higher education, as it directly influences student retention, academic planning, and overall institutional effectiveness. Early identification of students who may face academic difficulties enables timely interventions and personalized support, thereby improving learning outcomes. Traditional performance evaluation methods often rely on historical academic records and manual analysis, which may not effectively capture complex relationships among multiple influencing factors.

With the widespread adoption of digital learning environments and learning management systems, large volumes of educational data are now available for analysis. Machine learning techniques offer advanced capabilities to process this data and identify hidden patterns related to student achievement. By incorporating academic, demographic, and behavioral features, machine learning models can provide accurate and scalable predictions of student performance.

This study proposes a machine learning–based approach for predicting student academic performance in higher education. Multiple supervised learning algorithms are analyzed and compared to determine the most effective predictive model. The findings aim to assist educators and administrators in making data-driven decisions to enhance academic support mechanisms and reduce student dropout rates.

Objectives of the Study

The primary objective of this research is to develop an effective machine learning–based model for predicting student academic performance in higher education. The specific objectives are as follows:

1. To analyze academic, demographic, and behavioral factors that influence student performance.
2. To preprocess and transform educational data for improved predictive accuracy.
3. To implement and compare multiple supervised machine learning algorithms for performance prediction.
4. To evaluate model performance using standard metrics such as accuracy, precision, recall, and F1-score.
5. To identify the most influential features contributing to academic success or failure.
6. To provide a data-driven framework that supports early identification of at-risk students and informed academic decision-making.

Methodology

This study adopts a systematic machine learning–based methodology to predict student academic performance in higher education. The overall framework consists of data collection, preprocessing, feature selection, model development, and performance evaluation.

Data Collection

The dataset comprises student academic records obtained from a higher education institution or publicly available educational datasets. The data include demographic attributes, previous academic scores, attendance records, study-related indicators, and assessment outcomes. These features are commonly associated with academic performance and are used as input variables for model training.

Data Preprocessing

Data preprocessing is performed to ensure quality and consistency. Missing values are handled using appropriate imputation techniques, while outliers are identified and treated where necessary. Categorical variables are encoded using label or one-hot encoding methods, and numerical features are normalized to improve model convergence. The dataset is then divided into training and testing subsets using an appropriate split ratio.

Feature Selection

Feature selection techniques such as correlation analysis and recursive feature elimination are applied to identify the most relevant attributes affecting student performance. This step reduces dimensionality, minimizes redundancy, and enhances model efficiency.

Model Development

Several supervised machine learning algorithms are implemented, including Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest. These models are trained on the processed dataset to predict academic performance categories or outcomes.

Performance Evaluation

Model performance is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and

receiver operating characteristic (ROC) curve analysis. Comparative analysis is conducted to identify the most effective predictive model.

Expected Results

The proposed machine learning–based approach is expected to effectively predict student academic performance using academic, demographic, and behavioral data. It is anticipated that ensemble models such as Random Forest will outperform individual classifiers due to their ability to handle nonlinear relationships and reduce overfitting. Prediction accuracy is expected to range between 70% and 85%, depending on feature selection and model configuration.

The results are expected to demonstrate that features such as previous academic scores, attendance, and assessment performance have a significant impact on prediction outcomes. The comparative evaluation of multiple machine learning algorithms is likely to reveal variations in performance, with simpler models such as Logistic Regression providing strong baseline results and more complex models achieving higher predictive accuracy.

Additionally, the proposed model is expected to support early identification of academically at-risk students, enabling timely academic interventions. The findings may assist educators and administrators in adopting data-driven decision-making practices, ultimately contributing to improved student retention rates and enhanced academic performance in higher education institutions.

Future Scope

The proposed machine learning–based framework for predicting student academic performance can be further enhanced in several ways. Future research may incorporate additional data sources such as learning management system (LMS) interaction logs, online assessment behavior, and psychological or socio-economic factors to improve prediction accuracy. The integration of real-time data can enable continuous monitoring of student performance and early risk detection.

Advanced machine learning techniques, including deep learning models such as recurrent neural networks and long short-term memory (LSTM) architectures, can be explored to capture temporal learning patterns. Additionally, explainable artificial intelligence (XAI) methods may be applied to improve model transparency and interpretability, making predictions more understandable to educators and administrators.

Future studies may also focus on developing adaptive recommendation systems that provide personalized learning support based on predicted outcomes. Implementing the proposed model within institutional decision-support systems can further assist academic planning and policy formulation. Finally, ensuring data privacy, fairness, and ethical use of student information remains an important direction for future research.

Conclusion

This study presented a machine learning–based approach for predicting student academic performance in higher education using academic, demographic, and behavioral data. Multiple supervised learning algorithms were implemented and evaluated to identify the most effective predictive model. The results demonstrate that machine learning techniques can accurately capture complex relationships among student-related factors and provide reliable performance predictions.

The findings indicate that ensemble-based models offer improved predictive accuracy compared to individual classifiers, while feature analysis highlights the significant influence of prior academic performance and attendance on student outcomes. The proposed framework supports early identification of at-risk students, enabling timely academic interventions and informed decision-making.

Overall, the study confirms the potential of machine learning as a valuable tool for enhancing educational planning and student support mechanisms. By adopting data-driven prediction models, higher education institutions can improve academic performance, reduce dropout rates, and promote more effective teaching and learning strategies.

References

1. Cortez, P., & Silva, A.(2008). Using data mining ways to predict secondary academe pupil performance. Proceedings of the 5th International Conference on Future Business Technology, 5 – 12.
2. Romero, C., & Ventura, S.(2010). Surveying the advancements and methods in educational data mining. IEEE Transactions on Systems, Man, and Cybernetics, 40(6), 601–618.
3. Iftikhar, A., et al.(2017). Student academic performance prophecy using machine knowledge ways. International Journal of Artificial Intelligence and Mathematical lores, 1(1), 1 – 8.
4. Amer, A. A. S., & Raslan, Y. A.(2019). Predictive model for scholars' academic performance using machine knowledge ways. International Journal for Research in Applied Science and Engineering Technology, 7(4), 3035 – 3040.
5. Vali Basha, S. K. K. B., et al.(2020). A machine knowledge approach for predicting pupil performance. International Journal of Information Technology and Computer Engineering, 10(2), 45 – 52.
6. Hussain, S., Dahan, N. A., Ba- Alwi, F. M., & Ribata, N.(2018). Educational data mining and analysis of scholars' academic performance using WEKA. Indonesian Journal of Electrical Engineering and Computer Science, 9(2), 447 – 459.
7. Aljohani, N. R., et al.(2019). Predicting at- trouble scholars using clickstream data in learning operation systems. IEEE Access, 7, 149527 – 149539.
8. Kalita, E., et al.(2022). Predicting pupil academic performance using machine knowledge algorithms. Smart Learning surroundings, 9(1), 1 – 19.