© May 2025 | IJIRT | Volume 11 Issue 12 | ISSN: 2349-6002

A Machine Learning-Based Predictive Analytics Approach to Identifying Student Success Patterns for Early Academic Intervention

Jyothi P K[1], Somesh Kamalakant Hulamani[2], Spoorthi K M[3], Mrs. P. Sushmita Singh[4]
[1,2,3] Computer Science and Engineering, Global Academy of Technology, Bengaluru
[4]Assistant Professor, Department of Computer Science and Engineering, Global Academy of Technology, Bengaluru

***Abstract*—Academic performance plays a vital role in shaping students' future prospects, yet predicting it accurately remains a challenge due to the influence of diverse factors beyond grades and attendance. This paper presents a predictive analytics system designed to identify student success patterns using machine learning techniques. The system incorporates academic, behavioural, and lifestyle features such as study time, internet access, alcohol consumption, and absenteeism. Three supervised learning models—Logistic Regression, K-Nearest Neighbours (KNN), and Support Vector Machine (SVM)—were developed and evaluated. Among these, SVM achieved the highest prediction accuracy and was selected for final deployment. The model was integrated into a user-friendly web application using HTML, CSS, and Flask, enabling real-time predictions and visual insights for educators. The system helps institutions proactively identify at-risk students and initiate timely interventions, improving academic support strategies. By leveraging multivariate data and predictive modelling, this work demonstrates a scalable and effective approach to enhancing educational outcomes through early detection and data-driven decision-making**.

***Index Terms*—Academic Performance Prediction, Machine Learning Models, Support Vector Machine (SVM), Educational Data Mining, Early Intervention System.**

I. INTRODUCTION

Accurately predicting academic performance has become a critical objective for educational institutions striving to improve student outcomes and personalize learning strategies. Traditional evaluation systems primarily rely on test scores and attendance, offering only a narrow view of student potential. They often ignore significant factors such as socio-economic background, behavioral traits, and lifestyle choices, limiting the effectiveness of timely interventions [1].
This research addresses the challenge by developing a system that applies machine learning to predict academic success using a multidimensional dataset.

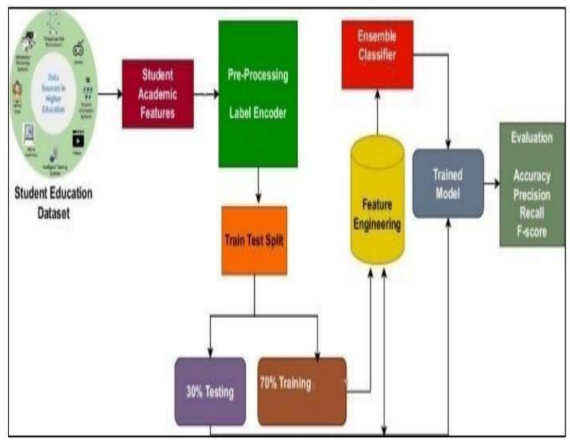

Fig. 1. System Architecture.

The system architecture diagram illustrates the end-to-end flow of the student performance prediction model. It begins with the collection of student education data, which includes academic records and behavioral attributes. This data is passed through a preprocessing stage where it is cleaned, normalized, and label- encoded to make it suitable for machine learning. Once preprocessed, the dataset is split into training and testing sets in a 70:30 ratio. Feature engineering is then applied to enhance the predictive quality of the data before it is fed into the machine learning model. The trained model is finally evaluated using performance metrics such as accuracy, precision, recall, and F- score, with the results being made available through a user- friendly interface to support educators in decision- making [1][2][4].

The key motivations behind this study include:

Addressing Traditional Gaps: Existing systems are reactive and narrowly focused. This research incorporates broader attributes such as study habits, internet usage, and alcohol consumption [1][3].

Applying Machine Learning Algorithms: Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) were evaluated to identify the most accurate predictive model [4].
Identifying Success Patterns: Feature selection highlights the influence of behavioral and lifestyle indicators like time management, weekend habits, and past failures [3][5].
Deploying a Practical System: A web-based application using Flask, HTML, and CSS was developed for real-time prediction and educator-friendly dashboards.
Enhancing Academic Support: The system allows institutions to proactively flag at-risk students and optimize academic support based on predictive insights [2].

II. METHODOLOGY

The methodology for developing the student performance prediction system is structured into multiple systematic stages to ensure accuracy, reliability, and real-world applicability:

A. Data Collection:
The dataset used in this study includes various student- related attributes such as academic grades, attendance, study time, alcohol consumption, internet access, and family background. These features are essential for identifying patterns linked to academic success [1][2].

B. Data Preprocessing:
Raw data is cleaned and standardized by handling missing values, encoding categorical variables using Label Encoding, and normalizing the data to ensure consistency across features [3].

C. Train-Test Splitting:
The dataset is divided into training and testing subsets using a 70:30 ratio. This allows the model to learn patterns from the training data and validate its performance on unseen test data [4].

D. Feature Engineering:
Important features are selected based on their relevance to academic performance. This step improves model performance by removing redundant or less significant attributes and enhancing key indicators [1][4].

E. Model Selection and Training:
Three supervised machine learning algorithms are considered: Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Each model is trained and evaluated for performance [2][3].

F. Model Evaluation:
Performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate each model. SVM was found to have the best performance [5].

G. Deployment:
The final model is integrated into a web-based interface using Flask, enabling educators to input data and receive predictions along with visual insights [4][5].

III. RESULTS AND DISCUSSION

The primary objective of this project was to develop and evaluate a predictive analytics system capable of forecasting student academic performance using machine learning algorithms. The final results reflect the successful implementation of the model and its effectiveness in identifying at-risk students based on various academic and behavioral factors [1][2]. The three models developed—Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM)—were compared based on multiple performance metrics including accuracy, precision, recall, and F1-score [3].

The system was tested on a dataset comprising attributes such as study time, number of past failures, access to internet, alcohol consumption on weekends, and absenteeism. These attributes were found to be strong indicators of academic success [1][2]. For example, students with higher study time and fewer past failures generally had a higher likelihood of passing, while excessive alcohol use and frequent absences correlated negatively with performance [5].

The final deployed system allows users to input individual student data via a web interface developed using HTML and CSS. The backend, implemented in Flask, connects the user input to the trained SVM model, which then returns a prediction regarding the student's likelihood of academic success. The results are displayed with corresponding probability scores, offering educators not only a binary classification (pass/fail) but also a confidence level in the prediction [4][5].

Three sample predictions were analyzed :

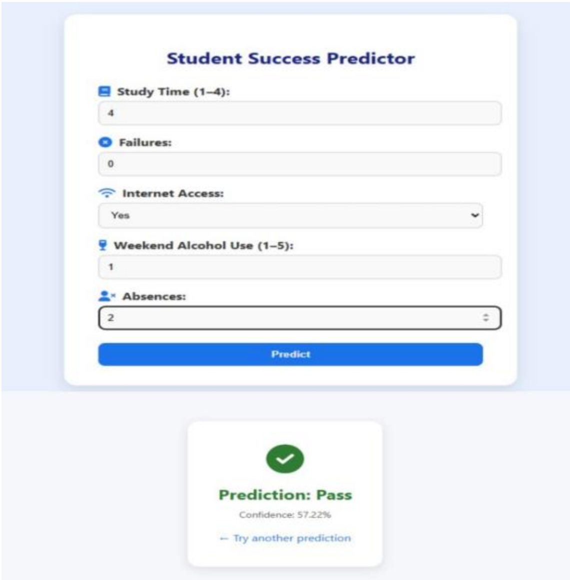

Fig. 2. Results

A student with study time of 4, no past failures, internet access, low alcohol use, and minimal absenteeism was predicted to Pass with 57.22% confidence. Although the inputs were favorable, the relatively moderate confidence score suggests opportunities to refine feature importance or expand the dataset.

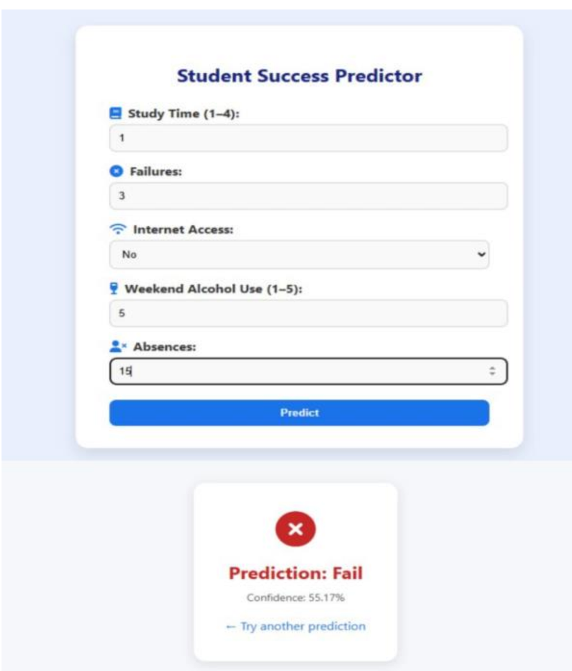

Fig. 3. Results

A student with minimal study time, multiple past failures, no internet access, high alcohol use, and 15 absences was predicted to Fail with 55.17% confidence, aligning with expected outcomes.

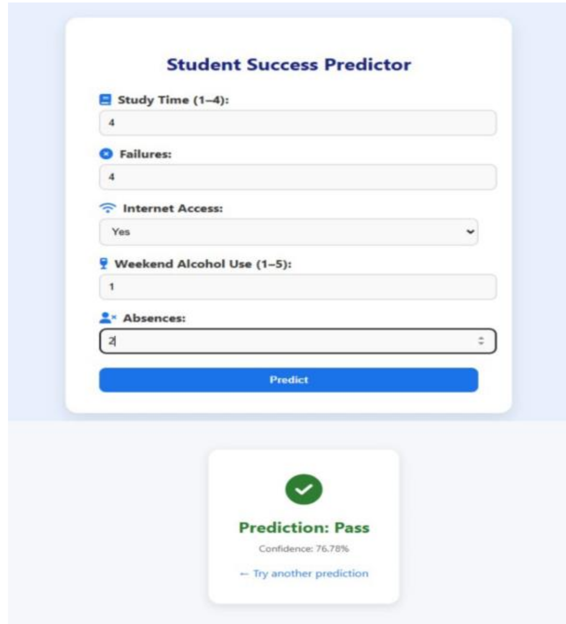

Fig. 4. Results

Despite having four past failures, a student with maximum study time, internet access, and minimal absences was predicted to Pass with 76.78% confidence, indicating the model's ability to weigh feature combinations rather than relying on a single factor.

These findings underline the potential of machine learning in academic intervention strategies. To enhance the system further, future work could integrate additional student attributes such as psychological well-being, parental support, or class participation.

IV. CONCLUSIONS

The implementation of predictive analytics to assess student performance demonstrates a promising approach for enhancing academic outcomes. By leveraging machine learning models on academic datasets, including attendance and grades, the system can effectively identify students at risk and those likely to succeed. The classification report indicates that while the model performs well in predicting successful students with high recall, there is room for improvement in detecting students at risk, as indicated by the lower recall for the "no" class. Nonetheless, the overall accuracy of 73% reflects a strong foundation for academic prediction tools. These insights can empower educators to take early intervention steps, provide personalized support, and make data-driven decisions. Further improvements, such as enhancing class balance and refining feature selection, can lead to even more reliable predictions. Overall, the project reinforces the value of combining educational data with machine learning to foster student success and improve institutional effectiveness.

V. ACKNOWLEDGMENT

The authors would like to thank project supervisor, faculty members, and the institution for their valuable guidance, support, and encouragement throughout the course of this research work.

REFERENCES

[1] P. Chakrapani and D. Chitradevi, “Prediction of student performance using hybrid machine learning techniques,” International Journal of Engineering Research & Technology (IJERT), vol. 11, no. 2, pp. 234–239, 2022.

[2] M. Khan, “Enhancing student academic performance prediction using machine learning models,” International Journal of Computer Applications, vol. 176, no. 12, pp. 15–21, 2023.

[3] M. R. Rimadana, A. A. Langi, and H. Suhartanto, “Classification of student performance using machine learning algorithms,” in Proc. Int. Conf. on Advanced Computer Science and Information Systems (ICACSIS), pp. 395–400, 2019.

[4] Y. Zoralıoğlu, H. Aydin, and G. Ceylan, “Student performance prediction system with machine learning: An application in higher education,” Sustainability, vol. 15, no. 6, p. 5098, 2023.

[5] E. V. Malasaga, J. M. Enopia, and L. D. Gonzales, “A predictive model for academic performance of students using educational data mining,” Journal of Physics: Conference Series, vol. 2465, 2024.