International Journal of Engineering & Scientific Research
Vol. 14 Issue 5, May 2026
ISSN: 2347-6532 Impact Factor: 6.660
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Student Academic Performance Prediction Using Machine Learning Model

[1]Nisha Madaan, [2]Sanjay Kumar Sharma
[1]pgphb23021_nisha@banasthali.in
[1,2]Banasthali Vidyapith, Rajasthan

Abstract

Forecasting student achievement is essential in educational environments to enhance academic success and reduce dropout rates. This research focused on improving student performance prediction through integrating advanced machine learning (ML) techniques. The dataset from Kaggle contains 20 features of the student's academic, demographic, motivational, and other. The study showed a strong relationship between the students' academic achievement and their academic data, like previous exam scores, regularity, sleeping hours, study hours, and demographic features, as well as some other features. Students' academic achievement is predicted using the ML model. Then, these features were input into various ML models, such as LR, RF, KNN, SVM, DT, and NB. The experimental results show that the SVM model outperformed other models after applying hyperparameter tuning on the models and achieved the maximum classification accuracy of 0.9440. Also, a comparison of ML models evaluated through metrics has been presented. Also, graphical analysis shows that attendance, previous exam score, and study hours have a higher impact on a student's academic achievement. This paper promotes the enhancement of quality education and the development of student skills as essential components for workforce advancement and sustainable industrialization, following the UN Sustainable Development Goals (SDGs) aims. By using this technique, teachers will enable students to monitor their academic progress according to their performance and adjust their study habits accordingly.

Keywords: Academic Achievements, Machine Learning, Exam Score, Student Performance, Sustainability

1. Introduction

Sustainable development refers to progress that fulfills current needs while ensuring that future generations can also meet their own needs without limitations,' United Nations [30]. However, sustainable development is not an easy thing to do; it requires addressing major global problems in a deeply set, complex, and multidimensional social [30]. And as the demand for sustainable development and environmental rejuvenation increases among the general public, so does the demand for academics to contribute to it by involving environmental protection, social inclusion, and economic well-being (Baghdadi et al., 2020; Tarhini et al., 2022; Pradhan et al., 2022) [30]. Data science and technology can help solve the world's sustainable development challenges. Although these studies provide valuable information, there are still many open questions on how to align technology with the SDGs. To jointly develop practical answers, their study must place a high priority on interdisciplinary collaboration across disciplines like information systems, environmental science, and policy studies. Second, more research is required to fully understand how emerging technologies—like blockchain and artificial intelligence—interact and what it means for resilient and equitable sustainable practices [30].

The world is changing very fast and carrying technological developments and theoretical models. Artificial intelligence (AI) is a development that has already changed human lives, and technology produces the outcome based on input data into the computer after analyzing that data. It includes the ability to learn, analyze, solve problems, and give the output that is in human-understandable language. We are using so many advanced technologies related to AI, such as digital phones, social media with algorithms, and booking systems for railways, banking, and hospitals, with the continuous technological evolution uniquely taking place. "The concept of sustainable development must be interpreted as a form of economic growth that guarantees the satisfaction of present-day wants without endangering the capacity of future generations to satisfy their own needs and appropriate actions intended to have long-term effects and intervals," according to some Romanian studies. Stefanescu (Stefanescu et al. 2009; [28]). The aim of the competition for institutions of higher education is to ensure that learners are educated for their future positions in society. These environments should enable students to have priceless scientific knowledge, which helps them acquire the skills needed to function in multicultural and multidisciplinary perspectives within intricate processes in a globalized world.

Education is an essential element of economic progress, so different strategies are used to improve student outcomes. A way to do this is by monitoring student progress [24]. The key purpose of all universities is to prepare young students with relevant knowledge and skills to be competitive in the labor market. An additional effective way of achieving this goal is to predict student performance on time. As a result, it is simple to recognize students who require assistance and take steps to achieve better educational results. This would enable teachers to adopt effective teaching techniques [25]. The students are hired for in-field research or experimental studies in a related parallel area. For instance, professors can pay special attention to undergraduates whose performance achievement is projected to be low or below average, and this can help the student greatly, as well as the educational institute in general, as its aggregated result percentage can increase [24].

A significant number of studies identified in the literature suggest that various factors may impact a student's academic success. These features include gender, high school grades, parental educational background, financial status, the medium of instruction, geographical location, prior semester grades, performance in seminars, test scores, assignment outcomes, class and laboratory attendance, overall proficiency, interest in specific courses, study habits, time dedicated to studying, family support, previous academic records, type of admission, and accommodation arrangements. A scholar's academic performance is influenced by their previous grades, and class performance is an important indicator of their potential for academic achievement. Students' consistent performance in their academic careers may help to explain this conclusion. Early on in their college career, students who set the habit of performing well will carry that through to their academic careers. This is also the case for students who have a history of academic difficulties; they might carry on with the same conduct throughout their academic careers, which could negatively affect their performance in both present and upcoming schoolwork [8] and [11]

The power or guidance that pushes you to come closer to your goal and destination. It comes into two categories: i) Intrinsic Motivation and ii) Extrinsic Motivation. Intrinsic motivation refers to doing something because it is intrinsically interesting. Extrinsic motivation refers to the drive that originates from external factors. This type of motivation can encompass intangible rewards such as recognition or acclaim, as well as tangible benefits like financial compensation or academic grades. Lastly, giving students a variety of learning and assignment options may help to promote motivation since students are more likely to internalize behaviors in environments that support their sense of autonomy and freedom to carry out plans [1], [6].

We create a detailed framework derived from our research, highlighting the essential connections among sustainability, education, and artificial intelligence (AI), particularly within the realm of sustainable education. This diagram will depict how AI can strengthen educational sustainability by enhancing student engagement, forecasting academic performance, and strengthening institutional support mechanisms. The framework's key elements include “Sustainability in Education”, which addresses lifelong learning, equitable access to education, and the promotion of student engagement to secure enduring academic and professional achievements. “AI in Education” serves a pivotal function by facilitating personalized learning experiences, employing predictive analytics to separate trends in student performance, and integrating intelligent tutoring systems that offer customized educational assistance. Another crucial component is “Academic Performance Prediction”, which entails the examination of students' learning behaviors, social interactions, and psychological characteristics to evaluate their academic development and identify potential obstacles. Lastly, “Institutional Support” utilizes AI-driven strategies to detect at-risk students early, implement focused interventions for student success, and formulate policies that encourage sustainable educational advancement. This cohesive framework illustrates how AI-enhanced solutions can refine learning environments, improve student outcomes, and foster a more inclusive and sustainable educational system. Fig. 1 shows the relationship of sustainability with higher education, AI, and Institutional support.

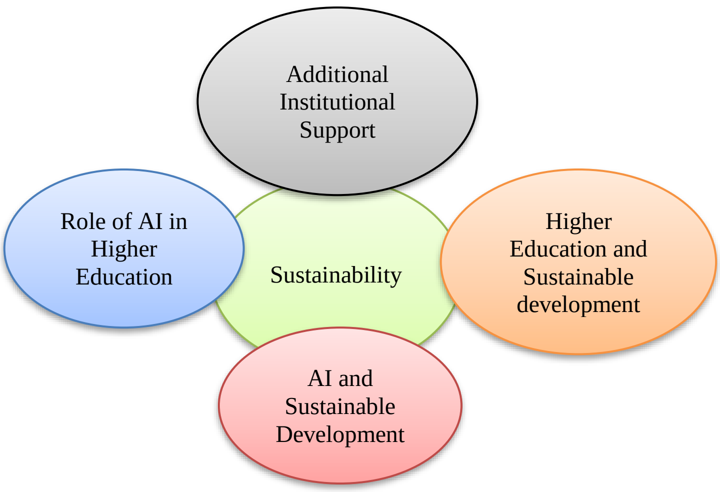

Fig.1 AI, Higher Education, Institutional Support, and Sustainable Development Relations

The system can predict the academic success score (grade) of the student. We also analyze the features that affect the students' academic achievement. The designed system can be used in institutions to document low-performing students. The institution can run a program for these students to motivate them and provide additional support that helps to increase their academic performance and make them successful.

1.1 Features Influencing the Academic

The basic goal of education has always been to improve students' academic achievement. Researchers and educators have conducted several studies over the years to examine the elements that influence (positively or negatively) student progress in their educational track.

It was suggested that evaluating student academic achievement would be challenging due to socioeconomic, psychological, and environmental factors influencing student success. Exams are a unique function in assessing a student's academic success. In truth, three elements strongly impact students' exam performance: demography, academic atmosphere (Study environment), and socioeconomic considerations.

- **Demographic factors:** Personal features, including age, gender, weight, eating preferences, and impairments, as well as the kind of family system, living area, and sibling arrangement, are examples of demographic variables.
- **Academic Manner factors:** Those features that directly impact a student's academic achievements at the postsecondary education level are known as academic environmental factors. Continual evaluations of students' performance, the kind of education, the kind of institution, the location of the organization, the language of teaching, sequestered tutoring, final grades, the kind of community selected for higher education, and additional activities are all included.
- **Social Status factors:** Families possessing a high socioeconomic status tend to be more engaged in preparing their young children for school, largely due to their access to diverse resources that promote and support early childhood development. These families can offer their children high-quality childcare, as well as a variety of books and educational toys, thereby enabling them to engage in numerous learning activities within the home environment. Families with a high socioeconomic status also tend to have a higher parental education level and occupation.

Other Factors

1) **Motivation is defined as the power or guidance that pushes you to reach** your goal and destination. It comes into categories: i) Intrinsic Motivation, ii) Extrinsic Motivation
2) **Study Strategy** is students' planning to achieve their aims and goals.
3) **Psychological Factor:** Those factors include students' cognitive, emotional, physical, personality, and self-esteem attributes.
4) **Student Behavior:** A student's manner or actions that are used to achieve the learning process. It includes all the factors related to their academic environment, like time management, claims, peer support, responsibility towards their work, respect, self-motivation, and many others related to studies.
5) **Human Personality Traits:** Those traits that make a student more responsible about their duties, work, skills, attitude, and personality that make a good human being.

Providing students with the best learning environment and knowledge is the goal of any educational institution. It can be crucial to categorize the students who require extra aid and to take the necessary steps to improve their results. The ML can be used to identify students' accomplishments and poor-performing students and to improve study strategies. This research aims to explore the existing techniques used for students' academic performance analysis and identify the features used in the ML model and metrics for performance analysis of the model. We methodically approached the review process.

The key goals of this work are to forecast students' academic performance, identify the characteristics that most influence academic achievement through correlation analysis, and assess the effectiveness of various models to identify the best one for predicting student success and directing academic interventions. The rest of the paper is organized as follows. The literature review section first offers a thorough analysis of previous research on ML-based models for student performance prediction. The procedures of feature extraction, parameter optimization, model evaluation, and data collection and preprocessing are covered in detail in the study methods section.

2. Literature Review

We study and explore many research papers related to student academic performance prediction in this work. Some of the work has been briefly described, along with the technique used for predicting students' performance and results. Finally, the analysis of these papers allows us to draw some conclusions about the work presented by different authors.

Orji et al. (2022) proposed an ML method for forecasting students' performance attainment and education approaches based on their inspiration (motivation). Their study aimed to develop models applicable to any subject in higher education by utilizing key features such as intrinsic and extrinsic motivation, independence, relatedness, competence, and self-esteem. Using Scikit-learn in Python, they implemented DT, KNN, RF, LR, and SVM for both classification and regression tasks. Their results indicated that tree-based models, particularly RF with an accuracy of 94.9%, outperformed other techniques. Similarly, Bendangnuksung et al. (2018) applied a DNN to classify students into pass/fail categories, demonstrating an 85% accuracy in identifying at-risk students. Altabrawee and Ali (2019) further explored the impact of students' internet and social media usage on academic achievement using ML techniques such as NB, DT, and LR. Their results show that the ANN model is the most effective, achieving 80.7% accuracy.

Gajwani and Chakraborty (2021) examined demographic, psychological, and academic factors influencing performance, employing feature selection and supervised learning methods like RF, DT, and ensemble techniques. Their findings revealed that ensemble methods, particularly boosting and bagging, achieved the highest accuracy (75%). Rai and Shastry (2021) focused on addressing challenges posed by large educational datasets, comparing RF and SVM classifiers and concluding that RF yielded superior predictive performance. Bujang et al. (2021) addressed the issue of imbalanced datasets in multi-class student grade prediction using techniques such as SMOTE and feature selection. Their study reported that an RF-integrated model significantly improved prediction accuracy, achieving an F-score of 99.5%.

Kale et al. (2022) introduced the 4Q&S model, which integrates intelligence, adversity, spiritual, emotional, and stress factors to enhance educational outcomes. Their research emphasized the importance of a holistic approach to student performance prediction. Likhitha et al. (2022) compared logistic regression, DT, and SVM models, reporting that DT achieved the highest accuracy (96%) in predicting student performance. Abulhaija et al. (2023) leveraged data mining techniques to analyze student performance at Princess Sumaya University for Technology, identifying course grades, high school averages, and semesters attended as key predictors. Their findings demonstrated that RF and DT performed better than NB and ANN. In a review study, Yadav et al. (2023) examined various ML techniques, such as ANN, SVM, NB, LR, and DT in student performance prediction, emphasizing the role of attribute selection in improving accuracy. Similarly, Sri Lalitha et al. (2023) analyzed students' personality traits and behaviors, classifying them as introverts or extroverts using SVM, RF, and DT, achieving 95% accuracy.

Dhilipan et al. (2021) employed binomial logistic regression, entropy-based DT, and KNN classifiers to predict student performance based on previous academic records. Collectively, these studies underscore the increasing role of machine learning in academic performance prediction, offering insights into the most effective models and features for identifying at-risk students and enhancing learning outcomes. Rastrollo-Guerrero et al. [26] reviewed more than 70 papers to show the many approaches that are currently widely utilized to predict student achievements and the objectives that must be accomplished to do so. ML, collaborative filtering, recommender systems, and ANNs are some of the AI-based approaches and techniques that have been researched. Conversational agents embodied in social robots and other forms of communicative AI engage with students in ways that go beyond encouraging their cognitive growth. They can also act as supportive peers or tutors to promote affective development, which includes enhancing learning motivation, interest, self-control, empathy, and teamwork (Chin et al., 2014).

UNESCO IESALC representative Ms. Victoria Galán-Muros unveiled the new Sustainability Evaluation Tool for Higher Education Institutions (SET4HEI). The tool establishes a broad framework that enables universities to evaluate their SDG contributions through teaching operations, research activities, and management procedures. According to Galán-Muros, the tool maintains an open and adaptable design that enables institutions to customize their evaluation processes according to their particular contexts. Students in relevant fields experience multiple advantages through research and experimental studies. The effectiveness of this fact demonstrates that the faculty should target students with low-performance predictions or those near performance standards to benefit both the students and the institution through improved overall high-grade percentages.

Junejo et al. (2025) developed a 1D-CNN-based multiclass model that predicts student outcomes ("Distinction," "Pass," "Fail," and "Withdrawn") using OULAD data. Their model achieved 92% accuracy in early course stages, outperforming traditional models like RF and ANN-LSTM by focusing on clickstream and demographic features [31]. Abdullah et al. (2025) applied ensemble ML models, including LightGBM, to predict students' online exam answer correctness and choices using the NeurIPS 2020 dataset. LightGBM achieved 74.17% accuracy in predicting correctness and 65.7% for multiple-choice selection, demonstrating the strength of ensemble learning in question-level prediction [32]. Sandeepa & Mohottala (2025) evaluated traditional ML classifiers on school-student data—behavioral, academic, and demographics—with an MLP classifier achieving 86.46% on the test set (79.58% in 10-fold CV). Their work underscores the effectiveness of neural networks and explainable models (e.g., feature-importance, explainable ML) in improving performance forecasts [33].

Althaqafi, Saleem, & Al-Ghamdi (2025) emphasize the impact of class imbalance handling (e.g., oversampling, weighted loss) in ML-based student performance prediction. Their results show that addressing imbalance significantly improves accuracy, recall, and fairness—making previously underrepresented student groups more accurately modelled [34]. Chen et al. (forthcoming 2025) offer a systematic literature review on ML applications

predicting student performance, engagement, and self-efficacy in higher education. They synthesize findings across hundreds of empirical studies and identify key predictive variables such as engagement metrics and learning motivation, while also pointing out methodological gaps and future research directions [35]. Moubayed et al. (2024) evaluate DL (CNN, LSTM-RNN) models to predict student performance at mid-course across three geographically diverse datasets. They find deep models outperform traditional ML in two datasets and remain comparable in the third, demonstrating the global applicability and robustness of such approaches in online learning contexts [36].
Mr. Mohammed Alshehri from Majmaah University explained their approach to merging AI-based educational advancements with their sustainability goals. The university merges AI applications with sustainability-focused projects to prompt students to apply data analysis and forecasting models for resolving environmental and societal issues. The speaker emphasized the experimental knowledge gained from students' use of AI tools in real-world projects like resource administration and ecological monitoring to develop practical explanations consistent with the SDGs. The HESI Global Forum 2024 panel reached a consensus on developing partnerships among universities, manufacturing sectors, government bodies, and civil society to advance SD and promote innovation within higher education. The panel emphasized how inclusive interdisciplinary approaches and emerging technologies, such as AI combined with student-centered education, ensure no one is left behind in reaching SDG objectives. The discussion underlined how universities must lead SD advancement while partnerships remain essential for transformative progress.

3. Research Methodology

In this research work, the methodology contains 3 parts, shown in Fig. 2: (i) data collection, (II) data preprocessing, and (III) ML model formation and evaluation. Data was collected from an open repository system, that is Kaggle, and the data preprocessing stage includes steps such as handling missing values, solving inconsistencies, removal of redundancy, feature selection, and normalization of the dataset. Then, the transformed data record was converted into a normalized data form. In the model construction and evaluation, the normalized dataset is divided into two parts: the train and the test datasets(30%). The training data is used to build a predictive model—potentially using algorithms such as LR, RF, KNN, DT, and SVM by capturing underlying patterns. The model is subsequently validated on the testing set to evaluate its generalization capability. Finally, model performance is assessed using technical metrics like accuracy, precision, recall, and F1-score, ensuring the robustness and effectiveness of the predictive system.

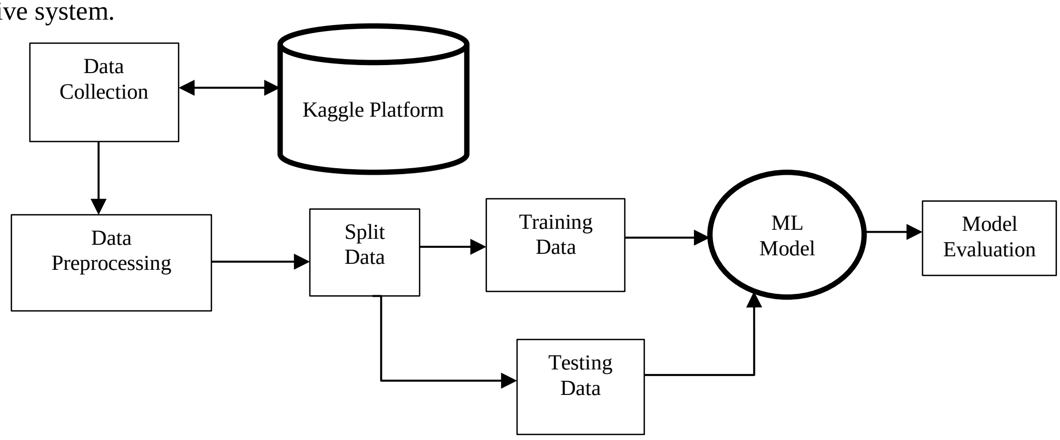

Fig. 2 Proposed Research Work

Data Collection: In this research, the dataset was sourced from Kaggle and encompasses a range of academic, demographic, and additional characteristics of student performance. Initially, the dataset comprised 6,607 variables, including studied hours, Attendance (regularity), Sleeping Hours, Previous Exam Scores, Tutoring Sessions, Sports Activity, and Current Exam Scores. Furthermore, the dataset features categorical variables such as Parent Participation, resource access, Extracurricular Activities, Motivation, Internet usage, Family Revenue, Educator Teaching Quality, University Type, Peer Impact, Learning Disabilities, Parent Education Level, Distance from Home, and Gender, which show socio-economic and demographic factors. Tutor Quality, Parental Qualification Level, and Distance from Home exhibit missing values. This dataset serves as a robust basis for examining the various academic, personal, and socio-economic elements of student performance.

Data Preprocessing: For each dataset, we first use Label Encoder and the OneHot Encoding technique to convert the categorical data into numerical ones. For the categorical values "Male" and "Female" of the attributes "Gender," we translate the numerical values "0" and "1." Additionally, characteristics are normalized by removing the mean and scaling them to unit modification using the StandardScaler [31], and missing data is handled with SimpleImputer. 70% of the dataset is used for model training, while 30% is used for model testing.

Our data set contains a mix of variables and classifications. To choose a subset of features that are appropriate for resolving the intended learning problem, the attribute selection task attempts to eliminate redundant and superfluous features from the dataset. These unimportant characteristics lower the learning algorithm's accuracy. Using this approach, we compute the key characteristics, which also aid in assessing the model's output. We determined each ML model's optimal accuracy. Before presenting the findings or creating a model for an outcome, it is crucial to optimize the algorithm's parameters to improve algorithm performance. In machine learning, parameter optimization entails modifying hyperparameters to enhance model performance.

Five ML algorithms were applied for model validation: (i) Logistic Regression (LR), (ii) Random Forest (RF), (iii) K-Nearest Neighbour (KNN), (iv) Decision Trees (DT), and (v) Support Vector Machine (SVM). The Python programming language and Google Colab Environment were used to obtain results that are presented using metrics.

4. Analysis and Discussion

Several ML techniques are used to build the model to determine which model best predicts the scholar's success. The best model is identified by estimating the model's performance and comparing the outcomes. Below are the specifics of the research that was conducted and the measures that were employed to examine the model's success.

1) **Logistic Regression (LR) Model:** The relationship between the target variable and the logistic function's input features is measured by the LR. The other variables are input variables, and grades are the target variable. To make predictions based on the test data, the logistic regression technique is utilized. After that, use the hyperparameter. The penalty is set to 'none', while the solver parameter is set to 'liblinear'. The model is then trained using the training datasets X_train and y_train, and it is fitted using the lr_model object. The value of the hyperparameter max_iter is 1000. The LR model's accuracy is 93%.
2) **Random Forest (RF) Model:** Applied the RF method to datasets belonging to students. Since the model performs better when max_features is increased, we set max_leaf_nodes to 9 and utilize the "criterion" value for entropy. Max_depth (6) and random_state (50) have specified values. The model's prediction is enhanced by the RF in training data using the randomForest() technique with parameter adjustment. With a 92% accuracy rate, this model is more dependable when a grid search strategy is used.
3) **K-Nearest Neighbour (KNN) Model:** To construct the model, the KNN method is employed. Using cross-validation parameter adjustment, we optimize the choice of k in this instance to attain high accuracy and minimal misclassification error. The model performs 86% with the lowest misclassification error when k is 5. With the lowest misclassification error and the best prediction, k=5 is the ideal value.
4) **Decision Tree (DT) Model:** The Splitting principle and the minimum size for split attributes are determined by the model using the hyperparameter tuning. More pruning results from the higher values of the ccp_alpha, which controls pruning. class_weight modifies each class's relative relevance, and we utilize a criterion to calculate the split quality measure using 'entropy'. The maximum depth of the tree is 10; the maximum number of features taken into account in splits is max_features; and the maximum number of leaf nodes is max_leaf_nodes. The DT model has an accuracy of 89%.
5) **Support Vector Machine (SVM) Model:** The kernel, regularization parameter (C), gamma, and degree are the four parameters that are implemented by the SVM method. It locates data points from every class and establishes their boundaries. In this model, the "rbf" kernel was employed. Using the GridSearch technique, which provides the optimal kernel parameter, we tune the function for the C value, which is assumed to be 10. The polynomial kernel's degree parameter of 3 is the optimal one. A hyperplane of higher degree indicates greater flexibility. 94% is the accuracy of the SVM model.

The Python and the Scikit-Learn, Pandas, Numpy, matplotlib, and Seaborn libraries [39] were used to implement the model. The academic datasets are used to train and test the model that was created using a particular ML algorithm. Each model's classification reports and model performances. Fig. 3 shows the accuracy and F1-scores for models used in this work.

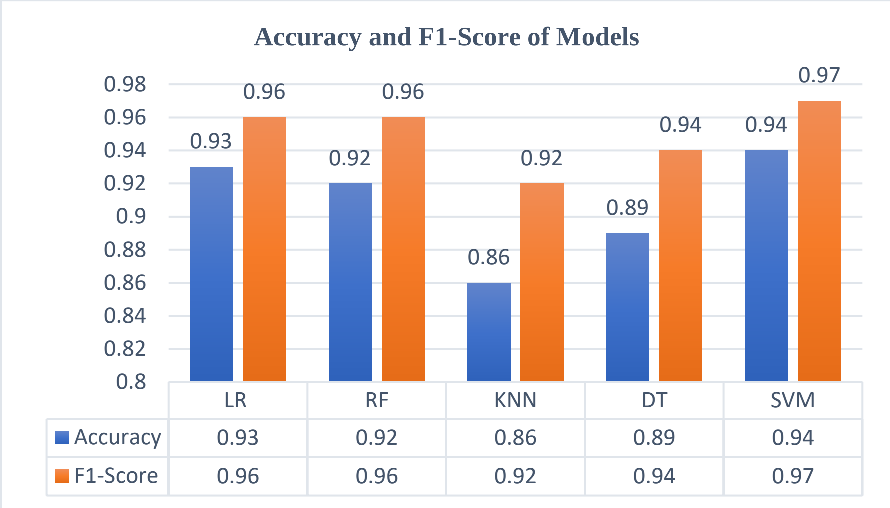

Fig. 3 Comparison of Accuracy and F1-Score of Different Models

To eliminate sample-bias in our work, we used k-fold CV, k=5. As indicated in Table 1, hyperparameter optimization significantly improves the academic attainment of the students' performance prediction accuracy. The results show that, with an accuracy of 94% the SVM model is the best; the accuracy of the logistic regression model is 93%. And 86%, KNN has the lowest accuracy.

Table 1: Performance Comparison

Models	Accuracy	Precision	Recall	F1-Score
LR	93%	94%	97%	96%
RF	92%	93%	99%	96%
KNN	86%	87%	98%	92%
DT	89%	92%	95%	94%
SVM	94%	96%	98%	97%

4.1 Graphical Analysis

Understanding the factors that affect academic achievement is crucial for effective education. The three factors have an impact on scholars' academic performances: attendance, study hours, and previous scores. Also, some other factors impact performance, like tutoring sessions, parental education level, peer influence, and extracurricular activity. This work examines the relationship between these features and students' academic accomplishment. The graphical visualization of the dataset in Fig. 4 reveals that attendance has a higher impact on performance.

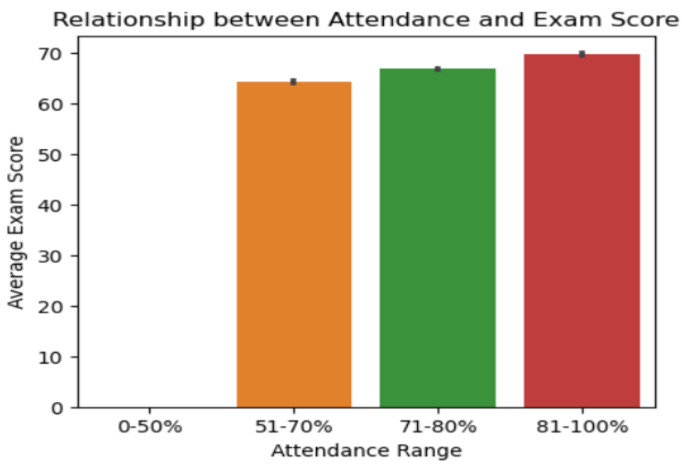

Fig. 4 Exam score vs Attendance

This scatter plot shows in Fig. 5 the **Hours Studied** and **Exam Scores**, having an **optimistic correlation** between the two features. The hours studied increase, and exam scores tend to rise, indicating that students who dedicate more time to studying attain better results. Some outliers exist—students who studied fewer hours but scored exceptionally high, as well as those who studied extensively but received moderate scores. This suggests that factors beyond study hours, such as learning techniques, individual aptitude, and exam difficulty, might also influence performance. After analysis, it is suggested that **more study time is beneficial**, but the **quality of study plays an equally important role** in academic success. This suggests that **consistent classroom engagement** significantly enhances learning outcomes, reinforcing the importance of attendance policies and early intervention for irregular attendees.

The graphical analysis shows that increased academic engagement quantified through higher attendance and more study hours positively influences student performance. These findings provide empirical justification for including these features as dominant predictors in ML models aimed at academic performance prediction. Moreover, the insights emphasize the importance of institutional strategies that promote consistent attendance and effective study habits to enhance student learning outcomes.

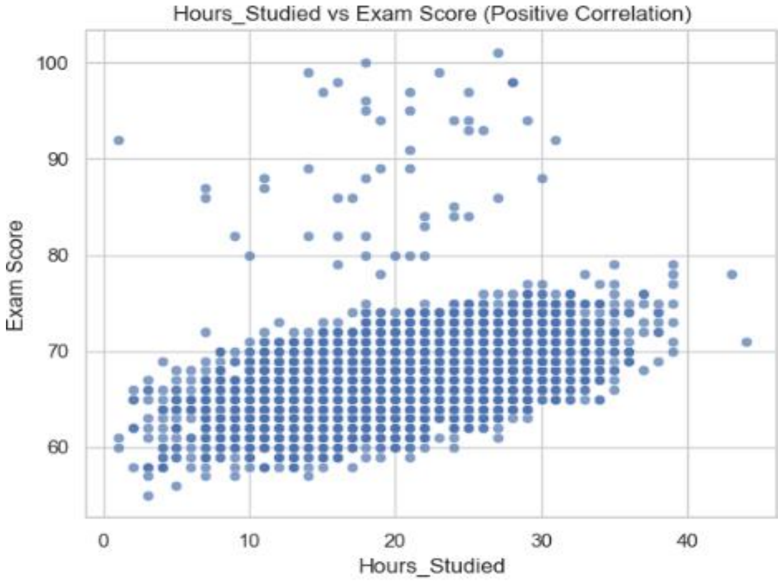

Fig. 5 Exam score vs Hours studied

The heatmap of the correlation matrix shown in Fig. 6 reveals various factors influencing student exam scores. Among these, **attendance (0.58)** has the strongest positive correlation, indicating that students with higher class participation tend to perform better. **Hours studied (0.45)** also play an important role in exam success, reinforcing the importance of dedicated study time. Additionally, **previous scores (0.18)** and **tutoring sessions (0.16)** have moderate correlations; prior academic performance and additional guidance can contribute to better outcomes. However, factors like **sleep hours (-0.02)** and **physical activity (0.03)** show no correlation, implying that they have little to no direct impact on exam performance. Overall, this suggests that **consistent attendance and study habits are the most influential factors** in achieving higher exam scores.

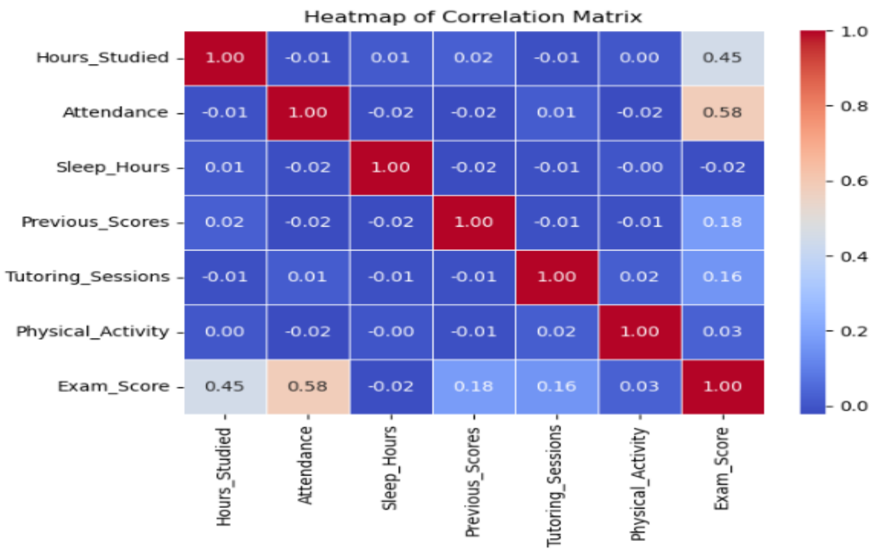

Fig. 6 Heatmap Correlation

5. Conclusion and Future Work

Student achievement plays a crucial role in the education sector by improving academic success rates and minimizing dropouts. This study aimed to boost student achievement using advanced machine learning (ML) techniques. The dataset, sourced from Kaggle, comprises 20 features, including academic, demographic, and behavioral attributes. Several ML models, including LR, RF, KNN, SVM, DT, and NB, were used for classification. The Support Vector Machine (SVM) model achieved the highest classification accuracy of 0.9440 after hyperparameter tuning. A comparative study of ML models using different evaluation metrics was conducted, highlighting key influential factors in academic success. The analysis demonstrated a significant correlation between students' academic achievement and features such as prior exam grades, attendance, study hours, and sleep patterns. Regularity and previous exam scores influence the student's academic achievement. Study hours are positively related to students' performance. An early assessment of students' achievement can help teachers make wise choices and raise academic success. The scholars can also put in a lot of effort to improve their academic performance and reach their goals in the upcoming semesters. Teachers can enhance the academic achievement of underperforming students by using the ML model in the classroom. The research supports the enhancement of education quality and skill development, aligning with the Sustainable Development Goals. Implementing this predictive approach allows educators to track student progress effectively and provide tailored academic guidance to optimize learning outcomes.

Future work: The design model can be strengthened by applying the ensemble method or Deep learning models and taking additional features, i.e., skill-based data, project work, certifications, and internships for the placement possibility.

Declaration:

DataAvailability

Publicly available data online on Kaggle database here:
https://www.kaggle.com/datasets/lainguyn123/student-performance-factors

FundingInformation

Not applicable.

AuthorContributions

Nisha (corresponding author): Conceptualization, methodology, data collection, analysis, and manuscript writing. Sanjay Kumar Sharma: Supervision and review

Acknowledgments

The author extends their thanks and regards to their teacher for being the constant pillar of strength, encouragement, and support throughout their journey.

ORCID ID

[1]Nisha Madaan**: https://orcid.org/0009-0006-3510-9788**

[2]Sanjay Kumar Sharma**: https://orcid.org/0000-0002-2546-8262**

6. References

1. Hassanbeigi, A., Askari, J., Nakhjavani, M., Shirkhoda, S., Barzegar, K., Mozayyan, M. R., & Fallahzadeh, H. (2011). The relationship between study skills and academic performance of university students. Procedia-Social and Behavioral Sciences, 30, 1416-1424.
2. Muraina, M. B., Nyorere, I., Emana, I. E., & Muraina, K. O. (2014). Impact of note taking and study habit on academic performance among selected secondary school students in Ibadan, Oyo State, Nigeria. International Journal of Education and Research, 2(6), 437-448.
3. Cerna, M. A., & Pavliushchenko, K. (2015). Influence of Study Habits on Academic Performance of International College Students in Shanghai. Higher Education Studies, 5(4), 42-55.
4. Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behaviour. *Memory*, *25*(6), 784-792.
5. Rabia, M., Mubarak, N., Tallat, H., & Nasir, W. (2017). A study on study habits and academic performance of students. International Journal of Asian Social Science, 7(10), 891-897.
6. Tus, J., Lubo, R., Rayo, F., & Cruz, M. A. (2020). THE LEARNERS'STUDY HABITS AND ITS RELATION ON THEIR ACADEMIC PERFORMANCE. *International Journal Of All Research Writings*, *2*(6), 1-19.

7. Rahman, M. M., Watanobe, Y., Kiran, R. U., Thang, T. C., & Paik, I. (2021). Impact of practical skills on academic performance: A data-driven analysis. *IEEE Access*, *9*, 139975-139993.
8. Al Husaini, Y. N. S., & Shukor, N. S. A. (2022). Factors affecting students' academic performance: A review. *RES MILITARIS*, *12*(6), 284-294.
9. Walck-Shannon, E. M., Rowell, S. F., & Frey, R. F. (2021). To what extent do study habits relate to performance?. *CBE—Life Sciences Education*, *20*(1), ar6.
10. Barcenas, J. M. A., & Bibon, M. B. (2022). The influence of study habits in the academic performance of senior high school students in Cagraray Island, Philippines. *Int. J. Sci. Res. in Multidisciplinary Studies Vol*, *8*(1).
11. Rubeena, Dr. (2019). STUDY HABITS AND ITS EFFECT ON THE ACADEMIC PERFORMANCE OF GRADE X TH STUDENTS. Review of Research in Education. 8. 1-5.
12. Orji, F. A., &Vassileva, J. (2022), Machine Learning Approach for Predicting Students Academic Performance and Study Strategies based on their Motivation. arXiv preprint arXiv: 2210.08186.
13. Bendangnuksung, P. P. (2018), Students' performance prediction using deep neural network. International Journal of Applied Engineering Research, 13(2), 1171-1176.
14. Altabrawee, Hussein & Ali, Osama & Qaisar, Samir. (2019). Predicting Students' Performance Using Machine Learning Techniques. JOURNAL OF UNIVERSITY OF BABYLON for pure and applied sciences. 27. 194-205. 10.29196/jubpas.v27i1.2108.
15. Gajwani, J., & Chakraborty, P. (2021). Students' performance prediction using feature selection and supervised machine learning algorithms. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Volume 1 (pp. 347-354). Springer Singapore.
16. Rai, S., Shastry, K. A., Pratap, S., Kishore, S., Mishra, P., & Sanjay, H. A. (2021). Machine learning approach for student academic performance prediction. In Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 1 (pp. 611-618). Springer Singapore.
17. Bujang, S.D., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E.E., Fujita, H., & Ghani, N.A. (2023). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. IEEE Access, 9, 95608-95621.Herrera-Viedma, E.E., Fujita, H., & Ghani, N.A. (2023). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. IEEE Access, 9, 95608-95621.
18. Kale, R. A., & Rawat, M. K. (2022, February). Machine learning-based improvement in quality of learning using 4Q & S approach. In 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 284-289). IEEE.
19. Talla Sai Likhitha et al,[8] "Student Academic Performance Prediction Using Machine Learning", 2022.International Journal of Creative Research Thoughts, ISSN:2320-2882
20. Abulhaija, S., Hattab, S., &Etaiwi, W. (2023, August). Predicting Students' Performance Using Machine Learning. In 2023 International Conference on Information Technology (ICIT) (pp. 470-475). IEEE
21. Yadav, Nitin & Deshmukh, Sonal. (2023). Prediction of Student Performance Using Machine Learning Techniques: A Review. 10.2991/978-94-6463-136-4_63.
22. Sri Lalitha Y et al,[11] "An Efficient Novel Approach on Machine Learning Paradigms for Analysis and Prediction of Academic Performance Based On Student Behaviour Approach", 2023, ECB, 10.48047/ecb/2023.12.si1o.00152
23. Dhilipan, J., Vijayalakshmi, N., Suriya, S., & Christopher, A. (2021, February). Prediction of student's performance using machine learning. In IOP conference series: Materials science and engineering (Vol. 1055, No. 1, p. 012122). IOP Publishing.
24. Verma, U., Garg, C., Bhushan, M., Samant, P., Kumar, A., & Negi, A. (2022, March). Prediction of students' academic performance using Machine Learning Techniques. In *2022 International Mobile and Embedded Technology Conference (MECON)* (pp. 151-156). IEEE.
25. Nedeva, V., & Pehlivanova, T. (2021). Students' performance analyses using machine learning algorithms in WEKA. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1031, No. 1, p. 012061). IOP Publishing.
26. Rastrollo-Guerrero, J.L.; Gómez-Pulido, J.A.; Durán-Domínguez, A. Analyzing and predicting students' performance by means of machine learning: A review. Appl. Sci. 2020, 10, 1042.

27. Chin, K. Y., Hong, Z. W., & Chen, Y. L. (2014). Impact of using an educational robot-based learning system on students' motivation in elementary education. IEEE Transactions on Learning Technologies, 7(4), 333–345.
28. Stefanescu, Daniela, Emilia Herman, and Maria-Ana Georgescu. 2009. Sustainable Development and Business Opportunities, Series: European Entrepreneurship Education. Targu Mures: Petru Maior University of Targu Mures Printing House
29. Baghdadi, Y., Harfouche, A., & Musso, M. (2020). ICT for an Inclusive World: Industry 4.0–Towards the Smart Enterprise, lecture notes in Information Systems and Organisation (LNISO) (p. 35). Springer
30. Harfouche, A., Merhi, M. I., Albizri, A., Dennehy, D., & Thatcher, J. B. (2024). Sustainable development through technological innovations and data analytics. *Information Systems Frontiers*, *26*(6), 1989-1996.
31. Junejo, N. U. R., Nawaz, M. W., Huang, Q., Dong, X., Wang, C., & Zheng, G. (2025). Accurate multi-category student performance forecasting at early stages of online education using neural networks. *Scientific Reports*, *15*(1), 16251.
32. Abdullah, M. W., Abu Agolah, R., Al Ajlouni, M., & Abdullah, M. (2025). Predicting Students' Online Exam Performance Using Machine Learning. In *Proceedings of the 16th International Conference on Information and Communication Systems (ICICS)*. IEEE.
33. Sandeepa, A. G. R., & Mohottala, S. (2025, February). Evaluation of Machine Learning Models in Student Academic Performance Prediction. In *2025 5th International Conference on Advanced Research in Computing (ICARC)* (pp. 1-6). IEEE.
34. Althaqafi, T., Saleem, F., & AL-Ghamdi, A. A. M. (2025). Enhancing student performance prediction: the role of class imbalance handling in machine learning models. *Discover Computing*, *28*(1), 79.
35. Chen, J., Zhou, X., Yao, J., & Tang, S. K. (2025). Application of machine learning in higher education to predict students' performance, learning engagement and self-efficacy: a systematic literature review. *Asian Education and Development Studies*, *14*(2), 205-240.
36. Moubayed, A., Injadat, M., Alhindawi, N., Samara, G., Abuasal, S., & Alazaidah, R. (2023, December). A deep learning approach towards student performance prediction in online courses: Challenges based on a global perspective. In *2023 24th International Arab Conference on Information Technology (ACIT)* (pp. 01-06). IEEE.
37. D. Boominath, S. Dhinakaran, "*Analysis of Students Performance Prediction Models Using Machine Learning Approaches*", International Journal of Computer Sciences and Engineering, Vol.12, Issue.9, pp.9-13, 2024.
38. N. Charuhasini, P. Drakshayani, P. Dhana Sri Aparna, P. Pravalika, Ch. Praneeth, "Analysing Privacy-Preserving Techniques in Machine Learning for Data Utility," *International Journal of Computer Sciences and Engineering*, Vol.13, Issue.2, pp.64-70, 2025.
39. Madaan, N., & Sharma, S. K. (2024, October). An Efficient Model for Academic Performance Prediction of the University Students. In *International Conference on Advancements in Smart Computing and Information Security* (pp. 149-163).