Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 2 Page No 217-222

Forecasting Academic Performance using Data Mining and Decision Support

Bhavna Sharma [1], Himanshu Vasnani [2], Sagar Damodar Padiya [3]

[1] Research Scholar, Faculty of Science, Suresh Gyan Vihar University, Jaipur

[2] Assistant Professor, Suresh Gyan Vihar University, Jaipur

[3] Associate Professor & Head, Shri Sant Gajanan Maharaj College of Engineering, Shegaon

Abstract- Education is undergoing a revolution that is made possible by the assistance of artificial intelligence (AI). This transition makes it possible to obtain a deeper understanding of student performance and to identify kids who are at risk of dropping out of school. Using patterns in students' academic records, attendance, online activity, and engagement levels, this study analyzes how learning analytics driven by artificial intelligence might be used to estimate student success and dropout rates. The ppaper will demonstrate how artificial intelligence can enable timely interventions to aid students in maintaining on track by analyzing data from massive open online courses (MOOCs), online learning platforms, and university databases. This will be accomplished through the evaluation of data. It is possible that these interventions will contain recommendations for adaptive learning and tailored feedback respectively. The goal is to improve student outcomes, reduce the number of students who drop out of school, and support educators in making decisions that are well-informed. This will be accomplished through the utilization of a trustworthy and ethical model of artificial intelligence. Additionally, in order to ensure that artificial intelligence is utilized in educational settings in a manner that is consistent with ethical standards, this research will take into consideration challenges such as the security of data, the elimination of bias, and the promotion of openness. At the end of the day, the objective of our research is to make education more tailored, more efficient, and accessible to all students.

Keywords: AI-powered learning analytics, student performance prediction, dropout rate analysis, machine learning in education, predictive analytics.

1. INTRODUCTION

A. Background

In recent years, educational institutions have produced extensive data concerning students' academic, behavioral, and demographic attributes via learning management systems, online tests, and administrative records. Nonetheless, a significant portion of this data is underexploited in conventional educational decision-making procedures, which predominantly depend on manual analysis, historical averages, and subjective assessment. Consequently, the early identification of kids susceptible to inadequate academic performance continues to pose a considerable difficulty. [6] Predicting academic success has gained significance due to escalating worries regarding student retention, graduation rates, and overall educational quality. Precise forecasting of student outcomes allows educators and administrators to implement timely interventions, customize learning trajectories, and enhance academic resources. Traditional assessment systems tend to be reactive, detecting performance deficiencies only after pupils have failed or underachieved. Data mining techniques provide robust instruments for revealing concealed patterns, correlations, and trends across extensive educational datasets. Techniques such as classification, regression, clustering, and association rule mining can be utilized to examine elements affecting academic achievement, including attendance, assessment scores, socio-economic status, learning behavior, and engagement levels. These strategies facilitate predictive modeling that underpins evidence-based educational planning. [7]

Decision Support Systems (DSS) augment the practical utilization of data mining methods by converting predictive insights into implementable suggestions. Integrating forecasting models with Decision Support Systems enables educational stakeholders to monitor student development in real time, assess different academic methodologies, and make educated decisions to enhance learning outcomes. These

Correspondence to: Bhavna Sharma, Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur
Corresponding author email: reply2bhavna21084@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 2 Page No 217-222

solutions facilitate the connection between data analysis and policy execution.

B. Literature Review

Recent studies have demonstrated the growing importance of data-driven approaches for forecasting academic performance. In 2024, Dalia Khairy et al. employed Random Forest and Decision Tree classifiers to predict the academic performance of first-year undergraduate students in the Computer Department over the period from 2016 to 2021. Their study aimed to identify the most effective algorithm for analyzing educational data to improve future student performance. The results showed a high level of accuracy, with only three misclassifications out of 253 testing instances, indicating the strong predictive capability of ensemble-based classification models. [1]

Similarly, in 2024, M. A. Arul Rozario and Dr. R. GunaSundari proposed a data mining–based framework for predicting the academic performance of first-year bachelor's students in computer science courses. Their work contributed valuable insights to the field of educational data mining, emphasizing the role of predictive analytics in enhancing student outcomes. [2]

Mohamed Bellaj et al. (2024) conducted a comprehensive evaluation of machine learning and data mining techniques applied in education, highlighting that machine learning models outperform traditional ensemble approaches. Their findings demonstrated the effectiveness of ML models in predicting student performance and supporting proactive academic decision-making through timely alerts. [3]

Esmael Ahmed (2024) addressed the challenges faced by modern educational institutions in analyzing student performance and developing evaluation strategies. His study showed that parameter optimization significantly improved the accuracy of four prediction models, while the Naïve Bayes classifier exhibited the lowest accuracy due to its assumption of feature independence. [4]

Extending this research direction, Athanasios Angeioplastis et al. (2025) investigated the application of educational data mining techniques to predict student performance in higher Consequently, predicting academic success using data mining and decision support systems constitutes a proactive and astute strategy for enhancing educational efficacy. It facilitates early warning systems, improves institutional accountability, and aids in data-informed educational administration, rendering it a vital domain of inquiry in contemporary educational analytics. [9].

In a broader technological context, Mehmooda et al. (2024) explored the role of edge computing in the post-cloud smart city era, emphasizing the need for improved technology integration, which indirectly highlights the importance of scalable and efficient computing infrastructures for implementing advanced educational analytics systems. [6]

C. Research Gaps

There have been a number of studies that have been carried out over the course of a number of years with the purpose of predicting the level of academic and behavioral success of children based on the data that was obtained from those pupils. The purpose of these studies was to make a prediction about the level of academic and behavioral achievement that children have achieved. The evaluation processes that are now in place are not particularly helpful when it comes to identifying children who are having issues at an early stage. Specifically, this is due to the fact that the evaluation mechanisms are not very effective. For the purpose of achieving the objective of enabling timely intervention, it is important to create a machine learning model that is capable of properly forecasting the performance of pupils. This stage of the procedure is quite important. It is necessary to fulfill a condition in order to use this particular type of model. Two examples of deep learning models that have been investigated in order to examine learning patterns are convolutional neural networks and recurrent neural networks. [7]

D. Research Objectives

1. To design and implement a machine learning-based predictive model for forecasting students' academic performance with high accuracy, enabling early identification of at-risk learners.
2. To develop a classification framework using deep learning algorithms to identify slow learners and fast learners, with the aim of supporting individual learner's capability-based teaching practices.
3. To design and implement a machine learning based predictive model for suggesting suitable career choice to the students.

E. Scope of the research

This study examines the utilization of data mining techniques and decision support systems to predict students' academic success in higher education institutions. The research entails the collection and analysis of past academic data, including internal assessment scores, attendance records, demographic information, and markers of learning behavior, to discern patterns that affect student success. Diverse data mining and

Correspondence to: Bhavna Sharma, Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur
Corresponding author email: reply2bhavna21084@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 2 Page No 217-222

machine learning strategies, including classification and regression models, are used to forecast pupils' academic performance at an early stage of their educational programs. The suggested approach amalgamates predictive models with a decision support framework to aid educators and academic administrators in making informed judgments. The decision support system offers early warning alarms, performance insights, and actionable recommendations to facilitate academic planning, individualized interventions, and resource allocation. The scope encompasses model evaluation through performance indicators like accuracy, precision, recall, and error rates to evaluate the efficacy of various methods. The study highlights the system's scalability and applicability for various courses and academic levels. Ethical aspects, including data privacy, security, and fairness, are addressed to guarantee the proper utilization of student data. The scope is confined to forecasting academic success and excludes psychological or medical evaluations of students. The study seeks to enhance proactive academic management via data-informed decision-making. [8]

II. MATERIALS AND METHODS

Collecting the data with the prescribed format so that all points related to objective will be covered, also we generate the link for collection of same data with the category of Graduation students

A. Data Collection and Preprocessing:

1. Academic Records: Historical grades (specific subjects, cumulative GPA), attendance data standardised test results (if applicable), course completion statistics, disciplinary history.
2. Demographic Data: Age, gender, socioeconomic situation (anonymized), familial background (e.g., parents educational attainment).
3. Behavioral Data: Involvement in extracurricular activities, interaction with online learning systems (e.g., login frequency, duration of participation, patterns of assignment submissions).
4. Psychometric Data (Optional but Advisable): Self-efficacy metrics, motivational indices, learning modalities (by surveys or questionnaires) [15-16].

B. Data Preparation:

1. Addressing Missing Values: Employing imputation techniques (mean, median, mode, K-nearest neighbors imputation) or eliminating rows/columns with substantial missing data.
2. Categorical Encoding: One-hot encoding for nominal variables (e.g., gender, course type) and ordinal encoding for ordinal variables (e.g., education levels).
3. Outlier Identification and Management: Detect and address outliers utilizing methods such as Z-score, Interquartile Range (IQR), or robust scaling.
4. Data Normalization/Standardization: Adjust numerical features to a uniform range (e.g., Min-Max scaling or StandardScaler) to ensure that features with greater magnitudes do not overshadow the model. [11]

C. Deployment and Surveillance of Models:

1. Real-time Prediction: Incorporate the model into an internet platform for immediate forecasts as fresh data is received.
2. Detailed Evaluation Metrics: Scores from individual assignments, quizzes, and tests (both formative and summative); scores on specific topics or learning objectives within a subject; project evaluations, laboratory reports, and presentations; standardized test scores (if accessible); pre-assessment scores for particular units or topics; grades from prior courses, semesters, or academic years (e.g., GPA); number of failed courses or attempts at specific courses; and academic progress over time (e.g., fluctuations in grades) [13].

D. Technical Description

Python is widely used as a primary programming language for implementing data mining and decision support systems due to its simplicity, flexibility, and extensive library support. Its easy-to-read syntax reduces development time and makes it suitable for rapid prototyping and experimentation, especially in data-driven applications. Python supports seamless integration of data preprocessing, model development, evaluation, and deployment within a single environment. [12]

For academic performance forecasting, Python provides powerful libraries such as Pandas and NumPy for data handling and preprocessing, Matplotlib and Seaborn for data visualization, and Scikit-learn for implementing data mining and machine learning algorithms such as Decision Trees, Random Forests, Naïve Bayes, Support Vector Machines, and Neural Networks. These libraries enable efficient analysis of large educational datasets and facilitate accurate predictive modelling.

Correspondence to: Bhavna Sharma, Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur
Corresponding author email: reply2bhavna21084@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 2 Page No 217-222

Python also supports the development of Decision Support Systems (DSS) through frameworks like Flask or Django, which allow predictive models to be integrated into interactive dashboards and web-based applications. This enables real-time monitoring, early warning alerts, and data-driven recommendations for educators and administrators. Furthermore, Python's strong community support, open-source nature, and cross-platform compatibility make it a cost-effective and scalable solution for educational analytics. Overall, Python serves as an effective and reliable programming language for implementing intelligent academic performance forecasting systems, supporting both analytical accuracy and practical decision-making. [8].

TABLE I
DESCRIPTION OF LIBRARIES AND ITS PURPOSE OF USE IN THE RESEARCH [9]

Purpose	Library	Description
Programming Language	Python	Core language used for data analysis, model development, and decision support system implementation
Data Collection & Handling	Pandas	Used to load, clean, preprocess, and manage student academic datasets efficiently
Numerical Computation	NumPy	Supports mathematical operations, array handling, and numerical processing
Data Visualization	Matplotlib	Generates graphs such as performance trends, distribution plots, and prediction results
Advanced Visualization	Seaborn	Provides statistical visualizations for better data interpretation
DM and ML	Scikit-learn	Implements algorithms such as Decision Tree, Random Forest, Naïve Bayes, SVM, and regression models
Model Evaluation	Scikit-learn	Used to evaluate models using accuracy, precision, recall, F1-score, and confusion matrix
Feature Selection	Scikit-learn	Helps identify significant academic factors influencing student performance
DSS	Flask / Django	Enables development of web-based DSS dashboards and interfaces
Data Storage	CSV / Excel	Stores historical student academic records and prediction results
Environment	VS Code	Used for coding, testing, visualization, and documentation

VI. RESULT AND DISCUSSION

In order to improve prediction accuracy, this research's future scope calls for integrating advanced deep learning techniques, such as hybrid models, recurrent neural networks, and artificial neural networks. Online tests, behavioral analytics for continuous performance tracking, and real-time data from learning management systems can all be included to the system. Future efforts might focus on developing adaptive decision support systems that independently recommend intervention strategies based on particular student characteristics. While including explainable AI techniques can increase transparency and confidence in prediction outcomes, expanding the model to more institutions and disciplines would improve generalizability. Additionally, cloud and edge computing technologies can be investigated to improve system deployment efficiency, scalability, and responsiveness. [14]

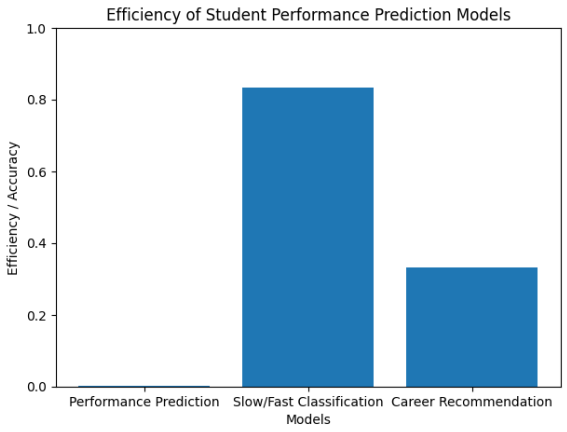

Fig. 1 Efficiency of Student performance

Figure 2 illustrates the efficiency comparison of the three proposed machine learning models. The slow and fast learner classification model achieved the highest accuracy of approximately 83%, indicating strong performance in

Correspondence to: Bhavna Sharma, Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur
Corresponding author email: reply2bhavna21084@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 2 Page No 217-222

identifying learner categories. The career recommendation model achieved moderate accuracy due to the complexity of multi-class prediction and limited dataset size. The academic performance prediction model showed low efficiency, highlighting the need for larger training data and improved feature selection. Overall, the results demonstrate the effectiveness of the proposed framework in learner classification and provide scope for further enhancement in performance prediction and career guidance.

VII. CONCLUSION

The proposed academic performance prediction model demonstrates the effectiveness of machine learning in early identification of at-risk students, enabling educational institutions to shift from reactive measures to proactive interventions that reduce dropout rates and improve overall academic outcomes. The deep learning–based classification of slow and fast learners provides detailed insights to support capability-based and personalized education, enhancing student engagement, understanding, and performance irrespective of learning pace. Additionally, the career recommendation model shows that integrating academic, psychometric, and interest-based data can generate accurate and personalized career guidance, helping students make informed decisions and improving workforce alignment. Overall, the study establishes a foundation for intelligent, equitable, and data-driven educational practices that benefit students, institutions, and the wider community.

REFERENCES

[1] Dalia Khairy, Nouf Alharbi, Mohamed A. Amasha, Marwa F. Areed, Salem Alkhalaf, Rania A. Abougalala. (2024). Prediction of student exam performance using data mining classification algorithms. *Education and Information Technologies*, 29:21621–21645

[2] M. AArul Rozario, Dr. R. GunaSundari. (2024). Predicting and Analysis of Students' Academic Performance using Hybrid Techniques. *International Journal of Intelligent Systems and Applications in Engineering IJISAE*, 12(3), 4189–4195

[3] Mohamed Bellaj, Ahmed Ben Dahmane, Said Boudra, Mohammed Lamarti Sefian. (2024). Educational Data Mining: Employing Machine Learning Techniques and Hyperparameter Optimization to Improve Students' Academic Performance. *International Journal of Online and Biomedical Engineering (iJOE)* iJOE, Vol. 20 No. 3

[4] Esmael Ahmed. (2024). Student Performance Prediction Using Machine Learning Algorithms. *Hindawi Applied Computational Intelligence and Soft Computing*, Volume, Article ID 4067721

[5] Athanasios Angeioplastis, John Aliprantis, Markos Konstantakis, Alkiviadis Tsimpiris. (2025). Predicting Student Performance and Enhancing Learning Outcomes: A Data-Driven Approach Using Educational Data Mining Techniques. *Computers*, 14, 83

[6] Mehmooda, H., Khalid, A., Kostakos, P., Gilman, E., & Pirttikangas, S. (2024). A novel edge architecture and solution for detecting concept drift in smart environments. *Future Generation Computer Systems,* 150, 127–143.

[7] Mahdia, O. A., Ali, N., Pardede, E., & Al-Quraishi, T. (2024). Online concept drift detector: Optimally balancing delay detection, runtime, memory, and accuracy. *Procedia Computer Science*, 237, 559–567.

[8] Yawen Chen, Jiande Sun, Jinhui Wang, Liang Zhao, Xinmin Song, Linbo Zhai. (2025). Machine Learning-Driven Student Performance Prediction for Enhancing Tiered Instruction. *arXiv:2502.03143v1* [cs.LG] 5 Feb.

[9] Swapnil Raj. (2019). An Overview on Python Programming. *JETIR February 2019, Volume 6, Issue 2 www.jetir.org ISSN-2349-5162.*

[10] Sazol Sarker, Mahit Kumar Paul, Sheikh Tasnimul Hasan Thasin, Md. Al Mehedi Hasan. (2024). Analyzing students' academic performance using educational data mining. *Computers and Education: Artificial Intelligence*, 7, 100263

[11] Md. Mahmudul Islam, Farhad Hossain Sojib, Md. Fazle Hasan Mihad, Mahmudul Hasan, Mahfujur Rahman. (2025). The integration of explainable AI in Educational Data Mining for student academic performance prediction and support system. *Telematics and Informatics Reports*, Volume 18, June, 100203

[12] Ahmed Kord, Ahmed Aboelfetouh, Samaa M. Shohieb. (2025). Academic course planning recommendation and students' performance prediction multimodal based on educational data mining techniques. *Journal of Computing in Higher Education,* Vol 2, 324

[13] Venera Nakhipova, Yerzhan Kerimbekov, Zhanat Umarova, Laura Suleimenova, Saule Botayeva, Almira Ibashova, and Nurlybek Zhumatayev. (2024). Use of the Naive Bayes Classifier Algorithm in Machine Learning for Student Performance Prediction. *International Journal of Information and Education Technology*, Vol. 14, No. 1.

[14] Arash Khosravi, Ahmad Azarnik. (2024). Leveraging Educational Data Mining: XGBoost and Random Forest for Predicting Student Achievement. *International*

Correspondence to: Bhavna Sharma, Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur
Corresponding author email: reply2bhavna21084@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 2 Page No 217-222

Journal of Data Science and Advanced Analytics (ISSN: 2563-4429).

[15] Munaf Salim Najim Al-Din and Hussein Ali Al Abdulqader. (2024). *Students' Academic Performance Prediction Using Educational Data Mining and Machine Learning: A Systematic Review. International Journal of Research and Innovation in Social Science (Ijriss)* ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VIII Issue VIII August, Page 1264

[16] Linqiang Tang, Chen Sian. (2025). Educational Data Mining For Student Performance Prediction. *Scalable Computing: Practice and Experience*, ISSN 1895-1767, SCPE. Volume 26, Issues 3, pp. 1551–1558

[17] Abdullah Al-Ameri, Aniello Castiglione, Michele Nappi, Chiara Pero, Muhammad Umer. (2024). *Student Academic Success Prediction Using Learning Management Multimedia Data With Convoluted Features and Ensemble Model. ACM J. Data Inform. Quality*, ACM 1936-1963/2024/8-ART

[18] Bello, H. O., Ige, A. B., & Ameyaw, M. N. (2024). Adaptive machine learning models: Concepts for real-time financial fraud prevention in dynamic environments *World Journal of Advanced Engineering Technology and Sciences*, 12(2), 21–34.

[19] Popoola, N. T. (2023). Big data-driven financial fraud detection and anomaly detection systems for regulatory compliance and market stability. *International Journal of Computer Applications Technology and Research*, 12(9), 32–46. https://doi.org/10.7753/IJCATR1209.1004

[20] Saleem Malik, S. Gopal Krishna Patro, Chandrakanta Mahanty, Rashmi Hegde, Quadri Noorulhasan Naveed, Ayodele Lasisi, Abdulrajak Buradi, Addisu Frinjo Emma, Naoufel Kraiem. (2025). Advancing educational data mining for enhanced student performance prediction: a fusion of feature selection algorithms and classification techniques with dynamic feature ensemble evolution. *Scientific Reports,* (2025) 15:8738

[21] Evelina Staneviciene, Daina Gudoniene, Vytenis Punys, Arturas Kukstys. (2024). A Case Study on the Data Mining-Based Prediction of Students' Performance for Effective and Sustainable E-Learning. *Sustainability*, 16, 10442

[22] Jedidi Y., Ibriz A., Benslimane M., Hachmoud A., Tmimi M., Hajjioui Y., Rahhali M. (2024). Predicting students' academic performance and modeling using data mining techniques. *Mathematical Modeling And Computing*, Vol. 11, No. 3, pp. 814–825

[23] Esmael Ahmed. (2024). Student Performance Prediction Using Machine Learning Algorithms. *Applied Computational Intelligence and Soft Computing Volume*, Article ID 4067721, 15 pages

[24] Murnawan, Sri Lestari, Rosalim Samihardjo, Deshinta Arrova Dewi. (2024). Sustainable Educational Data Mining Studies: Identifying Key Factors and Techniques for Predicting Student Academic Performance. *Journal of Applied Data Sciences*, Vol. 5, No. 3, September, pp. 1325-1342 ISSN 2723-6471 1325

[25] Furkan AYDIN. (2024). Prediction of Academic Success and Selection of Effective Features with Machine Learning. *BUEFAD,* Volume 14, Issue 1, 14(1), 181-195

[26] Şenay Kocakoyun Aydoğan, Turgut Pura, Fatih Bingül. (2024). Predicting Students' Academic Performances Using Machine Learning Algorithms in Educational Data Mining. *Malaysian Online Journal of Educational Technology*, (Volume 12 - Issue 4)

Correspondence to: Bhavna Sharma, Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur
Corresponding author email: reply2bhavna21084@gmail.com