VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2025, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 13, Number 1, January-March 2025 pp:111-120

Data-Driven Analysis of Student Performance Using Machine Learning:Insights from Demographics and Educational Records

Ashi Mehmood [1], **Bushra Mehmood** [1], **Muhammad Ammar** [2], **Haq Nawaz** [3], **Muhammad Latif** [2*], **Abdul Basit Abro** [4]

[1]Department of Computer Science, Iqra University Islamabad, Pakistan; [2]Department of Computer Science, Iqra University Karachi, Pakistan; [3]Department of Computer Science, National University of Sciences & Technology,Karachi, Pakistan; [4]Department of Computer System Engineering, Mehran University,Jamshoro,Pakistan

Keywords: *Machine Learning Student Performance Prediction Classification Algorithms Educational Data Mining Support Vector Machine*

Journal Info:
Submitted:
February 14, 2025
Accepted:
March 10, 2025
Published:
March 27, 2025

Abstract

The utilization of machine learning (ML) methods has the potential to address the challenges posed by the rapid growth of student-related data, enabling better predictions of student performance and supporting informed managerial decisions. These techniques analyze data through advanced models and algorithms to predict academic outcomes. This research identifies key factors that influence student performance using ML approaches. Using statistical and classification algorithms, machine learning improves the accuracy of predictions. The research explores relevant factors and applies them in state-of-the-art models to achieve precise performance predictions. Various studies have used ML techniques to predict student success, highlighting its broad applicability. This research proposes a framework for assessing students' academic achievements. The dataset includes information such as demographic details, previous academic records, and family history. Data was sourced from students across multiple universities using online surveys, comprising 24 attributes adapted from prior research. The objective is to identify critical attributes that significantly affect student performance. It also evaluates distinguish classification techniques to enhance prediction accuracy. Experimental findings reveal that the Support Vector Machine (SVM) outperforms other methods, achieving a maximum accuracy of 62.50 %.

***Correspondence author email address:** mlatif714@gmail.com
DOI: 10.21015/vtse.v13i1.2062

1 Introduction

Anticipating the outcomes of tasks performed by individuals has become a captivating topic, attracting significant attention in recent years. Predictive Analysis offers a systematic approach to examining past events and forecasting future outcomes. This approach integrates various methods, such as data mining, ML, statistical modeling, and artificial intelligence, to analyze data and generate predictions. By utilizing these techniques, student grade performance can be predicted with precision. Emerging fields like educational data analysis, learning analytics, academic analytics, educational data mining, predictive analytics, and learner analytics have become crucial research areas. Numerous personal, socioeconomic, psychological, and environmental factors shape students' academic performance. Accurate prediction models must account for these factors to deliver reliable results. High-accuracy predictions can help identify students requiring additional academic support.

This research investigates the critically analyzes existing models for predicting student performance, and determines the impact of various features on academic outcomes. The objectives of this research include identifying key attributes that enhance student performance, designing a more precise machine-learning algorithm, and validating the proposed model by benchmarking it against existing systems. The research question is whether it is possible to predict a student's performance in an academic setting. For instance, can one determine if a student will successfully complete their degree? [1]. However, learning is increasingly seen as an individual effort, making it challenging to develop models that evaluate a student's academic efforts effectively. Recent advancements in ML techniques have introduced fresh perspectives on this challenge.

Studies emphasize that success in life is closely tied to education, which plays a pivotal role in securing a brighter future [2]. Governments and educational institutions engage in various initiatives to enhance the quality of education nationwide. Education contributes to personal growth, improved social status, economic progress, global awareness, and solutions to environmental issues. ML, a subfield of artificial intelligence, has advanced the creation of algorithms that enable computers to analyze datasets and predict outcomes [3]. The application of these algorithms in education has grown rapidly in recent years. ML focuses on pattern recognition and decision-making, identifying patterns and generating rules from data inputs. It is a distinct field within computer science, differing from traditional computing methods in problem-solving [4]. Its algorithms are designed to allow systems to process data, create training sets, and deliver outputs based on statistical analysis. Supervised learning techniques, particularly classification algorithms, are widely used to predict academic success [5]. This research examined various classification methods to analyze and forecast final-year student grades. The algorithm's performance, combined with the dataset quality, determines the success of these ML models. Statistical methods for predicting student outcomes were carefully assessed in this study [6]. It was found that a student's academic success is influenced by their background and other relevant factors.

Research has consistently shown that personal characteristics, family background, and previous academic achievements significantly affect a student's performance. Predicting academic outcomes is especially critical in educational contexts, as it facilitates the creation of strategies to improve academic results and reduce dropout rates [7]. ML techniques are frequently applied to predict final grades by analyzing data from similar students' past performances [8]. Information such as age, gender, and field of study is often readily available in university databases, making it a common resource for researchers. Several techniques have been explored to enhance student's academic outcomes and promote their holistic development. The study also evaluated student's behavior to anticipate academic challenges and suggested personalized interventions to improve their learning experience [9]. Accurate predictions of student performance are highly beneficial, as they enable educators to identify at-risk students early and provide targeted support [10]. While predicting academic success is critical for fostering growth, it remains complex due to the interplay of multiple influencing factors.

2 RELATED WORK

ML encompasses a range of techniques that allow computers to learn and make decisions autonomously, without direct human intervention. This technology has been widely applied across various domains, including medical diagnostics, stock market predictions, DNA sequence classification, gaming, robotics, and predictive analysis, among others. In the educational sector, machine learning offers opportunities to enhance understanding of the learning process by identifying, extracting, and analyzing factors that influence students' educational journeys. Predicting student performance involves identifying key attributes and examining the relationships between these features and the factors that impact academic outcomes. This study evaluates whether learners can complete specific tasks or achieve defined learning objectives. In recent years, significant research efforts have focused on developing models to predict student performance, emphasizing its importance in improving educational outcomes. A framework was developed to predict the academic performance of first-year bachelor students in computer science courses, emphasizing the identification of critical factors contributing to success [11]. The dataset, spanning eight years from July 2003/2004 to July 2013/2014, included information such as students' academic records, family backgrounds, and demographics. Classification methods like Decision Tree, Naïve Bayes, and Rule-Based (RB) models were applied to this data to identify the most effective prediction model. The finding concluded that the Rule-Based model achieved the highest accuracy compared to the others.

This model was particularly effective for students with poor or average grades, helping teachers identify and support students needing academic assistance. Clustering analysis was introduced to study student behavior, yielding insightful findings. Data mining algorithms have proven instrumental in understanding and consistently improving student performance [12]. These analyses play a crucial role in the admissions and placement processes, leveraging parameters such as projects, internships, skill sets, and academic marks from 10th, 12th, and undergraduate programs. The K-means algorithm was used for clustering due to its simplicity and computational efficiency, with additional clustering strategies proposed to further enhance performance. Ranking or classification within clusters was also suggested to better evaluate student performance.

The increasing volume of information in academic databases has been noted, and classification algorithms have been applied to educational datasets to extract insights into student performance [13]. A prediction model based on K-Nearest Neighbor (KNN) and Naïve Bayes algorithms was proposed to evaluate students' academic outcomes. These algorithms were compared using achievement parameters, with Naïve Bayes achieving an accuracy rate of 93.17%. The study [13] identified a strong relationship between significant features and student performance. Another analysis used various analytical methods to assess student achievements, abilities, and development [14]. The Student Attribute Matrix (SAM) was employed to model performance-related and non-performance-related attributes. BP-NN algorithms were utilized to create a performance estimation tool using current student data, such as achievements and features. Data from 60 high schools were analyzed, providing indicators and predictors to understand factors influencing student outcomes. The evaluation showed strong results, offering valuable insights into the educational process.

Data mining techniques have also been applied to analyze undergraduate students' performance and reveal patterns [15]. Academic achievements were assessed after completing a four-year program, categorizing students into low- and high-achieving groups. This research highlighted the importance of focusing on fewer courses to improve academic performance and provided early warnings for students at risk of underachievement. A comparative study of linear regression and multilayer perceptron algorithms was conducted to predict student outcomes [16]. Based on mean values and absolute error, the multilayer perceptron outperformed linear regression. The inclusion of online discussion forums was shown to enhance students' learning experiences.

Categorization techniques were used to analyze students' knowledge, applying past performance data to predict semester outcomes [17]. Various classification methods were implemented, with localized theorems playing a key role. The study enabled educators to identify students requiring additional support and reduce failure rates through targeted interventions. The research also informed strategies for improving academic outcomes in subsequent semesters.

A system to predict secondary school grades employed data mining methods, including binary, regression, and five-level classification techniques, along with models such as decision trees, random forests, neural networks, and support vector machines (SVM) [18]. The findings indicated that predictive accuracy was highest when first- and second-term grades were available. The study confirmed that earlier academic performance strongly influences future success. The use of data mining techniques also provided insights into additional attributes, such as prior academic records and feedback from educational institutions.

3 DETAILS OF PROPOSED METHOD

Predicting a student's academic performance requires a model that incorporates all relevant features affecting their outcomes. The base line model used for this study is inaccurate. as shown in Figure 1.

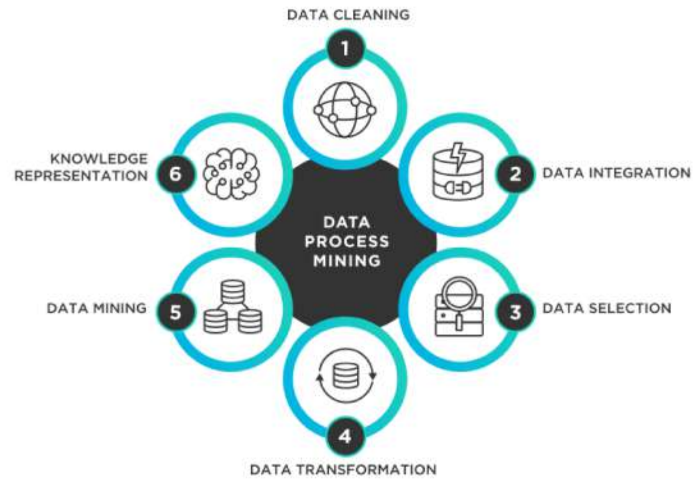

Figure 1. Baseline Model.

This study employs a quantitative research methodology, analyzing numerical data collected via surveys distributed to students across various universities. A pilot study validated the questionnaire before full deployment, ensuring clarity and relevance. Data preprocessing involved cleaning missing values through mean and mode imputation, encoding categorical variables, and normalizing numerical data. Feature selection identified key academic and socio-economic factors influencing student performance, with a target classification of high, medium, and low achievers. Machine learning models such as Decision Trees, Random Forest, and Support Vector Machines (SVM) were trained and tested using an 80-20 data split. To establish a performance benchmark, baseline methods such as Logistic Regression, K-Nearest Neighbors (KNN), Naïve Bayes, Linear Regression, and Majority Class Baseline were implemented for comparison. Educational data mining techniques were applied to extract patterns, highlighting geographic disparities in academic performance, particularly in a local context compared to existing datasets from Europe and America. Data analysis utilized statistical methods and visualizations through tables, graphs, and figures to illustrate key trends. The structured workflow involved survey distribution, data collection, preprocessing, feature selection, model training and testing, and results interpretation, ensuring a methodical approach to deriving meaningful insights, as outlined in Figure 3.

4 EXPERIMENTAL WORK

This research aims to evaluate the effectiveness of popular machine learning algorithms for early detection of students likely to underperform and to assess the impact of process mining variables on the performance of these techniques. The study employs classification algorithms suitable for datasets with imbalances. Specifically, Random Forest (RF), Logistic Regression (LR), Naïve Bayes, and K-Nearest Neighbors (KNN) were utilized in the experiments.

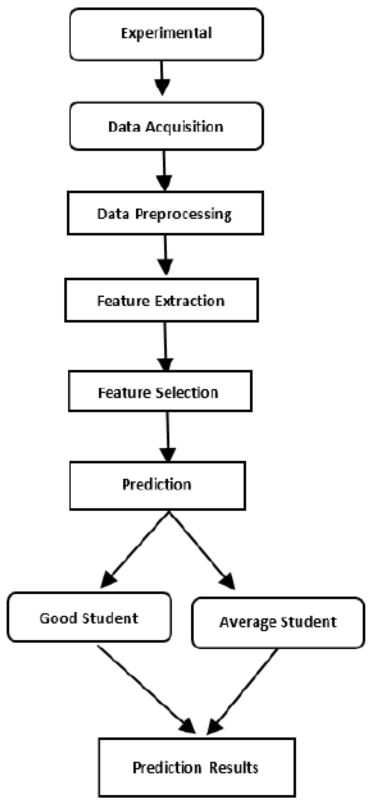

Figure 2. Dataset processing workflow diagram.

4.1 Data Collection

This study employed a structured survey methodology to gather student data from multiple universities. The dataset includes demographic details, academic records, and socio-economic factors, with 24 attributes identified as potential predictors of student performance. Data was collected through online surveys and preprocessed to handle inconsistencies such as missing values and noisy data.

4.2 Sample Size

The sample size reflects the number of completed responses from the survey, representing the target population. A random sample of students from three universities was chosen, initially targeting 450 students. After applying preprocessing techniques, 318 valid student responses remained, each characterized by 24 unique attributes. This sample size aligns with previous studies that used similar datasets. The decision to use this sample size was based on its successful application in four to five comparable studies.

4.3 Data Preprocessing

Data preprocessing is a critical step in machine learning, as it transforms raw data into a format suitable for analysis. The dataset collected from students contained various inconsistencies, such as missing values and noisy data. The preprocessing phase addressed these issues to enhance data quality. Techniques such as handling missing values, correcting errors, and filtering noisy data were implemented to ensure the dataset's reliability. This step also eliminated any anomalies that could obscure patterns in the analysis.

4.4 Feature Extraction

Feature extraction involves reducing the number of features in a dataset by creating new ones from existing attributes and discarding the originals. This technique ensures that the dataset retains meaningful information while reducing complexity. In this study, the extracted features effectively summarized the original dataset, enabling a more concise analysis.

4.5 Feature Selection

Although the dataset does not include an excessive number of features, some attributes were deemed irrelevant to predicting student achievement. The WEKA tool was employed for feature selection, offering automated algorithms to identify and remove unnecessary features. Feature selection refines the dataset by eliminating irrelevant attributes, enhancing the accuracy and efficiency of the predictive models. This process also ensured that the dataset included only significant variables, such as demographics, socioeconomics, and course-related factors.

4.6 Feature Engineering

Feature engineering was conducted to optimize the dataset for machine learning models. Categorical variables such as gender and parental occupation were encoded, while numerical attributes were normalized to ensure uniformity. The WEKA tool was used for feature selection, removing irrelevant or redundant features to enhance model performance.

4.7 Data Description

The dataset included 24 attributes categorized into demographic, learning environment, socioeconomics, and course-related features. Examples include student gender, living address, family size, parental job status, daily travel time, study hours, internet connectivity, co-curricular activities, health status, and teacher encouragement. The study utilized the 10-fold cross-validation technique to evaluate the generalizability of the models. This method splits the dataset into ten equal parts, iteratively training on nine parts and testing on the tenth, ensuring robust validation.

4.8 Model Selection and Hyperparameter Tuning

Classification is a key component of supervised machine learning. This study used four algorithms to predict student performance: Naïve Bayes, J48 Decision Tree, Random Forest, and Support Vector Machine (SVM). These algorithms were chosen based on their prevalence in existing research. The inclusion criteria for features considered geographical location, social culture, and alignment with study objectives. Features irrelevant to the context or lacking predictive significance were excluded. This approach ensures the dataset aligns with the research goals, enhancing the reliability of the results. This section provides a comprehensive overview of the data gathered from students, the results obtained, and the methodology used to address the research questions. After analyzing the dataset, the outcomes were reproduced. WEKA (Waikato Environment for Knowledge Analysis), an open-source software, was utilized for this study. WEKA offers a collection of machine learning algorithms designed specifically for tasks such as data preprocessing, classification, regression, visualization, and other data mining applications. Many researchers rely on WEKA for implementing prediction algorithms. Various machine-learning classification techniques were tested using WEKA for data mining purposes. The classifiers in WEKA were used to predict different levels of accuracy.

Four classification algorithms were employed, which are Naive Bayes, Decision Tree, Random Forest, and Support Vector Machine. Hyperparameter tuning was performed using grid search to optimize parameters such as the number of trees for Random Forest, Kernel type for SVM (linear, polynomial, RBF), Minimum instances per leaf for J48 Table 1 outlines the accuracy of the features in the dataset.

Table 1. Attributes with Their Accuracy

Attribute	Accuracy %
Gender	62.5
Previous Degree	68.75
Level of Study	68.75
Current Semester	68.75
Address	46.875
Family Size	62.5
Order in Family	56.25
Parents' Relationship Status	62.5
Occupation of Father	40.625
Occupation of Mother	65.62
Guardian	68.75
Travel Time	40.625
Weekly Study Hours	50
Extracurricular Activities	62.5
Internet Facility	40.625
Go Outing with Friends	40.625
Educational Expense	62.5

Table 1 presents the accuracy of individual attributes using the classifier. Four classifiers were tested to determine accuracy, with the best-performing classifier providing the final results. The research dataset yielded these accuracies based on the target class "CGPA." Accuracy for each attribute was evaluated against the target class, though some features did not produce significant results as anticipated. For instance, the feature "Father's Occupation" demonstrated lower-than-expected accuracy due to specific factors, such as the distribution of accuracy across three trials, which impacted the results. Certain attributes in the dataset also displayed minimal accuracy. Among the classifiers, SVM achieved the highest accuracy at 62.5%. The confusion matrix for the SVM classifier is shown below.

Table 2. Values of SVM Classifier

Parameter	Value %
Health Status	50
Encourage by Teacher	62.5
Marks Increase/Decrease	56.25
Suitable System	62.5
Coursework Provides Sufficient Information	43.75

4.8.1 Model Evaluation and Confusion Matrix

The models were assessed based on multiple performance metrics such as accuracy [19] which measures correct predictions across all instances[20].The precision [21] evaluates the proportion of correctly predicted positive instances[22]. recall [23] determines the proportion of actual positives correctly identified [24] and F1-score [25] harmonic mean of [26] precision and recall [27].

The confusion matrix of an SVM classifier shows the sensitivity, specificity, and false positive values. Here are the values of the SVM classifier confusion matrix.

The confusion matrix is generated by the WEKA tool. The confusion matrix gives you a brief overview of the model's performance and accuracy level. The matrix contains detailed information about the model's predictions. The confusion matrix table, which comprises information on actual and expected classifications, is then generated. For the pass and fail categories, the prediction is correct, as seen in the matrix. The cross-tabulation of true labels from the labelled dataset and predicted labels from a classifier is called a confusion matrix. A confusion matrix was generated for each classifier. Below is the confusion matrix for the best-performing model (SVM).

From this, we calculated:

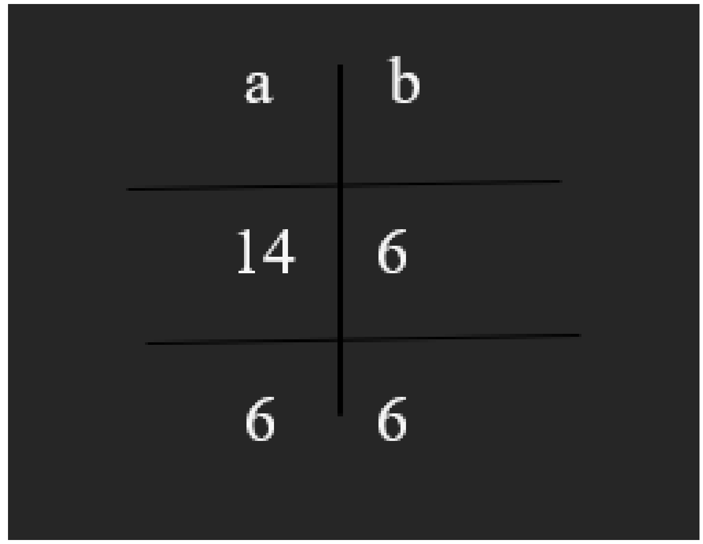

Figure 3. Confusion Matrix of SVM Classifier

- **Precision** = 6 / (6+6) = 50%
- **Recall** = 6 / (6+6) = 50%
- **F1-score** = 50%

Table 3. Confusion Matrix with parameters and theirs respective values

Parameter	Value
Total Samples (N)	32
Predicted No	
True Negative (TN)	14
False Negative (FN)	6
Predicted Yes	
False Positive (FP)	6
True Positive (TP)	6

Table 4 highlights the accuracies of the proposed and existing studies using four classifiers commonly employed in prior research. These classifiers were selected as they are frequently utilized in various existing studies to determine accuracy [28]. The table demonstrates that while both studies used the same classifiers, they yielded different results. The existing study achieved a maximum accuracy of 64.10%, while the proposed study's highest accuracy was 62.5% with the SVM classifier. The accuracy of the proposed study closely aligns with that of the existing study[29]. The

Table 4. Comparison of Proposed Study with Existing Study

Parameter	Value
Student's CGPA	
Classifier	Naive Bayes
Proposed Study Accuracy	53.125%
Existing Study Accuracy	55.3%
Students Final Grade G3	
Classifier	J48
Proposed Study Accuracy	56.25%
Existing Study Accuracy	58.97%
Classifier	Random Forest
Proposed Study Accuracy	59.37%
Existing Study Accuracy	61.5%
Classifier	Support Vector Machine
Proposed Study Accuracy	62.5%
Existing Study Accuracy	64.10%

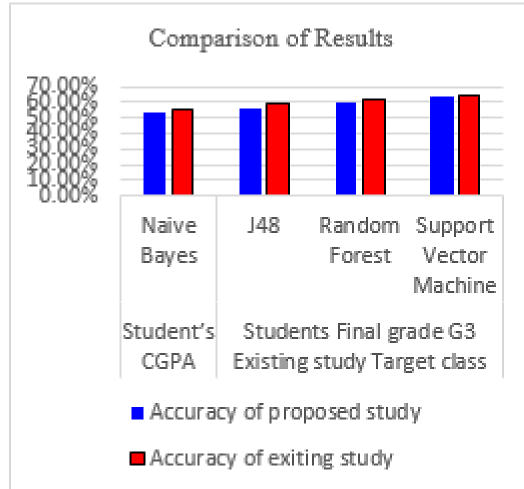

Figure 4. Comparison of results of existing and proposed studies

figure below provides a graphical comparison of the accuracies from both studies.

Figure 2 illustrates the accuracies of the proposed study dataset compared to the existing study, both of which utilized different classifiers. To ensure accuracy, this research employed four classifiers: Naive Bayes, J48, Support Vector Machine (SVM), and Random Forest, each producing varying levels of accuracy. In the bar chart, blue represents the accuracy of the proposed study, while red represents the accuracy of the existing research. The chart demonstrates that the SVM classifier achieved the highest accuracy in both studies [30]. The highest accuracy in the existing study was greater than that of the proposed study, which reached 62.5%.

5 Results

The dataset was analyzed using four machine learning classifiers: Naive Bayes, Random Forest, Support Vector Machine (SVM), and J48. The results of our dataset were compared with an existing dataset using these classifiers. Data collected from university-level students underwent preprocessing techniques such as data cleaning and handling missing values. The SVM classifier achieved the highest accuracy on our dataset, outperforming the other classifiers. Some features from the existing dataset were selected because they were utilized in 4 to 5 research papers, while others were excluded due to demographic differences. Modifying input data and selecting suitable methods were crucial factors in analyzing student performance. Applying various approaches to student data revealed that SVM consistently outperformed other methods in predicting student performance.

6 Conclusion

The study found that Support Vector Machine (SVM) achieved the highest accuracy in predicting student performance. Proper data preprocessing and careful feature selection significantly impacted the results. The comparison with an existing dataset further emphasized the importance of choosing relevant attributes. Overall, SVM proved to be the most effective model, highlighting the role of data handling in educational analysis.

7 Future Work

In the future, advanced machine learning algorithms, such as deep learning and Reinforcement Learning, could be utilized to enhance the prediction of student performance. The research scope can be expanded to include college-level students, incorporating additional features and a larger sample size for more robust analysis.

Author Contributions

Ashi Mehmood:Conceptualization, Methodology, Software.**Bushra Mehmood**:Data curation, Writing- Original draft preparation.**Muhammad Ammar**: Investigation, Validation.**Haq Nawaz**:Visualization**Muhammad Latif**: Supervision **Abdul Basit Abro**: Reviewing and Editing.

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] G. Feng and M. Fan, "Research on learning behavior patterns from the perspective of educational data mining: Evaluation, prediction and visualization," *Expert Systems with Applications*, vol. 237, p. 121555, 2024.

[2] G. Al-Tameemi, J. Xue, I. H. Ali, and S. Ajit, "A Hybrid Machine Learning Approach for Predicting Student Performance Using Multi-class Educational Datasets," *Procedia Computer Science*, vol. 238, pp. 888-895, 2024.

[3] M. Shoaib, N. Sayed, J. Singh, J. Shafi, S. Khan, and F. Ali, "AI student success predictor: Enhancing personalized learning in campus management systems," *Computers in Human Behavior*, vol. 158, p. 108301, 2024.

[4] H. Pallathadka, A. Wenda, E. Ramirez-Asís, M. Asís-López, J. Flores-Albornoz, and K. Phasinam, "Classification and prediction of student performance data using various machine learning algorithms," *Materials today: proceedings*, vol. 80, pp. 3782-3785, 2023.

[5] M. Nachouki, E. A. Mohamed, M. R. Mehdi, and M. Abou Naaj, "Student course grade prediction using the random forest algorithm: Analysis of predictors' importance," *Trends in Neuroscience and Education*, p. 100214, 2023.

[6] S. Hussain and M. Q. Khan, "Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning," *Annals of data science*, vol. 10, no. 3, pp. 637-655, 2023.

[7] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H. Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Education and Information Technologies*, vol. 28, no. 1, pp. 905-971, 2023.

[8] I. Issah, O. Appiah, P. Appiahene, and F. Inusah, "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance," *Decision analytics journal*, vol. 7, p. 100204, 2023.

[9] A. Quílez-Robres, A. González-Andrade, Z. Ortega, and S. Santiago-Ramajo, "Intelligence quotient, short-term memory and study habits as academic achievement predictors of elementary school: A follow-up study," *Studies in Educational Evaluation*, vol. 70, p. 101020, 2021.

[10] R. Bertolini, S. J. Finch, and R. H. Nehm, "Quantifying variability in predictions of student performance: Examining the impact of bootstrap resampling in data pipelines," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100067, 2022.

[11] S. A. Alwarthan, N. Aslam, and I. U. Khan, "Predicting student academic performance at higher education using data mining: A systematic review," *Applied Computational Intelligence and Soft Computing*, vol. 2022, no. 1, p. 8924028, 2022.

[12] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, p. 11, 2022.

[13] A. Anwarudin, W. Andriyani, B. P. DP, and D. Kristomo, "The Prediction on the students' graduation timeliness using naive bayes classification and k-nearest neighbor," *Journal of Intelligent Software Systems*, vol. 1, no. 1, pp. 75-88, 2022.

[14] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, "Educational data mining techniques for student performance prediction: method review and comparison analysis," *Frontiers in psychology*, vol. 12, p. 698490, 2021.

[15] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H. Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Education and Information Technologies*, vol. 28, no. 1, pp. 905-971, 2023.

[16] S. Ranjeeth, T. P. Latchoumi, and P. V. Paul, "Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model," *Recent Advances in Computer Science and Communications*, vol. 14, no. 6, pp. 1728-1741, 2021.

[17] G. Feng, M. Fan, and Y. Chen, "Analysis and prediction of students' academic performance based on educational data mining," *IEEE Access*, vol. 10, pp. 19558-19571, 2022.

[18] M. M. Arcinas, G. S. Sajja, S. Asif, S. Gour, E. Okoronkwo, and M. Naved, "Role of data mining in education for improving students performance for social change," *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32, no. 3, pp. 6519-6526, 2021.

[19] Abro, A. A., Taşcı, E., & Ugur, A. (2020). A stacking-based ensemble learning method for outlier detection. *Balkan Journal of Electrical and Computer Engineering*, *8*(2), 181-185.

[20] Abro, A. A., Khan, A. A., Talpur, M. S. H., Kayijuka, I., Yaşar, E. (2021). Machine learning classifiers: a brief primer. University of Sindh Journal of Information and Communication Technology, 5(2), 63-68.

[21] Abro, A. A., Sıddıque, W. A., Talpur, M. S. H., Jumani, A. K., Yaşar, E. (2023). A combined approach of base and meta learners for hybrid system. Turkish Journal of Engineering, 7(1), 25-32.

[22] Abro, A. A., Yimer, M. A., & Bhatti, Z. (2020). Identifying the machine learning techniques for classification of target datasets. *Sukkur IBA Journal of Computing and Mathematical Sciences*, *4*(1), 45-52.

[23] Abro, A. A., Talpur, M. S. H., Jumani, A. K., Sıddıque, W. A., Yaşar, E. (2021). Voting combinations-based ensemble: A hybrid approach. Celal Bayar University Journal of Science, 18(3), 257-263.

[24] Abro, A. A., Abro, A. B., Abro, M., Siddique, A. A. (2018). Model Valuation of MPLS utilization Physical and virtual Network on GNS3. University of Sindh Journal of Information and Commmunication Technology (USJICT), 2(2).

[25] Latif, M., Mehmood, B., Khalid, N., Ahmed, S., Abro, A. B., Ahmed, M. (2024). Protection Issues and Challenges within the Cloud: A Survey. VFAST Transactions on Software Engineering, 12(4), 167-179.

[26] Daniyal, S. M., Abbasi, M. M., Hussain, D., Amjad, U., Abro, A. B., Naeem, M. (2024). A hybrid approach for simultaneous effective automobile navigation with DE and PSO. VAWKUM Transactions on Computer Sciences, 12(2), 01-15.

[27] Moiz, S., Abro, A. A., Ebrahim, M., Abro, A. B. (2024, January). Unveiling the arsenal of user data protection tools and practices. In 2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC) (pp. 1-7). IEEE.

[28] Moiz, S., Abro, A. A., Ebrahim, M., Abro, A. B. (2024, January). Unveiling the arsenal of user data protection tools and practices. In 2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC) (pp. 1-7). IEEE.

[29] Abro, A. A. (2021). Vote-based: Ensemble approach. Sakarya University Journal of Science, 25(3), 858-866.

[30] Latif, M., Ebrahim, M., Abro, A. S., Ahmed, M., Abbasi, M. D., Tunio, I. A. (2024). Face recognition from video by matching images using deep learning-based models. VAWKUM Transactions on Computer Sciences, 12(2), 50-64.