VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

Study of educational achievement based on examining machine learning techniques: Systematic Literature

Kshiteeja Mhatre[1], Hiren Dand[2]
[1](Research Scholar, Department of Computer Science, Shri JJT University, Rajasthan India)
[2](Associate Professor, Department of Computer Science, Shri JJT University, Rajasthan India)

Abstract : *Students, courses, and academic activities produce a lot of data for educational institutions. The amount, diversity, and complexity of this data make effective analysis difficult. Learners may be categorized according to their skills and abilities, learning behavior can be understood, and academic performance can be predicted using sophisticated data analysis tools. In order to solve these issues, this study uses the machine learning technique to analyze 15 relevant research papers published between 2019 and 2025 as part of a systematic literature review on machine learning technology for forecasting student performance. Examine machine learning techniques based on classification, data mining, ensemble approaches, k-nearest neighbor, and K-means. According to the survey, the most popular method for forecasting student academic performance also yielded the best accuracy findings, ranging from 70% to 98.6%. The study emphasizes how important it is to identify students who are at risk of learning challenges early on so that educators and policymakers may offer timely and appropriate support. It also highlights a number of difficulties, such as inconsistent performance metrics, restricted model application in different situations, skewed datasets, and ethical challenges, underscoring the need for more reliable research techniques and well-rounded prediction strategies in educational analytics.*

Keywords - *academic performance, classification, data mining, ensemble approaches, K-means, k-nearest neighbor, literature review, machine learning, student.*

INTRODUCTION

In order to analyze vast amounts of academic and behavioral data produced by students and institutions, machine learning is being used more and more in the education sector. It aids in analyzing individual learning patterns, identifying students who could be at danger, and forecasting student success. Tutors can enhance learning results by tailoring their lesson plans and offering prompt assistance through the use of clever algorithms. Despite its advantages, implementation requires careful consideration of issues including bias, data privacy, and model reliability.

Due to its direct impact on students' future prospects and institutional success, student dropout is still a significant concern in educational institutions. Finding worthy students for scholarships is equally crucial to helping those who are struggling financially and promoting academic continuity. Students that perform better academically overall acquire greater knowledge and abilities, which improves learning outcomes. Effective teaching methods are also essential for improving students' employability and placement prospects. In today's environment, education plays a crucial role in both career advancement and social mobility. High-quality education is essential for long-term success and financial security in a knowledge-based economy. An essential basis for comprehending educational progress is provided by student academic records, learning behavior, and psychometric data. To facilitate efficient decision-making, a substantial amount of this data—which is gathered from students and their educational activities—must be precisely examined. Socio-

demographic and study-environment variables, in addition to academic data, are crucial in explaining variations in student performance. These data systems are also used by educational institutions to methodically track students, instructors, and courses. Organized academic administration is further aided by information about academic circulars, exam papers, curricula, and faculty profiles. When combined, these many data sources assist institutions in enhancing overall educational quality, administrative effectiveness, and learning outcomes. [1]

Students are the essential building elements of a country and have a significant influence on its destiny. However, a lot of students deal with issues like anxiousness, hopelessness, and trouble controlling their mental health. A thorough examination of machine learning methods can aid in comprehending these problems and improving student support.

By employing academic insights to spot possible problems before they have an impact on student performance, educational institutions can take preventative action. Strong faculty competence and insightful teacher evaluation are essential for enhancing the teaching-learning process. Data-driven evidence can be used by academic policymakers to create and assess successful policies that improve the general quality of education. Students can better grasp which subjects and abilities are crucial for their future academic and professional aspirations by following clear and organized criteria. Additionally, educators can better support learning activities and promote better student outcomes when they are equipped with knowledge and analytical insights. [2]

In both research and industry, machine learning has grown quickly. It now includes both sophisticated deep learning models that can learn complicated representations and more conventional techniques like support vector machines and k-nearest neighbors (K-NN) [3]. When compared to single models, ensemble approaches like Random Forests, bagging, and boosting combine several classifiers to increase prediction accuracy, stability, and minimize overfitting [3, 4].Classical machine learning techniques are still prized for their interpretability and reduced computing cost, even when deep learning delivers great performance in representation learning [3]. According to comparative research, performance is dependent on the features of the data and no algorithm consistently outperforms others across all datasets [3, 4]. As a result, ensemble-based methods are frequently chosen to produce solid and trustworthy classification outcomes [4].

II Literature Review

Walia et al. (2020) [5] used dataset including 395 records with 33 attributes was subjected to a variety of data mining and machine learning classification approaches, including Naive Bayes, Decision Tree, Random Forest, JRip, and ZeroR. The findings demonstrated that Decision Tree-based models obtained accuracy above 80%, suggesting that these machine learning techniques are useful for detecting potentially vulnerable students utilizing academic data from the school. Yağcı et al. (2022) [6] study examines the academic records of 1,854 students using a variety of machine learning and educational data mining approaches, such as Random Forest, k-nearest neighbor, support vector machines, logistic regression, and Naïve Bayes. The findings demonstrate these models' efficacy in predicting student performance and identifying students who could need early academic support, with prediction accuracies of roughly 70–75%. Chen et al. (2023) [7] study Machine learning, deep learning, and ensemble models like SVM, Random Forest, and Decision Trees achieve high predictive performance, achieving 93.1% accuracy and a 90.45% F1-score, while effectively supporting early identification of at-risk students for timely academic intervention, according to a study done on university information school data from 2017–2018. Verma et al. (2022) [8] used Classifiers like Decision Tree, Naïve Bayes, k-Nearest Neighbor, Support Vector Machine, and the ensemble method Random Forest were assessed using data mining and machine learning techniques on academic records of institutional students. Naïve Bayes performed best, with an accuracy of roughly 89%, while overall accuracy across models ranged from 60% to 95%. Yıldız et al. (2020) [9] study compared k-NN, Logistic Regression, SVM, Decision Tree, Random Forest, Naive Bayes, and Neural Network models using data mining techniques on data gathered from 421 ninth-grade students. It found that the Neural Network model achieved the highest classification accuracy (98.6%), while parental education, family income, students' learning beliefs, and study habits significantly influenced academic success. Nafea et al. (2023) [10] examined the OULAD dataset of 32,593 students gathered by the Open University (UK) using data mining and machine learning techniques like Random Forest, Decision Tree, AdaBoost, and Support Vector Machine. Accuracy levels ranged from roughly 78% to 96%, where ensemble methods showed superior and more consistent predictive performance by successfully combining the strengths of individual models. Begum et al. (2022) [11] study shows that a variety of machine learning classifiers, especially ensemble-based methods, are successful in accurately forecasting students' academic performance, allowing educational institutions to put timely

intervention tactics into place. Prahmana et al. (2025) [12] used 59 student academic and behavioral datasets to compare several machine learning models, including K-Nearest Neighbors (KNN), KNN with NGO (KNNG), KNN with TDO (KNTD), Naive Bayes Classification (NBC), and various regression and classification methods, finding that ensemble and tree-based models achieved higher accuracy (around 70%–90%) and were more effective for early identification of at-risk students.Yang et al. (2022) [13] evaluated on two public student learning datasets, the suggested Adaptive K-Means with Deep Neural Network model showed better prediction performance than benchmark approaches and, through feature selection, revealed important influencing elements including mother's education and prior failure history. Sekeroglu et al. (2019) [14] found an accuracy of roughly 70% to 90%, machine learning classification models such as Decision Tree, Random Forest, Naïve Bayes, and SVM can accurately predict student performance using educational records, assisting in the early identification of at-risk students and supporting focused academic interventions. Khan et al. (2022) [15] According to the study, a number of machine learning classifiers, such as Random Forest, Gradient Boosted Trees, Decision Tree, k-NN, and Naïve Bayes, can predict secondary school student performance with an accuracy of roughly 75%–90%. This allows for the early identification of at-risk students and the implementation of focused educational interventions. Yağcı et al. (2022) [16] Using midterm results and departmental data, the study examined the academic grades of 1854 students in Turkish Language-I at a Turkish state university and discovered that machine learning classifiers like Random Forest, SVM, and k-Nearest Neighbor could predict final exam outcomes with a respectable accuracy (70%–75%), supporting early identification of at-risk students and enhancing learning analytics. Karale et al. (2022) [17] has been demonstrated that machine learning classifiers like Decision Tree, Random Forest, SVM, and KNN can effectively predict students' academic performance, frequently reaching 75% to 92% accuracy. This allows for the early identification of at-risk kids for prompt intervention. [18] Alhazmi et al. (2023) [18] found approximately 70–90% accuracy was attained when machine learning classification models like Decision Tree, Random Forest, SVM, KNN, Logistic Regression, and XGBoost were applied to student academic datasets (such as grades and test scores). Random Forest and SVM demonstrated strong performance in identifying at-risk students, while t-SNE was utilized for feature exploration. Ahmed et al. (2024) [19] used several classification techniques, including SVM, Decision Tree, Naïve Bayes, K-NN, and K-Means clustering, were assessed in a single-institution higher education dataset. Following parameter adjustment, SVM achieved the highest accuracy (96%), while Naïve Bayes performed the worst because of its independence assumptions.[19]

III METHODOLOGY

The study attempts to identify the optimal algorithm for predicting eventual educational attainment by evaluating the accuracy of multiple machine learning models and using student data.A thorough examination and comparison are necessary to determine the effectiveness of different machine learning techniques in prediction tasks. The primary tool for searching literature is Google Scholar. As part of a comprehensive literature review on machine learning techniques for predicting performance among learners, 15 pertinent research papers published in SSRN Electronic Journal, Mathematics (MDPI), IEEE Access, International Journal of Emerging Technologies in Learning (iJET), etc. between 2019 and 2025 were used for this study. This review focused primarily on empirical and data-driven research related to forecasting students' academic success using machine learning techniques. A methodical search approach and an organized information source were used to guarantee accurate and trustworthy results. A comprehensive overview of the body of existing literature in this field was produced by the research selection process, which included a first screening of titles and abstracts, a conclusion, and subsequent reviews in areas of uncertainty.

The study on machine learning and classification is presented in Table 1, which mainly splits down into three categories: publication year, various classification and machine learning algorithms with their minimum and maximum accuracy, and six research articles from 2019 to 2024. The range of accuracy is 70% to 96%.

Table 1: ML Technique : Machine Learning & Classification

Sr. No.	Year of Publication	Algorithm	Accuracy (min–max)
1	2019	Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine (SVM)	~70% – ~90% (varies by classifier)
2	2022	Purposed Model (Random Forest), GB Tree, Decision Tree, k-NN, Naïve Bayes	~75% – ~90% (varies by classifier)
3	2022	Random Forest, Nearest Neighbour (kNN), Support Vector Machine (SVM), Logistic Regression, Naïve Bayes	70%–75% (classification accuracy)
4	2022	Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN)	Approximately 75%–92% (varies by model)
5	2023	Decision Tree; Random Forest; Support Vector Machine (SVM); K-Nearest Neighbors (KNN); Logistic Regression; XGBoost	Approximately 70–90% depending on classifier
6	2024	K-Means clustering (Davies-Bouldin), Support Vector Machine (SVM), Decision Tree, Naïve Bayes, K-Nearest Neighbors (KNN)	77.3% – 96.0%

Five research articles from 2020 to 2023 are included in Table 2, which mostly focuses on different machine learning and data mining algorithms with their minimum and maximum accuracy. The accuracy ranges from a minimum of 70% to a maximum of 98.6%.

Table 2: ML Technique : Machine Learning & Data Mining

Sr. No.	Year of Publication	Algorithm	Accuracy (min–max)
1	2020	Naive-Bayes, Decision Tree, Random-Forest, JRip, and ZeroR	~80%
2	2020	kNN, Logistic Regression, SVM, Decision Tree,Random Forest, and Naive Bayes	Neural Network algorithm (98.6%), kNN (86.2%), Logistic Regression (78.4%), SVM (90.3%), Decision Tree (91.9%), Random Forest (90.0%), and Naive Bayes (81.7%).
3	2022	random forests, nearest neighbour, support vector machines, logistic regression, Naïve Bayes, and k-nearest neighbour algorithms	70–75%

4	2022	Decision tree, Naïve Bayes, k-Nearest Neighbor, and Support Vector Machine along with an ensemble method called "Random Forest".	Naïve Bayes gives the highest accuracy with 89%, Approximately 70%–90% depending on the classifier
5	2023	SVM, Random Forest, Decision Tree, Logistic Regression, Ensemble Models, Deep Learning architectures	93.1% and an F1-score of 90.45%

Table 3 contains two research articles from 2022 and 2023 that mostly discuss various machine learning and ensemble methods and their accuracy. The accuracy range in the results is 72% to 96%.

Table 3: ML Technique : Machine Learning & Ensemble Method

Sr. No.	Year of Publication	Algorithm	Accuracy (min–max)
1	2022	Logistic Regression, Decision Tree, Random Forest, Naïve Bayes	Approximately 70%–90% depending on algorithm
2	2023	random forest (RF), decision tree (DT), AdaBoosts, and support vector machine (SVM).	Approximately 78%–96% depending on the algorithm

A single research publication from 2025 that discusses several machine learning and KNN techniques and their accuracy is included in Table 4. range varies by model and is roughly 70% to 90%.

Table 4: ML Technique : Machine Learning & KNN

Sr. No.	Year of Publication	Algorithm	Accuracy (min–max)
1	2025	K-Nearest Neighbors (KNN) + NGO (KNNG) and KNN + TDO (KNTD), Naive Bayes Classification (NBC) and K-Nearest Neighbor (KNN) algorithms. Regression & Classification	≈70% – 90% (varies by model)

The accuracy of K-means algorithms and machine learning is not specified in the 2022 study article.

Table 5: ML Technique : Machine Learning & K-means

Sr. No.	Year of Publication	Algorithm	Accuracy (min–max)

1	2022	Adaptive-K-means-DNN, the adaptive K-means	Not explicitly reported in abstract; model achieves improved prediction performance compared to benchmarks

Figure 1 shows the distribution of the selected studies by year. By examining recent publications that provide minimum and maximum reliability for every technique, researchers or professionals can ascertain which machine learning algorithms are used to forecast student performance between 2019 and 2025.

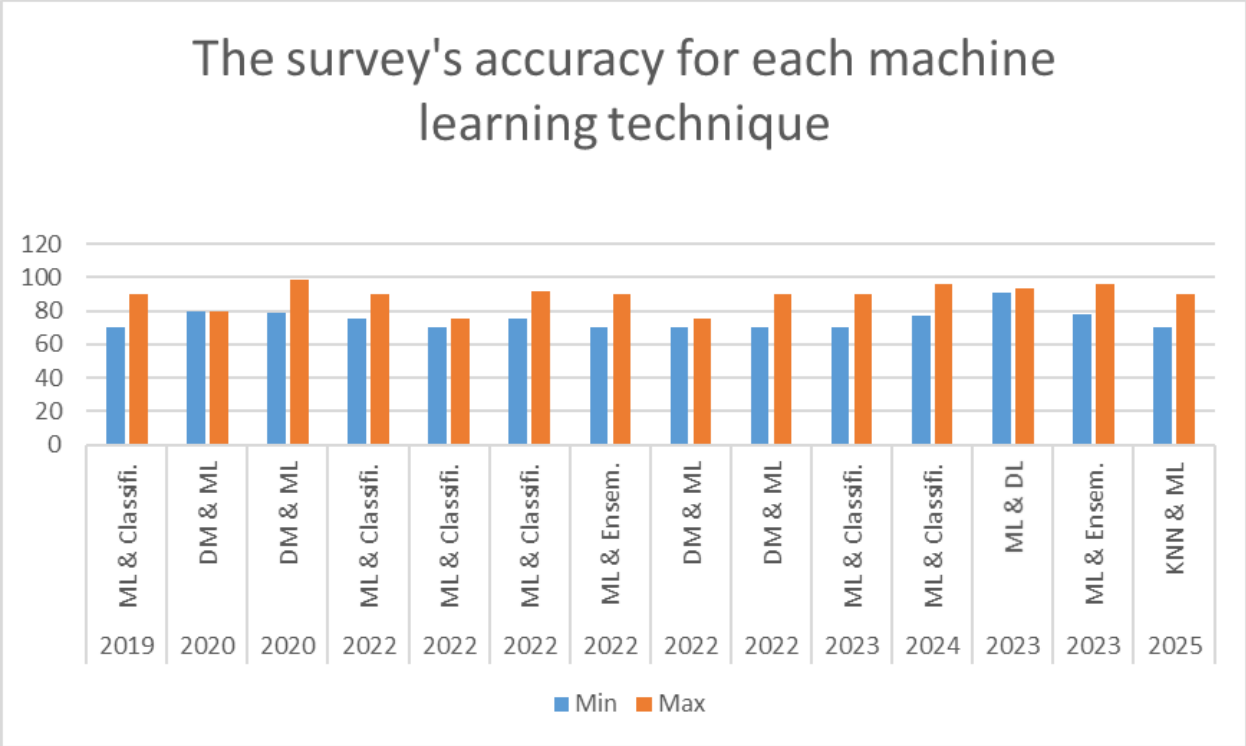

Fig.1: Year wise survey accuracy of each machine learning technique

Figure 2 shows how well machine learning predicts student academic achievement using classification, data mining, ensemble, and K-nearest neighbor techniques, along with the minimum and maximum ranges and total number of articles utilized for each technique.

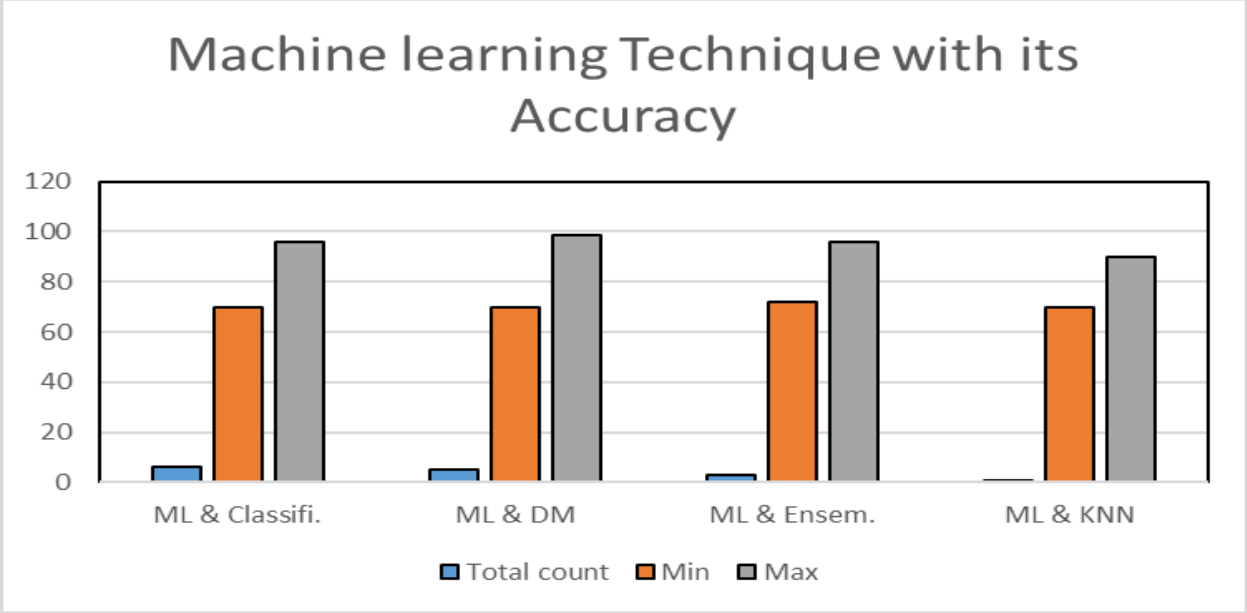

Fig. 2 The accuracy and total number of articles for each machine learning technique

Conclusion:

A comprehensive examination of works on machine learning-based prediction of student educational achievement is necessary for evaluating the proposed study with prior work in order to show its efficacy. This study conducted a comprehensive literature evaluation of fifteen research articles released between 2019 and 2025, which focused on machine learning with a review of categorization techniques. From 2019 to 2024 six research articles that utilized machine learning with classification techniques focused on the Random Forest, Naïve Bayes, Support Vector Machine (SVM), and decision tree algorithms with an accuracy range of 70% to 96%. Decision trees, logistic regression, support vector machines, Random Forest, and Naive Bayes were the subjects of five studies that used machine learning and data mining techniques. The accuracy ranged from a minimum of 70% to a maximum of 98.6% between 2020 and 2023.In the years 2022 and 2023, two research paper that used machine learning with ensemble techniques concentrated on Logistic Regression, AdaBoosts, Decision Tree, Support Vector Machine (SVM), and Random Forest, with accuracy ranging from a low of 70% to a high of 96%.The accuracy of machine learning using KNN techniques with K-Nearest Neighbors (KNN) + NGO (KNNG) and KNN + TDO (KNTD) is between 70% to 90%, according to a single study report from 2025.One 2022 study that addresses machine learning using K-means algorithms concentrated on Adaptive-K-means-DNN and adaptive K-means in cases while accuracy was not stated.To get more thorough insights, future study should take into account a greater variety of studies and examine a variety of parameters. These initiatives can help educators, organizations, and legislators make informed choices and use timely tactics to improve outcomes for learning and the achievement of students.

The availability of trustworthy and well-structured academic data is essential for making accurate predictions about students' academic development, and new technologies can improve the efficacy of early warning systems. Model reliability and utility can be strengthened by employing methods of selecting features and extending data gathering across several institutions. To effectively support educators and policymakers in enhancing learning results, future research should concentrate on bigger, more varied datasets and sophisticated analytical methods backed by reliable guidance.

References

[1] Kumari, V., Meghji, A. F., Qadir, R., Gianchand, U., & Shaikh, F. B. (2024). Predicting Student Performance Using Educational Data Mining: A Review. *KIET Journal of Computing and Information Sciences*, *7*(1).[1]

[2] Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, *16*(1), 1-20.[2]

[3] Ali, K. M., Ahmed Khan, T., Ali, S. M., Aziz, A., Khan, S. A., & Ahmad, S. (2024). An exhaustive comparative study of machine learning algorithms for natural language processing applications. *Engineering Proceedings*, *76*(1), 79.

[4] Sayed, A. F. A., Arafa, M. A., El-Nimr, N. A., Banawan, K. A. S., & Abdou, M. S. (2025). Comparison of machine learning classification and regression models for prediction of academic performance among postgraduate public health students. *Scientific Reports*, *15*(1), 44056.[3]

[5] Walia, N., Kumar, M., Nayar, N., & Mehta, G. (2020, April). Student's academic performance prediction in academic using data mining techniques. In Proceedings of the international conference on innovative computing & communications (ICICC).[5]

[6] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 11.

[7] Chen, Z., Cen, G., Wei, Y., & Li, Z. (2023). Student performance prediction approach based on educational data mining. IEEE Access, 11, 131260-131272.

[8] Verma, S., Yadav, R. K., & Kholiya, K. (2022). Prediction of academic performance of engineering students by using data mining techniques. International Journal of Information and Education Technology, 12(11), 1164-1171.

[9] Yıldız, M., & Börekci, C. (2020). Predicting academic achievement with machine learning algorithms. Journal of educational technology and online learning, 3(3), 372-392.

[10] Nafea, A. A., Mishlish, M., AL-Ani, M. M., Alheeti, K. M. A., & Mohammed, H. J. (2023). Enhancing Student'sPerformance Classification Using Ensemble Modeling. Iraqi Journal For Computer Science and Mathematics, 4(4), 16.

[11] Begum, S., & Padmannavar, S. S. (2022). Genetically optimized ensemble classifiers for multiclass student performance prediction. Int. J. Intell. Eng. Syst, 15(2), 316-328.

[12] Prahmana, I. G., Maulidya, A., Sitepu, K. A., & Habibi, R. (2025). Application of the K-Nearest Neighbor (KNN) Algorithm in Machine Learning to Predict the Selection of Undergraduate Study Programs Based on New KIP Lecture Students. Journal of Artificial Intelligence and Engineering Applications (JAIEA), 4(2), 1522-1526.

[13] Yang, X., Zhang, H., Chen, R., Li, S., Zhang, N., Wang, B., & Wang, X. (2022). Research on Forecasting of Student Grade Based on Adaptive K-Means and Deep Neural Network. Wireless Communications and Mobile Computing, 2022(1), 5454158.

[14] Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. In Proceedings of the 2019 8th international conference on educational and information technology (pp. 7-11).

[15] Khan, M. I., Khan, Z. A., Imran, A., Khan, A. H., & Ahmed, S. (2022, May). Student performance prediction in secondary school education using machine learning. In 2022 8th International Conference on Information Technology Trends (ITT) (pp. 94-101). IEEE.

[16]Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 11.

[17]Karale, A., Narlawar, A., Bhujba, B., & Bharit, S. (2022). Student performance prediction using AI and ML. International Journal for Research in Applies Science and Engineering Technology, 10(6), 1644-1650.

[18]Alhazmi, E., & Sheneamer, A. (2023). Early predicting of students performance in higher education. Ieee Access, 11, 27579-27589.

[19]Ahmed, E. (2024). Student performance prediction using machine learning algorithms. Applied computational intelligence and soft computing, 2024(1), 4067721.