World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:19, No:1, 2025

Machine Learning Algorithms in Study of Student Performance Prediction in Virtual Learning Environment

Shilpa Patra, Pinaki Pratim Acharjya

Abstract**—**One of the biggest challenges in education today is accurately forecasting student achievement. Identifying learners who require more support early on can have a big impact on their educational performance. Developing a theoretical framework that forecasts online learning outcomes for students in a virtual learning environment (VLE) using machine learning techniques is the aim of this study. Resolving the flaws in different forecasting models and increasing accuracy are major goals of the study.

Keywords**—**Virtual Learning Environments, K-Nearest Neighbors, KNN, Random Forest, Extra Trees.

I. INTRODUCTION

THE ability to predict student achievement is essential to education and significantly influences the organization of academic institutions. Recognizing the significance of predicting student outcomes enables institutions to make data-driven, strategic choices that enhance overall academic success. One benefit of precisely predicting student performance is that it enables schools to promptly identify and assist students who might need it. With a comprehensive understanding of the intended student outcomes, schools and institutions can more effectively allocate resources and concentrate support on students who need it most. With this proactive approach, failure rates can be reduced and overall academic performance can be enhanced [1]. The expected performance of students can also be used to optimize planning strategically in educational institutions. These forecasts can be used by educational institutions to develop customized curricula that will assist students in enhancing their academic performance and focusing on areas where they may need more support. Additionally, educational goals can be met, and students can stay motivated and on track for success by scheduling tests and assessments in accordance with these projections [2].

Assessing student achievement can improve education by promoting more meaningful and successful teaching strategies. When educators are aware of the performance standards for their students, they can adjust their classes and employ instructional tactics that are more suitable for the ability levels of their students [3].

Teachers can alter their teaching methods and evaluation strategies to help students understand concepts better and perform better academically. One helpful technique that teachers can use to identify each student's areas of strength and growth is the capacity to forecast student performance. This allows them to provide individualized support to improve performance. Additionally, projections of student performance could be useful for comprehensive assessments of the school [4]. To improve students' comprehension and academic achievement, teachers can alter their teaching and assessment strategies. A helpful tool for teachers to identify each student's strengths and weaknesses is the ability to predict student performance. This allows them to provide tailored, focused support to help students perform better. Additionally, forecasts of student achievement can be utilized to conduct thorough evaluations of the educational institution [5].

Numerous machine-learning algorithms have been discovered to be efficient for specific learning tasks. They are particularly helpful in poorly understood fields where people might lack the expertise necessary to create efficient knowledge-engineering algorithms [6]-[12].

However, traditional approaches sometimes have problems providing precision, scalability, and flexibility in many learning environments, especially in rapidly changing educational environments [13]. In recent years, a number of researchers have attempted to forecast the performance of students utilizing different machine learning methods. Despite the improvements, several current models still have issues like overfitting, a lack of flexibility when working with several forms of educational data, and inefficiencies when it comes to effectively managing large datasets [14]. This work offers a better strategy that uses machine learning techniques, Specifically, K-Nearest Neighbors, Random Forest, and Extra Trees are used to solve these issues. The goal is to increase the precision and effectiveness of student performance prediction models to provide administrators and educators with more trustworthy data.

The utilization of VLEs as training platforms has enabled an approach to education and training delivery. E-learning is more effective with the VLE [15]. Online education is made possible via platforms. Because of its user-friendly interface, affordability, and ability to facilitate productive connection between teachers and students, virtual classrooms are growing increasingly and becoming more popular in the educational field [16]. It is necessary to take the trainees' educational requirements into account [17].

Ms. Shilpa Patra is with the Department of CSE, Haldia Institute of Technology, Haldia, West Bengal, India (e-mail: patrashilpa20@gmail.com).

Dr. Pinaki Pratim Acharjya is with the Department of CSE, Haldia Institute of Technology, Haldia, West Bengal, India (corresponding author, phone: +918348841764; e-mail: ppacharjya@gmail.com).

II. Literature Review

Numerous studies within the data mining field have been carried out in order to improve the way the system of education operates and guarantee that students will have a prosperous future. This contains studies that focused on utilizing machine learning methods to forecast student achievements between 2017 and 2024.

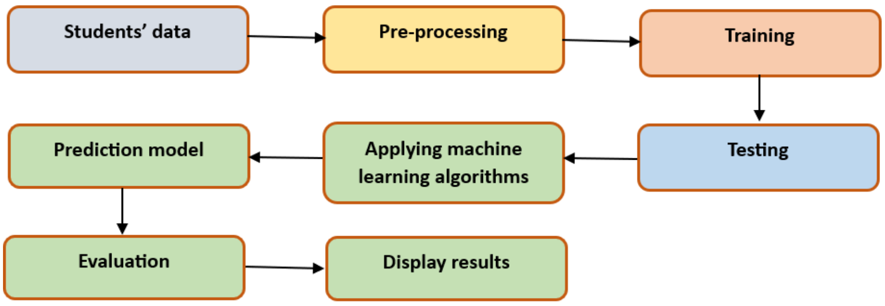

Fig. 1 Students' performance prediction [18]

In order to enhance student performance prediction models, Al-Shehri et al. [19] recommended looking at better model designs. The study used a well-known dataset on math student performance from Minho University to test two algorithms with the Weka tool: SVM and KNN. The SVM outperformed KNN by a small margin, as evidenced by its correlation coefficient of 0.96 vs. 0.95 for KNN. This study highlights the importance of assessing different algorithms to improve the accuracy of prediction models. The study's findings might also be useful for selecting the most qualified student for a particular task, for prompt interventions, or for early preventive measures.

Tanuar et al. [20] recommended using historical data from Bina Nusantara University graduates in their analysis. In order to assist students and prepare proactively, this study used methods including decision trees (DT), deep learning (DL), and generalized linear models to uncover critical elements influencing final outcomes. According to the results, the DT model was the easiest to use when it came to successfully informing users.

Hussein et al. [21] created a predictive model to anticipate potential challenges that students may face in upcoming sessions of a digital design course. This was accomplished by applying machine learning algorithms to data collected from the very advanced educational system DEEDS. The input data for each session of the Digital Design course included metrics like average session duration, total number of activities, mean number of keystrokes, average idle moments, and overall amount of pertinent activity for each exercise. The study employed a variety of machine learning (ML) approaches, such as DTs, logistic regression (LR), support vector machines (SVM), artificial neural networks (ANN), and Naïve Bayes (NB). The results showed that when it came to prediction accuracy, ANN and SVM performed better than other methods. It is simple to include these high-performing algorithms into the TEL system to raise student achievement in subsequent classes.

In order to find traits that can improve student performance, Hammoud et al. [22] suggested creating a model using DT algorithms. A questionnaire comprising 60 questions about relationships, health, social activities, and academic performance was used to gather data from 161 students. Three classifiers — Random Tree, J48, and REP Tree — were employed for the analysis using the Weka 3.8 tool. The results showed that J48 performed better than the other two algorithms. This research demonstrates how data mining and DT algorithms can uncover hidden patterns and provide recommendations to raise student achievement in learning environments.

Vijayalakshmi and Venkatachalapathy developed a deep neural network-based model to forecast student performance. In several fields, including educational data mining, which assesses and improves students' academic performance using data mining concepts and algorithms, machine learning techniques are becoming more and more important. In order to compare the accuracy of different R programming techniques, including DTs (C5.0), SVM, random forest (RF), NB, KNN, and the Deep Neural Network (DNN), they used a Kaggle dataset to train and evaluate the model. The DNN algorithm achieved the highest accuracy and outperformed the other methods with an accuracy rate of 84% [23].

According to Mubarak et al., video click data can be used to predict MOOC participants' success and examine learning patterns. To forecast students' weekly performance, they created a DNN model dubbed LSTM based on a set of implicit features extracted from video streaming data. The objective is to give educators the resources they need to carry out prompt interventions that improve the learning environment. This work tackles the problem of classifying time series data by predicting student performance through the analysis of streaming video data. The suggested LSTM model outperformed basic ANN, SVM, and LR in the real-world datasets used, with an accuracy of 93% [24].

An association between online LMS activity data and students' performance evaluation scores was suggested by a study by Hassan et al. This study used a number of machine learning techniques, such as LR, DT s (J48), RF, multi-layer perceptrons (MLP), and sequential minimum optimization (SMO). The Deanship of E-Learning and King Abdulaziz University's Distance Education program provided the data. With a rate of 99.17%, the results showed that the RF algorithm achieved the maximum accuracy [25].

Hernández-Blanco et al. conducted a systematic study that emphasized the EDM tasks that have benefited by DL and those that require additional research. Predicting student achievement, classifying and profiling students, creating suggestions, and identifying undesired student conduct are a few of these activities. Additionally, both public and private datasets used to test and train DL models in EDM tasks were categorized and explained in the paper. Furthermore, the paper offered a summary of the core concepts, main frameworks, and configurations of DL as they relate to EDM. Lastly, they found that in 67% of the experiments they looked at, DL performed better than traditional ML baselines [26].

The prediction of student outcomes by data mining and machine learning models was examined by Namoun and Alshanqiti. Their three main areas of focus were the prediction of learning outcomes, predictive analytics models for student learning prediction, and important factors affecting student performance. The degree to which pupils had learnt their goals was determined by their achievement grades and performance class requirements. Online learning activities, term assessment marks, and student academic mood were the most obvious indications of learning outcomes, according to the analysis [27].

The relevant EDM study on dropout rates and students at risk of failing was found by Albreiki et al. in another systematic literature review. According to the review's conclusions, a variety of machine learning techniques have been applied to understand and resolve fundamental problems, such as predicting student dropout rates and students who are in danger of failing current courses in educational institutions. The review evaluated student performance using both static and dynamic data, and it showed that machine learning approaches were essential for determining dropout rates and students who were at risk of failing. Additionally, it suggested fixes, such as implementing timely feedback, which can assist teachers and students in resolving problems and improving student performance [28].

The use of the EDM, LA, and ML to forecast student performance in secondary schools and higher education institutions was covered by Nawang et al. The assessment identified the features that have the biggest impacts on the prediction's success and included a summary of the techniques and algorithms used. When forecasting student achievement, the bulk of characteristics are used in the academic feature category, which performs better than other feature categories. The analysis also found that GPA, exam scores, and marks are the best measures of performance when compared to academic performance [29].

The ANN methods used to forecast student success at higher education institutions were studied by Baashar et al. It showed that while other studies used cognitive traits, the most frequently used input variables in the studies were student test scores and demographic characteristics. Therefore, by assessing how well their findings measure academic accomplishment, ANNs, data mining algorithms, and data analysis have been used to specify high accuracy in predicting student outcomes and performance. According to the authors, the methodology's degree of accuracy was unaffected by the sample size, level, educational background, or study setting [30].

III. Proposed Model

In order to forecast a student's chances of success, possible dropout, or increasing involvement in academic activities, this study uses machine learning algorithms. This paper's main goal is to compare several machine learning approaches, address the problem of class imbalance, and examine how each approach enhances predicted accuracy. Three machine learning algorithms are used in the study to categorize student results. The suggested system design, which includes essential elements including data collection, pre-processing, and classification methods, is thoroughly explained in this section. Fig. 2 describes the model's setup and shows the procedures for putting it into practice.

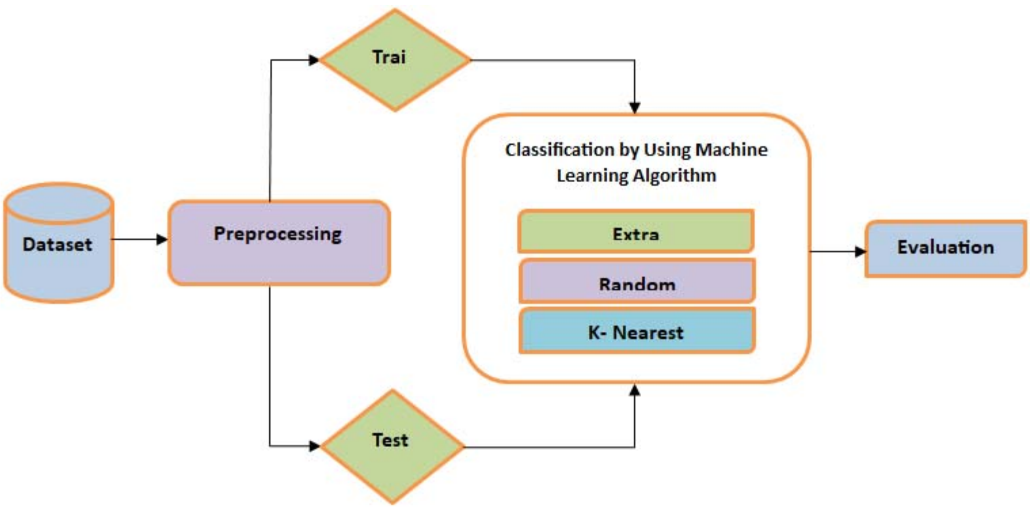

Fig. 2 The Flow Chart of the Proposed Model

IV. Machine Learning Algorithms for Student Performance Prediction

A subfield of artificial intelligence called machine learning (ML) uses labeled datasets to train algorithms [31]. The model can learn to correlate inputs with the right outputs thanks to these datasets, which provide pairs of inputs and associated outputs. A variety of student-related criteria, including attendance, prior exam scores, class involvement, and socioeconomic background, may be used as inputs when predicting student performance. The output is the anticipated result, such as future test scores or general academic achievement. In order to predict student performance, this study used supervised ML, which involves a number of crucial procedures [30].

- Data Collection
- Data Preprocessing
- Label Encoding
- Feature Scaling
- Splitting the Data
- Applying Machine Learning Algorithms

A. Data Collection

The study's dataset, which comes from multiple colleges worldwide, spans the academic years 2021 and 2022. Data were collected through a survey targeted at distance learners, yielding a dataset of 30,000 student records with seventeen different features. These traits can be divided into five groups: study habits, family relationships, academic success indicators, personal and lifestyle data, and contentment with the learning environment. Students are classified as either "pass," "fail," or "drop out" according to their academic standing. A student is deemed to have passed a course or program if they successfully finish it and achieve the minimum score or percentage needed to pass. A failing student, on the other hand, could have to improve their performance or retake the course in order to achieve passing requirements because they did not receive the grade or percentage needed to pass. A dropout is a student who leaves a course or program early, either voluntarily or involuntarily, because of personal problems, financial hardships, or scholastic difficulties.

B. Data Preprocessing

Data preprocessing/preparation/cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a dataset and refers to identifying incorrect, incomplete, or irrelevant parts of the data and then modifying, replacing, or deleting the dirty or coarse data. Since it involves arranging and changing the data, this step is quite important. Reducing the amount of information for best use is the aim of data preparation. Given the limited analytical usefulness of raw data, the method of pre-processing is used to get data ready for analysis. This study used preprocessing procedures before to entering the data into the classification system in order to prevent overfitting. Preprocessing is a crucial phase in the entire process since it is necessary to create a predictive model that produces reliable results.

C. Label Encoding

Numerical input is typically required for ML algorithms. Nevertheless, categorical attributes expressed as strings are included in the dataset used in this study. The'sklearn preprocessing' module's Label Encoder is used in the study to solve this problem. A helpful tool for standardizing label formats and making sure they only contain values between 0 and n_classes-1 is the Label Encoder. Both the target column and the feature columns undergo this modification.

D. Feature Scaling

Generally speaking, ML algorithms perform poorly when the numerical attributes in the input data very much in size. To get around this problem, this study makes use of the "Standard Scaler," a preprocessing tool that is part of the "sklearn preprocessing" package. This tool ensures a mean of 0 and a variance of 1 by standardizing the dataset's characteristics to a common scale, maximizing efficacy across a range of ML techniques.

E. Splitting the Data

In ML, data splitting is typically done to avoid overfitting. That is an instance where a ML model fits its training data too well and fails to reliably fit additional data. The original data in a ML model are typically taken and split into three or four sets. The dataset's 'train_test_split' function from the 'model_selection' module in 'sklearn' was used to construct a training set and a testing set. The training set is used to train ML models, while the testing set is used to assess how well they perform. 80% of the data went to the training set, while only 20% went to the testing set. Additionally, a specific value was supplied to the 'random_state' option to ensure that the data splits could be successfully reproduced.

F. Applying Machine Learning Algorithms

The fifth stage in developing a model for solving problems is to apply ML techniques. Applying ML algorithms involves a process of collecting data, preparing it for analysis, selecting the appropriate algorithm based on the problem, training the model on the data, evaluating its performance, and finally using it to make predictions on new data; essentially, teaching a computer to identify patterns and make decisions based on the information provided. This study employed ensemble methods and three different ML algorithms —Extra Trees, RF, and K-Nearest Neighbours—to predict student performance. These specific methods were picked in order to enable comparison with earlier research that tackled the same subject. The system was then updated to incorporate each of these algorithms. Three supervised algorithms are used in the paper: RF, K-Nearest Neighbours, and Extra Trees.

- *Extra trees:* The RF algorithm serves as the foundation for the Extra Trees Classifier, a ML methodology that is classified as an ensemble method. Often called "Extremely Randomized Trees," it builds individual DTs using randomization. This versatile approach has gained popularity in industries like natural language processing, healthcare, and finance since it can be used for both classification and regression applications. Using the aggregated judgments of several trees to increase resilience to outliers, it works remarkably well with high-dimensional datasets or those with noisy features. The number of trees in the ensemble and the maximum depth of each tree are important factors to consider while training an Extra Trees Classifier since they affect the bias to variance ratio. Two strategies that can further enhance the model's performance are cross-validation and hyperparameter tuning. For a variety of ML applications, the Extra Trees Classifier is a helpful tool since it can efficiently balance model complexity, bias, and variation. It was necessary to correct for class imbalance in order to predict student performance in this study with an accuracy rate of 98.15%.

- *KNN (K-Nearest Neighbor):* A non-parametric, supervised learning classifier, the KNN algorithm classifies or predicts how to group a single data point based on closeness. This is among the most widely used and straightforward [24]. The KNN algorithm is a non-parametric, supervised learning classifier that uses proximity to categorize or predict how to group a single data point. It is among the most widely used and straightforward regression and classification classifiers in ML today. The KNN technique, which operates on the premise that comparable points can be discovered close to one another, is commonly employed as a classification algorithm, though it can be applied to regression or classification problems [32].
- *Random Forest:* Regression and classification are used by ML classifiers. The KNN technique, which may be applied to regression or classification issues, is widely used as a classification approach since it is based on the notion that comparable points can be discovered close to one another. RF is a supervised ML method that performs exceptionally well in both classification and regression tasks. The RF approach is well-known for its effectiveness and simplicity of use. It uses DTs, which are constructed by selecting a root node from a feature set and dividing it into different branches according to the entropy values determined for each feature [33]. The fundamental idea of this technique is building several little DTs, which is a computationally efficient way. By building these trees concurrently, techniques such as majority voting can be used to merge them into a single, trustworthy model. RF has a number of noteworthy advantages. It can successfully manage missing values and tackle regression and classification problems, ensuring accuracy even with imperfect data. It also does well on big datasets with lots of features. Given a well-balanced dataset and excellent data quality, we particularly adjusted the RF classifier in our study to predict student performance.

V. Conclusion

The study focuses on developing a practical application for predicting student achievement. KNN, RF, and Extra Trees make up the integrated model, which has shown potential in tailoring interventions to each student's unique needs. These methods demonstrate that ML can accurately predict student performance, which encourages its use in educational settings. Future research should attempt to use and evaluate the model in a range of educational and cultural contexts to confirm its universality and adaptability. By identifying critical components that contribute to academic success, this method can assist educators and organizations in developing targeted plans.

References

[1] Farrington CA, Roderick M, Allensworth E, Nagaoka J, Keyes TS, Johnson DW, et al. Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance--A Critical Literature Review. ERIC. 2012; 38 p.

[2] Hale EL, Moorman HN. Preparing school principals: A national perspective on policy and program innovations. ERIC. 2003; p 11.

[3] Del Río S, López V, Benítez JM, Herrera F. On the use of mapreduce for imbalanced big data using random forest. Inf Sci (Ny). 2014; 285: 112–37. https://doi.org/10.1016/j.ins.2014.03.043

[4] Thai-Nghe N, Drumond L, Horváth T, Krohn Grimberghe A, Nanopoulos A, Schmidt-Thieme L. Factorization techniques for predicting student performance. In: Educational recommender systems and technologies: Practices and challenges. IGI Global. 2012; p. 129–53. https://doi.org/10.4018/978- 1-61350-489-6.ch006

[5] Daud A, Aljohani NR, Abbasi RA, Lytras MD, Abbas F, Alowibdi JS. Predicting student performance using advanced learning analytics. In: Proc 26th int conf world wide web comp. 2017; p. 415–21. https://doi.org/10.1145/3041021.3054164

[6] Koley, S., Acharjya, P. P., Mukharjee, R., Roy, S., Das., S., Deep Learning: An Approach to Encounter Pandemic Effect of Novel Corona Virus (COVID19), in Convergence of Deep Learning in Cyber-IoT Systems and Security, pp. 183-205, Print ISBN:9781119857211 |Online ISBN:9781119857686, 04 November 2022, DOI: https://doi.org/10.1002/9781119857686.ch9, Wiley Publications.

[7] Chattopadhyay, S., Acharjya, P. P., Koner, C., A Morphological Approach to Match Two Color Images Using Deep Learning, Proc. IEEE sponsored International Conference on Smart Systems and Inventive Technology, pp. 217-220, Tirunelveli, Tamilnadu, India, December 2018.DOI: 10.1109/ICSSIT.2018.8748396

[8] Chattopadhyay, S., Acharjya, P. P., Koner, C., A Deep Learning Approach to implement slab based image steganography algorithm of RGB images, Proc. IEEE sponsored International Conference on Inventive Computation Technologies, pp. pp. 734-736, Coimbatore, Tamilnadu, India, November, 2018.DOI: 10.1109/ICICT43934.2018.9034299

[9] Koley, S., Acharjya, P.P., Concepts & Techniques in Deep Learning Applications in the field of IoT Systems and Security, in Convergence of Deep Learning in Cyber-IoT Systems and Security, pp. 303-348, Print ISBN:9781119857211 |Online ISBN:9781119857686, 04 November 2022, DOI: https://doi.org/10.1002/9781119857686.ch15, Wiley Publications.

[10] Das, S., Acharjya, P. P., Integrating Machine Learning to Enhance Online Learning Student Performance, in book, Social Reflections of Human-Computer Interaction in Education, Management, and Economics, pp. 1-15, ISBN13: 9798369330333, EISBN13: 9798369330340, IGI Global: International Academic publisher, July 2024. DOI: 10.4018/979-8-3693-3033-3

[11] Das, S., Acharjya, P. P., Machine Learning Integration to Analyse the Students Behavior in an Online Digital Learning System, book, "Integrating Artificial Intelligence in Education: Enhancing Teaching Practices for Future Learning, pp-pages 217-242, ISBN13: 9798369339442, EISBN13: 9798369339459, IGI Global: International Academic publisher, 2024. DOI: 10.4018/979-8-3693-3944-2

[12] Das, S., Acharjya, P. P., Utilizing Machine Learning to Increase Online Learning Instructors' Efficiency, in book, Social Reflections of Human-Computer Interaction in Education, Management, and Economics, pp. 40-54, ISBN13: 9798369330333, EISBN13: 9798369330340, IGI Global: International Academic publisher, July 2024. DOI: 10.4018/979-8-3693-3033-3

[13] Nafea AA, Mishlish M, Muwafaq A, Shaban S, Alani MM, Alheeti KMA, et al. Enhancing Student’s Performance Classification Using Ensemble Modeling. Iraqi J Comput Sci Math. 2023; 4(4): 204– 14. https://doi.org/10.52866/ijcsm.2023.04.04.016

[14] Xu Z, Yuan H, Liu Q. Student performance prediction based on blended

learning. IEEE Trans Educ. 2020; 64(1): 66–73. https://doi.org/10.1109/TE.2020.3008751

[15] J. Mc Nelis An exploratory study into virtual learning environments as training platforms in the workplace Int. J. Adv. Corp. Learn, 7 (3) (2014), pp. 8-11.

[16] M. El Boghdady, B.M. Ewalds-Kvist, A. Alijani A review of online platforms in training and surgical education Eur. Surg., 51 (2) (2019), pp. 41-48.

[17] J. Sneha and G. Nagaraja, "Virtual learning environments-a survey," arXiv preprint arXiv:1402.2404, 2014.

[18] Alboaneen, D.; Almelihi,M.; Alsubaie, R.; Alghamdi, R.;Alshehri, L.; Alharthi, R., Development of a Web-Based Prediction System for Students' Academic Performance. Data 2022, 7, 21. https://doi.org/10.3390/ data7020021

[19] Al-Shehri H, Al-Qarni A, Al-Saati L, Batoaq A, Badukhen H, Alrashed S, et al. Student performance prediction using support vector machine and knearest neighbor. IEEE. Canadian Conference on Electrical and Computer Engineering (CCECE). 2017.

[20] Tanuar E, Heryadi Y, Abbas BS, Gaol FL. Using machine learning techniques to earlier predict student's performance. IEEE. Indonesian Association for Pattern Recognition International Conference (INAPR). 2018; p.p. 85–9. https://doi.org/10.1109/INAPR.2018.8626856

[21] Hussain M, Zhu W, Zhang W, Abidi SMR, Ali S. Using machine learning to predict student difficulties from learning session data. Artif Intell Rev. 2019; 52: 381–407. https://doi.org/10.14569/IJACSA.2016.070531

[22] Hamoud A, Hashim AS, Awadh WA. Predicting student performance in higher education institutions using decision tree analysis. Int J Interact Multimed Artif Intell. 2018; 5: 26–31. https://ssrn.com/abstract=3243704

[23] Vijayalakshmi V, Venkatachalapathy K. Comparison of predicting student's performance using machine learning algorithms. Int J Intell Syst Appl. 2019; 11(12): 34. https://doi.org/10.5815/ijisa.2019.12.04

[24] Mubarak AA, Cao H, Ahmed SAM. Predictive learning analytics using deep learning model in MOOCs' courses videos. Educ Inf Technol. 2021; 26(1): 371–92.

[25] Hasan R, Palaniappan S, Mahmood S, Abbas A, Sarker KU, Sattar MU. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Appl Sci. 2020; 10(11): 3894. https://doi.org/10.3390/app10113894

[26] Hernández-Blanco, A., Herrera-Flores, B., Tomás, D. and Navarro-Colorado, B., 2019. A systematic review of deep learning approaches to educational data mining. Complexity, 2019(1), p.1306039.

[27] Namoun, A. and Alshanqiti, A., 2020. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), p.237.

[28] Albreiki, B., Zaki, N. and Alashwal, H., 2021. A systematic literature review of student'performance prediction using machine learning techniques. Education Sciences, 11(9), p.552.

[29] Nawang, H., Makhtar, M. and Hamzah, W.M., 2021. A systematic literature review on student performance predictions. International Journal of Advanced Technology and Engineering Exploration, 8(84), pp.1441-1453.

[30] Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A.A., Alsariera, Y.A., Ali, A.Q., Hashim, W. and Tiong, S.K., 2022. Toward predicting student's academic performance using artificial neural networks (ANNs). Applied Sciences, 12(3), p.1289.

[31] Kemper L, Vorhoff G, Wigger BU. Predicting student dropout: A machine learning approach. Eur J High Educ. 2020; 10(1): 28–47.

[32] Mucherino A., Papajorgji P. J., Pardalos P. M., Mucherino A., Papajorgji P. J., and Pardalos P. M., K-nearest neighbor classification, *Data Mining in Agriculture*. (2009) 83–106.

[33] Pinto A, Pereira S, Correia H, Oliveira J, Rasteiro DMLD, Silva CA. Brain tumour segmentation based on extremely randomized forest with high-level features. Annu Int Conf IEEE Eng Med Biol Soc. IEEE. 2015; p. 3037–40.

Open Science Index, Computer and Information Engineering Vol:19, No:1, 2025 publications.waset.org/10013966.pdf