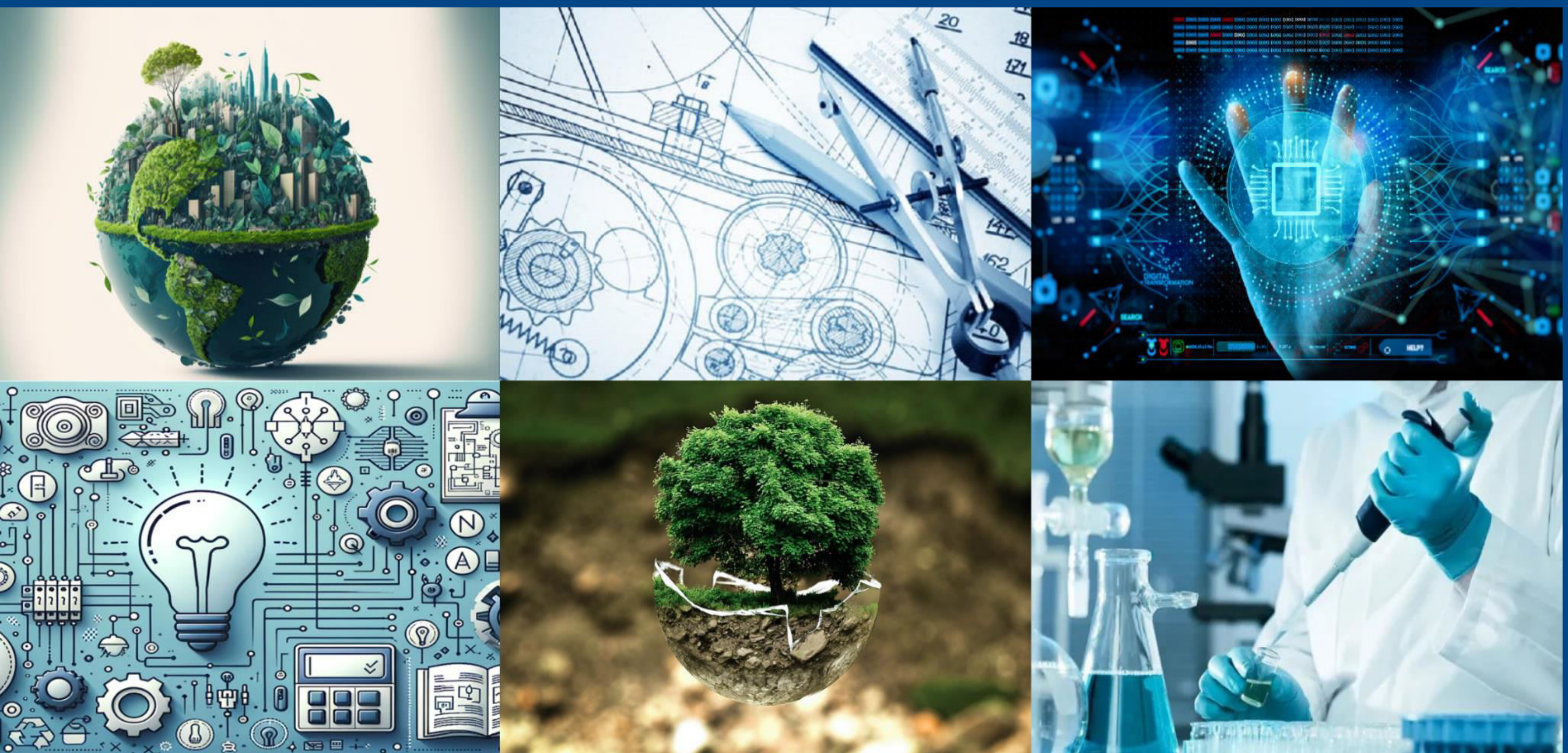
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 3, March 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Predicting Student Academic Performance using Machine Learning

T.Kiruba Rani[1], H. Santhu Mohamed[2], S.Sarang[3]

Assistant Professor, Dept. of CS, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India[1]

B.Sc CS Students, Dept. of CS, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India[2],[3]

ABSTRACT: The emergence of technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and Big Data has significantly contributed to research aimed at enhancing the student learning experience and addressing challenges within the education system. Machine Learning enables data analysis to identify patterns and generate predictions. This study presents an ML model designed to classify and forecast students' academic performance using supervised ML techniques, including Random Forest, Support Vector Machines, Gradient Boosting, Decision Tree, Logistic Regression, Regression, Extreme Gradient Boosting, and Deep Learning. The primary objective is to predict academic success based on past data while determining key factors influencing student performance. The proposed approach enhances prediction accuracy by evaluating multiple ML models against a Deep Learning model. Findings indicate that Extreme Gradient Boosting achieves a 97.12% accuracy rate in forecasting student academic outcomes. This research concludes that integrating Machine Learning in education can assist educators in identifying learning gaps and detecting struggling students early, enabling data-driven decision-making to improve learning outcomes.

KEYWORDS: Machine Learning, Deep Learning, Student Academic Success, Educational Data Analysis, Data Analytics, Convolutional Neural Networks (CNN)

I. INTRODUCTION

Educational Data Mining (EDM) utilizes data mining, machine learning, and deep learning techniques to analyze academic data and enhance student learning experiences. The interaction between students and educational platforms generates vast amounts of data, which, when analyzed, provides valuable insights into learning processes and academic performance. Further investigation can reveal academic, demographic, and social factors influencing student success. Student academic achievement is typically evaluated through various metrics, including final grades, Grade Point Average (GPA), and standardized test scores. Reports from the U.S. Department of Education and the National Assessment of Educational Progress (NAEP) highlight ongoing challenges within the education system, such as low academic performance, rising university dropout rates, delayed graduations, and insufficient workforce preparedness. These issues have persisted over time, with a more pronounced impact on minority students. Currently, various ML tools such as R Software, Python Scikit-learn, and TensorFlow are widely used. Additionally, multiple ML algorithms, including Random Forest, Support Vector Machines (SVM), AdaBoost, Decision Tree, Naïve Bayes, and K-Nearest Neighbors, are employed to predict student academic performance effectively.

A. Objective

The Student Performance Prediction System leverages Machine Learning (ML) to analyze various factors such as attendance, study habits, parental education and past academic records to forecast student success. By processing structured and unstructured data, the system applies feature engineering techniques to identify key predictors, including parental involvement, economic background, and cognitive skills. Various ML models such as Linear Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and Neural Networks are trained to enhance prediction accuracy. Additionally, Natural Language Processing (NLP) is used to analyze student feedback and written assignments for deeper insights. The system provides real-time academic performance predictions, personalized learning recommendations, and an early warning system for at-risk students. An interactive dashboard enables educators to track student progress, make data-driven decisions, and implement targeted interventions. By fostering a proactive approach, this AI-powered solution aims to enhance student engagement, improve success rates, and optimize

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

educational outcomes.

B. Significance and Impact

The application of Machine Learning (ML) in predicting student academic performance presents a transformative opportunity to address key challenges in the education sector. With the increasing need for data-driven insights in education, ML models can analyze various factors influencing student outcomes and provide early predictions to support educators and institutions.By leveraging historical academic records, behavioral patterns, and socio-economic factors, ML-powered prediction systems can offer accurate and personalized assessments of student performance. These insights enable proactive interventions, helping educators identify students at risk, tailor learning experiences, and enhance overall academic success.

C. Scope of the Paper

This study explores a machine learning-based system that goes beyond traditional academic evaluation, offering a proactive approach to student performance prediction through data-driven insights, personalized learning recommendations, and early intervention strategies. Designed as a versatile tool, it aids educators and institutions in optimizing student success and retention.

1. Exploration of ML techniques for analyzing student data and predicting academic outcomes.
2. Early identification of at-risk students to enable timely interventions.
3. Personalized learning recommendations based on performance trends and behavioral patterns.

II. LITERATURE REVIEW

The integration of machine learning (ML) in predicting student academic performance has gained significant attention as a means to enhance educational outcomes, provide early interventions, and personalize learning experiences. These methods often struggled to capture the intricate relationships between various academic and behavioral factors, leading to less effective predictions.However, advancements in ML techniques have enabled the development of more sophisticated and accurate models capable of analyzing diverse datasets, including student demographics, attendance records, engagement levels, and previous academic performance. Initial studies focused on basic ML techniques, such as decision trees and linear regression, to identify patterns and trends in student performance (Romero & Ventura, 2010). While these methods provided valuable insights, they were often insufficient for capturing the dynamic and evolving nature of student learning.

With the advent of deep learning and ensemble methods, more advanced algorithms, such as Random Forest, Support Vector Machines (SVMs), and artificial neural networks, have been employed to improve the accuracy and interpretability of student performance predictions (Kotsiantis et al., 2007). In conclusion, the evolution of ML techniques—from traditional statistical approaches to deep learning and transformer-based models—has significantly improved the ability to predict student academic performance.

III. METHODOLOGY

A. Data Collection

Data Collection involves gathering student-related information from various sources, including demographic details such as age, gender, and socio-economic background, along with academic records like grades, test scores, and historical performance.

B. Data Preprocessing

Data Preprocessing is a crucial step in handling inconsistencies, missing values, and data imbalances. Missing data is addressed using imputation techniques such as mean or mode replacement, ensuring completeness. Numerical features are normalized to maintain uniformity, while categorical variables are converted into numerical representations using encoding methods like one-hot encoding or label encoding.

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803191

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

C. Model Selection and Training

Model Selection and Training involves testing multiple machine learning algorithms to determine the best-performing model for predicting student success. Traditional ML models such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Logistic Regression are explored, alongside advanced deep learning techniques like Artificial Neural Networks (ANNs) and Long Short-Term Memory (LSTM) networks.

D. Model Evaluation

Model Evaluation is conducted using various performance metrics to assess the effectiveness of each model. Accuracy, precision, recall, and F1-score are used for classification-based predictions, while regression models are evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). To enhance model reliability, cross-validation techniques such as k-fold cross-validation are employed, ensuring that the model performs well across different subsets of data.

E. Model Interpretation and Deployment

Model Interpretation and Deployment ensure that the selected model is not only accurate but also interpretable for real-world applications. Feature importance analysis is conducted to determine the key factors influencing academic performance predictions. Visualization techniques such as heatmaps and graphs are used to present insights in an easily understandable manner. Finally, the model is integrated into a web or mobile application, providing educators, students, and administrators with real-time predictions and recommendations.

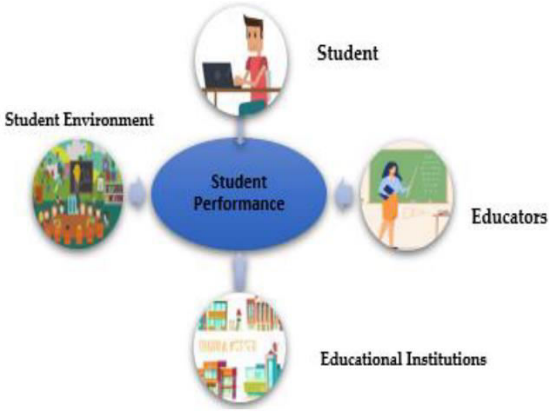


Fig 1. System Methodology

IV. RELATED WORK

Several advancements in machine learning (ML) and data analytics have contributed to the development of models that predict student academic performance. These models leverage artificial intelligence to analyze student data, identify at-risk students, and provide personalized learning recommendations. Below are some of the notable studies and applications in this field:

A. Early Machine Learning Models

Early research in student performance prediction relied on traditional statistical and machine learning techniques. Kotsiantis et al. (2007) compared multiple ML algorithms, including Decision Trees, Naïve Bayes, and Artificial Neural Networks, to predict student success based on academic and demographic features. Their findings demonstrated that ensemble methods improved classification accuracy.

B. Ethical Considerations and Explainable AI

Recent studies have focused on improving the interpretability and fairness of ML models in education. Asif et al. (2017) applied Explainable AI (XAI) techniques to ensure that academic institutions can understand the factors influencing predictions. Additionally, Bowers et al. (2018) addressed concerns regarding bias in student data, advocating for transparent and ethical AI models in academic performance prediction.

C. Deep Learning and Advanced ML Techniques

The rise of deep learning has further enhanced the accuracy and adaptability of academic performance prediction models. Shahiri et al. (2015) reviewed various ML approaches in educational data mining, concluding that Artificial Neural Networks (ANNs) outperform traditional models in capturing complex learning patterns. Al-Barrak and Al-Razgan (2016) applied deep learning techniques to student engagement data, showing improvements in predicting final grades and dropout rates. Furthermore, Tomasevic et al. (2020) integrated socio-economic factors into ML models, highlighting the importance of holistic data analysis in educational predictions.

V. EXISTING SYSTEM

The traditional approach to predicting student academic performance primarily relies on manual methods, which include teacher assessments, paper-based reports, parental feedback, and basic statistical analysis. Teachers evaluate students based on classroom behavior, test scores, and personal observations, while schools generate periodic progress reports that summarize academic performance. Additionally, parental feedback during meetings provides insights into student progress, and some institutions utilize Excel sheets or simple statistical models to track performance trends. However, these methods come with significant limitations.

VI. PROPOSED SYSTEM

To overcome the limitations of traditional student performance assessment methods, our project proposes a Machine Learning-based Student Performance Prediction System that leverages advanced techniques to enhance accuracy, scalability, and efficiency. This system integrates key features such as automated predictions, where machine learning algorithms analyze student data—including exam results, demographics, and study habits—to provide data-driven insights into academic performance. The proposed system offers multiple benefits that improve decision-making in education. By utilizing, it achieves a high prediction accuracy of approximately 85%, significantly reducing human error in student assessments.

A. System Architecture

The system architecture for predicting student academic performance consists of multiple layers, ensuring seamless data collection, processing, machine learning model training, and prediction generation.

1. Data Collection Layer: This layer gathers essential student data from various sources, including academic records such as exam scores, coursework grades, and attendance records.

2. Data Preprocessing Layer: Raw data is often inconsistent and requires preprocessing to enhance model accuracy. This stage involves handling missing values using imputation techniques, feature engineering to extract meaningful attributes, and normalizing numerical data for uniformity.

3. Machine Learning Model Layer: This layer applies various machine learning algorithms to analyze student data and predict performance. Decision Trees and Random Forests help in identifying key performance indicators. These models enable a comprehensive analysis of factors influencing academic success.

4. Prediction and Decision-Making Layer: Once trained, the models generate real-time predictions based on student input data. Educators and administrators can enter student details through the system and receive predictive insights about academic success or potential risk of failure.

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803191

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5. Web-Based User Interface Layer: A user-friendly web-based interface is designed for accessibility and usability. Teachers and administrators can use the platform to analyze student performance trends and plan interventions, while students and parents can receive personalized insights and recommendations. The interface ensures real-time interaction with the predictive system, enhancing its practical implementation in educational institutions.

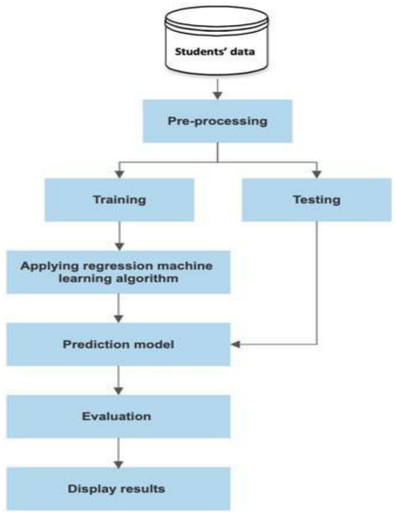


Fig 2. System Architecture

B. System Workflow

The workflow of the Student Performance Prediction System is designed to be intuitive, efficient, and user-friendly. Educators, administrators, and students can easily navigate the system to access academic insights and predictions. The core steps of the workflow include the following:

1. Users, such as teachers or school administrators, can enter student-related data, including academic records (exam scores, coursework performance), demographic details, attendance records, and behavioral patterns. The system may also integrate with existing school databases for automatic data retrieval.

2. Once the data is provided, the system cleans and preprocesses it, handling missing values, normalizing numerical data, and encoding categorical attributes. This step ensures data quality and consistency before passing it to the machine learning model.

3. The system processes the data using advanced machine learning algorithms such as Random Forest, XGBoost, and Artificial Neural Networks (ANNs). The model analyzes patterns, correlations, and historical trends to generate predictions regarding student performance and risk levels.

4. After analyzing the data, the system generates predictions, identifying students who are on track for success and those at risk of academic difficulties. The output includes performance forecasts, personalized recommendations, and early intervention alerts for struggling students.

C. User Privacy and Security

Ensuring the privacy and security of student data is a fundamental aspect of the Student Performance Prediction System. The system employs robust encryption mechanisms to protect sensitive student information, including academic records, demographic details, and behavioral data, both in transit and at rest. Access to student data is strictly controlled, allowing only authorized personnel (such as educators, administrators, and system operators) to retrieve or modify information based on their access level. Role-based authentication ensures that each user interacts only with the data relevant to their responsibilities.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VII. USER INTERFACE DESIGN

The User Interface (UI) of the Student Performance Prediction System is designed to be simple, intuitive, and educator-friendly, ensuring that users such as teachers, school administrators, and students can easily navigate the system and access performance insights. The UI prioritizes usability and efficiency, enabling seamless interaction with predictive analytics and academic data.

Within the performance prediction workflow, the system provides an interactive, step-by-step process where users input academic and behavioral data. The UI presents real-time analytics, visualizing trends through graphs, heatmaps, and progress indicators to make insights more digestible

The interface is designed with accessibility in mind, supporting customizable font sizes, color contrast settings, and mobile-friendly layouts to ensure usability for all users, including individuals with visual impairments. The overall design emphasizes clarity and functionality, reducing complexity while ensuring that educators and students can easily access meaningful academic insights.

VIII. IMPLEMENTATION

The Student Performance Prediction System is developed using a combination of advanced machine learning techniques and structured software development methodologies to ensure high scalability, accuracy, and maintainability. The system follows a modular architecture, allowing seamless integration with school databases and learning management systems (LMS) for enhanced data collection and analysis.At its core, the system leverages machine learning algorithms to analyze academic records, and student demographics to generate accurate performance predictions.

A. Tools and Technologies

The Student Performance Prediction System is developed using a combination of modern programming languages, frameworks, and cloud-based solutions to ensure efficiency and reliability. Python is the primary development language, utilizing machine learning libraries such as Scikit-Learn, TensorFlow, and XGBoost for predictive modeling.The backend services are built using Flask/Django for API management, while React.js or Angular is used for the front-end dashboard, providing educators with an interactive and user-friendly interface. The system integrates with SQL and NoSQL databases (PostgreSQL, Firebase, or MongoDB) to securely store student data, historical records, and prediction results. Cloud-based services such as AWS or Google Cloud ensure scalability and remote access for educators and school administrators.

B. Dataset

The Student Academic Performance Dataset forms the foundation of the prediction model. It consists of historical student performance records, including exam scores, attendance rates, behavioral patterns, study habits, and demographic information. These datasets are sourced from educational institutions, government reports, and publicly available learning analytics datasets.To enhance accuracy, the dataset is regularly updated and preprocessed using techniques such as data normalization, feature selection, and outlier detection. The system also allows integration with school management software to collect real-time student performance data, ensuring continuous learning and improvement of the predictive model.

C. Deployment

The Student Performance Prediction System is deployed using scalable cloud-based infrastructure to ensure high availability, security, and efficiency. The deployment process follows best practices in software engineering, including continuous integration and continuous deployment (CI/CD) pipelines to automate testing, monitoring, and updates.

IX. FUTURE SCOPE

The future of Predicting Student Academic Performance Using Machine Learning lies in enhancing personalized learning, real-time monitoring, and AI-driven recommendations. Future advancements will integrate adaptive learning systems, multimodal data analysis, and explainable AI to provide deeper insights into student performance. Expanding the system to multiple institutions will enable cross-institutional comparisons and global educational benchmarking.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Additionally, AI-powered career guidance, gamification, and bias mitigation will ensure fair and effective academic support. These improvements will make the system more scalable, accurate, and impactful in shaping student success.

X. CONCLUSION

The integration of Machine Learning in predicting student academic performance has transformed traditional evaluation methods, offering data-driven insights for proactive educational interventions. By leveraging advanced algorithms such as Random Forest, SVM, and deep learning models, the system can accurately identify at-risk students, predict academic outcomes, and personalize learning recommendations. Compared to manual assessments, this approach provides greater accuracy, scalability, and efficiency, enabling educators to take timely actions to support student success. Despite its advancements, challenges such as data privacy, model interpretability, and potential biases must be carefully addressed to ensure fair and ethical AI-driven predictions. With continued improvements, including adaptive learning integration, real-time analytics, and global scalability, this system has the potential to revolutionize education, enhance learning experiences, and drive better academic outcomes for students worldwide.

REFERENCES

1. Jayaprakash, S., Krishnan, S., Jaiganesh, V.: Predicting students academic performance using an improved random forest classifier. In: 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, pp. 238–243, March 2020. https://doi.org/ 10.1109/ESCI48226.2020.9167547
2. Bhutto, E.S., Siddiqui, I.F., Arain, Q.A., Anwar, M.: Predicting students' academic performance through supervised machine learning. In: 2020 International Conference on Information Science and Communication Technology (ICISCT), Karachi, Pakistan, pp. 1–6, February 2020. https://doi.org/10.1109/ICISCT49550.2020.9080033
3. Jacob, J., Jha, K., Kotak, P., Puthran, S.: Educational data mining techniques and their applications. In: 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), pp. 1344–1348, October 2015. https://doi.org/10.1109/ICGCIoT.2015.7380675
4. Al Mayahi, K., Al-Bahri, M.: Machine learning based predicting student academic success. In: 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Brno, Czech Republic, pp. 264–268, October 2020. https://doi.org/ 10.1109/ICUMT51630.2020.9222435
5. Olaperi, Y., Fernandez-Sanz, L., Medina, J., Misra, S.: Framework for academic advice through mobile applications (2016)
6. Statement from Secretary DeVos on 2019 NAEP Results. U.S. Department of Education. https://www.ed.gov/news/press-releases/statement-secretary-devos-2019-naep-results. Accessed 24 Feb 2021
7. Rimadana, M.R., Kusumawardani, S.S., Santosa, P.I., Erwianda, M.S.F.: Predicting student academic performance using machine learning and time management skill data. In: 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, pp. 511–515, December 2019. https://doi.org/10.1109/ISRITI48646. 2019.9034585
8. bin Mohd Nasir, M.A.H., bin Asmuni, M.H., Salleh, N., Misra, S.: A review of student attendance system using near-field communication (NFC) technology. In: Gervasi, O., et al. (eds.) ICCSA 2015. LNCS, vol. 9158, pp. 738–749. Springer, Cham (2015). https://doi.org/ 10.1007/978-3-319-21410-8_56
9. Xu, X., Wang, J., Peng, H., Wu, R.: Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. Comput. Hum. Behav. 98, 166– 173 (2019). https://doi.org/10.1016/j.chb.2019.04.015
10. Hasan, H.M.R., Rabby, A.S.A., Islam, M.T., Hossain, S.A.: Machine learning algorithm for student's performance prediction. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, pp. 1–7, July 2019. https://doi.org/10.1109/ICCCNT45670.2019.8944629
11. Scikit-learn: machine learning in Python—scikit-learn 0.24.2 documentation. https://scikitlearn.org/stable/. Accessed 04 May 2021
12. TensorFlow. https://www.tensorflow.org/. Accessed 04 May 2021
13. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/index.php. Accessed 28 Feb 2021
14. Nurafifah, M.S., Abdul-Rahman, S., Mutalib, S., Hamid, N.H.A., Malik, A.M.A.: Review on predicting students' graduation time using machine learning algorithms. Int. J. Mod. Educ. Comput. Sci. 11(7),1(2019).https://doi.org/10.5815/ijmecs.2019.07.01 Predicting Student Academic Performance Using M

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY