

Research Paper

Early Student Performance Prediction and Progress Tracking Using Machine Learning Techniques

K Naga Chaithanya¹, Nathi Uday Kiran², Tammineni Sri Teja³, Pachipala Nagalakshmi⁴,
Mrs. J. Supraja⁵

^{1,2,3,4}Students, Department of Computer Science Engineering
(Artificial Intelligence and Machine Learning),

⁵Assistant Professor, Department of Computer Science Engineering
(Artificial Intelligence and Machine Learning),

Sree Dattha Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, Telangana, India

Abstract— In today's educational environment, a significant amount of student-related data is generated through digital learning platforms and academic systems. Effectively analyzing this data can provide meaningful insights into student learning patterns and outcomes. This study focuses on applying data mining techniques, particularly within the field of Educational Data Mining to predict student performance at an early stage. The proposed system is designed to evaluate students enrolled in an introductory programming course and estimate their likely final performance based on their initial assessments, accounting for approximately 15% of total grading. To achieve accurate predictions, multiple machine learning algorithms from different categories were implemented and tested using the WEKA tool. These algorithms were evaluated based on performance metrics such as accuracy, precision, recall, and F-measure. Among all the models tested, the Decision Tree algorithm demonstrated superior performance in identifying correct outcomes and maintaining balanced evaluation metrics. The developed model not only predicts academic results but also acts as a tracking mechanism, enabling timely feedback. This allows students to recognize their academic standing early and take corrective actions to improve their performance. Ultimately, this approach supports better academic planning and enhances overall student success.

Keywords— Data Mining, Education Data Mining, Student Performance Prediction, Machine Learning, WEKA, Decision Tree, J48 Algorithm, F-Measure, Predictive Modelling.

I. INTRODUCTION

The rapid growth of digital technologies in education has led to the generation of vast amounts of student-related data. This data includes academic records, attendance, assessment scores, and interaction patterns from learning management systems. Extracting meaningful insights from such data has become increasingly important for improving educational outcomes. Educational Data Mining (EDM) and machine learning techniques have emerged as powerful tools for analysing this data and predicting student performance effectively [1], [3]. These predictive models enable educators and institutions to identify students at risk and provide timely interventions to enhance their academic success.

Early prediction of student performance is particularly valuable, as it allows corrective actions to be taken before final evaluations. Research has shown that early assessment data can be a strong indicator of overall academic outcomes, making it possible to guide students toward better learning strategies [2], [12].

Machine learning approaches such as decision trees, support vector machines, neural networks, and ensemble methods have been widely applied to this problem, each offering different levels of accuracy and interpretability [5], [6]. Among these, decision tree-based models are often preferred due to their simplicity, transparency, and strong predictive capabilities [8], [14].

Several studies have focused on analysing student data to predict grades and classify performance levels. For instance, classification techniques have been used to group students into categories such as high, medium, and low performers, enabling targeted academic support [9], [10]. Additionally, feature extraction and selection methods play a crucial role in improving model performance by identifying the most relevant factors influencing student success [11], [22].

Recent advancements also highlight the effectiveness of hybrid and ensemble models, which combine multiple algorithms to achieve higher prediction accuracy [17], [18], [21]. Furthermore, predictive systems are not limited to academic forecasting but also support institutional decision-making. They assist in curriculum planning, resource allocation, and personalized learning strategies [4], [20]. Tracking student progress over time using machine learning models helps in understanding learning behaviours and designing adaptive educational systems [7], [19].

Despite these advancements, challenges such as data quality, model generalization, and interpretability remain significant concerns [15], [16]. In this context, the proposed student performance predictor and tracker aims to leverage machine learning techniques to provide early and accurate predictions. By doing so, it helps students understand their academic standing and take proactive steps to improve their performance, ultimately contributing to better educational outcomes.

II. RELATED WORK

[1] B. Albreiki, N. Zaki, and H. Alashwal (2021) This study presents a comprehensive systematic literature review on student performance prediction using machine learning techniques. The authors analysed numerous research works to identify commonly used algorithms, datasets, and evaluation methods in Educational Data Mining (EDM). Their findings highlight that classification algorithms such as Decision Trees, Support Vector Machines, and Neural Networks are widely applied due to their effectiveness in predicting academic outcomes. The paper also emphasizes the importance of feature selection, noting that demographic, academic, and

behavioural attributes significantly influence prediction accuracy. Additionally, the study discusses challenges such as data imbalance, overfitting, and lack of standard datasets, which affect the generalizability of models. One of the key contributions of this work is its structured comparison of different techniques, providing insights into their strengths and limitations.

[2] C. Romero and S. Ventura (2019) Romero and Ventura focus on the importance of early prediction and support of student learning performance. This work highlights how timely identification of at-risk students can significantly improve educational outcomes. The authors discuss the role of learning analytics and data mining techniques in monitoring student activities and predicting performance during the early stages of a course. They emphasize that early prediction models rely heavily on continuous assessment data, such as quizzes, assignments, and participation metrics. The paper also explores how predictive systems can be integrated into learning management systems to provide real-time feedback to both students and instructors. One of the major contributions of this study is its focus on actionable insights, where predictions are not only used for analysis but also for intervention strategies. The authors argue that combining predictive analytics with personalized feedback mechanisms can enhance student engagement and motivation.

[3] A. Khan and S. K. Ghosh (2021) This paper provides a detailed review of educational data mining techniques used for analysing and predicting student performance in classroom settings. The authors examine various machine learning approaches, including classification, clustering, and regression methods, and compare their effectiveness in different educational scenarios. They highlight that classification techniques are the most commonly used due to their ability to categorize students based on performance levels. The study also discusses the importance of data preprocessing steps such as cleaning, normalization, and transformation, which significantly impact model accuracy. Furthermore, the authors explore the role of different features, including attendance, assignment scores, and socio-economic factors, in influencing student outcomes. The paper identifies several challenges, such as handling large-scale data and ensuring model interpretability, especially in real-world applications. It concludes by suggesting future research directions, including the integration of deep learning and hybrid models. This reference is valuable as it provides a broad understanding of the methodologies and challenges involved in student performance analysis.

[4] M. Al-kmal et al. (2020) This study focuses on the application of machine learning techniques to

support academic decision-making in higher education institutions. The authors propose a framework that uses predictive models to analyse student data and assist administrators in making informed decisions. The research highlights how machine learning can be used to identify patterns in student performance, enabling institutions to improve teaching strategies and curriculum design. Various algorithms were evaluated to determine their effectiveness in predicting student outcomes, with a focus on achieving high accuracy and reliability. The study also emphasizes the importance of data-driven decision-making in modern education systems, where traditional methods may not be sufficient to handle large volumes of data. One of the key contributions of this work is its practical approach to implementing predictive models in real-world educational settings.

[5] F. Ofori, E. Maina, and R. Gitonga (2020) This literature-based review examines the use of machine learning algorithms to predict student performance and improve learning outcomes. The authors analyse various studies to identify the most effective techniques and highlight the growing importance of predictive analytics in education. They discuss commonly used algorithms such as Decision Trees, Naïve Bayes, and Artificial Neural Networks, noting their strengths in handling different types of data. The paper also emphasizes the role of data quality and feature selection in achieving accurate predictions. Additionally, the authors explore how predictive models can be used to enhance teaching and learning processes by providing personalized feedback and identifying students who require additional support. The study highlights the potential of machine learning to transform traditional education systems into more adaptive and student-centered environments. However, it also points out challenges such as data privacy concerns and the need for ethical considerations in data usage. This reference is significant as it reinforces the practical benefits and challenges of applying machine learning in education.

III. DATASET DETAILS

The system is designed with three main user roles: Admin, Teacher, and Student, each having specific functionalities to manage and track academic performance effectively. Initially, the Admin logs into the system through the “Admin Login” interface. After successful authentication, the admin dashboard is displayed, providing access to various management features.

The admin can add new teachers by entering their details through the “Add Teacher” module. Similarly, student information can be added using the “Add Student” option, where valid email IDs are

required to enable communication between teachers, students, and parents.

The admin can also view the list of registered teachers and students through the “View Teachers” and “View Students” sections. Using the “Allocation” module, resources such as subjects or classes are assigned to teachers. Additionally, the “School Performance” feature provides a graphical representation of students’ average performance under each teacher, enabling comparative analysis.

After admin operations, teachers can log in to their accounts. Teachers can mark student attendance, create assignments with deadlines, upload study materials, and record marks for different subjects. The system also includes a messaging feature that allows teachers to communicate with students and automatically send updates to parents via email. Teachers can further analyse student progress by viewing total and average marks.

Students can log in to access assignments, download study materials, view marks, and communicate with teachers through the messaging system. They can also review all messages sent by faculty. Overall, the system provides a comprehensive platform to manage academic activities and continuously monitor student performance using structured data.

IV. PROPOSED METHODOLOGY

The proposed Student Performance Predictor and Tracker system is designed to analyse academic data and predict student outcomes at an early stage using machine learning techniques. The methodology follows a structured approach consisting of data collection, preprocessing, model building, evaluation, and deployment. Initially, data is collected from various academic sources such as student records, attendance logs, assignment scores, and internal assessment marks. This data forms the foundation for building the prediction model. Since raw data may contain inconsistencies, missing values, or irrelevant attributes, a preprocessing step is applied. This includes data cleaning, normalization, and transformation to ensure that the dataset is suitable for analysis. Feature selection techniques are also used to identify the most influential factors affecting student performance, such as attendance percentage, assignment completion rate, and internal test scores.

Once the data is prepared, multiple machine learning algorithms are implemented to predict student performance. These algorithms include Decision Tree (J48), Naïve Bayes, Support Vector Machine, and Random Forest. Each model is trained using historical student data and tested to evaluate its

predictive capability. Among these, the Decision Tree algorithm is chosen as the primary model due to its interpretability and high accuracy in classification tasks.

The dataset is divided into training and testing subsets to ensure proper validation of the model. Performance metrics such as accuracy, precision, recall, and F-measure are used to compare the effectiveness of each algorithm. The model that provides the best results is selected for deployment. The system is then integrated into a web-based application where three types of users Admin, Teacher, and Student interact with the platform. Teachers continuously update student data, including attendance and marks, which is used by the model to generate predictions. The system provides early alerts to students regarding their expected performance, enabling them to take corrective actions. Overall, the methodology combines data-driven analysis with user interaction to create a dynamic and efficient system for predicting and tracking student performance.



Figure [1] : System Architecture

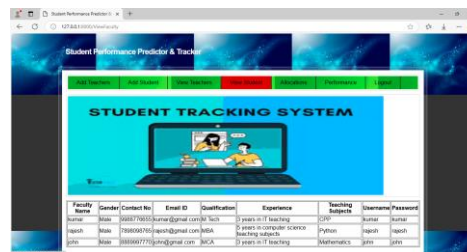
Figure[1] illustrates diagram represents a cloud-based platform integrating Admin, Teacher, and Student dashboards. The Admin manages users and resources, Teachers handle attendance, assignments, and progress tracking, while Students access materials, marks, and communication. All components interact through an API/microservices layer, ensuring seamless data flow, centralized control, and efficient academic management across web and mobile interfaces..

V. RESULT AND DISCUSSION

The experimental results of the proposed Student Performance Predictor and Tracker system demonstrate the effectiveness of machine learning techniques in predicting academic outcomes. Multiple algorithms, including Decision Tree (J48), Naïve Bayes, Support Vector Machine, and Random Forest, were implemented and evaluated using the prepared dataset. The performance of these models was measured using standard evaluation metrics such as accuracy, precision, recall, and F-measure.

Among all the algorithms tested, the Decision Tree (J48) model achieved the highest accuracy in correctly classifying student performance levels. It also showed better consistency in terms of precision and recall, indicating its ability to minimize both false positives and false negatives. The model effectively identified patterns between input features such as attendance, assignment scores, and internal marks, and the final performance of students.

The system successfully generated early predictions based on partial evaluation data , proving its capability to provide timely insights. Additionally, the graphical representation of student performance under different teachers helped in visual analysis and comparison. Overall, the results confirm that the proposed system is reliable and efficient in predicting student outcomes, enabling early intervention and supporting improved academic performance.



Figure[2] : View Teachers

The Figure[2] screen admin can view list of available teachers.



Figure [3] View Students

Figure[3] The image shows screen admin can view list of available students.

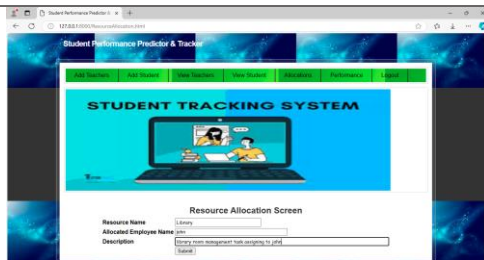


Figure [4]: Resource Allocation

The Figure[4] screen admin is allocating resource to employee.

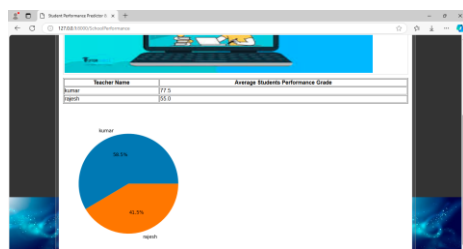


Figure [5]: School Performance

The Figure [5] screen admin can view overall students average grades obtained under teacher name and in graph can see which faculty score which is obtained by all students under that faculty.



Figure [6] Add Attendance

Figure[6] The image shows screen faculty can select present multiple students names by holding CTRL key from keyboard.

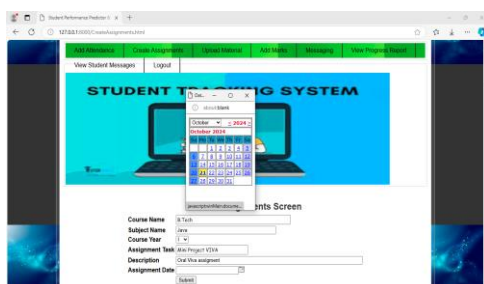


Figure [7] Create Assignment

Figure[7] The image shows screen faculty will create assignment task and then select date.

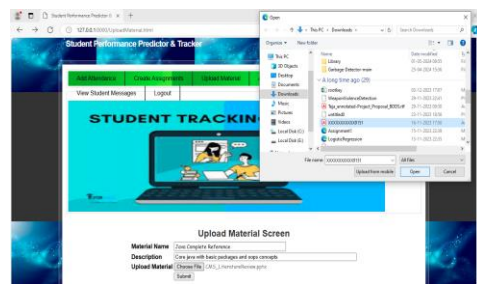


Figure [8] Upload Material

Figure[8] The image shows screen teacher uploading study material .



Figure [9] Marks Details

Figure[9] The image shows screen marks details successfully added and similarly you can add for all subjects.

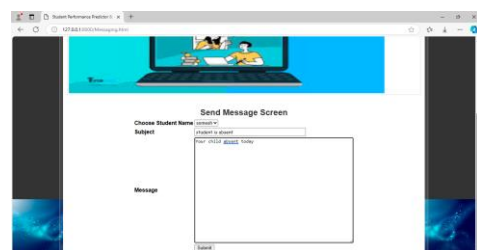


Figure [10] Messaging

Figure[10] The image shows screen teacher can post message which will send to student and their parent's mail.



Figure [11] See image

Figure[11] The image shows screen can see message sent to student parent.

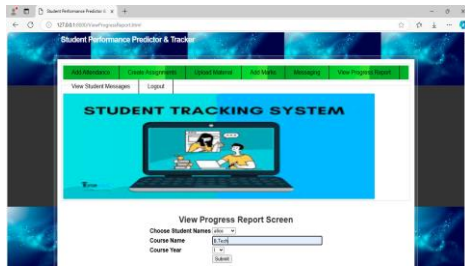


Figure [12] View Progress Marks

Figure[12] The image shows screen teacher can select student name along with course details.

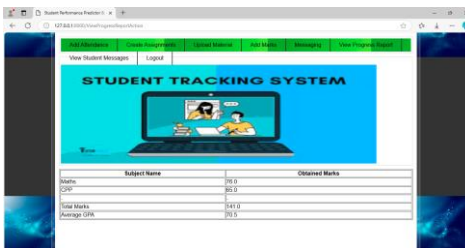


Figure [13] View Marks

Figure[13] The image shows screen teacher can view total and average marks secured by students along with subject details.

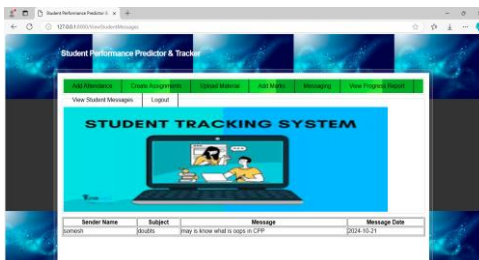


Figure [14] View Student Messages

Figure[14] The image shows screen faculty can view list of messages posted by students and now logout and login as student.

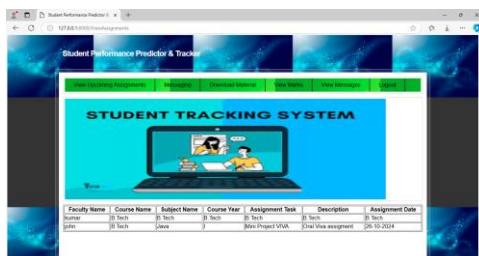


Figure [15] View Upcoming Assignments

Figure[15] The image shows screen student can view list of assignment and now click on 'Messaging' link to send message to faculty.



Figure [16] Messaging

Figure[16] The image shows screen student can post messages to faculty and now click on 'Download Material' link to view list of available material and can download desired material.



Figure [17] Download Material

Figure[5] The image shows screen user can view list of materials and can click on 'Download' link to download desired material.

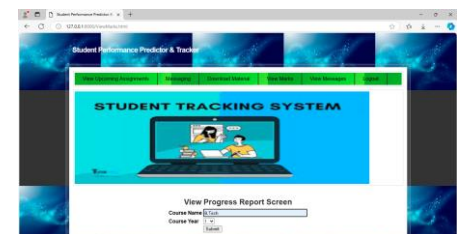


Figure [18] View Marks

Figure[18] The image shows screen student can select course name and year and then press button to get below page.

DISCUSSION

The proposed system demonstrates the practical application of machine learning in the education sector, particularly in predicting student performance. By utilizing historical and real-time academic data, the system provides valuable insights into student progress and learning patterns. One of the key strengths of this approach is its ability to generate early predictions, allowing both students and educators to take timely corrective measures.

The comparison of multiple machine learning algorithms highlights the effectiveness of

the Decision Tree (J48) model, which offers a balance between accuracy and interpretability. Unlike complex models, decision trees provide clear decision rules, making it easier for educators to understand the factors influencing student performance.

However, the system also faces certain limitations. The accuracy of predictions largely depends on the quality and completeness of the input data. Missing or biased data may affect the model's reliability. Additionally, the system may need further improvements to handle large-scale datasets and adapt to diverse educational environments.

Despite these challenges, the system serves as a strong foundation for integrating predictive analytics into academic management systems, enhancing both teaching strategies and student learning outcomes.

VI. CONCLUSION

In conclusion, the Student Performance Predictor and Tracker system offers an effective solution for analysing and forecasting student academic outcomes using machine learning techniques. By leveraging data from attendance, assignments, and assessments, the system provides early predictions that help students understand their academic standing and improve their performance proactively. The implementation of multiple machine learning algorithms ensures a comprehensive evaluation process, with the Decision Tree model emerging as the most suitable due to its high accuracy and simplicity. The integration of this model into a user-friendly web application allows seamless interaction among administrators, teachers, and students, making the system practical and accessible. Furthermore, the system not only predicts performance but also acts as a continuous monitoring tool, enabling better academic planning and decision-making. It enhances communication between teachers, students, and parents through its messaging features, contributing to a more transparent and supportive learning environment. Future enhancements may include the use of advanced models such as deep learning and the incorporation of additional data sources to improve prediction accuracy. Overall, the system plays a significant role in promoting data-driven education and improving student success rates.

REFERENCES

[1] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student performance

prediction using machine learning techniques," *Education Sciences*, vol. 11, no. 9, p. 552, 2021.

[2] C. Romero and S. Ventura, "Guest editorial: Special issue on early prediction and supporting of learning performance," *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 145–147, 2019.

[3] A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," *Education and information technologies*, vol. 26, pp. 205–240, 2021.

[4] M. Al-kmal, H. Mugahed, W. Boulila, M. Al-Sarem, and A. Abuhamdah, "A machine-learning based approach to support academic decision making at higher educational institutions," in *2020 international symposium on networks, computers and communications (ISNCC)*, IEEE, 2020, pp. 1–5.

[5] F. Ofori, E. Maina, and R. Gitonga, "Using machine learning algorithms to predict students' performance and improve learning outcome: A literature based review," *Journal of Information and Technology*, vol. 4, no. 1, pp. 33–55, 2020.

[6] B. Sekeroglu, K. Dimililer, and K. Tuncal, "Student performance prediction and classification using machine learning algorithms," in *Proceedings of the 2019 8th International Conference on Educational and Information Technology*, 2019, pp. 7–11.

[7] J. Xu, K. H. Moon, and M. VanDerSchaar, "A machine learning approach for tracking and predicting student performance in degree programs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 742–753, 2017.

[8] M. I. Hoque, A. K. Azad, M. A. H. Tuhin, and Z. U. Salehin, "University students result analysis and prediction system by decision tree algorithm," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 5, pp. 115–122, 2020.

[9] B. Khan, M. S. H. Khiyal, and M. D. Khattak, "Final grade prediction of secondary school student using decision tree," *International Journal of Computer Applications*, vol. 115, no. 21, 2015.

[10] T. Jeevalatha, N. Ananthi, and D. S. Kumar, "Performance analysis of undergraduate students placement selection using decision tree algorithms," *International Journal of Computer Applications*, vol. 108, no. 15, 2014.

[11] A. Polyzou and G. Karypis, "Feature extraction for classifying students based on their academic performance," *International Educational Data Mining Society*, 2018.

[12] A. Acharya and D. Sinha, "Early prediction of students performance using machine learning techniques," *International Journal of Computer Applications*, vol. 107, no. 1, pp. 37–43, 2014.

[13] T. Anderson and R. Anderson, "Applications of machine learning to student grade prediction in quantitative business courses," *Global Journal of Business Pedagogy*, vol. 1, no. 3, pp. 13–22, 2017.

- [14] M. A. Al-Barrak and M. Al-Razgan, "Predicting students final gpa using decision trees: a case study," *International journal of information and education technology*, vol. 6, no. 7, p. 528, 2016.
- [15] A.Zohair and L. Mahmoud, "Prediction of student's performance by modelling small dataset size," *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–18, 2019.
- [16] L. Cagliero, L. Canale, L. Farinetti, E. Baralis, and E. Venuto, "Predicting student academic performance by means of associative classification," *Applied Sciences*, vol. 11, no. 4, p. 1420, 2021.
- [17] A.Siddique, A. Jan, F. Majeed, A. I. Qahmash, N. N. Quadri, and M. O. A. Wahab, "Predicting academic performance using an efficient model based on fusion of classifiers," *Applied Sciences*, vol. 11, no. 24, p. 11845, 2021.
- [18] W. Wunnasri, P. Musikawan, and C. So-In, "A two-phase ensemble based method for predicting learners' grade in moocs," *Applied Sciences*, vol. 13, no. 3, p. 1492, 2023.
- [19] A. Zollanvari, R. C. Kizilirmak, Y. H. Kho, and D. Hernández-Torrano, "Predicting students' gpa and developing intervention strategies based on self-regulatory learning behaviors," *IEEE Access*, vol. 5, pp. 23792–23802, 2017.
- [20] T. M. Alam, M. Mushtaq, K. Shaukat, I. A. Hameed, M. Umer Sarwar, and S. Luo, "A novel method for performance measurement of public educational institutions using machine learning models," *Applied Sciences*, vol. 11, no. 19, p. 9296, 2021.
- [21] S. Sakri and A. S. Alluhaidan, "Rhem: A robust hybrid ensemble model for students' performance assessment on cloud computing course," *Int. J. Adv. Comput. Sci. Appl*, vol. 11, pp. 388–396, 2020.
- [22] A.Polyzou and G.Karypis, "Feature extraction for next-term prediction of poor student performance," *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 237–248, 2019.