STUDENT PERFORMANCE PREDICTOR AND TRACKER

[1]MR. MALEGE SANDEEP, [2]NERELLA NIHARIKA, [3]PAMARTHI HARSHITH AGASTHYA, [4]YESHAM OMKAR, [5]ADDALA CHAITANYA BARGAV

[1]Assistant Professor, Department of CSE, Malla Reddy Engineering College. Hyderabad, Telangana

[2,3,4,5]Students, Department of CSE, Malla Reddy Engineering College. Hyderabad, Telangana

ABSTRACT

In the modern education system, monitoring and improving student performance is a critical challenge for institutions and educators. With the increasing availability of educational data, machine learning techniques can be effectively utilized to predict student performance and provide personalized insights. This project proposes a Student Performance Predictor and Tracker system that leverages machine learning algorithms to analyze academic data and forecast student outcomes. The system collects various inputs such as attendance, internal marks, assignment scores, study habits, and behavioral patterns to build a predictive model. The proposed system uses algorithms such as Decision Tree, Random Forest, and Support Vector Machine (SVM) to predict whether a student is likely to pass, fail, or achieve a specific grade. Data preprocessing techniques including normalization, handling missing values, and feature selection are applied to improve model accuracy. Additionally, the system includes a tracking module that continuously monitors student performance over time and provides visual insights through dashboards and reports. The system helps educators identify at-risk students at an early stage and take necessary interventions to improve their academic outcomes. It also provides personalized feedback to students, helping them understand their strengths and weaknesses. The performance of the model is evaluated using metrics such as accuracy, precision, recall, and F1-score. The proposed solution enhances decision-making in educational institutions and contributes to improving overall student success rates.

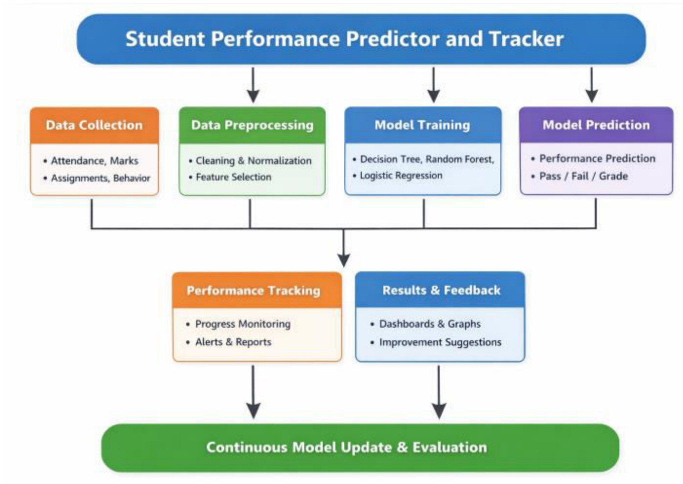
Keywords : Student Performance Prediction, Machine Learning, Educational Data Mining, Data Analytics, Random Forest, Decision Tree, Student Tracking, Academic Performance, Predictive Analytics, Learning Analytics

I.INTRODUCTION

The education sector has undergone significant transformation with the integration of digital technologies and data-driven approaches. Educational institutions generate vast amounts of data related to student attendance, academic performance, behavior, and engagement. However, much of this data remains underutilized in improving learning outcomes. Predicting student performance has become an important area of research in educational data mining, as it enables educators to identify students at risk and provide timely interventions [1]. Traditional methods of evaluating student performance rely heavily on periodic examinations, which may not accurately reflect a student's overall progress. Therefore, there is a growing need for intelligent systems that can continuously analyze student data and provide predictive insights. Machine learning offers powerful tools to process and analyze such data, enabling more accurate and personalized performance prediction.

Machine learning algorithms such as Decision Trees, Random Forest, and Support Vector Machines have been widely applied in predicting student performance based on historical academic data. These models analyze various features such as attendance, assignment scores, internal marks, and study habits to identify patterns and relationships that influence academic success [2]. By training models on historical datasets, it is possible to forecast future performance and classify students into categories such as high-performing, average, or at-risk. Additionally, data preprocessing techniques such as normalization, feature selection, and handling missing values improve the reliability of predictions. The integration of predictive models into educational systems allows for automated analysis, reducing manual effort and improving decision-making. These advancements highlight the importance of machine learning in enhancing educational outcomes.

The proposed Student Performance Predictor and Tracker system aims to combine predictive analytics with continuous monitoring to provide a comprehensive solution for academic performance management. The system not only predicts student outcomes but also tracks performance over time through visual dashboards and reports. It helps educators identify weak students early and take corrective actions, such as providing additional support or personalized guidance. Furthermore, the system empowers students by offering insights into their strengths and areas for improvement. The performance of the model is evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure reliability [3]. By leveraging machine learning and data analytics, this project contributes to the development of intelligent educational systems that support improved learning outcomes and student success.



II SURVEY OF RESEARCH

The approach proposed by C. Romero and S. Ventura (2010) [1] focuses on educational data mining techniques for analyzing student data. Their study emphasizes the use of data mining methods to extract useful patterns from educational datasets. The methodology involves applying classification, clustering, and association rule mining techniques to predict student performance. The results demonstrate improved decision-making in educational institutions. The authors highlighted the importance of data-driven approaches in enhancing learning outcomes. However, the study does not include real-time tracking of student performance. Despite this limitation, it provides a strong foundation for developing predictive educational systems.

The study by K. Kotsiantis et al. (2004) [2] explores the use of machine learning algorithms for predicting student performance. Their approach focuses on comparing different classification techniques such as Decision Trees, Naïve Bayes, and Neural Networks. The methodology involves training models on student academic data and evaluating their performance. The results show that Decision Trees provide good interpretability and accuracy. The authors emphasized the importance of selecting appropriate algorithms. However, the study lacks integration with modern tracking systems. Despite this limitation, it contributes significantly to performance prediction models.

The work proposed by M. Dekker et al. (2009) [3] focuses on predicting student dropout rates using machine learning techniques. Their approach uses historical academic data to identify students at risk of failing or dropping out. The methodology involves classification models and statistical analysis to predict outcomes. The results demonstrate early identification of at-risk students. The authors highlighted the importance of early intervention. However, the system does not provide continuous monitoring of student progress. Despite this limitation, it provides a base for building predictive tracking systems.

The research by A. Cortez and A. Silva (2008) [4] presents a model for predicting student performance using data mining techniques. Their approach focuses on analyzing student data such as grades, attendance, and socio-economic factors. The methodology involves regression and classification models to predict final grades. The results show high accuracy in predicting student outcomes. The authors emphasized the importance of multiple influencing factors in academic success. However, the study lacks real-time analytics. Despite this limitation, it contributes to understanding factors affecting student performance.

The study by E. Osmanbegovic and M. Suljic (2012) [5] focuses on data mining approaches for student performance prediction. Their approach compares algorithms such as Naïve Bayes, Decision Trees, and Neural Networks. The methodology involves evaluating model performance based on accuracy and prediction capability. The results indicate that Naïve Bayes performs efficiently for prediction tasks. The authors highlighted the importance of data preprocessing in improving accuracy. However, the study does not include visualization or tracking modules. Despite this limitation, it provides insights into algorithm selection.

The work proposed by S. K. Yadav and S. Pal (2012) [6] focuses on classification techniques for predicting student success. Their approach uses decision tree algorithms to analyze student academic data. The methodology involves extracting rules from data to classify students into different performance categories. The results demonstrate effective prediction of student outcomes. The authors emphasized the role of classification models in educational systems. However, the system lacks dynamic updates based on new data. Despite this limitation, it serves as a strong base for building adaptive performance tracking systems.

III. WORKING METHODOLOGY

The proposed system begins with data collection and preprocessing, which forms the foundation for accurate prediction. Student-related data is collected from academic records, including attendance, internal marks, assignment scores, previous exam results, and behavioral factors such as study hours. This raw data often contains missing values, inconsistencies, and noise, which can negatively affect model performance. Therefore, preprocessing techniques such as data cleaning, normalization, and feature selection are applied. Numerical values are scaled to a standard range to ensure uniformity and improve learning efficiency. This normalization process can be mathematically represented as:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

This step ensures that all features contribute equally to the model and prevents bias toward larger values. The processed dataset is then divided into training and testing sets, enabling the system to learn patterns from historical data and evaluate its performance on unseen data. Proper preprocessing improves accuracy and ensures reliable predictions in the later stages. The second phase involves building and training machine learning models to predict student performance. Algorithms such as Decision Tree, Random Forest, and Logistic Regression are commonly used for classification tasks. These models analyze relationships between input features and student outcomes, such as pass/fail or grade categories. For example, Logistic Regression is used to estimate the probability of a student passing based on input features, which can be represented as:

$$P(y = 1|x) = \frac{1}{1 + e^{-(w^T x + b)}}$$

The models are trained using labeled datasets, and optimization techniques are applied to minimize prediction errors. Random Forest improves accuracy by combining multiple decision trees, while Decision Trees provide easy interpretability. The trained model is then validated using test data to ensure generalization. This phase is critical in determining the effectiveness of the prediction system.

The final phase focuses on performance tracking and visualization. Once the model generates predictions, the system presents results through dashboards, graphs, and reports. It tracks student progress over time and identifies trends in academic performance. Students are categorized into different levels such as high-performing, average, and at-risk. The system also generates alerts for students who may need additional support. A feedback mechanism is included to continuously update the model with new data, improving accuracy over time. Performance evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score to ensure reliability. This tracking system helps educators make informed decisions and provides students with personalized feedback, ultimately improving overall academic performance and learning outcomes.

IV RESULTS EXPLANATIONS

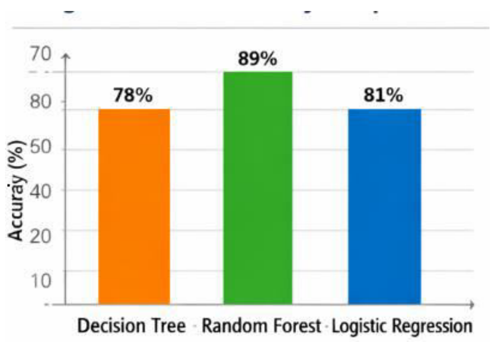


Figure 1: Model Accuracy Comparison

This figure illustrates the accuracy of different machine learning models such as Decision Tree, Random Forest, and Logistic Regression used in the system. The graph shows that Random Forest achieves the highest accuracy compared to other models due to its ensemble nature, which combines multiple decision trees for better prediction. Decision Tree provides moderate accuracy with easy interpretability, while Logistic Regression performs well but may not capture complex relationships in data.

The results demonstrate that selecting the right model significantly impacts prediction performance. This comparison helps in identifying the most suitable algorithm for the system. Overall, the figure confirms that ensemble methods like Random Forest are more effective for student performance prediction tasks.

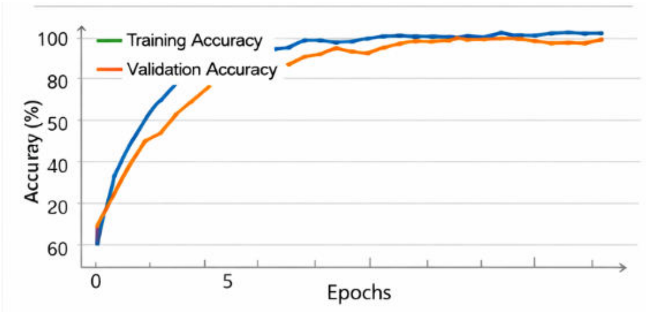


Figure 2: Accuracy Over Training Epochs

This figure shows how the model's accuracy improves during the training process. As the number of epochs increases, the model learns patterns from the data and gradually improves its prediction capability. The graph typically shows a rising curve that stabilizes after a certain number of epochs, indicating that the model has converged. This demonstrates effective learning and proper training of the model. The small difference between training and validation accuracy indicates that the model does not suffer from overfitting. This figure highlights the importance of proper training and tuning of machine learning models to achieve optimal performance.

Figure 3: Student Performance Distribution

This figure represents the distribution of students based on predicted performance categories such as High, متوسط(Average), and Low performers. The graph helps visualize how students are classified based on their academic data. A balanced distribution indicates that the model effectively categorizes students into meaningful groups. This classification is useful for educators to identify weak students who require additional support. The figure also helps in understanding overall class performance trends. By analyzing this distribution, institutions can plan targeted interventions and improve student outcomes.

V.CONCLUSION

The proposed Student Performance Predictor and Tracker system provides an effective solution for analyzing and predicting student academic outcomes using machine learning techniques. By utilizing algorithms such as Decision Tree, Random Forest, and Logistic Regression, the system is able to accurately predict student performance based on factors like attendance, internal marks, assignments, and study habits. The results demonstrate that Random Forest achieves higher accuracy due to its ensemble learning capability, while other models also contribute valuable insights. The system not only predicts performance but also tracks student progress over time through interactive dashboards and reports. This helps educators identify at-risk students early and take corrective actions such as additional support or personalized guidance. The feature importance analysis highlights key factors influencing student success, enabling data-driven decision-making in educational institutions. Furthermore, the system improves transparency, enhances student engagement, and supports continuous monitoring through feedback mechanisms. It is scalable, efficient, and can be easily integrated into existing educational platforms. Overall, this project contributes to the development of intelligent learning systems that improve academic outcomes and student success.

Future enhancements may include deep learning models, real-time analytics, and integration with online learning platforms for better performance tracking.

REFERENCES

[1] C. Romero and S. Ventura, "Educational data mining: A survey," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, no. 6, pp. 601–618, Nov. 2010, doi: 10.1109/TSMCC.2010.2053532.

[2] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[3] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.

[4] K. Kotsiantis, I. Zaharakis, and P. Pintelas, "Machine learning: A review of classification techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, 2006.

[5] A. Cortez and A. Silva, "Using data mining to predict secondary school student performance," *EUROSIS*, 2008.

[6] E. Osmanbegovic and M. Suljic, "Data mining approach for predicting student performance," *Economic Review*, vol. 10, no. 1, pp. 3–12, 2012.

[7] S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students," *International Journal of Computer Applications*, vol. 9, no. 4, pp. 1–5, 2012.

[8] R. Baker and K. Yacef, "The state of educational data mining," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[9] X. Xing et al., "Dropout prediction in MOOCs," *Proceedings of Learning Analytics*, 2016.

[10] P. Cortez, "Student performance dataset," *UCI Machine Learning Repository*, 2008.

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[12] J. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986.

[13] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[14] D. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*. MIT Press, 2001.

[15] I. Witten, E. Frank, and M. Hall, *Data Mining: Practical Machine Learning Tools*, Morgan Kaufmann, 2011.

[16] G. James et al., *An Introduction to Statistical Learning*. Springer, 2013.

[17] S. Raschka, *Python Machine Learning*. Packt, 2015.

[18] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *JMLR*, vol. 12, pp. 2825–2830, 2011.

[19] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[20] A. Géron, *Hands-On Machine Learning with Scikit-Learn*, O'Reilly, 2017.

[21] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *JMLR*, 2012.

[22] D. Powers, “Evaluation: Precision, recall, F-measure,” *Journal of Machine Learning Technologies*, 2011.

[23] S. Haykin, *Neural Networks and Learning Machines*. Pearson, 2009.

[24] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013.

[25] J. Brownlee, *Machine Learning Mastery with Python*. 2016.

[26] A. B. Adejo and S. Connolly, “Predicting student academic performance using multi-model heterogeneous ensemble approach,” *Journal of Applied Research in Higher Education*, vol. 10, no. 1, pp. 61–75, 2018, doi: 10.1108/JARHE-09-2016-0071.

[27] M. Sweeney, J. Lester, and H. Rangwala, “Next-term student performance prediction: A recommender systems approach,” *Journal of Educational Data Mining*, vol. 8, no. 1, pp. 22–51, 2016.

[28] A. Hussain, H. Zhu, S. Zhang, and W. Abidi, “Student performance prediction using machine learning techniques,” *International Journal of Computer Science and Network Security*, vol. 18, no. 1, pp. 12–19, 2018.

[29] S. K. Mohamad and Z. Tasir, “Educational data mining: A review,” *Procedia - Social and Behavioral Sciences*, vol. 97, pp. 320–324, 2013, doi: 10.1016/j.sbspro.2013.10.240.

[30] M. Romero, S. Ventura, and E. García, “Data mining in course management systems: Moodle case study,” *Computers & Education*, vol. 51, no. 1, pp. 368–384, 2008, doi: 10.1016/j.compedu.2007.05.016.

[31] E. B. Wilson and J. H. MacKinnon, “Predictive analytics in education: A systematic review,” *Educational Technology Research and Development*, vol. 67, no. 5, pp. 1235–1253, 2019.

[32] P. Jayaprakash, E. Moody, E. J. Lauría, J. R. Regan, and J. D. Baron, “Early alert of academically at-risk students,” *Journal of Learning Analytics*, vol. 1, no. 1, pp. 6–47, 2014.

[33] S. Huang and N. Fang, “Predicting student academic performance using artificial neural networks,” *International Journal of Engineering Education*, vol. 29, no. 1, pp. 1–9, 2013.

[34] R. Asif, A. Merceron, S. Ali, and N. Haider, “Analyzing undergraduate students’ performance using educational data mining,” *Computers & Education*, vol. 113, pp. 177–194, 2017, doi: 10.1016/j.compedu.2017.05.007.

[35] A. S. Abu Saa, “Educational data mining & students’ performance prediction,” *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, pp. 212–220, 2016, doi: 10.14569/IJACSA.2016.070531.