IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 3, May-June 2024

Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103021

Prediction of Student Performance Using Machine Learning

Prerna Kale[1], Sakshi Mandhare[2], Sakshi Khaire[3], Drishti Dhawan[4], Prof.D. S. Rakshe[5]

[1,2,3,4]B.E.Students, Dept. of Computer Engineering, Pravara Rural Engineering College, Loni, Maharashtra, India

[5] Assistant Professor, Dept. of Computer Engineering, Pravara Rural Engineering College, Loni, Maharashtra, India

ABSTRACT: We discovered a method to forecast students' academic performance based on their grades, behavior, and personal information. While previous studies also explored this avenue, they fell short in offering actionable insights for improvement. Hence, we introduced 'Learning Coefficients' to empower students with strategies for enhancing their performance. Initially, we examined the interconnections between various factors. Subsequently, we experimented with multiple computer programs to determine the most effective predictor of grades. This innovative approach empowered students with personalized insights and practical recommendations for improving their academic performance, transforming abstract data into tangible pathways for growth and achievement. Our research represents a convergence of rigorous empirical inquiry, advanced computational analysis, and a commitment to empowering students with the tools they need to thrive in their academic endeavors.

KEYWORDS: Machine Learning Models, Support Vector Machine, Random Forest, Hidden Layers, etc.

I.INTRODUCTION

Within educational institutions, a wealth of student data is meticulously recorded, encompassing details such as their origins, academic progress, and extracurricular engagements. Leveraging advanced computer programs, we can harness this trove of information to employ machine learning algorithms for predictive analysis. This innovative approach enables educators and administrators to anticipate students' academic trajectories more accurately, facilitating targeted interventions to enhance support mechanisms. Of particular interest is our focus on forecasting students' performance within specific courses, including their grades and overall academic proficiency. By harnessing the power of predictive analytics, we endeavor to provide educators with invaluable insights into students' educational journeys, ultimately fostering an environment conducive to their holistic development and success. This process aids schools in identifying students who may require additional assistance or guidance. By leveraging these tools, educators can proactively detect potential issues and intervene early to support students in achieving their academic goals. Various methods, such as Decision Trees, Random Forests, and others, have proven effective in predicting student performance with high accuracy. Before further analysis, this information undergoes processing using specialized learning methods, enhancing its utility and relevance in guiding educational interventions and strategies.

II.LITERATURE REVIEW

[1] Demographic and personality data encompass various factors such as age, gender, parental education and income, emotional intelligence, student interests, motivation levels, communication skills, sports participation, hobbies, and ethnicity. These details are instrumental as they heavily influence students' motivation levels. However, obtaining accurate data in this category can be challenging, often leading to biases against students from specific backgrounds. Moreover, ethical permissions are required for collecting and analyzing such sensitive data. Institutional data pertains to the resources provided by educational institutions, including facilities, laboratory conditions, experiment status, infrastructure quality, teaching methodologies, transportation options, and communication mediums. This contextual information is crucial for understanding the learning environment. Factors such as resource availability for specific courses can significantly impact the accuracy of predictive models used in educational analysis.

[2] [3] Academic data serves as a cornerstone in predicting students' academic performance reliably. This dataset typically encompasses a student's Cumulative Grade Point Average (CGPA), internal assessment results, internal examination scores, and the specific courses opted for by the students. However, despite its significance, a notable challenge arises due to the non-uniformity inherent in this type of data across different educational institutions. This variability underscores the importance of tailoring predictive models and analysis techniques to the unique characteristics of each educational setting. By acknowledging and accounting for the nuances inherent in academic data across

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103021

institutions and courses, researchers and educators can develop more robust and contextually relevant predictive models. Such efforts are essential for ensuring the accuracy and effectiveness of interventions aimed at supporting student success and improving educational outcomes. Each institute employs its unique assessment methods, which introduces variability into the data. Consequently, a predictive model that demonstrates accuracy when trained on a dataset from one institute may not yield the same level of precision when applied to data from another institution. Moreover, the evaluation methods employed within a single institution may also vary across different courses, further complicating the predictive analysis. For instance, consider the scenario of secondary school data. Some institutions might have stringent admission criteria, resulting in a narrower range of variation in academic performance data. Conversely, institutions with more lenient admission standards may exhibit greater diversity in their data. Consequently, the reliability of specific features within the dataset may fluctuate depending on the institutional context.

[3] This dataset encompasses various aspects of student behavior, such as study patterns, attention span, engagement with study materials (particularly in flipped classroom settings), social interactions, time spent on social media, and participation in educational computer games. Understanding these features is crucial for accurately predicting student performance. Much of this data is subjective and gathered through surveys or questionnaires. Researchers often use this data as primary information or extract secondary insights for predictive purposes. To address this limitation, there is a pressing need to develop an effective predictive tool capable of assessing student performance before final assessments, thereby enhancing overall learning outcomes. Our aim in this study is to predict students' performance throughout the semester, offering opportunities for timely intervention and improvement. This objective can be achieved by incorporating dynamic features alongside other relevant factors to ensure accurate assessment and measured progress over time. Relying solely on student responses can introduce significant diversification, highlighting the need for precise mapping of various factors, including demographic and academic profiles, study habits, and academic benchmarks. While these factors can indeed forecast student performance to some extent, they may not adequately capture ongoing learning progress throughout the semester.

[4] Educational institutions gather diverse sets of student data, including demographic, academic, and educational information. Additionally, various programs and teaching-learning activities generate additional data, such as study behaviors, patterns, and participation in extracurricular activities. This wealth of data can be effectively utilized to predict student performance and establish correlations between its different features through the application of machine learning (ML) algorithms, provided it is processed using appropriate tools.

[5] These correlations, derived from the analysis of various student data sets, serve as powerful tools for predicting students' performance by the culmination of their academic program. This predictive capability is profoundly significant, as it empowers educators and institutions to adapt their approaches to the learning environment in a manner that is tailored to the individual needs of each student. By foreseeing how well students are likely to fare academically, educators can proactively provide customized academic assistance, guidance, and mentoring, thereby maximizing each student's potential for success. This process allows for informed decision-making regarding instructional practices, curriculum design, and resource allocation, with the ultimate aim of optimizing the learning experience for all students. Additionally, the predictive insights derived from student data analysis facilitate the provision of meaningful feedback to both students and educators alike. By understanding the factors that contribute to students' academic performance, educators can offer constructive feedback that is specific, actionable, and conducive to growth. Similarly, students can benefit from insights into their learning behaviors and patterns, allowing them to make informed decisions about their academic pursuits and areas for improvement.

[6] In academic research, there's a notable emphasis on evaluating student performance across various metrics, each shedding light on different aspects of their educational journey. Dropout rates, retention rates, and assignment performance are among the key indicators often scrutinized to gauge student success and engagement within educational programs. These metrics serve as quantitative measures of academic achievement, providing insights into the overall proficiency and competence of students in their chosen field of study. The CGPA, in particular, holds significant importance as it encapsulates a student's performance across multiple courses and semesters, offering a comprehensive overview of their academic standing.

III.METHODOLOGY OF PROPOSED SURVEY

To create a Prediction of student performance by using the machine learning model we will implement the following Basic steps:

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103021

• Introduction:
Begin by emphasizing the criticality of forecasting student performance and underscore the importance of employing machine learning models in educational contexts. Acknowledge the challenge of integrating university-provided data into meta-learning models directly.
• Data Collection:
Detail the dataset utilized for prediction, elucidating its origin and the variables it encompasses. Delve into the process of data cleaning and preprocessing, addressing strategies employed to address missing values and outliers effectively.
• Feature Selection:
In the realm of predictive modeling, feature selection plays a crucial role in enhancing the accuracy and efficiency of the model. The process involves identifying the subset of input variables (features) that are most relevant for making accurate predictions.
One commonly used criterion for feature selection is correlation analysis, which examines the relationships between different features and the target variable (in this case, student performance). Features that exhibit strong correlations with the target variable are considered important for prediction and are retained, while those with weak or insignificant correlations may be discarded.
• Model Selection:
The choice of a machine learning model significantly impacts the predictive performance of the system. Several factors influence the selection of an appropriate model, including the nature of the dataset, the complexity of the prediction task, and the computational resources available.
• Model Training:
Detail the procedure of partitioning the dataset into training and testing subsets, emphasizing techniques like cross-validation for robust model evaluation and hyperparameter tuning for optimizing model performance. The training dataset comprises an N-dimensional input vector containing various input features, such as the number of clicks on Virtual Learning Environment (VLE) activities until a student completes their initial course assessment.
• Interpretability and Explainability:
Explore methodologies for interpreting model predictions and comprehending the factors influencing student performance. Emphasize the importance of transparency and fairness in model predictions, particularly within educational contexts.
• Deployment and Integration:
Examine approaches for implementing the predictive model within educational environments. Explore potential hurdles concerning integration with current educational systems and workflows.
• Ethical Considerations:
Explore ethical dilemmas surrounding the prediction of student performance, including issues of bias and privacy. Propose strategies for mitigating bias and safeguarding privacy to ensure fairness and equity in model predictions.

Proposed Work:

In this project,
• SVM (Support Vector Machine): This algorithm is utilized to forecast student performance by identifying patterns within their data.
• Decision Tree: A model that assesses student performance by posing a series of questions about their data and making decisions accordingly.
• Random Forests: This ensemble method comprises multiple decision trees working collaboratively to enhance the accuracy of student performance predictions. Random Forest is an ensemble learning method that combines multiple decision trees to make more accurate predictions about student performance. Each decision tree in the forest independently makes predictions, and the final prediction is determined by averaging the predictions of all the trees or by taking a vote among them. This approach helps reduce overfitting and improves the overall predictive accuracy compared to a single decision tree.
• Multilayer Perception (MP): A form of artificial neural network adept at discerning patterns within student data to facilitate predictions.
• Linear Regression (LR): This straightforward model predicts student performance by fitting a straight line through their data points.
• Bagging and Boosting: These methodologies involve combining multiple models to improve the accuracy of predictions regarding student performance.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103021

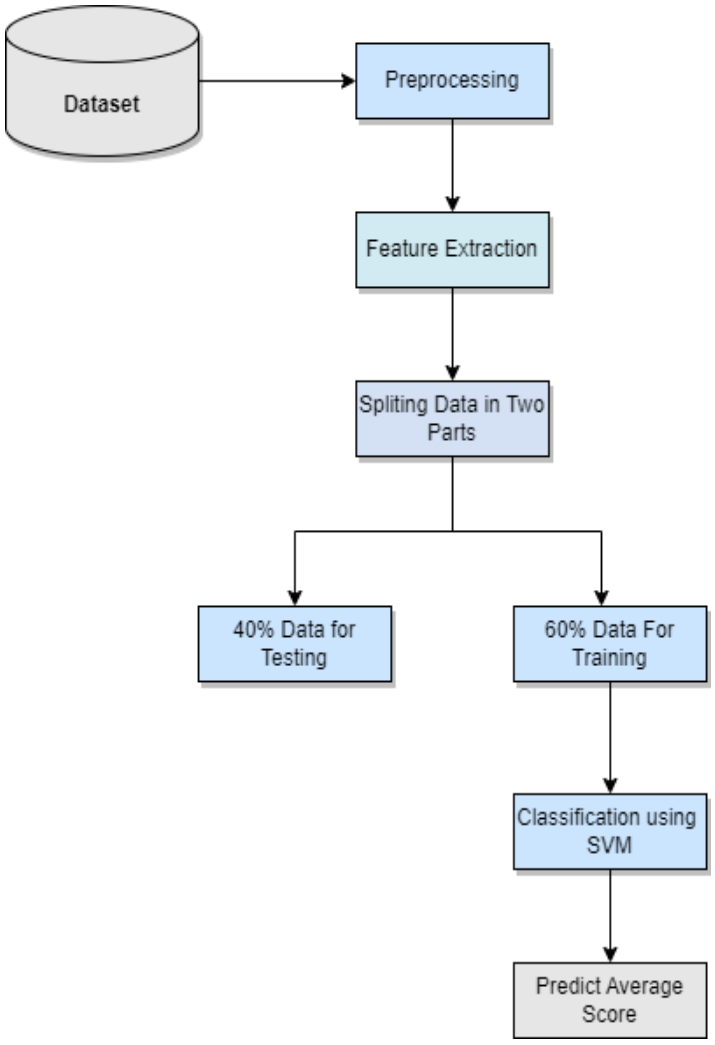

Fig 1. System Architecture Diagram

Performance Analysis

First, we gather lots of information about students, like their grades, how often they come to class and other important details.

Then, we figure out how important different classes are for predicting how well a student will do in other classes. For example, we see if being good at Math means someone will do well in other subjects too.

After that, we use fancy math to calculate "learning coefficients." These coefficients help us understand how doing well in one class can predict doing well in another.

Machine learning techniques are employed to construct a model predicting student performance based on their historical achievements and the significance of specific courses.

The model's efficacy is assessed by testing it with novel data, ensuring the accuracy of predictions.

Model refinement occurs based on evaluation results, involving adjustments to data selection and algorithmic approaches to enhance accuracy.

Upon validation, the model is deployed for utilization by students, educators, and educational institutions, facilitating real-world predictions of student performance.

Ongoing monitoring of model performance prompts updates as necessary, adapting to evolving student learning patterns and additional data availability.

The objective is continuous enhancement of the predictive system's accuracy and utility, achieved through sustained research efforts, collaboration with educators, and feedback from students to optimize effectiveness.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103021

IV.CONCLUSION AND FUTURE WORK

In this study the proposed work exploring the factors influencing students' performance in academic institutions is a compelling endeavor, offering educators insights to improve their teaching methodologies. To this end, we've devised a methodology meticulously analyzing students' attributes and prioritizing the most significant ones for constructing a prediction model. Our approach involves employing various techniques to identify key variables, which are subsequently utilized to develop Lasso linear regression models. Investigating the determinants of students' academic performance holds considerable promise for refining educational practices. Our proposed methodology, focused on discerning critical attributes and leveraging them for predictive modeling, lays a foundation for enhancing educational outcomes.

REFERENCES

[1] F. Qiu, G. Zhang, X. Sheng, L. Jiang, L. Zhu, Q. Xiang, B. Jiang, and P.-K. Chen, "Predicting students' performance in e-learning using learning process and behavior data," Sci. Rep., vol. 12, no. 1, p. 453, Jan. 2022.

[2] A. Hellas, P. Ihantola, A. Petersen, V. V. Ajanovski, M. Gutica, T. Hynninen, A. Knutas, J. Leinonen, C. Messom, and S. N. Liao, "Pre- dictating academic performance: A systematic literature review," in Proc. Companion 23rd Annu. ACM Conf. Innov. Technol. Comput. Sci. Educ., Jul. 2018, pp. 175-199.

[3] M. Bilal, M. Omar, W. Anwar, R. H. Bokhari, and G. S. Choi, "The role of demographic and academic features in a student performance prediction," Sci. Rep., vol. 12, no. 1, p. 12508, Jul. 2022.

[4] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student performance prediction using machine learning techniques," Educ. Sci., vol. 11, no. 9, p. 552, Sep. 2021.

[5] C. Romero and S. Ventura, "Guest editorial: Special issue on early pre-diction and supporting of learning performance," IEEE Trans, Learn.Technol., vol. 12, no. 2, pp. 145-147. Apr. 2019.

[6] Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," Educ.Inf. Technol., vol. 26, no. 1, pp. 205-240, Jan. 2021.

[7] F. Ofori, E. Maina, and R. Gitonga, "Using machine learning algorithms to predict students' performance and improve learning outcome: A literature-based review," J. Inf. Technol., vol. 4, no. 1, pp. 33-55, 2020.

[8] Abdala, G. A., Meira, M. D. D., Rodrigo, G. T., Fróes, M. B. D. C., Ferreira, M. S.,Abdala, S. A., & Koenig, H. G. (2021). Religion, age, education, lifestyle, and health: Structural equation modeling. Journal of Religion and Health, 60(1), 517-528.

[9] Rivas, A., González-Briones, A., Hernández, G., Prieto, J., & Chamoso, P. (2021). Artificial neural network analysis of the academic performance of students in virtual learning environments. Neurocomputing, 423, 713-720.

[10] 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, no. 11, pp. 388-396, 2020.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 **Impact Factor: 6.421**

www.ijarety.in editor.ijarety@gmail.com