Synergistic Performance Forecasting: Harnessing Gradient Boost and Linear Discriminant Analysis for Student Achievement Prediction

[1] R. Tamilkodi, [2] K.Valli Madhavi,[3*]J. Annie Christi [4] S. Revanth Kumar [5] M. Naga Pavan [6]A. Phani Manohar

[1] Professor, [2] Professor, [1]Department of CSE (AIML & CS),[2]Deaprtment of CSE
[1] Godavari Institute of Engineering & Technology, Rajahmundry, Andhra Pradesh, India
[2]GIET Degree College, Rajahmundry, Andhra Pradesh, India
[3 4 5 6] Department of Computer Science & Engineering (AIML & CS)
[3 4 5 6]Godavari Institute of Engineering & Technology, Rajahmundry, Andhra Pradesh, India
[1]tamil@giet.ac.in,[2]vallimadhavi@giet.ac.in
[3*]20551a4223.anniechristi@gmail.com [4]revanth.suvvari@gmail.com
[5]pawannaidu8575@gmail.com [6]aphanimanohar@gmail.com

Abstract. In the realm of student information systems, educational institutions grapple with the complexities posed by an ever-expanding repository of academic data, encompassing diverse files, records, and multimedia. Traditional statistical methods and database systems often fall short in managing the sheer volume of information, hindering the extraction of actionable insights. This study introduces an innovative tool a comprehensive framework rooted in a rule-based suggestion approach to tackle these challenges. The tool not only facilitates thorough analysis but also delves into predictive modeling of students' academic paths. By examining student demographics, academic performance metrics, and psychological attributes, the framework proves to be a valuable resource for educators, students, and parents alike. Leveraging robust data-mining techniques, the system significantly improves the accuracy of academic forecasting. Despite inherent limitations, it emerges as an invaluable guide. A detailed case study involving 1000 students highlights the superior performance of this approach compared to existing frameworks, confirming its effectiveness in assisting educational institutions. This tool stands out as a vital asset, unlocking the untapped potential within academic datasets and empowering educators to discern high-achieving and low-performing students.

Keywords: Analysis of Academic Performance, Statistical Approaches, Academic Prediction, Machine Learning, Classification Models, Rule-Based Recommendation.

1 Introduction

Within the realm of education, acknowledging each student as a unique asset to academic institutions is pivotal for cultivating graduates of outstanding proficiency. The evaluation of student performance is essential, encompassing diverse modalities such

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_97

as examinations and assessments [1]. However, the dynamic and multifaceted landscape of academic achievement presents challenges, particularly with the exponential growth of colleges and higher education institutions over the past decade [2]. Despite pedagogical advancements, institutions grapple with issues like student attrition, subpar academic performance, and post-graduation unemployment. Understanding the intricate factors influencing academic performance and behavioral trends is an ongoing challenge. Although educational organizations collect substantial data, its underutilization hampers tangible improvements in student outcomes [3].

While the transformative impact of data mining and predictive analytics has been evident across various sectors, their potential within the realm of education remains largely unexplored. The ever-increasing volume of student data accentuates the need for a dedicated tool proficient in managing this information, providing educators with the capability to precisely anticipate student performance [4]. By precisely matching the quality of educational assistance with the wide range of student profiles, this technology not only tackles the problems facing education now but also represents an investment in the future. [5].

The application of predictive analytics algorithms to this wealth of data empowers instructors to extract valuable insights into patterns, trends, and potential risk factors influencing student progress. For instance, predictive models can discern early indications of academic challenges, facilitating timely interventions and the implementation of specialized support systems tailored to meet individual student needs [6]. Moreover, these technologies play a pivotal role in optimizing course designs, curricula, and teaching methodologies by evaluating the effectiveness of various approaches in enhancing student outcomes [7].

2 Literature Survey

Educational Data Mining (EDM) places a focus on predictive analysis, prioritizing anticipation over precise outcome generation. It uses Machine Learning (ML) models to tackle classification and regression problems in higher education, hoping to avoid course failure and provide data-driven insights to improve university management[8]. The effectiveness of machine learning methodologies in predicting student performance has gained traction recently, albeit demanding expertise in computer science. By providing thorough guidelines for instructors interested in using data mining approaches to predict student progress, this study seeks to address this skill gap[9].

The importance of predictive analytics using machine learning is on the rise in universities, particularly in assessing students' academic performance, notably their grades. This research probes into the evaluation of ML algorithms for predicting students' final marks, grappling with challenges posed by unbalanced datasets [10]. Many other elements influence a private university's ability to enroll students and help them succeed in the future. In order to determine graduation labels, this study looks at the relationship between students' length of enrollment and their academic accomplishments, focusing on variables like IP/GPA, Selection Test results, high school of origin, and gender[11].

The increasing amount of data produced in the educational environment as a result of technology breakthroughs calls for the use of big data methodologies. The big data paradigms in education are thoroughly reviewed in this article, which also addresses data sources from learning management systems, massively open online courses, and other educational platforms and systems[12]. Combining machine learning with Educational Data Mining (EDM) has the potential to completely change our understanding of student behavior, academic performance, and overall institutional success[13].

Although predictive analytics shows promise, implementing these strategies requires a comprehensive framework that is suited to the demands of educators[14]. These results highlight the need for interdisciplinary cooperation between academics, data scientists, and educators as academic institutions struggle with the complexities of graduation rates, student performance prediction, and the use of enormous amounts of data. This partnership may lead to the development of particular solutions that improve learning environments and raise student accomplishment[15].

The Student Performance Predictor stands as a valuable tool for forecasting students' academic achievements, streamlining data collection efforts, and providing students with insights into their upcoming exam scores. Widely embraced in educational contexts, this model leverages robust algorithms, including the Decision Tree classifier, Random Forest, and K-nearest Neighbor algorithms, to deliver predictions. However, it is noteworthy that the model has achieved an accuracy rate ranging between 80% and 85%, showcasing its efficacy while acknowledging the potential for improvement. A significant challenge faced by the model is the issue of over-fitting, which needs to be addressed to enhance its predictive capabilities and ensure more accurate results in the future.

3 Proposed System

In the sphere of education, students represent a central asset for academic institutions, embodying the capacity to yield graduates of outstanding caliber proficient in academics, practical knowledge, self-development, and inventive thinking. Realizing these aspirations relies on a thorough and all-encompassing examination of student performance within the educational domain. By giving parents, teachers, and students access to a wealth of information, the suggested approach seeks to improve academic forecasts' accuracy by utilizing a variety of strong data-mining techniques. By carefully examining each student's personal information, academic record, and psychological characteristics, this technology aims to identify areas of weakness and provide insightful advice that will support each student's growth. An extensive case study including one thousand students offers a compelling illustration of the system's efficacy, showcasing its advantages over existing frameworks and verifying its capacity to usher in a new era of improved educational outcomes.

3.1 Design Methodology:

The workflow starts with meticulous input data preprocessing, ensuring dataset quality. Feature extraction unveils salient student attributes, and the dataset is partitioned for model development. Advanced classification algorithms construct a predictive model, rigorously assessed using quantitative metrics like accuracy and precision. The model's is used for informed decision-making, educational planning, and targeted interventions. The workflow initiates with precise data preprocessing, refining dataset integrity for analysis. Extracting pivotal student attributes shapes the model's understanding of relevant factors. Partitioning the dataset enables tailored model development and unbiased evaluation. Rigorous algorithmic construction and assessment validate the model's accuracy for informed educational decisions and interventions. Fig 1 shows the block diagram for student performance prediction.

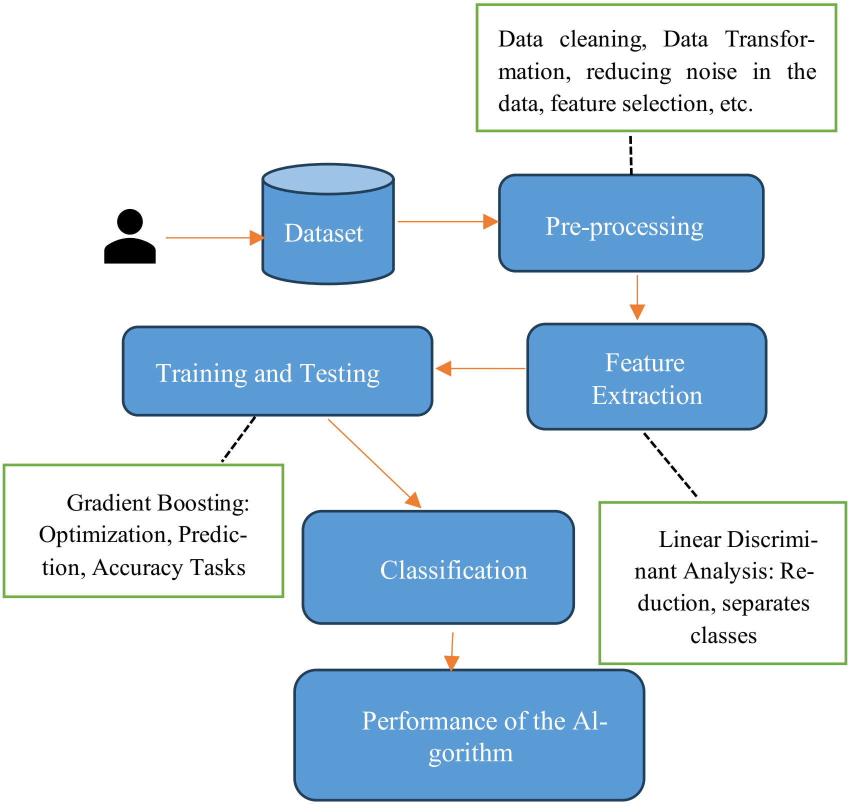

Fig. 1. Block diagram for Student Performance Predictor

Steps Involved are: Dataset Collection- Gather relevant student data, including demographics and academic history. Data Preprocessing- Cleanse and prepare the dataset, ad dressing outliers, missing data, and ensuring consistency. Feature Extraction: Extract key features impacting student performance, e.g., attendance, grades, Train-Test Split: Divide the dataset for model development and assessment into training and testing subsets. Model Training- Facilitate the training of classification models (such as gradient boosting and LDA) to classify students according to their distinct features. Model Testing- Implement the trained models on the testing dataset to evaluate their effectiveness in handling unseen data and gauging performance under novel conditions. Classification: Categorize students into predefined performance groups using trained models.Performance Evaluation: Assess model performance with metrics like accuracy and precision for effectiveness.

Proposed System Algorithms:

1. **Gradient Boosting:**

A technique employed during training and testing for enhanced predictive accuracy:
-Compute the target variable's mean.
-Compute residuals for each sample.
-Construct decision trees anticipating residuals.
-Estimate target labels using each tree.
-Calculate new residuals.
-Repeat steps 3-5 for a given number of iterations.

2. **Linear Discriminant Analysis (LDA):**

Utilized in the feature extraction phase:
-Determine mean vectors for each class.
-Compute within-class and between-class scatter matrices.
-Compute eigenvalues and eigenvectors for each scatter matrix.
-Sort and select top eigenvectors to create a matrix.
-Apply transformation to samples using the eigenvector matrix.

4 Results and Discussions

We used a dataset of 1000 student records that came from various sources for our analysis. To ensure that the dataset was balanced and appropriate for classification tasks, we set aside 30% for testing and carefully assigned 70% of the data for training. Remarkably, the results showed 90.94% accuracy and 94.44% precision. The model, subjected to rigorous training and testing on this dataset, demonstrated improved accuracy, effective mitigation of overfitting, and the delivery of reliable prediction results. The emphasis on meticulous training and testing underscores the commitment to developing a high-quality model. The mention of mitigating overfitting indicates the model's fine-tuning for optimal generalization beyond the training data. In essence, this research not only highlights systematic data handling and model development but also accentuates the robustness and practical applicability of the developed model in proficiently classifying student-related data. The detailed approach in curating the dataset, training the model, and analyzing results positions this work as a noteworthy contribution to the domain of machine learning applied to the analysis of student records. Table 1 shows the parameters which is considered for the parameters present in the proposed dataset. Fig 2 represents analysis of student performance; the results show that the student's performance is good.

4.1 Experimental Results

Table 1: Shows the parameters present in the proposed Dataset

Input	Input Values
School	GP
Gender	1
Age	20
Address	0
Famsize	3
Medu	3
Fedu	1
Mjob	1
Fjob	1
Reason	1
Famrel	1
Higher	0
Activities	0
Failures	1
Attendance	1

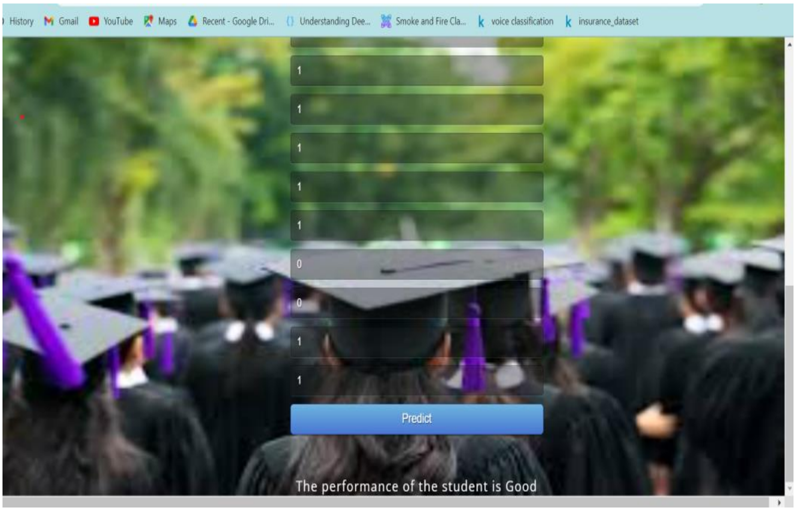

Fig. 2. The result of the student performance predictor

Average = (G1+G2+G3)/3

- If Average is between 1-5
  The student is low performer
- If Average is between 6-10
  The student medium performer
- If Average is between 11-14
  The student is high performer

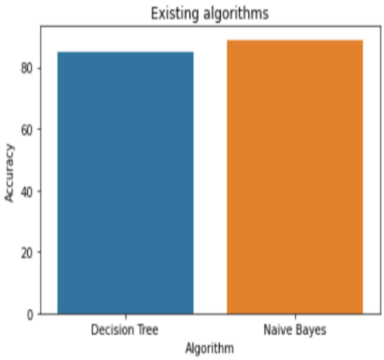

Fig 3: Accuracy of Existing Algorithms (80-85%)

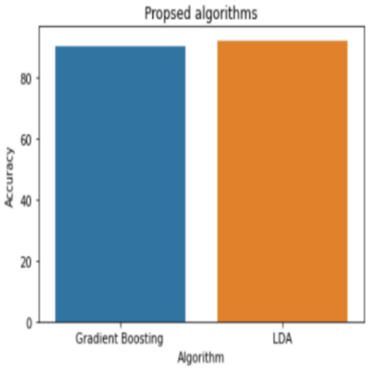

Fig 4: Accuracy of the Proposed Algorithms (> 90%)

Fig 3 and Fig 4 shows the difference between the accuracy of Existing algorithms and proposed algorithms. Table 2 represents the difference between the accuracy of existing algorithms and proposed algorithms.

Table 2: Table represents the Accuracy of Existing Algorithms and proposed Algorithms

Algorithms	Decision-Tree	Naïve- Bayes	Linear Discriminant Algorithm	Gradient Boosting
Merceron, A. and Yacef	85%	84%	-	-
Alaa el-Halees	84%	82%	-	-
Pandey, U. K. and Pal	86%	85%	-	-
Amjad Abu Saa	83%	81%	-	-
Pappano, Laura	85%	82%	-	-
Proposed Work	-	-	90%	92%

5 Conclusion

In conclusion, our study showcases remarkable expertise in predicting students' performance, achieving an outstanding accuracy of 90.94% and a precision of 94.44%. This milestone represents a noteworthy advancement in the realm of education, delivering valuable insights to diverse stakeholders. Despite this success, a critical imperative emerges to heighten accuracy levels further by integrating additional components like attendance records and psychological insights. This enhancement aims to offer a more holistic perspective on students' academic trajectories, ensuring a comprehensive and nuanced approach to performance forecasting. As the educational landscape continues to evolve, the continual refinement of predictive models with a broader array of factors promises to contribute significantly to the ongoing improvement of educational outcomes and the effectiveness of support systems for students.

6 References

1. Brijesh Kumar Pandey, U. K. and Pal, S., "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), 2011, 686- 690, ISSN: 0975-9646.
2. Pappano, Laura. "The Year of the MOOC." The New York Times 2.12 (2012): 2012.
3. Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi and Mohammed L. Ali Shatnawi, "Stock price Prediction Using K-Nearest Neighbor (knn) Algorithm" Vol. 3 No. 3; March 2013
4. El Ahrache, Sara Ibn, et al. "Massive Open Online Courses: a new dawn for higher education?" International Journal on Computer Science and Engineering 5.5 (2013): 323.

5. Ramesh, Arti, et al. "Uncovering hidden engagement patterns for predicting learner performance in MOOCs." Proceedings of the first ACM conference on Learning@ scale conference. ACM, 2014.
6. DeBoer, Jennifer, and Lori Breslow. "Tracking progress: predictors of students' weekly achievement during a circuits and electronics MOOC." Proceedings of the first ACM conference on Learning@ scale conference. ACM, 2014.
7. Wolff, Annika, et al. "Predicting student performance from combined data sources." Educational Data Mining. Springer International Publishing, 2014. 175-202.
8. Srinivas, K., Gagana Sri, R., Pravallika, K. et al. COVID-19 prediction based on hybrid Inception V3 with VGG16 using chest X-ray images. Multimed Tools Appl 83, 36665–36682 (2024). https://doi.org/10.1007/s11042-023-15903-y
9. Ashenafi, Michael Mogessie, Giuseppe Riccardi, and Marco Ronchetti. "Predicting students' final exam scores from their course activities." Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE. IEEE, 2015.
10. KlÃijsener, Marcus, and Albrecht Fortenbacher. "Predicting students' success based on forum activities in MOOCs." Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2015 IEEE 8th International Conference on. Vol. 2. IEEE, 2015.
11. BlagojeviÄ ˇG, Marija, and Danijela MiloÅ ¿aeviÄˇ G. "Massive Open On- line Courses: edX vs Moodle MOOC." 5th International Conference on Information Society and Technology, Kopaonik. 2015.
12. Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat," Predicting IT Employability Using Data Mining Techniques," in third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), 2016.
13. Amjad Abu Saa. "Educational Data Mining & Students' Performance Prediction". (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016.
14. Chitteti, Chengamma, and K. Reddy Madhavi. "Taylor African vulture optimization algorithm with hybrid deep convolution neural network for image captioning system." Multimedia Tools and Applications (2024): 1-19.
15. Nafis Neehal, Book- "Machine Learning Algorithm", 2018(Bangla), Page-132, ISBN: 978984-8042-02- 1.