International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

LEVERAGING ARTIFICIAL INTELLIGENCE FOR PREDICTIVE LEARNING ANALYTICS IN ENHANCING STUDENT PERFORMANCE IN THE EDUCATION SECTOR

Akshay Rajshekar Shiraguppi
University of Illinois at Springfield
Springfield, Illinois
akshayrs1993@gmail.com

Abstract

In today's rapidly evolving educational landscape, leveraging technology to enhance student outcomes has become crucial. In order to ensure that students have a satisfactory learning experience, teachers need early signs of their development so that they may maximize their learning tactics and concentrate on different instructional practices. Using machine learning, teachers may foresee their students' likely areas of weakness in the learning process and proactively work with them to improve their learning outcomes. This study explores the effectiveness of AI-driven models in predicting student outcomes using the Student Study Performance dataset from Kaggle. The research involves comprehensive data preprocessing, including handling missing values, label encoding, and feature engineering. Important performance metrics, including F1-score, recall, precision, and accuracy, are used for implementing and assessing machine learning models like neural networks, random trees, and logistic regression. The most reliable technique for forecasting student performance is logistic regression, according to experimental data, which shows that it achieves 95.50% accuracy, 97.21% precision, and an F1-score of 95.91%. The research demonstrates how AI predictive analytics functions as a base to enhance educational results when using data-driven selections.

Keywords: Student Performance Prediction, Educational Data, Academic Success, Personalized Learning, Machine Learning, Predictive Learning Analytics, Student study performance data.

I. INTRODUCTION

In the modern era, social-economic development shows growing dependence on educational advancements [1]. The quality of education received by the people functions as an essential engine of national development, which shapes readiness for workforce jobs, advances technology development and strengthens economic development. A knowledge-based economy development requires the education sector to provide essential skills and competencies for individuals who need success in global competition. The education sector needs regular development of modern teaching approaches along with better assessment methods and improved student assistance systems to obtain maximum learning results. The education industry faces extensive modification because digital technologies now combine with data-driven methods. The basic educational methods used for teaching and evaluating students through standardized exams and assignments fail to grasp the various learning needs that students possess [2]. Educational institutions now

International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

focus more on technological implementations to create tailored learning content because academic support requirements and educational effectiveness need additional enhancement. The major struggle in this field involves recognizing students who show signs of academic difficulties before they fail, along with delivering swift aid to help them succeed academically[3].

Student academic outcomes depend on multiple elements that consist of mental capabilities alongside financial status and mental health status, as well as the learning circumstances. Educators, along with institutions and policymakers, require a thorough comprehension of such influencing factors to improve academic achievement [4]. The ability to predict student success with precision grants instructors the opportunity to spot students who encounter difficulties so they can provide suitable intervention strategies[5]. The educational system should provide guidance to high-performing students for accessing advanced educational options to maximize their learning potential. Predictive learning analytics experienced a transformation through AI which delivers improved solutions to evaluate extensive educational databases[6][7]. In order to forecast student performance outcomes, machine learning (ML) approaches can analyze past academic records, behavioral trends, and engagement measures. Educational institutions may improve student engagement, provide individualized learning experiences, and make data-driven choices by utilizing AI-driven models. In order to improve overall learning efficiency, AI-powered systems may dynamically adjust to the demands of each individual student, suggest customized study schedules, and issue early warnings for academic difficulties.

A. Motivation and Contribution of the Study

Modern educational institutions require innovative approaches to use data analysis for decision-making which leads to a need for better student outcome enhancement. The current identification methods for at-risk students depend mainly on manual evaluations as well as simple statistical methods that miss fundamental relationships within educational data sets. The implementation of artificial intelligence in predictive learning analytics generates a transformative answer through its capability to identify unknown connections, which helps teachers deliver preventive support measures. However, challenges such as data pre-processing, feature engineering, and model optimization remain critical to ensuring accurate and reliable predictions. This study is motivated by the necessity to develop an AI-powered framework that improves predictive learning analytics, helping educators personalize learning strategies and support students effectively. The contributions of this study are as follows:

- Experiment performed on the student performance data from student records, assessments, and behavioural patterns collected from online learning platforms.
- Handling missing values, label encoding, and Min-Max Scaling for improved model accuracy.
- Enhances predictive performance by incorporating composite metrics, such as total scores, to capture nuanced academic patterns.
- Takes use of several AI technologies, such as neural networks, logistic regression, and Random Forest, to classify student performance and identify at-risk learners.
- Makes sure that predictive models are robustly assessed by evaluating their performance using measures such as F1-Score, Accuracy, Precision, and Recall.

B. Structure of the Paper

The following is the paper's structure: Section II explores related studies on student performance enhancement. Section III discusses the proposed approach used in the study. Section IV provides a comparative analysis of model performance with visual representations. Finally, Section V summarizes key findings and offers future research directions.

II. LITERATURE REVIEW

In this section, the literature review on machine learning methodologies and strategies for predicting student performance is presented.

Amrieh, Hamtini and Aljarah (2016) provide a new model for predicting a student's behavioural characteristics are extra data attributes and features that are included in a student's performance based on data mining techniques. random forest (RF), bagging, and boosting—all of which are popular ensemble techniques in the literature. The accuracy of the suggested model with behavioural features was improved by up to 22.1% when compared to the results when those features were removed and by as much as 25.8% when employing group methods. Tested on inexperienced students, the model achieved an acquired accuracy of over 80%. This outcome demonstrates the suggested model's dependability [8].

Mueen et al. (2016) Make use of data mining techniques to predict and analyze students' academic performance based on their academic background and forum participation. Data from two undergraduate classes was gathered for this investigation. Three different data mining classification algorithms—Neuro Network, Decision Tree, and Naïve Bayes—were used for the dataset. With an overall prediction accuracy of 86%, the Naïve Bayes classifier was shown to perform better than the other two classifiers [9].

Purwaningsih and Suwarno (2016) forecast a student's success in vocational school by looking at their motivation. The naïve Bayes method is used to classify data based on input parameters that describe motivation level. The association between students' motivation and achievement is determined using the RMSE value derived from the categorization experiments. The n-ACH and GPA variables had RMSE values of 0.3696 and 0.4049, respectively, according to the results [10].

Amrieh, Hamtini and Aljarah (2015) proposed a new class of characteristics for a new student performance model called behavioural features. They obtain the information from the e-learning platform Kalboard 360 via the Experience API Web service (XAPI). Next, they use data mining techniques such as Artificial Neural Network, Naïve Bayesian, and Decision Tree classifiers to evaluate how these variables affect students' academic achievement. Results using different classification algorithms containing these parameters indicated an increase in classification accuracy of up to 29% when compared to the same data set when behavioural aspects were omitted [11].

Ahmad, Ismail and Aziz (2015) provide a paradigm for forecasting the academic achievement of first-year computer science students pursuing bachelor's degrees. To determine the optimal prediction model for students' academic performance, the data is put through Decision Tree, Naïve

International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

Bayes, and Rule Based categorization techniques. As the experiment's outcome demonstrates, the Rule Based model outperforms the other approaches with the greatest accuracy value of 71.3%[12]. Sorour et al. (2014) use student-written, free-form comments following each class to forecast performance. Semantic information is extracted from student comments using the LSA latent semantic analysis approach, which uses statistically determined conceptual indexes instead of specific words. Then, an ANN model is applied to the analyzed comments to predict students' performance, revealing the high accuracy of predicting students' grades. To anticipate a student's ultimate score, they used five grades rather than the mark itself. The F-measure of students' grades and the average prediction accuracy for their proposed method is 76.1% and 82.6%, respectively [13].

The reviewed studies summaries in Table I use a variety of machine learning techniques, such as deep learning, ensemble methods, and Naïve Bayes, to predict student performance using academic, behavioural, and textual data, achieving accuracies up to 86%. Future research should focus on larger datasets, deep learning approaches, and multimodal data integration for enhanced predictive accuracy.

TABLE I. SUMMARY OF THE PREVIOUS STUDIES ON STUDENT PERFORMANCE PREDICTION USING MACHINE LEARNING

References	Methodology	Dataset	Performance	Limitations & Future Work
Amrieh, Hamtini & Aljarah (2016)	ANN, Naïve Bayesian, DT, Ensemble methods (Bagging, Boosting, Random Forest)	e-Learning management system data	Using group techniques, accuracy can increase by up to 25.8%; with novice students, it can increase by 80%	Restricted to behavioral features, future work should test generalization on other platforms and add more sophisticated feature engineering techniques.
Mueen et al. (2016)	Decision trees, neural networks, and Naïve Bayes	Students' data from two undergraduate courses	Naïve Bayes achieved 86% accuracy	Small sample size; future work should expand dataset and explore deep learning models.
Purwaningsih & Suwarno (2016)	Naïve Bayes	Vocational education students' motivation data	RMSE for n-ACH: 0.3696, GPA: 0.4049	Limited variables for motivation; future work could incorporate psychological and engagement metrics.
Amrieh, Hamtini & Aljarah (2015)	Naïve Bayesian, Artificial Neural Network, and Decision Tree	Information obtained using XAPI from the Kalboard 360 e-Learning system	Up to 29% improvement in accuracy with behavioral features	Limited to behavioral features from Kalboard 360, future work could explore integrating more diverse data sources and advanced classifiers.
Ahmad, Ismail & Aziz (2015)	Naïve Bayes, Rule-Based Classification, and Decision Tree	Data from 8-year period of Computer Science Bachelor students	Rule-Based method achieved 71.3% accuracy	Focused on first-year students only, future research could include multi-year student data and hybrid models.

International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

Sorour et al. (2014)	LSA, or latent semantic analysis, and artificial neural networks (ANN)	Students' free-form remarks following every class (Student Comments Dataset)	82.6% prediction accuracy, F-measure: 76.1%	Focused on text-based data, future work could combine text analysis with quantitative academic data.

III. METHODOLOGY

The methodology for this research analyses and predicts student performance based on various academic and demographic features. The investigation starts by gathering information from the Student Investigation Performance dataset, which was obtained from Kaggle, followed by extensive Pre-processing data to address missing values, label encoding categorical features and Feature engineering is conducted by creating composite metrics such as total scores to enhance predictive power. An 80:20 ratio split approach is used to separate the available dataset into training and testing halves for accurate model evaluation. AI-driven models, including Random Forest, Neural Networks, and Logistic Regression, are implemented to classify and predict student performance. To evaluate model performance, evaluation metrics such as F1-score, recall, accuracy, and precision are used. The insights derived from these predictive models support data-driven decision-making in educational institutions, enabling targeted interventions to enhance student learning outcomes. Figure 1 illustrates the methodology steps.

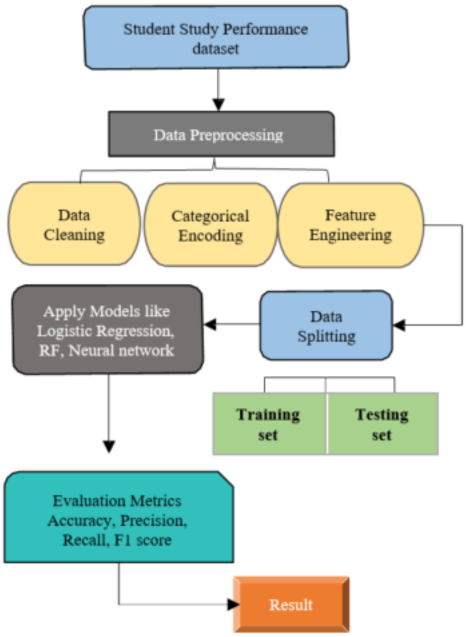

Fig. 1. Flowchart for Enhancing the Student Performance

Each step of the flowchart in Figure 1 is elaborate in the next section.

A. Data Collection

The Student Study Performance dataset utilized in this study was obtained from Kaggle and contains a variety of student demographic, academic, and parental data. The dataset's salient characteristics include test preparation course, lunch type, gender, race/ethnicity, parental

International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

educational attainment, and reading, writing, and math results. EDA was used to comprehend the data distribution and find important connections and patterns. This included Figure 2 displays the results of calculating summary statistics for each characteristic.

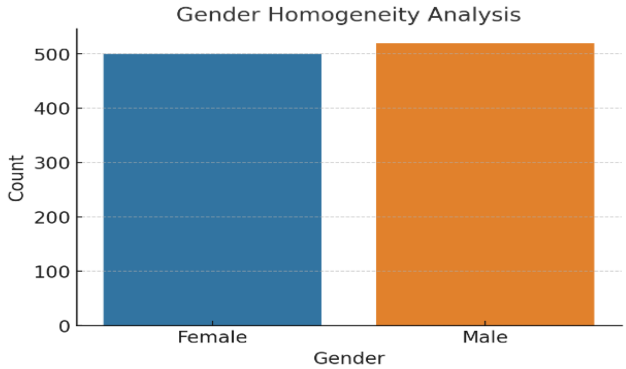

Fig. 2. Bar Graph for Gender homogeneity analysis

The bar chart illustrates in Figure 2 the gender homogeneity analysis, comparing the counts of male and female participants. The data reveals a nearly equal distribution, with females having a count of 500 and males slightly higher at 520. The chart uses distinct colors for better differentiation and includes gridlines for enhanced readability. The balanced representation suggests minimal gender disparity in the dataset, ensuring fairness in gender-based analyses.

B. Data Pre-processing

Data preparation is essential to machine learning and cannot be emphasized enough. Data pre-processing is a cleaning process that creates a tidy, well-structured dataset from unstructured raw data that may be utilized for more study. The following data data pre-processed in various steps that are listed in below:

C. Data Cleaning

Data cleaning entails a number of procedures, including addressing discrepancies, detecting or eliminating outliers, smoothing noisy data, and filling in missing information [14]. The cleansed data is then converted into an appropriate tabular format.

D. Categorical Encoding

Categorical data and special characters are transformed into numerical values in order to enhance model performance. Numerical data types are created from categorical data types using label encoding techniques. Using label encoding, categorical data is converted into numerical data by assigning a unique number label to each category.

E. Feature engineering for Creating new features

Created new features to enhance predictive power, such as calculating the total score by summing math, reading, and writing scores. The total score of features is depicted in Figure 3.

International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

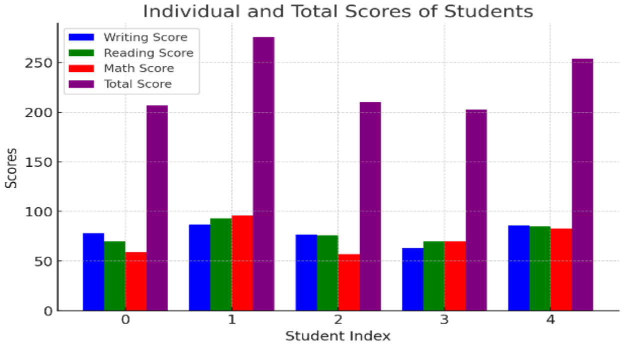

Fig. 3. Bar Graph for Total score of students

The bar chart in Figure 3 displays Writing, Reading, Math, and Total Scores for students. Total Scores are: Student 0 (207), Student 1 (276), Student 2 (210), Student 3 (203), and Student 4 (254). Student 1 has the highest score (276), and Student 3 has the lowest (203), highlighting performance variations.

F. Data Splitting

Data splitting partitions a dataset into training and test sets for machine learning evaluation; an 80:20 data split allocates 80% for training AI models to learn student performance patterns and 20% for testing, ensuring objective evaluation and reliable predictions to support data-driven educational interventions.

G. Classification of Logistic Regression Model

A linear model for binary classification is called logistic regression. In this case, it uses input characteristics to forecast the likelihood of a student's performance category[15]. For the following, assume that every data vector xi has an extra component 1. This will make notation simpler by allowing us to write a simple dot product an α.x for a linear combination of vector components instead of the more complex an α.x+α_0. A logistic regression model is frequently used to calculate the class membership probability for one of the two categories in the data set, as indicated by Equation 1:

$$P(1|x,\alpha) = \frac{1}{1+e^{-(\alpha.x)}} \quad (1)$$

and P(0│x,α)=1-P(1│x,α). Writing P(1│x,α) makes it clear how the parameters affect the posterior distribution. When both the class-conditional densities, p(x│1) and p(x│0), have the same covariance matrices and are multinomial, it may be demonstrated that this model is accurate. The equation an α is satisfied by the all-points-x hyperplane. P(1│x,α)=P(1│x,α)=0.5 defines the decision boundary between the two classes, which is generated by x=0. A logistic regression model that just includes the original variables is called a main effects model; the model becomes more flexible by including interaction terms like products, which cause the covariates to be nonlinear. Greater flexibility may be preferable overall, but it also increases the possibility of model overfitting or "memorizing the training cases," which may lower a model's accuracy in situations that haven't been seen before. Fitting the training instances is only one aspect of predictive modeling; the primary objective is accurately identifying incoming cases.

H. Classification of Performance Measures

The confusion matrix displayed in Table II has been utilized to assess the efficiency of the categorization models used in this investigation. An effective method for measuring results and determining the accurate categorization rate is A matrix of perplexity. These are the confusion matrix's values:

- **True Positive (TP):** The percentage of samples that were marked as positive both in anticipation and in fact.
- **True Negative (TN):** The number of samples that had both expected and actual negative labels.
- **False Positive (FP):** The quantity of samples that were anticipated to be positive but had actual labels that were negative. (Error type 1).
- **False Negative (FN):** The number of samples that were projected to be negative but were actually labelled as positive. (Error type 2).

TABLE II. FOR A TWO-CLASS CLASSIFICATION PROBLEM CONFUSION MATRIX

		Predicted Class		
		Positive	***Negative***	***Total***
Actual Class	Positive Negative	TP FP	FN TN	P N
	Total	P′	N′	P+N

Accuracy: The evaluation of classification models occurs through accuracy as a performance metric standard[16]. The measurement calculates correct predictions as a ratio within the total instance count as follows in Equation 2:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad \textbf{(2)}$$

Precision: The precision measures the percentage of binaries that were found to be failure prone but were categorized as such, thereby quantifying the type I mistakes. Precision is computed as follows in Equation 3:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall: Recall is the proportion of binaries that are actually prone to failure and are identified as such (type II faults). This is calculated as shown in Equation 4:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F1 Score: The F score is a statistic that has been used traditionally to evaluate both accuracy and recall. For a given model, It is determined by taking the accuracy's harmonic mean and recall in Equation 5.

International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

ROC curve: A non-parametric method of assessing 2-class discriminant models is using receiver operating characteristic (ROC) curves. The real positive rate is shown against the false positive rate on the curve. The formula is presented in Equation 6.

$$AUC = \int_0^1 TPR(x)dx \quad (6)$$

The values suggest that the model fits the training data remarkably well. On the testing dataset, the model demonstrated strong generalization and produced impressive results.

IV. RESULT & DISCUSSION

The equipment used for the trials has two Nvidia Tesla K40 12GB GPUs, 64GB of RAM, and a dual Intel E5-2680 CPU. Even with 12GB of RAM on each Tesla K40 GPU, the implementation is carried out in a Python module that utilizes GPU capabilities in an efficient manner. A student performance dataset serves as the basis for the performance evaluation, using important measures including F1-score, recall, accuracy/loss, and precision.

TABLE III. PERFORMANCE OF LOGISTIC REGRESSION MODEL FOR ENHANCING STUDENT PERFORMANCE

Performance Metrics	Logistic Regression
Accuracy	95.50
Precision	97.21
Recall	94.80
F1-score	95.91

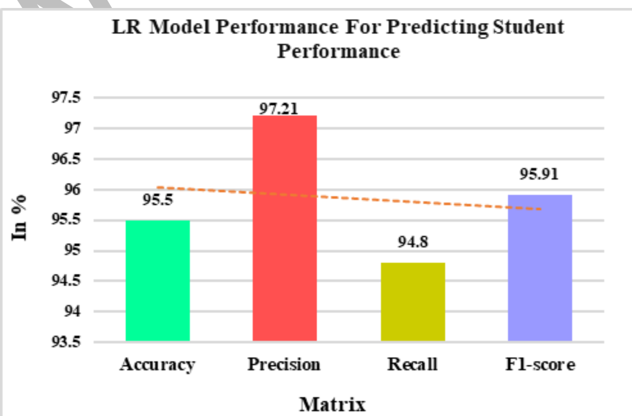

Fig. 4. Bar Graph for LR model performance

Table III and Figure 4 demonstrate the strong performance of an LR model used for predictive learning analytics in education. With 95.5% accuracy, the model correctly predicts student

International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

outcomes most of the time. Its high precision of 97.21% and recall of 94.8% suggest reliable identification of both struggling and successful students. The robustness of the model is further shown by the balanced F1-score of 95.91%. These results imply that educators could leverage this AI-driven approach to proactively identify at-risk students, personalize learning, and ultimately improve educational outcomes.

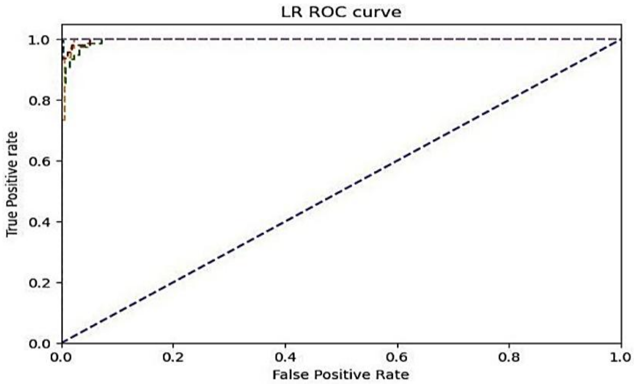

Fig. 5. ROC Curve for LR model

The ROC curve depicted in Figure 5, plots the TPR against the FPR. Different colored lines represent variations of an LR model, aiming to predict a binary outcome (e.g., at-risk vs. not at-risk). The prediction power of the model increases as a curve gets closer to the upper-left corner. The diagonal dotted line indicates a random classifier. In education, this visualization aids in optimizing models that identify struggling students, enabling timely interventions and personalized support to enhance overall student performance.

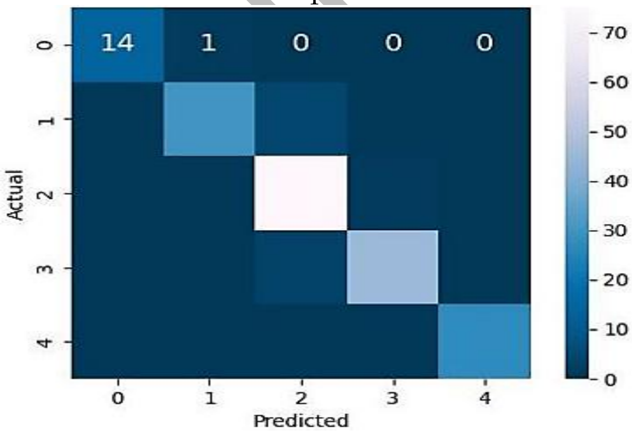

Fig. 6. Confusion Matrix for LR Model

The matrix visualizes in Figure 6, The quantity of genuine positive, real negative, false positive, and false negative forecasts. In this context, it shows how well a model predicts student categories (0 through 4). Correct predictions are represented by diagonal cells, whereas mistakes are indicated by off-diagonal cells. For instance, 14 students in Category 0 were correctly predicted, but 1 student in Category 1 was incorrectly assigned to Category 0.

International Journal Of Core Engineering & Management
Volume-4, Issue-2, May-2017, ISSN No: 2348-9510

A. Comparative Analysis and Discussion

The comparison between propose model LR with existing models such as Random Tree [17], Neural net [18], performance based on performance matrix are provide in Table IV.

TABLE IV. COMPARISON OF EVALUATION OF MODELS FOR ENHANCING STUDENT PERFORMANCE

Model	Accuracy	Precision	Recall	F1-score
Random Tree [17]	94.4	94.5	94.4	94.4
Neural net [18]	75.28	75.63	75.42	75.48
Logistic Regression	95.50	97.21	94.80	95.91

Table IV presents a comparison between proposed and existing model performance. In this comparison, the Logistic Regression model demonstrates the highest accuracy 95.50% and precision 97.21%, along with a strong recall 94.80% and F1-score 95.91%, making it the most effective among the evaluated models. The Random Tree model follows closely with 94.4% accuracy, maintaining balanced precision, recall, and F1-score values of 94.5%, 94.4%, and 94.4%, respectively. In contrast, the Neural Net model underperforms, achieving the lowest accuracy, 75.28% and less than ideal F1-score of 75.48%, recall of 75.42%, and accuracy of 75.63%, relatively lower precision, 75.63%, recall of 75.42%, and F1-score, 75.48%. This comparison highlights the superior predictive capability of Logistic Regression for student performance enhancement.

The proposed AI-driven approach for student performance prediction offers several advantages. The model uses Logistic Regression methods to identify students at risk with 95.50% accuracy and 97.21% precision, thus ensuring reliable detection results. The model delivers flexible interpretation and scalability, which enables its simple incorporation into various educational organizations. A practical, real-world integration becomes possible since the efficient computational implementation supports existing educational analytics systems without hurdles.

V. CONCLUSION & FUTURE WORK

The unique way people learn differs for each individual student. Correct teaching techniques hold essential value for better student performance combined with improved academic outcomes. Academic forecasting about student performance exists in most educational organizations running from primary schools to universities to boost system effectiveness. The recent educational learning systems together with educational technology benefit highly from locating valuable patterns that hide in student data. An AI-based research demonstrates machine learning techniques for student performance prediction through Logistic Regression which proves superior to Random Tree and Neural Networks. The LR model delivers reliable performance by reaching 95.50% accuracy, 97.21% precision, and an F1-score of 95.91% for identifying at-risk students, thus improving educational results. The superior predictive ability demonstrated by comparative analysis proves that the model represents a strong asset for data-based decisions in educational institutions. The research faces restriction because it uses one data set for analysis which may reduce its ability to achieve universal application in educational environments. The model analyzes academic features and demographic characteristics only, which might prevent the identification of psychological and behavioural aspects. The analysis using Logistic Regression might fail to uncover complete nonlinear patterns within the data. Future studies must engage various institutions with additional features about student involvement and social-economic aspects to boost predictive model reliability.

REFERENCES

1. E. Zakaria and Z. Iksan, "Promoting cooperative learning in science and mathematics education: A Malaysian perspective," Eurasia J. Math. Sci. Technol. Educ., vol. 3, no. 1, pp. 35–39, 2007.
2. K. Kavyashree and L. Durga, "A Review on Mining Students' Data for Performance Prediction," Int. J. Adv. Res. Comput. Commun. Eng., 2016.
3. M. Lowis and A. Castley, "Factors affecting student progression and achievement: Prediction and intervention. A two-year study," Innov. Educ. Teach. Int., 2008, doi: 10.1080/14703290802377232.
4. K. G. Preetha, S. Saritha, and S. Asokan, "Enhanced Performance of Engineering Students through REA: A Comparative Analysis," in 2012 International Conference on Advances in Computing and Communications, 2012, pp. 178–181.
5. M. Ramaswami and R. Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining," Int. J. Comput. Sci. Issues, vol. 7, no. 1, 2010, doi: 10.48550/arXiv.1002.1144.
6. du Boulay, "Artificial intelligence as an effective classroom assistant," IEEE Intell. Syst., vol. 31, no. 6, pp. 76–81, 2016.
7. Kolluri, "A Comprehensive Analysis on Explainable and Ethical Machine: Demystifying Advances in Artificial Intelligence," Int. Res. J., vol. 2, no. 7, 2015.
8. E. A. Amrieh, T. Hamtini, and I. Aljarah, "Mining educational data to predict student's academic performance using ensemble methods," Int. J. database theory Appl., vol. 9, no. 8, pp. 119–136, 2016.
9. Mueen, B. Zafar, U. Manzoor, and others, "Modeling and predicting students' academic performance using data mining techniques," Int. J. Mod. Educ. Comput. Sci., vol. 8, no. 11, pp. 36–42, 2016.
10. J. N. Purwaningsih and Y. Suwarno, "Predicting students achievement based on motivation in vocational school using data mining approach," in 2016 4th International Conference on Information and Communication Technology (ICoICT), 2016, pp. 1–5. doi: 10.1109/ICoICT.2016.7571880.
11. E. A. Amrieh, T. Hamtini, and I. Aljarah, "Preprocessing and analyzing educational data set using X-API for improving student's performance," in 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 2015, pp. 1–5. doi: 10.1109/AEECT.2015.7360581.
12. F. Ahmad, N. H. Ismail, and A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques," Appl. Math. Sci., vol. 9, no. 129, pp. 6415–6426, 2015.

13. S. E. Sorour, T. Mine, K. Goda, and S. Hirokawa, “Predicting students’ grades based on free style comments data by artificial neural network,” in 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, 2014, pp. 1–9. doi: 10.1109/FIE.2014.7044399.
14. M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang, and H. Y. Seong, “Predicting NDUM student’s academic performance using data mining techniques,” in 2009 Second International Conference on Computer and Electrical Engineering, 2009, pp. 357–361.
15. S. Dreiseitl and L. Ohno-Machado, “Logistic regression and artificial neural network classification models: a methodology review,” J. Biomed. Inform., vol. 35, no. 5, pp. 352–359, 2002, doi: https://doi.org/10.1016/S1532-0464(03)00034-0.
16. C> Bird, N. Nagappan, H. Gall, B. Murphy, and P. Devanbu, “Putting It All Together: Using Socio-technical Networks to Predict Failures,” in 2009 20th International Symposium on Software Reliability Engineering, IEEE, Nov. 2009, pp. 109–119. doi: 10.1109/ISSRE.2009.17.
17. T. Mishra, D. Kumar, and S. Gupta, “Mining Students’ Data for Prediction Performance,” in International Conference on Advanced Computing and Communication Technologies, ACCT, 2014, pp. 255–262. doi: 10.1109/ACCT.2014.105.
18. S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, “Improving accuracy of students’ final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique,” Decis. Anal., vol. 2, pp. 1–25, 2015.