PREDICTION OF STUDENT PERFORMANCE USING LEARNING ALGORITHMS

P.VIJAYA DURGA, Y.ANJIREDDY,S.ANUSHKA,R.SRUJAN

Under The Guidance of Dr S.PAVAN M.Tech,Ph.D

Department of Electronics and Communication Engineering, Tirumala Engineering College, Narasaraopet, Palnadu Dist. Andhra Pradesh, India,522601

ABSTRACT:

The ever increasing importance of education has driven researchers and educators to seek innovative methods for enhancing student performance and understanding the factors that contribute to academic success.The dataset used in this study comprises anonymized student records collected from educational institutions, including information on grades, attendance, socio- economic background, and other relevant factors. In this work we are using learning algorithms like SVM ,Random forest, XGBOOST, LogisticRegression, Gradient Boosting and KNN these algorithms are give more accuracy so we can easily predict the student performance.

.***Keywords -*** *Student prediction, Datasets, feature Selection*

I. INTRODUCTION

In the education, prediction of student performance and understanding how well students will perform academically can significantly impact educational strategies, curriculum development, and student support systems. Traditionally, statistical models and deep learning algorithms have been employed for this task. However, with the rise of Machine learning algorithms, particularly random forest, SVM, XGBoost, there has been a growing interest in leveraging their capabilities for prediction of student performance.

Predicting student performance using learning algorithms has emerged as a powerful tool in educational research and practice. By leveraging machine learning techniques, researchers and educators can analyze vast amounts of data to gain insights into various factors that influence student outcomes. These algorithms can identify patterns, trends, and correlations within data sets comprising student demographics, academic records, socioeconomic backgrounds, and even behavioral attributes. Through predictive modeling, stakeholders in education can anticipate potential challenges, tailor interventions, and optimize educational strategies to enhance student success. This approach not only aids in early identification of at-risk students but also enables personalized learning experiences, ultimately contributing to improved educational outcomes and a more effective educational system overall. The predictive power of learning algorithms extends beyond identifying at-riskstudents; it also encompasses the ability to forecast future academic achievements and guide strategic decision- making in educational institutions. By analyzing historical data and real-time inputs, these algorithms can predict not only students' likelihood of success in specific courses but also their potential career pathways based on their academic strengths and interests. This predictive modeling is particularly valuable in designing personalized learning experiences, recommending relevant courses and resources, and fostering student engagement and motivation. Moreover, learning algorithms can assist educators in understanding the effectiveness of teaching methodologies, curriculum designs, and assessment practices, thereby facilitating continuous improvement and innovation in educational delivery.

II. LITERATURE REVIEW

The literature survey encompasses a diverse range of studies that delve into various aspects of predictive modeling, including algorithm selection, feature engineering, model evaluation, and practical applications in educational settings. These studies contribute to a comprehensive understanding of the capabilities and limitations of learning algorithms in predicting student success.

Ahmad, Z., & Shahzadi, E. (2018). Prediction of student's academic performance using artificial neural network but it deals with small data sets only. Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020)Predicting key educational outcomes in academic trajectories using machine-learning approach in Higher Education in this work it predicts only student educational outcomes only. Waheed, H.,Hassan, S.U.,Aljohani, N.R., Hardman,.(2020)Predicting academic performance of students from VLE big data using deep learning models in this model , it may produce unfair predictionsof the student performance in academics. Lukáš Falát & Terézia Piscová (2022). Predicting GPA of UniversityStudents with Supervised Regression Machine Learning Models in this work Regression models were simple and better captured the data , but it does not give accurate results for giving more datasets. Zheng, L., Wang, C., Chen, X., Song, (2023). Prediction or evolutionary machine learning builds smart education big data platform using machine learning in data driven approach.

III. EXISTING METHODS

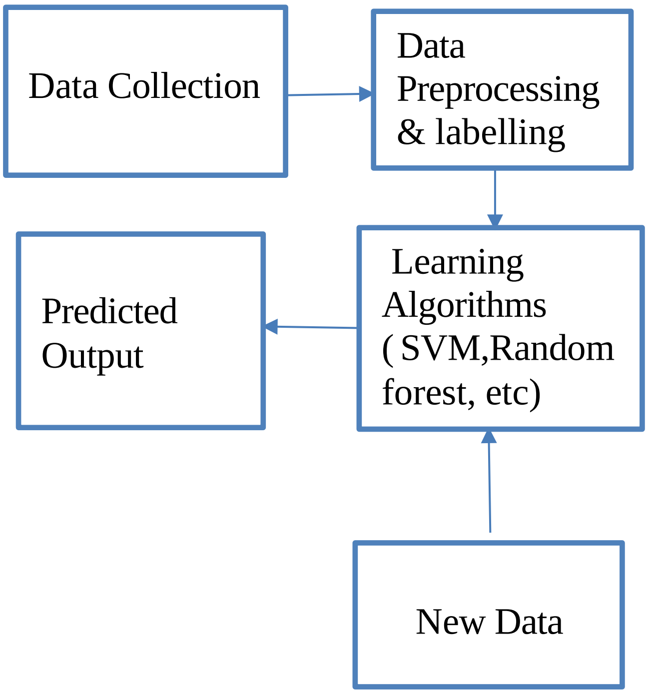
In exploring the existing systems or approaches that contribute to academic success, it's important to recognize the multifaceted nature of this endeavor. One existing system involves traditional educational structures, including curriculum design, teaching methodologies, and assessment practices. These systems provide a framework for delivering education and evaluating student performance. However, they may vary widely across different educational institutions and contexts, leading to disparities in academic outcomes. Another existing system involves academic support services provided by schools, colleges, and universities. These services may include tutoring programs, writing centers, academic advising, counseling services, and peer mentoring initiatives. These support systems aim to address students' diverse learning needs, provide guidance and assistance, and promote academic success. However, their effectiveness may depend on factors such as accessibility, availability, and quality of support. Furthermore, technological advancements have led to the emergence of online learning platforms, digital resources, and educational technologies that supplement traditional educational systems. These tools offer opportunities for personalized learning, flexibility, and access to educational content. However, they also present challenges related to digital literacy, equitable access, and the quality of online learning experiences.

IV. METHODOLOGY

proposed system involves implementing a comprehensive framework that integrates various strategies, support mechanisms, and interventions to support students' academic journey effectively. This proposed system encompasses the following key components: Implementing personalized support services tailored to students' individual needs and circumstances can help address inequities and provide targeted assistance where it's most needed. This may include offering academic tutoring, counseling, mentoring, and academic advising services that are accessible, culturally responsive, and inclusive .

BLOCK DIAGRAM

The block of prediction of student performance using learning algorithms .It have five blocks those are data collection ,data preprocessing, learning algorithms, and predicted output.



a) Data collection

Data collection refers to the systematic process of gathering and measuring information on variables of interest in a structured and organized manner.

Selected features:

1) Academic Records
2) Attendance Patterns
3) Student Demographics
4) Extra Curricular Activities

b) Data preprocessing And Labelling

Data preprocessing involves cleaning and transforming raw data into a format suitable for analysis or model training. In this Data preprocessing we are using normalization method.

Data Labelling

Data labeling refers to the process of assigning predefined categories or class labels to the data points in a dataset. In a classification task, labels represent the categories or classes that the model aims to predict.

C) Learning Algorithms

The learning algorithms are used to predict the student performance accurately.

Those are :

1) KNN

The k-Nearest Neighbors (KNN) algorithm is a versatile supervised learning method used for both classification and regression tasks. During training, it memorizes the entire training dataset. When predicting the label or value for a new data point, KNN calculates the distance between that point and all training points, typically using metrics like Euclidean distance or cosine similarity. It then selects the k nearest neighbors based on these distances. For classification, it employs a majority vote among the k neighbors to assign the class label to the new point, while for regression, it predicts the average value of the target variable. KNN's simplicity and effectiveness make it a popular choice, although optimal parameter selection and computational cost considerations are important for its performance.

RANDOM FOREST

Random Forest is an ensemble learning technique that combines multiple decision trees to make predictions. Each decision tree in the Random Forest is built independently and operates by splitting the data into subsets based on features. However, unlike traditional decision trees, Random Forest introduces randomness both in the selection of data points and features during the tree-building process. During training, each tree is constructed using a bootstrap sample (randomly selected with replacement) from the training data, which introduces diversity among the trees. Additionally, at each split in a tree, only a random subset of features is considered, further enhancing diversity and reducing overfitting.

2) SVM(Support Vector Machine)

Support Vector Machines (SVM) are powerful supervised learning models used for both classification and regression tasks. The core idea behind SVM is to find the optimal hyperplane that separates data points of different classes with the maximum margin. In the case of linearly separable data, this hyperplane is a straight line in two dimensions or a hyperplane in higher dimensions. SVM works by transforming the input data into a higher-dimensional space using a kernel function, which enables it to find a linear decision boundary that wasn't possible in the original feature space.

One of the key strengths of SVM is its ability to work well in high-dimensional spaces, making it effective for tasks with a large number of features. Additionally, SVM is less affected by overfitting, especially in cases where the margin is properly regularized using techniques like soft margin SVM (allowing for some misclassification) or using appropriate kernel parameters.

3) GRADIENT BOOSTING

Gradient Boosting is an ensemble learning technique used for supervised learning tasks like regression and classification. Unlike other ensemble methods that combine multiple models in parallel, Gradient Boosting builds an ensemble of weak learners (usually decision trees) sequentially, where each new learner focuses on correcting the errors made by the previous ones.

4) LOGISTIC REGRESSION

Logistic Regression is a fundamental statistical method used for binary classification tasks, where the goal is to predict the probability of an input belonging to one of two classes. Despite its name, logistic regression is a linear model that uses a logistic (sigmoid) function to map the output of a linear combination of input features to a probability score between 0 and 1. This probability score represents the likelihood of the input belonging to the positive class, with values closer to 1 indicating higher confidence in the positive class and values closer to 0 indicating higher confidence in the negative class.

One of the key advantages of logistic regression is its simplicity and interpretability. The model's output can be easily understood as a probability, making it straightforward to interpret the impact of each feature on the classification decision. Logistic regression also performs well when the decision boundary between classes is linear or when features are linearly separable.

5) XGBOOST

XGBoost (Extreme Gradient Boosting) is a high- performance implementation of the gradient boosting algorithm, renowned for its speed, accuracy, and scalability. It excels in both classification and regression tasks, making it a preferred choice in machine learning competitions and real- world applications alike. One of its key strengths lies in its ability to build an ensemble of decision trees sequentially, refining the model's predictions with each new tree by focusing on the residuals of the previous ones. XGBoost incorporates advanced features such as regularization techniques (L1 and L2), optimized tree building through pruning, handling of missing values, and support for parallel processing, which enhances its efficiency on large datasets. Additionally, it offers

built-in cross-validation and flexibility to customize loss functions and evaluation metrics, empowering users to fine- tune models for optimal performance. With its speed, accuracy, and scalability, XGBoost continues to be a top choice for data scientists tackling complex predictive modeling tasks.
XGBoost uses a regularized objective function that combines a loss function with a penalty on the complexity of the model. The objective function for regression typically looks like

Advantages

Firstly, these algorithms can analyze vast amounts of data, including historical academic records, demographic information, and even behavioral patterns, to identify key factors influencing student outcomes And it can detect early warning signs of academic challenges or potential dropout risks.

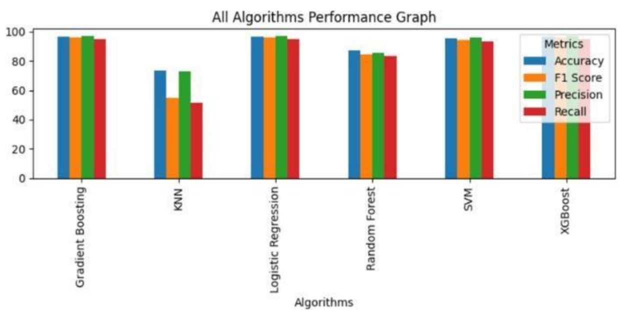
Applications

The applications of predicting student performance using learning algorithms are vast and have numerous implications for education and student support. Here are several specific applications:

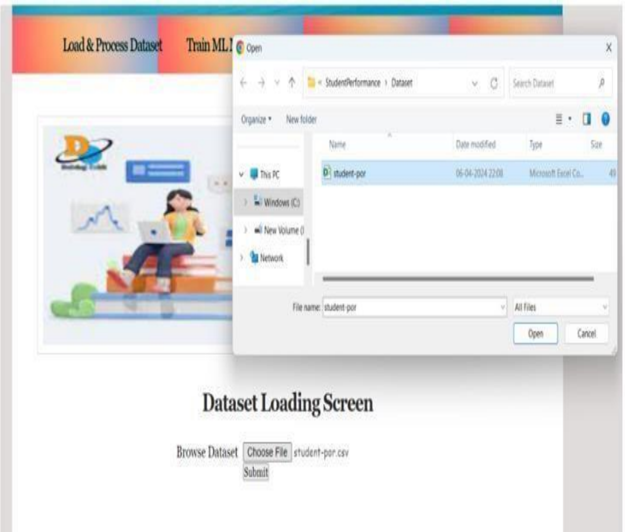
1) Early Intervention and Support
2) Personalized Learning
3) Curriculum Enhancement
4) Institutional Decision-Making

V . RESULTS

The results of predicting student performance using learning algorithms can have a significant impact on various aspects of education and student outcomes. Here are some key results and outcomes that can be achieved through the application of learning algorithms in predicting student performance. The accuracy table
drawn below
Obj=*i*=1∑*nL*(*yi*,*y*^*i*)+*k*=1∑*K*Ω(*fk*)
Here, *LL* is the loss function (like squared error for regression), *yiyi* are the actual target values, *y*^*iy*^*i* are the predicted values, Ω(*fk*)Ω(*fk*) is the regularization term for each tree *fkfk*, and *KK* is the total number of trees.
The learning algorithms are give more accurate predictions of student performance .In this algorithms XGBOOST produce high accuracy,precision,recall and Fscore.

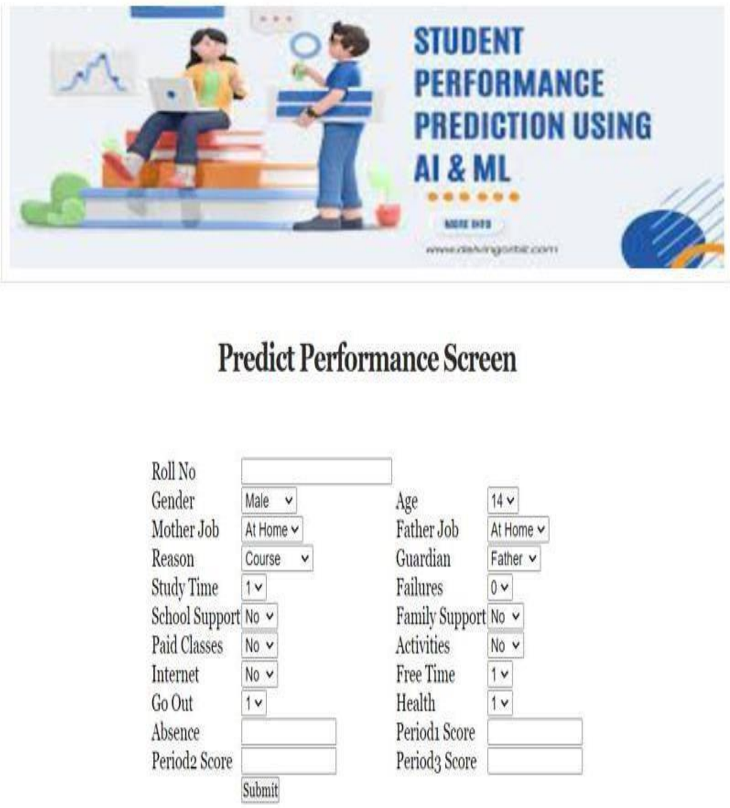


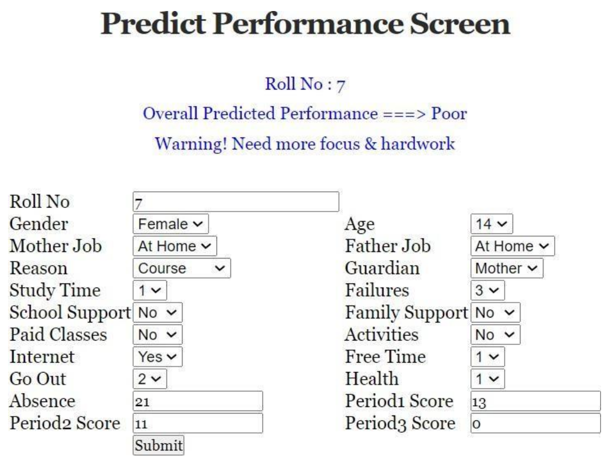
The bar graph represents all algorithms performance graph.



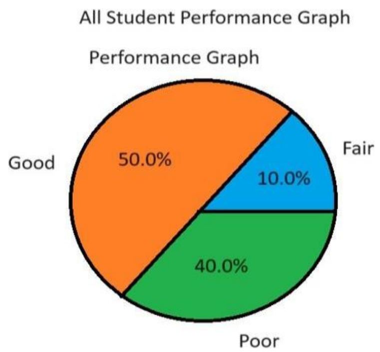
Algorithm Name	Accuracy	Precision	Recall	Fscore
KNN	73.54	72.74	51.25	54.92
Random Forest	87.07	85.81	83.16	84.37
SVM	95.38	96.1	93.08	94.4
Gradient Boosting	96.31	97.03	94.91	95.87
Logistic Regression	93.85	95.02	91.24	92.81
XGBoost	95.38	94.08	95.76	94.89

In the above screen user can click on 'Load & Process Dataset' link to get below page and select load dataset file and this dataset file available inside 'dataset' folder and then click on 'Open' and 'Submit' button to get below page.In the above screen user will enter roll number and select academic details and then click on 'Submit' button get below output.





In the above screen in blue color can see user academic data and then can predicted performance as 'Poor' with alert message to improve. Similarly you can give any input details and get performance predicted like good, fair, etc.



In the above pie chart graph can see overall performance of student and the half of the students are performing Good means 50% of the students and in the remaining students 10% of the students are performing Fair means they need to improve some more and the remaining 40% of the students are Poor in there both academics and extra curricular activities They need to improve a lot work hard.

VI . CONCLUSION

In conclusion, the prediction of student performance using learning algorithms offers several advantages. It enables early intervention and personalized learning, helping educators identify struggling students and tailor educational content to their needs. Learning algorithms also optimize resource allocation, support data-driven decision making, and provide continuous feedback for improvement. These applications demonstrate the potential for technology to positively impact education by enhancing student outcomes and promoting individualized learning experiences. And to see how learning algorithms can contribute to a more effective and inclusive educational system with the help of learning algorithms, educators can gain valuable insights into student performance. By analyzing data such as academic records, test scores, attendance, and even social factors, these algorithms can identify patterns and trends that may impact student success. This information can then be used to provide targeted interventions and support to students who may be at risk of falling behind. Additionally, learning algorithms can help in identifying areas where teaching methods and curriculum can be improved, leading to more effective instruction and better student outcomes. It's amazing how technology can assist educators in better understanding and supporting their students' academic journey.

REFERENCE

[1]. Zheng, L., Wang, C., Chen, X., Song, Y., Meng, Z. and Zhang, R., 2023. Evolutionary machine learning builds smart education big data platform: Data-driven higher education. *Applied Soft Computing*, *136*, p.110114.

[2]. Duong, H.T.H., Tran, L.T.M., To, H.Q. and Van Nguyen, K., 2023. Academic performance warning system basez on data driven for higher education. *Neural Computing and Applications*, *35*(8), pp.5819-5837.

[3]. Kumar, G.S., De la Cruz-Cámaco, D., Ravichand, M., Joshi, K., Gupta, Z. and Gupta, S., 2023, March. Monitoring and Predicting Performance of Students in Degree Programs using Machine Learning. In *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1311-1315). IEEE.

[4] Althaph, B., Sreenivasu, S.V.N. and Reddy, D.V., 2023, January. Student Performance Analysis with Ensemble Progressive Prediction. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1513- 1517). IEEE.

[5] Ouatik, F., Erritali, M., Ouatik, F. and Jourhmane, M., 2022. Predicting student success using big data and machine learning algorithms. International Journal of Emerging Technologies in Learning (iJET), 17(12), pp.236-251

[6] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís- López, M., Flores-Albornoz, J. and Phasinam, K., 2023. Classification and prediction of student performance data using various machine learning algorithms. Materials tzday: proceedings, 80, pp.3782-3785.

[7] Amra, I.A.A. and Maghari, A.Y., 2017, May. Students performance prediction using KNN and Naïve Bayesian. In *2017 8th international conference on information technology (ICIT)* (pp. 909-913). IEEE.

[8] Kumar, R.S. and Kumar, J., 2018. Analysis of student performance based on classification and map reduce approach in big data. *International Journal of pure and applied mathematics*, *118*(14), pp.141-148.

[9] Shanmugarajeshwari, V. and Lawrance, R., 2016, January. Analysis of students' performance evaluation using classification techniques. In *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)* (pp. 1-7). IEEE.

[10] Asif, R., Merceron, A., Ali, S.A. and Haider, N.G., 2017. Analyzing undergraduate students' performance using educational data mining. *Computers & education*, *113*, pp.177-194.

[11] Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R. and Van Erven, G., 2019. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of business research*, *94*, pp.335-343

[12] Hoffait, A.S. and Schyns, M., 2017. Early detection of university students with potential difficulties. *Decision Support Systems*, *101*, pp.1-11.