STUDENT GRADE PREDICTION USING GRADIENT BOOSTING CLASSIFIER

By **Annakula Srikanth**
Student
B V Raju Institute of Technology
Narsapur, Hyderabad 502313

By **Alluri Abhi karthikeya**
student
B V Raju Institute of Technology
Narsapur, Hyderabad 502313

By **Dr.B Venkateswara Rao**
Associate Professor
B V Raju Institute of Technology
Narsapur, Hyderabad 502313

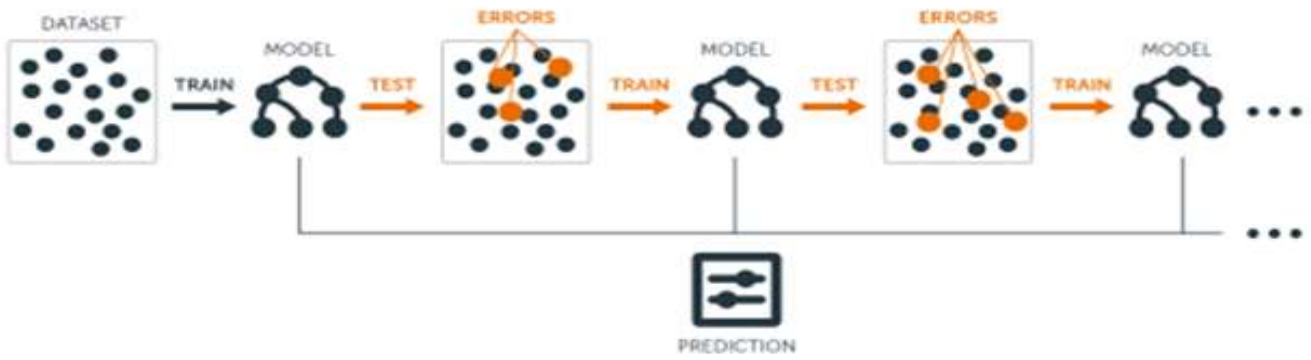
ABSTRACT

Educational institutions are increasingly interested in predictive analytics applications, which utilize advanced analytics, including machine learning, to provide high-quality performance data and meaningful insights across all education levels. One crucial performance indicator for students is their grades, and over the past decade, researchers have explored various machine learning techniques to predict them. However, imbalanced datasets have posed significant challenges to achieving accurate predictions. Our project aims to analyze machine learning techniques comprehensively to improve the accuracy of predicting final grades for students in their respective courses, specifically addressing the issue of imbalanced multi-classification. Our proposed model shows promising results that can enhance the predictive accuracy of student grade predictions.

1.INTRODUCTION

Student grade prediction using gradient boosting classifier is a machine learning approach that uses historical student performance data to predict future academic success. This approach involves training a model using various features such as student demographics, academic history, and other relevant factors, to predict a student's grade in a particular course or subject. The gradient boosting classifier algorithm is particularly useful in this application because it is a powerful, ensemble learning method that can handle complex, non-linear relationships between the input variables and the output label. By accurately predicting a student's grade, this approach can be used to identify students who are at risk of failing a course or subject, and provide interventions and support to help them succeed. It can also be used to identify high-performing students and provide opportunities for them to challenge themselves academically.

Gradient boosting classifier is a powerful machine learning technique that utilizes a boosting algorithm to improve the accuracy of predictions. It is particularly effective in handling imbalanced datasets and can handle both numeric and categorical data.In this project, we will use gradient boosting classifier to predict final student grades in their respective courses.We will conduct a comprehensive analysis of the machine learning technique and its application to student grade prediction. We will use a dataset containing features such as student demographics, past performance, and class attendance to train the algorithm and make predictions on a test set Overall, Gradient Boosting Classifier is a powerful tool for student grade prediction that can help educators and administrators make data-driven decisions to improve student outcomes. By using machine learning to analyze and predict student performance, schools and universities can identify at-risk students early and provide targeted support to help them succeed.



2.RESEARCH METHODOLOGY

Apart from existing system, In our project we have taken a dataset of various attributes so that the predictive accuracy is high.We are using the features from our dataset and define classification algorithms to identify whether the student performs good in final grade exam, We used many Attributes in our dataset and we performed many techniques and selected best machine learning algorithm for

high accuracy and consistency. Predictive Accuracy is high.

2.1 Assumptions for Gradient Boosting Classifier:-

Gradient boosting is a popular machine learning approach that is used for solving classification and regression problems. The technique involves creating an ensemble of multiple weak prediction models, usually in the form of decision trees, to generate a final prediction model.

In simpler terms, the algorithm combines many simple models, each having relatively low predictive power, to create a stronger model that can make accurate predictions. The boosting process involves iteratively improving the weak models by focusing on the misclassified data points in each iteration, until the overall model performs well on the entire dataset.

2.2Why Gradient Boosting Classifier? :-

- The model has a high degree of accuracy in predicting output, even when handling large datasets, and it performs efficiently.
- No need for scaling or Normailizing the data
- The function is highly flexible as it allows for optimization on various loss functions and provides multiple hyperparameter options, making it adaptable to various use cases.
- It takes less training time as compared to other algorithms.

2.3 Working process of Gradient Boosting Classifier:-

Step-1: The first step includes building a base model to predict the dataset.

Step-2: Find the pseudo residuals (observed value – predicted value)

Step-3: This step includes finding the output values of each leaf of a decision tree

Step-4: The final step involves the update of the forecasts of the previous model.

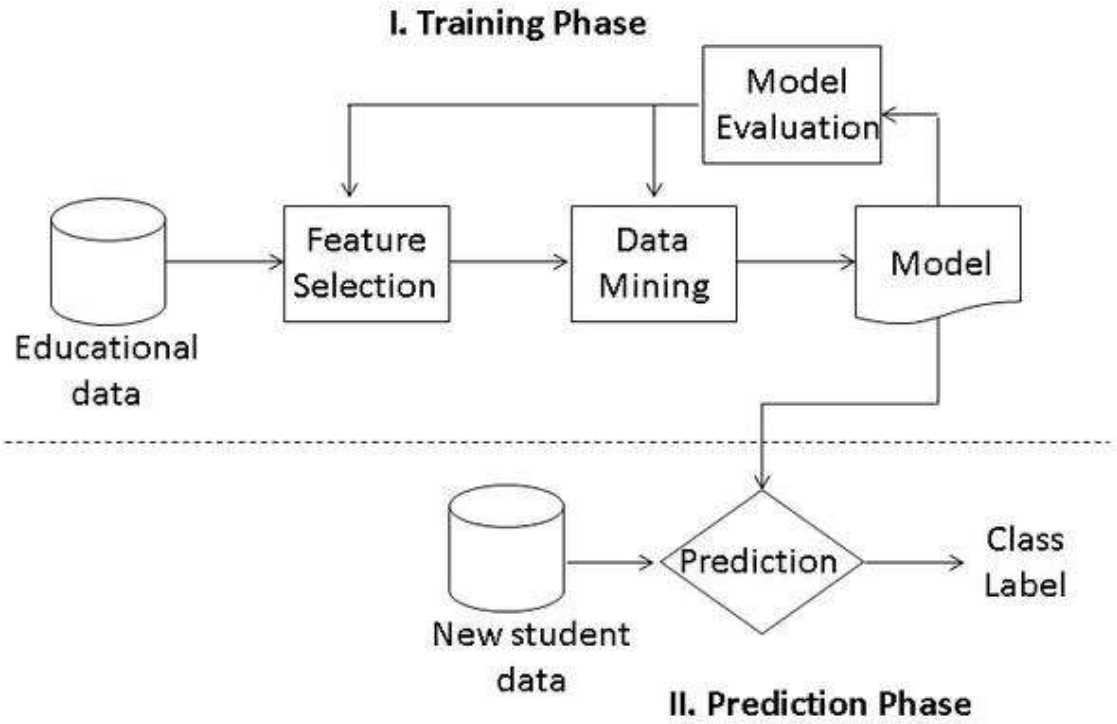
2.4 Description of algorithm:-

Step 1: Consider a dataset having different data points and initialize it.

Step 2: Now, give equal weight to each of the data points.Assume this weightas an input for the model.

Step 3: Identify the data points that are incorrectly classified and increase the weight for data points

Step 4: If you get appropriate output then terminate this process else follow Step2.



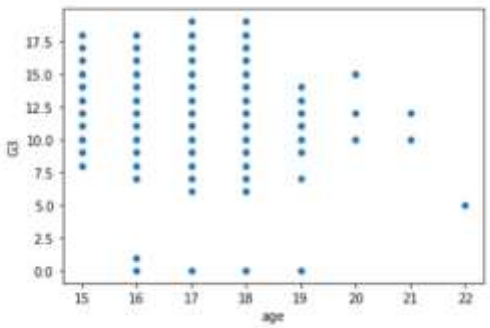
3.RESULTS

3.1 Algorithm Accuracy

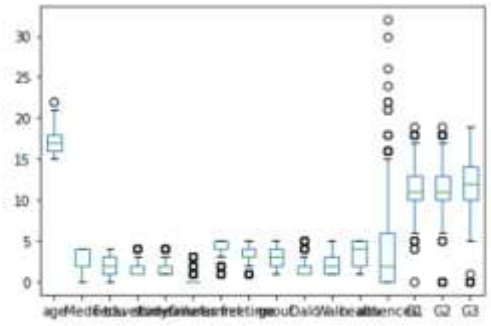
```
In [ ]: import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.ensemble import GradientBoostingClassifier
gb_clf = GradientBoostingClassifier(n_estimators=20, learning_rate=0.2, max_features=2, max_depth=2, random_state=0)
gb_clf.fit(X_train, y_train)
print("Accuracy score (training): {0:.3f}".format(gb_clf.score(X_train, y_train)))
print("Accuracy score (validation): {0:.3f}".format(gb_clf.score(X_test, y_test)))

Accuracy score (training): 0.762
Accuracy score (validation): 0.749
```

3.2 Data Visualization

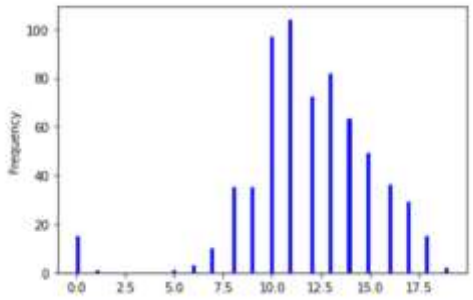
```
In [ ]: sns.scatterplot(data=df,x='age',y='G3')
Out[352]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb26c6a1750>
```



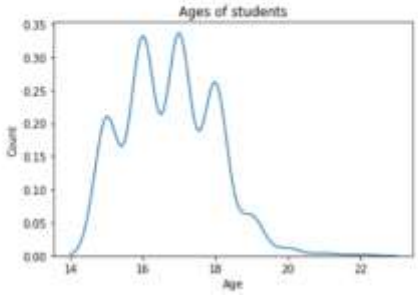

```
In [ ]: df.plot(kind='box')
Out[353]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb26c77f0d0>
```



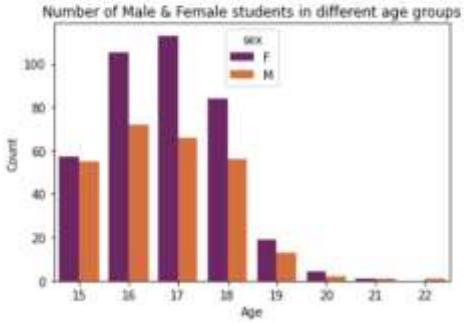
```
In [ ]: df['G3'].plot(kind='hist',bins=100,color='blue')
Out[354]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb26c719710>
```



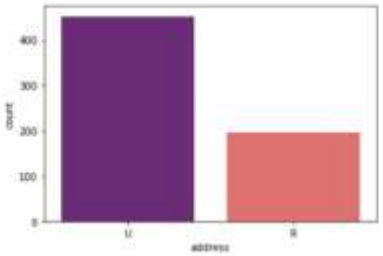

```
In [ ]: b = sns.kdeplot(df['age'])
b.axes.set_title('Ages of students')
b.set_xlabel('Age')
b.set_ylabel('Count')
plt.show()
```



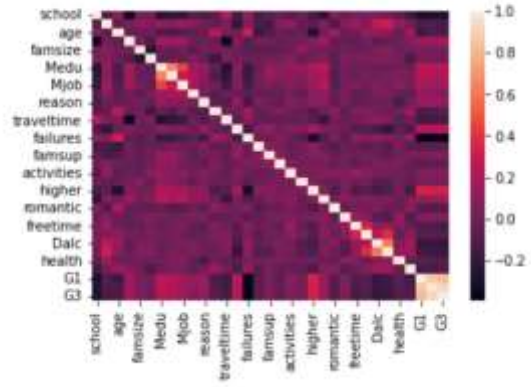

```
In [ ]: b = sns.countplot(x='age',hue='sex', data=df, palette='inferno')
b.axes.set_title('Number of Male & Female students in different age groups')
b.set_xlabel("Age")
b.set_ylabel("Count")
plt.show()
```



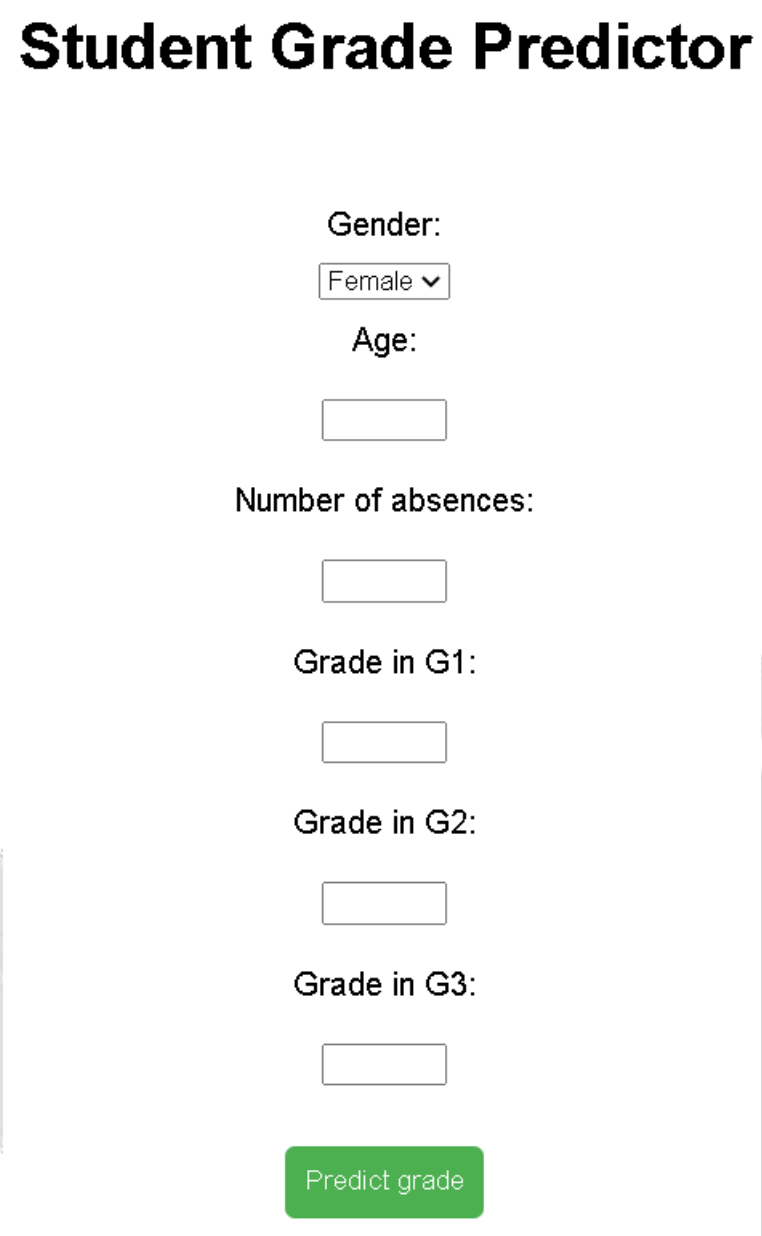

```
In [ ]: u = len(df[df['address'] == 'U'])
print('Number of Urban students:',u)
r = len(df[df['address'] == 'R'])
print('Number of Rural students:',r)
sns.countplot(x='address',data=df,palette='magma')

Number of Urban students: 452
Number of Rural students: 197
Out[357]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb26c1a0d50>
```




```
In [ ]: #Correlation graph
cormat = df.corr()
sns.heatmap(cormat);
```



3.3 Input Screen



4.CONCLUSION:

One way for educators to track their students' academic progress is by predicting their grades, which serves as a crucial performance indicator. Therfore, Having a predictive model is crucial in reducing the uncertainty in the outcome of an imbalanced dataset.

In recent times, there has been a significant increase in the adoption of e-learning management systems, and many developed countries have transitioned their educational systems to fully or partially automated systems. As a result, these systems generate vast amounts of data containing hidden patterns and knowledge that can be leveraged to generate meaningful insights, helping students improve their academic performance and achievements. So, From our project we conclude that Gradient Boosting Classifier is more accurate and gave better output compared to other algorithms.

5.REFERENCES

[1] I. Khan,A. Al Sadiri, A. R. Ahmad, and N. Jabeur,‘‘Tracking student performance in introductory programming by Means of machine learning,’’

[2] T. Anderson and R. Anderson, ‘‘Applications of machine learning to student grade prediction in quantitative business courses.’’

[3] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, ‘‘Using machine learning to predict student difficulties from learning session data,’’

[4] D. Solomon , S. Patil, and P.Agarwal, "Predicting performance and potential difficulties of university student using classification.

[5] F. Ahmad, N. H. Ismail, and A. Aziz,''The prediction of students'academic performance using classification data mining techniques,''

[6] B. Predic,G. Dimic,D. Ranic,P. Strbac,N.Macek,and P. Spalevic''Improving final grade prediction accuracy in blended learning environment using voting ensembles,'' Comput.Z.

[7] Iqbal,J. Qadir,A. N. Mian,and F. Kamiran,''Machine learning based student grade prediction:A case study,'' 2017

[8] E. Alyahyan and D. Dustegor, "Predicting academic success in higher education: Literature review and best practices,"Int.J.Educ.Technol.Higher Educ.,