



ABSTRACT

This poster aims to highlight the importance of predicting student test scores and the various methods that can be used to achieve accurate predictions. By using machine learning algorithms and data preprocessing techniques, teachers and institutions can understand the strengths and weaknesses of each student and provide targeted interventions to enhance their learning outcomes. Through visualizations, such as scatter plots and heat maps, the poster will demonstrate the correlation between predicted scores and various factors that influence student learning. Ultimately, predicting student test scores can lead to personalized learning, targeted interventions, and improved student outcomes, making it a valuable tool for educators and educational institutions.

INTRODUCTION

The ability to predict student test scores has become increasingly important in the education sector. Predicting student test scores can help educators and institutions better understand their students' strengths and weaknesses, identify areas that need improvement, and provide targeted interventions to enhance learning outcomes. With the advancements in technology, machine learning algorithms have become a popular tool for predicting student test scores. These algorithms can analyze vast amounts of data and identify patterns that can help predict how a student will perform on a test. In this poster, we will explore the importance of predicting student test scores, the different methods used to achieve accurate predictions, and the benefits that these predictions can provide to educators and institutions. We will also discuss the role of data preprocessing and feature selection in achieving accurate predictions and demonstrate how visualizations can help educators interpret and communicate the predicted scores. By understanding the importance of predicting student test scores, educators and institutions can take the necessary steps to improve the quality of teaching and learning and provide students with the support they need to succeed.

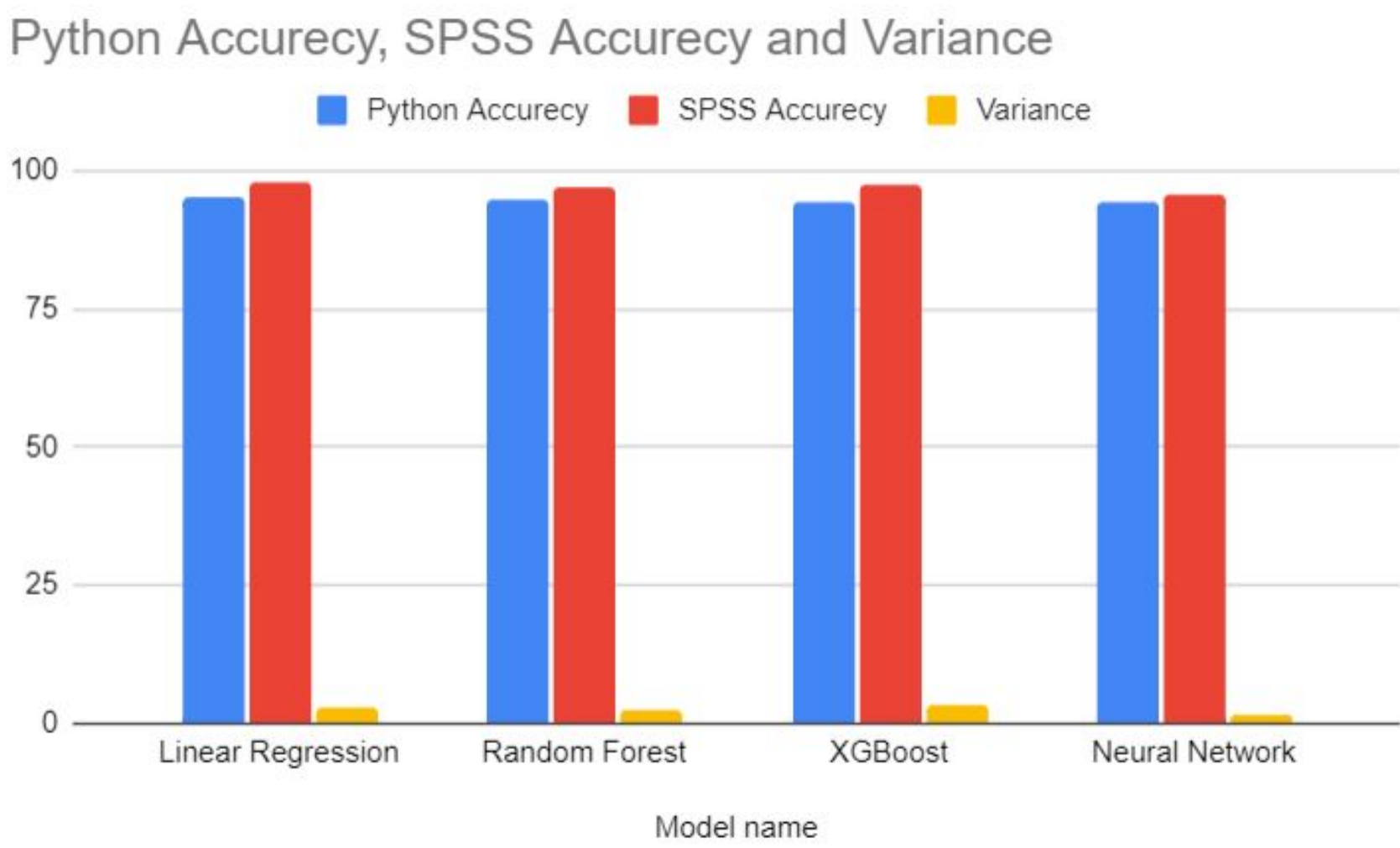
OBJECTIVE

- To demonstrate the importance of predicting student test scores and how it can benefit educators and institutions in improving the quality of teaching and learning.
- To showcase the different machine learning algorithms that can be used to predict student test scores, including linear regression, decision trees, and random forests, and highlight the importance of data preprocessing and feature selection in achieving accurate predictions.
- To use visualizations, such as scatter plots, heat maps, and bar charts, to present the predicted scores and their correlation with various factors that influence student learning, such as demographics, attendance, and past academic performance.

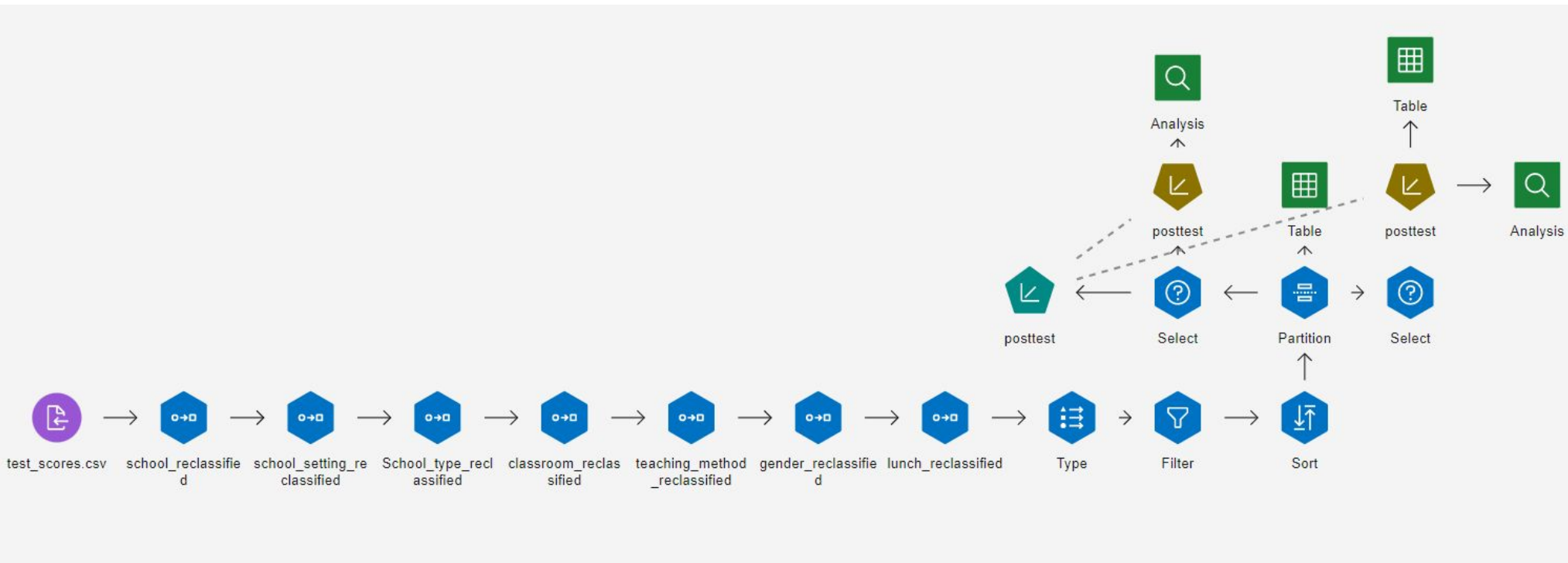
Proposed Method

- We have made stream for predicting the heart failure using SPSS Modeler from IBM. As for our given dataset we were given target attribute so we used supervised learning algorithm.
- Model we tried are linear Regression, Lasso Regression, Random Forest, XGBoost Regression and Neural Network provided by SPSS Modeler.
- Below are the Screenshot of accuracy achieved from the respective model and there python code:

Model name	Python Accuracy	SPSS Accuracy	Variance
Linear Regression	95	97.8	2.8
Random Forest	94.54	97	2.46
XGBoost	94.34	97.5	3.16
Neural Network	94.39	95.8	1.41



- Linear regression is a commonly used algorithm for predicting student test scores because it is simple, easy to implement, and can model the relationship between the independent variables and the dependent variable (test score) using a straight line.
- Linear regression assumes that there is a linear relationship between the independent variables and the dependent variable, which is often the case in education research.
- Overall, the proposed method involves collecting and preprocessing data, selecting important features, using machine learning algorithms to predict test scores, and presenting the results using visualizations to enhance the quality of teaching and learning.
- As we can see that the highest accuracy provide by Python and SPSS is linear regression. that's why we used so it suits best for the given dataset and below is the stream of the Linear regression model prepared in IBM SPSS modeler

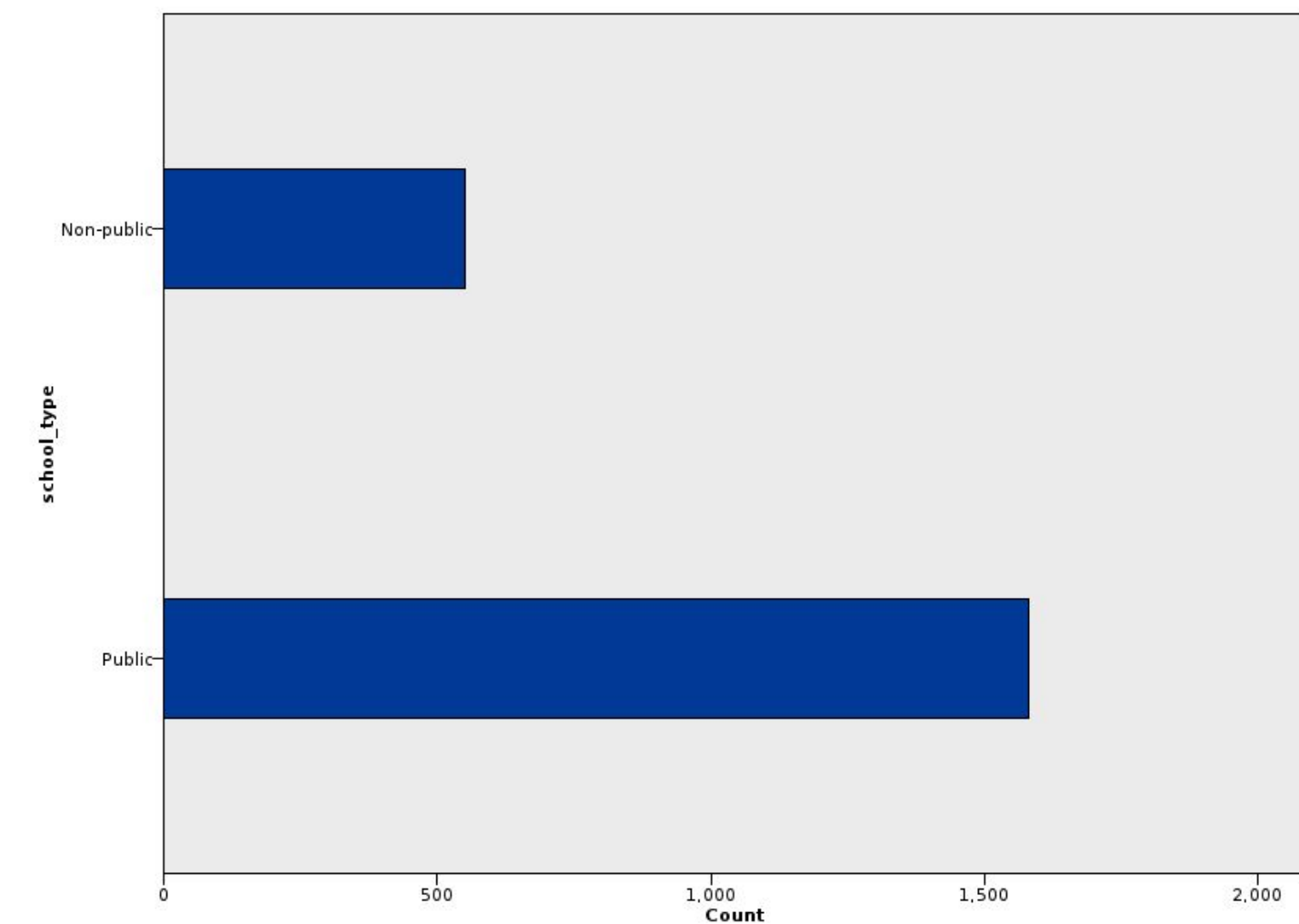
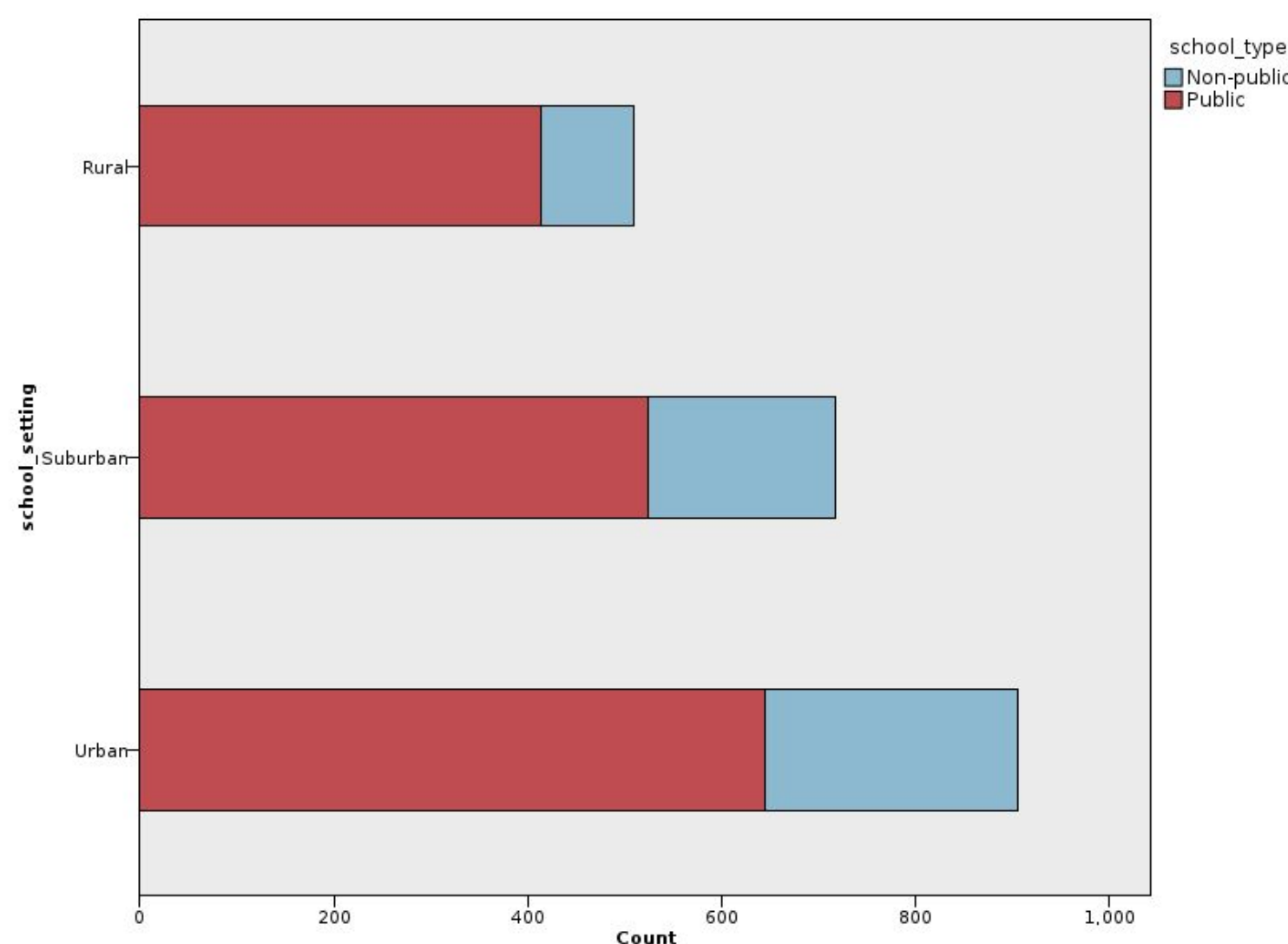


Data Description

- Number of fields = 11
- Number of Records = 2133
- Attributes contained in the dataset school,school_setting,School_type,Classroom,Teaching_method ,n_student,student_id,gender,Lunch,pretest,posttest.
- n_student,,pretest and posttest are fields having continuous measurement level.
- school,school_setting,School_type,Classroom,Teaching_method ,student_id,gender and Lunch are fields with nominal measurement level.

Data Exploration

School_setting(Rural-Suburban-Urban) Vs school_type(Private-Public)



Working Flow

- Used Python language for coding.
- Firstly loaded the dataset using pandas
- Then We dropped the target attribute from the variable in which the dataset was loaded and formed a new variable in which only the target attribute was present.
- After this the splitting was done as 70% training data and 30% testing data which is done using sklearn library.
- After splitting the the training and testing dataset was brought into same scale using StandardScaler.
- After scaling into same range Linear regression model is applied on dataset and the accuracy observed on training and testing dataset is 98.10 and 97.80 respectively.

Results

- The linear regression model was found to be an effective method for predicting student test scores, with past academic performance being the most important factor affecting the scores. These results can help educators and institutions make informed decisions and develop targeted interventions to improve the learning outcomes of students.
- The selected model is Linear regression and the accuracy obtained after partitioning the model into 70% of training data and 30% of testing data is 98.10 for training and 97.80 for testing with variance of 2.80.
- Below is the screenshot of the accuracy of training and testing data:

Training:

'Partition'	1_Training
Minimum Error	-8.707
Maximum Error	10.034
Mean Error	0.0
Mean Absolute Error	2.145
Standard Deviation	2.7
Linear Correlation	0.981
Occurrences	1,468

Testing:

'Partition'	2_Testing
Minimum Error	-9.697
Maximum Error	10.641
Mean Error	-0.261
Mean Absolute Error	2.303
Standard Deviation	2.948
Linear Correlation	0.978
Occurrences	665

Conclusion

In conclusion, our study demonstrated that linear regression is a useful method for predicting student test scores by identifying the most significant predictors of student success. Our proposed method involved exploring student demographics, attendance, and past academic performance as independent variables, and using linear regression to model the relationship between these variables and test scores. Our analysis revealed that past academic performance had the most significant positive impact on test scores, followed by attendance and demographics.

References

Prediction of Student's performance by modelling small dataset size | Published: 02 August 2019 | Lubna Mahmoud Abu Zohair

Link:<https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-019-0160-3>

A Review on Predicting Student's Performance using Data Mining Techniques | Amirah Mohamed Shahiria,Wahidah Husaina,Nur'aini Abdul Rashid | Publish by Elsevier Ltd. Selection and/or peer-review under responsibility of the sc

Link:<https://core.ac.uk/download/pdf/235050621.pdf>