SIBTAIN KHAN (20JE0946)

MINI PROJECT

STUDENT PERFORMANCE PREDICTION:

⇛➤INTRODUCTION:

Description: This dataset contains information on the performance of high school students in mathematics, including their grades and demographic information. The data was collected from three high schools in the United States.
"This dataset was created for educational purposes and was generated, not collected from actual data sources."

Columns:
• **Gender:** The gender of the student (male/female)
• **Race/ethnicity:** The student's racial or ethnic background (Asian, African-American, Hispanic, etc.)
• **Parental level of education:** The highest level of education attained by the student's parent(s) or guardian(s)
• **Lunch:** Whether the student receives free or reduced-price lunch (yes/no)
• **Test preparation course:** Whether the student completed a test preparation course (yes/no)
• **Math score:** The student's score on a standardized mathematics test
• **Reading score:** The student's score on a standardized reading test
• **Writing score:** The student's score on a standardized writing test

Sample Data

Sample Output

gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
male	group C	some college	standard	none	79	76	77
male	group C	associate's degree	standard	completed	82	78	80
female	group D	high school	standard	none	60	71	69
female	group D	some college	free/reduced	completed	59	79	77
male	group B	some college	free/reduced	completed	59	67	65
male	group B	some high school	standard	none	53	60	51
female	group C	high school	standard	none	44	54	55
female	group D	some high school	standard	none	68	69	74
female	group B	some college	standard	completed	88	88	97
male	group C	associate's degree	standard	none	65	50	50
...							

⇛OBJECTIVE:

This dataset could be used for various research questions related to education, such as examining the impact of parental education or test preparation courses on student performance. It could also be used to develop machine learning models to predict student performance based on demographic and other factors.

⇛METHODOLOGY:

Train/Test is a method to measure the accuracy of your model. It is called Train/Test because we split the data set into two sets: a training set and a testing set. 80% for training, and 20% for testing. We train the model using the training set.

Steps Involved:

- Loading the dataset.
- Understanding the dataset.
- Data preprocessing.
- Data visualization.
- Building a regression model.
- Model evaluation.
- Model prediction.

Linear regression:

Regression searches for relationships among variables.Regression problems usually have one continuous and unbounded dependent variable. The inputs, however, can be continuous, discrete, or even categorical data such as gender, nationality, or brand.
It's a common practice to denote the outputs with y and the inputs with x. If there are two or more independent variables, then they can be represented as the vector $\mathbf{x} = (x_1, ..., x_r)$, where r is the number of inputs.

Regression Performance

The variation of actual responses y_i, $i = 1, ..., n$, occurs partly due to the dependence on the predictors $\mathbf{x}_i$. However, there's also an additional inherent variance of the output.

The coefficient of determination, denoted as R^2, tells you which amount of variation in y can be explained by the dependence on $\mathbf{x}$, using the particular regression model. A larger R^2 indicates a better fit and means that the model can better explain the variation of the output with different inputs.

The value $R^2 = 1$ corresponds to SSR = 0. That's the perfect fit, since the values of predicted and actual responses fit completely to each other.

⇛RESULT AND DISCUSSION:

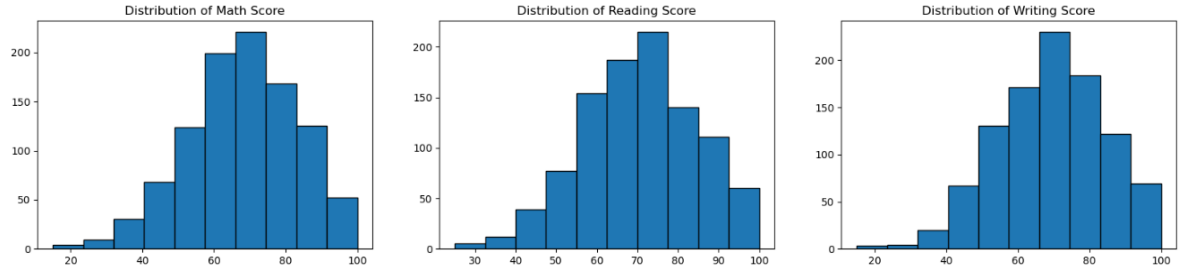
Mean squared error: 28.38885289905444

Root Mean squared error: 5.328119077034075

Mean absolute error: 4.379877672766591

R-squared score: 0.8702740399180283

The value R^2=0.87 means that our model and actual data match by 87% and hence the values of predicted and actual responses fit to each other.



⇛CONCLUSION:

We have trained our model to predict whether the student will perform good or not and use this model to predict the performance.From the result we can conclude that male students have better probability to perform good in the exam.Also the test preparation course is quite effective in determining the performance.

⇛REFERENCES:

-KAGGLE.COM

-YOUTUBE.COM