**OWN DATASET**

**Dataset Description:**

The dataset used for this analysis comprises year-wise records of CO2 emissions and temperature differences, spanning over 50 years. It is structured into three columns:

Year: Representing the specific year for each data entry.

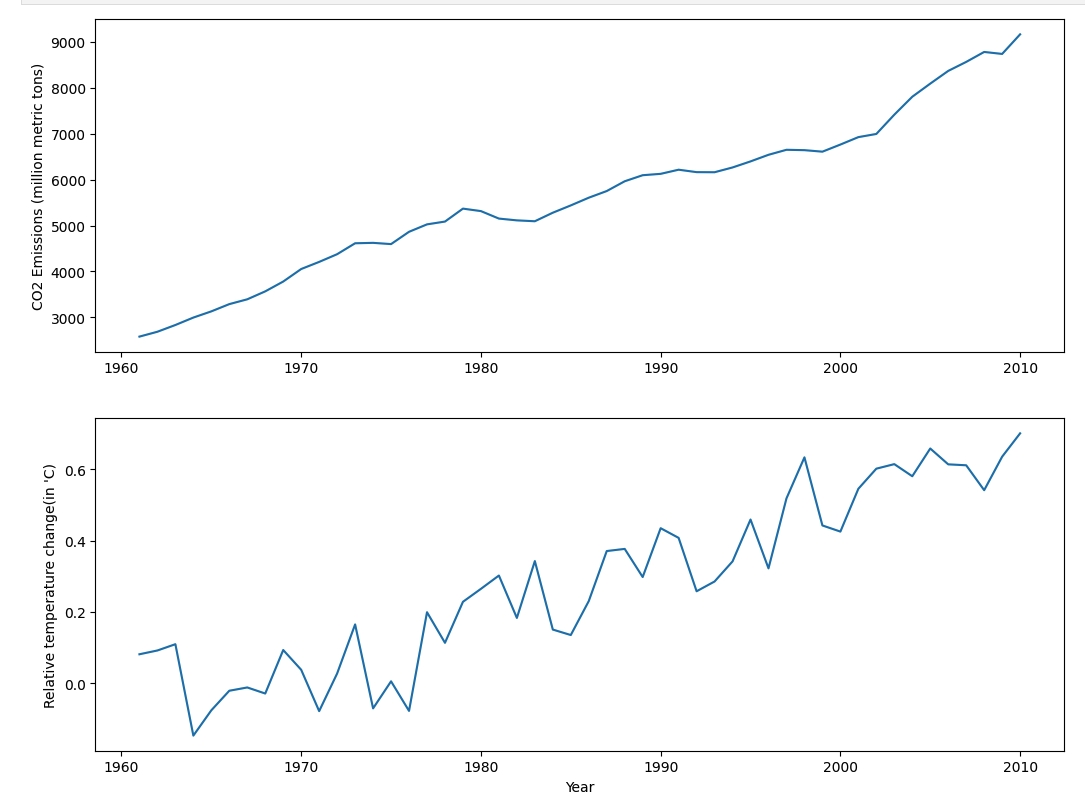
CO2 Emission: Quantified in metric tons, this column captures the annual CO2 emissions.

Temperature Difference: Measured in degrees Celsius (°C), this column reflects the change in temperature relative to a baseline year.

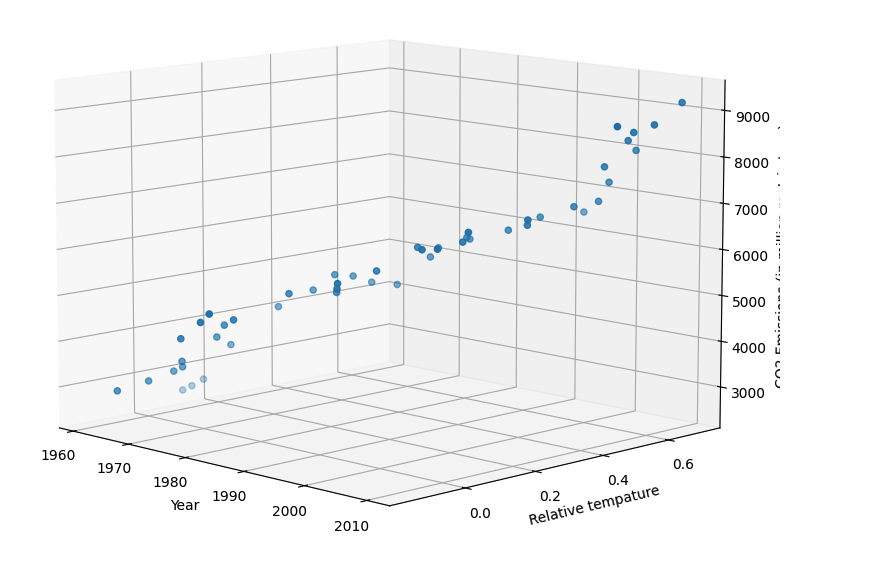
In total, the dataset consists of 50 rows, each corresponding to a distinct year's data. The primary objective of this dataset is to investigate the impact of CO2 emissions on temperature changes, making the 'Temperature Difference' the target variable for analysis. This dataset offers a valuable resource for understanding the long-term trends and correlations between human-induced emissions and climatic shifts.

**EDA and Data Visualisation:**

* Next, we try to analyze the trend temperature change with both variables. We observe that somehow it is an approximately linear trend so we can apply linear regression for this.
* Below are the visualizations we did for the dataset.



[line plot between each variable]



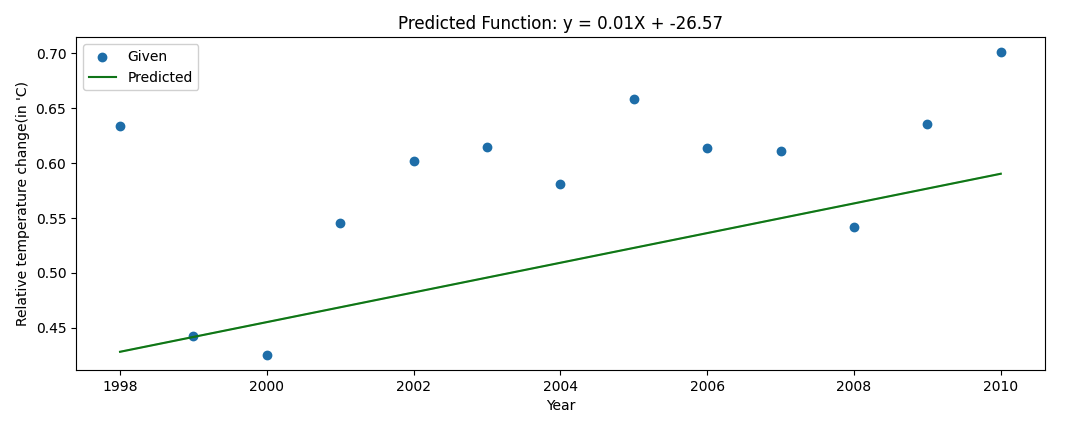
[3d scatter plot between each variable]

**Splitting the Dataset**

* The dataset was divided into training and testing subsets to facilitate model training and evaluation.
* Our training set and testing set ratio we kept at 75:25
* And then we scaled the dataset.

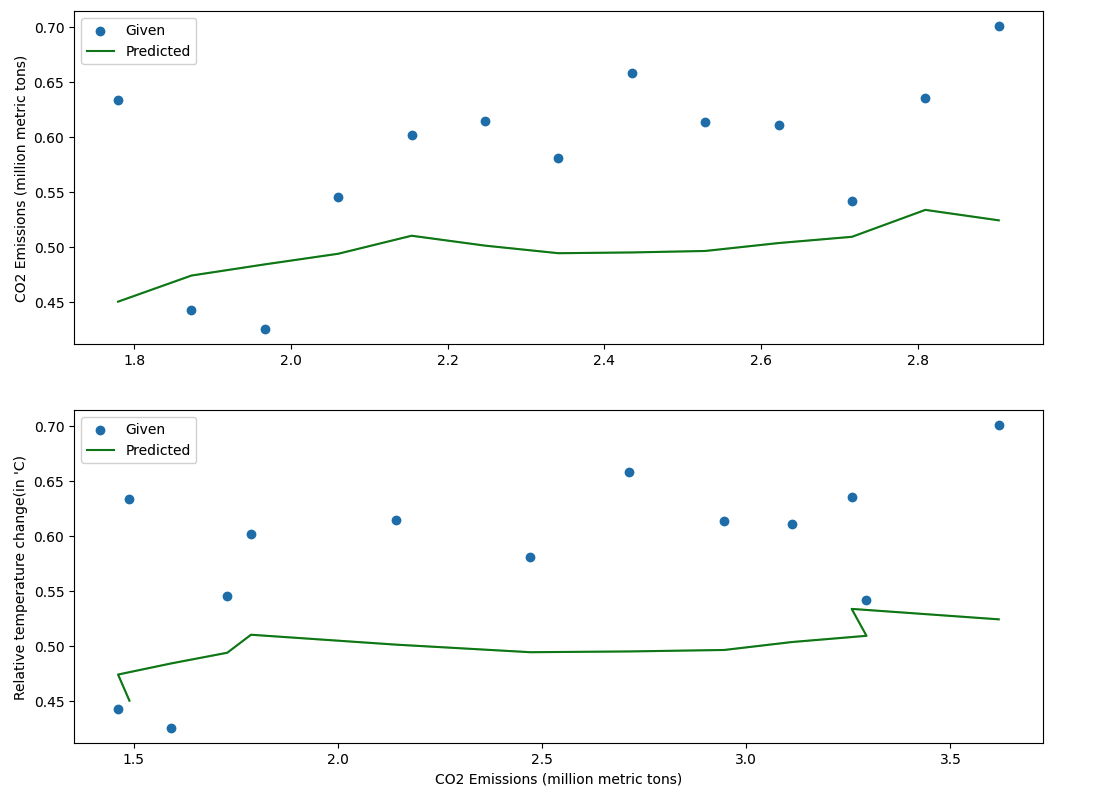
**Applying Linear Regression**

* Linear regression was applied to the dataset as a means to understand and predict the relationship between Year and temperature change. This technique, ideal for examining linear relationships, involves fitting a linear equation to the observed data.
* The following visualization is the plot between the Actual and predicted values we are getting after testing it on the test set.
* Mean squared error: 0.0097
* R2 score: -0.6807



**Multilinear Regression**

* Multilinear regression was also employed to delve deeper into the data. This approach extends the simplicity of linear regression to accommodate multiple independent variables, offering a more nuanced view of the dataset. Here we have included the Co2 emission along with the year to build a robust model.
* Mean squared error: 0.0126
* R2 score: -1.1758
* Below is the visualization of prediction vs actual for each variable.



**Given DATASET**

**Dataset Description:**

The dataset in focus is designed to evaluate the attributes influential in the application process for graduate programs. It encompasses a range of variables, each playing a pivotal role in assessing a candidate's suitability. The target variable, 'Chance of Admit,' is crucial for understanding the overall likelihood of a candidate's acceptance, making this dataset invaluable for predictive modeling in academic admissions scenarios.

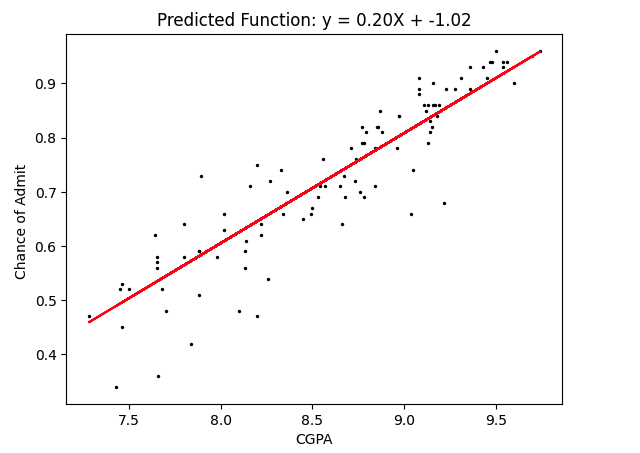
We have a total of 500 observations from the data.

**Splitting the Dataset**

* We will drop the serial no column as that is not necessary.
* Do the normalisation of data to get them into same range.
* The dataset was divided into training and testing subsets to facilitate model training and evaluation.
* Our training set and testing set ratio we kept at 75:25.

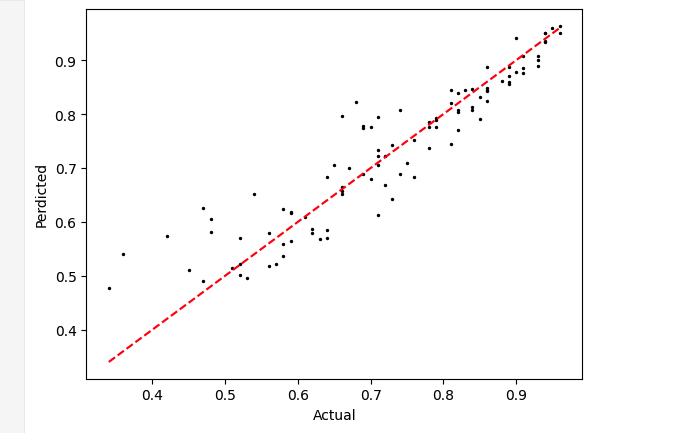
**Applying Linear Regression**

* We apply Linear regression between only CGPA and chances of admission.
* The following visualization is the plot between the Actual and predicted values we are getting after testing it on the test set.
* Mean squared error: 0.0038
* R2 score: 0.8290
* Below is the prediction vs actual value of the Chance of admin and the CGPA column.



**Multilinear Regression**

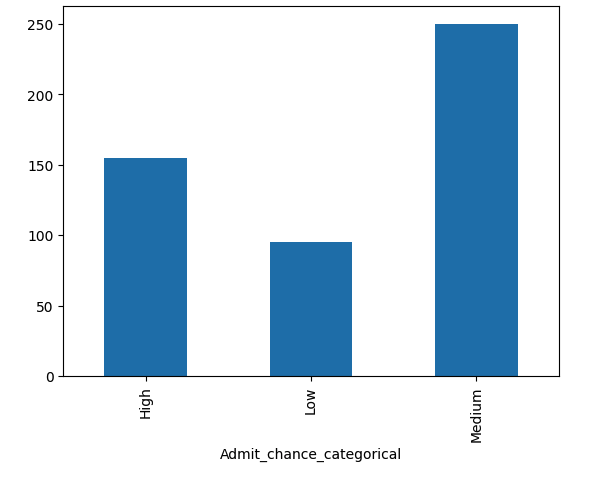
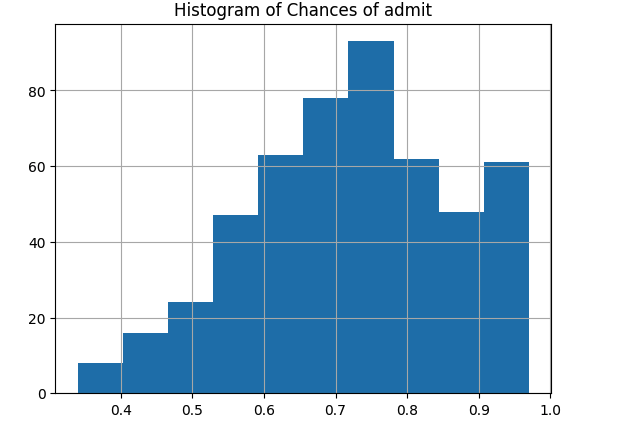
* Mean squared error: 0.0030
* R2 score: -0.8633
* Below is the visualization of prediction vs actual for each variable.



**Classification**

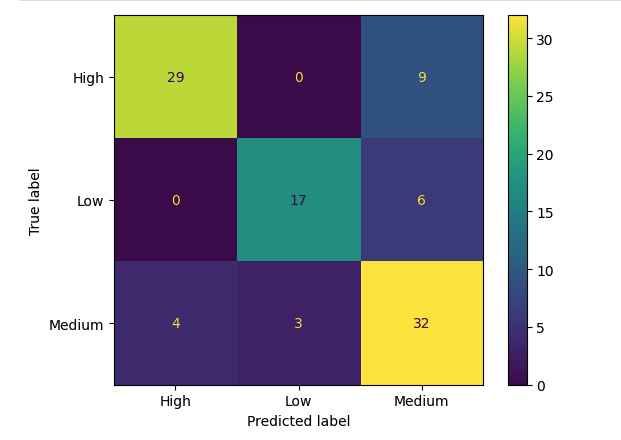
**Dataset Transformation:**

* To convert the data from the regression column to the classification problem, we created different bins of chances of admission and made it a categorical variable that contains values like high, medium, and low based on the probability chance.
* Below is the visualization for the histogram of the chances of admission before and after the transformation.

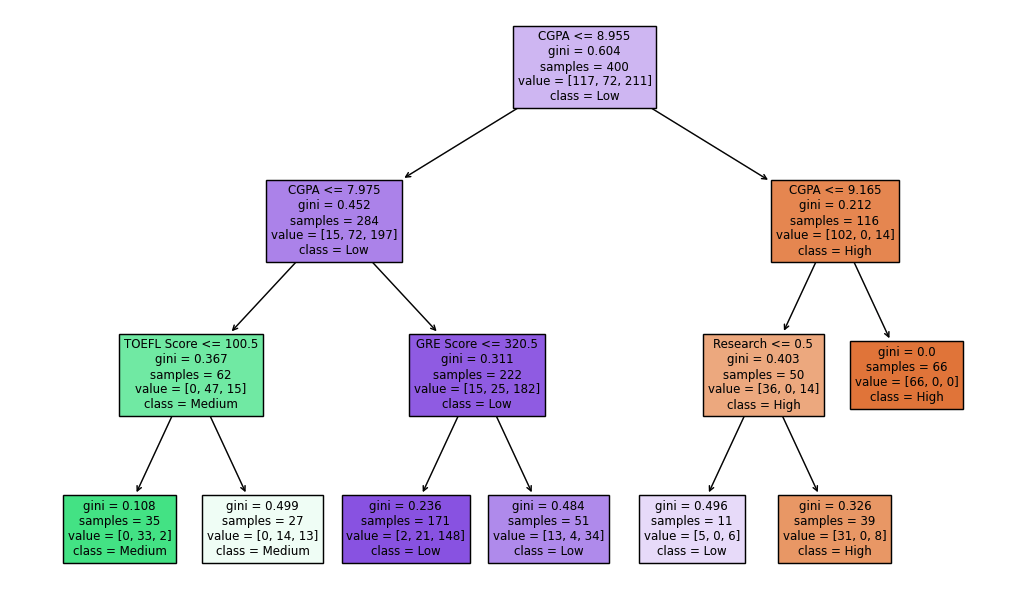
****

**Applying Decision Tree:**

* As we are applying the decision tree we didn’t normalize the dataset as it can handle it without normalization of the data.
* We got an accuracy score: of 0.78
* Confuson matrix:

****

* Below is the decision tree we are getting after training the data, for better understanding we kept the max depth as 3.



# **Conclusion**

* From the above rule it is pretty much clear that above CGPA 9.16, the chance of admission is high.
* If the CGPA is above 8.95 and less than 9.16 and someone has an experience in research then also the chance of admission is high.
* If CGPA scores more than 7.97 and less than 8.95 then the chances are medium.
* Similarly, if the CGPA is less than 7.97, then the chances are Low.