**Lab3 - Quantitative Metrics**

Previously I was able to resolve the missing value problem and the data type problem. After solving this problem, I saved the current state of the dataset. Then I read the data again and standardized it by using ‘’LabelEncoder’‘ from Scikit-Learn library. Previously I was able to find the best columns/features by feature selection using different techniques. I was able to find 30 best features that were best related to the target variable. So I dropped the rest of the columns and perform a Linear Regression algorithm on it.

I checked the accuracy score on the training data. The results was equal to 1. By looking at this I thought my model had overfit, so I checked the results with the testing data.

To do so I visualized the results on a scatter plot that compared the difference between the actual output and predicted output. I found out that they had almost the same graph. This showed that the results were very much accurate. I then performed three quantitative metrics on it – MSE, MAE, RMSE. They all showed the error was very close to zero almost equal to zero. This showed that my model was performing very well.

This high accuracy is justifiable, because the correlation of different features with the target variable was very high. Some had correlations of more that 0.85. Plus by repeated and recursive process of feature selection I was able to find best 30 features out of 55 columns. This would I have improved the results by a lot.