**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ**

**ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРОЗРВАНИЯ**

****

**«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ»**

**(Национальный исследовательский университет)**

Кафедра 101

«Проектирование и сертификация авиационной техники»

Лабораторная работа № 1&2

По дисциплине ;Моделирование инженерных задач

Тема: Laboratory report

Выполнил:

Студент гр; М10-415Бки-19

Студент ; Кодзанайи Грациано

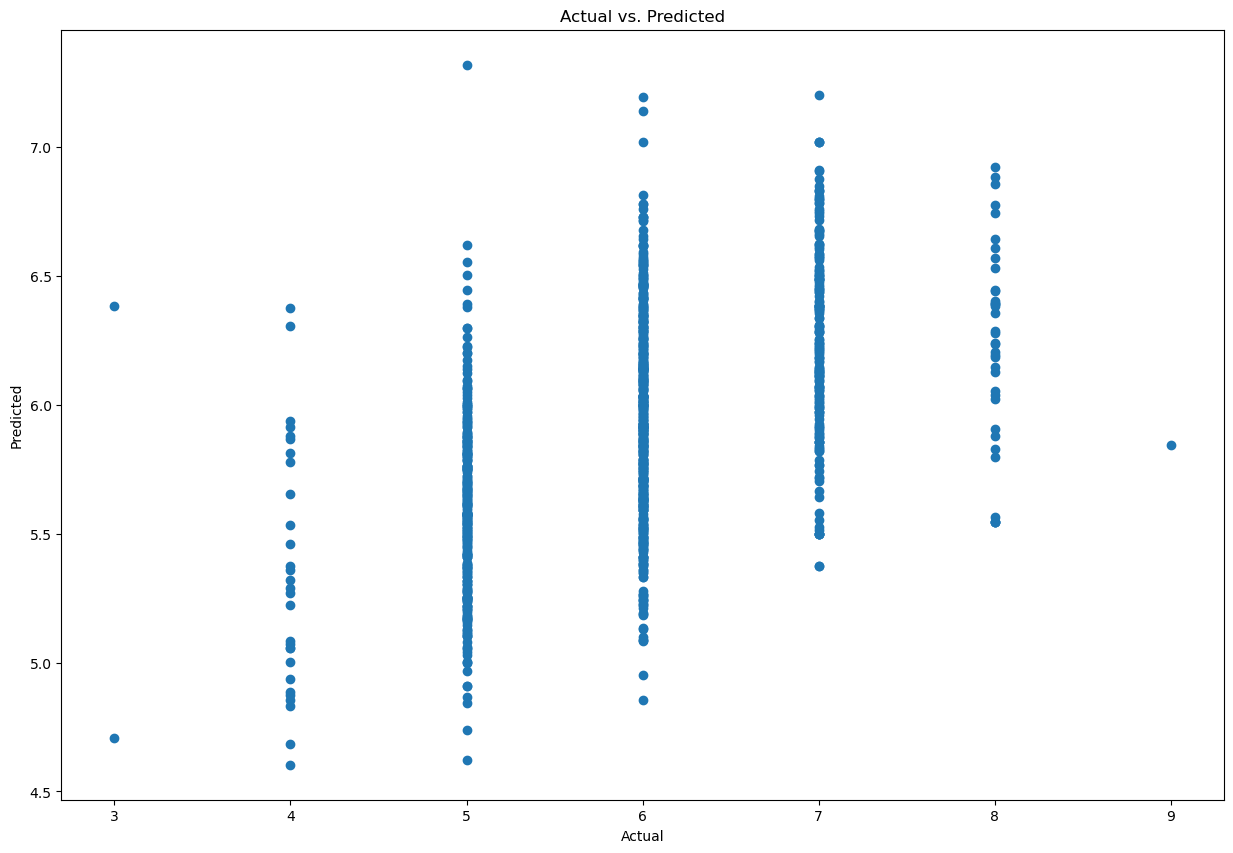
2022г

LABORATORY REPORT

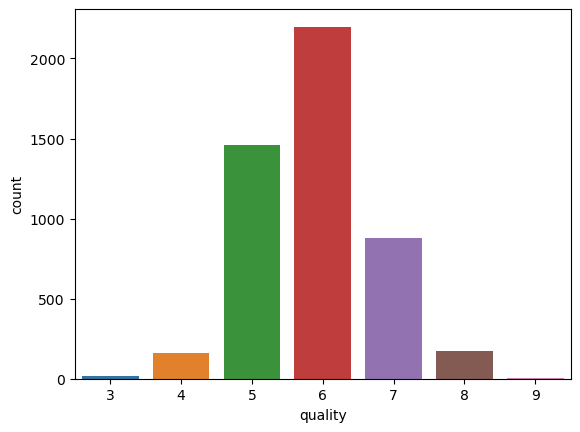
I created a model which will predict the type of the wine(white), according to it's features.I used White Wine Quality dataset which is a tidy data set. This dataset contains 4 898 white wines with 12 variables on quantifying the chemical properties of each wine. At least 3 wine experts rated the quality of each wine, providing a rating between 0 (very bad) and 10 (very excellent) Objective: To identify which chemical properties influence the quality of white wines? I used the method of multiple linear regression at first, then after obtaining my model, i was under the presumption that if i ufurther clean and analyse my data and use the most correlated features on a relative scale instead of all the features theat would make the model better but that was not it. Most of the wines get a quality rating of five or six, while having good and bad wines seems more unlikely. Excellent wines >8 are absent.

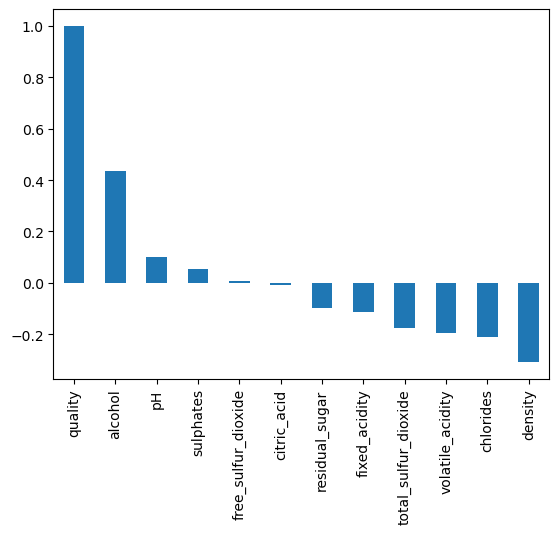
Firstly I  upload the dataset and find variables in the dataset.

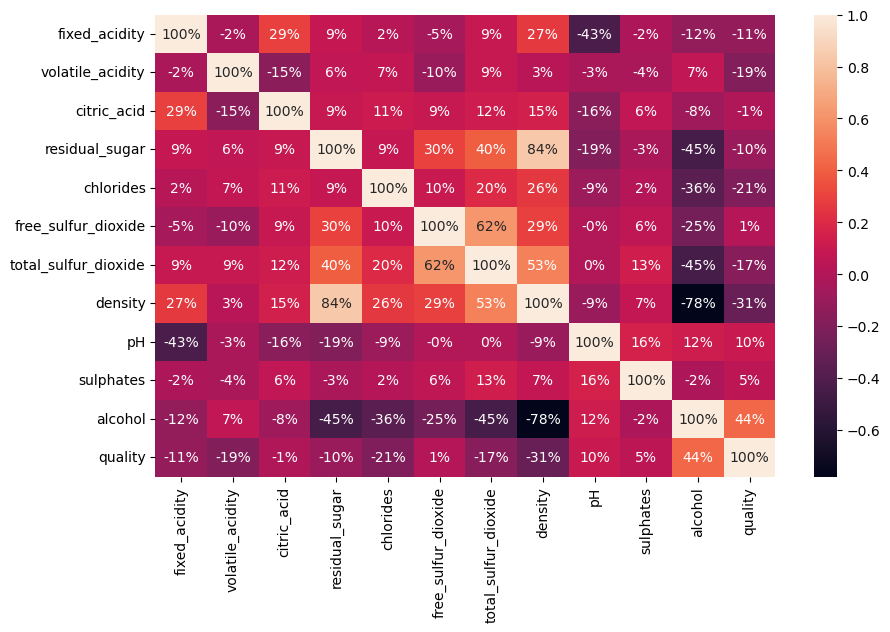
Secondly was the classificatoin of wine , the rating as 0 very bad and 10 very excellent and the results was as follows;



Next was visualizing and correlation the different quality values and how many wines have that rating in our dataset and the results where as follows







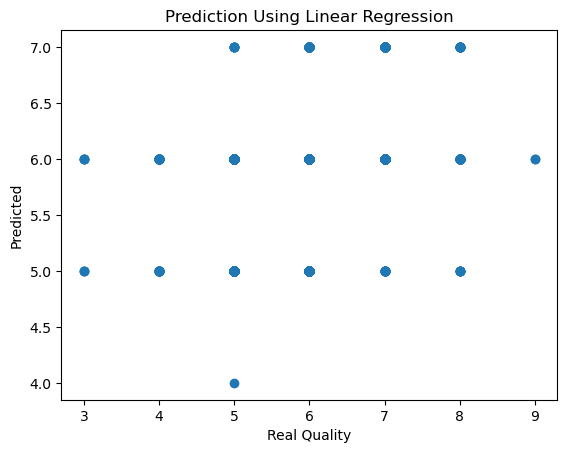
Quality is moderately correlated with alcohol

Alcohol is moderately correlated with the density of wine apart from being moderaltely correlated with quality

Density is strongly correlated with residual sugar quantity and moderately correlated with pH

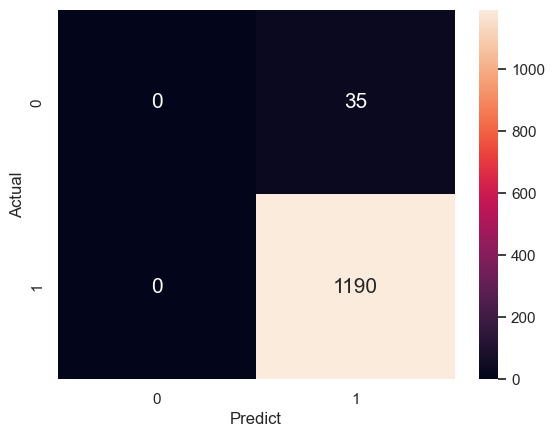
Free sulfur dioxide and total sulfur dioxide are strongly correlated

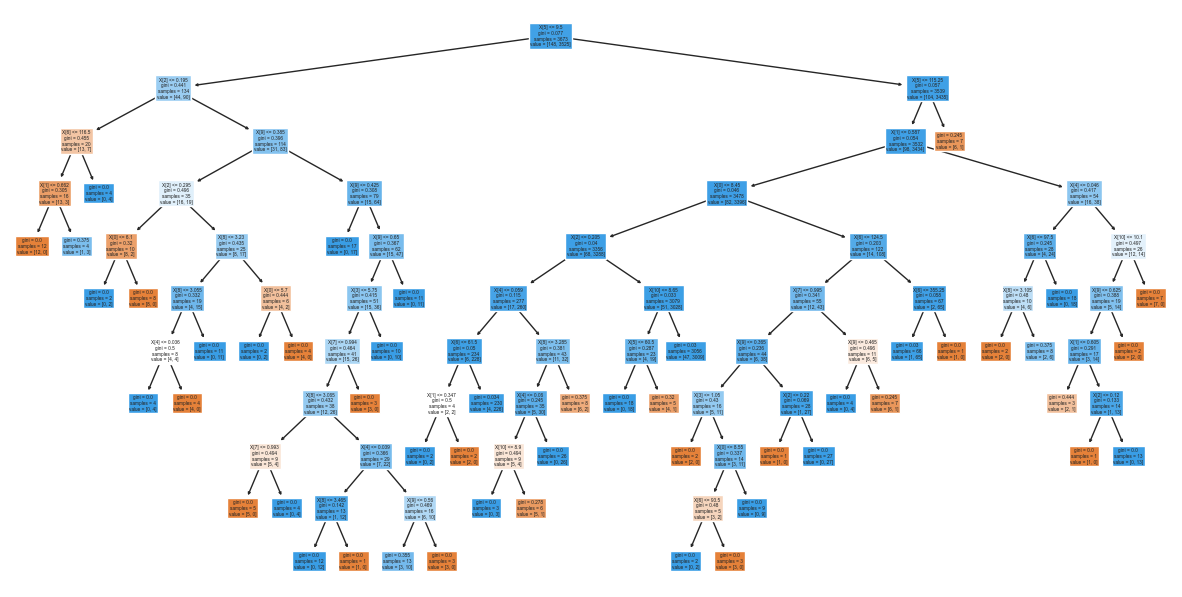
now was the ****Linear Regression Model****  
I created a linear regression model and find out which variables are significant. I improved my model by removing the insignificant variables and checking for improvement in R-Squared value. So, first I created a model to predict the quality of the wine. First I have to split dataset into training and testing sets. Training set I used to craete our model and predict over testing set.



## A graph showing the performance of an ambiguous classifier(ROC Curve), due to the decision threshold set for it. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) under different acceptance thresholds

#### After I used cross validatoin and the goal of cross-validation is to test the model's ability to predict new data that was not used in estimating it, in order to flag problems like overfitting or selection bias and to give an insight on how the model will generalize to an independent dataset.





#### *I used a decision tree which is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. Criterion is the function to measure the quality of a split. Supported criteria are “gini” for the Gini impurity and “entropy” for the information gain. Splitter is the strategy used to choose the split at each node. Supported strategies are “best” to choose the best split and “random” to choose the best random split.*

Lastly was*****Quality Rating Adjustment*****Random Forest and Conditional inference Tree model predicted with very astonishing accuracy.