# Logistic Regression

**About The Model**

It’s a machine learning algorithm which is used to predict the probability of categorical dependent variable.

The dependent variable in the data contain as 0 (failure,etc..) & 1 (success,etc..).

**Binary or Binomial Regression:** The dependant variable has just two potential outcomes in this sort of classification: 1 or 0. For example, these variables may indicate success or failure, yes or no, victory or loss, and so on.

**Multinomial Regression:** The dependent variable in this form of classification might have three or more potential unordered outcomes, or it can have no quantitative meaning. These variables might stand for "Type A," "Type B," or "Type C," for example.

**Ordinal Regression:** In this classification, the dependent variable may have three or more quantitatively significant ordered outcomes or outcomes. With ratings ranging from 0 to2, these factors might indicate "poor" or "excellent," "very good," or "Excellent."

**Dataset**

The given data is red wine quantity, 1599 rows with 12 columns. Data set loaded as **df** using pandas package.

The column name are ['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',

'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',

'pH', 'sulphates', 'alcohol', 'quality']

A screenshot of a computer

Description automatically generated with medium confidence

**Data Pre-Processing**

Getting the all the data columns, their types and the count of rows. The float values are 11 columns and the other is int type

**A screenshot of a computer

Description automatically generated with medium confidence**

Checking the data with any null values.

**Text

Description automatically generated with medium confidence**

Creating new column using the quality, as Quality of Wine with quality less than 5 is **“0”** & other as **“1”**.



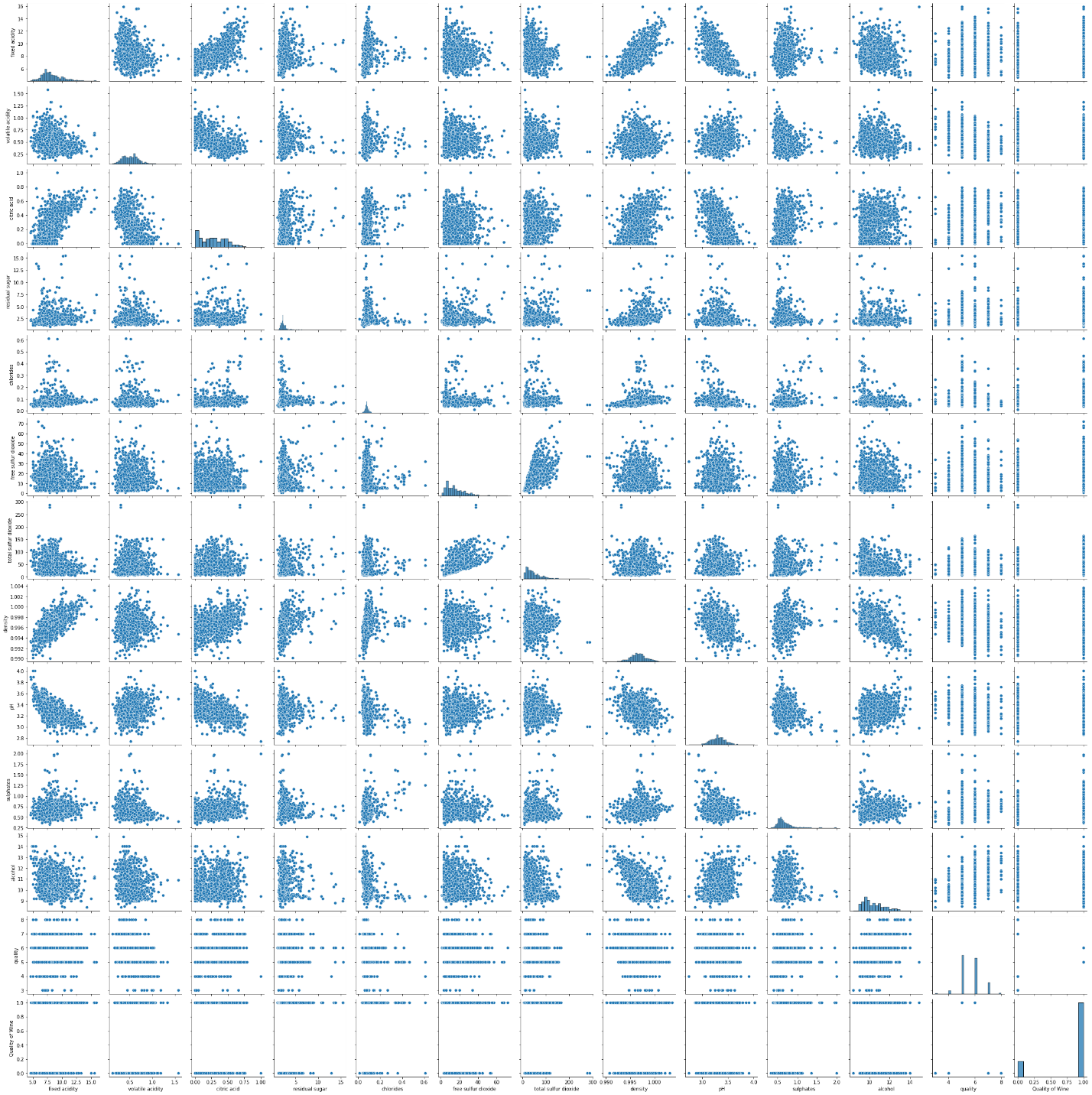
Getting count values of new column

Graphical user interface, text

Description automatically generated

**Data visualization**

In the below fig., created plots for the each variable in the data set.



The fig. below shows correlation matrix with plot and the heat map.

Text

Description automatically generated with medium confidence

**Splitting the data**

The given data set is spitted into 2 parts initially **X** as all the data columns with out “Quality of Wine” & **y**  as only “Quality of Wine” column.

The data set splitted int training dataset & test dataset with train data as 80% and test data as 20% (where test\_size =0.2)

The training and testing is done by sklearn package.

Text

Description automatically generated

**Applying the model**

Applying the logistic regression from sklearn package by importing, with maximum iteration is 10000, cross validation is 80 & class type is **multinomial.**

**Text

Description automatically generated**

**Prediction results and test score**

Text

Description automatically generated

The accuracy for the predicted data is 90%.

**Metrics**

Calculating the accuracy score, Confusion matrix, Recall score & Precision score.

Text

Description automatically generated

Calendar

Description automatically generated

**Conclusion**

The result of confusion matrix are, that we have 30+259 correct predictions and 28+3 incorrect predictions( the values are taken from results of confusion matrix above).

The accuracy of the test data is 90%, which shows the quality of the wine is red.

The Recall score for the test data is 0.5172 and predicted data score is 0.988

The precision score for the test data is 0.90 and the predicted data score is 0.90