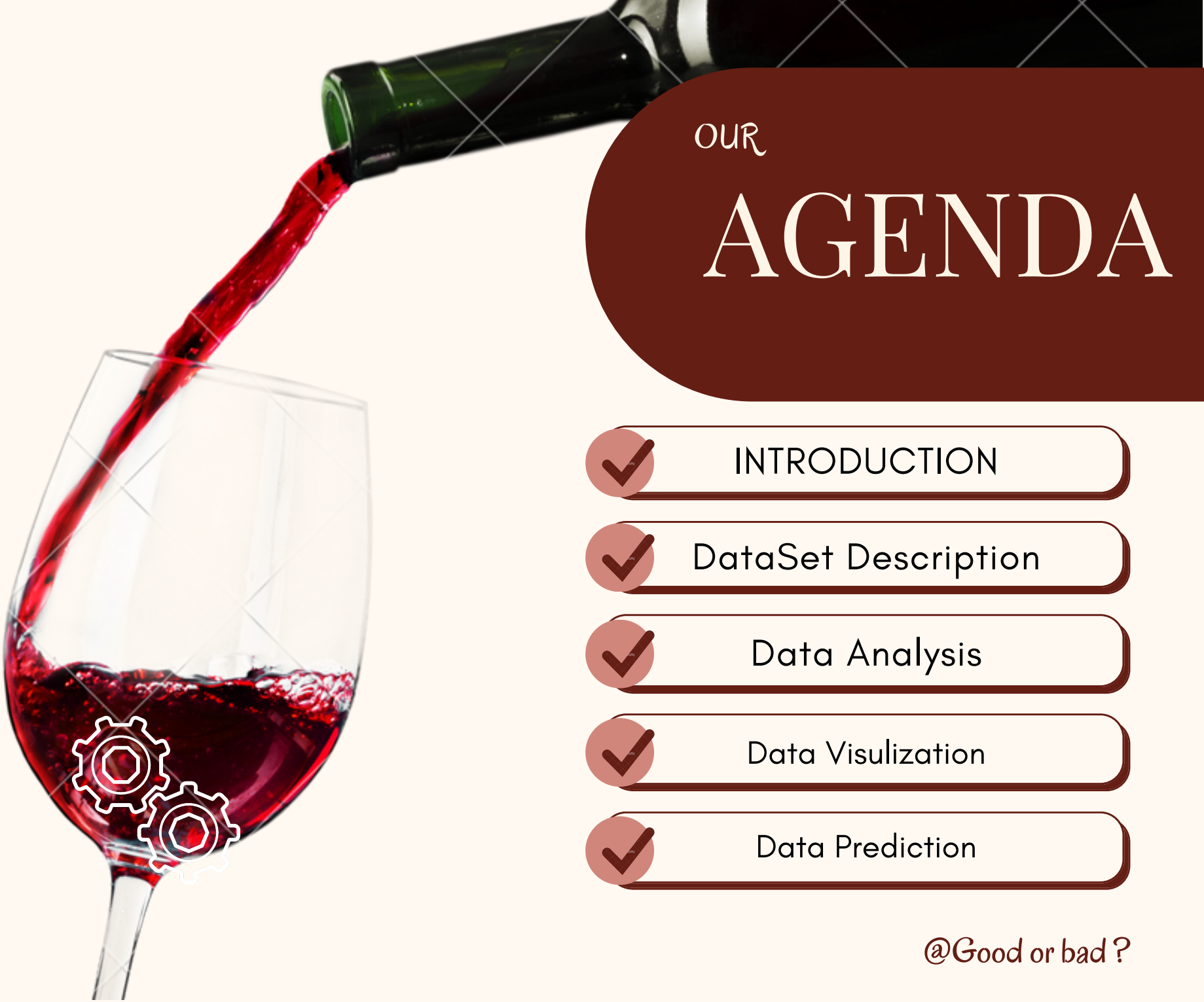
WINE QUALITY PREDICITION

Team Members

TEAM 32

Yash Choudhary



OUR

AGENDA

- INTRODUCTION
- DataSet Description
- Data Analysis
- Data Visulization
- Data Prediction

@Good or bad ?

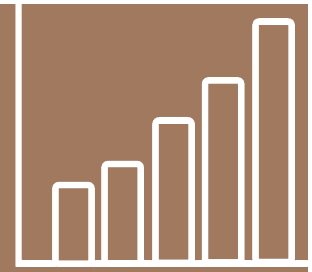
INTRODUCTION

The quality of the wine is a very important part for the consumers as well as the manufacturing industries. Industries are increasing their sales using product quality certification

Nowadays, all over the world wine is a regularly used beverage and the industries are using the certification of product quality to increases their value in the market.

The wine quality dataset is publically available on the Kaggle repository.

Dataset Description



The Attributes are as follows

1)fixed acidity

2) volatile acidity

3) citric acid

4) residual sugar

5) chlorides

6)free sulfur dioxide

7)total sulfur dioxide

8)density

9)pH

10) sulphates

11) alcohol Output variable (based on sensory
data):

12)quality (score between 0 and 10)

Data Analysis

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792	0.996747	3.311113	0.658149	10.422983
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	0.001887	0.154386	0.169507	1.065668
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.990070	2.740000	0.330000	8.400000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	0.995600	3.210000	0.550000	9.500000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	0.996750	3.310000	0.620000	10.200000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	0.997835	3.400000	0.730000	11.100000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000	1.003690	4.010000	2.000000	14.900000

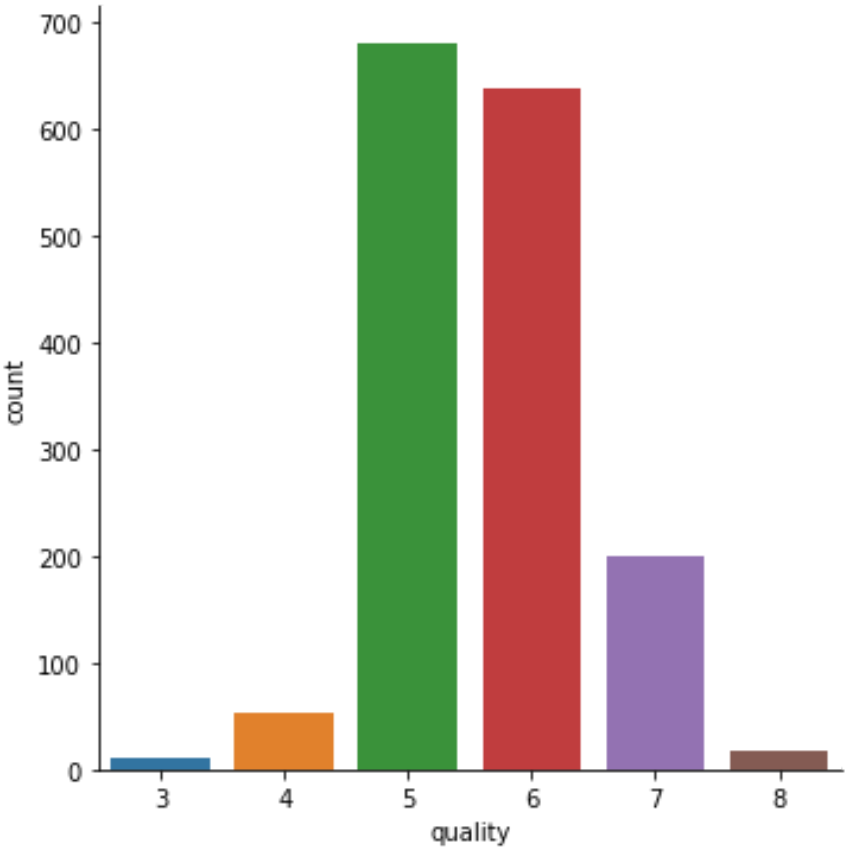
Descriptive statistical measures of the dataset.

Data Analysis

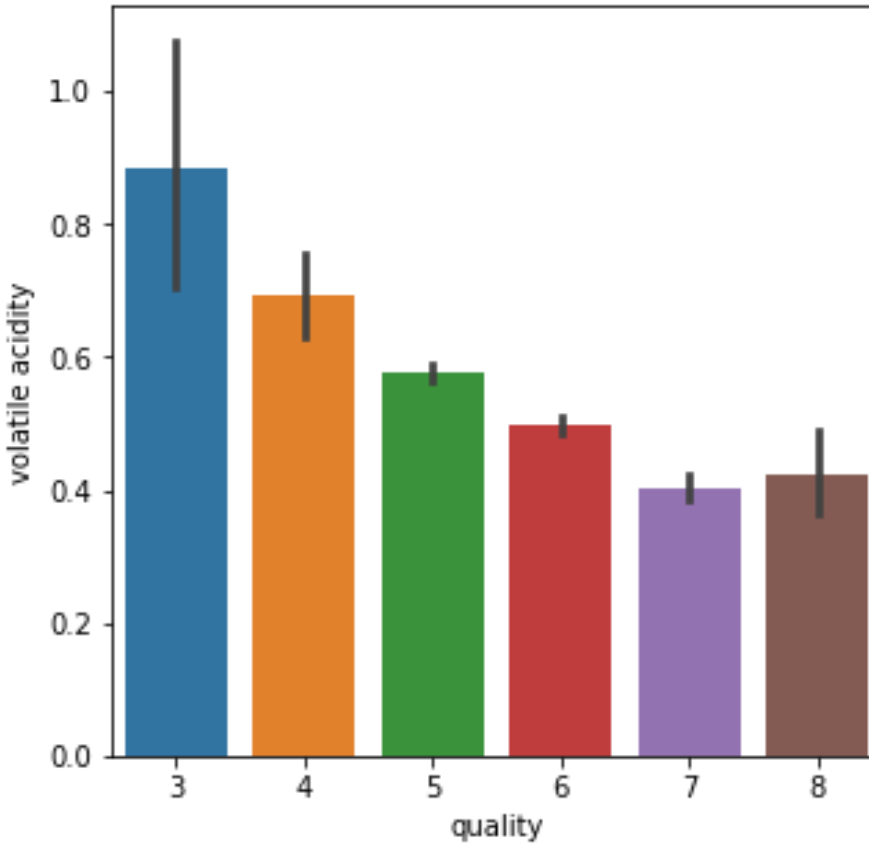
Sum of all the NULL values in the dataset(if present) of all the attributes

```
fixed acidity           0
volatile acidity        0
citric acid             0
residual sugar          0
chlorides               0
free sulfur dioxide     0
total sulfur dioxide    0
density                 0
pH                      0
sulphates               0
alcohol                 0
quality                 0
dtype: int64
```

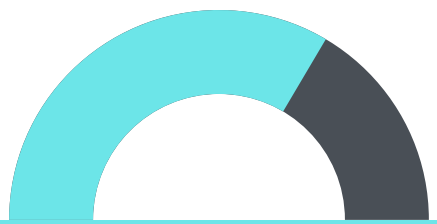
Data Visulization



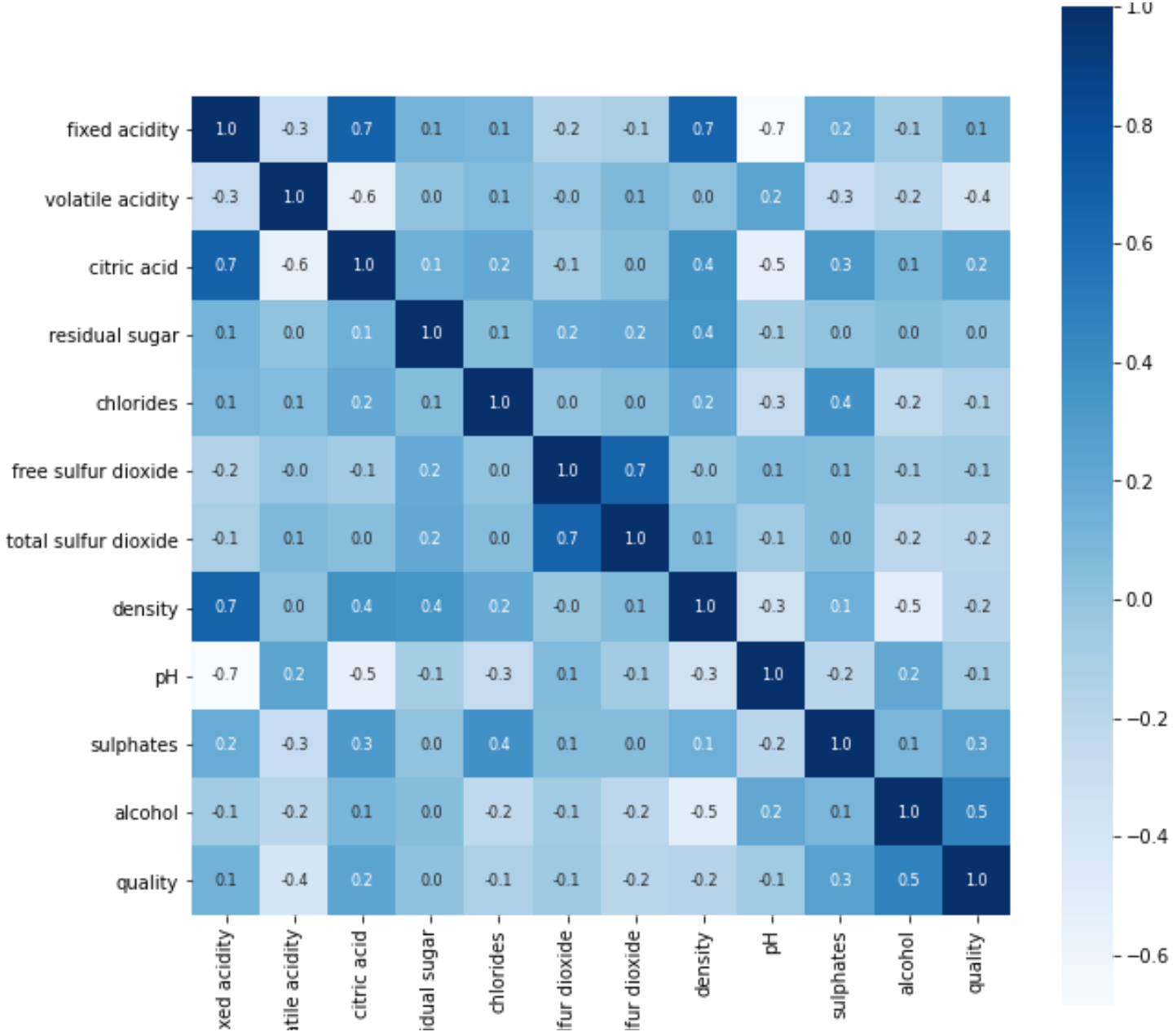
count of different quality of wine



Graph representing the relationship between quality and volatile acidity.



HEAT MAP


fixed acidity 1.0 -0.3 0.7 0.1 0.1 -0.2 -0.1 0.7 -0.7 0.2 -0.1 0.1
volatile acidity -0.3 1.0 -0.6 0.0 0.1 -0.0 0.1 0.0 0.2 -0.3 -0.2 -0.4
citric acid 0.7 -0.6 1.0 0.1 0.2 -0.1 0.0 0.4 -0.5 0.3 0.1 0.2
residual sugar 0.1 0.0 0.1 1.0 0.1 0.2 0.2 0.4 -0.1 0.0 0.0 0.0
chlorides 0.1 0.1 0.2 0.1 1.0 0.0 0.0 0.2 -0.3 0.4 -0.2 -0.1
free sulfur dioxide -0.2 -0.0 -0.1 0.2 0.0 1.0 0.7 -0.0 0.1 0.1 -0.1 -0.1
total sulfur dioxide -0.1 0.1 0.0 0.2 0.0 0.7 1.0 0.1 -0.1 0.0 -0.2 -0.2
density 0.7 0.0 0.4 0.4 0.2 -0.0 0.1 1.0 -0.3 0.1 -0.5 -0.2
pH -0.7 0.2 -0.5 -0.1 -0.3 0.1 -0.1 -0.3 1.0 -0.2 0.2 -0.1
sulphates 0.2 -0.3 0.3 0.0 0.4 0.1 0.0 0.1 -0.2 1.0 0.1 0.3
alcohol -0.1 -0.2 0.1 0.0 -0.2 -0.1 -0.2 -0.5 0.2 0.1 1.0 0.5
quality 0.1 -0.4 0.2 0.0 -0.1 -0.1 -0.2 -0.2 -0.1 0.3 0.5 1.0
xed acidity
itile acidity
citric acid
idual sugar
chlorides
lfur dioxide
lfur dioxide
density
pH
sulphates
alcohol
quality
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

Data Processing

- Train and Testing

```
print(Y.shape, Y_train.shape, Y_test.shape)
```

```
(1599,) (1279,) (320,)
```

Model Training:

Random Forest Classifier

```
model = RandomForestClassifier()
```

```
model.fit(X_train, Y_train)
```

```
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                       criterion='gini', max_depth=None, max_features='auto',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100,
                       n_jobs=None, oob_score=False, random_state=None,
                       verbose=0, warm_start=False)
```

Accuracy testing

```
[ ]  # accuracy on test data
     X_test_prediction = model.predict(X_test)
     test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
```

```
[ ]  print('Accuracy : ', test_data_accuracy)
```

```
     Accuracy :  0.928125
```

CONCLUSION

IN CONCLUSION, WE HAVE FOUND THE INTER-RELATION BETWEEN THE QUALITY OF THE WINE AND THE DIFFERENT COMPONENTS USED IN THE MAKING OF THE WINE.

FOR CLASSIFYING THE WINE QUALITY, WE HAVE IMPLEMENTED RANDOM FOREST CLASSIFIER ALGORITHMS,

In the end, we have implemented the predictivity system to find the quality of the wine based on the given dataset and with a accuracy of around 92%.

Thank You