Red Wine Quality Prediction

-Dhruv Darda

Index

Understanding Dataset

- volatile acidity : Volatile acidity is the gaseous acids present in wine. Leading to unpleasant taste
- fixed acidity : Primary fixed acids found in wine are tartaric, succinic, citric, and malic
- residual sugar : Amount of sugar left after fermentation.
- citric acid : It is weak organic acid, found in citrus fruits naturally. Adds freshness and flavour
- chlorides : Amount of salt present in wine.
- free sulfur dioxide : So2 is used for prevention of wine by oxidation and microbial spoilage.
- total sulfur dioxide: total amount of SO2 present.
- pH : In wine pH is used for checking acidity
- density: sweeter wines have higher density.
- sulphates : Added sulfites preserve freshness and protect wine from oxidation, and bacteria.
- alcohol : Amount of alcohol present in wine.

Data Description

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
count	1599.000000	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	5.636023
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	0.001887	0.154386	0.169507	1.065668	0.807569
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.990070	2.740000	0.330000	8.400000	3.000000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	0.995600	3.210000	0.550000	9.500000	5.000000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	0.996750	3.310000	0.620000	10.200000	6.000000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	0.997835	3.400000	0.730000	11.100000	6.000000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000	1.003690	4.010000	2.000000	14.900000	8.000000

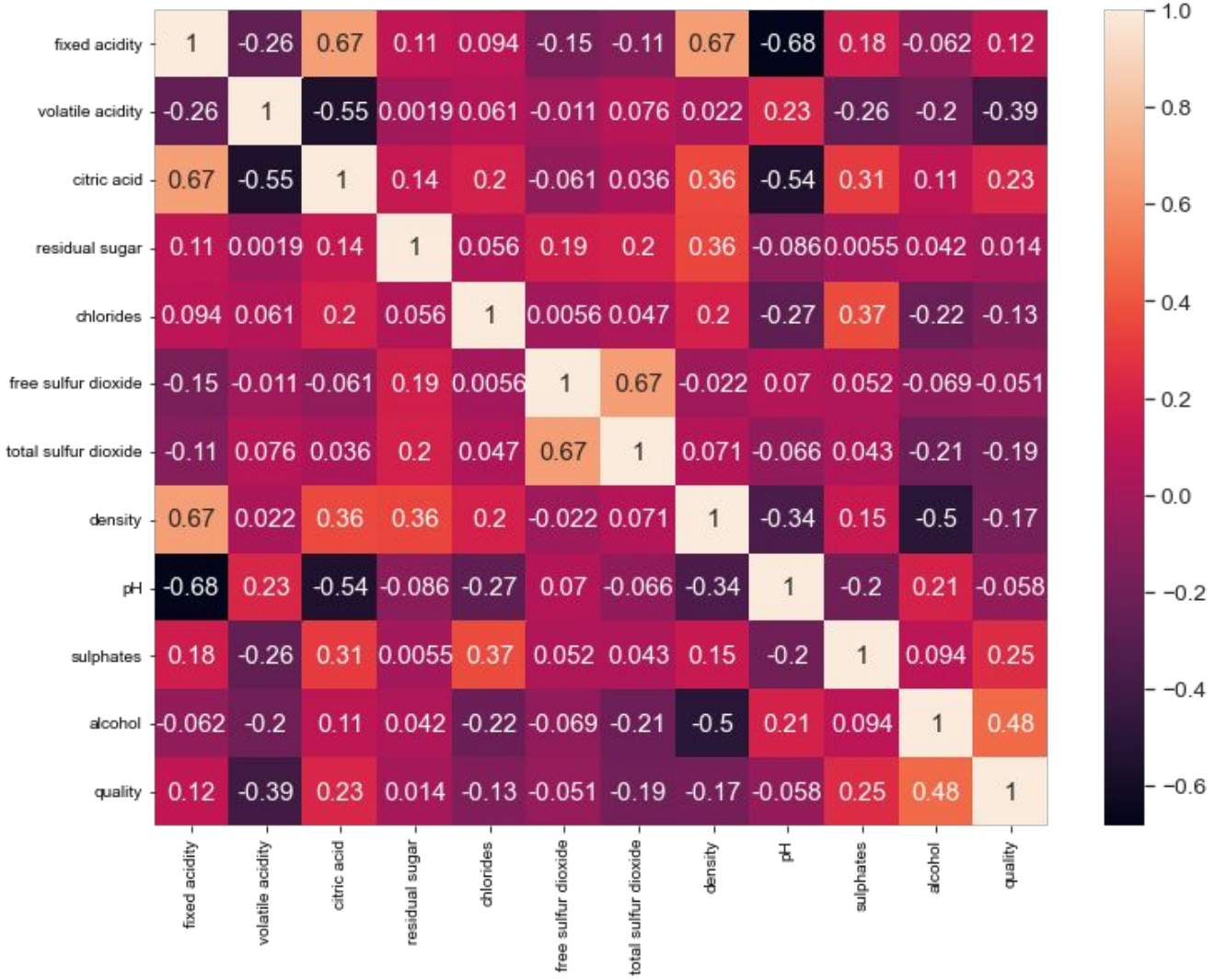
Data Description

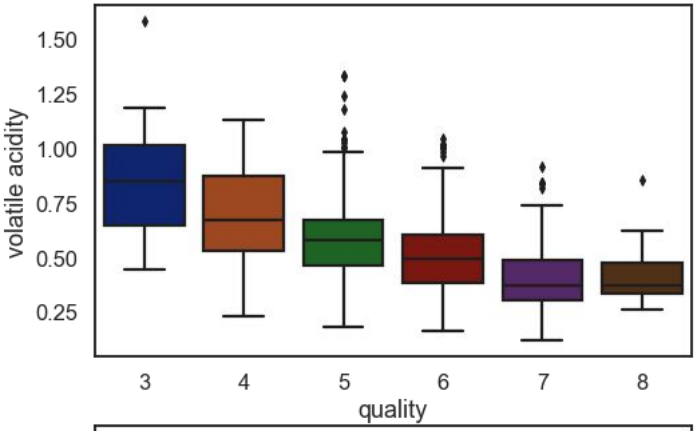
```
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   fixed acidity         1599 non-null   float64
 1   volatile acidity      1599 non-null   float64
 2   citric acid           1599 non-null   float64
 3   residual sugar        1599 non-null   float64
 4   chlorides             1599 non-null   float64
 5   free sulfur dioxide   1599 non-null   float64
 6   total sulfur dioxide  1599 non-null   float64
 7   density               1599 non-null   float64
 8   pH                    1599 non-null   float64
 9   sulphates             1599 non-null   float64
 10  alcohol               1599 non-null   float64
 11  quality               1599 non-null   int64
dtypes: float64(11), int64(1)
```

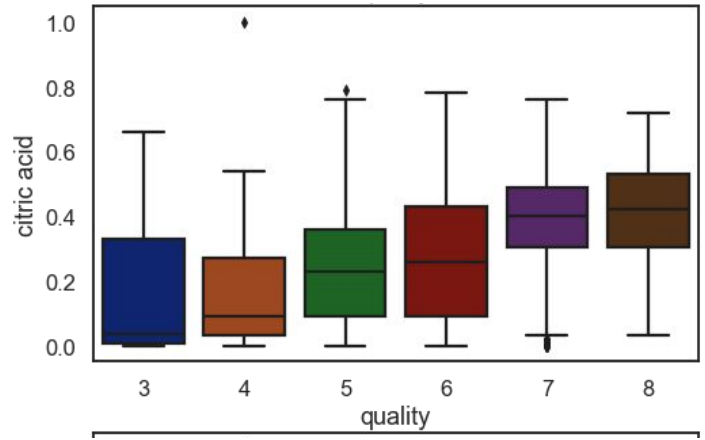
Visualisation

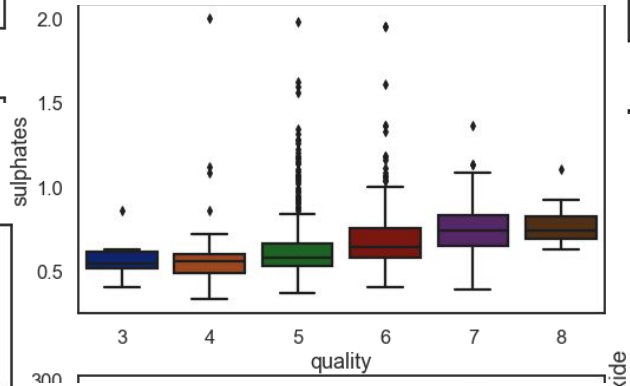
Correlations

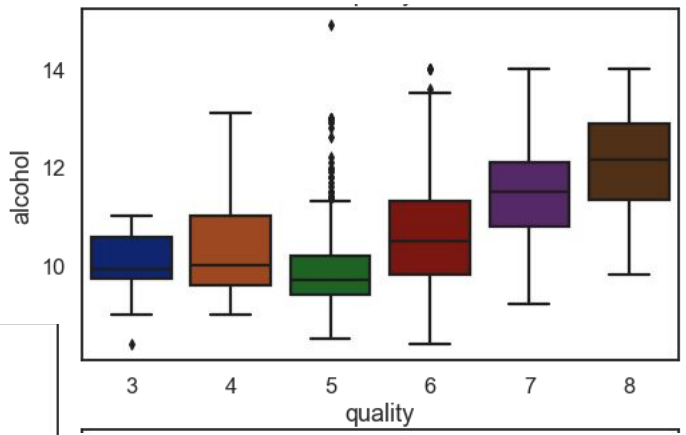
- Volatile acidity
- Total SO_2
- Alcohol
- Sulphates
- Citric acid

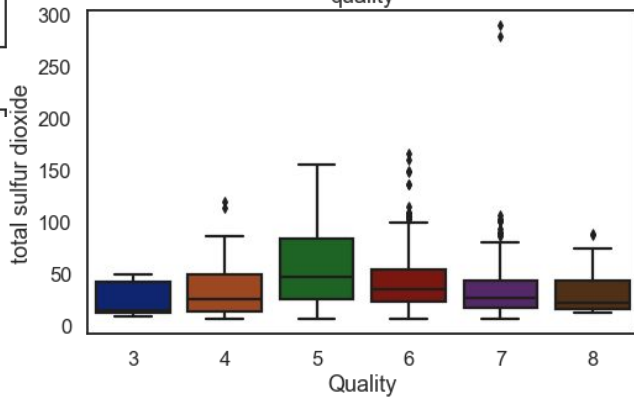



volatile acidity
1.50
1.25
1.00
0.75
0.50
0.25
3
4
5
6
7
8
quality


citric acid
1.0
0.8
0.6
0.4
0.2
0.0
3
4
5
6
7
8
quality


sulphates
2.0
1.5
1.0
0.5
3
4
5
6
7
8
quality


alcohol
14
12
10
3
4
5
6
7
8
quality


total sulfur dioxide
300
250
200
150
100
50
0
3
4
5
6
7
8
Quality

Box plots

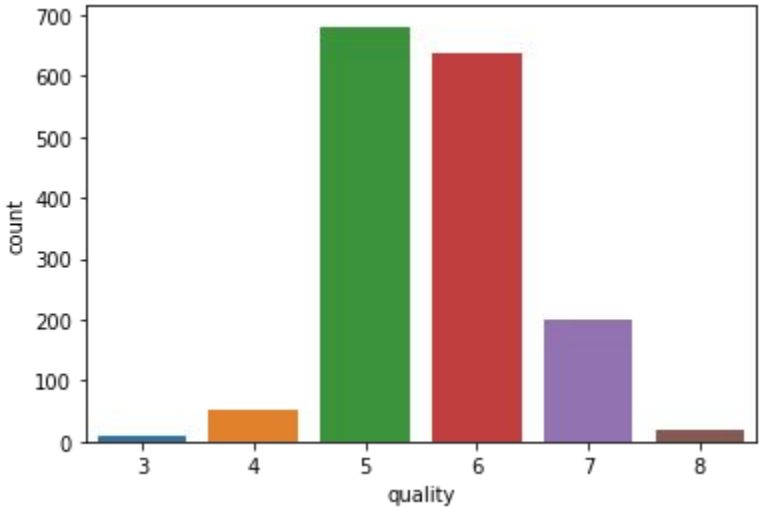
Data Normalization and Transformation

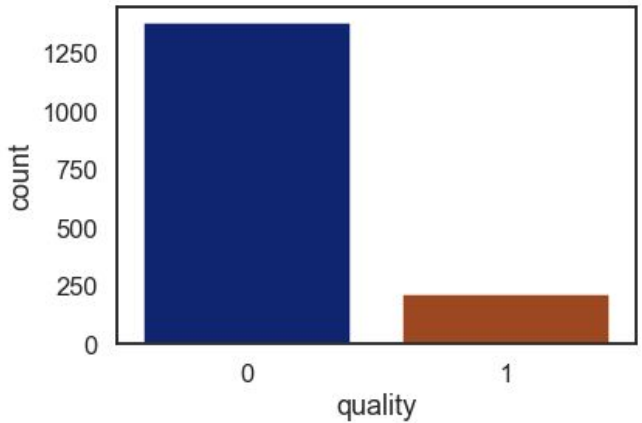
Normalization:

StandardScaler

MinMaxScaler

Transformation:

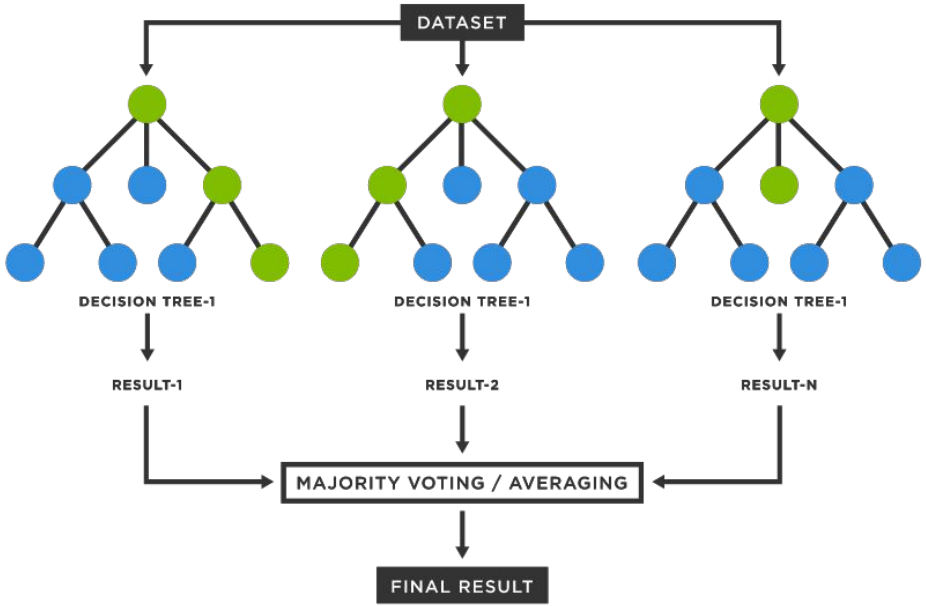




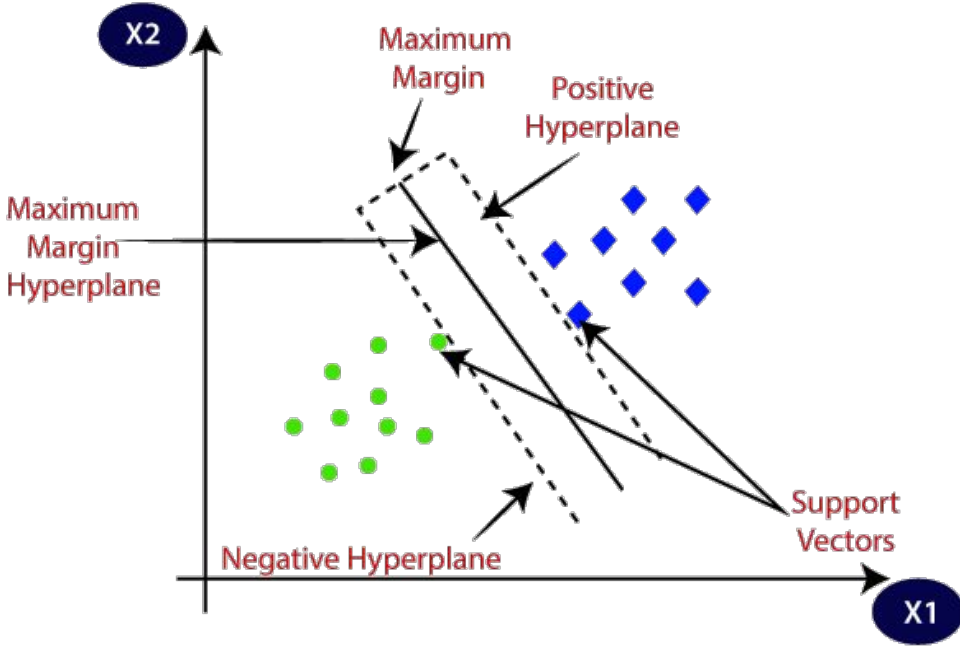
Total quality count

Models:

Random Forest Classifier



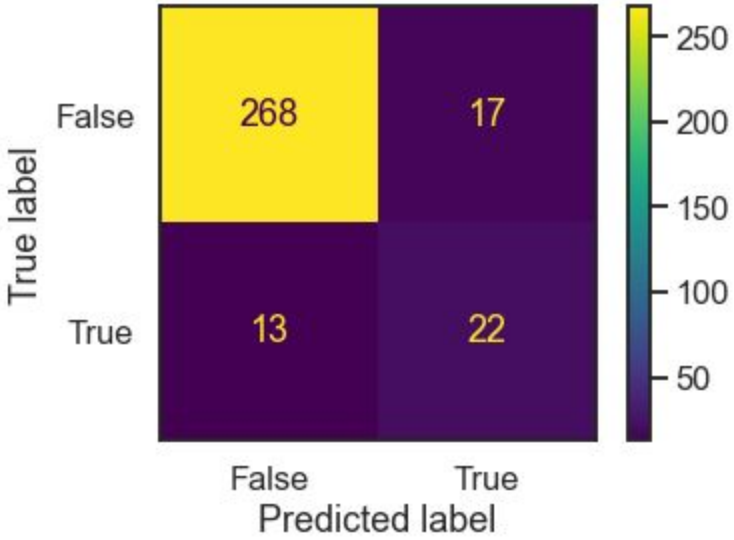
SVC



Results

Results

Confusion Matrix:



Classification Report:

	precision	recall	f1-score	support
0	0.95	0.94	0.95	285
1	0.56	0.63	0.59	35
accuracy			0.91	320
macro avg	0.76	0.78	0.77	320
weighted avg	0.91	0.91	0.91	320

Cross-Validation