

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn import metrics

In [2]: wine = pd.read_csv('D:\\24 - Machine_Learning\\download_files\\winequality-red.csv', sep=';')
wine

Out[2]:
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	
0	7.4	0.700	0.00	1.9	0.076		11.0	34.0	0.99780	3.51	0.56	9.4	5
1	7.8	0.880	0.00	2.6	0.098		25.0	67.0	0.99680	3.20	0.68	9.8	5
2	7.8	0.760	0.04	2.3	0.092		15.0	54.0	0.99700	3.26	0.65	9.8	5
3	11.2	0.280	0.56	1.9	0.075		17.0	60.0	0.99800	3.16	0.58	9.8	6
4	7.4	0.700	0.00	1.9	0.076		11.0	34.0	0.99780	3.51	0.56	9.4	5
...
1594	6.2	0.600	0.08	2.0	0.090		32.0	44.0	0.99490	3.45	0.58	10.5	5
1595	5.9	0.550	0.10	2.2	0.062		39.0	51.0	0.99512	3.52	0.76	11.2	6
1596	6.3	0.510	0.13	2.3	0.076		29.0	40.0	0.99574	3.42	0.75	11.0	6
1597	5.9	0.645	0.12	2.0	0.075		32.0	44.0	0.99547	3.57	0.71	10.2	5
1598	6.0	0.310	0.47	3.6	0.067		18.0	42.0	0.99549	3.39	0.66	11.0	6

1599 rows × 12 columns

```
In [3]: wine.info()

<class 'pandas.core.frame.DataFrame'>
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)
memory usage: 150.0 KB

In [4]: wine.isnull

Out[4]:
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	
0	7.4	0.700	0.00	1.9	0.076		11.0	34.0	0.99780	3.51	0.56	9.4	5
1	7.8	0.880	0.00	2.6	0.098		25.0	67.0	0.99680	3.20	0.68	9.8	5
2	7.8	0.760	0.04	2.3	0.092		15.0	54.0	0.99700	3.26	0.65	9.8	5
3	11.2	0.280	0.56	1.9	0.075		17.0	60.0	0.99800	3.16	0.58	9.8	6
4	7.4	0.700	0.00	1.9	0.076		11.0	34.0	0.99780	3.51	0.56	9.4	5
...
1594	6.2	0.600	0.08	2.0	0.090		32.0	44.0	0.99490	3.45	0.58	10.5	5
1595	5.9	0.550	0.10	2.2	0.062		39.0	51.0	0.99512	3.52	0.76	11.2	6
1596	6.3	0.510	0.13	2.3	0.076		29.0	40.0	0.99574	3.42	0.75	11.0	6
1597	5.9	0.645	0.12	2.0	0.075		32.0	44.0	0.99547	3.57	0.71	10.2	5
1598	6.0	0.310	0.47	3.6	0.067		18.0	42.0	0.99549	3.39	0.66	11.0	6

```
In [5]: sns.heatmap(wine.isna())

<AxesSubplot:~>

Out[5]:
```

```
In [6]: wine.shape

Out[6]: (1599, 12)

In [7]: wine.describe()

Out[7]:
```

	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	2.538906	0.067467	15.874922	46.467792	0.996747	3.211113	0.658149	10.422983	5.626023
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

```
In [8]: features=wine[['fixed acidity', 'volatile acidity', 'citric acid', 'chlorides', 'total sulfur dioxide', 'density', 'sulphates']].values
classes = wine['quality'].values

In [9]: (train_feat, test_feat, train_classes, test_classes) = train_test_split(features, classes, train_size = 0.9, random_state=100)

In [10]: # TRAINING
dectree = DecisionTreeClassifier(criterion = 'gini') #- used to calculate gini
# dectree = DecisionTreeClassifier(criterion = 'entropy') #- used to calculate entropy
dectree.fit(train_feat, train_classes)

Out[10]: DecisionTreeClassifier()

In [11]: pred = dectree.predict(test_feat)
print("Accuracy:",metrics.accuracy_score(test_classes, pred))

Accuracy: 0.64375

In [12]: print(classification_report(test_classes, pred))

              precision    recall  f1-score   support

     3         0.00         0.00         0.00         2
     4         0.00         0.00         0.00         5
     5         0.70         0.75         0.72         64
     6         0.65         0.59         0.62         68
     7         0.68         0.79         0.68         19
     8         0.00         0.00         0.00         2

 accuracy          0.32         0.35         0.34         168
 macro avg         0.32         0.64         0.63         168

D:\24-Annaconda\lib\site-packages\sklearn\metrics\_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use 'zero_division' parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
D:\24-Annaconda\lib\site-packages\sklearn\metrics\_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use 'zero_division' parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
D:\24-Annaconda\lib\site-packages\sklearn\metrics\_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use 'zero_division' parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))

In [13]: sns.pairplot(wine, hue='quality')

Out[13]: <seaborn.axisgrid.PairGrid at 0x19bf2092e0>
```

```
In [14]: fig = plt.figure(figsize=(25, 20))
_ = tree.plot_tree(dectree,
                  feature_names = features,
                  filled = True)
```

