In [1]:
```
import pandas as pd
```

In [3]:
```
df = pd.read_csv('D:\\dataset\\winequality-red.csv')
```

In [4]:
```
df.head(5)
```

Out[4]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	q
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	

In [5]:
```
df.isnull().any()
```

Out[5]:
```
fixed acidity           False
volatile acidity        False
citric acid             False
residual sugar          False
chlorides               False
free sulfur dioxide     False
total sulfur dioxide    False
density                 False
pH                      False
sulphates               False
alcohol                 False
quality                 False
dtype: bool
```

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
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
```

In [7]:
```
df.columns
```

Out[7]:
```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol', 'quality'],
      dtype='object')
```

In [8]:
```
df.shape
```

Out[8]:
```
(1599, 12)
```

In [9]:
```
x=df[['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol']]
```

In [10]:
```
y=df[['quality']]
```

In [11]:
```
from sklearn.model_selection import train_test_split
```

In [12]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

In [13]:
```
x_train.shape
```

Out[13]:
```
(1119, 11)
```

In [15]:
```
from sklearn.linear_model import LogisticRegression
```

In [18]:
```
model=LogisticRegression()
```

In [19]:
```
model.fit(x_train,y_train)
```

```
D:\data science\lib\site-packages\sklearn\utils\validation.py:73: DataConversionWarn
ing: A column-vector y was passed when a 1d array was expected. Please change the sh
ape of y to (n_samples, ), for example using ravel().
  return f(**kwargs)
D:\data science\lib\site-packages\sklearn\linear_model\_logistic.py:762: Convergence
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

Out[19]:
```
LogisticRegression()
```

In [20]:
```
y_pred=model.predict(x_test)
```

In [21]:
```
from sklearn import metrics
```

In [22]:
```
cm = metrics.confusion_matrix(y_test,y_pred)
```

In [23]:
```
metrics.accuracy_score(y_test,y_pred)
```

Out[23]:
```
0.5729166666666666
```

In [24]:
```
from sklearn.feature_selection import RFE
```

In [26]:
```
rfe= RFE(model,10)
```

```
D:\data science\lib\site-packages\sklearn\utils\validation.py:68: FutureWarning: Pas
s n_features_to_select=10 as keyword args. From version 0.25 passing these as positi
onal arguments will result in an error
  warnings.warn("Pass {} as keyword args. From version 0.25 "
```

In [27]:
```
fit=rfe.fit(x_train,y_train)
```

```
D:\data science\lib\site-packages\sklearn\utils\validation.py:73: DataConversionWarn
ing: A column-vector y was passed when a 1d array was expected. Please change the sh
ape of y to (n_samples, ), for example using ravel().
  return f(**kwargs)
```

```
D:\data science\lib\site-packages\sklearn\linear_model\_logistic.py:762: Convergence
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
D:\data science\lib\site-packages\sklearn\utils\validation.py:73: DataConversionWarn
ing: A column-vector y was passed when a 1d array was expected. Please change the sh
ape of y to (n_samples, ), for example using ravel().
  return f(**kwargs)
D:\data science\lib\site-packages\sklearn\linear_model\_logistic.py:762: Convergence
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

In [28]:
```
fit.ranking_
```

Out[28]:
```
array([1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1])
```

In [29]:
```
x.columns
```

Out[29]:
```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol'],
      dtype='object')
```

In [30]:
```
x1=x[['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide','density',
       'pH', 'sulphates', 'alcohol']]
```

In [31]:
```
x1_train,x1_test,y_train,y_test=train_test_split(x1,y,test_size=0.3)
```

In [32]:
```
model=LogisticRegression()
```

In [33]:
```
model.fit(x1_train,y_train)
```

```
D:\data science\lib\site-packages\sklearn\utils\validation.py:73: DataConversionWarn
ing: A column-vector y was passed when a 1d array was expected. Please change the sh
ape of y to (n_samples, ), for example using ravel().
  return f(**kwargs)
D:\data science\lib\site-packages\sklearn\linear_model\_logistic.py:762: Convergence
Warning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

Out[33]:
```
LogisticRegression()
```

In [34]:
```
y_pred1=model.predict(x1_test)
```

In [35]:
```
cm = metrics.confusion_matrix(y_test,y_pred1)
```

In [36]:
```
metrics.accuracy_score(y_test,y_pred1)
```

Out[36]: 0.5708333333333333

In [ ]: