In [1]:

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier,export_graphviz
from sklearn.model_selection import train_test_split,GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, confusion_matrix,roc_curve,roc_auc_score

import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
data=pd.read_csv("https://raw.githubusercontent.com/training-ml/Files/main/wine.csv")
data.head()
```

Out[2]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcoh
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9

In [3]:

```
# Any missing values?

data.isna().sum()
```

Out[3]:

```
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
Alcohol_content         0
quality                 0
dtype: int64
```

In [4]:

```
data.shape
```

Out[4]:

```
(1599, 13)
```

In [5]:

```
from sklearn.preprocessing import OrdinalEncoder
```

In [6]:

```
ord_encoder=OrdinalEncoder(categories=[['Low','Medium','High']])
df1=ord_encoder.fit_transform(data[['Alcohol_content']])
df1
```

Out[6]:

```
array([[0.],
       [1.],
       [1.],
       ...,
       [2.],
       [1.],
       [2.]])
```

In [7]:

```
data['Alcohol_content']=df1
data.head()
```

Out[7]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcoh
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9

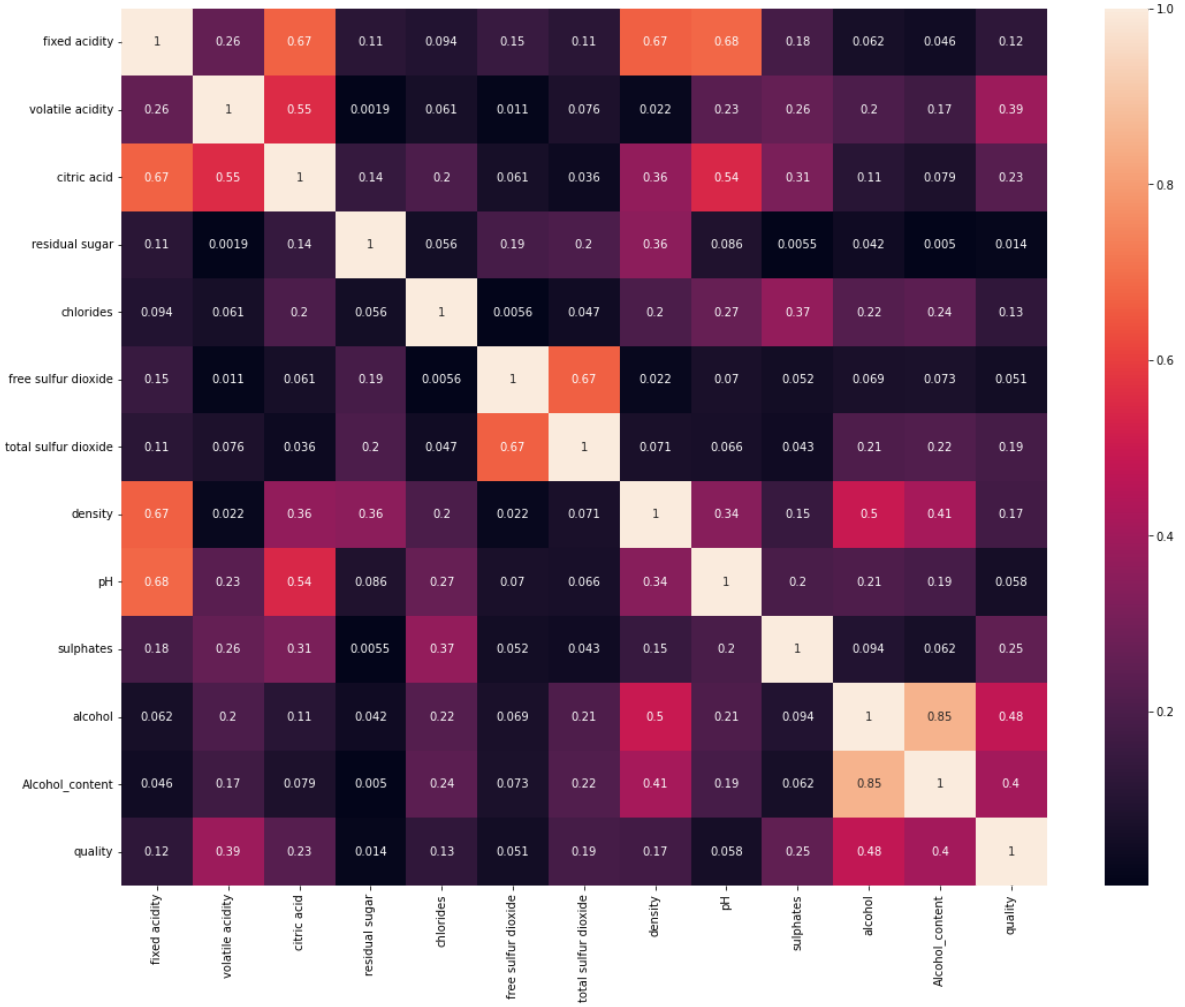
In [8]:

```
df_corr=data.corr().abs()  # This code will get the coefficient of one variable vs all othe

plt.figure(figsize=(18,14))
sns.heatmap(df_corr,annot=True, annot_kws={'size': 10})
```

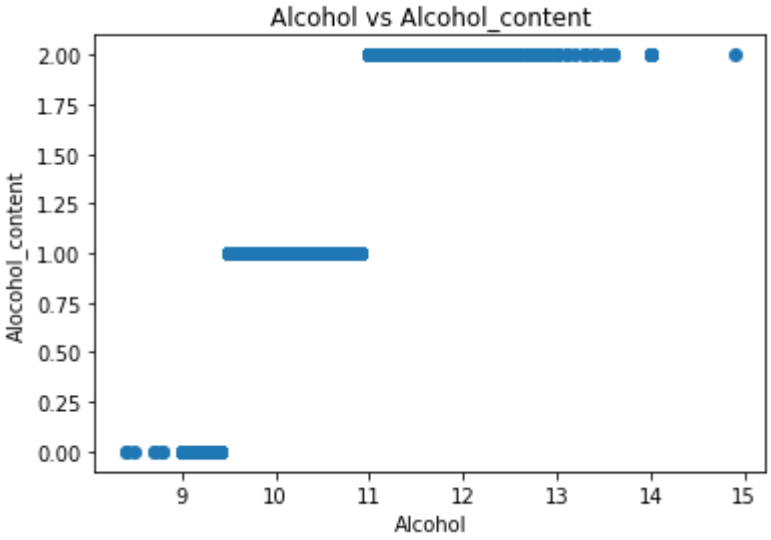
Out[8]:

<AxesSubplot:>



In [9]:

```
plt.scatter(data.alcohol,data.Alcohol_content)
plt.xlabel('Alcohol')
plt.ylabel('Alocohol_content')
plt.title('Alcohol vs Alcohol_content')
plt.show()
```



In [10]:

```
x= data.drop(columns=['quality','Alcohol_content'])
y= data['quality']
```

In [11]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25, random_state=41)
```

In [12]:

```
# Let's first visualize the tree on the data without doing any pre processing

clf=DecisionTreeClassifier()
clf.fit(x_train,y_train)
```

Out[12]:

```
DecisionTreeClassifier()
```

In [13]:

```
feature_name=list(x.columns)
class_name=list(y_train.unique())
feature_name
```

Out[13]:

```
['fixed acidity',
 'volatile acidity',
 'citric acid',
 'residual sugar',
 'chlorides',
 'free sulfur dioxide',
 'total sulfur dioxide',
 'density',
 'pH',
 'sulphates',
 'alcohol']
```

import graphviz from sklearn.tree import export_graphviz from sklearn import tree

import pydotplus

#create a dot_file which stores the tree structure
dot_data=export_graphviz(clf,feature_names=feature_name,rounded= True, filled= True)

draw graph

graph=pydotplus.graph_from_dot_data(dot_data) graph.write_png("myTree.png")

show graph

Image(graph.create_png())

In [14]:

```
clf.score(x_train,y_train)  # This is Training score
```

Out[14]:

```
1.0
```

In [15]:

```
y_pred=clf.predict(x_test)

# check the accuracy
accuracy_score(y_test,y_pred)
```

Out[15]:

```
0.61
```

In [16]:

```
# we are tuning four Important hyperparameters right now, we are passing the different valu

grid_param={
     'criterion':['gini','entropy'],
     'max_depth': range(10,15),                # The maximum depth of the tree
     'min_samples_leaf' : range(2,10),          # The minimumn number of samples required to
     'min_samples_split': range(3,10),         # The minimum number of sample required to sp
     'max_leaf_nodes': range(5,10)              # Best nodes are defined as relative reducti
 }
```

In [17]:

```
grid_search=GridSearchCV(estimator=clf,
                     param_grid=grid_param,

                         cv=5,
                         n_jobs= -1)   # use all the cores in your system. For performance i
```

In [18]:

```
grid_search.fit(x_train,y_train)
```

Out[18]:

```
GridSearchCV(cv=5, estimator=DecisionTreeClassifier(), n_jobs=-1,
             param_grid={'criterion': ['gini', 'entropy'],
                         'max_depth': range(10, 15),
                         'max_leaf_nodes': range(5, 10),
                         'min_samples_leaf': range(2, 10),
                         'min_samples_split': range(3, 10)})
```

In [19]:

```
best_parameters=grid_search.best_params_
print(best_parameters)
```

```
{'criterion': 'entropy', 'max_depth': 10, 'max_leaf_nodes': 7, 'min_samples_
leaf': 2, 'min_samples_split': 3}
```

In [20]:

```
clf=DecisionTreeClassifier(criterion= 'gini',min_samples_split=3, max_depth=10,min_samples_
clf.fit(x_train,y_train)
```

Out[20]:

```
DecisionTreeClassifier(max_depth=10, min_samples_leaf=2, min_samples_split=
3)
```

In [21]:

```
y_pred=clf.predict(x_test)
# Check the accuracy
accuracy_score(y_test,y_pred)
```

Out[21]:

```
0.6175
```

feature_name=list(x.columns) class_name=list(y_train.unique())

create a dot_file w

dot_data=export_graphviz(clf,rounded=True,filled=True)

#Draw graph graph=pydotplus.graph_from_dot_data(dot_data) graph.write_png("tree_hype_png") #show graph Image(graph.create_png())

Step 1: Run the below command in your anaconda prompt

conda install python-graphviz

give it some time it will get installed in your machine..

Step 2: Locate to your Library File in your system where you have your anaconda Folder(anaconda>Library>bin>graphviz>)

In the Library Folder you will see the bin folder open it and again you will see the graphviz folder open that graphviz folder and copy the path of that location..

If you don't see the graphviz folder in the bin file then therre must be some error while installing run the command in step 1 again in your anaconda prompt

anaconda>Library>bin>graphviz>Step 1: Run the below command in your anaconda prompt

conda install python-graphviz

give it some time it will get installed in your machine..

Step 2: Locate to your Library File in your system where you have your anaconda Folder(anaconda>Library>bin>graphviz>)

In the Library Folder you will see the bin folder open it and again you will see the graphviz folder open tha blog.ineuron.ai/How-to-Fix-Inv...nstalled%20in%20your%20machine.. Follow this link and do setup and revert back ok ma'am

In [22]:

```
conda install python-graphviz
```

```
Collecting package metadata (current_repodata.json): ...working... done
Note: you may need to restart the kernel to use updated packages.
Solving environment: ...working... done

# All requested packages already installed.
```

In [23]:

```
pip install graphviz
```

```
Requirement already satisfied: graphviz in c:\users\user\anaconda3\lib\site-
packages (0.20)
Note: you may need to restart the kernel to use updated packages.
```

In [24]:

```
conda install graphviz
```

```
Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done

# All requested packages already installed.


Note: you may need to restart the kernel to use updated packages.
```

In [29]:

```
import graphviz
from sklearn.tree import export_graphviz
from sklearn import tree

import pydotplus

feature_name=list(x.columns)
class_name=list(y_train.unique())

# create a dot_file which stores the tree structure
dot_data=export_graphviz(clf,rounded=True,filled=True)

#Draw graph
graph=pydotplus.graph_from_dot_data(dot_data)
graph.write_png("Tree_hype_png")
#show graph
Image(graph.create_png())
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_3212/3660970494.py in <module>
----> 1 import graphviz
      2 from sklearn.tree import export_graphviz
      3 from sklearn import tree
      4
      5 import pydotplus

~\anaconda3\lib\site-packages\graphviz\__init__.py in <module>
     25 """
     26
---> 27 from .dot import Graph, Digraph
     28 from .files import Source
     29 from .lang import escape, nohtml

~\anaconda3\lib\site-packages\graphviz\dot.py in <module>
     30
     31 from . import backend
---> 32 from . import files
     33 from . import lang
     34

~\anaconda3\lib\site-packages\graphviz\files.py in <module>
     20
     21
---> 22 class Base(object):
     23
     24     _engine = 'dot'

~\anaconda3\lib\site-packages\graphviz\files.py in Base()
     26     _format = 'pdf'
     27
---> 28     _encoding = backend.ENCODING
     29
     30     @property

AttributeError: module 'graphviz.backend' has no attribute 'ENCODING'
```

In [26]:

```
conda install -c conda-forge python-graphviz
```

```
Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done

# All requested packages already installed.


Note: you may need to restart the kernel to use updated packages.
```

In [27]:

```
pip install graphviz
```

```
Requirement already satisfied: graphviz in c:\users\user\anaconda3\lib\site-
packages (0.20)
Note: you may need to restart the kernel to use updated packages.
```

In [ ]:

```

```