decisionTree

March 22, 2023

1 DECISION TREE CLASSIFICATION USING ENTROPY AND GINI INDEX:

1.0.1 Dataset: Red Wine classification dataset

1.0.2 Contents

Input variables (based on physicochemical tests):

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):

1. quality (score between 0 and 10)

1.1 1. LOAD AND EXPLORATION

[21]:
```
#importing libraries

import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
%matplotlib inline
```

```
import seaborn as sns
from sklearn import tree

from sklearn import metrics
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,precision_score, recall_score,␣
↪classification_report,confusion_matrix

from sklearn.tree import DecisionTreeClassifier
```

[22]:
```
#loading the data

df = pd.read_csv('./winequality-red.csv')
print('The Dataset contains {} rows and {} columns '.format(df.shape[0], df.
↪shape[1]))
```

```
The Dataset contains 1599 rows and 12 columns
```

[23]:
```
df.head()
```

[23]:
```
   fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  \
0            7.4              0.70         0.00             1.9      0.076
1            7.8              0.88         0.00             2.6      0.098
2            7.8              0.76         0.04             2.3      0.092
3           11.2              0.28         0.56             1.9      0.075
4            7.4              0.70         0.00             1.9      0.076

   free sulfur dioxide  total sulfur dioxide  density    pH  sulphates  \
0                 11.0                  34.0   0.9978  3.51       0.56
1                 25.0                  67.0   0.9968  3.20       0.68
2                 15.0                  54.0   0.9970  3.26       0.65
3                 17.0                  60.0   0.9980  3.16       0.58
4                 11.0                  34.0   0.9978  3.51       0.56

   alcohol  quality
0      9.4        5
1      9.8        5
2      9.8        5
3      9.8        6
4      9.4        5
```

[24]:
```
#getting the statiscal information

df.describe()
```

[24]:

```
       fixed acidity  volatile acidity  citric acid  residual sugar  \
count    1599.000000       1599.000000  1599.000000     1599.000000
mean        8.319637          0.527821     0.270976        2.538806
std         1.741096          0.179060     0.194801        1.409928
min         4.600000          0.120000     0.000000        0.900000
25%         7.100000          0.390000     0.090000        1.900000
50%         7.900000          0.520000     0.260000        2.200000
75%         9.200000          0.640000     0.420000        2.600000
max        15.900000          1.580000     1.000000       15.500000

         chlorides  free sulfur dioxide  total sulfur dioxide      density  \
count  1599.000000          1599.000000           1599.000000  1599.000000
mean      0.087467            15.874922             46.467792     0.996747
std       0.047065            10.460157             32.895324     0.001887
min       0.012000             1.000000              6.000000     0.990070
25%       0.070000             7.000000             22.000000     0.995600
50%       0.079000            14.000000             38.000000     0.996750
75%       0.090000            21.000000             62.000000     0.997835
max       0.611000            72.000000            289.000000     1.003690

                pH    sulphates      alcohol      quality
count  1599.000000  1599.000000  1599.000000  1599.000000
mean      3.311113     0.658149    10.422983     5.636023
std       0.154386     0.169507     1.065668     0.807569
min       2.740000     0.330000     8.400000     3.000000
25%       3.210000     0.550000     9.500000     5.000000
50%       3.310000     0.620000    10.200000     6.000000
75%       3.400000     0.730000    11.100000     6.000000
max       4.010000     2.000000    14.900000     8.000000
```

1.2 2. DATA CLEANING

[25]:

```
#counting the frequency of each element from the 'quality'

df['quality'].value_counts().index
```

[25]:

```
Int64Index([5, 6, 7, 4, 8, 3], dtype='int64')
```

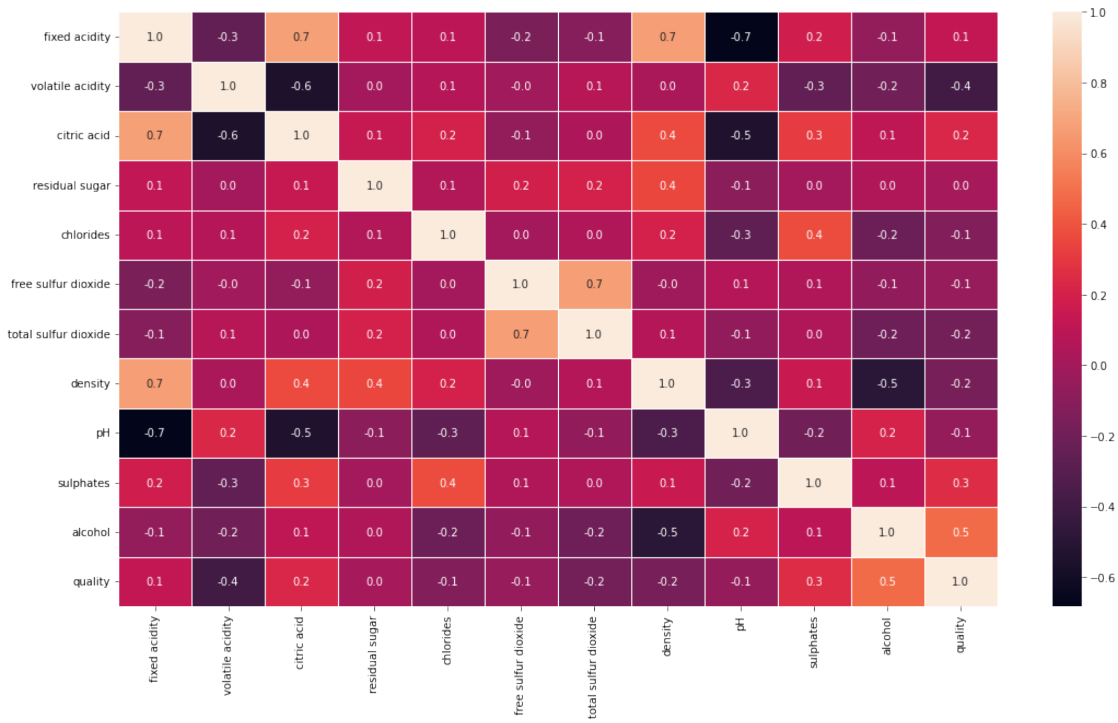
So the ratings are 3,4,5,6,7 and 8 making only 6 values in quality column

[26]:

```
#correlation between the columns

plt.figure(figsize=(18,10))
sns.heatmap(df.corr(), annot=True, fmt = ".1f", linewidths = .7)
```

[26]:

```
<AxesSubplot:>
```



Checking for missing values:

[27]:
```
df.isnull().sum()
```

[27]:
```
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
```

1.2.1 *There are no missing values in the dataset*

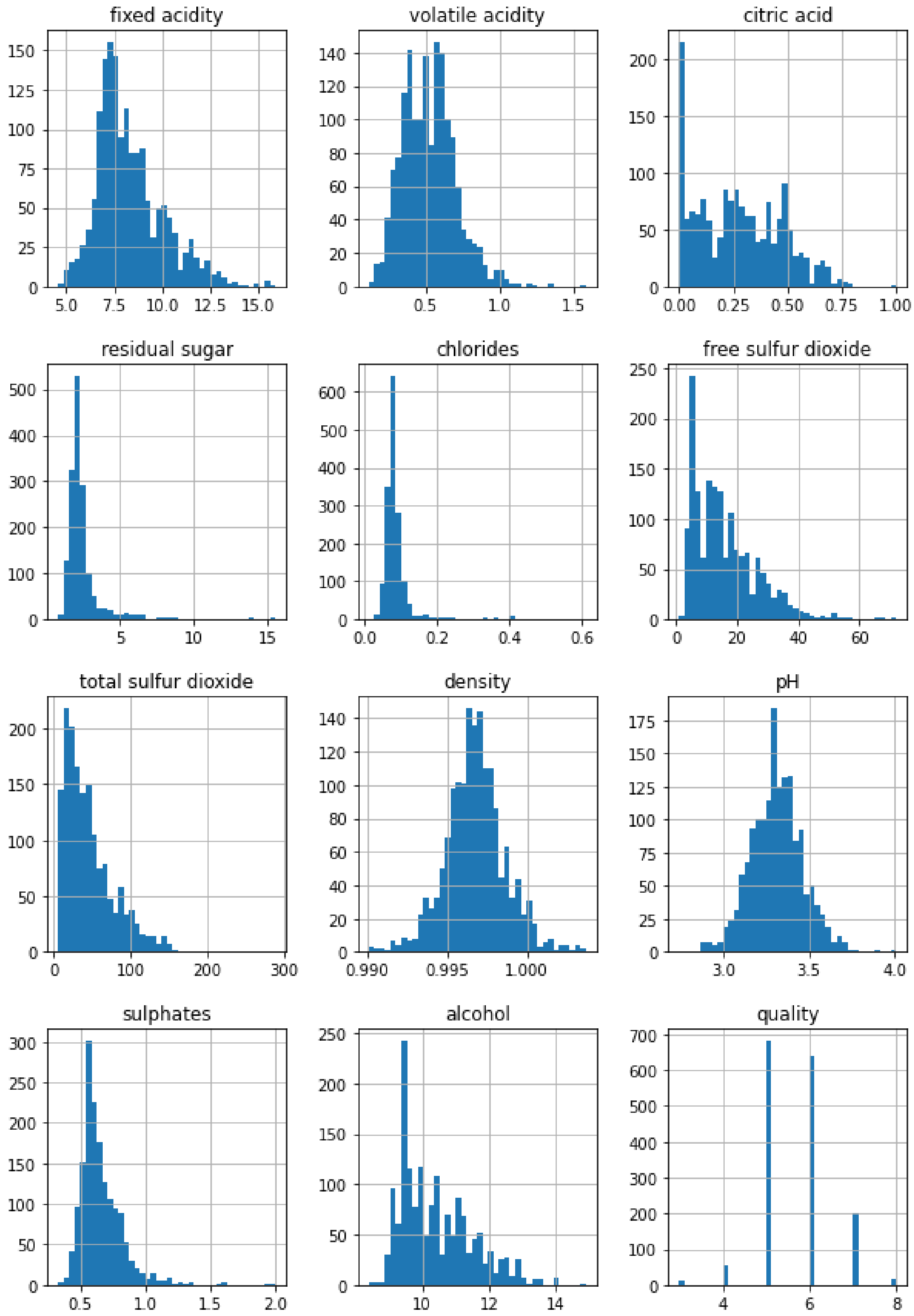
[28]:
```
df.info()
```

```
<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
```

1.2.2 Now Showing the distribution of each feature:

[29]:
```
df.hist(bins=40, figsize=(10,15))
plt.show()
```


fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality

1.2.3 What do we Understand?

Data distribution for attribute "alcohol" is positively skewed, for attribute "density" data quite normally distributed. Take attention to the wine quality data distribution. It's a bimodal distribution and there are more wines with average quality than wines with 'good' or 'bad' quality.

[30]:
```python
#counting the frequency of each element from the 'class'

df['quality'].value_counts()
```

[30]:
```
5    681
6    638
7    199
4     53
8     18
3     10
Name: quality, dtype: int64
```

Human wine preferences scores varied from 3 to 8, so it's straightforward to categorize answers into 'bad' or 'good' quality of wines. We assign for categorizes corresponding discrete values 0 or 1.

1.2.4 Good - 1, Bad - 0

[31]:
```python
# Dividing wine as good and bad by giving the limit for the quality

bins = (2, 6, 8)
group_names = ['bad', 'good']
df['quality'] = pd.cut(df['quality'], bins = bins, labels = group_names)
```

[32]:
```python
print(df['quality'].value_counts())
```

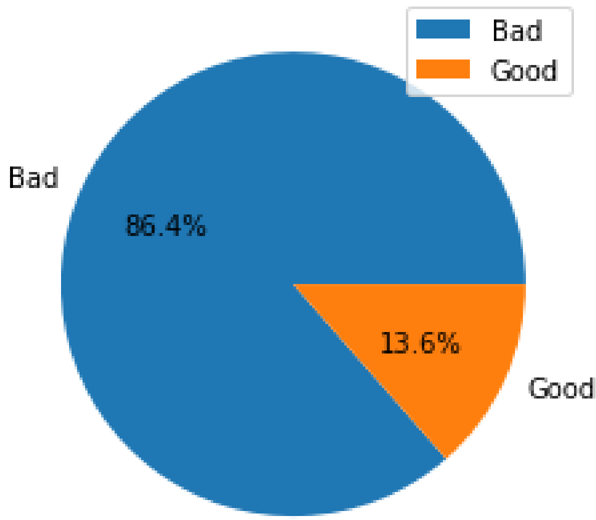
```
bad     1382
good     217
Name: quality, dtype: int64
```

[33]:
```python
# assign labels to our quality variable

label_quality = LabelEncoder()

# Bad becomes 0 and good becomes 1

df['quality'] = label_quality.fit_transform(df['quality'])
```

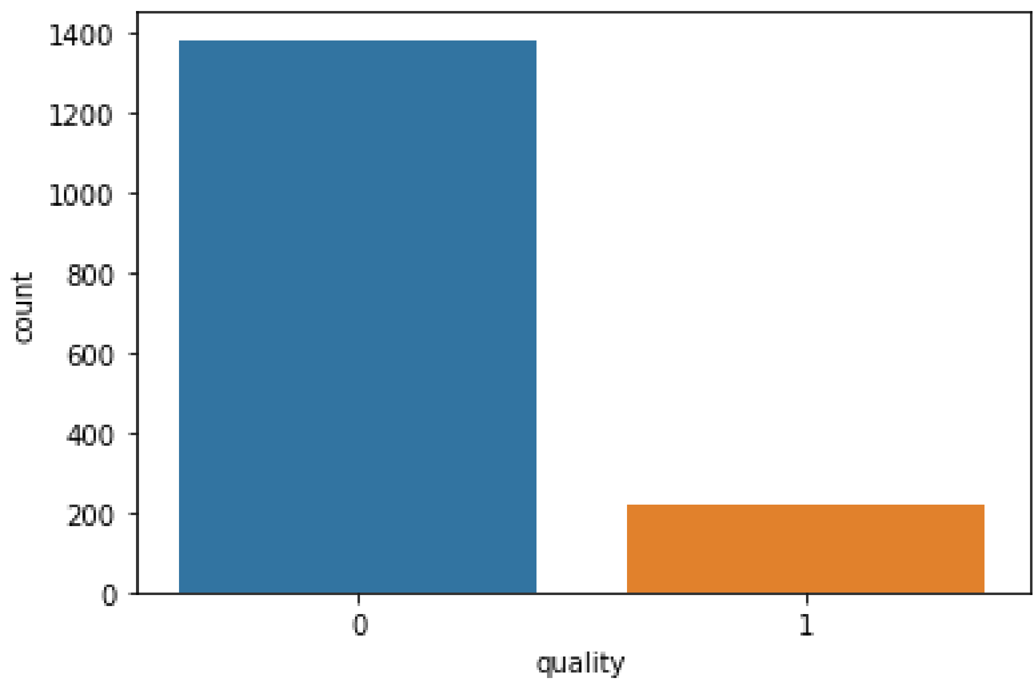
[34]:
```python
df['quality'].value_counts()
```

[34]:
```
0    1382
1     217
Name: quality, dtype: int64
```

[35]:
```
#proportion of different elements of the class


plt.pie(df['quality'].value_counts(),autopct="%1.1f%%",labels=['Bad','Good'])
plt.legend();
```



[36]:
```
#plot to show count of labels

sns.countplot(x=df['quality'])
plt.show()
```



[37]:
```
#Filter the dataset for input and output

x = df.drop(['quality'], axis=1)
y = df['quality']
```

[38]:
```
#splitting the processed dataset into train and test dataset

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.25,␣
↪random_state = 50)
```

2 A) ALGORITHM FOR DECISION TREE IMPLEMENTED FROM SCRATCH

2.1 Entropy and GINI INDEX

[41]:
```
#class node for representing each node of the decision tree
class Node():
    #constructor
    def __init__(self, feature_index=None, threshold=None, left=None,␣
↪right=None, info_gain=None, value=None):

        # for decision node
        self.feature_index = feature_index
        self.threshold = threshold
```

```
        self.left = left
        self.right = right
        self.info_gain = info_gain

        # for leaf node
        self.value = value

class MyDecisionTreeClassifier():
    #constructor
    def __init__(self, criterion="gini", min_samples_split=2, max_depth=2):

        # initialize the root of the tree
        self.root = None

        # stopping conditions
        self.criterion = criterion
        self.min_samples_split = min_samples_split
        self.max_depth = max_depth

    #recursive function to build the tree
    def build_tree(self, dataset, curr_depth=0):

        X, Y = dataset[:,:-1], dataset[:,-1]
        num_samples, num_features = np.shape(X)

        # split until stopping conditions are met
        if num_samples>=self.min_samples_split and curr_depth<=self.max_depth:
            # find the best split
            best_split = self.get_best_split(dataset, num_samples, num_features)
            # check if information gain is positive
            if "info_gain" in best_split and best_split["info_gain"]>0:
                # recur left
                left_subtree = self.build_tree(best_split["dataset_left"],␣
 ↪curr_depth+1)
                # recur right
                right_subtree = self.build_tree(best_split["dataset_right"],␣
 ↪curr_depth+1)
                # return decision node
                return Node(best_split["feature_index"],␣
 ↪best_split["threshold"],
                            left_subtree, right_subtree,␣
 ↪best_split["info_gain"])

        # compute leaf node
        leaf_value = self.calculate_leaf_value(Y)
        # return leaf node
        return Node(value=leaf_value)
```

```
    #function to find the best split
    def get_best_split(self, dataset, num_samples, num_features):

        # dictionary to store the best split
        best_split = {}
        max_info_gain = -float("inf")

        # loop over all the features
        for feature_index in range(num_features):
            feature_values = dataset[:, feature_index]
            possible_thresholds = np.unique(feature_values)
            # loop over all the feature values present in the data
            for threshold in possible_thresholds:
                # get current split
                dataset_left, dataset_right = self.split(dataset,␣
↪feature_index, threshold)
                # check if childs are not null
                if len(dataset_left)>0 and len(dataset_right)>0:
                    y, left_y, right_y = dataset[:, -1], dataset_left[:, -1],␣
↪dataset_right[:, -1]
                    # compute information gain
                    curr_info_gain = self.information_gain(y, left_y, right_y,␣
↪self.criterion)
                    # update the best split if needed
                    if curr_info_gain>max_info_gain:
                        best_split["feature_index"] = feature_index
                        best_split["threshold"] = threshold
                        best_split["dataset_left"] = dataset_left
                        best_split["dataset_right"] = dataset_right
                        best_split["info_gain"] = curr_info_gain
                        max_info_gain = curr_info_gain

        # return best split
        return best_split

    #function to split the data
    def split(self, dataset, feature_index, threshold):

        dataset_left = np.array([row for row in dataset if␣
↪row[feature_index]<=threshold])
        dataset_right = np.array([row for row in dataset if␣
↪row[feature_index]>threshold])
        return dataset_left, dataset_right

    #function to compute information gain
    def information_gain(self, parent, l_child, r_child, mode="gini"):
```

```
        weight_l = len(l_child) / len(parent)
        weight_r = len(r_child) / len(parent)
        if mode=="gini":
            gain = self.gini_index(parent) - (weight_l*self.gini_index(l_child) + weight_r*self.gini_index(r_child))
        else:
            gain = self.entropy(parent) - (weight_l*self.entropy(l_child) + weight_r*self.entropy(r_child))
        return gain

    #function to compute entropy
    def entropy(self, y):

        class_labels = np.unique(y)
        entropy = 0
        for cls in class_labels:
            p_cls = len(y[y == cls]) / len(y)
            entropy += -p_cls * np.log2(p_cls)
        return entropy

    #function to compute gini index
    def gini_index(self, y):

        class_labels = np.unique(y)
        gini = 0
        for cls in class_labels:
            p_cls = len(y[y == cls]) / len(y)
            gini += p_cls**2
        return 1 - gini

    #function to compute leaf node
    def calculate_leaf_value(self, Y):

        Y = list(Y)
        return max(Y, key=Y.count)

    #function to print the tree
    def print_tree(self, tree=None, indent=" "):

        if not tree:
            tree = self.root

        if tree.value is not None:
            print(tree.value)

        else:
```

```
            print("X_"+str(tree.feature_index), "<=", tree.threshold, "?", tree.
↪info_gain)
            print("%sleft:" % (indent), end="")
            self.print_tree(tree.left, indent+indent)
            print("%sright:" % (indent), end="")
            self.print_tree(tree.right, indent+indent)

    #function to train the tree
    def fit(self, X, Y):

        dataset = np.concatenate((X, Y), axis=1)
        self.root = self.build_tree(dataset)

    #function to predict new dataset
    def predict(self, X):

        preditions = [self.make_prediction(x, self.root) for x in X]
        return preditions

    #function to predict a single data point
    def make_prediction(self, x, tree):

        if tree.value!=None: return tree.value
        feature_val = x[tree.feature_index]
        if feature_val<=tree.threshold:
            return self.make_prediction(x, tree.left)
        else:
            return self.make_prediction(x, tree.right)
```

[42]:

```
x = df.iloc[:, :-1].values
y = df.iloc[:, -1].values.reshape(-1,1)
```

[44]:

```
#split train and test samples

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.25,␣
↪random_state = 50)
```

2.2 4. TRAIN THE MODEL

2.2.1 A) DECISION TREE IMPLEMENTED FROM SCRATCH

2.2.2 USING ENTROPY

[45]:
```
myClassifier_entropy = MyDecisionTreeClassifier(criterion="entropy",␣
 ↪max_depth=50)
myClassifier_entropy.fit(x_train, y_train)
myClassifier_entropy.print_tree()
```

```
X_10 <= 10.4 ? 0.09345652303969632
 left:X_9 <= 0.61 ? 0.02742215521756683
  left:X_9 <= 0.56 ? 0.008250172890607414
    left:0.0
    right:X_5 <= 10.0 ? 0.024400785009314543
        left:X_5 <= 8.0 ? 0.09575537889326563
                left:0.0
                right:X_2 <= 0.02 ? 0.25767880510333147
                                left:1.0
                                right:X_4 <= 0.078 ? 0.1935068433729344
                                                                left:0.0
                                                                right:X_8 <=
3.24 ? 0.9182958340544896
                                                left:1.0
                                                right:0.0
        right:0.0
  right:X_10 <= 9.6 ? 0.0412359652963899
    left:X_0 <= 14.3 ? 0.04638979071433067
        left:X_9 <= 0.62 ? 0.021314443909423664
                left:X_6 <= 35.0 ? 0.15649412347457692
                                left:X_5 <= 7.0 ? 0.9182958340544896
                                                                left:0.0
                                                                right:1.0
                                right:0.0
                right:0.0
        right:1.0
    right:X_2 <= 0.08 ? 0.039542966228935605
        left:0.0
        right:X_6 <= 67.0 ? 0.045874225349620046
                left:X_4 <= 0.074 ? 0.04112938525068455
                                left:X_7 <= 0.99769 ? 0.10911003076268089
                                                                left:X_4 <=
0.069 ? 0.17404892546009187
                                                left:X_1 <= 0.3 ?
0.16251125329718275
                left:0.0
                right:X_8 <= 3.23 ? 0.2621549647380449
                                left:1.0
```

```
                                right:X_4 <= 0.059 ? 0.3435794213678428
                                                                left:0.0
                                                                right:X_1 <=
0.48 ? 0.46956521111470695
                                                left:X_5 <= 31.0 ?
0.7219280948873623
                left:1.0
                right:0.0
                                                right:0.0
                                                right:1.0
                                                                right:X_0 <=
12.5 ? 0.4394969869215134
                                                left:0.0
                                                right:1.0
                                right:X_4 <= 0.081 ? 0.08363242229665335
                                                                left:0.0
                                                                right:X_6 <=
11.0 ? 0.1075414071369365
                                                left:1.0
                                                right:X_1 <= 0.89 ?
0.061704192549288606
                left:X_4 <= 0.096 ? 0.08986867817719413
                                left:X_2 <= 0.52 ? 0.08556114743657839
                                                                left:X_8 <= 3.51
? 0.13824095117944724
                                                left:X_4 <= 0.091 ?
0.1648677151048581
                left:0.0
                right:X_0 <= 9.5 ? 0.46956521111470695
                                left:0.0
                                right:X_0 <= 10.1 ? 0.9182958340544896
                                                                left:1.0
                                                                right:0.0
                                                right:1.0
                                                                right:X_0 <=
12.5 ? 0.4591479170272448
                                                left:1.0
                                                right:X_1 <= 0.415 ?
0.8112781244591328
                left:0.0
                right:1.0
                                right:0.0
                right:1.0
                right:0.0
 right:X_9 <= 0.68 ? 0.06912096376765586
  left:X_1 <= 0.38 ? 0.06103333983734327
    left:X_0 <= 6.4 ? 0.06198490656686273
        left:0.0
```

```
        right:X_0 <= 6.8 ? 0.08252524785132997
                left:1.0
                right:X_3 <= 2.9 ? 0.0880316723464829
                                left:X_9 <= 0.54 ? 0.10862163149277804
                                                                left:0.0
                                                                right:X_8 <=
3.26 ? 0.13822251414224762
                                                left:X_6 <= 16.0 ?
0.21906026991893968
                left:X_2 <= 0.34 ? 0.5032583347756457
                                left:0.0
                                right:1.0
                right:X_5 <= 20.0 ? 0.31668908831502096
                                left:X_3 <= 1.7 ? 0.3828503397420074
                                                                left:1.0
                                                                right:X_2 <=
0.34 ? 0.2373974097831018
                                                left:X_0 <= 10.1 ? 1.0
                left:1.0
                right:0.0
                                                right:0.0
                                right:1.0
                                                right:X_4 <= 0.0579999999999999
? 0.3502090290998975
                left:X_0 <= 7.1 ? 0.9182958340544896
                                left:0.0
                                right:1.0
                right:0.0
                                right:X_10 <= 11.0 ? 0.7642045065086203
                                                                left:0.0
                                                                right:1.0
    right:X_6 <= 19.0 ? 0.033748229932452856
        left:X_1 <= 0.66 ? 0.18220283362606426
                left:X_4 <= 0.086 ? 0.2655234166823892
                                left:X_9 <= 0.58 ? 0.25620623685627303
                                                                left:0.0
                                                                right:X_5 <= 5.0
? 0.5216406363433185
                                                left:0.0
                                                right:X_4 <= 0.043 ?
0.8112781244591328
                left:0.0
                right:1.0
                                right:X_4 <= 0.1009999999999999 ?
0.6052891061068587
                                                                left:1.0
                                                                right:X_2 <=
0.49 ? 0.7219280948873623
```

```
                                                left:0.0
                                                right:1.0
                right:0.0
        right:X_8 <= 2.92 ? 0.028044537047481777
                left:1.0
                right:X_4 <= 0.065 ? 0.03593861065895981
                                left:X_4 <= 0.053 ? 0.13061142961974004
                                                                left:0.0
                                                                right:X_5 <=
31.0 ? 0.22625794497561413
                                                left:X_3 <= 1.7 ?
0.2025070034547547
                left:X_1 <= 0.47 ? 0.9182958340544896
                                left:0.0
                                right:1.0
                right:X_1 <= 0.4 ? 0.22002600168808803
                                left:X_0 <= 5.9 ? 1.0
                                                                left:0.0
                                                                right:1.0
                                right:0.0
                                                right:1.0
                                right:X_1 <= 0.42 ? 0.04560409401861014
                                                                left:X_4 <= 0.09
? 0.22994744641573744
                                                left:0.0
                                                right:X_5 <= 9.0 ?
0.4199730940219749
                left:0.0
                right:X_2 <= 0.66 ? 0.9182958340544896
                                left:1.0
                                right:0.0
                                                                right:0.0
  right:X_10 <= 11.6 ? 0.10044876766219746
    left:X_1 <= 0.4 ? 0.07567434015082874
        left:X_9 <= 0.75 ? 0.08277568191155193
                left:X_10 <= 11.0 ? 0.2771196685025808
                                left:0.0
                                right:X_1 <= 0.34 ? 0.9910760598382222
                                                                left:0.0
                                                                right:1.0
                right:X_7 <= 0.9974 ? 0.09849899432197895
                                left:X_7 <= 0.99572 ? 0.20972714405924964
                                                                left:X_6 <= 44.0
? 0.5297257989969673
                                                left:1.0
                                                right:X_0 <= 9.0 ?
0.8112781244591328
                left:0.0
```

```
                right:1.0
                                                                right:X_2 <=
0.38 ? 0.27621156854915635
                                                left:X_2 <= 0.34 ?
0.4040097573248599
                left:X_9 <= 0.78 ? 0.8112781244591328
                                left:1.0
                                right:0.0
                right:1.0
                                                right:0.0
                                right:X_4 <= 0.075 ? 0.2651749506101608
                                                                left:X_2 <= 0.47
? 0.5216406363433185
                                                left:1.0
                                                right:X_2 <= 0.49 ?
0.8112781244591328
                left:0.0
                right:1.0
                                                                right:1.0
        right:X_10 <= 11.4 ? 0.10523421790669629
                left:X_2 <= 0.16 ? 0.13666642690501452
                                left:X_0 <= 6.3 ? 0.23193334876682492
                                                                left:0.0
                                                                right:X_2 <=
0.06 ? 0.3435794213678428
                                                left:0.0
                                                right:X_10 <= 10.8 ?
0.5487949406953987
                left:1.0
                right:X_0 <= 6.4 ? 0.8112781244591328
                                left:1.0
                                right:0.0
                                right:0.0
                right:X_3 <= 3.1 ? 0.31127812445913283
                                left:X_4 <= 0.062 ? 0.45810589515712374
                                                                left:1.0
                                                                right:X_2 <=
0.56 ? 0.3059584928680418
                                                left:0.0
                                                right:X_0 <= 9.9 ? 1.0
                left:1.0
                right:0.0
                                right:1.0
    right:X_5 <= 18.0 ? 0.13723548885905645
        left:X_7 <= 0.99468 ? 0.13755617370705508
                left:1.0
                right:X_7 <= 0.9948 ? 0.15375242402031997
                                left:0.0
```

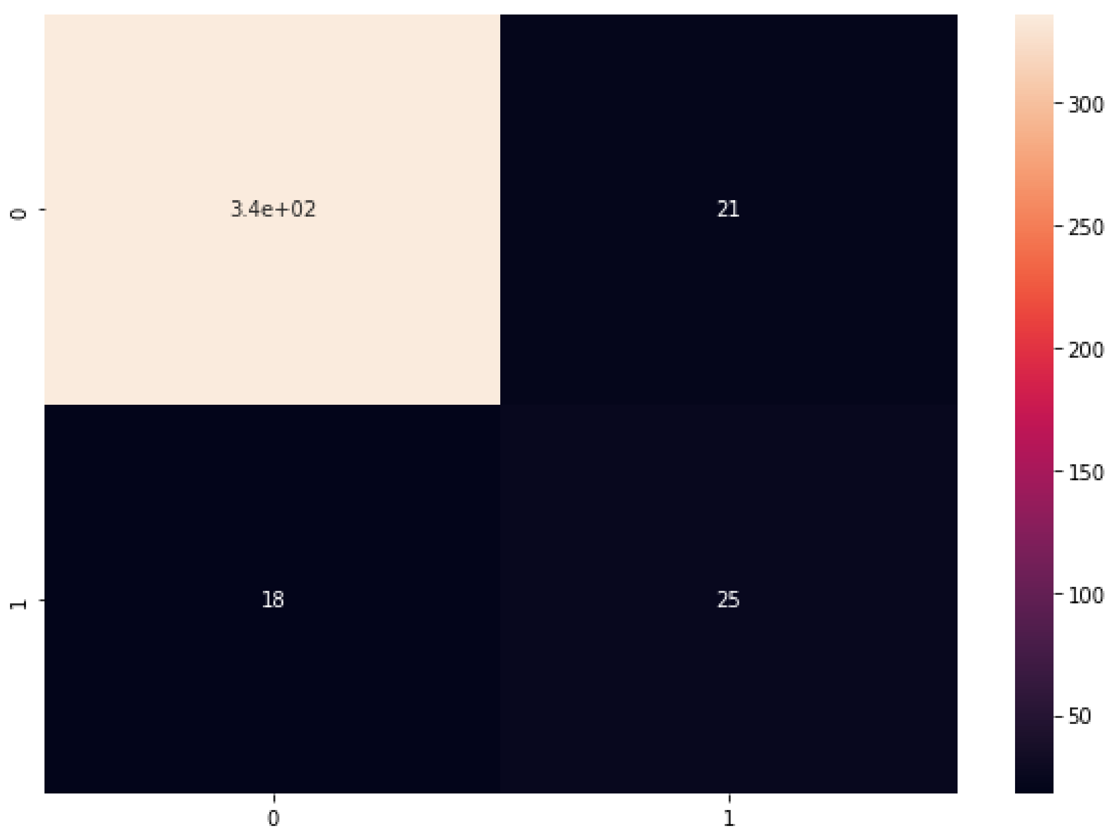
```
                                right:X_7 <= 0.9962 ? 0.14157309748501146
                                                                left:X_4 <= 0.11
? 0.3095434291503252
                                                left:1.0
                                                right:0.0
                                                                right:X_3 <= 2.2
? 0.3814444125401065
                                                left:0.0
                                                right:X_5 <= 7.0 ?
0.3178113757536235
                left:X_2 <= 0.5 ? 0.4591479170272448
                                left:0.0
                                right:X_4 <= 0.088 ? 0.8112781244591328
                                                                left:0.0
                                                                right:1.0
                right:1.0
        right:X_1 <= 0.57 ? 0.10958835569188075
                left:X_5 <= 27.0 ? 0.16402047076084547
                                left:X_6 <= 50.0 ? 0.40590730096336636
                                                                left:X_0 <= 7.3
? 0.9709505944546686
                                                left:1.0
                                                right:0.0
                                                                right:0.0
                                right:X_5 <= 45.0 ? 0.19350684337293445
                                                                left:X_1 <= 0.42
? 0.31976006206417584
                                                left:X_1 <= 0.33 ? 1.0
                left:1.0
                right:0.0
                                                right:1.0
                                                                right:0.0
                right:0.0
```

2.2.3 Predicting and Performance metrics of the model

[46]:
```python
y_pred = myClassifier_entropy.predict(x_test)
print("Accuracy = ", round(accuracy_score(y_test, y_pred)*100, 2), "%")
print("Precision = ",precision_score(y_test, y_pred))
print("Recall = ",recall_score(y_test, y_pred),"\n")

print(classification_report(y_pred,y_test))

cm=metrics.confusion_matrix(y_test,y_pred)
print("\nconfusion matrix: \n",cm)
plt.figure(figsize = (10,7))
sns.heatmap(cm, annot=True)
```

```python
print("---------------------------------------")
```

```
Accuracy =  90.25 %
Precision =  0.5434782608695652
Recall =  0.5813953488372093

              precision    recall  f1-score   support

         0.0       0.94      0.95      0.95       354
         1.0       0.58      0.54      0.56        46

    accuracy                           0.90       400
   macro avg       0.76      0.75      0.75       400
weighted avg       0.90      0.90      0.90       400


confusion matrix:
 [[336  21]
 [ 18  25]]
---------------------------------------
```



2.2.4 TRAINING MODEL USING DECISION TREE IMPLEMENTED FROM SCRATCH

2.2.5 USING GINI INDEX

[47]:
```
myClassifier_gini = MyDecisionTreeClassifier(criterion="gini", max_depth=50)
myClassifier_gini.fit(x_train, y_train)
myClassifier_gini.print_tree()
```

```
X_10 <= 11.5 ? 0.03625483343863817
 left:X_1 <= 0.4 ? 0.013073495960650688
  left:X_10 <= 10.4 ? 0.03624076396118492
    left:X_9 <= 1.06 ? 0.015289095986072615
        left:X_4 <= 0.075 ? 0.015524624715379298
                left:X_0 <= 11.6 ? 0.08908202314537333
                                left:X_3 <= 1.4 ? 0.04684972395295184
                                                                left:X_0 <= 6.4
? 0.1111111111111111
                                                left:1.0
                                                right:X_0 <= 8.7 ? 0.5
                left:0.0
                right:1.0
                                                                right:X_1 <=
0.39 ? 0.04585610541289219
                                                left:X_9 <= 0.82 ?
0.009307833632158063
                left:X_2 <= 0.28 ? 0.004224058769513185
                                left:X_0 <= 7.9 ? 0.17999999999999994
                                                                left:0.0
                                                                right:1.0
                                right:0.0
                right:X_4 <= 0.067 ? 0.375
                                left:0.0
                                right:1.0
                                                right:1.0
                                right:X_2 <= 0.72 ? 0.31999999999999984
                                                                left:1.0
                                                                right:0.0
                right:X_3 <= 2.4 ? 0.0007687057248696658
                                left:0.0
                                right:X_2 <= 0.31 ? 0.03969754253308119
                                                                left:X_0 <= 6.9
? 0.5
                                                left:0.0
                                                right:1.0
                                                                right:0.0
        right:X_10 <= 9.5 ? 0.5
                left:0.0
```

```
                right:1.0
    right:X_9 <= 0.75 ? 0.050909762825298244
        left:X_8 <= 3.28 ? 0.04692462962257682
                left:X_2 <= 0.39 ? 0.09095533865720212
                                left:X_7 <= 0.99552 ? 0.3111111111111111
                                                                left:1.0
                                                                right:X_3 <= 2.2
? 0.31999999999999984
                                                left:0.0
                                                right:1.0
                                right:X_0 <= 10.4 ? 0.08664146187775673
                                                                left:0.0
                                                                right:X_7 <=
0.9972 ? 0.2268518518518518
                                                left:1.0
                                                right:X_2 <= 0.65 ?
0.19753086419753085
                left:0.0
                right:X_0 <= 11.9 ? 0.4444444444444444
                                left:0.0
                                right:1.0
                right:X_6 <= 10.0 ? 0.05619146722164431
                                left:X_1 <= 0.31 ? 0.5
                                                                left:1.0
                                                                right:0.0
                                right:0.0
        right:X_7 <= 0.9974 ? 0.06347146701817968
                left:X_7 <= 0.99572 ? 0.1371485657199944
                                left:X_6 <= 44.0 ? 0.2396449704142011
                                                                left:1.0
                                                                right:X_0 <= 9.0
? 0.375
                                                left:0.0
                                                right:1.0
                                right:X_8 <= 3.38 ? 0.12144168962350785
                                                                left:X_3 <= 1.6
? 0.12345679012345678
                                                left:X_0 <= 8.0 ?
0.4444444444444444
                left:0.0
                right:1.0
                                                right:0.0
                                                                right:X_0 <= 6.1
? 0.375
                                                left:0.0
                                                right:1.0
                right:X_6 <= 21.0 ? 0.10596955128205143
                                left:X_0 <= 10.6 ? 0.4444444444444444
```

```
                                                                left:0.0
                                                                right:1.0
                                openright:X_7 <= 1.0002 ? 0.14201183431952646
                                                                left:1.0
                                                                right:0.0
  right:X_10 <= 11.4 ? 0.003314260580281378
    left:X_10 <= 9.8 ? 0.0011420669860130253
        left:X_5 <= 12.0 ? 0.00017046813296177701
                left:X_5 <= 8.0 ? 0.0008583773774161821
                                left:0.0
                                right:X_6 <= 18.0 ? 0.0073973429951689346
                                                                left:X_3 <= 1.9
? 0.4444444444444444
                                                left:0.0
                                                right:1.0
                                                                right:X_1 <=
0.52 ? 0.004116782188615824
                                                left:X_1 <= 0.49 ?
0.029999999999999916
                left:0.0
                right:X_4 <= 0.071 ? 0.20833333333333334
                                left:0.0
                                right:X_2 <= 0.09 ? 0.4444444444444444
                                                                left:0.0
                                                                right:1.0
                                                right:0.0
                right:0.0
        right:X_3 <= 5.6 ? 0.002981655912649414
                left:X_9 <= 0.63 ? 0.0025556105893931313
                                left:X_2 <= 0.01 ? 0.0017302179614422375
                                                                left:X_1 <= 0.58
? 0.05753968253968264
                                                left:X_2 <= 0.0 ?
0.48979591836734704
                left:0.0
                right:1.0
                                                right:0.0
                                                                right:X_8 <=
3.51 ? 0.0006228373702421124
                                                left:0.0
                                                right:X_7 <= 0.99648 ?
0.05190311418685113
                left:0.0
                right:X_0 <= 7.0 ? 0.5
                                left:1.0
                                right:0.0
                                right:X_5 <= 3.0 ? 0.011772905933159467
                                                                left:1.0
```

```
                                                            right:X_3 <= 1.5
? 0.010096038415366354
                                                left:X_2 <= 0.43 ? 0.5
                left:0.0
                right:1.0
                                                right:X_9 <= 0.71 ?
0.005737763813556396
                left:X_10 <= 10.0 ? 0.025417106623011776
                                left:X_1 <= 0.48 ? 0.5
                                                                left:1.0
                                                                right:0.0
                                right:X_4 <= 0.118 ? 0.03122447464171685
                                                                left:X_8 <= 3.57
? 0.015176005747126589
                                                left:X_3 <= 1.7 ?
0.015479021572957177
                left:X_0 <= 7.0 ? 0.5
                                left:1.0
                                right:0.0
                right:X_3 <= 5.1 ? 0.011267006802721108
                                left:0.0
                                right:X_0 <= 7.1 ? 0.4444444444444444
                                                                left:0.0
                                                                right:1.0
                                                right:X_9 <= 0.68 ?
0.4444444444444444
                left:0.0
                right:1.0
                                                                right:X_1 <=
0.53 ? 0.4444444444444444
                                                left:0.0
                                                right:1.0
                right:X_2 <= 0.09 ? 0.0017636684303352149
                                left:X_0 <= 8.1 ? 0.13265306122448983
                                                                left:0.0
                                                                right:1.0
                                right:0.0
                right:X_3 <= 6.0 ? 0.2603305785123967
                                left:X_1 <= 0.45 ? 0.375
                                                                left:0.0
                                                                right:1.0
                                right:0.0
    right:X_2 <= 0.29 ? 0.15052083333333338
        left:X_4 <= 0.088 ? 0.07999999999999993
                left:0.0
                right:X_0 <= 6.5 ? 0.5
                                left:0.0
                                right:1.0
```

```
        right:X_0 <= 9.8 ? 0.2222222222222222
                left:1.0
                right:X_1 <= 0.5 ? 0.4444444444444444
                                left:0.0
                                right:1.0
 right:X_9 <= 0.68 ? 0.08000739820667302
  left:X_6 <= 15.0 ? 0.04634753181112827
    left:X_9 <= 0.58 ? 0.21566162731442162
        left:X_3 <= 1.8 ? 0.11999999999999983
                left:X_4 <= 0.063 ? 0.5
                                left:0.0
                                right:1.0
                right:0.0
        right:X_6 <= 8.0 ? 0.24489795918367352
                left:0.0
                right:1.0
    right:X_5 <= 31.0 ? 0.07255173079087426
        left:X_8 <= 3.27 ? 0.024997095879298714
                left:X_0 <= 11.9 ? 0.10741138560687441
                                left:X_0 <= 10.0 ? 0.15195884447962002
                                                                left:X_1 <= 0.31
? 0.27551020408163274
                                                left:1.0
                                                right:X_0 <= 9.4 ? 0.375
                left:0.0
                right:1.0
                                                                right:0.0
                                right:1.0
                right:X_6 <= 99.0 ? 0.016759854529209375
                                left:X_4 <= 0.086 ? 0.012302960399846262
                                                                left:0.0
                                                                right:X_0 <= 5.4
? 0.4444444444444444
                                                left:1.0
                                                right:0.0
                                right:X_0 <= 4.7 ? 0.5
                                                                left:0.0
                                                                right:1.0
        right:X_1 <= 0.21 ? 0.375
                left:0.0
                right:1.0
  right:X_5 <= 18.0 ? 0.07739488813292295
    left:X_7 <= 0.99468 ? 0.043660719823041705
        left:1.0
        right:X_7 <= 0.9948 ? 0.10955198647506342
                left:0.0
                right:X_5 <= 5.0 ? 0.07213358070500941
                                left:0.0
```

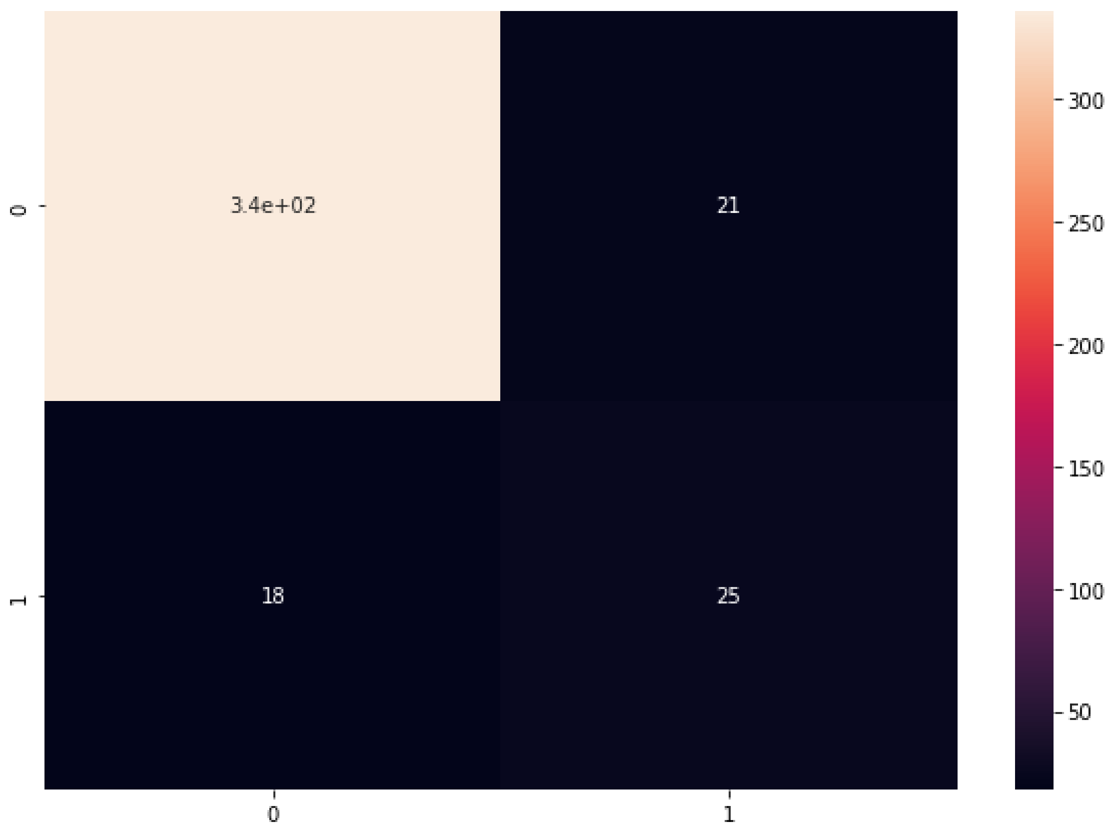
```
                                right:X_4 <= 0.128 ? 0.041838842975206514
                                                                left:X_7 <=
0.9962 ? 0.055338541666666685
                                                left:1.0
                                                right:X_3 <= 2.2 ?
0.2222222222222227
                left:0.0
                right:X_9 <= 0.86 ? 0.16666666666666657
                                left:1.0
                                right:X_5 <= 7.0 ? 0.4444444444444444
                                                                left:0.0
                                                                right:1.0
                                                                right:0.0
    right:X_5 <= 27.0 ? 0.08498959417273677
        left:X_6 <= 50.0 ? 0.1171875
                left:X_4 <= 0.068 ? 0.5
                                left:1.0
                                right:0.0
                right:0.0
        right:X_1 <= 0.58 ? 0.17999999999999994
                left:X_5 <= 38.0 ? 0.11574074074074076
                                left:X_1 <= 0.42 ? 0.08641975308641975
                                                                left:X_0 <= 8.2
? 0.5
                                                left:1.0
                                                right:0.0
                                                                right:1.0
                                right:X_0 <= 8.2 ? 0.4444444444444444
                                                                left:0.0
                                                                right:1.0
                right:0.0
```

2.2.6 Predicting and Performance metrics of the model

[48]:
```python
y_pred = myClassifier_gini.predict(x_test)
print("Accuracy = ", round(accuracy_score(y_test, y_pred)*100, 2), "%")
print("Precision = ",precision_score(y_test, y_pred))
print("Recall = ",recall_score(y_test, y_pred),"\n")

print(classification_report(y_pred,y_test))

cm=metrics.confusion_matrix(y_test,y_pred)
print("\nconfusion matrix: \n",cm)
plt.figure(figsize = (10,7))
sns.heatmap(cm, annot=True)
print("---------------------------------------")
```

```
Accuracy =  90.25 %
```

```
Precision =  0.5434782608695652
Recall =  0.5813953488372093

              precision    recall  f1-score   support

         0.0       0.94      0.95      0.95       354
         1.0       0.58      0.54      0.56        46

    accuracy                           0.90       400
   macro avg       0.76      0.75      0.75       400
weighted avg       0.90      0.90      0.90       400


confusion matrix:
 [[336  21]
 [ 18  25]]
---------------------------------------
```



2.3 B) DECISION TREE from SKLEARN library (Inbuilt modules)

2.4 3. TRAIN THE MODEL

2.4.1 Entropy

[66]:
```
x = df.drop(['quality'], axis=1)
y = df['quality']
```

[61]:
```
#splitting the processed dataset into train and test dataset

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.25,␣
↪random_state = 50)
```

[62]:
```
builtInClassifier_entropy =␣
↪DecisionTreeClassifier(criterion="entropy",max_depth=50)
builtInClassifier_entropy.fit(x_train, y_train)
preds = builtInClassifier_entropy.predict(x_test)
```

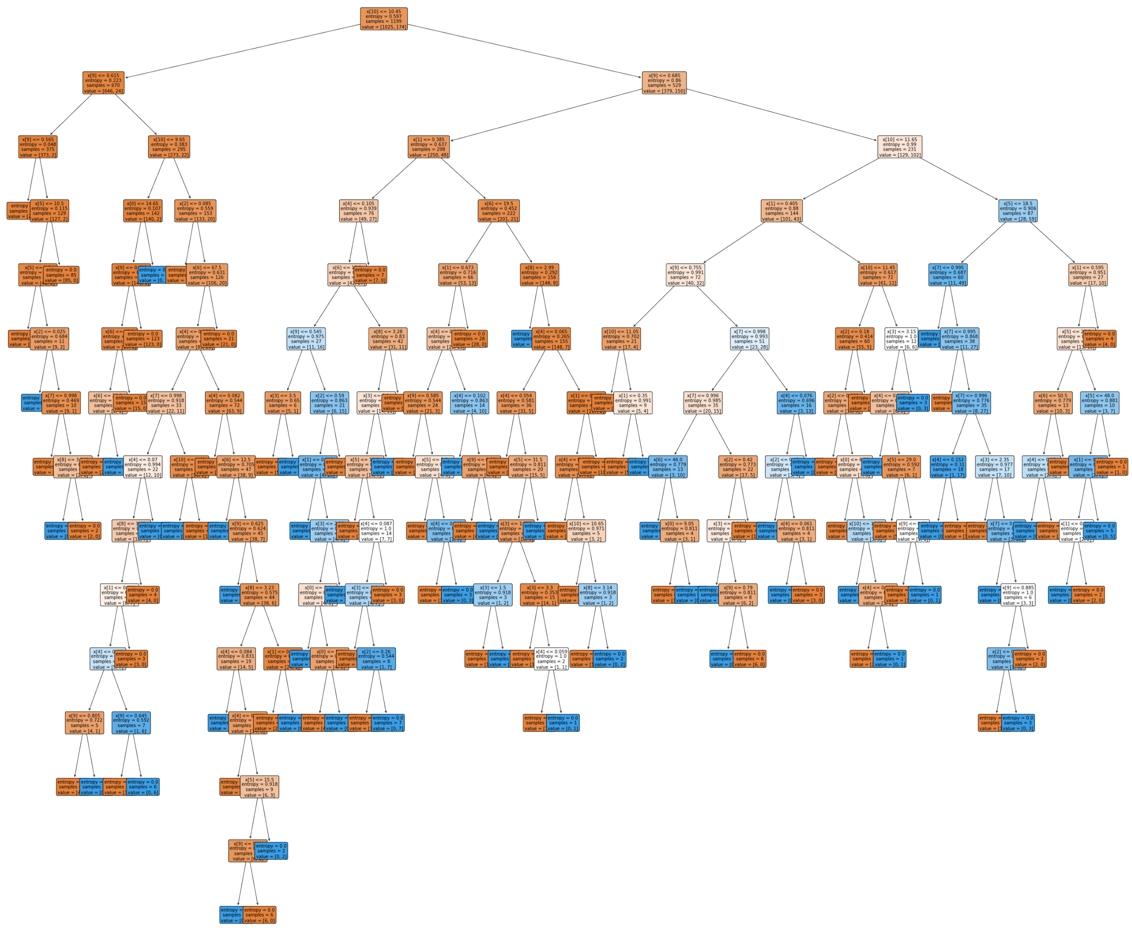
2.4.2 Predicting and Performance metrics of the model

[63]:
```
y_pred = builtInClassifier_entropy.predict(x_test)
print("Accuracy = ", round(accuracy_score(y_test, y_pred)*100, 2), "%")
print("Precision = ",precision_score(y_test, y_pred))
print("Recall = ",recall_score(y_test, y_pred),"\n")

print(classification_report(y_pred,y_test))

cm=metrics.confusion_matrix(y_test,y_pred)
print("\nconfusion matrix: \n",cm)
print("---------------------------------------")
```

```
Accuracy =  93.25 %
Precision =  0.6904761904761905
Recall =  0.6744186046511628

              precision    recall  f1-score   support

           0       0.96      0.96      0.96       358
           1       0.67      0.69      0.68        42

    accuracy                           0.93       400
   macro avg       0.82      0.83      0.82       400
weighted avg       0.93      0.93      0.93       400


confusion matrix:
 [[344  13]
```

```
 [ 14  29]]
----------------------------------------
```

[64]:
```python
fig = plt.figure(figsize=(45,40))
tree.plot_tree(builtInClassifier_entropy,  filled=True, rounded=True,␣
 ↪fontsize=10)
plt.show()
```



[67]:
```python
print(tree.export_text(builtInClassifier_entropy, feature_names = x.columns.
 ↪tolist()))
```

```
|--- alcohol <= 10.45
|   |--- sulphates <= 0.62
|   |   |--- sulphates <= 0.56
|   |   |   |--- class: 0
|   |   |--- sulphates >  0.56
|   |   |   |--- free sulfur dioxide <= 10.50
|   |   |   |   |--- free sulfur dioxide <= 8.50
```

```
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- free sulfur dioxide >  8.50
|   |   |   |   |   |--- citric acid <= 0.02
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- citric acid >  0.02
|   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |--- pH <= 3.25
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- pH >  3.25
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |--- free sulfur dioxide >  10.50
|   |   |   |   |--- class: 0
|   |--- sulphates >  0.62
|   |   |--- alcohol <= 9.65
|   |   |   |--- fixed acidity <= 14.65
|   |   |   |   |--- sulphates <= 0.62
|   |   |   |   |   |--- total sulfur dioxide <= 37.50
|   |   |   |   |   |   |--- total sulfur dioxide <= 26.50
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- total sulfur dioxide >  26.50
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- total sulfur dioxide >  37.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- sulphates >  0.62
|   |   |   |   |   |--- class: 0
|   |   |   |--- fixed acidity >  14.65
|   |   |   |   |--- class: 1
|   |   |--- alcohol >  9.65
|   |   |   |--- citric acid <= 0.09
|   |   |   |   |--- class: 0
|   |   |   |--- citric acid >  0.09
|   |   |   |   |--- total sulfur dioxide <= 67.50
|   |   |   |   |   |--- chlorides <= 0.07
|   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |--- chlorides <= 0.07
|   |   |   |   |   |   |   |   |--- pH <= 3.43
|   |   |   |   |   |   |   |   |   |--- volatile acidity <= 0.53
|   |   |   |   |   |   |   |   |   |   |--- chlorides <= 0.06
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- chlorides >  0.06
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |--- volatile acidity >  0.53
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- pH >  3.43
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- chlorides >  0.07
```

```
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |--- alcohol <= 9.85
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- alcohol >  9.85
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- chlorides >  0.07
|   |   |   |   |   |   |--- chlorides <= 0.08
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- chlorides >  0.08
|   |   |   |   |   |   |   |--- total sulfur dioxide <= 12.50
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- total sulfur dioxide >  12.50
|   |   |   |   |   |   |   |   |--- sulphates <= 0.62
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- sulphates >  0.62
|   |   |   |   |   |   |   |   |   |--- pH <= 3.23
|   |   |   |   |   |   |   |   |   |   |--- chlorides <= 0.08
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- chlorides >  0.08
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 4
|   |   |   |   |   |   |   |   |   |--- pH >  3.23
|   |   |   |   |   |   |   |   |   |   |--- volatile acidity <= 0.86
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- volatile acidity >  0.86
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- total sulfur dioxide >  67.50
|   |   |   |   |   |--- class: 0
|--- alcohol >  10.45
|   |--- sulphates <= 0.69
|   |   |--- volatile acidity <= 0.38
|   |   |   |--- chlorides <= 0.11
|   |   |   |   |--- total sulfur dioxide <= 17.50
|   |   |   |   |   |--- sulphates <= 0.55
|   |   |   |   |   |   |--- residual sugar <= 3.50
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- residual sugar >  3.50
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- sulphates >  0.55
|   |   |   |   |   |   |--- citric acid <= 0.59
|   |   |   |   |   |   |   |--- volatile acidity <= 0.31
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- volatile acidity >  0.31
|   |   |   |   |   |   |   |   |--- residual sugar <= 2.35
|   |   |   |   |   |   |   |   |   |--- fixed acidity <= 7.40
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- fixed acidity >  7.40
|   |   |   |   |   |   |   |   |   |   |--- fixed acidity <= 9.80
```

```
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- fixed acidity >  9.80
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- residual sugar >  2.35
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- citric acid >  0.59
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- total sulfur dioxide >  17.50
|   |   |   |   |   |--- pH <= 3.28
|   |   |   |   |   |   |--- residual sugar <= 3.30
|   |   |   |   |   |   |   |--- free sulfur dioxide <= 12.50
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- free sulfur dioxide >  12.50
|   |   |   |   |   |   |   |   |--- chlorides <= 0.09
|   |   |   |   |   |   |   |   |   |--- residual sugar <= 1.60
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- residual sugar >  1.60
|   |   |   |   |   |   |   |   |   |   |--- citric acid <= 0.26
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- citric acid >  0.26
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- chlorides >  0.09
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- residual sugar >  3.30
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- pH >  3.28
|   |   |   |   |   |   |--- class: 0
|   |   |   |--- chlorides >  0.11
|   |   |   |   |--- class: 0
|   |   |--- volatile acidity >  0.38
|   |   |   |--- total sulfur dioxide <= 19.50
|   |   |   |   |--- volatile acidity <= 0.67
|   |   |   |   |   |--- chlorides <= 0.09
|   |   |   |   |   |   |--- sulphates <= 0.58
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- sulphates >  0.58
|   |   |   |   |   |   |   |--- free sulfur dioxide <= 5.50
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- free sulfur dioxide >  5.50
|   |   |   |   |   |   |   |   |--- chlorides <= 0.05
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- chlorides >  0.05
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- chlorides >  0.09
|   |   |   |   |   |   |--- chlorides <= 0.10
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- chlorides >  0.10
|   |   |   |   |   |   |   |--- sulphates <= 0.62
```

```
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- sulphates >  0.62
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- volatile acidity >  0.67
|   |   |   |   |   |--- class: 0
|   |   |   |--- total sulfur dioxide >  19.50
|   |   |   |   |--- pH <= 2.99
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- pH >  2.99
|   |   |   |   |   |--- chlorides <= 0.07
|   |   |   |   |   |   |--- chlorides <= 0.05
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- chlorides >  0.05
|   |   |   |   |   |   |   |--- free sulfur dioxide <= 31.50
|   |   |   |   |   |   |   |   |--- residual sugar <= 1.73
|   |   |   |   |   |   |   |   |   |--- residual sugar <= 1.50
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- residual sugar >  1.50
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- residual sugar >  1.73
|   |   |   |   |   |   |   |   |   |--- residual sugar <= 3.30
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- residual sugar >  3.30
|   |   |   |   |   |   |   |   |   |   |--- chlorides <= 0.06
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- chlorides >  0.06
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- free sulfur dioxide >  31.50
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- chlorides >  0.07
|   |   |   |   |   |   |--- volatile acidity <= 0.42
|   |   |   |   |   |   |   |--- chlorides <= 0.09
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- chlorides >  0.09
|   |   |   |   |   |   |   |   |--- alcohol <= 10.65
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- alcohol >  10.65
|   |   |   |   |   |   |   |   |   |--- pH <= 3.14
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- pH >  3.14
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- volatile acidity >  0.42
|   |   |   |   |   |   |   |--- class: 0
|   |--- sulphates >  0.69
|   |   |--- alcohol <= 11.65
|   |   |   |--- volatile acidity <= 0.41
|   |   |   |   |--- sulphates <= 0.75
|   |   |   |   |   |--- alcohol <= 11.05
```

```
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- alcohol >  11.05
|   |   |   |   |   |   |--- volatile acidity <= 0.35
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- volatile acidity >  0.35
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- sulphates >  0.75
|   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |--- total sulfur dioxide <= 46.00
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- total sulfur dioxide >  46.00
|   |   |   |   |   |   |   |   |--- fixed acidity <= 9.05
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- fixed acidity >  9.05
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |--- citric acid <= 0.42
|   |   |   |   |   |   |   |   |--- residual sugar <= 2.00
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- residual sugar >  2.00
|   |   |   |   |   |   |   |   |   |--- sulphates <= 0.79
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- sulphates >  0.79
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- citric acid >  0.42
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |--- chlorides <= 0.08
|   |   |   |   |   |   |   |--- citric acid <= 0.48
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- citric acid >  0.48
|   |   |   |   |   |   |   |   |--- chlorides <= 0.06
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- chlorides >  0.06
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- chlorides >  0.08
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |--- volatile acidity >  0.41
|   |   |   |   |--- alcohol <= 11.45
|   |   |   |   |   |--- citric acid <= 0.18
|   |   |   |   |   |   |--- citric acid <= 0.09
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- citric acid >  0.09
|   |   |   |   |   |   |   |--- fixed acidity <= 6.35
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- fixed acidity >  6.35
|   |   |   |   |   |   |   |   |--- alcohol <= 10.85
```

```
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- alcohol >  10.85
|   |   |   |   |   |   |   |   |   |--- chlorides <= 0.10
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- chlorides >  0.10
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- citric acid >  0.18
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- alcohol >  11.45
|   |   |   |   |   |--- residual sugar <= 3.15
|   |   |   |   |   |   |--- chlorides <= 0.07
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- chlorides >  0.07
|   |   |   |   |   |   |   |--- free sulfur dioxide <= 29.00
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- free sulfur dioxide >  29.00
|   |   |   |   |   |   |   |   |--- sulphates <= 0.73
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- sulphates >  0.73
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- residual sugar >  3.15
|   |   |   |   |   |   |--- class: 1
|   |   |--- alcohol >  11.65
|   |   |   |--- free sulfur dioxide <= 18.50
|   |   |   |   |--- density <= 0.99
|   |   |   |   |   |--- class: 1
|   |   |   |   |--- density >  0.99
|   |   |   |   |   |--- density <= 0.99
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- density >  0.99
|   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |--- chlorides <= 0.15
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- chlorides >  0.15
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |--- residual sugar <= 2.35
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- residual sugar >  2.35
|   |   |   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |   |   |--- sulphates <= 0.88
|   |   |   |   |   |   |   |   |   |   |--- citric acid <= 0.51
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- citric acid >  0.51
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- sulphates >  0.88
```

```
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |--- free sulfur dioxide >  18.50
|   |   |   |   |--- volatile acidity <= 0.59
|   |   |   |   |   |--- free sulfur dioxide <= 27.50
|   |   |   |   |   |   |--- total sulfur dioxide <= 50.50
|   |   |   |   |   |   |   |--- chlorides <= 0.07
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- chlorides >  0.07
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- total sulfur dioxide >  50.50
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- free sulfur dioxide >  27.50
|   |   |   |   |   |   |--- free sulfur dioxide <= 48.00
|   |   |   |   |   |   |   |--- volatile acidity <= 0.43
|   |   |   |   |   |   |   |   |--- volatile acidity <= 0.34
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- volatile acidity >  0.34
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- volatile acidity >  0.43
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- free sulfur dioxide >  48.00
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- volatile acidity >  0.59
|   |   |   |   |   |--- class: 0
```

2.4.3 DECISION TREE from SKLEARN library (Inbuilt modules)

2.4.4 GINI INDEX

[68]:
```python
builtInClassifier_gini = DecisionTreeClassifier(criterion="gini",max_depth=50)
builtInClassifier_gini.fit(x_train, y_train)
preds = builtInClassifier_gini.predict(x_test)
score = builtInClassifier_gini.score(x_test, y_test)
score
```

[68]:
```
0.9175
```

2.4.5 Predicting and Performance metrics of the model

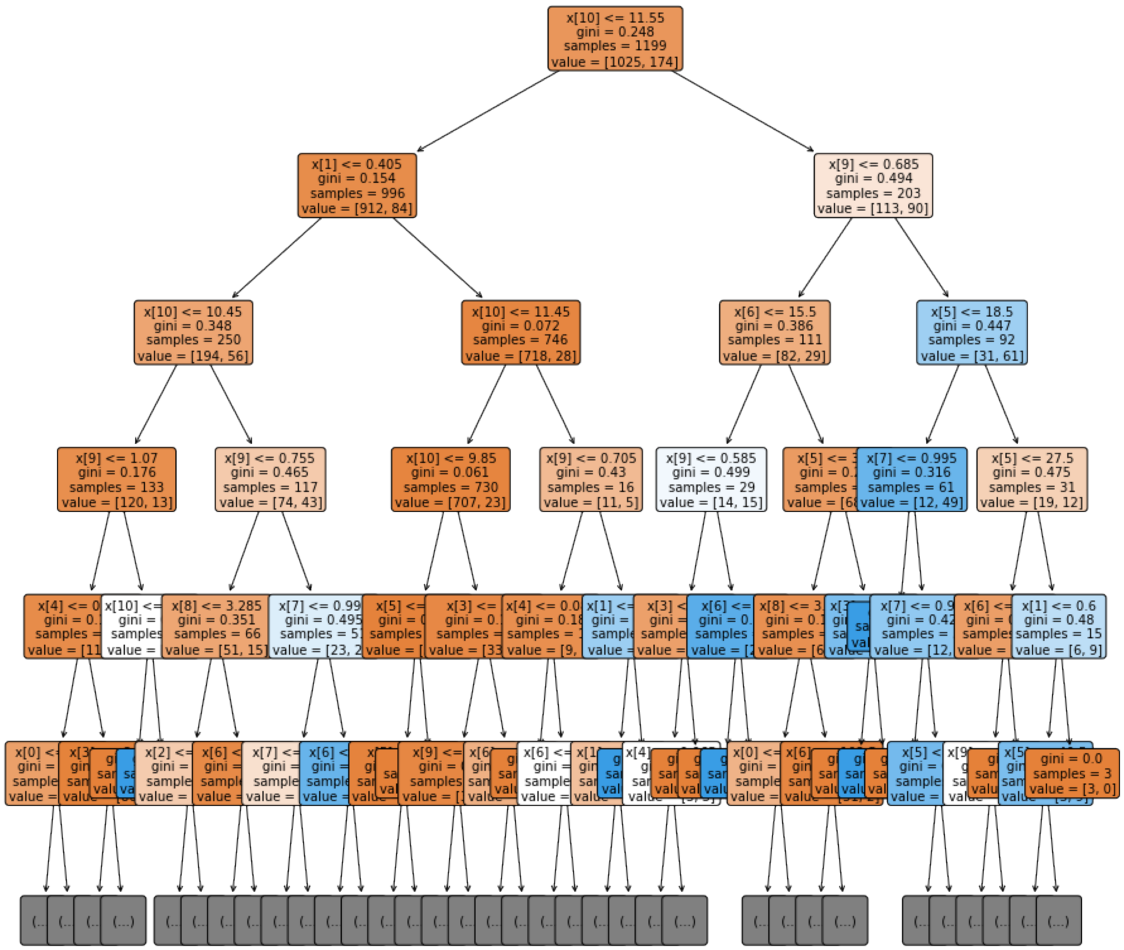
[69]:
```python
y_pred = builtInClassifier_gini.predict(x_test)
print("Accuracy = ", round(accuracy_score(y_test, y_pred)*100, 2), "%")
print("Precision = ",precision_score(y_test, y_pred))
print("Recall = ",recall_score(y_test, y_pred),"\n")

print(classification_report(y_pred,y_test))

cm=metrics.confusion_matrix(y_test,y_pred)
```

```
print("\nconfusion matrix: \n",cm)
print("----------------------------------------")
```

```
Accuracy =  91.75 %
Precision =  0.6190476190476191
Recall =  0.6046511627906976

              precision    recall  f1-score   support

           0       0.96      0.95      0.95       358
           1       0.60      0.62      0.61        42

    accuracy                           0.92       400
   macro avg       0.78      0.79      0.78       400
weighted avg       0.92      0.92      0.92       400


confusion matrix:
 [[341  16]
 [ 17  26]]
---------------------------------------
```

[70]:

```
fig = plt.figure(figsize=(15,15))
tree.plot_tree(builtInClassifier_gini,  filled=True, rounded=True, max_depth=5,
 ↪fontsize=10)
plt.show()
```



[71]:

```
print(tree.export_text(builtInClassifier_gini, feature_names = x.columns.
↪tolist()))
```

```
|--- alcohol <= 11.55
|   |--- volatile acidity <= 0.41
|   |   |--- alcohol <= 10.45
|   |   |   |--- sulphates <= 1.07
|   |   |   |   |--- chlorides <= 0.08
|   |   |   |   |   |--- fixed acidity <= 11.70
|   |   |   |   |   |   |--- residual sugar <= 1.45
|   |   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |   |--- fixed acidity <= 9.15
|   |   |   |   |   |   |   |   |   |--- class: 0
```

```
|   |   |   |   |   |   |   |   |--- fixed acidity >  9.15
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- residual sugar >  1.45
|   |   |   |   |   |   |   |--- volatile acidity <= 0.39
|   |   |   |   |   |   |   |   |--- sulphates <= 0.84
|   |   |   |   |   |   |   |   |   |--- citric acid <= 0.29
|   |   |   |   |   |   |   |   |   |   |--- pH <= 3.25
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |   |--- pH >  3.25
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- citric acid >  0.29
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- sulphates >  0.84
|   |   |   |   |   |   |   |   |   |--- sulphates <= 0.87
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- sulphates >  0.87
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- volatile acidity >  0.39
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- fixed acidity >  11.70
|   |   |   |   |   |   |--- pH <= 2.87
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- pH >  2.87
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- chlorides >  0.08
|   |   |   |   |   |--- residual sugar <= 2.45
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- residual sugar >  2.45
|   |   |   |   |   |   |--- citric acid <= 0.35
|   |   |   |   |   |   |   |--- citric acid <= 0.27
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- citric acid >  0.27
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- citric acid >  0.35
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |--- sulphates >  1.07
|   |   |   |   |--- alcohol <= 9.65
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- alcohol >  9.65
|   |   |   |   |   |--- class: 1
|   |   |--- alcohol >  10.45
|   |   |   |--- sulphates <= 0.75
|   |   |   |   |--- pH <= 3.28
|   |   |   |   |   |--- citric acid <= 0.39
|   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |--- density <= 1.00
```

```
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- citric acid >  0.39
|   |   |   |   |   |   |--- fixed acidity <= 10.50
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- fixed acidity >  10.50
|   |   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |   |--- citric acid <= 0.66
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- citric acid >  0.66
|   |   |   |   |   |   |   |   |   |--- fixed acidity <= 11.95
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- fixed acidity >  11.95
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- pH >  3.28
|   |   |   |   |   |--- total sulfur dioxide <= 10.50
|   |   |   |   |   |   |--- alcohol <= 11.25
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- alcohol >  11.25
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- total sulfur dioxide >  10.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |--- sulphates >  0.75
|   |   |   |   |--- density <= 1.00
|   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |--- total sulfur dioxide <= 46.00
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- total sulfur dioxide >  46.00
|   |   |   |   |   |   |   |--- fixed acidity <= 9.05
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- fixed acidity >  9.05
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |--- pH <= 3.39
|   |   |   |   |   |   |   |--- residual sugar <= 1.65
|   |   |   |   |   |   |   |   |--- citric acid <= 0.50
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- citric acid >  0.50
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- residual sugar >  1.65
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- pH >  3.39
|   |   |   |   |   |   |   |--- alcohol <= 10.75
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- alcohol >  10.75
```

```
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- density >  1.00
|   |   |   |   |   |--- total sulfur dioxide <= 21.50
|   |   |   |   |   |   |--- pH <= 3.28
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- pH >  3.28
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- total sulfur dioxide >  21.50
|   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |--- class: 0
|   |--- volatile acidity >  0.41
|   |   |--- alcohol <= 11.45
|   |   |   |--- alcohol <= 9.85
|   |   |   |   |--- free sulfur dioxide <= 12.50
|   |   |   |   |   |--- free sulfur dioxide <= 8.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- free sulfur dioxide >  8.50
|   |   |   |   |   |   |--- total sulfur dioxide <= 18.50
|   |   |   |   |   |   |   |--- sulphates <= 0.61
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- sulphates >  0.61
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- total sulfur dioxide >  18.50
|   |   |   |   |   |   |   |--- volatile acidity <= 0.52
|   |   |   |   |   |   |   |   |--- volatile acidity <= 0.50
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- volatile acidity >  0.50
|   |   |   |   |   |   |   |   |   |--- chlorides <= 0.08
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- chlorides >  0.08
|   |   |   |   |   |   |   |   |   |   |--- citric acid <= 0.12
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- citric acid >  0.12
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- volatile acidity >  0.52
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- free sulfur dioxide >  12.50
|   |   |   |   |   |--- class: 0
|   |   |   |--- alcohol >  9.85
|   |   |   |   |--- residual sugar <= 5.70
|   |   |   |   |   |--- sulphates <= 0.63
|   |   |   |   |   |   |--- citric acid <= 0.01
|   |   |   |   |   |   |   |--- volatile acidity <= 0.58
|   |   |   |   |   |   |   |   |--- citric acid <= 0.00
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- citric acid >  0.00
```

```
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- volatile acidity >  0.58
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- citric acid >  0.01
|   |   |   |   |   |   |   |--- pH <= 3.51
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- pH >  3.51
|   |   |   |   |   |   |   |   |--- density <= 1.00
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- density >  1.00
|   |   |   |   |   |   |   |   |   |--- sulphates <= 0.62
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |--- sulphates >  0.62
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- sulphates >  0.63
|   |   |   |   |   |   |--- free sulfur dioxide <= 3.50
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- free sulfur dioxide >  3.50
|   |   |   |   |   |   |   |--- residual sugar <= 1.55
|   |   |   |   |   |   |   |   |--- pH <= 3.04
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |--- pH >  3.04
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |--- residual sugar >  1.55
|   |   |   |   |   |   |   |   |--- sulphates <= 0.72
|   |   |   |   |   |   |   |   |   |--- alcohol <= 10.05
|   |   |   |   |   |   |   |   |   |   |--- citric acid <= 0.28
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- citric acid >  0.28
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- alcohol >  10.05
|   |   |   |   |   |   |   |   |   |   |--- chlorides <= 0.12
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 5
|   |   |   |   |   |   |   |   |   |   |--- chlorides >  0.12
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |--- sulphates >  0.72
|   |   |   |   |   |   |   |   |   |--- citric acid <= 0.10
|   |   |   |   |   |   |   |   |   |   |--- fixed acidity <= 8.40
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- fixed acidity >  8.40
|   |   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- citric acid >  0.10
|   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- residual sugar >  5.70
|   |   |   |   |   |--- total sulfur dioxide <= 43.50
|   |   |   |   |   |   |--- residual sugar <= 6.15
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- residual sugar >  6.15
```

```
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- total sulfur dioxide >  43.50
|   |   |   |   |   |   |--- class: 0
|   |   |--- alcohol >  11.45
|   |   |   |--- sulphates <= 0.70
|   |   |   |   |--- chlorides <= 0.09
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- chlorides >  0.09
|   |   |   |   |   |--- total sulfur dioxide <= 36.50
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- total sulfur dioxide >  36.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |--- sulphates >  0.70
|   |   |   |   |--- volatile acidity <= 0.51
|   |   |   |   |   |--- volatile acidity <= 0.46
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- volatile acidity >  0.46
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- volatile acidity >  0.51
|   |   |   |   |   |--- class: 1
|--- alcohol >  11.55
|   |--- sulphates <= 0.69
|   |   |--- total sulfur dioxide <= 15.50
|   |   |   |--- sulphates <= 0.58
|   |   |   |   |--- residual sugar <= 1.90
|   |   |   |   |   |--- chlorides <= 0.07
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- chlorides >  0.07
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- residual sugar >  1.90
|   |   |   |   |   |--- class: 0
|   |   |   |--- sulphates >  0.58
|   |   |   |   |--- total sulfur dioxide <= 8.50
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- total sulfur dioxide >  8.50
|   |   |   |   |   |--- class: 1
|   |   |--- total sulfur dioxide >  15.50
|   |   |   |--- free sulfur dioxide <= 31.50
|   |   |   |   |--- pH <= 3.27
|   |   |   |   |   |--- fixed acidity <= 11.95
|   |   |   |   |   |   |--- fixed acidity <= 10.05
|   |   |   |   |   |   |   |--- chlorides <= 0.08
|   |   |   |   |   |   |   |   |--- fixed acidity <= 9.70
|   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |--- fixed acidity >  9.70
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- chlorides >  0.08
|   |   |   |   |   |   |   |   |--- class: 1
```

```
|   |   |   |   |   |   |--- fixed acidity >  10.05
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- fixed acidity >  11.95
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- pH >  3.27
|   |   |   |   |   |--- total sulfur dioxide <= 102.50
|   |   |   |   |   |   |--- chlorides <= 0.09
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- chlorides >  0.09
|   |   |   |   |   |   |   |--- chlorides <= 0.09
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- chlorides >  0.09
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- total sulfur dioxide >  102.50
|   |   |   |   |   |   |--- pH <= 3.69
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- pH >  3.69
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |--- free sulfur dioxide >  31.50
|   |   |   |   |--- residual sugar <= 1.65
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- residual sugar >  1.65
|   |   |   |   |   |--- class: 1
|   |--- sulphates >  0.69
|   |   |--- free sulfur dioxide <= 18.50
|   |   |   |--- density <= 0.99
|   |   |   |   |--- class: 1
|   |   |   |--- density >  0.99
|   |   |   |   |--- density <= 0.99
|   |   |   |   |   |--- class: 0
|   |   |   |   |--- density >  0.99
|   |   |   |   |   |--- free sulfur dioxide <= 5.50
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |--- free sulfur dioxide >  5.50
|   |   |   |   |   |   |--- total sulfur dioxide <= 39.50
|   |   |   |   |   |   |   |--- chlorides <= 0.16
|   |   |   |   |   |   |   |   |--- free sulfur dioxide <= 9.50
|   |   |   |   |   |   |   |   |   |--- total sulfur dioxide <= 20.00
|   |   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |   |   |--- total sulfur dioxide >  20.00
|   |   |   |   |   |   |   |   |   |   |--- fixed acidity <= 10.80
|   |   |   |   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |   |   |   |   |--- fixed acidity >  10.80
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |--- free sulfur dioxide >  9.50
|   |   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- chlorides >  0.16
|   |   |   |   |   |   |   |   |--- class: 0
```

```
|   |   |   |   |   |   |--- total sulfur dioxide >  39.50
|   |   |   |   |   |   |   |--- class: 0
|   |   |--- free sulfur dioxide >  18.50
|   |   |   |--- free sulfur dioxide <= 27.50
|   |   |   |   |--- total sulfur dioxide <= 50.50
|   |   |   |   |   |--- sulphates <= 0.77
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- sulphates >  0.77
|   |   |   |   |   |   |--- class: 0
|   |   |   |   |--- total sulfur dioxide >  50.50
|   |   |   |   |   |--- class: 0
|   |   |   |--- free sulfur dioxide >  27.50
|   |   |   |   |--- volatile acidity <= 0.60
|   |   |   |   |   |--- free sulfur dioxide <= 40.50
|   |   |   |   |   |   |--- volatile acidity <= 0.43
|   |   |   |   |   |   |   |--- volatile acidity <= 0.38
|   |   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |   |--- volatile acidity >  0.38
|   |   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- volatile acidity >  0.43
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |--- free sulfur dioxide >  40.50
|   |   |   |   |   |   |--- density <= 0.99
|   |   |   |   |   |   |   |--- class: 0
|   |   |   |   |   |   |--- density >  0.99
|   |   |   |   |   |   |   |--- class: 1
|   |   |   |   |--- volatile acidity >  0.60
|   |   |   |   |   |--- class: 0
```

2.4.6 Checking the best suited depth of decision tree for the dataset (this time inbuilt module is used !!)

[74]:
```python
Ks = 100
mean_acc = np.zeros((Ks-1))
for n in range(1,Ks):

    #Train Model and Predict
    builtInClassifier = DecisionTreeClassifier(max_depth = n).
 ↪fit(x_train,y_train)
    yhat=builtInClassifier.predict(x_test)
    mean_acc[n-1] = metrics.accuracy_score(y_test, yhat)

print(mean_acc)
```

```
[0.8925 0.9075 0.9125 0.905  0.915  0.9175 0.9075 0.9125 0.9125 0.9025
 0.8925 0.9025 0.8925 0.915  0.915  0.9025 0.9025 0.8975 0.905  0.9
 0.9025 0.9    0.9    0.905  0.9075 0.895  0.9    0.91   0.9    0.9125
```

```
0.89   0.9025 0.9075 0.9    0.91   0.915  0.9025 0.905  0.9125 0.8975
0.9125 0.91   0.9    0.91   0.915  0.9    0.9125 0.9025 0.9075 0.9125
0.9075 0.915  0.895  0.905  0.915  0.9075 0.905  0.905  0.9075 0.9
0.9125 0.9075 0.9025 0.915  0.9075 0.905  0.8975 0.9075 0.9075 0.9025
0.9125 0.8975 0.9    0.9075 0.895  0.9025 0.905  0.92   0.9075 0.91
0.9    0.905  0.8975 0.91   0.895  0.9025 0.91   0.9    0.9125 0.9125
0.9    0.9025 0.905  0.9025 0.905  0.9025 0.9025 0.905  0.905 ]
```

[75]:
```
print( "The best accuracy was with", mean_acc.max(), "with depth =", mean_acc.
 ↪argmax()+1)
```

```
The best accuracy was with 0.92 with depth = 78
```

The best accuracy was with 0.9225 with depth = 6

2.5 4. EVALUATE THE PERFORMANCE OF THE ALGORITHMS:

2.6 COMPARISON OF PERFORMANCE OF BOTH IMPLEMENTATION:

2.6.1 (i) DECISION TREE from scratch:

Entropy:

1. Accuracy = 90.25 %
2. Precision = 0.5434782608695652
3. Recall = 0.5813953488372093

Gini:

1. Accuracy = 90.25 %
2. Precision = 0.5434782608695652
3. Recall = 0.5813953488372093

2.6.2 (ii) DECISION TREE using SKlearn module:

Entropy:

1. Accuracy = 91.5 %
2. Precision = 0.6097560975609756
3. Recall = 0.5813953488372093

Gini:

1. Accuracy = 90.25 %
2. Precision = 0.5434782608695652
3. Recall = 0.5813953488372093

[ ]:

[ ]: