Name :-

Roll No :-

Practical No:- 5.2

Practical Name:- Write a program implement to decision tree to popular attribute selection measure like information gain,gini index etc.for decision tree

import matplotlib.pyplot as plt  
import pandas as pd  
from sklearn.datasets import load\_iris  
data\_b = load\_iris()  
df=pd.DataFrame(data\_b.data,columns=data\_b.feature\_names)  
df['target'] = data\_b.target  
#df['target']  
print(df)  
#print(data\_b)  
print("Dataset Labels=",data\_b.target\_names)  
from sklearn.tree import DecisionTreeClassifier  
from sklearn import metrics  
from sklearn import tree  
from sklearn.model\_selection import train\_test\_split  
x\_train, x\_test, y\_train, y\_test = train\_test\_split(df[data\_b.feature\_names], df['target'])  
print(x\_train)  
print(x\_test)  
print(y\_train)  
print(y\_test)  
clf = DecisionTreeClassifier(max\_depth = 5,random\_state =1, criterion='gini') #'gini'  
clf = clf.fit(x\_train, y\_train)  
y\_pred = clf.predict(x\_test)  
print(y\_test, y\_pred)  
print("Accuracy:",metrics.accuracy\_score(y\_test, y\_pred))  
fn=['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)']  
cn=['setosa', 'versicolor', 'virginica']  
  
fig, axes = plt.subplots(nrows = 1, ncols = 1, figsize = (4, 4), dpi = 300)  
tree.plot\_tree(clf, feature\_names = fn, class\_names = cn, filled = True); fig.savefig('dstimg.png')

output:-

C:\Users\patil\PycharmProjects\ml\venv\Scripts\python.exe C:\Users\patil\PycharmProjects\ml\ml4.py

sepal length (cm) sepal width (cm) ... petal width (cm) target

0 5.1 3.5 ... 0.2 0

1 4.9 3.0 ... 0.2 0

2 4.7 3.2 ... 0.2 0

3 4.6 3.1 ... 0.2 0

4 5.0 3.6 ... 0.2 0

.. ... ... ... ... ...

145 6.7 3.0 ... 2.3 2

146 6.3 2.5 ... 1.9 2

147 6.5 3.0 ... 2.0 2

148 6.2 3.4 ... 2.3 2

149 5.9 3.0 ... 1.8 2

[150 rows x 5 columns]

Dataset Labels= ['setosa' 'versicolor' 'virginica']

sepal length (cm) sepal width (cm) petal length (cm) petal width (cm)

15 5.7 4.4 1.5 0.4

73 6.1 2.8 4.7 1.2

53 5.5 2.3 4.0 1.3

104 6.5 3.0 5.8 2.2

69 5.6 2.5 3.9 1.1

.. ... ... ... ...

96 5.7 2.9 4.2 1.3

18 5.7 3.8 1.7 0.3

77 6.7 3.0 5.0 1.7

88 5.6 3.0 4.1 1.3

24 4.8 3.4 1.9 0.2

[112 rows x 4 columns]

sepal length (cm) sepal width (cm) petal length (cm) petal width (cm)

148 6.2 3.4 5.4 2.3

1 4.9 3.0 1.4 0.2

30 4.8 3.1 1.6 0.2

107 7.3 2.9 6.3 1.8

123 6.3 2.7 4.9 1.8

149 5.9 3.0 5.1 1.8

132 6.4 2.8 5.6 2.2

9 4.9 3.1 1.5 0.1

112 6.8 3.0 5.5 2.1

117 7.7 3.8 6.7 2.2

75 6.6 3.0 4.4 1.4

102 7.1 3.0 5.9 2.1

89 5.5 2.5 4.0 1.3

127 6.1 3.0 4.9 1.8

37 4.9 3.6 1.4 0.1

16 5.4 3.9 1.3 0.4

29 4.7 3.2 1.6 0.2

83 6.0 2.7 5.1 1.6

133 6.3 2.8 5.1 1.5

135 7.7 3.0 6.1 2.3

40 5.0 3.5 1.3 0.3

59 5.2 2.7 3.9 1.4

43 5.0 3.5 1.6 0.6

106 4.9 2.5 4.5 1.7

131 7.9 3.8 6.4 2.0

23 5.1 3.3 1.7 0.5

26 5.0 3.4 1.6 0.4

74 6.4 2.9 4.3 1.3

70 5.9 3.2 4.8 1.8

109 7.2 3.6 6.1 2.5

90 5.5 2.6 4.4 1.2

99 5.7 2.8 4.1 1.3

139 6.9 3.1 5.4 2.1

20 5.4 3.4 1.7 0.2

62 6.0 2.2 4.0 1.0

147 6.5 3.0 5.2 2.0

116 6.5 3.0 5.5 1.8

118 7.7 2.6 6.9 2.3

15 0

73 1

53 1

104 2

69 1

..

96 1

18 0

77 1

88 1

24 0

Name: target, Length: 112, dtype: int32

148 2

1 0

30 0

107 2

123 2

149 2

132 2

9 0

112 2

117 2

75 1

102 2

89 1

127 2

37 0

16 0

29 0

83 1

133 2

135 2

40 0

59 1

43 0

106 2

131 2

23 0

26 0

74 1

70 1

109 2

90 1

99 1

139 2

20 0

62 1

147 2

116 2

118 2

Name: target, dtype: int32

148 2

1 0

30 0

107 2

123 2

149 2

132 2

9 0

112 2

117 2

75 1

102 2

89 1

127 2

37 0

16 0

29 0

83 1

133 2

135 2

40 0

59 1

43 0

106 2

131 2

23 0

26 0

74 1

70 1

109 2

90 1

99 1

139 2

20 0

62 1

147 2

116 2

118 2

Name: target, dtype: int32 [2 0 0 2 2 2 2 0 2 2 1 2 1 2 0 0 0 2 2 2 0 1 0 1 2 0 0 1 2 2 1 1 2 0 1 2 2

2]

Accuracy: 0.9210526315789473

Process finished with exit code 0

