2_DT_Iris

January 2022

[ ]:
```
#Import scikit-learn dataset library
from sklearn import datasets
from sklearn.tree import DecisionTreeClassifier
import pandas as pd
import numpy as np

#Load dataset
iris = datasets.load_iris()
```

[ ]:
```
# print the names of the 4 features


# print the label type of iris(class_0, class_1, class_2)


# print data(feature)shape
```

[ ]:
```
#import the necessary module
from sklearn.model_selection import train_test_split

#split data set into train and test sets
```

[ ]:
```
#Create a Decision Tree Classifier (using Gini)


#Train the model using the training sets
```

[ ]:
```
# Predict the classes of test data


#print(test_pred.dtype)
from sklearn import metrics
```

[ ]:
```
# Model Accuracy, how often is the classifier correct?
```

[ ]:
```
from sklearn.tree import export_graphviz
#export_graphviz(clf,out_file='iris_tree.dot',feature_names=list(iris.
 ↪feature_names),
```

```
 #                class_names=list(iris.target_names), filled=True)

# Convert to png
#from subprocess import call
#call(['dot', '-Tpng', 'iris_tree.dot', '-o', 'iris_tree.png', '-Gdpi=600'])
```

[ ]:
```
# Display in python
import matplotlib.pyplot as plt
#plt.figure(figsize = (14, 18))
#plt.imshow(plt.imread('iris_tree.png'))
#plt.axis('off')
#plt.show()
```

[ ]:
```

```

Exercise:

1) 1<=Rollnumber<=25: #Task 1: Try the algo on Same Weather dataset - OneHotEncoding of features: and Train test Division 70%-30%

#Task 2: Apply algorithm on wine dataset - LabelEncoding of features: and Train test Division 80%-20%

2) 26<=Rollnumber<=50: #Task 1: Try the algo on Same Weather dataset - LabelEncoding of features: and Train test Division 80%-20% with Gini Index as attribute selection measure

#Task 2: Apply algorithm on digits dataset - One Hot Encoding of features: and Train test Division 65%-35%

3) 51<=Rollnumber<=75: #Task 1: Try the algo on Same Weather dataset- LabelEncoding of features: and Train test Division 90%-10%

#Task 2: Apply algorithm on breast cancer wisconsin dataset - One Hot Encoding of features: and Train test Division 60%-40%

4) 76<=Rollnumber<=100: #Task 1: Try the algo on Same Weather dataset - OneHotEncoding of features: and Train test Division 75%-25%

#Task 2: Apply algorithm on digits dataset - LabelEncoding of features: and Train test Division 80%-20%

5) 101<=Rollnumber<=125: #Task 1: Try the algo on Same Weather dataset - OneHotEncoding of features:and Train test Division 85%-15% and Gini Index as attribute selection measure

#Task 2: Apply algorithm on breast cancer wisconsin dataset - One Hot Encoding of features: and Train test Division 50%-50%

6) 126<=Rollnumber + All with No RollNumbers: #Task 1: Try the algo on Same Weather dataset - LabelEncoding of features:and Train test Division 95%-5%

#Task 2: Apply algorithm on wine dataset - LabelEncoding of features: and Train test Division 66%-34%

Instruction for Task-1 & 2:

i) Set Random state and maximum allowed leaf of model equals to your roll number (or last 2 digit of your id -if you don't have roll number)

Questions: For Task - 1

(1) What will be the value of Play, if Outlook is 'Rainy', Temperature is 'Mild', Humidity ='Normal', and Wind = 'False'?

(2) What will be the value of Play, if Outlook is 'Sunny', Temeprature is 'Cool', Humidity ='High', and Wind = 'True'?

(3) Accuracy , precision and recall of both Models?