Data Driven Computing and Networking (DDCN-2019)

Classification using Decision Tree Algorithm

A. Write a python script to perform the following tasks:-

1. **Load all the required packages to implement Decision Tree Classifier based algorithm**

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn import tree
```

2. **Import packages to load the datasets**

```
import pandas as pd
```

3. **Load the dataset "balance.csv"**

```
balance_data=pd.read_csv('C:/Users/ pc/Desktop/DDCN
2019/Classification/Datasets/balance.csv',header=None,sep=',')
```

4. **Printing Dataset characterstics and five records**

```
print("Dataset Length:: ",len(balance_data))

print("DatasetShape:: ",balance_data.shape)

print ("Dataset::\n ", balance_data.head())
```

5. **Slice data set to create feature set X by taking first second to five columns and target set Y as the first column**

```
X = balance_data.values[:, 1:5]
Y = balance_data.values[:,0]
```

6. **Spliting feature set and target set both into trainingset and test set**

```
X_train, X_test, y_train, y_test = train_test_split( X, Y, test_size = 0.3, random_state = 100)
```

7. **Create Decision Tree Classifier using gini index**

```
clf_gini = DecisionTreeClassifier(criterion = "gini", random_state = 100,
                                  max_depth=3, min_samples_leaf=5)
```

8. **Train the created Decision Tree classifier model on training data set of feature setand target set**

```
clf_gini.fit(X_train, y_train)
```

9. **Make prediction using gini index based Decision Tree Classifier on a random newdata [4, 4, 3, 3]**

```
clf_gini.predict([[4, 4, 3, 3]])
```

10. **Make prediction of target using gini index based Decision Tree Classifier for test data set of feature set data**

```
y_pred = clf_gini.predict(X_test)
```

11. **Print Predictions using Gini_index criteria**

```
print ("Predictions using Gini_index")
print (y_pred)
```

12. **Print accuracy of gini index based Decision Tree Classifier for the test data set of target set**

```
print ("Accuracy of Predictions using Gini_index is ", accuracy_score(y_test,y_pred)*100)
```

13. **Create Decision Tree Classifier using entropy**

```
clf_entropy = DecisionTreeClassifier(criterion = "entropy", random_state = 100,
 max_depth=3, min_samples_leaf=5)
```

14. **Train Decision Tree classifier model on training data set of feature set and target set**

```
clf_entropy.fit(X_train, y_train)
```

15. **Make prediction of target using information gain based Decision Tree Classifier for test data set of feature set data**

```
y_pred_en = clf_entropy.predict(X_test)
```

16. **Print Predictions using Entropy Measure**

```
print ("Predictions using Entropy Measure"
print y_pred_en)
```

17. **Print accuracy of information gain based Decision Tree Classifier for the target set**

```
print ("Accuracy of Predictions using Entropy Measure is ",
accuracy_score(y_test,y_pred_en)*100)
```