Name: Virti Bipin Sanghavi

Student ID : 1001504428

Course: CSE 6363 – Machine Learning

Packages used :

Numpy

Pandas

Sklearn

Math

CSV

C45

1)Explanation of the interfaces:

In Assignment1.py

1)**function saveContent**

It takes filename and training dataset as the input

This function uses the training dataset and stores the content in order for sending it for preprocessing, cleaning, then integration etc.

2)**function kFoldValidation**

Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples.

Here we perform splitting of data and appending the split data and setting the fold size.

The function returns the splitted Data.

3) **function dataSplitting(dataset)**

Here the training dataset is splitted 80% and test data is splitted 20%

The function returns the accuracy of the kfold test data.

4) retrieveData() is used to fetch the data which has been preprocessed and printed.

5) constructTree() – constructing the Decision tree

6) accuracy() – To calculate the accuracy of the functions k-fold and splitting attribute is performed on the Test Data.

Name: Virti Bipin Sanghavi

Student ID : 1001504428

Course: CSE 6363 – Machine Learning

7) **get_index(self, label):** This function returns the index of the attribute w.r.t the label passed.
Parameter: Self: it is an object of the class.
Label: class label

8) **allSameClass(self, data):** This function checks if the predicted class and the actual class is same.

2) Steps of Experiments

Import numpy

Import Pandas

Import CSV

Import Math

Import C45

Working of the Algorithm:

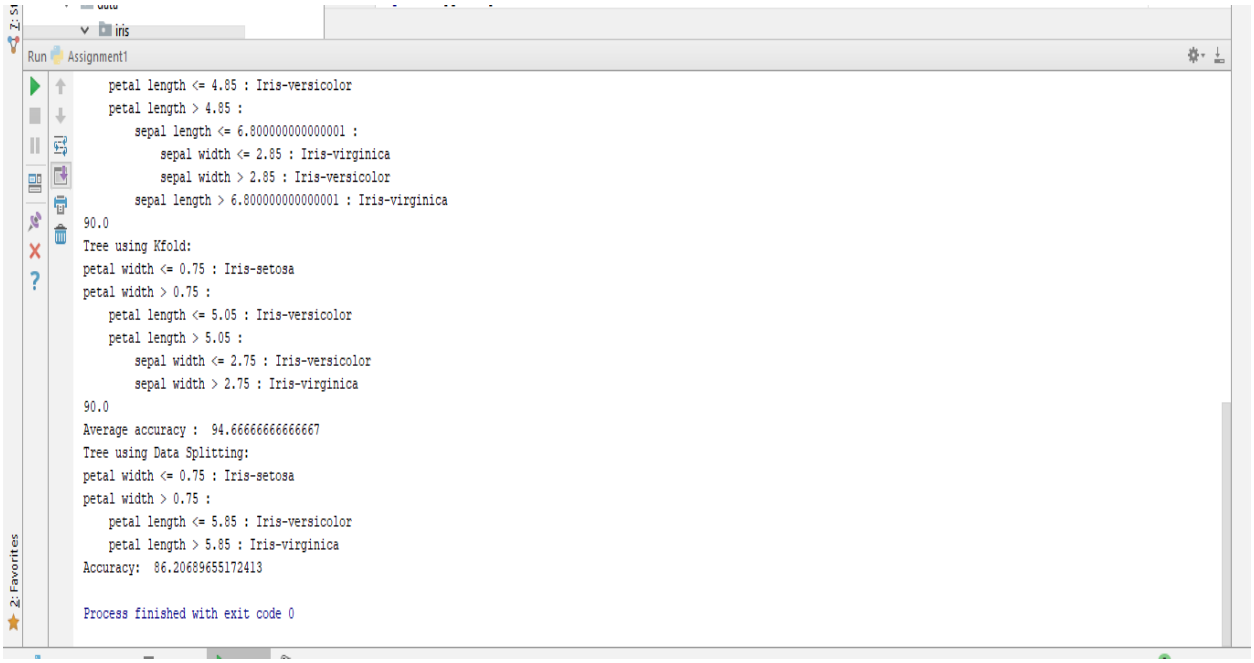
1. Split the data into train and test using k-fold and data splitting function 80-20.
2. Create C45 object with train data and labels.
3. Fetch the Data – calling retrieveData()
4. Clean the data and preprocess the data
5. Generate tree recursively using printTree()
6. Calculate the accuracy by predicting the classes and comparing it with the actual class.

3) Visualization of the Decision Tree and Predicting Accuracy

```
Tree using Kfold:
petal width <= 0.8 : Iris-setosa
petal width > 0.8 :
    petal length <= 4.85 : Iris-versicolor
    petal length > 4.85 :
        sepal length <= 6.15 : Iris-versicolor
        sepal length > 6.15 :
            sepal width <= 2.85 : Iris-virginica
            sepal width > 2.85 : Iris-virginica
100.0
Tree using Kfold:
petal width <= 0.8 : Iris-setosa
petal width > 0.8 :
    petal length <= 4.800000000000001 : Iris-versicolor
    petal length > 4.800000000000001 :
        sepal length <= 7.25 :
            sepal width <= 3.05 : Iris-virginica
            sepal width > 3.05 : Iris-versicolor
        sepal length > 7.25 : Iris-virginica
96.66666666666667
Tree using Kfold:
petal width <= 0.8 : Iris-setosa
petal width > 0.8 :
    petal length <= 4.85 : Iris-versicolor
```

```
Run  Assignment1
        sepal length > 6.15 :
            sepal width <= 3.05 : Iris-virginica
            sepal width > 3.05 : Iris-virginica
96.66666666666667
Tree using Kfold:
petal width <= 0.8 : Iris-setosa
petal width > 0.8 :
    petal length <= 4.85 : Iris-versicolor
    petal length > 4.85 :
        sepal length <= 6.800000000000001 :
            sepal width <= 2.85 : Iris-virginica
            sepal width > 2.85 : Iris-versicolor
        sepal length > 6.800000000000001 : Iris-virginica
90.0
Tree using Kfold:
petal width <= 0.75 : Iris-setosa
petal width > 0.75 :
    petal length <= 5.05 : Iris-versicolor
    petal length > 5.05 :
        sepal width <= 2.75 : Iris-versicolor
        sepal width > 2.75 : Iris-virginica
90.0
Average accuracy :  94.66666666666667
Tree using Data Splitting:
petal width <= 0.75 : Iris-setosa
```

Python Console Terminal 4: Run 6: TODO Event Log



```
    petal length <= 4.85 : Iris-versicolor
    petal length > 4.85 :
        sepal length <= 6.800000000000001 :
            sepal width <= 2.85 : Iris-virginica
            sepal width > 2.85 : Iris-versicolor
        sepal length > 6.800000000000001 : Iris-virginica
90.0
Tree using Kfold:
petal width <= 0.75 : Iris-setosa
petal width > 0.75 :
    petal length <= 5.05 : Iris-versicolor
    petal length > 5.05 :
        sepal width <= 2.75 : Iris-versicolor
        sepal width > 2.75 : Iris-virginica
90.0
Average accuracy :  94.66666666666667
Tree using Data Splitting:
petal width <= 0.75 : Iris-setosa
petal width > 0.75 :
    petal length <= 5.85 : Iris-versicolor
    petal length > 5.85 : Iris-virginica
Accuracy:  86.20689655172413

Process finished with exit code 0
```

References : Used some online coding repositories and blogs as a reference.