CS561 - ARTIFICIAL INTELLIGENCE LAB
ASSIGNMENT-4: Decision Trees

GroupID : 1801cs12_1801cs16_1801cs22
Date : 05/10/2021

1801CS12 : Bablu Kumar
1801CS16 : Mangesh Chandrawanshi
1801CS22 : Hrishabh Raj

Report the 10-fold cross-validation results in terms of precision, recall, and F-score

See kfold10cvreport.txt

Report results of feature ablation study and state which feature has contributed most towards correctly predicting a particular class

See featureablationreport.txt

In this we have considered different combinations of features to classify
00000 : stands for none of the features used
……..
11111 : all the features are used

We have printed out accuracy values for each of the class predictions using all possible feature combinations which will help to determine the dependency of a particular feature in predicting a particular class. e.g.

```
Labels for prediction: ['DESC', 'ENTY', 'NUM', 'LOC', 'ABBR', 'HUM']
Feature ablation report(avgLength, unigrams, bigrams, trigrams, postags):


Training DecisionTree with No Unigrams; No Bigrams; No Trigrams; No POS_Tags; No AvgLength;
------------------------------------------------------------
Accuracy Report: {'DESC': 1.0, 'ENTY': 0.0, 'NUM': 0.0, 'LOC': 0.0, 'ABBR': 0.0, 'HUM': 0.0}
```

Report precision, recall, and F-score measures on test sets using models based on the gini index, mis-classification error and cross-entropy

See modelreport.txt

Show whether errors propagated by one model are corrected by other models or not. If yes, then report how many percent of samples are corrected.
Ex. Observe how many samples are mis-classified using gini index based model but correctly classified by mis-classification error and cross-entropy based model

Values taken from confusion matrix on next page

```
class    miserror   gini    entropy

DESC       114       133      133

ENTY       72        69        54

'gini' and 'entropy' improve upon 'miserror'
for class DESC by about 15%

'miserror' improves upon 'entropy' for
class ENTY by about 25%
```

```
Labels for prediction: ['ABBR', 'DESC', 'NUM', 'ENTY', 'LOC', 'HUM']
```

```
Training model with miserror information gain...
'miserror' Model Report
---------------------------
Precision: 0.8501638034606601
Recall: 0.7695624126429937
F-Score: 0.7990780553903548
Confusion Matrix:
 [[  6   3   0   0   0   0]
 [  0 114   0  23   1   0]
 [  0  14  90   6   3   0]
 [  0  16   2  72   3   1]
 [  0   5   1  14  58   3]
 [  0   0   0   8   2  55]]
```

```
Training model with entropy information gain...
'entropy' Model Report
---------------------------
Precision: 0.8165712979504854
Recall: 0.7620693051830414
F-Score: 0.7806341472102495
Confusion Matrix:
 [[  6   3   0   0   0   0]
 [  1 133   0   4   0   0]
 [  0  14  91   6   2   0]
 [  0  30   0  54   5   5]
 [  1  14   1   6  58   1]
 [  0   2   0   7   1  55]]
```

```
Training model with gini information gain...
'gini' Model Report
---------------------------
Precision: 0.8317008708851444
Recall: 0.785518260384753
F-Score: 0.802431531675766
Confusion Matrix:
 [[  6   3   0   0   0   0]
 [  1 133   0   4   0   0]
 [  0  13  92   6   2   0]
 [  0  18   0  69   4   3]
 [  1   5   0  17  57   1]
 [  0   1   0   8   2  54]]
```