CS561 Artificial Intelligence Lab
Assignment 4: Decision Trees

Group ID: 1801cs03_1801cs07_1801cs46
Members:

1801CS03	Abhishek Chopra
1801CS07	Amish Mittal
1801CS46	Shashwat Mahajan

Q1: Report the 10-fold cross-validation results in terms of precision, recall, and F-score.
A: Refer to kfold10cvreport.txt

Q2: Report results of feature ablation study and state which feature has contributed most towards correctly predicting a particular class.
A: Refer to featureablationreport.txt
To find out the different ablations, we know that there are 5 features in total. Now this means, there can be 2^5 different combinations possible. Therefore, we iterate from 00000 to 11111 (in binary) and use these to carry out ablations.
00000: None of the features selected
.....
11111: All features selected

In ablation studies, we calculate the accuracy values for each of the classes for a given combination of features. Now this is helpful to determine whether a particular class is heavily dependent on a given feature or not.

```
Feature ablation report(avgLength, unigrams, bigrams, trigrams, postags):


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

Q3: Report precision, recall, and F-score measures on test sets using models based on the gini index, mis-classification error and cross-entropy.
A: Refer to modelreport.txt

Q4: 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 the gini index based model but correctly classified by mis-classification error and cross-entropy based model.
A: Using the confusion matrix:

Class	**miserror**	**gini**	**entropy**
DESC	114	133	133
ENTY	72	69	54

We can see that gini and entropy improve upon miserror for DESC by ~15%. Similarly, miserror improves upon entropy for ENTY by ~25%

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

```
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]]
```

```
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]]
```