CS561- Assignment-06

1.) After performing 10-fold cross-validation, we obtained the following results:

Gini Index
Precision Score = 0.8040501395611818
Recall Score = 0.7487842295806157
F Score = 0.767009336109572

2. **Feature Ablation Study:**
We have conducted the analysis for both the length feature and the POS Tag feature using three different metrics: Gini Index, Misclassification Error, and Entropy.

We are using accuracy while using 3 metrics.

Feature Removing	Gini Index	Entropy	Misclassification
Including all Feature	0.82	0.79	0.79
Length	0.82	0.79	0.79
POS Tag	0.79	0.64	0.76

Based on the results, we can conclude that removing the length feature does not significantly affect accuracy. However, when we remove the POS (Part-of-Speech) tag feature, accuracy decreases noticeably.

3). Results when we are using 3 different metrics:

Gini index:

```
Prediction done...
{'ABBR': 0.6666666666666666, 'DESC': 0.9710144927536232, 'ENTY': 0.723404255319149, 'HUM': 0.8461538461538461, 'LOC': 0.7037037037037037, 'NUM': 0.8
141592920353983}
              precision    recall  f1-score   support

        ABBR       0.86      0.67      0.75         9
        DESC       0.76      0.97      0.85       138
        ENTY       0.68      0.72      0.70        94
         HUM       0.92      0.85      0.88        65
         LOC       0.89      0.70      0.79        81
         NUM       0.99      0.81      0.89       113

    accuracy                           0.82       500
   macro avg       0.85      0.79      0.81       500
weighted avg       0.84      0.82      0.82       500
```

Entropy:

```
Prediction done...
{'ABBR': 0.6666666666666666, 'DESC': 0.9710144927536232, 'ENTY': 0.5, 'HUM': 0.8615384615384616, 'LOC': 0.7283950617283951, 'NUM': 0.805309734513274
4}
              precision    recall  f1-score   support

        ABBR       0.86      0.67      0.75         9
        DESC       0.66      0.97      0.79       138
        ENTY       0.69      0.50      0.58        94
         HUM       0.90      0.86      0.88        65
         LOC       0.88      0.73      0.80        81
         NUM       0.98      0.81      0.88       113

    accuracy                           0.79       500
   macro avg       0.83      0.76      0.78       500
weighted avg       0.81      0.79      0.78       500
```

Miscalculation:

```
Prediction done...
{'ABBR': 0.6666666666666666, 'DESC': 0.8260869565217391, 'ENTY': 0.7978723404255319, 'HUM': 0.8461538461538461, 'LOC': 0.691358024691358, 'NUM': 0.7
876106194690266}
              precision    recall  f1-score   support

        ABBR       0.86      0.67      0.75         9
        DESC       0.77      0.83      0.79       138
        ENTY       0.57      0.80      0.66        94
         HUM       0.92      0.85      0.88        65
         LOC       0.92      0.69      0.79        81
         NUM       0.98      0.79      0.87       113

    accuracy                           0.79       500
   macro avg       0.83      0.77      0.79       500
weighted avg       0.82      0.79      0.80       500
```

4.) Where entropy correctly classifies but gini does not correctly classified and also Where misclassification correctly classifies but gini does not correctly classified.

```
:
1  print('no of samples Entropy correctly classifies but gini dont :', (len(gini_wrong-entropy_wrong)))
2  print('no of samples wrong classified in gini and Entropy ', len(gini_wrong),len(entropy_wrong))
3
4
```

```
no of samples Entropy correctly classifies but gini dont : 10
no of samples wrong classified in gini and Entropy  88 107
```

```
:
1  print('no of samples misclaffication correctly classifies but gini dont :', (len(gini_wrong- mis_wrong)))
2  print('no of samples wrong classified in gini and misclassification :', (len(gini_wrong), len(mis_wrong)))
3
```

```
no of samples misclaffication correctly classifies but gini dont : 21
no of samples wrong classified in gini and misclassification : (88, 105)
```


Members:

1.Devanand 2001CS19

2.Teja Vardhan 2001CS26

3.Gnaneshwar 2001CS15