

CS561- ARTIFICIAL INTELLIGENCE LAB

ASSIGNMENT-6: Decision Trees

Intaj Choudhury - 2211MC09

Ankit Anand – 2311MC04

Khushbu Bharti – 2311MC21

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

```
print('Acc: {}, Prec: {}, Rec: {}, Fscore: {}'.format(*get_scores(Y_test_idx, y_pred_test)))
✓ 0.0s
Acc: 0.784, Prec: 0.7975703944602568, Rec: 0.784, Fscore: 0.7852263537777933

precision_recall_fscore_support(Y_test_idx, y_pred_test, average='macro')
✓ 0.0s
(0.8227507784372343, 0.7488553732220372, 0.7763058087746822, None)
```

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

```
Features: frozenset({'pos', 'tri', 'len', 'bi'}), Missing Feature: uni
Acc: 0.762, Prec: 0.8041078055243244, Rec: 0.762, Fscore: 0.7653980349989458

Features: frozenset({'uni', 'pos', 'len', 'bi'}), Missing Feature: tri
Acc: 0.776, Prec: 0.8052230983083352, Rec: 0.776, Fscore: 0.7790498535919935

Features: frozenset({'uni', 'tri', 'pos', 'bi'}), Missing Feature: len
Acc: 0.79, Prec: 0.8058626920363204, Rec: 0.79, Fscore: 0.7877892981525293

Features: frozenset({'uni', 'tri', 'len', 'pos'}), Missing Feature: bi
Acc: 0.766, Prec: 0.802194505921939, Rec: 0.766, Fscore: 0.7684352903409354

Features: frozenset({'uni', 'tri', 'len', 'bi'}), Missing Feature: pos
Acc: 0.782, Prec: 0.8153808058404247, Rec: 0.782, Fscore: 0.7849296877605018

Features: frozenset({'pos', 'len'}), Missing Feature: pos
Acc: 0.646, Prec: 0.7365468507023657, Rec: 0.646, Fscore: 0.6429398022180843
```

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.

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

```
Gain type: entropy
Acc: 0.798, Prec: 0.8190360638144951, Rec: 0.798, Fscore: 0.8022986412927636

Gain type: gini
Acc: 0.818, Prec: 0.8191442574257426, Rec: 0.818, Fscore: 0.8136217953772946

Gain type: misclassification
Acc: 0.782, Prec: 0.7946936773118786, Rec: 0.782, Fscore: 0.7802440665363832

Best gain type is gini for class LOC, LOC-only-f1-score: 0.8148148148148148
Best gain type is gini for class HUM, HUM-only-f1-score: 0.832116788321168
Best gain type is gini for class NUM, NUM-only-f1-score: 0.8878504672897195
Best gain type is entropy for class DESC, DESC-only-f1-score: 0.9257950530035336
Best gain type is misclassification for class ENTY, ENTY-only-f1-score: 0.6994535519125683
Best gain type is entropy for class ABBR, ABBR-only-f1-score: 0.8
```

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.

```
Examples that entropy got wrong,and everything else got right:
Percentage of corrected examples: 62.37623762376238%
```

```
LOC:city What county is Modesto , California in ?
HUM:gr George Bush purchased a small interest in which baseball team ?
LOC:city What city had a world fair in 1900 ?
HUM:ind What person 's head is on a dime ?
NUM:other What is the life expectancy for crickets ?
LOC:other What river in the US is known as the Big Muddy ?
LOC:other What hemisphere is the Philippines in ?
ENTY:plant Material called linen is made from what plant ?
ENTY:animal What is the proper name for a female walrus ?
HUM:gr What is the name of the chocolate company in San Francisco ?
---
```

```
Examples that gini got wrong,and everything else got right:
Percentage of corrected examples: 58.24175824175824%
```

```
ENTY:plant What is Australia 's national flower ?
HUM:ind What person 's head is on a dime ?
NUM:other What is the life expectancy for crickets ?
LOC:other What river in the US is known as the Big Muddy ?
ENTY:plant Material called linen is made from what plant ?
ENTY:animal What is a group of turkeys called ?
HUM:ind What was W.C. Fields ' real name ?
LOC:other What river flows between Fargo , North Dakota and Moorhead , Minnesota ?
LOC:other What strait separates North America from Asia ?
ENTY:other What is the longest major league baseball-winning streak ?
---
```

```
Examples that misclassification got wrong,and everything else got right:
Percentage of corrected examples: 65.13761467889908%
```

```
LOC:city What county is Modesto , California in ?
NUM:dist How tall is the Sears Building ?
ENTY:plant What is Australia 's national flower ?
ENTY:substance What metal has the highest melting point ?
HUM:ind Who developed the vaccination against polio ?
HUM:ind Who developed the Macintosh computer ?
LOC:other What hemisphere is the Philippines in ?
ENTY:animal What is the proper name for a female walrus ?
ENTY:animal What is a group of turkeys called ?
HUM:ind Who discovered x-rays ?
---
```