OUTPUTS AND INFERENCES

In [ ]:

```
Training Data class distribution
[('ABBR', 0.02), ('DESC', 0.21), ('ENTY', 0.23), ('HUM', 0.22), ('LOC', 0.15), ('NUM', 0.16)]

Testing Data class distribution
[('ABBR', 0.02), ('DESC', 0.28), ('ENTY', 0.19), ('HUM', 0.13), ('LOC', 0.16), ('NUM', 0.23)]
```

K-fold Validation

In [ ]:

```
                10-fold Validation using Gini Index



****************Iteration  0 **************

        Confusion Matrix:
             ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR        3.00   1.00   4.00   0.00   2.00   2.00
DESC        1.00  58.00  35.00  11.00   8.00   7.00
ENTY        1.00   9.00  60.00  29.00  11.00  15.00
HUM         1.00  11.00  45.00  40.00   9.00  12.00
LOC         3.00   5.00  30.00  13.00  25.00  10.00
NUM         0.00  11.00  29.00  13.00   4.00  28.00
Precision  33.33  61.05  29.56  37.74  42.37  37.84
Recall     25.00  48.33  48.00  33.90  29.07  32.94
f1_score   28.57  53.95  36.59  35.71  34.48  35.22


****************Iteration  1 **************

        Confusion Matrix:
             ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR        1.00   1.00   0.00   1.00   1.00   2.00
DESC        1.00  61.00  10.00  15.00   6.00  19.00
ENTY        0.00  19.00  49.00  31.00  15.00  32.00
HUM         0.00  18.00  23.00  44.00  14.00  19.00
LOC         0.00  11.00  17.00  12.00  20.00  18.00
NUM         0.00   9.00  17.00  11.00   8.00  41.00
Precision  50.00  51.26  42.24  38.60  31.25  31.30
Recall     16.67  54.46  33.56  37.29  25.64  47.67
f1_score   25.00  52.81  37.40  37.93  28.17  37.79


****************Iteration  2 **************

        Confusion Matrix:
             ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR        1.00   2.00   1.00   0.00   0.00   2.00
DESC        0.00  53.00  21.00  14.00   9.00  11.00
ENTY        0.00  12.00  44.00  32.00  22.00  18.00
HUM         0.00  11.00  21.00  60.00  18.00  20.00
LOC         0.00  12.00  15.00  21.00  26.00  17.00
NUM         1.00  10.00  17.00  15.00   9.00  30.00
Precision  50.00  53.00  36.97  42.25  30.95  30.61
Recall     16.67  49.07  34.38  46.15  28.57  36.59
f1_score   25.00  50.96  35.63  44.12  29.71  33.33


****************Iteration  3 **************

        Confusion Matrix:
              ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR         1.00   3.00   2.00   0.00   2.00   2.00
DESC         0.00  51.00  20.00  31.00   9.00  16.00
ENTY         0.00   4.00  62.00  35.00   6.00  23.00
```

```
HUM          0.00   8.00  27.00  47.00   8.00  22.00
LOC          0.00   5.00  29.00  15.00  16.00  17.00
NUM          0.00  13.00  15.00   5.00  15.00  36.00
Precision  100.00  60.71  40.00  35.34  28.57  31.03
Recall      10.00  40.16  47.69  41.96  19.51  42.86
f1_score    18.18  48.34  43.51  38.37  23.19  36.00


****************Iteration  4 **************

        Confusion Matrix:
             ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR        2.00   2.00   2.00   0.00   1.00   2.00
DESC        1.00  60.00  10.00  21.00   6.00  27.00
ENTY        0.00  20.00  43.00  32.00  10.00  20.00
HUM         0.00  12.00  17.00  49.00  11.00  24.00
LOC         1.00   7.00  15.00  25.00  19.00  21.00
NUM         0.00   7.00  10.00  19.00  11.00  38.00
Precision  50.00  55.56  44.33  33.56  32.76  28.79
Recall     22.22  48.00  34.40  43.36  21.59  44.71
f1_score   30.77  51.50  38.74  37.84  26.03  35.02


****************Iteration  5 **************

        Confusion Matrix:
            ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR        0.0   4.00   1.00   0.00   0.00   0.00
DESC        3.0  52.00  32.00  11.00   7.00  14.00
ENTY        0.0  10.00  44.00  33.00   6.00  19.00
HUM         0.0  11.00  46.00  51.00  11.00  16.00
LOC         0.0   9.00  26.00  20.00  10.00  14.00
NUM         0.0   6.00  33.00  16.00   2.00  38.00
Precision   0.0  56.52  24.18  38.93  27.78  37.62
Recall      0.0  43.70  39.29  37.78  12.66  40.00
f1_score    0.0  49.29  29.93  38.35  17.39  38.78


****************Iteration  6 **************

        Confusion Matrix:
              ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR         4.00   2.00   6.00   0.00   0.00   0.00
DESC         0.00  58.00  22.00  16.00   2.00  12.00
ENTY         0.00  19.00  50.00  28.00   3.00  26.00
HUM          0.00  11.00  32.00  65.00   3.00   8.00
LOC          0.00   7.00  30.00  18.00  16.00  13.00
NUM          0.00  13.00  30.00  22.00   2.00  27.00
Precision  100.00  52.73  29.41  43.62  61.54  31.40
Recall      33.33  52.73  39.68  54.62  19.05  28.72
f1_score    50.00  52.73  33.78  48.51  29.09  30.00


****************Iteration  7 **************

        Confusion Matrix:
             ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR        3.00   6.00   0.00   1.00   1.00   0.00
DESC        0.00  51.00  25.00  14.00   4.00  14.00
ENTY        0.00  14.00  45.00  42.00  10.00  14.00
HUM         0.00   8.00  28.00  62.00  11.00  12.00
LOC         0.00  20.00  13.00  20.00  20.00   8.00
NUM         1.00  13.00  20.00  22.00   9.00  34.00
Precision  75.00  45.54  34.35  38.51  36.36  41.46
Recall     27.27  47.22  36.00  51.24  24.69  34.34
f1_score   40.00  46.36  35.16  43.97  29.41  37.57


****************Iteration  8 **************

        Confusion Matrix:
             ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR         1.0   1.00   2.00   0.00   0.00   0.00
DESC         0.0  53.00  36.00  14.00   4.00  12.00
ENTY         0.0   9.00  58.00  27.00   5.00  20.00
HUM          0.0  10.00  47.00  60.00   6.00  12.00
LOC          0.0  11.00  31.00  11.00  11.00  16.00
NUM          0.0   9.00  34.00   9.00   2.00  34.00
Precision  100.0  56.99  27.88  49.59  39.29  36.17
Recall      25.0  44.54  48.74  44.44  13.75  38.64
f1_score    40.0  50.00  35.47  46.88  20.37  37.36


****************Iteration  9 **************

        Confusion Matrix:
             ABBR   DESC   ENTY    HUM    LOC    NUM
ABBR        4.00   3.00   0.00   1.00   1.00   2.00
DESC        1.00  45.00  16.00  24.00   7.00  21.00
ENTY        1.00   6.00  34.00  39.00   7.00  27.00
HUM         0.00   9.00  24.00  65.00   9.00  15.00
LOC         1.00   5.00  19.00  23.00  20.00  18.00
NUM         1.00   5.00  14.00  30.00   5.00  43.00
Precision  50.00  61.64  31.78  35.71  40.82  34.13
Recall     36.36  39.47  29.82  53.28  23.26  43.88
f1_score   42.11  48.13  30.77  42.76  29.63  38.39




Accuracy of 10-fold Cross validation are:

[39.19, 39.56, 39.27, 39.08, 38.72, 35.78, 40.37, 39.45, 39.82, 38.72]

Average ~ 39 %
```

For all three models over different feature combinations

In [ ]:

```
*******************************For  gini index****************************


Considering feature set 0 :
['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram', 'Lex_Prob_Scores_3gram', 'Syntactic_Prob_Scores']

        Accuracy: 51.0

        Confusion Matrix:
             ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR        3.00    2.00   3.00   0.00   0.00   1.00
DESC        1.00  125.00   2.00   5.00   0.00   5.00
ENTY        0.00   26.00  27.00  17.00   7.00  17.00
HUM         0.00    5.00   3.00  37.00  10.00  10.00
LOC         0.00   15.00  16.00  16.00  26.00   8.00
NUM         0.00   31.00  16.00  13.00  16.00  37.00
Precision  75.00   61.27  40.30  42.05  44.07  47.44
Recall     33.33   90.58  28.72  56.92  32.10  32.74
f1_score   46.15   73.10  33.54  48.37  37.14  38.74


Considering feature set 1 :
    ['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram', 'Lex_Prob_Scores_3gram']

        Accuracy: 48.2

        Confusion Matrix:
             ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR        3.00    1.00   3.00   0.00   1.00   1.00
DESC        1.00  129.00   1.00   1.00   0.00   6.00
ENTY        0.00   26.00  30.00  13.00   8.00  17.00
HUM         1.00   11.00   8.00  24.00  10.00  11.00
```

```
LOC          4.00   13.00  19.00  17.00  22.00   6.00
NUM          0.00   29.00  23.00  13.00  15.00  33.00
Precision   33.33   61.72  35.71  35.29  39.29  44.59
Recall      33.33   93.48  31.91  36.92  27.16  29.20
f1_score    33.33   74.35  33.71  36.09  32.12  35.29


Considering feature set 2 : ['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram']

        Accuracy: 43.8

        Confusion Matrix:
             ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR         3.00    3.00   2.00   0.00   0.00   1.00
DESC         1.00  123.00   1.00   6.00   1.00   6.00
ENTY         0.00   21.00  25.00  16.00   6.00  26.00
HUM          1.00    9.00  15.00  19.00  14.00   7.00
LOC          4.00   17.00  16.00  13.00  15.00  16.00
NUM          1.00   22.00  31.00  12.00  13.00  34.00
Precision   30.00   63.08  27.78  28.79  30.61  37.78
Recall      33.33   89.13  26.60  29.23  18.52  30.09
f1_score    31.58   73.87  27.17  29.01  23.08  33.50


Considering feature set 3 : ['Length', 'Lex_Prob_Scores_1gram']

        Accuracy: 36.2

        Confusion Matrix:
             ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR         3.00    2.00   1.00   1.00   0.00   2.00
DESC         2.00  110.00   4.00  21.00   0.00   1.00
ENTY         0.00   21.00  40.00  19.00   2.00  12.00
HUM          3.00   16.00  18.00  12.00   7.00   9.00
LOC          6.00   17.00  31.00  15.00   3.00   9.00
NUM          4.00   21.00  50.00  19.00   6.00  13.00
Precision   16.67   58.82  27.78  13.79  16.67  28.26
Recall      33.33   79.71  42.55  18.46   3.70  11.50
f1_score    22.22   67.69  33.61  15.79   6.06  16.35


*******************************For  misclsfn_err index****************************


Considering feature set 0 :
['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram', 'Lex_Prob_Scores_3gram', 'Syntactic_Prob_Scores']

        Accuracy: 44.8

        Confusion Matrix:
            ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR         0.0    4.00   3.00   1.00   0.00   1.00
DESC         0.0  129.00   4.00   4.00   0.00   1.00
ENTY         0.0   28.00  30.00  21.00   6.00   9.00
HUM          0.0   13.00  13.00  30.00   2.00   7.00
LOC          1.0   22.00  23.00  16.00  12.00   7.00
NUM          0.0   26.00  51.00  10.00   3.00  23.00
Precision    0.0   58.11  24.19  36.59  52.17  47.92
Recall       0.0   93.48  31.91  46.15  14.81  20.35
f1_score     0.0   71.67  27.52  40.82  23.08  28.57


Considering feature set 1 :
    ['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram', 'Lex_Prob_Scores_3gram']

        Accuracy: 41.6

        Confusion Matrix:
            ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR         0.0    4.00   3.00   1.00   0.00   1.00
DESC         0.0  129.00   5.00   3.00   0.00   1.00
```

```
ENTY        0.0   29.00  34.00  17.00   4.00  10.00
HUM         0.0   15.00  20.00  21.00   2.00   7.00
LOC         0.0   24.00  24.00  17.00   7.00   9.00
NUM         0.0   29.00  52.00  12.00   3.00  17.00
Precision   0.0   56.09  24.64  29.58  43.75  37.78
Recall      0.0   93.48  36.17  32.31   8.64  15.04
f1_score    0.0   70.11  29.31  30.88  14.43  21.52


Considering feature set 2 : ['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram']

        Accuracy: 42.8

        Confusion Matrix:
            ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR        0.0    5.00   3.00   0.00   0.00   1.00
DESC        0.0  131.00   1.00   5.00   0.00   1.00
ENTY        0.0   27.00  31.00  12.00   2.00  22.00
HUM         0.0   17.00  16.00  20.00   5.00   7.00
LOC         0.0   28.00  19.00  14.00   8.00  12.00
NUM         0.0   34.00  40.00  13.00   2.00  24.00
Precision   0.0   54.13  28.18  31.25  47.06  35.82
Recall      0.0   94.93  32.98  30.77   9.88  21.24
f1_score    0.0   68.95  30.39  31.01  16.33  26.67


Considering feature set 3 : ['Length', 'Lex_Prob_Scores_1gram']

        Accuracy: 41.4

        Confusion Matrix:
            ABBR    DESC   ENTY    HUM   LOC    NUM
ABBR        0.0    4.00   5.00   0.00  0.00   0.00
DESC        0.0  127.00  10.00   1.00  0.00   0.00
ENTY        0.0   13.00  67.00   9.00  2.00   3.00
HUM         0.0   20.00  23.00   8.00  6.00   8.00
LOC         0.0   22.00  44.00   5.00  1.00   9.00
NUM         0.0   16.00  84.00   7.00  2.00   4.00
Precision   0.0   62.87  28.76  26.67  9.09  16.67
Recall      0.0   92.03  71.28  12.31  1.23   3.54
f1_score    0.0   74.71  40.98  16.84  2.17   5.84


*******************************For  cross_entropy index****************************


Considering feature set 0 :
['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram', 'Lex_Prob_Scores_3gram', 'Syntactic_Prob_Scores']

        Accuracy: 50.4

        Confusion Matrix:
             ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR        3.00    2.00   2.00   0.00   1.00   1.00
DESC        0.00  126.00   4.00   3.00   0.00   5.00
ENTY        0.00   26.00  31.00  10.00   8.00  19.00
HUM         1.00    3.00   8.00  29.00  10.00  14.00
LOC         0.00   15.00  15.00  15.00  27.00   9.00
NUM         2.00   31.00  21.00  10.00  13.00  36.00
Precision  50.00   62.07  38.27  43.28  45.76  42.86
Recall     33.33   91.30  32.98  44.62  33.33  31.86
f1_score   40.00   73.90  35.43  43.94  38.57  36.55


Considering feature set 1 :
    ['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram', 'Lex_Prob_Scores_3gram']

        Accuracy: 47.6

        Confusion Matrix:
             ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR        4.00    1.00   2.00   0.00   1.00   1.00
DESC        1.00  129.00   2.00   1.00   0.00   5.00
ENTY        2.00   24.00  27.00  15.00   6.00  20.00
HUM         1.00    9.00   9.00  23.00   9.00  14.00
LOC         1.00   15.00  16.00  15.00  24.00  10.00
NUM         1.00   35.00  23.00  13.00  10.00  31.00
Precision  40.00   60.56  34.18  34.33  48.00  38.27
Recall     44.44   93.48  28.72  35.38  29.63  27.43
f1_score   42.11   73.50  31.21  34.85  36.64  31.96


Considering feature set 2 : ['Length', 'Lex_Prob_Scores_1gram', 'Lex_Prob_Scores_2gram']

        Accuracy: 46.2

        Confusion Matrix:
             ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR        3.00    3.00   1.00   0.00   1.00   1.00
DESC        1.00  125.00   3.00   5.00   2.00   2.00
ENTY        0.00   19.00  39.00  12.00   7.00  17.00
HUM         1.00    8.00  20.00  20.00  10.00   6.00
LOC         4.00   15.00  17.00  18.00  18.00   9.00
NUM         1.00   26.00  34.00  12.00  14.00  26.00
Precision  30.00   63.78  34.21  29.85  34.62  42.62
Recall     33.33   90.58  41.49  30.77  22.22  23.01
f1_score   31.58   74.85  37.50  30.30  27.07  29.89


Considering feature set 3 : ['Length', 'Lex_Prob_Scores_1gram']

        Accuracy: 37.6

        Confusion Matrix:
             ABBR    DESC   ENTY    HUM    LOC    NUM
ABBR        3.00    2.00   2.00   0.00   0.00   2.00
DESC        2.00  111.00   6.00  18.00   0.00   1.00
ENTY        0.00   20.00  48.00  15.00   1.00  10.00
HUM         3.00   22.00  15.00  10.00   5.00  10.00
LOC         6.00   19.00  33.00  12.00   3.00   8.00
NUM         4.00   18.00  58.00  17.00   3.00  13.00
Precision  16.67   57.81  29.63  13.89  25.00  29.55
Recall     33.33   80.43  51.06  15.38   3.70  11.50
f1_score   22.22   67.27  37.50  14.60   6.45  16.56
```

In [ ]:

```
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.






.
```