Testing Default Parameters

LogisticRegression scored 0.7424452417652164

Best parameter (CV score=0.739):

{}

precision recall f1-score support

1 0.66 0.71 0.68 7066

2 0.40 0.20 0.26 4096

3 0.45 0.32 0.37 5845

4 0.49 0.26 0.33 10919

5 0.81 0.95 0.88 49277

accuracy 0.74 77203

macro avg 0.56 0.49 0.51 77203

weighted avg 0.70 0.74 0.71 77203

Cross validation with unseen test data

Model score on unseen data 0.6766680270768538

precision recall f1-score support

1 0.77 0.34 0.47 13075

2 0.45 0.06 0.10 7416

3 0.56 0.07 0.12 10647

4 0.65 0.01 0.02 20346

5 0.68 1.00 0.81 90630

accuracy 0.68 142114

macro avg 0.62 0.29 0.30 142114

weighted avg 0.66 0.68 0.57 142114

Testing with default parameters with weighted labels and C= 1 (Values obtained from GridSearch)

LogisticRegression scored 0.7410333795318835

Best parameter (CV score=0.739):

{'classifier\_\_C': 1}

precision recall f1-score support

1 0.65 0.71 0.68 7105

2 0.41 0.19 0.26 4090

3 0.44 0.31 0.36 5838

4 0.49 0.25 0.33 10963

5 0.81 0.95 0.87 49207

accuracy 0.74 77203

macro avg 0.56 0.48 0.50 77203

weighted avg 0.70 0.74 0.71 77203

Cross validation with unseen test data

Model score on unseen data 0.6784201415764809

precision recall f1-score support

1 0.77 0.35 0.48 13075

2 0.41 0.07 0.11 7416

3 0.52 0.07 0.13 10647

4 0.73 0.01 0.02 20346

5 0.68 1.00 0.81 90630

accuracy 0.68 142114

macro avg 0.63 0.30 0.31 142114

weighted avg 0.67 0.68 0.58 142114

Testing with Default parameters with no Stop Words removal (GridSearch for ngrams 1,2,3)

LogisticRegression scored 0.7571978086024013

Best parameter (CV score=0.757):

{'cv\_\_ngram\_range': (1, 2)}

precision recall f1-score support

1 0.68 0.74 0.71 7017

2 0.47 0.16 0.23 4106

3 0.49 0.37 0.42 5873

4 0.55 0.26 0.35 11078

5 0.81 0.97 0.88 49137

accuracy 0.76 77211

macro avg 0.60 0.50 0.52 77211

weighted avg 0.72 0.76 0.72 77211

Cross validation with unseen test data

Model score on unseen data 0.762901614197053

precision recall f1-score support

1 0.72 0.73 0.72 13075

2 0.58 0.17 0.27 7416

3 0.55 0.37 0.44 10647

4 0.58 0.25 0.35 20346

5 0.80 0.98 0.88 90630

accuracy 0.76 142114

macro avg 0.65 0.50 0.53 142114

weighted avg 0.73 0.76 0.72 142114