

MODEL TEST RESULTS

1 LinearSVC

Feature extraction parameters:

- max features: 8000
- ngram rang:(1,3)
- stop words: None
- max df: 1.0
- min df: 1
- norm: l2

c	train accuracy	test accuracy
1.0	76.49	66.55
0.8	76.42	66.64
0.6	76.24	66.93
0.4	76.03	67.24
0.2	75.26	67.69
0.05	73.14	68.11
0.001	66.22	64.52

Table 1: tol=0.001, penalty=l2

c	train accuracy	test accuracy
1.0	76.49	66.53
0.8	76.42	66.62
0.6	76.24	66.94
0.4	76.03	67.24
0.2	75.26	67.69

Table 2: tol=0.004, penalty=l2

c	train accuracy	test accuracy
1.0	76.50	66.54
0.8	76.42	66.64
0.6	76.24	66.94
0.4	76.03	67.24
0.2	75.26	67.70

Table 3: tol=0.007, penalty=l2

2 LogisticRegression

Feature extraction parameters:

- max features: 8000
- ngram rang:(1,3)
- stop words: None
- max df: 1.0
- min df: 1
- norm: l2

c	train accuracy	test accuracy
1.0	74.58	68.03
0.6	73.69	68.34
0.2	71.26	67.59

Table 4: tol=0.0001, penalty=l2

c	train accuracy	test accuracy
1.0	74.58	68.03
0.6	73.69	68.34
0.2	71.38	67.59

Table 5: tol=0.0006, penalty=l2

c	train accuracy	test accuracy
1.0	71.50	66.29
0.6	70.41	65.80
0.2	68.46	64.91

Table 6: tol=0.0011, penalty=l2

c	train accuracy	test accuracy
1.0	67.40	65.09
0.6	65.19	63.78
0.2	60.61	60.31

Table 7: tol=0.0001, penalty=l1

c	train accuracy	test accuracy
1.0	67.40	65.09
0.6	65.19	63.78
0.2	60.62	60.30

Table 8: tol=0.0006, penalty=l1

c	train accuracy	test accuracy
1.0	67.41	65.09
0.6	65.20	63.79
0.2	60.61	60.31

Table 9: tol=0.0011, penalty=l1

3 GaussianNB

Feature extraction parameters:

- max features: 8000
- ngram rang:(1,3)
- stop words: None
- max df: 1.0
- min df: 1
- norm: l2

varsmoothing	train accuracy	test accuracy
1e-9	71.15	63.94
1e-8	71.28	63.94
1e-7	71.38	63.93

4 RandomForestClassifier

Feature extraction parameters:

- max features: 15000
- ngram rang:(1,3)
- stop words: None
- max df: 1.0
- min df: 1
- norm: l1

4.1 1

RandomForestClassifier parameters:

- oob score: True
- n estimators: 10

Results:

- train score: 96.72
- test score: 64.54

4.2 2

RandomForestClassifier parameters:

- oob score: True
- n estimators: 30
- max depth:10

Results:

- train score: 98.18
- test score: 66.04