

Examples of bags of words without normalization (raw counts per document):

```
[[0 0 1 1 2 1 0 1 0 0 0 0 0 1 0 0 1 0 0]
 [0 1 0 0 0 2 1 0 1 0 0 1 1 1 1 0 0 1 0]
 [2 0 0 0 0 2 1 0 1 1 2 1 0 1 0 1 2 0 1]
 [0 1 1 1 0 0 1 1 0 0 0 0 0 0 1 1 1 1 0]
 [0 0 0 0 0 1 2 0 2 1 1 0 1 1 0 0 0 0 1]]
```

The same examples as above but with counts clipped to 1:

```
[[0 1 1 1 0 1 0 0 0 0 0 1 0 0 1 0 0]
 [1 0 0 1 1 0 1 0 0 1 1 1 1 0 0 1 0]
 [0 0 0 1 1 0 1 1 1 1 0 1 0 1 1 0 1]
 [1 1 1 0 1 1 0 0 0 0 0 0 1 1 1 1 0]
 [0 0 0 1 1 0 1 1 1 0 1 1 0 0 0 0 1]]
```

The same examples as above but with TF-IDF normalization:

```
[[0. 0. 0.3 0.3 0.8 0.2 0. 0.3 0. 0. 0. 0. 0.2 0. 0. 0.3 0.
 0. ]
 [0. 0.3 0. 0. 0. 0.5 0.2 0. 0.3 0. 0. 0.3 0.3 0.2 0.3 0. 0. 0.3
 0. ]
 [0.5 0. 0. 0. 0. 0.3 0.2 0. 0.2 0.2 0.4 0.2 0. 0.2 0. 0.2 0.4 0.
 0.2]
 [0. 0.3 0.3 0.3 0. 0. 0.2 0.3 0. 0. 0. 0. 0. 0. 0.3 0.3 0.3 0.3
 0. ]
 [0. 0. 0. 0. 0. 0.2 0.4 0. 0.5 0.3 0.3 0. 0.3 0.2 0. 0. 0. 0.
 0.3]]
```

test

	recall_macro	f1_pos_neg	accuracy	recall_neg
0	0.620937	0.602270	0.618093	0.611305
1	0.615800	0.599981	0.618009	0.588691
2	0.615914	0.595471	0.617768	0.592334
mean	0.617550	0.599241	0.617957	0.597443
std	0.002933	0.003459	0.000169	0.012142

test

	recall_macro	f1_pos_neg	accuracy	recall_neg
0	0.641287	0.619628	0.647080	0.613240
1	0.631539	0.610504	0.645126	0.575136
2	0.637179	0.611158	0.646998	0.596980
mean	0.636669	0.613763	0.646402	0.595119
std	0.004894	0.005089	0.001105	0.019121

test				
	recall_macro	f1_pos_neg	accuracy	recall_neg
0	0.656339	0.625988	0.653119	0.662408
1	0.656458	0.627115	0.655574	0.655306
2	0.658039	0.624048	0.656480	0.661634
mean	0.656945	0.625717	0.655058	0.659783
std	0.000949	0.001552	0.001739	0.003896

test				
	recall_macro	f1_pos_neg	accuracy	recall_neg
0	0.674462	0.645385	0.652455	0.745645
1	0.671421	0.640748	0.650079	0.742835
2	0.674438	0.644842	0.652857	0.745257
mean	0.673440	0.643659	0.651797	0.744579
std	0.001749	0.002535	0.001501	0.001523

0	0.001000	0.696250	0.676795	0.002522
1	0.010000	0.702407	0.673495	0.002558
2	0.000100	0.679654	0.670958	0.003752
3	1.000000	0.702548	0.670369	0.002391
4	0.100000	0.702850	0.670230	0.002478
5	None	nan	nan	nan

100%|██████████| 384/384 [00:36<00:00, 10.61it/s]

Total number of examples for testing: 12284

Distribution of classes:

neutral 0.483312

negative 0.323347

positive 0.193341

Name: label, dtype: float64

100%|██████████| 384/384 [00:36<00:00, 10.61it/s]

Total number of examples for testing: 12284

Distribution of classes:

neutral 0.483312

negative 0.323347

positive 0.193341

Name: label, dtype: float64

	precision	recall	f1-score	support
negative	0.58	0.81	0.68	3972
neutral	0.70	0.57	0.63	5937
positive	0.65	0.55	0.60	2375
accuracy			0.64	12284
macro avg	0.65	0.64	0.63	12284
weighted avg	0.66	0.64	0.64	12284