

Дорожкин Денис 425 группа

Отчет

Предобработка:

Перед обучением весь текст был обработан

1. Весь текст был преобразован в нижний регистр.
2. Были выкинуты лишние спецсимволы
3. Выкинуты лишние пробелы

Далее использовались различные методы классификации и выявления признаков. Далее приведено перечисление методов с их количественными показателями.

Binary + Multinomial Naive Bayes ngram_range=(1, 2) min_df = 1

precision recall f1-score support

+	0.64	0.47	0.54	2081
-	0.49	0.90	0.64	2173
0	1.00	0.00	0.00	1235

avg / total 0.66 0.53 0.46 5489

TF-IDF + Multinomial Naive Bayes ngram_range=(1, 2) min_df = 1

precision recall f1-score support

+	0.60	0.58	0.59	2081
-	0.52	0.84	0.64	2173
0	0.60	0.00	0.00	1235

avg / total 0.57 0.55 0.48 5489

TF-IDF + Multinomial Naive Bayes ngram_range=(1, 5) min_df = 5

precision recall f1-score support

+	0.54	0.70	0.61	2081
-	0.56	0.69	0.62	2173
0	0.53	0.05	0.09	1235

avg / total 0.55 0.55 0.50 5489

TF-IDF + Bernoulli Naive Bayes

precision recall f1-score support

+	0.63	0.52	0.57	2081
-	0.50	0.86	0.63	2173
0	0.51	0.02	0.03	1235

avg / total 0.55 0.54 0.47 5489

Binary + Linear classifiers with SGD training

precision recall f1-score support

+	0.53	0.67	0.59	2081
-	0.56	0.70	0.62	2173
0	0.59	0.07	0.13	1235

avg / total	0.56	0.55	0.50	5489
-------------	------	------	------	------

TF-IDF + Linear classifiers with SGD training

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

+	0.54	0.71	0.62	2081
-	0.58	0.69	0.63	2173
0	0.58	0.08	0.14	1235

avg / total	0.57	0.56	0.52	5489
-------------	------	------	------	------

TF-IDF + Logistic Regression

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

+	0.54	0.65	0.59	2081
-	0.61	0.61	0.61	2173
0	0.41	0.26	0.32	1235

avg / total	0.54	0.55	0.54	5489
-------------	------	------	------	------

Binary + Logistic Regression

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

+	0.54	0.66	0.60	2081
-	0.61	0.62	0.61	2173
0	0.42	0.26	0.32	1235

avg / total	0.54	0.55	0.54	5489
-------------	------	------	------	------