

Supplementary File 2

We trained five different classifiers (k-nearest neighbours, random-forest, SVM, gradient boosted trees, logistic regression) for each tweet category (i.e. content and sentiment) using grid-search for parameter tuning.

Model performance metrics for classification of content and sentiment.

1. NGP-related

Classifier	precision	recall	ROC-AUC	F1
SVM	0.86	0.76	0.81	0.81
Random forest	0.90	0.72	0.81	0.80
K-NN	0.80	0.64	0.73	0.71
Logistic-regression	0.81	0.68	0.75	0.73
Xg-boost	0.80	0.68	0.75	0.73

2. FCTC-related

Classifier	precision	recall	ROC-AUC	F1
SVM	0.50	0.33	0.56	0.40
Random forest	0.66	0.26	0.59	0.38
K-NN	0.50	0.33	0.56	0.40
Logistic-regression	0.50	0.26	0.55	0.34
Xg-boost	0.66	0.53	0.68	0.60

3. TC-related

Classifier	precision	recall	ROC-AUC	F1
SVM	0.46	0.77	0.78	0.58
Random forest	0.83	0.55	0.76	0.66
K-NN	0.23	0.66	0.57	0.34
Logistic - regression	0.50	0.88	0.84	0.64
Xg-boost	0.62	0.55	0.73	0.58

4. Argument

Classifier	precision	recall	ROC-AUC	F1
SVM	0.57	0.50	0.61	0.53
Random forest	0.64	0.45	0.64	0.53
K-NN	0.42	0.25	0.50	0.31
Logistic-regression	0.68	0.54	0.68	0.60
Xg-boost	0.57	0.66	0.65	0.61

5. Critical

Classifier	precision	recall	ROC-AUC	F1
SVM	0.81	0.68	0.80	0.74
Random forest	0.82	0.73	0.82	0.77
K-NN	0.34	0.42	0.51	0.38
Logistic-regression	0.50	0.36	0.59	0.42
Xg-boost	0.46	0.36	0.58	0.41

6. Informative/Neutral

Classifier	precision	recall	ROC-AUC	F1
SVM	0.38	0.62	0.71	0.62
Random forest	0.83	0.62	0.80	0.71
K-NN	0.25	0.75	0.65	0.37
Logistic-regression	0.35	0.75	0.73	0.48
Xg-boost	0.60	0.37	0.66	0.46