
Delivery 2 - Text Classification

CSE4095 - Special Topics in Comp. Eng. III

— Ali Reza Ibrahimzada
Muhammed Fatih Öztel
Basil Osama Hamid Elkhalfi

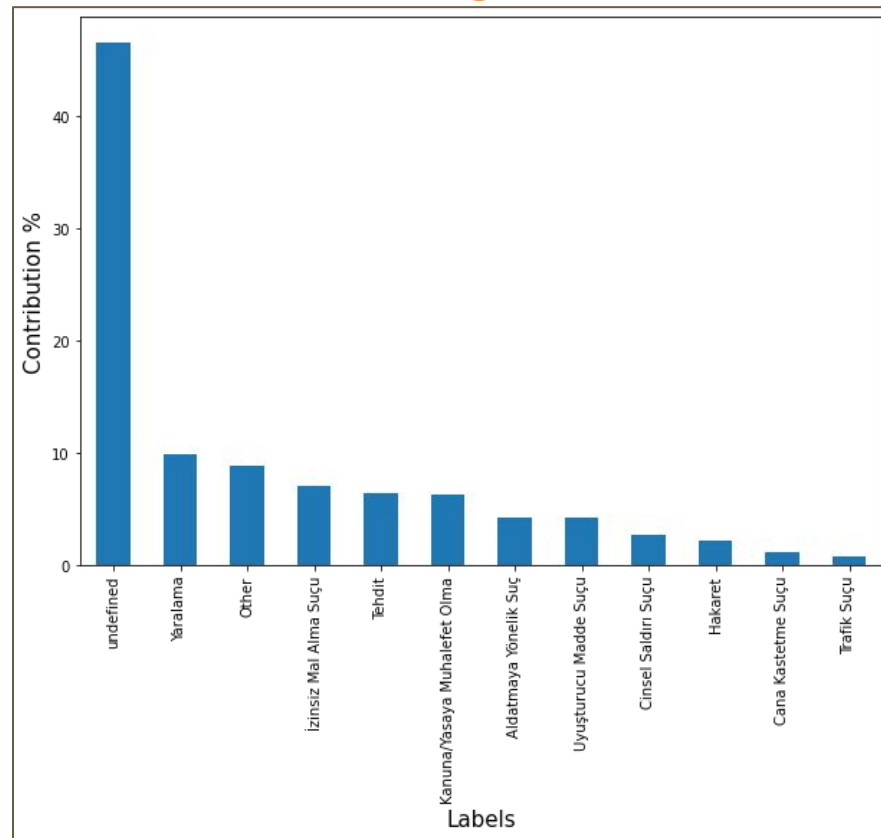
Huzeyfe Ayaz
Yunus Ahmed Stahlschmidt
Sameeh Kunbargi

Content

- Data Cleaning and Preprocessing
- Logistic Regression
- Support Vector Machines
- Multinomial Naive Bayes
- Random Forest
- AdaBoost
- Majority Voting
- Comparison

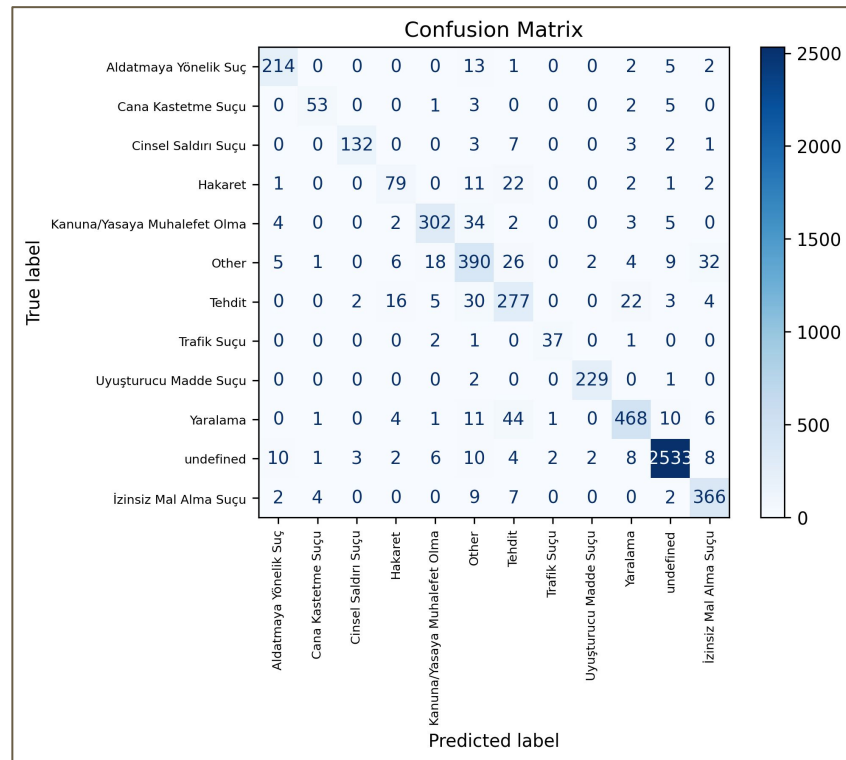
Data Cleaning and Preprocessing

- Cleaned Data from Delivery 1
- Less Frequent Labels in “Other”
- Empty Labels in “Undefined”
- Training (80%) and Test (20%) in Stratified Manner
- 10-Fold Stratified Cross-Validation



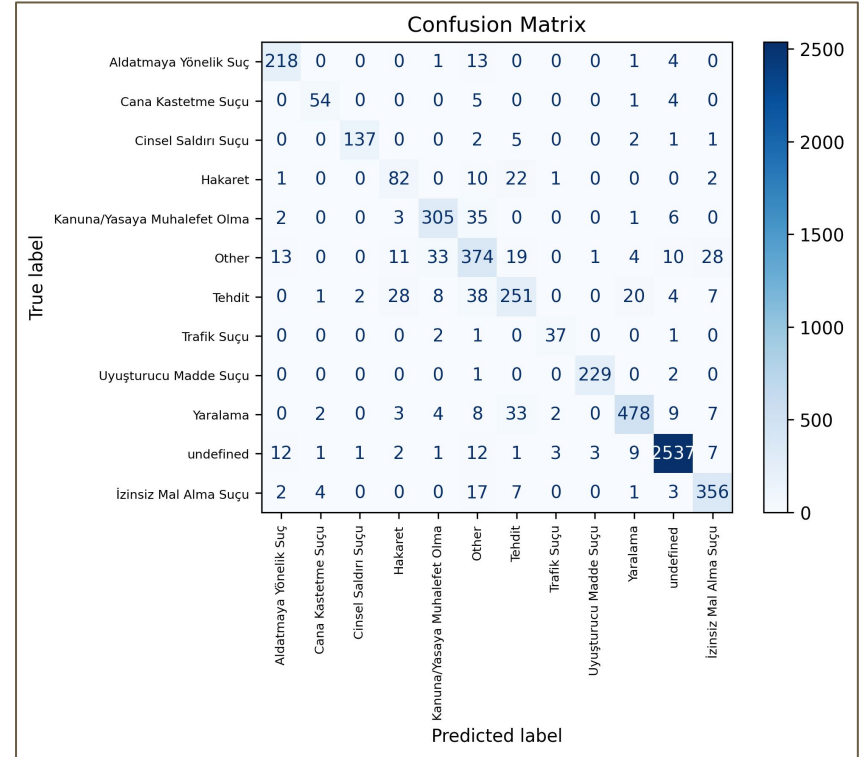
Logistic Regression

	precision	recall	f1-score	support
Aldatmaya Yönelik Suç	0.91	0.90	0.90	237
Cana Kastetme Suçu	0.88	0.83	0.85	64
Cinsel Saldırı Suçu	0.96	0.89	0.93	148
Hakaret	0.72	0.67	0.70	118
Kanuna/Yasaya Muhalefet Olma	0.90	0.86	0.88	352
Other	0.75	0.79	0.77	493
Tehdit	0.71	0.77	0.74	359
Trafik Suçu	0.93	0.90	0.91	41
Uyuşturucu Madde Suçu	0.98	0.99	0.98	232
Yaralama	0.91	0.86	0.88	546
undefined	0.98	0.98	0.98	2589
İzinsiz Mal Alma Suçu	0.87	0.94	0.90	390
accuracy			0.91	5569
macro avg	0.88	0.86	0.87	5569
weighted avg	0.91	0.91	0.91	5569



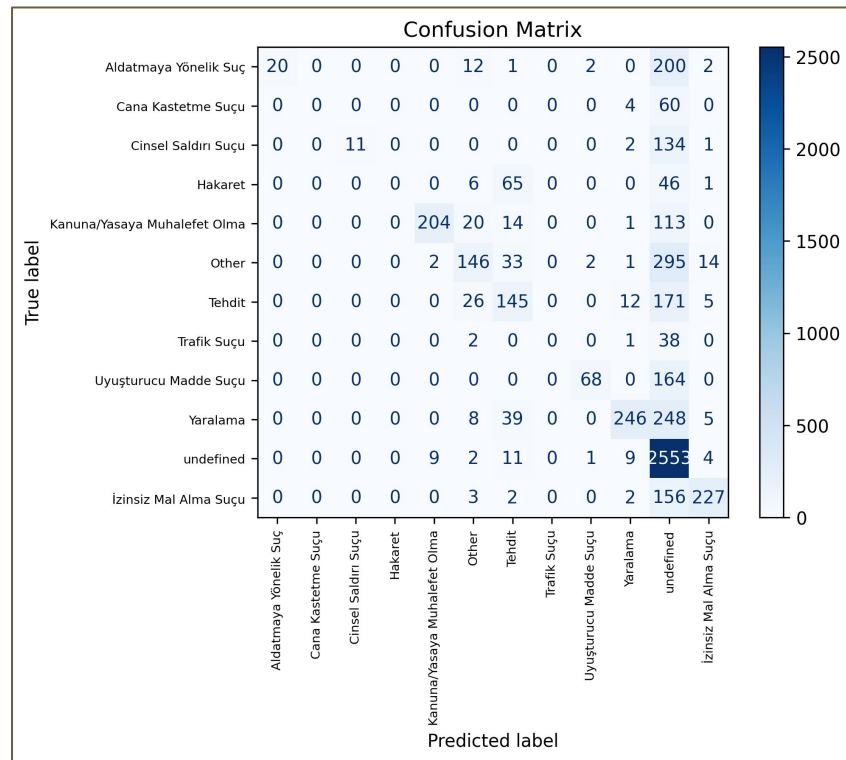
Support Vector Machine

	precision	recall	f1-score	support
Aldatmaya Yönelik Suç	0.88	0.92	0.90	237
Cana Kastetme Suçu	0.87	0.84	0.86	64
Cinsel Saldırı Suçu	0.98	0.93	0.95	148
Hakaret	0.64	0.69	0.66	118
Kanuna/Yasaya Muhalefet Olma	0.86	0.87	0.86	352
Other	0.72	0.76	0.74	493
Tehdit	0.74	0.70	0.72	359
Trafik Suçu	0.86	0.90	0.88	41
Uyuşturucu Madde Suçu	0.98	0.99	0.98	232
Yaralama	0.92	0.88	0.90	546
undefined	0.98	0.98	0.98	2589
İzinsiz Mal Alma Suçu	0.87	0.91	0.89	390
accuracy			0.91	5569
macro avg	0.86	0.86	0.86	5569
weighted avg	0.91	0.91	0.91	5569



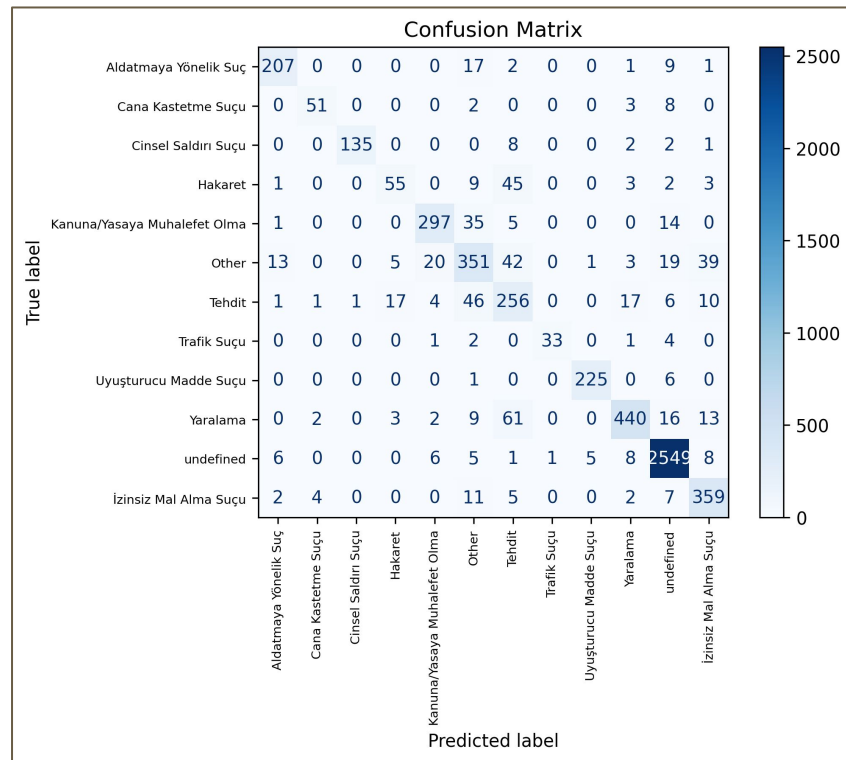
Multinomial Naive Bayes

	precision	recall	f1-score	support
Aldatmaya Yönelik Suç	1.00	0.08	0.16	237
Cana Kastetme Suçu	0.00	0.00	0.00	64
Cinsel Saldırı Suçu	1.00	0.07	0.14	148
Hakaret	0.00	0.00	0.00	118
Kanuna/Yasaya Muhalefet Olma	0.95	0.58	0.72	352
Other	0.65	0.30	0.41	493
Tehdit	0.47	0.40	0.43	359
Trafik Suçu	0.00	0.00	0.00	41
Uyuşturucu Madde Suçu	0.93	0.29	0.45	232
Yaralama	0.88	0.45	0.60	546
undefined	0.61	0.99	0.75	2589
İzinsiz Mal Alma Suçu	0.88	0.58	0.70	390
accuracy			0.65	5569
macro avg	0.61	0.31	0.36	5569
weighted avg	0.69	0.65	0.60	5569



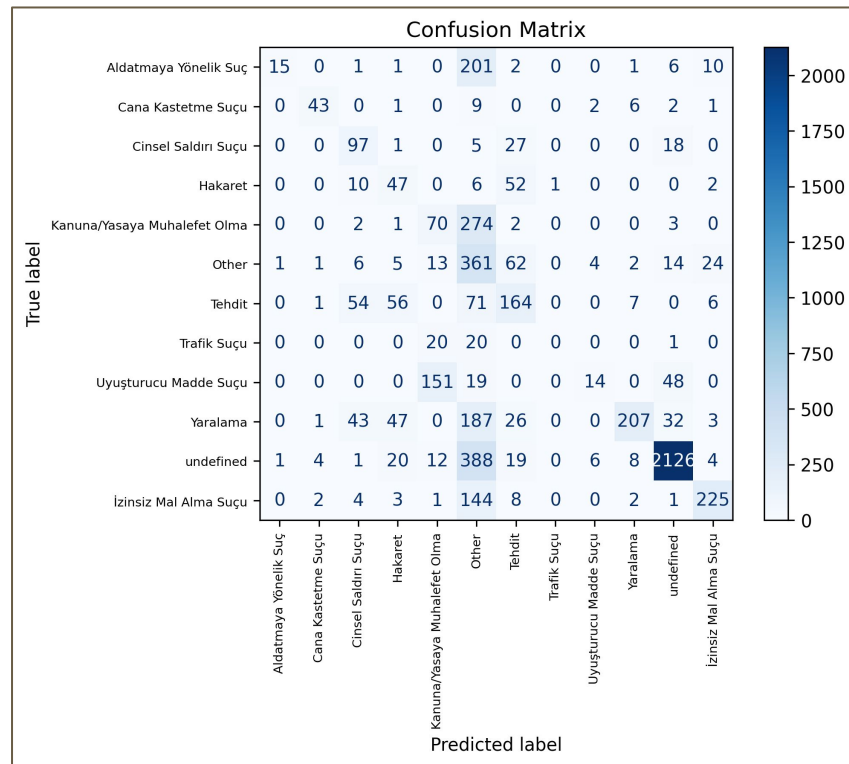
Random Forest

	precision	recall	f1-score	support
Aldatmaya Yönelik Suç	0.90	0.87	0.88	237
Cana Kastetme Suçu	0.88	0.80	0.84	64
Cinsel Saldırı Suçu	0.99	0.91	0.95	148
Hakaret	0.69	0.47	0.56	118
Kanuna/Yasaya Muhalefet Olma	0.90	0.84	0.87	352
Other	0.72	0.71	0.72	493
Tehdit	0.60	0.71	0.65	359
Trafik Suçu	0.97	0.80	0.88	41
Uyuşturucu Madde Suçu	0.97	0.97	0.97	232
Yaralama	0.92	0.81	0.86	546
undefined	0.96	0.98	0.97	2589
İzinsiz Mal Alma Suçu	0.83	0.92	0.87	390
accuracy			0.89	5569
macro avg	0.86	0.82	0.84	5569
weighted avg	0.89	0.89	0.89	5569



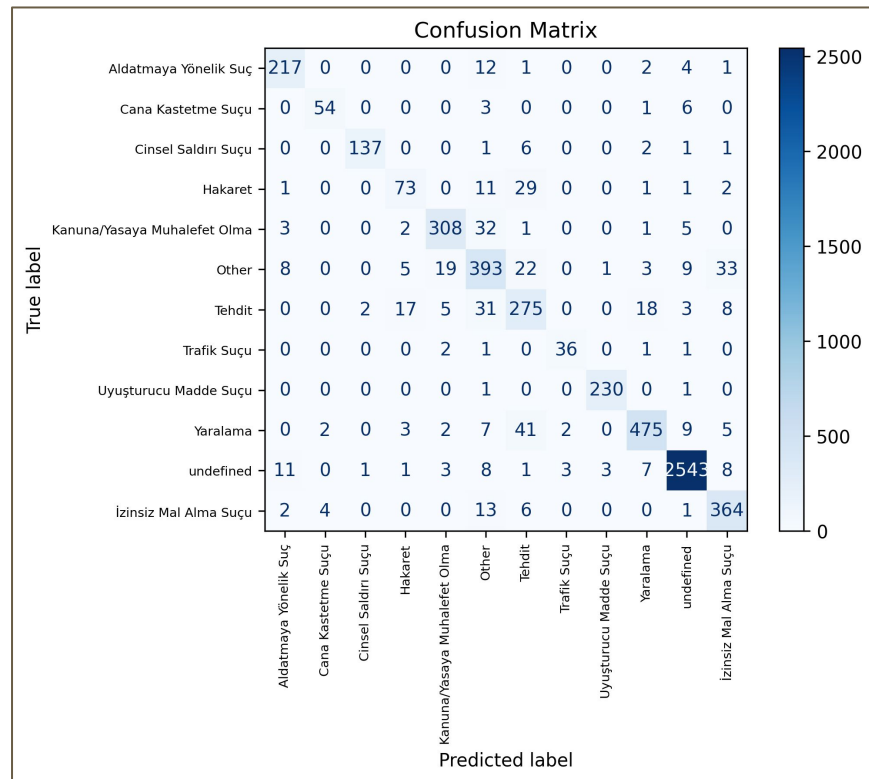
AdaBoost

Aldatmaya Yönelik Suç	0.88	0.06	0.12	237
Cana Kastetme Suçu	0.83	0.67	0.74	64
Cinsel Saldırı Suçu	0.44	0.66	0.53	148
Hakaret	0.26	0.40	0.31	118
Kanuna/Yasaya Muhalefet Olma	0.26	0.20	0.23	352
Other	0.21	0.73	0.33	493
Tehdit	0.45	0.46	0.45	359
Trafik Suçu	0.00	0.00	0.00	41
Uyuşturucu Madde Suçu	0.54	0.06	0.11	232
Yaralama	0.89	0.38	0.53	546
undefined	0.94	0.82	0.88	2589
İzinsiz Mal Alma Suçu	0.82	0.58	0.68	390
accuracy			0.60	5569
macro avg	0.54	0.42	0.41	5569
weighted avg	0.74	0.60	0.62	5569



Majority Voting

	precision	recall	f1-score	support
Aldatmaya Yönelik Suç	0.90	0.92	0.91	237
Cana Kastetme Suçu	0.90	0.84	0.87	64
Cinsel Saldırı Suçu	0.98	0.93	0.95	148
Hakaret	0.72	0.62	0.67	118
Kanuna/Yasaya Muhalefet Olma	0.91	0.88	0.89	352
Other	0.77	0.80	0.78	493
Tehdit	0.72	0.77	0.74	359
Trafik Suçu	0.88	0.88	0.88	41
Uyuşturucu Madde Suçu	0.98	0.99	0.99	232
Yaralama	0.93	0.87	0.90	546
undefined	0.98	0.98	0.98	2589
İzinsiz Mal Alma Suçu	0.86	0.93	0.90	390
accuracy			0.92	5569
macro avg	0.88	0.87	0.87	5569
weighted avg	0.92	0.92	0.92	5569



Comparison

Model	Accuracy	F1-Macro	F1-Weighted	Precision-Macro	Recall-Macro
LR	0.91	0.87	0.91	0.88	0.86
SVM	0.91	0.86	0.91	0.86	0.86
MNB	0.65	0.36	0.60	0.61	0.31
RF	0.89	0.84	0.89	0.86	0.82
AdaBoost	0.60	0.41	0.62	0.54	0.42
Majority Voting	0.92	0.87	0.92	0.88	0.87

THANK YOU FOR YOUR ATTENTION!!!

A Work By Members of Group 1