CS412-512 Machine Learning
SENTIMENT ANALYSIS - TURKISH TWEETS

Hasan Alp Boz 17842
Baran Köseoğlu 16963

Feature Set Analysis

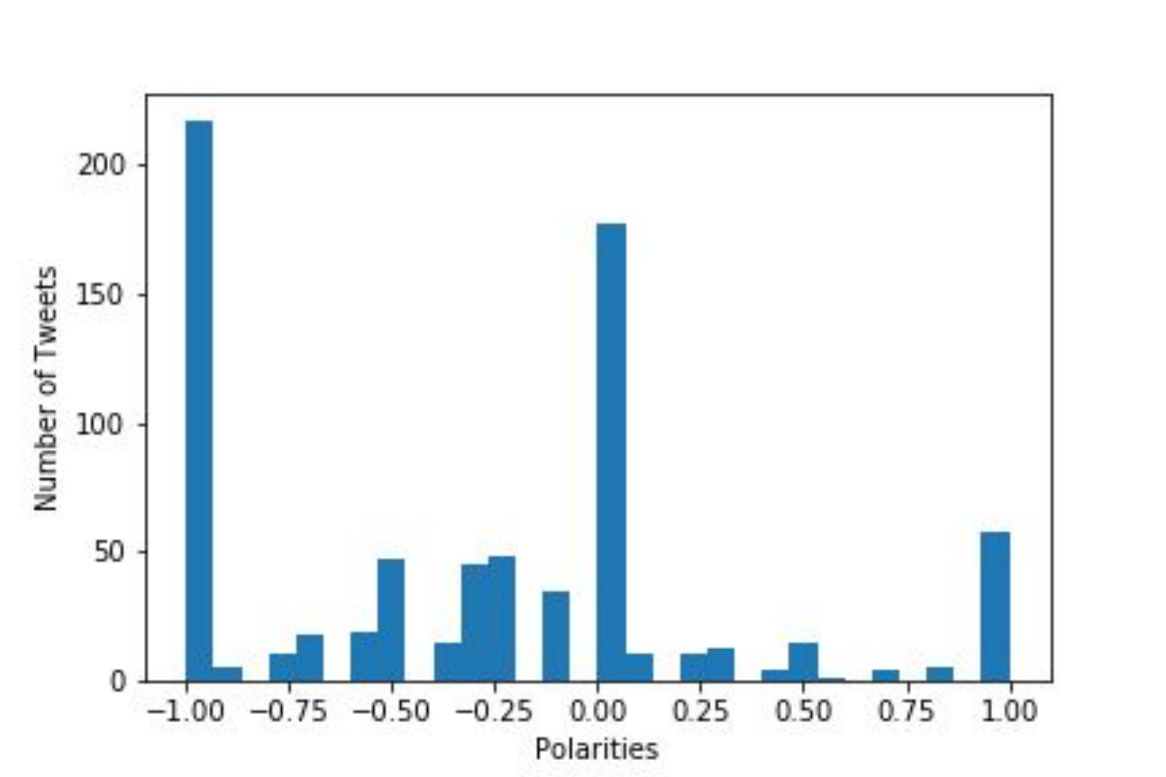

Figure 1. Distribution of Polarities among given Tweets

Feature Relations

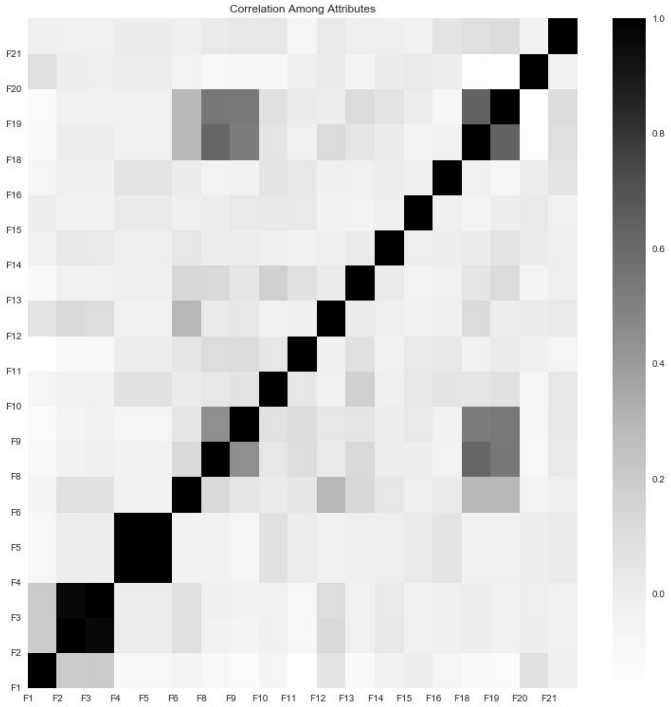

Figure 2. Correlation Matrix of Features

Model Selection For The Given Feature Set

Linear Regression → SVR → DTR → Random Forest → Ensemble Methods

Model	Parameters	MSE(10-fold cv)
Linear Regression	Default Parameters	0.368
SVR	'C': 100, 'gamma': 0.5, 'kernel': 'rbf'	0.243
DTR	'min_samples_leaf': 25, 'min_samples_split': 10	0.170
Random Forest	'min_samples_leaf': 10, 'min_samples_split': 12, 'n_estimators': 20	0.162
Bagging	'n_estimators': 15	0.158
Adaboosting	'learning_rate': 0.4, 'loss': 'exponential', 'n_estimators': 15	0.149

Feature Extraction

- Dotted and Undotted vowels in Turkish Language
- “Kavga”, “Savaş”, “Ayıp”, “Ahlaksızlık”
- “Sevgi”, “Teşekkür”, “İlgi”

Model	Parameters	MSE(10-fold CV)
Adaboosting	‘learning_rate’:1, ’loss’:’exponential’, ‘n_estimators’:10	0.146

Questions