

Objective

Our goal is to create a model that predicts depression risk using data on an individual’s personality dimensions based on the Five Factor Model of Personality (Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness) and the meaning of speech content on social networks using speech act types (Assertives, Commissives, Declaratives, Directives, Expressives)

Background

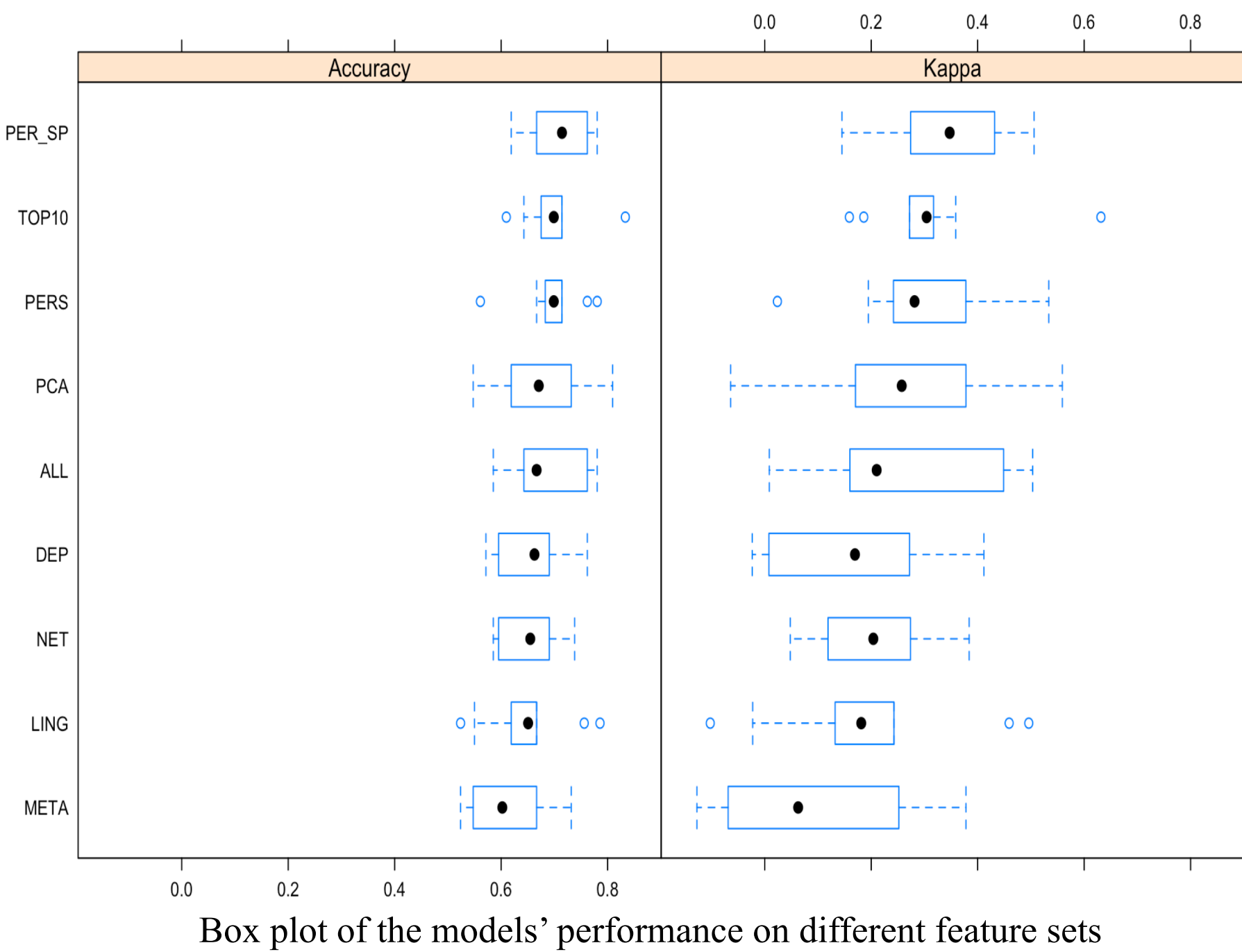
Depression affects 300 million individuals globally. It causes poor performance at work, school, and in relationships; can lead to suicide, the 2nd leading cause of death among 15-29 years old. It is associated with the personality traits: neuroticism, extraversion, and conscientiousness. Depressed individuals exhibit more negative statements and sadness, especially with friends.

Facebook has 2.2 billion monthly active users, with 64% aged 13-34 years old. Individual’s behaviors and speech on social networks may be useful to assessing depression risk.

Existing research use semantic and syntactic features from tweets or Facebook status updates to predict depression. Pragmatics, the study of the meaning of language, has not been explored. Pragmatics includes the observation that words can be used to perform an act, known as “speech acts.”

Our contribution is to examine the impact of using personality dimensions and speech act types of Facebook status messages to improve depression risk prediction.

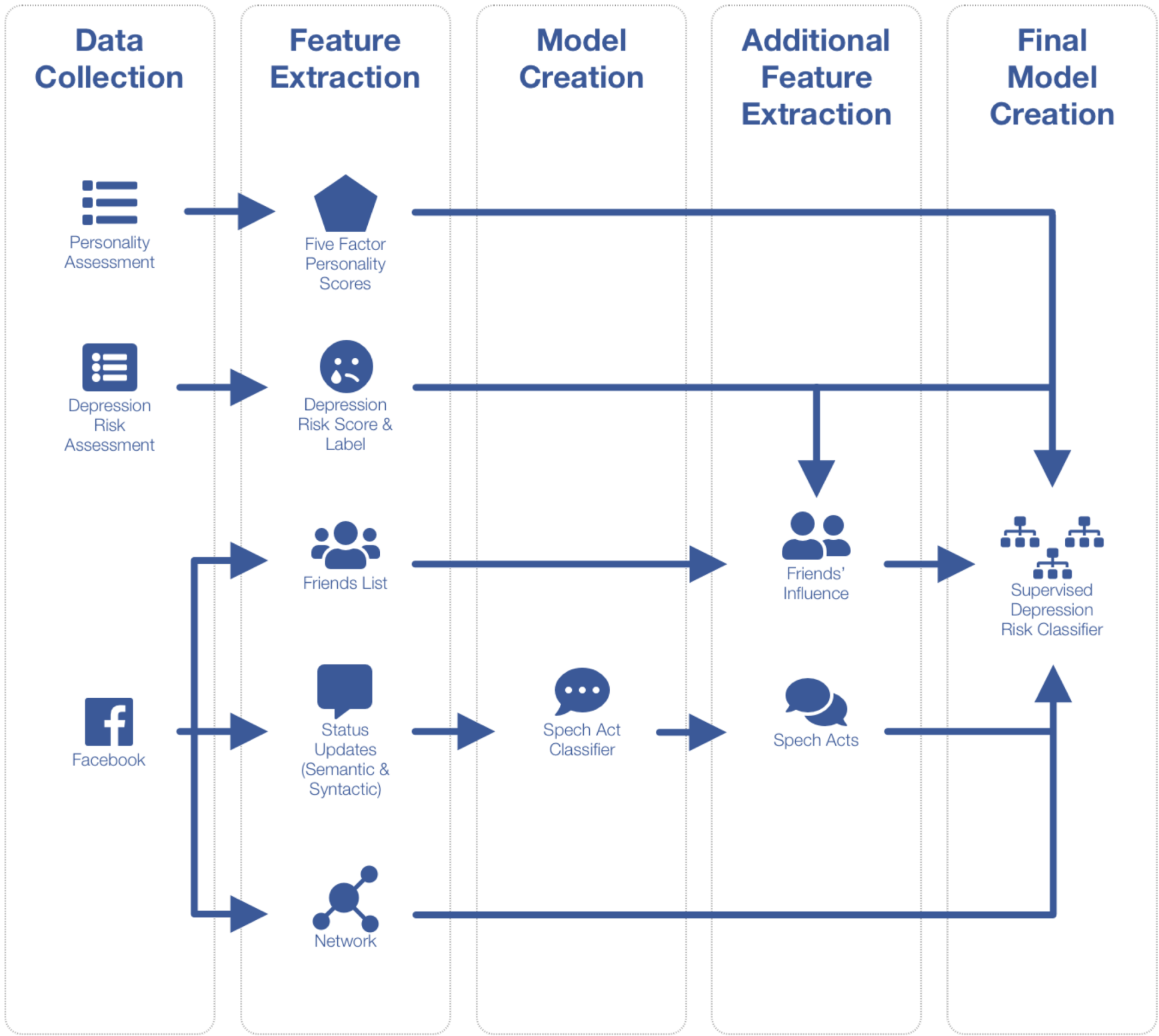
Results



Performance of the models using different feature sets on the test set

Features	Precision	Recall	F1	AUC	Accuracy
All features	0.649	0.848	0.735	0.630	0.657
Personality, linguistic, network, metadata	0.642	0.881	0.743	0.626	0.657
Top 10 features (4 personality, 2 speech acts, 3 LIWC, 1 network)	0.630	0.780	0.697	0.596	0.619
PCA on all features	0.630	0.780	0.697	0.596	0.619
Personality, Speech acts	0.629	0.746	0.682	0.590	0.610
Meta-data	0.613	0.831	0.705	0.578	0.610
Network	0.603	0.746	0.667	0.558	0.581
Number of depressed friends, linguistic, network, metadata	0.586	0.864	0.699	0.541	0.581
Linguistic	0.583	0.831	0.685	0.535	0.571

Methodology



Model	Method	Dataset	Features
Supervised Speech Act Classifier	Classification and Regression Trees	4395 status updates, 9 features	Number of characters, use of punctuation marks, use of swear words, use of speech act verbs
Supervised Depression Risk Classifier	Random Forest	521 Facebook users, 133,657 status updates, 25 features	Personality dimension scores, Use of speech act types, Number of friends also at risk of depression, LIWC categories (emotion, pronouns, articles), Network measures, Status updates meta data (insomnia index, number of updates, number of characters)

Conclusions

The best predictors for depression risk include personality scores on neuroticism, conscientiousness, and extraversion, consistent with studies relating these three personality traits with depression.

The speech act types assertives and expressives also contributed to improved prediction performance. This provides support to research indicating depressed individuals exhibit mostly negative statements and sadness, which may manifest in the use of more assertive and expressive types of speech acts.