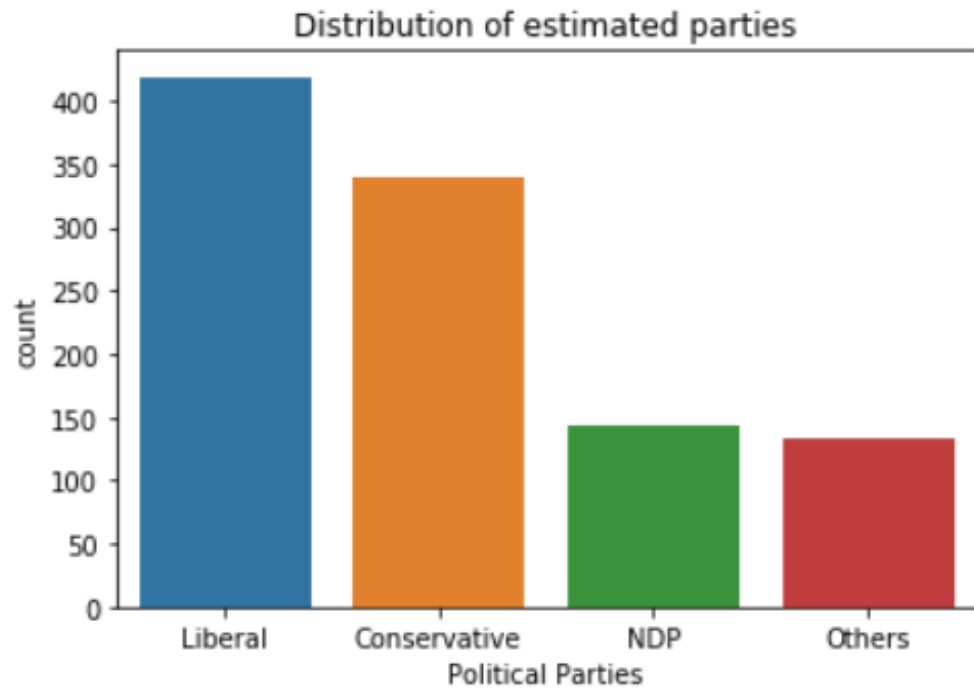


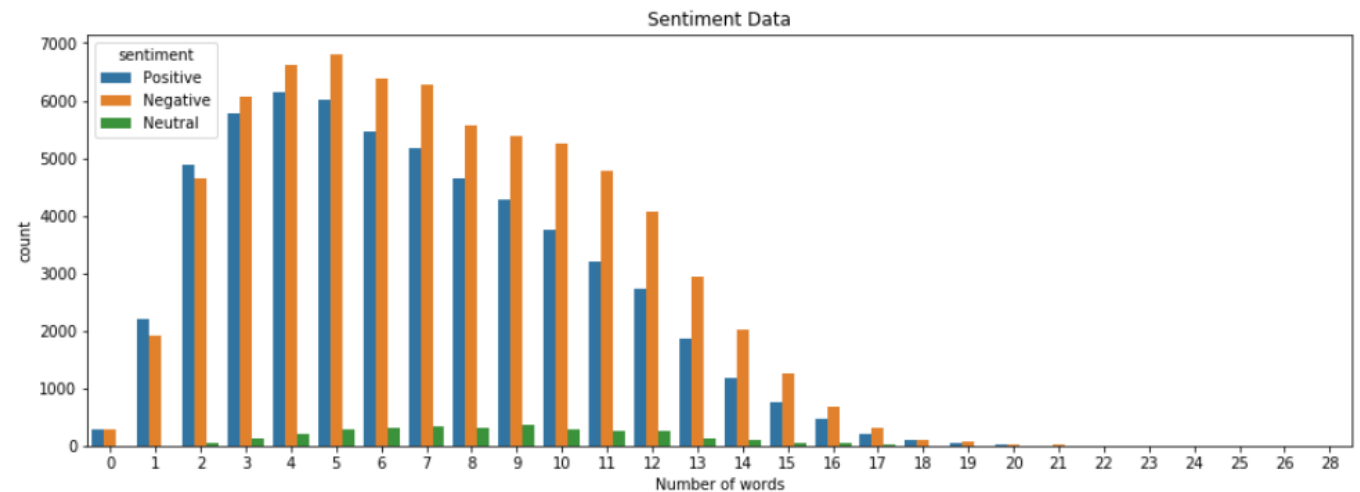
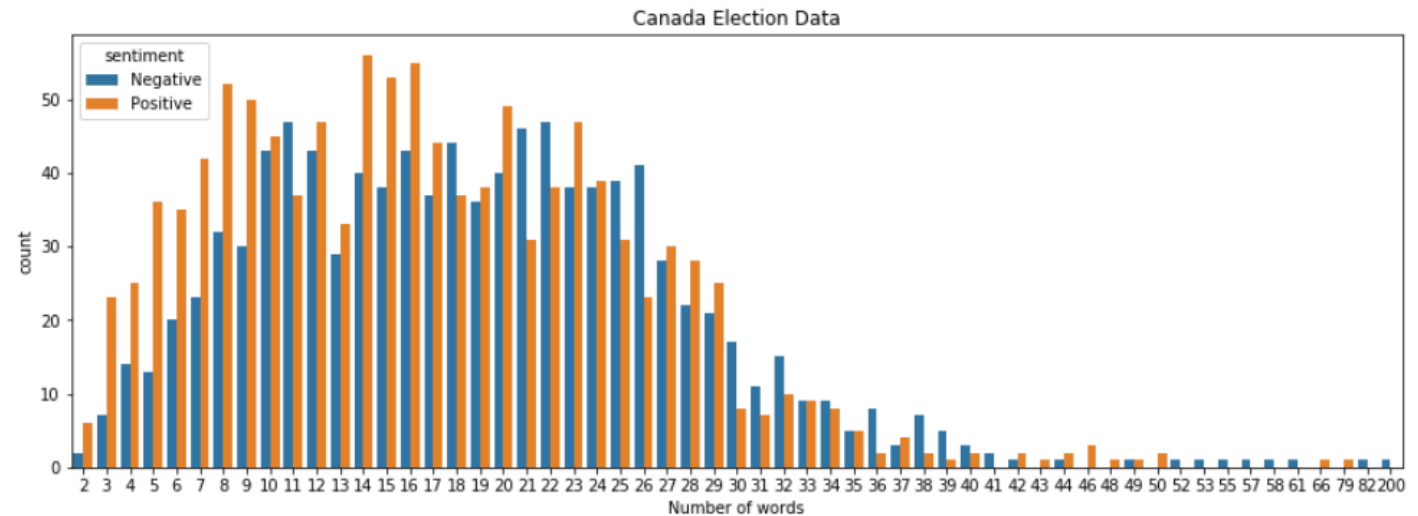
Data Cleaning

- 1. lower case the text
- 2. remove usernames and URLs
- 3. Unescape the HTML characters
- 4. Only keep English letter
- 5. remove stop words

Data Exploratory Analysis



1. Label the parties
2. Visualize the word length distributions.

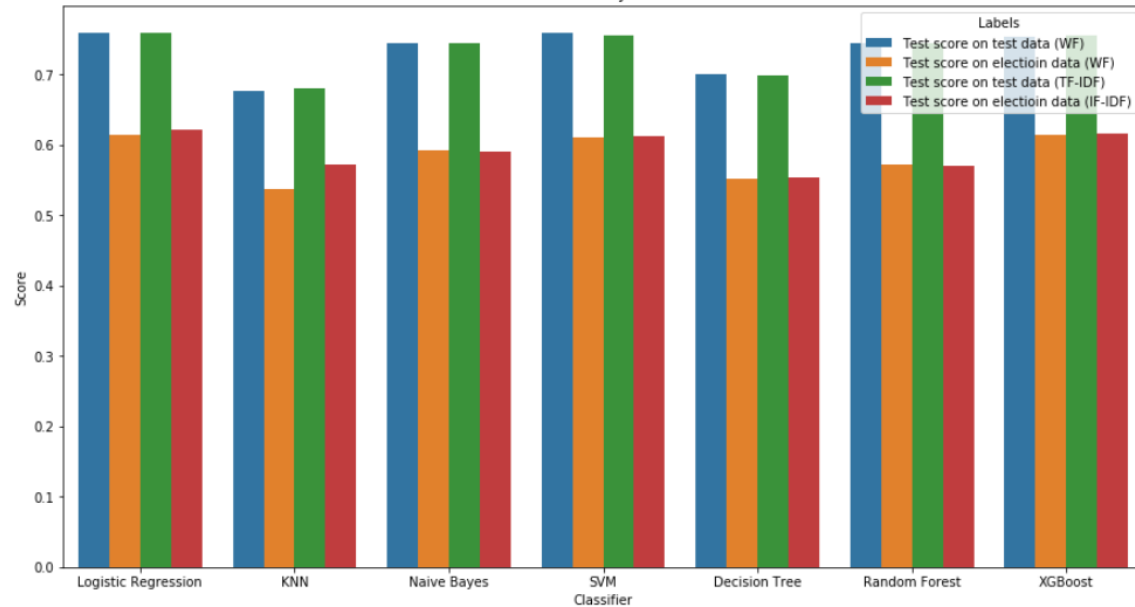


Model preparation

- 1. Binary Classifications
 - Performance score: `roc_auc_score`
 - Use WF and TF-IDF two kinds of features
 - Use GridSearchCV for parameter tuning and cross-validation
- 2. Multiclass Classifications
 - Performance score: `balanced_accuracy_score`
 - Use WF and TF-IDF two kinds of features
 - Use GridSearchCV for parameter tuning and cross-validation
 - Combine the last two labels into one.

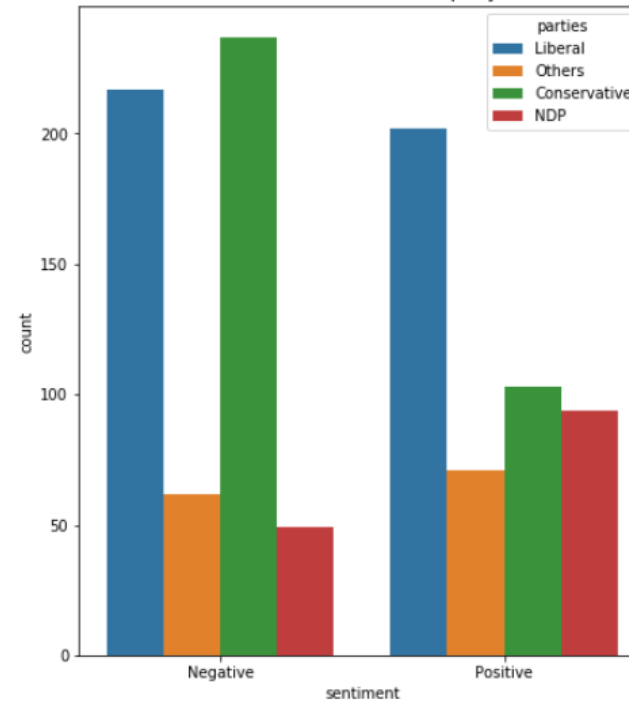
Results

Scores for Binary Classifiers

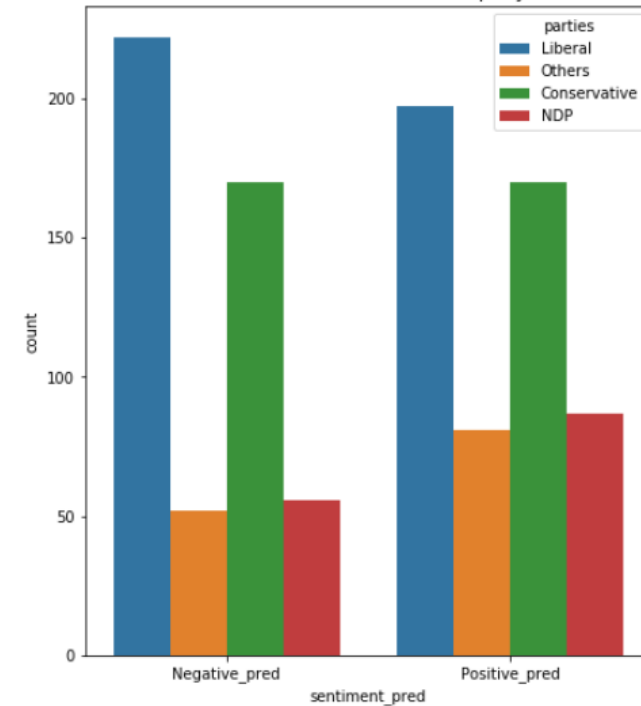


1. Models selection

Real Sentiment for each party

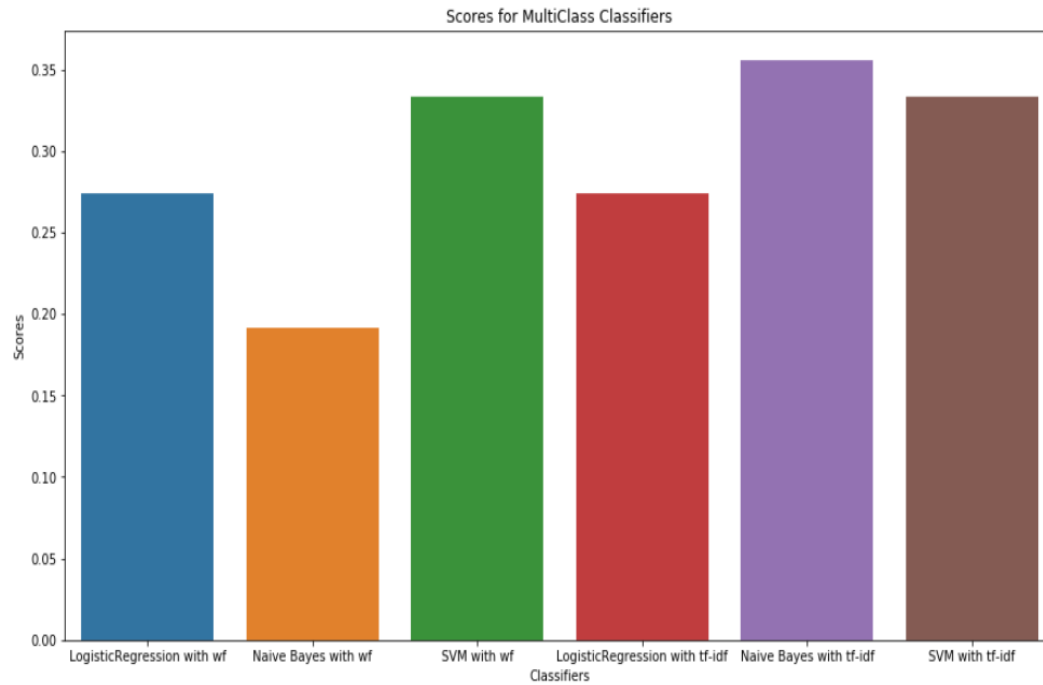


Predicted Sentiment for each party

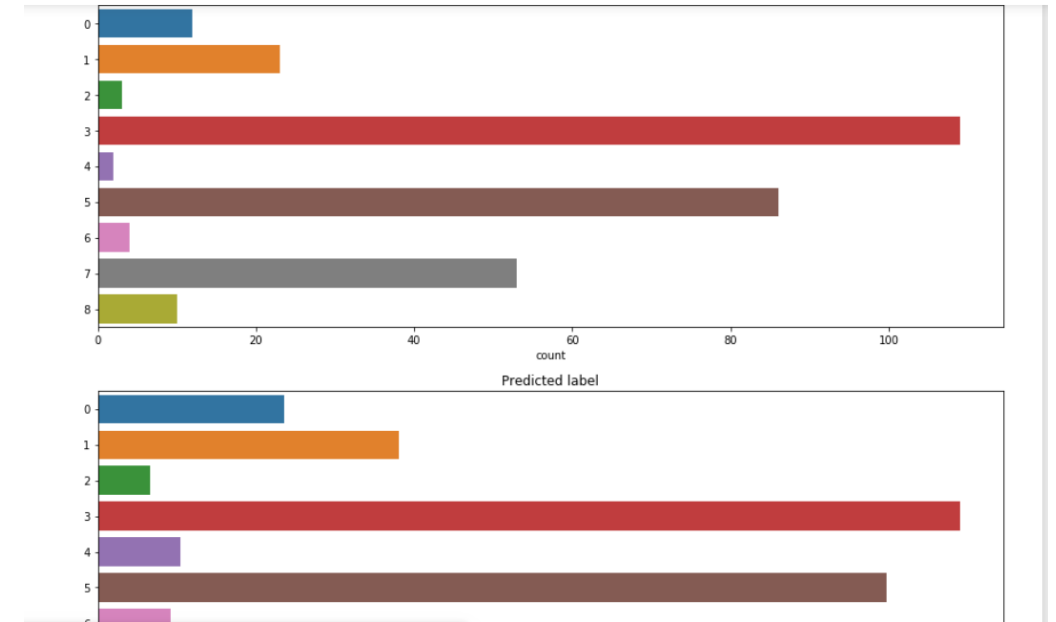


1. Check prediction performance

Results



1. Models selection



1. Check prediction performance