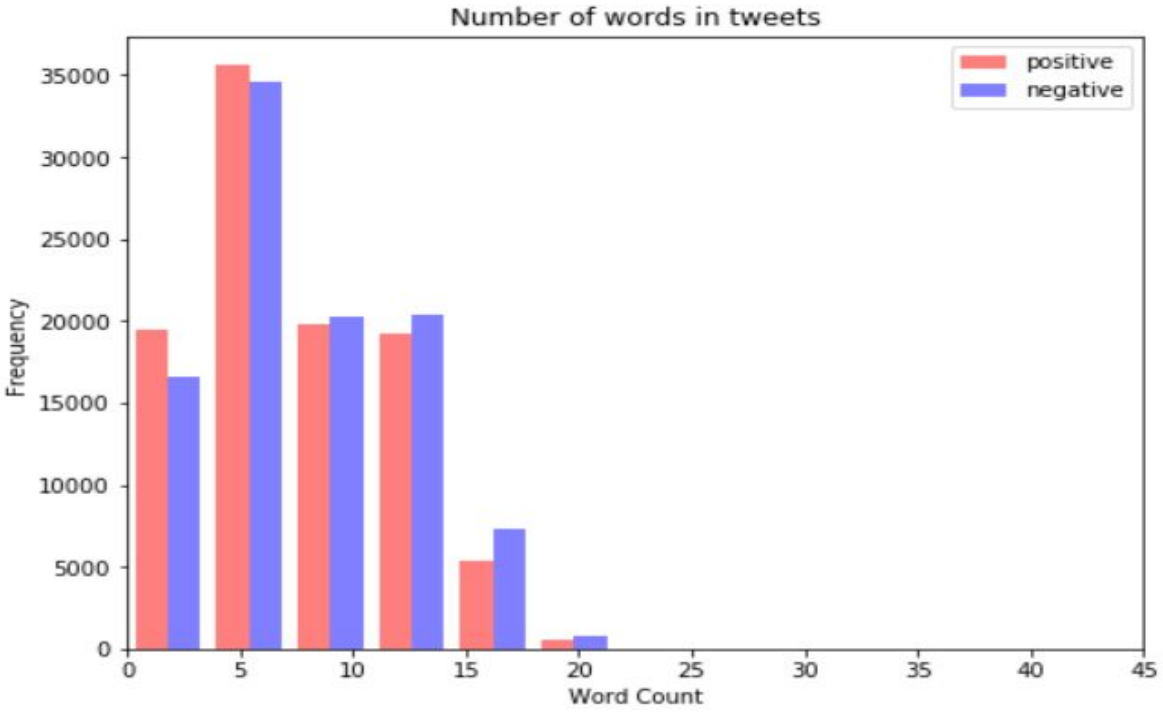
Sentiment Analysis of Twitter Data

An application of Natural Language Processing

Exploratory Analysis of Generic Tweets

UNDERSTANDING THE DATA



Word cloud for positive sentiments

Word cloud for negative sentiments

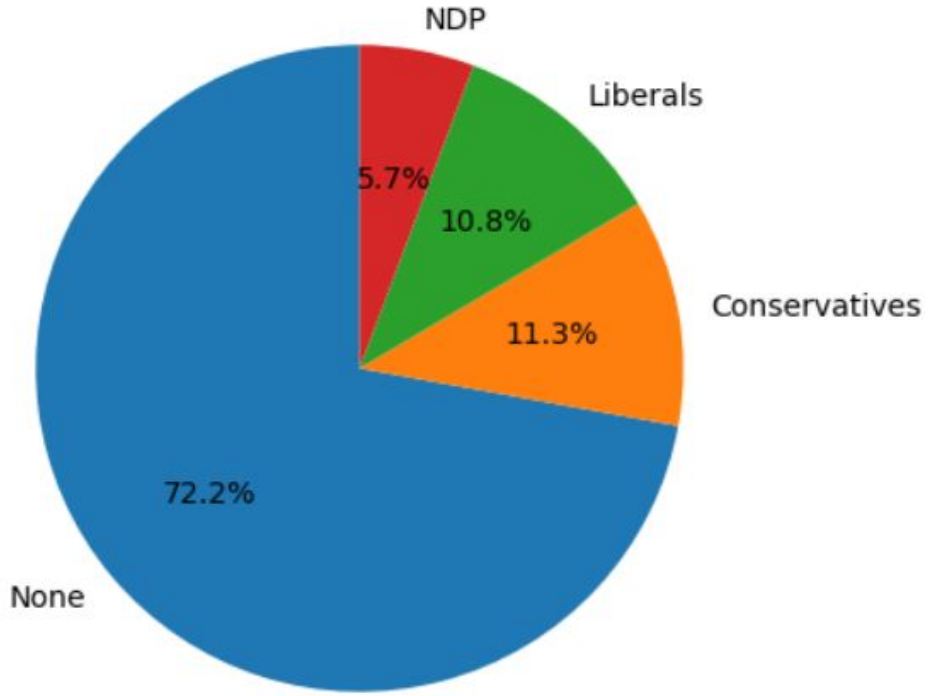
Inferences from word cloud regarding words in tweets:

Possible Subject words: work, today, going, love, time
Positive sentiments:- good, great, haha, fun
Negative sentiments:- dont, cant, hate

Average words in a tweet: 0-15

Exploratory Analysis of Canadian Elections Tweets

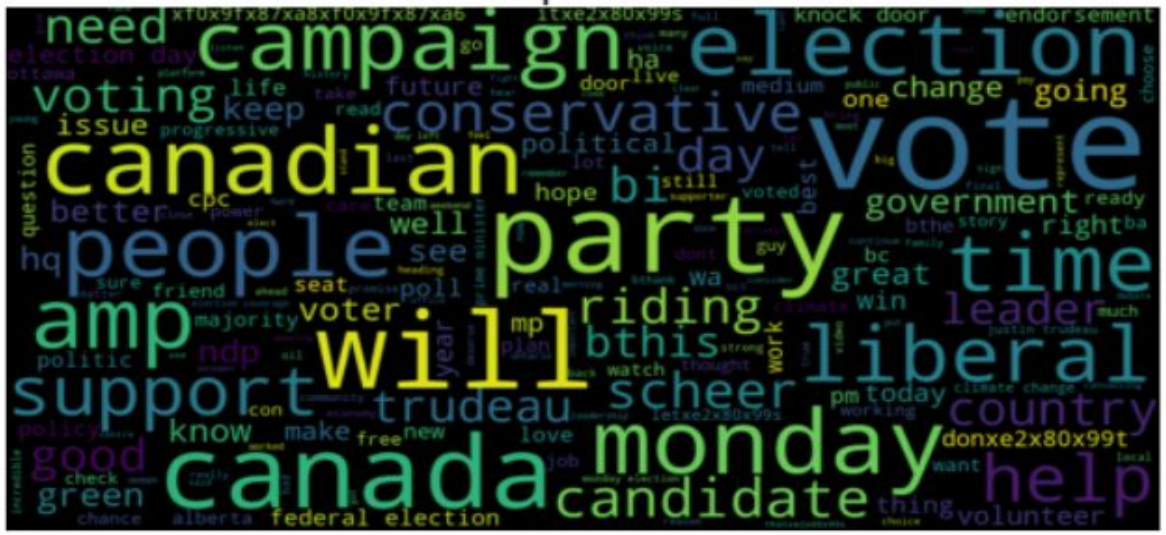
UNDERSTANDING THE DATA



Political Affiliation of tweets: (majority: none)
Conservatives > Liberals > NDP

Word cloud shows that conservatives and liberals have affiliation to both the sentiments

Word cloud for positive sentiments



Word cloud for negative sentiments

Model Results and Feature Importance

IMPLEMENTATION OF 7 MODELS ON BAG OF WORDS AND TFIDF

	Logistic Regression	Naive Bayes	KNN Classifier	SVM	Decision Trees	Random Forest	XGBoost
f1_score_bow	0.741036	0.745720	0.666657	0.748684	0.566456	0.565874	0.674799
f1_score_tfidf	0.734837	0.744240	0.634347	0.700912	0.565874	0.565874	0.675347
accuracy_bow	0.741400	0.745733	0.669800	0.749533	0.604617	0.693083	0.666650
accuracy_tfidf	0.734950	0.744300	0.640467	0.702600	0.604633	0.691817	0.666550

	Negative	Postive
0	the	the
1	to	you
2	my	to
3	not	good
4	it	my
...	...	...
95	there	amazing
96	long	girl
97	people	follow
98	see	sure
99	find	from

The figure above shows the accuracies and f1-score (macro- avg) for 7 models.

Comparing f1-score, the performance is as follows:

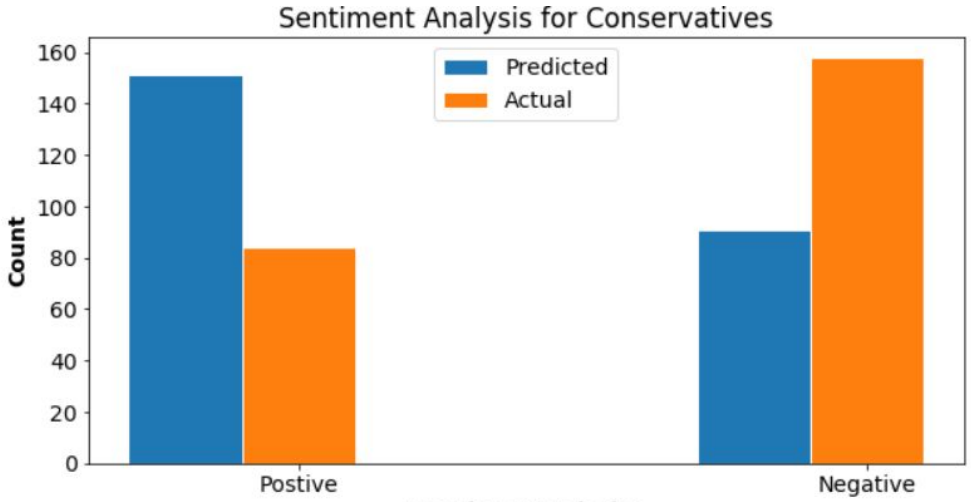
Naive Bayes > Logistic Regression >SVM > XGBoost > KNN Classifier > Decision Trees > Random Forest

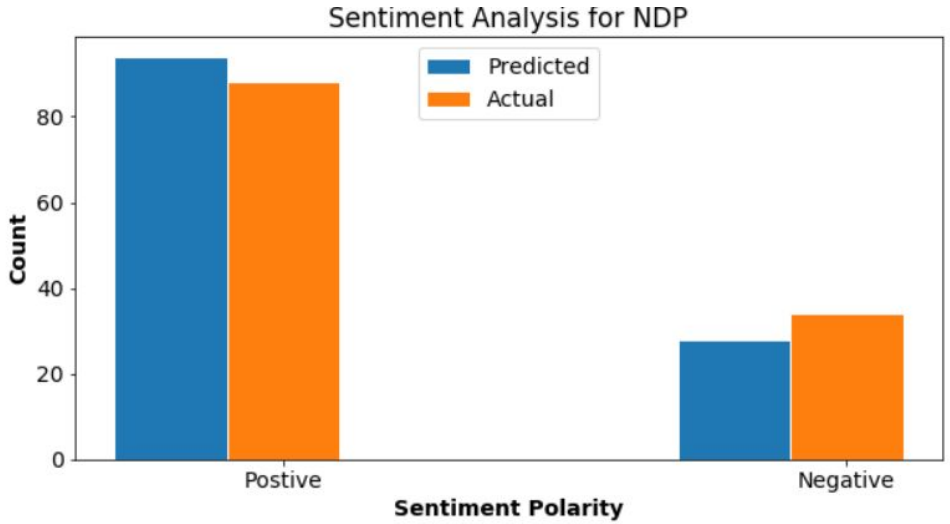
The figure on left shows the words of highest probability, i.e most predicted words for negative and positive sentiments.

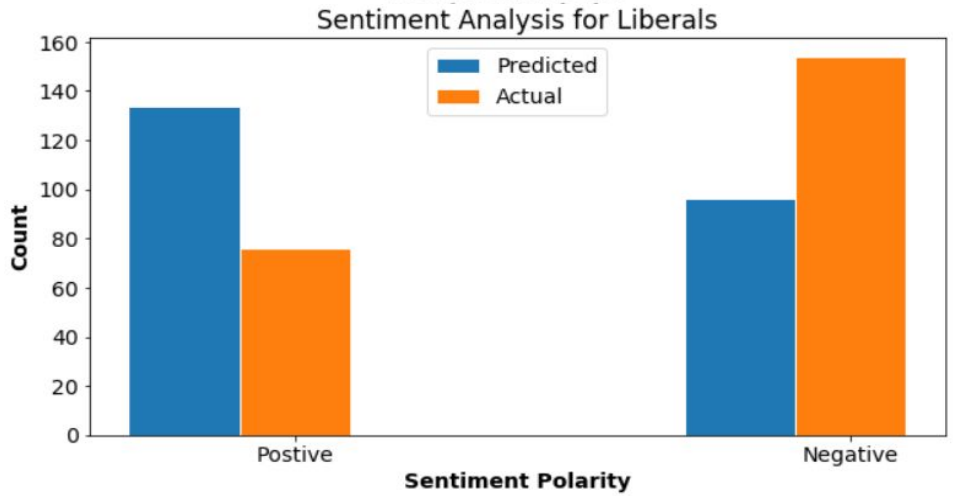
Inference: for both the sentiments, there are common words - with high importance- that should be removed. Accuracy of models can be increased.

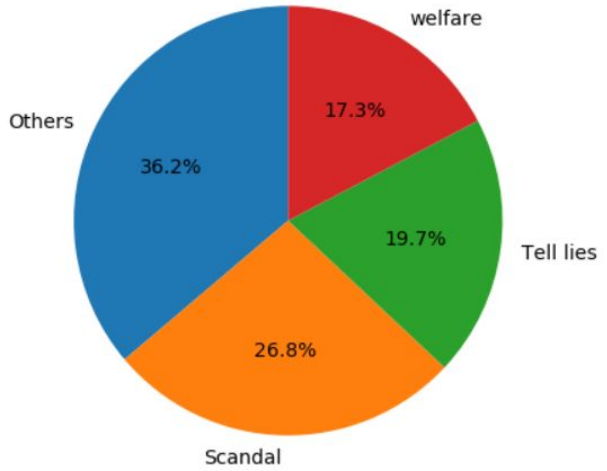
Sentiments towards political parties

PUBLIC VIEW PONT









Negative reason of the sentiment is not known for 37 % data, while 17 % is related to welfare of public, and the rest is due to scandal or hiding the facts

Model tends to predict positive class more than negative class. Common words should be removed which do not add weight to the sentiments. NDP seems to have less popularity but more positive sentiments than negative. Its reverse for other two partis