Public Relation Management Platform for the Canadian Election

Abhishek Sunnak, Sri Gayatri Rachakonda, Oluwaseyi Talabi

Simon Fraser University, Burnaby, Canada

Introduction

- Big Data and Machine learning is becoming an integral part of our lives, Political parties are now increasingly leveraging these tools to increase voter turnout and improve their chances of winning the elections.
- We developed a **NLP powered platform** to analyze the discussions about the **Canadian Election**. We aim to provide a automated platform to analyze how the perception of each party develops over time.
- We also developed models to detect **political bias** in articles to highlight publishers which do not give a balanced view about political parties to help reduce the spread of mis-information
- We have collected news and tweets about the 3 biggest parties in Canada to compare the popularity and sentiment about them leading up to the 2019 elections.
- We are augmenting our analysis by using poll data to help in understanding the reasons for swings in a candidates approval.

Objectives

We aim to answer the following questions through the platform:

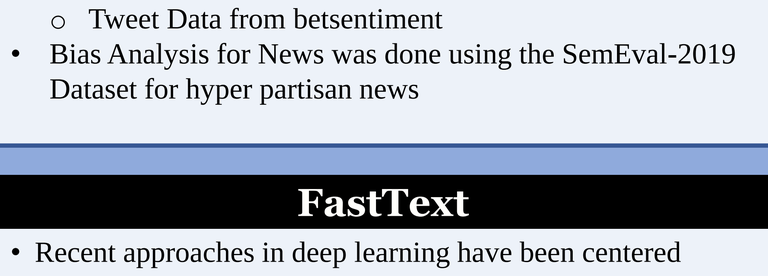
1. How does the popularity of a candidate change before the elections across different regions in Canada?
2. Which publishers have the highest share of biased news articles?
3. How does the Sentiment of news and tweets change across regions over time?

Data Sources and Tools

- The Source data for the dashboard was collected from the following sources:
  - News API and Bing Search API
  - Twitter
  - CBC
- We trained an ensemble of FastText models using labelled tweets from the following sources:
  - DataSemEval-2017 Task 4
  - GOP Debate
  - Tweet Data from betsentiment
- Bias Analysis for News was done using the SemEval-2019 Dataset for hyper partisan news

FastText

- Recent approaches in deep learning have been centered around neural networks, which achieve noticeable improvements in accuracy, **but they are highly dependent on GPU's** to improve training time
- FastText is a deep learning library developed by Facebook which allows training supervised and unsupervised representations of words and sentences
- However, **FastText** is written in **C++ and supports multiprocessing** which gives it **amazing speed without the use of GPU's**
- It finds word representations of the n-gram input features and then averages into hidden text representations. The representations are then fed through a linear classifier and a SoftMax output is used to classify the data



- The classifier is trained on **multiple CPUs with SGD** and a **linearly decaying learning rate**.
- FastText uses a **skipgram model with negative sampling** to significantly **improve the training speed** by having each training sample only modify a small percentage of the

$$\sum_{t=1}^{T}\left[\sum_{c\in\mathcal{C}_t}\ell(s(w_t,\ w_c))+\sum_{n\in\mathcal{N}_{t,c}}\ell(-s(w_t,\ n))\right]$$

Where, $N_{t;c}$: set of negative examples sampled from the vocabulary and l is the logistic loss function

Model Training Pipeline

The modelling pipeline for the bias and sentiment models consists of the following steps:

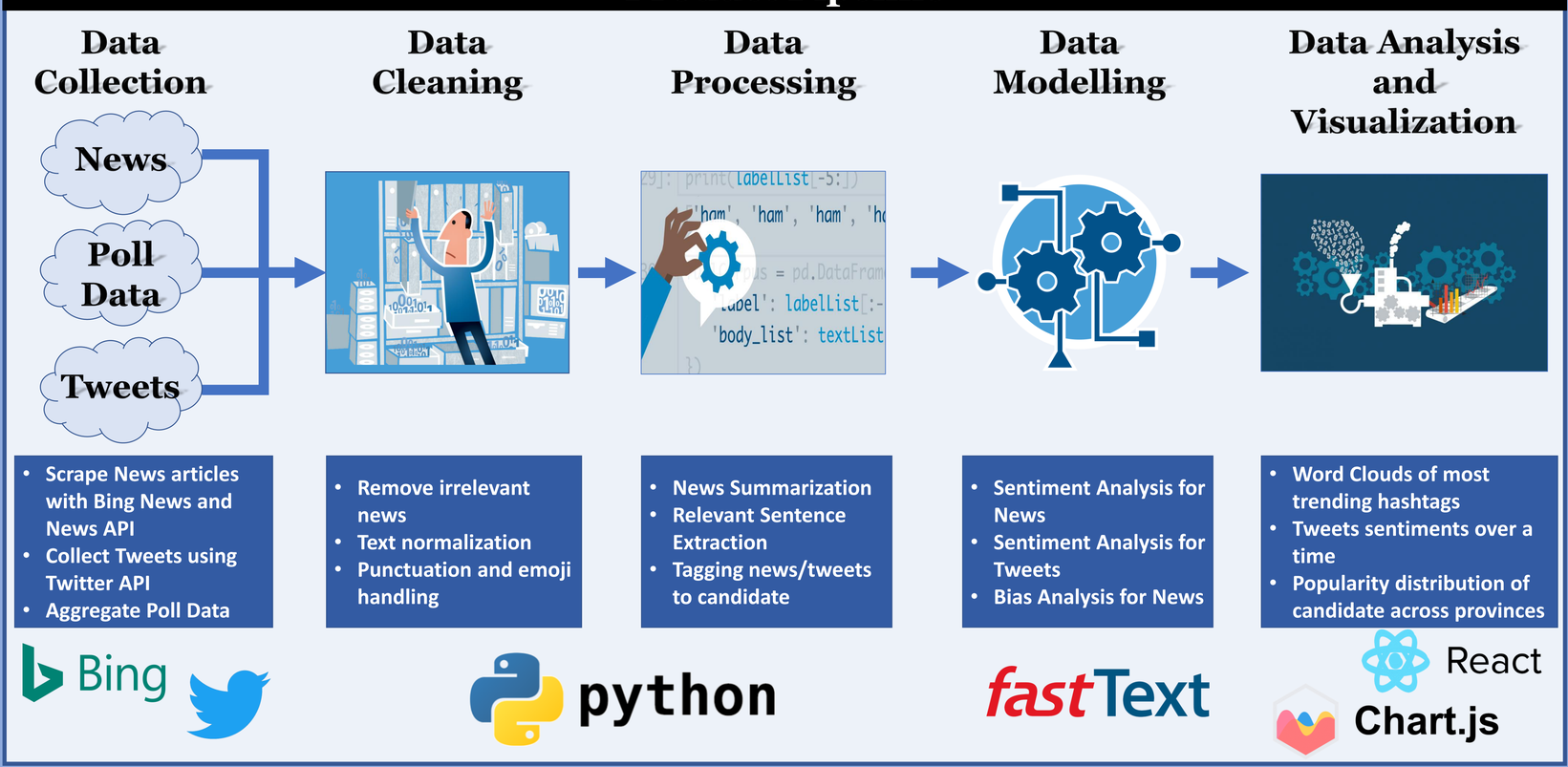
- **Data Aggregation:** Aggregate labelled data collected from multiple data sources to a common data format
- **Data Cleaning:** Clean the news/tweets to handle URL's, emojis, spelling mistakes and contractions
- **Data Preparation**: Up-sample minority classes to ensure a balanced target variable
- **Model Development:** Develop models using different hyperparameters and data samples
- **Model Ensemble:** Use an ensemble of the best 3 models to predict the sentiment of tweets

Results

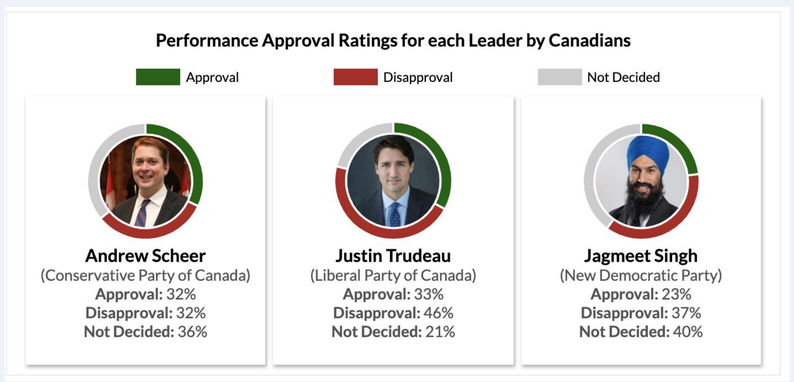
Distribution of Sentiment with Bias of News Articles

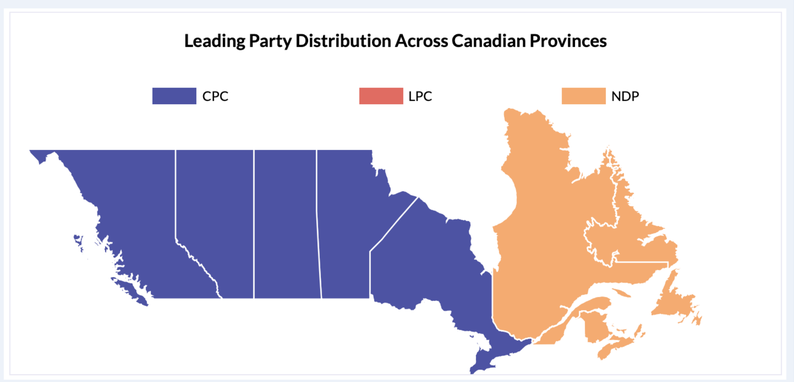
	Bias					
Sentiment	Left	Left Center	Balanced	Right Center	Right	Total
Negative	450	176	83	130	676	**1,515**
Moderately Negative	10	2	5	1	10	**28**
Neutral	133	47	29	62	94	**365**
Moderately Positive	457	313	130	247	389	**1,536**
Positive	193	148	39	87	180	**647**
Total	**1,243**	**686**	**286**	**527**	**1,349**	**4,091**

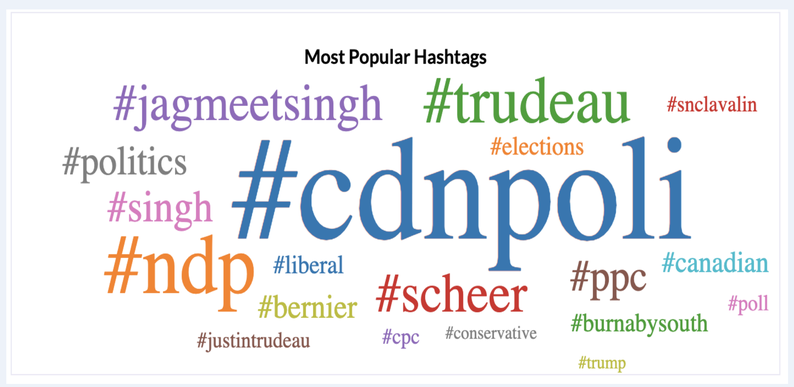
Product Pipeline

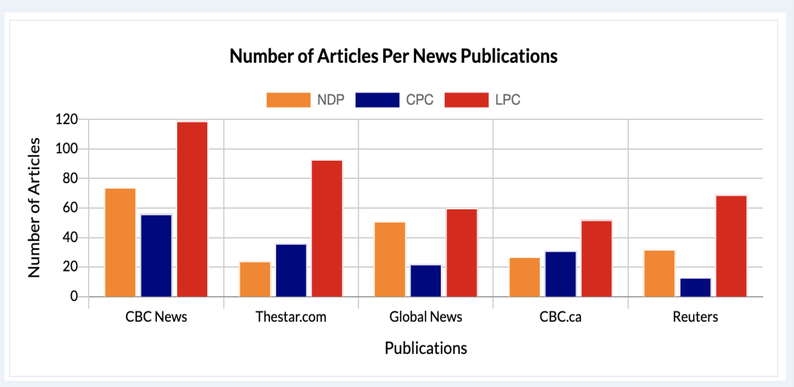


Dashboard Visualizations









Conclusions

- Most people have not decided or don't support the leading candidate for the top 3 parties in Canada, with no candidate having an approval rating over 33%
- Majority of the news and tweets about the top parties are negative with over 37% of the news about collected having a negative sentiment
- Conservatives are leading the polls across most provinces due to the dip in approval ratings for Liberals over the past few months. This can be primarily attributed to the backlash due to the SNC-Lavalin Scandal
- Most major news providers in Canada have both right and left leaning articles. However, a few publishers such as The Globe and Mail and VancouverSun have a comparatively larger share of hyperpartisan articles

References

- Enriching Word Vectors with Subword Information by Piotr Bojanowski, Edouard Grave, Armand Joulin, Tomas Mikolov
- Bag of Tricks for Efficient Text Classification; A. Joulin, E. Grave, P. Bojanowski, T. Mikolov
- CBC Historical Poll Data and Leadermeter
- Semeval-2019, Task 4: Hyperpartisan News Detection.
- react-d3-cloud (https://github.com/Yoctol)
- Twitter Sentiment Analysis using fastText by Sanket Doshi