**Twitter Election Party Sentiment Tracker**

**Description:** Using real time (streaming) Twitter data to track sentiment of the different riding/areas in Canada

\*Since the election is over, we will have to get data from a certain time frame ex. 3 months leading up to the election

**Data Mining Process**

Query by:

1. Country being Canada
2. Then split up into province
3. The find area/sections for ridings
4. At this point we need to categorize if a tweet is a political tweet about politics (use keywords).
5. Then we need to check the sentiment of that tweet. We can use a scale of -1.0 to 1.0 (good or bad sentiment and to what party it is referring to)
6. Split the sentiment up for each party and area

Use that data to predict based on the tweets what party that riding/area will be and output our own election map based on tweet sentiment

Data Collection > Political Analysis (Is it a political tweet) > Sentiment Analysis > Classify User > Analysis

**Tools**

Use

* MapReduce
* Streaming data
* Batch query
* Big queries
* Strom
* Spark
* Hadoop
* Supervised Machine Learning

(Try to use tools he gave us/suggested us to use and explain how we used them)

**Questions we want to answer?**

* What percentage of users are politically active?
* What area seems to be more interested in which political campaigns (Liberal, Conservatives, NDP)?
* What percentage of tweets are Negative/Neutral/Positive towards the current outcome of the election? Based on this, how satisfied are people with the election results?
* Compared to tweets of 4 years ago, how have people’s sentiments changed per political party?
* Does twitter sentiment analysis agree with opinion polls?
* What are the most frequent keywords related to each political campaign / political leader? Does that agree with other static data out there?
* Does housing market shift create sentiment on Twitter? How much of this sentiment is political? Towards which party are negative/positive sentiments being directed to?
* Does Tax return rate create sentiment on Twitter? How much of this sentiment is political? Towards which party are negative/positive sentiments being directed to?
* Does Insurance rate…
* Do all these shifts/changes combined create a sentiment on twitter?
* Does sentiment correlate to votes?

[Needs around 5 questions]

**Useful resources:**

Paper on scalable sentiment analysis using Spark: <https://www.researchgate.net/publication/299593681_Scalable_sentiment_analytics>

Paper on twitter political sentiment analysis:  
<https://www.aclweb.org/anthology/W13-1106.pdf>

Tips on sentiment analysis:  
<https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html>