Paper Link:

http://ijcai.org/papers15/Papers/IJCAI15-328.pdf

Title:

Tracking Political Elections on Social Media: Applications and Experience

Authors:

Bhupesh Chawda, Sameep Mehta, Venkata Subramanian, Tanveer Faruquie

Affiliations:

IBM

High level summary:

The paper discusses an implementation to predict polling and results of election based on terms that appear in social media. The overall datapath is very similar to our project, where a Twitter API is used to down tweets in realtimeand stored in HDFS.

The authors’ analysis algorithm involves identifying words and bigrams from tweets, finding a correlation between the occurrence of these words and polling data to predict future polls and results of the election using ordinary least square regression. Of particular interest is the use of lasso technique for a sparse output, where the most impactful keywords are identified. Identifying keywords that were likely to be long term was also part of the work.

This technique was applied to the 2012 US Presidential Election and 2013 Filipino General Elections. The authors were able to identify keywords such as “bin Laden”, “debt”, “deficit”, “birth certificate” that had the most impact on poll results when mentioned on social media. For US results, they were able to show predictions that were very close to the poll results on election night, and for the Filipino election, they were able to detect keywords that could potentially be scandals or breaking stories that might affect a candidate’s approval.