Predicting the results of the 15th Karnataka Legislative Assembly Elections using Twitter

Chaitanya Agarwal

Christ (Deemed to be University), Bangalore

Abstract: **With 330 million MAU's (Monthly Active Users) at the beginning of 2018, Twitter has been crowned the 13th most popular website in the world. Twitter as a data source is highly regarded due to its popularity and its infrastructure that provides almost 100% of its data through the API. Along with its own, third-party libraries have immensely contributed to the growth of Twitter as a source of information. Users are limited to post 280 characters per tweet, which often fall under numerous categories ranging from sports to politics. These tweets are suitable for various purposes such as research, analysis, marketing, campaigning and more. The goal of this paper is to provide a comparative analysis between the prediction obtained from Twitter and the opinion polls held with regards to the Karnataka State Elections and to highlight the feasibility of using social media to predict opinions, rather than the traditional methods of polling. I have crawled thousands of tweets with the help of Tweepy. Tweets are classified based on the location of origin or location the tweet is directed to. Using Natural Language Processing, each tweet is classified based on whether it was positive or negative with respect to the major political parties in the respective states.**

Keywords: **Twitter, NLP, State Elections**

I. INTRODUCTION

In this day and age data is readily available from a plethora of sources, the most popular being Social Media. Social Media has been exponentially growing as a source of news and discussion. These sites have provided a platform for people to voice their opinions with the shroud of anonymity encouraging them to be active and more engaged. Even though one is not truly anonymous, posting tweets behind a screen without having the fear of being judged has increased its subjectivity. This often results in tweets with more accurate views and opinions on everyday matters, which makes for an ideal topic for data analysis. It can be used to predict sales, box-office revenues, Stock prices, Politics and more. Politics has always been a topic that sparks controversies and discussions, which makes its way to the trending page on a daily basis. Users often have conflicting opinions based on a number of factors, which makes it a challenging task to accurately predict any political outcomes.

1.1 Hypothesis

Social Media is a feasible source to predict public opinion with a reasonable amount of accuracy.

II. DATA SCRAPING

Twitter's Application Programming Interface (API) facilitates the crawling of tweets and user data. I have used Twitter's most widely used API wrapper for Python, Tweepy.

2.1 Collecting tweets

Tweets have been scraped using the Twitter Streaming API using Tweepy. On an average, 60 tweets per second(TPS) are stored into a JSON file based on the search term. For eg: if the search term is "BJP", all tweets containing the term "BJP" will be stored in a JSON file. The Twitter Streaming API is limited to 1% of tweets, tweeted at that moment of time. With 350,000 tweets sent per minute, the 1% threshold of 3500 tweets makes for detail-rich study.
Tweets are classified based on their location, either provided by the API or filtered with the help of bounding boxes.

2.2 Preprocessing

The tweets are broken down into words. This process is called Tokenization. Tokenization is used to split a block of text into smaller segments called tokens.
Preprocessing removes unwanted entities such as URLs, HTML tags,@ mentions and other symbols. This process helps to clean the data for a more efficient and accurate result.

2.3 Token Frequency

Some words do not contribute to the overall subjectivity or polarity of a tweet. These words can be classified as stop-words. Words such as: 'to', 'and', 'from', etc do not convey a particular meaning and therefore it is logical for us to remove such words.

III. SENTIMENT ANALYSIS

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP).
TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

TextBlob returns two parameters, namely sentiment and subjectivity. Each tweet is passed through the TextBlob 'sentiment' function, and the sentiment value is stored. The values range from -1 to 1, where -1 is negative, 0 is neutral and 1 is positive.

IV. PREDICTION

Tweets have been divided and stored in separate files with respect to local majority Parties. In Karnataka, the presence of three major political parties has resulted in an interesting outcome.

Each political party has its own positive and negative tweets. The summation of positive and negative sentiment values in each party provides a foundation on which predictions can be made and further assessed. Predictions can be made based on the number of either positive or negative tweets, or by calculating the mean sentiment value of each party, comparing them to one another and converting them to a single unit of measurement. The latter provides for a more accurate and logical prediction and therefore I have implemented it.

V. PREDICTED VALUES

(As of March 5th, 2018)

BJP	CONGRESS	JDS
+ve = 33.33% -ve =12.82%	+ve = 27.27% -ve = 19.48%	+ve = 27.65% -ve = 17.02%

VI. COMPARATIVE ANALYSIS

5.1 TV9 Kannada-C-Voter poll

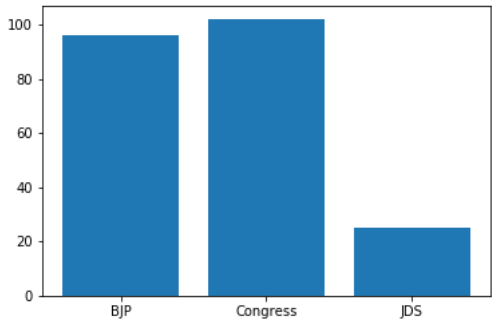

Fig 1.0 Results from TV9 Kannada-C-Voter poll.

TV9 Kannada-C-Voter opinion poll has conducted the survey and they suggest that Congress will win 102 seats out of 224 seats. BJP will be the second highest grosser with 96 seats in their basket.

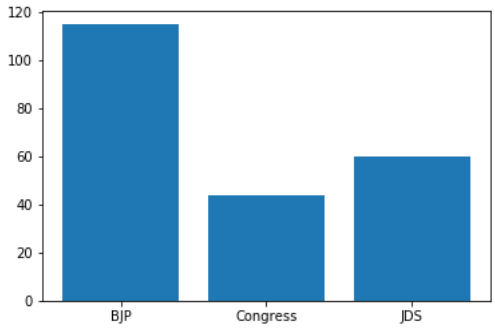

Fig 1.1 Results from the Twitter prediction.

5.2 Suvarna News poll

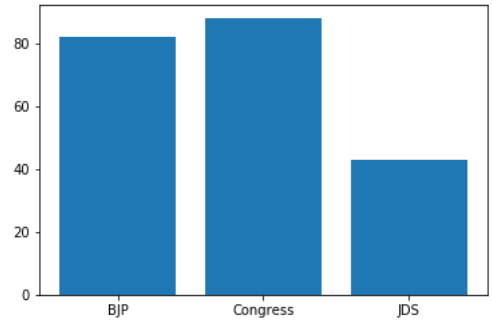

Fig 2.0 Results from Suvarna News Karnataka poll.

JD(S) might get 20% of vote share, good enough to make others notice their presence. Among the total 224 Karnataka assembly seats, no major party seems to have secured a majority.

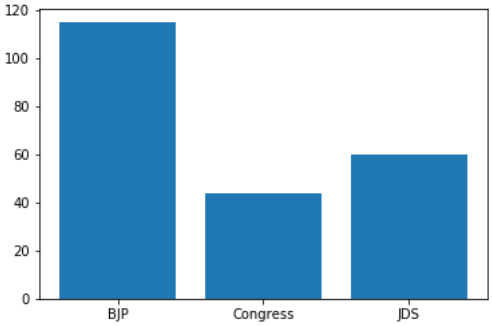

Fig 2.1 Results from the Twitter prediction.

5.3 Karnataka Opinion Survey by CHS

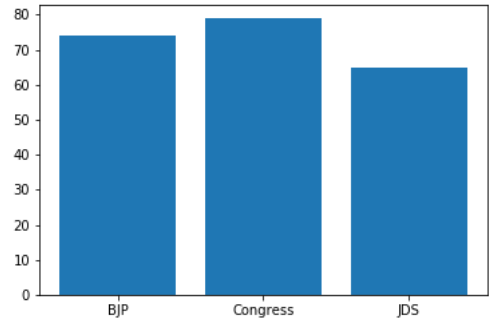

Fig 3.0 Results from the survey conducted by CHS.
CHS predicts that JDS will secure approximately 64-66 seats, Congress 77-81 seats and BJP securing 73-76 seats. The pre-poll prediction sets all three parties near each other by a small margin.

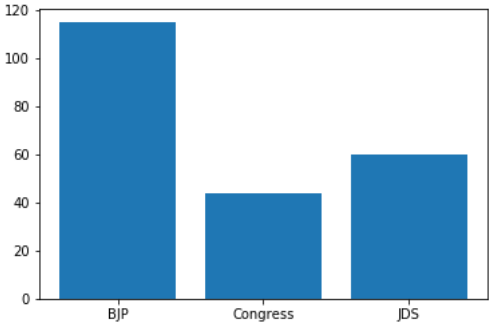

Fig 3.1 Results from the Twitter prediction.

5.4 Karnataka Opinion Poll by Creative Center for Political and Social Studies

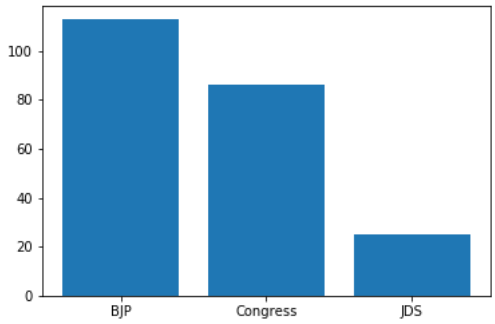

Fig 4.0 Results from an Opinion Poll conducted by Creative Center for Political and Social Studies.

As per the Creative Center for Political and Social Studies (COPS), BJP is expected to win 113 seats and INC will get 86 seats. Janata Dal-Secular is expected to secure 25 seats as per last years Karnataka election prediction.

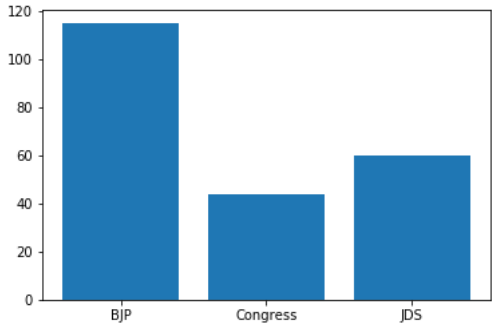

Fig 4.1 Results from the Twitter prediction.

5.5 Karnataka Opinion Poll by C-fore Survey Agency

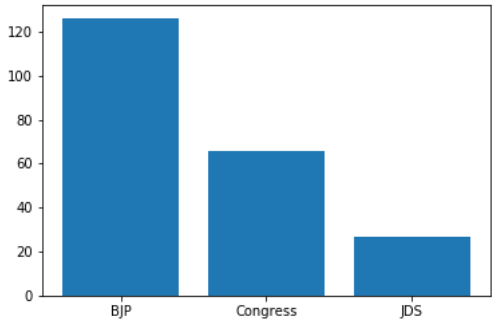

Fig 5.0 Results from an opinion poll conducted by C-fore Survey Agency.

This pre-poll survey was conducted on 24,679 voters of 165 assembly constituencies from 19 July to 10 August 2017. As per this Karnataka election prediction, Indian National Congress will win 120 to 132 seats and retains the power in the state. The Bharatiya Janata Party is expected to win 60 to 72 seats in upcoming 2018 polls. JD (S) is expected to win 24 to 30 seats

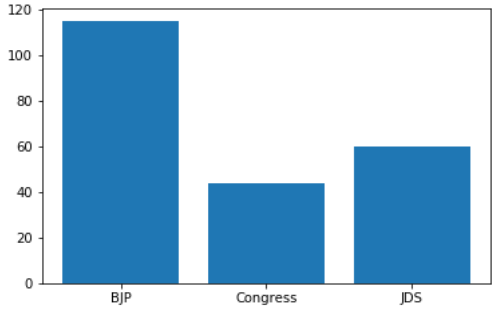

Fig 5.1 Results from the Twitter prediction.

5.6 Survey by the Karnataka State Unit of BJP

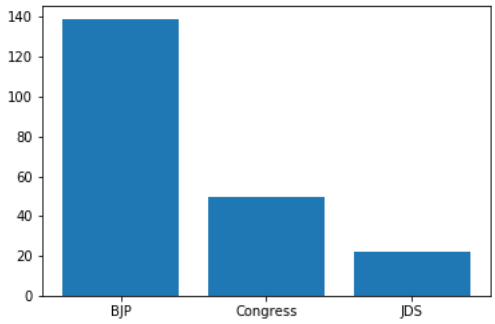

Fig 6.0 Results from an internal survey conducted by BJP.

Bharatiya Janata Party has conducted an internal survey to have an early idea of people's mind. As per the survey, BJP will win 129-150 seats, Congress will secure 50 and JDS will secure 22.

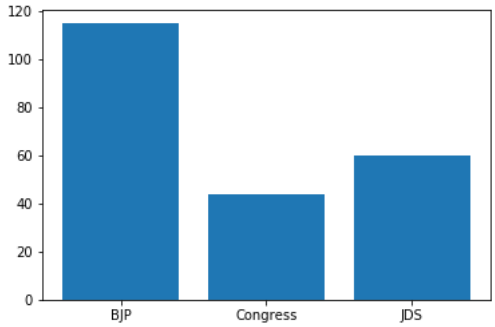

Fig 6.1 Results from the Twitter prediction.

VII. RESULTS

PARTY	PREDICTED SEATS	SEATS SECURED
BJP	118	104
CONGRESS	44	80
JDS	59	37

VIII. CONCLUSION

In this paper, I have concluded that Twitter or any Social Networking site for that matter, can be a feasible source to predict elections. In the 6 opinion polls considered in my paper, a majority of the polls do lean towards BJP securing the required number of seats. From the predictions obtained from Twitter, BJP has the highest possibility of securing a majority in the parliament. While my model is far from perfect, it has accurately predicted the party that secured the majority, paving the way for more improvements and more accurate predictions in the near future.

REFERENCES

[1] Kwak, Haewoon, et al. "What Is Twitter, a Social Network or a News Media?" *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, 2010, doi:10.1145/1772690.1772751.

[2] Sharma, Parul, and Teng-Sheng Moh. "Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter." *2016 IEEE International Conference on Big Data (Big Data)*, 2016, doi:10.1109/bigdata.2016.7840818.

[3] Patodkar, Vaibhavi N, and Sheikh I.r. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." *Ijarcce*, vol. 5, no. 12, 2016, pp. 320–322., doi:10.17148/ijarcce.2016.51274.

[4] Desai, Mitali, and Mayuri A. Mehta. "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey." *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 2016, doi:10.1109/ccaa.2016.7813707.