Opening Page (Team members &

Abstract / Introduction (One Page)

Why this topic is interesting or important to you

Data Collection:

In order to evaluate the current sentiments surrounding the 2016 Presidential Elections, the Twitter streaming API was used to sample tweets in real-time. Tweets containing the key-words mentioned below within the text body.

Key Words:

“ 'Trump2016', 'Hillary2016', '#imwithher', '#makeamericagreatagain', 'election2016', 'Donald Trump', 'Hillary Clinton', 'presidentialelections2016', 'donald trump', 'hillary clinton' “

These records were collected and stored in a JSON file (tweets.json).

A total of 127,367 tweets were obtained as a result of this process.

Data Storage and Visualization:

The JSON file was then imported into MongoDB. For this the mongo import functionality was used. MongoDB was the preferred choice as the document-oriented database allows for fast storage and retrieval. In addition it allows to summarize the results through its aggregation framework. In addition Elasticsearch was used for to allow a search engine based functionality on the dataset. Kibana Plugin was used with Elasticsearch to provide data visualizations capabilities on top of the content indexed on the Elasticsearch cluster.

Data Analysis:

1. Frequency Analysis

We performed frequency-based analysis on this collection of tweets. We determined the most commonly occurring words in the collection of tweets. To achieve this we used the NLTK.corpus library along with the use of regular expressions to filter the stop words, some common emoticons and punctuations.

The results for the top 30 words were as follows:

(PUT RESULTS)

Next, we tabulated the most popular tweet, i.e. the tweets with the largest number of retweet counts.

(PUT RESULTS)

We also tabulated the most popular tweet entities – hashtags and user mentions occurring in our dataset. This were obtained used the MongoDB aggregation framework and Elasticsearch, Kibana. These results were as follows:

(PUT RESULTS)

1. Twitter Users who are friends and followers of the Donald Trump campaign.
2. Analysis for Business Case. What did you find in the data?

Data Limitations

Conclusion