Presidential Election Prediction From Debate Tweets

Abhishek K Srivastava
asriv003@ucr.edu

UNIVERSITY OF CALIFORNIA UCRIVERSIDE

Ravdeep S Pasricha
rpasr001@ucr.edu

PROBLEM STATEMENT

Given Twitter Data we analyzed tweets during US Presidential debate to predict the political stance of the uses.

Why ??
- Valuable source of people opinions.
- Enormous number of tweets posts everyday.
- Audience Variety : Possible to collect posts of users from different social and interest groups.

Related Works:

Twitter as Corpus for Sentimental Analysis and Opinion Mining
- detecting and analysing overall sentiment of tweets.

Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment
- twitter as a forum for political discussion and whether what people tweets correlate to their opinion offline.

Techniques Used

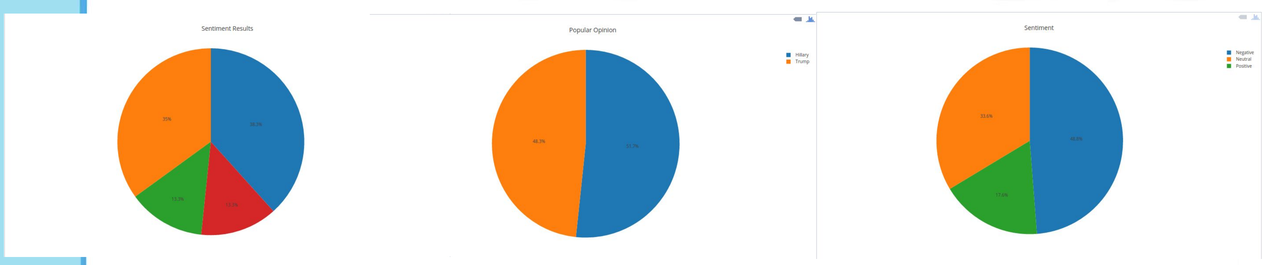
Data Collection :
- used Twitter APIs to fetch data using different hashtags.
- Collected aound 1.5 Million tweets during 2nd & 3rd Debate.

Data Cleaning:
- Parsed and selected relevant features from raw JSON data.

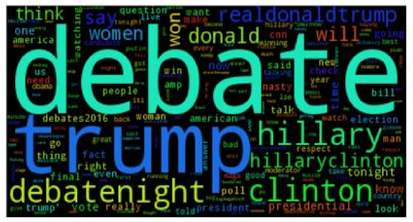
Sentiment Analysis:
- Used NLTK package for text sentiment analysis.
- Used Naive Bayes, LinearSVC, Multinomial Naive Bayes, Bernoulli Naive Bayes and Logistic Regression as classifiers.

Candidate Classification:
- Classification of tweets between Trump and Hillary.

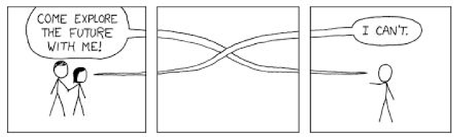
analysis techniques and Final results



summary of results and conclusions



- Out of Around 1.3 million tweets we classified 900k tweets among the candidates.
- Around 400k tweets were marked as Neutral.
- We ignored the tweets twetted either by candidates and major news outlets.
- A tweet is classified as a candidate if sentiment is positive for that candidate and negative for the opponent.
- Out of 900k tweets classified 460k as Hillary. 430k as Trump.
- One of the observation is that negative sentiment is more prevalent than positive sentiments in the tweets for both candidates.
- Each tweets has a sentiment confidence value associated with it based on our classifiers. Its value can range between [0.6,1].



Future Works:
- Better Sentiment Classifiers.
- Better Candidate Classifiers.
- Detecting Sarcasm.