**Trend Analysis of US Presidential Elections 2016 using Twitter Data**

**Project Group**:

1. Ankita Joshi
2. Madhura Gadgil
3. Narita Pandhe

**Introduction**:

**What we want to do**:

We want to analyze the different trends that can be seen from the US Presidential Elections of 2016. And we intend to do so, by analyzing the political ‘tweets’.

Why we are doing this:

Social media content is appealing in a way that we can develop a model which gives us trends which may help the candidates drive their campaign. Also, we could unmask people’s expectations of an ideal presidential candidate.

**Approach**:

Our first step will be cleaning the data, as Twitter data comes with a lot of noise, various NLP techniques like stemming, removing stop words will be utilized to clean the data and also label it. We plan to study the data using the following three different kinds of analyses

1. Statistical Analysis:

To understand what data we are dealing with, we will be conducting a statistical analysis of the data first. Our goal here will be to see underlying insights in the data. In this we plan to study

1. frequency distribution of the tweets
2. Candidate mentions in percentage
3. Hashtag distribution of the data

2. Sentiment Analysis:

We can determine the popularity of candidates by running sentiment analysis. We will be determining, for each candidate, whether people are talking positively or negatively of the said candidate. We will be using Naïve Bayes algorithm for this.

3. Topic Analysis:

We will be using the Latent Dirichlet Allocation (LDA) model to extract the underlying topical structure from the tweets. The main idea behind using LDA for tweets is that the words in each tweet are generated in the context of each topic, which will help us see, for example, if people are talking about Trump, what are the topics discussed by these people.

**Data**: We currently have ~ 30 to 40 gb. But we are not yet sure how much of this we will be using. We have obtained this data from Dr. Shannon Quinn.

**Reference**: http://arxiv.org/pdf/1407.0622v1.pdf