

AN ANALYSIS OF POLITICAL SOCIAL MEDIA DATA USING NATURAL LANGUAGE PROCESSING

Rahul Chandrashekhar '17

Faculty Sponsor: Peter Yoon

The increase of filtered content on the internet, particularly on social media, has left internet users in a “bubble”. In this bubble, the users only see such political content that aligns with their biases while the content with alternate viewpoints is filtered out. The recent election period was one which saw various prominent politicians use the power of social media to reach out to voters. As a result, there is an abundance of data coming in from political parties and figures. This study uses the power of Natural Language Processing and Machine Learning to analyze the opinions and views of politicians on Twitter in an attempt to reveal interesting trends and uncover biases. The Twitter API was used to download Tweets and articles over a 10 month period and organized in the form of a database. All the collected data was then preprocessed to filter out unnecessary terms while leaving only the essential ones. The filtered data was then analyzed using the Waikato Environment for Knowledge Analysis (WEKA) which is a Java based Machine Learning toolkit. The collected results were compiled using a charting library, Highcharts.