Project Proposal

Course:	CS410 Text Information System	MCS DS Fall 2022
Project Name	**US Midterm Elections Twitter Sentiment Analysis**	
Team Name	Googolplex	GitHub Repo: https://github.com/pd-illinois/CS410Googolplex.git
Team Members	Prateek Dhiman (Captain)	pdhiman2@illinois.edu
	Unnati Hasija	uhasija2@illinois.edu

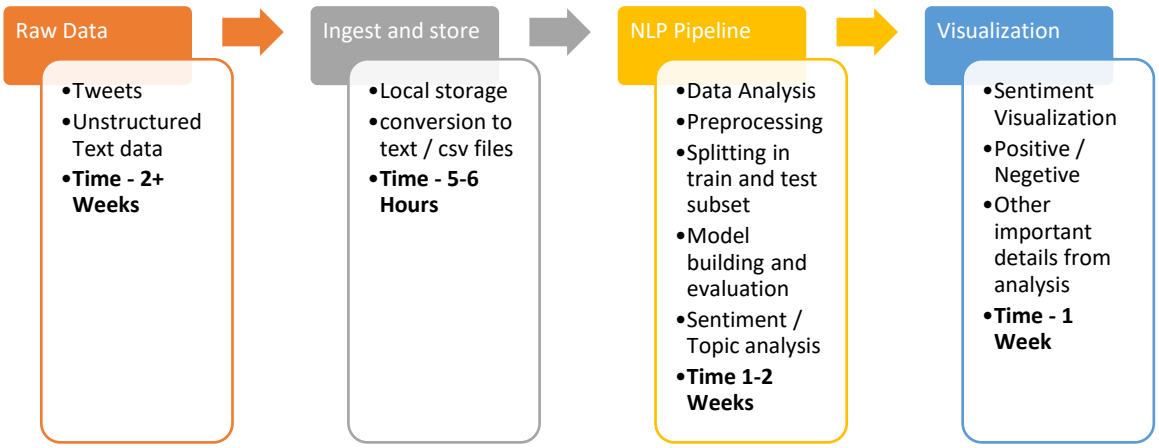
Project Description

We are living in an Information Age. With the invention and growth of social media humans came together on a common platform. They started to connect, share and post their daily thoughts and feelings. Having been the early users of most social networks and living in their growth and euphoria, most of us never understood its far-reaching impacts. Today, the social network has the capability to influence and change what people think, do, and react to everything happening around them. The social network has the capability to incept thoughts and ideas at an individual and societal level. As a result, not only they can create a viral trend, but they have the power to change governments. What we think and do today is a microcosm of what is happening around us and is due to the information we see on social media.

This brings us to our project proposal and the invisible impact of social media. We are specifically looking at one of the aspects of social media and its recommendation engine, which shows us the content relevant to user browsing. Each one of us has our timelines, posts, tweets etc. tailored to our interests, search history, and other criteria. Especially around election season, when there is a political environment around the nation, social media add fuel to political polarity across good people. Over the years this polarity is growing. Through this project, we would like to identify the political sentiments in favor or against the established political parties based on what work they have done, what top topics are being discussed, and how they impact people who are reading and sharing them. The idea is to derive a neutral and unbiased opinion for a variety of posts and topics. We plan to identify common hashtags and keywords and when they were being posted.

We plan to use Twitter as the main platform for either scraping the tweets or using Twitter APIs. Since, with the use of Twitter APIs, we can fetch only certain period worth of data, we can use this to show a dynamic dashboard. The initial idea includes using ‘snscrape’ or ‘tweepy’ to build the data/tweet store. This data will contain the polarity of the tweet, date, user information, and text. We plan to use python as the main language for this project. Once we ingest the data and store it locally, we would pass this to an NLP Pipeline. This pipeline would split, stem, and tokenize the tweets, start the document analysis, information extraction, and summarization. We also plan to implement a ranking mechanism for the relevance of tweets in relation to the US midterm elections. Finally, we can take this processed information and perform sentiment analysis and pass this to a visualization tool like matplotlib or tableau.

The overall approach is shown below. This is the initial conceptualization of our project and may change during its realization phase.



The core outcome we are targeting is to understand overall sentiments as we approach the midterm election. Based on this analysis we plan to gather a neutral and unbiased view of what is happening on one of the major social networks and see if our analysis echoes later with the results of the midterm elections in US. The expected outcome of this visualization is to show the people's sentiments on different topics w.r.t the different political parties.

For this project, we are a group of two, and both working full-time jobs with families. Based on our knowledge of Text information systems we plan to keep the topic simple yet impactful to today's trend. We would divide the workload equally into two parts where we work side by side as one team to gather the tweets, build our code base and work on refining the results. The documentation will be built side by side as the project progress. We plan to have small daily checkpoints followed by longer working sessions on the weekends to ensure the project moves forward. The end documentation would also include project progress, detailed project plan, architecture, and necessary documentation for using the project.

The total time we expect each task to take is also highlighted in the process flow above.