CS410 - Text Information Systems - Project Proposal

Sentiment Analysis of Tweets.

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

   Hitesh Yadav is the captain for the GroupX. Below are the details of the group members:

Name	NetId	Email
Hitesh Yadav	hitesh2	hitesh2@illinois.edu
Shubhendu Bhaskar	sb59	sb59@illinois.edu
Gopikrishnan Srinivasan	gs25	gs25@illinois.edu
Sanjeev Singh	sksingh6	sksingh6@illinois.edu
Chandan Goel	chandan3	chandan3@illinois.edu

2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?

   What is your free topic? Please give a detailed description?

   For our final project, we have chosen the topic - 'Sentiment Analysis of Tweets'. Sentiment analysis in general can be understood as the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. In our case, this piece of text would be a user's tweets.

   What is the task? Why is it important or interesting?

   The intent of the project is to allow a user to enter a tweet (text) on the UI interface and learn about its sentiment (positive, negative, neutral) using a number of different techniques/ methodologies. The idea is to serve the sentiment using multiple algorithms to provide users with enough information to make a decision.

   We have seen the evolution of social media in the last two decades. At the offset of this technology, the main focus was on its advantages. For example,

- Connect with our friends and family regardless of factors like physical distances, etc
- Product reviews
- Job search

It was only recently that we humans started to realize its dark side or disadvantages. There have been an number of research that have shown the following negative impact of social media platforms like twitter,

- Mental Health: Depression, Anxiety, Insomnia, etc.
- Cyberbullying
- Fear of Missing Out (FOMO)
- Unrealistic Expectations
- Negative Body Image
- General Addiction
- Political Polarization

We have chosen Twitter as the social media platform for this project not just for its popularity in recent years but also because it covers both positive and negative use case(s) of social media. While we understand that the current technologies might not be able to address all the use cases, we are confident that we can enhance the platform's functionality in future, as the technology around sentiment analysis further evolves (for example: fake news, etc).

With our project, we are hopeful to provide a platform to our users (student, youth, parent, buyer, seller, businesses, etc) to understand the sentiment of a tweet and then decide if they really want to get influenced by it or should they just simply choose to ignore it.

As of now, our scope is to provide the sentiment analysis to our users using a independent user interface but in future we can expand the scope to,

1. Include other social media platforms.
2. Integrate the technology directly with social media either using plug-ins or browser extensions.

What is your planned approach? What tools, systems or datasets are involved?

From a high level, we plan to use the following approach,

1. Build a user interface to enable a user to enter a tweet/ text for which they want to perform sentiment analysis.
2. Identify Tweet dataset - TBD
3. Perform Dataset Preprocessing: Data Cleanup, stemming, lemmatization, etc.
4. Leverage different technologies:
    - Python libraries like Textblob.
    - ML Models like - Naive Bayes.
    - Pre trained ML model - RoBERTa (available on huggingface.co).

5. Build API(s) to take in a tweet as a text and return the sentiments.
6. Integrate the UI with the sentiment analysis returned from our API.

What is the expected outcome?

We expect our platform to enable a user to understand the sentiment of a tweet and then decide if they really want to get influenced by it or should they just simply choose to ignore it.

How are you going to evaluate your work?

We plan to evaluate our work using both statistical (like - Precision, Recall, F1 Score) and empirical methods to decide how well the model is performing.

Note: The above can change/ evolve as we progress through the implementation phase.

3. Which programming language do you plan to use?
   We are planning to use Python in this project.

4. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

Task Name	Hours
Data Curation and Pre Processing	20
Build User Interface for Sentiment Analysis	30
Sentiment Analysis Methods	30
Integration with UI	20
Preparing the presentation and documentation	20