CS410 - Text Information Systems - Project Progress Report

Sentiment Analysis of Tweets.

Original Tasks:

The following are the tasks identified in the project proposal.

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

1) Which tasks have been completed?

- ☑ **Data Curation**
  - Identify Tweet dataset
  - Reference: Kaggle

2) Which tasks are pending?

- ☐ **Data Pre Processing**: Data Cleanup, stemming, lemmatization, etc
  - In Progress
  - Expected Completion: November 18, 2022
- ☐ **Build User Interface for Sentiment Analysis**
  - In Progress
  - Expected Completion: November 20, 2022
- ☐ **Sentiment Analysis Methods**
  - Yet to start
  - Expected Completion: November 27, 2022
- ☐ **Integration with UI**
  - In Progress
  - Expected Completion: November 30, 2022
- ☐ **Preparing the presentation and documentation**
  - Yet to start
  - Expected Completion: Dec 06, 2022

3) Are you facing any challenges?

We have been facing the following challenges:

- None of the team members have experience in building UI applications. We are trying to learn and explore different technologies like React, Angular, Flash to build the UI for the sentiment analysis application.
- We ran into a dilemma to either pull the tweets directly from the twitter API or use an existing publicly available dataset. The issue with fetching tweets from twitter won't have the sentiment label assigned to them and would make the team spend hours to manual classify them,

Detailed Progress Update

Curate Dataset: (Completed)

We reviewed multiple publicly available datasets on different websites and finally decide to use one from Kaggle.

Reference: https://www.kaggle.com/datasets/imrandude/twitter-sentiment-analysis

Data Pre Processing: (In Progress)

We are currently working on building the method to perform the following tasks:

- Data Cleanup - handle missing values
- Split data into training and test subsets
- Apply Stemmer
- Apply Lemmatizer

Build User Interface for Sentiment Analysis: (In Progress)

The plan is to create a ReactJS based UI running on a NodeJS server. It would be a responsive web application which can run on any kind of mobile device.
Here are the steps completed so far:

1. Created a React application which would take the user input tweet message
2. The user interface can query the backend server and fetch the response
3. For now we are rendering the results in a text format (yet to finalize on this)

Sentiment Analysis Methods: (Yet to start)

We haven't started on this yet.

Integration with UI: (In Progress)

To integrate the UI and the model, we are planning to use Flask. We would be having a method to receive the input from UI and query the model and respond back with the sentiment.

Below are the tasks completed so far:

1. Created a flask module which receives input from the UI
2. Responds back with a static JSON which has sentiment decision and scores

Preparing the presentation and documentation: (Yet to start)

We will start with this task once we have the basic building blocks like UI, Sentimental Analysis Methods, etc completed.

Note: The above can change/ evolve as we progress through the implementation of the sentiment analysis application.

Team:

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