Social Media Data Science Pipelines Project 3 Proposal: Answering Questions

Devang Jagdale
djagdale@binghamton.edu
Binghamton University
Binghamton, New York, USA

Tejas Hiremath
thiremath@binghamton.edu
Binghamton University
Binghamton, New York, USA

Chaitanya Jha
cjha@binghamton.edu
Binghamton University
Binghamton, New York, USA

1 INTRODUCTION

With a live data collection system established and a deep understanding of the dataset characteristics, this project involves addressing key research questions. This project focuses on answering the research questions and developing an interactive tool for querying and visualizing results from analyses conducted in the previous project.

2 PROJECT DESCRIPTION

The primary objectives of this project are:

- Answer at least one research question proposed in Project 2.
- Create an interactive tool to enable querying and visualization for at least two analyses from Project 2 (three analyses for graduate groups). The tool can take various forms, such as a web-based dashboard, a command-line tool, or a web API.

3 RESEARCH QUESTION

The research question we plan to address in this project is:

- How do temporal patterns in social media activity correlate with engagement metrics like comments?
- How do shifts in sentiment metrics—such as happy, sad, angry, and hopeful correlate with key events and discussions in U.S. politics ?
- How do the tones and volumes of discussions differ between mentions of different political candidates or parties?

4 ANALYSIS FOR DEMO

- Country Based Analysis
- Sentimental Analysis
- Live Data of Db for comments from 4Chan and Reddit

5 INTERACTIVE TOOL

The interactive tool will support the following analyses:

(1) Time-series analysis of engagement metrics (e.g., likes, shares, comments) over specific time periods.
(2) Sentiment analysis of user comments across various posts.
(3) Comparative analysis of engagement metrics across different post types (text, image, video).

The tool will allow users to:

- Query data by specifying time ranges and post types.
- Adjust parameters like the granularity of time periods (daily, weekly, etc.) and sentiment thresholds.
- Visualize results through interactive plots and summaries.

6 PROPOSED TOOLS AND FRAMEWORKS

The project will utilize the following tools and libraries:

- **Flask/Django**: To develop a web-based interactive dashboard.
- **Plotly/Matplotlib**: For creating interactive visualizations.
- **Pandas**: For data manipulation and analysis.
- **Scikit-learn**: For implementing sentiment analysis models.
- **PostgreSQL**: For querying and managing the dataset.

7 CONCLUSION

This project will build upon prior work by answering key research questions and providing an interactive framework for data exploration. It will also demonstrate the practical application of data science techniques for real-world social media analysis.