Analysis of Longitudinal Facebook Posts Before and After a Tornado Using Large Language Models

Abdullah Al Reza

University of Oulu

1. Abstract

The study focuses on analyzing Facebook posts made before, during, and after the 2011 tornado in the USA. My aim is to understand shifts in sentiment, behavior, and engagement over time by leveraging Large Language Models (LLMs). The study will conduct a sentiment and behavior analysis on longitudinal data. The process will begin by converting raw text data into structured datasets and then applying LLM-based methods to reveal patterns. (Miah,2024) We will focus on how users react and communicate around natural disasters. (Hara, 2015) The insights gained can help improve disaster response strategies and community engagement in future crises. I will process, organize and propose a new social media dataset around the timeline of 2011 tornado. After that I will do sentiment and behavior analysis using LLM on that dataset.

2. Background

Social media platforms like Facebook serve as critical channels for real-time information exchange, especially during disasters. However, understanding the behavioral shifts in user's communication patterns before, during, and after such events requires detailed analysis (Hara, 2015). Traditional data analysis methods often fall short in grasping the complexity of human emotions expressed through posts, comments, and messages. On the top of that it is time consuming. The use of LLMs presents an opportunity to automatically capture these emotional and behavioral transitions (Assiri, 2024).

This thesis aims to investigate the behavioral and emotional transitions observed on Facebook around the time of a tornado, using LLMs for sentiment analysis and pattern detection.

3. Objectives

This research has the following objectives:

Objective 1: To categorize Facebook posts and texts. Then I will divide that into "before," "during" and "after" the tornado event.

Objective 2: To preprocess raw text data and convert it into a structured format suitable for analysis.

Objective 3: To apply LLMs for sentiment analysis and behavioral trend identification across different time periods.

Objective 4: To provide insights into the public's reaction before, during, and after the tornado based on social media data.

4. Research questions

* Does the emotional valence change over the timeline of tornado?

* How can sentiment analysis of social media posts during natural disasters help improve community engagement and support services?
* What role do social media platforms play in spreading real-time information during natural disasters?

*How can this be optimized through behavioral analysis using Large Language Models (LLMs)?

5. Methodology

5.1 Data Pre-processing

The initial step involves converting the rich text file data into an Excel sheet for a structured analysis. Here are the steps I will follow:

Data Conversion: I will extract relevant data such as date, time, post content, and likes using **Python scripting**. I will use the **pandas** library to handle data extraction efficiently.

Data Cleaning: Unnecessary content like hyperlinks, special characters, unnecessary syntaxes and other irrelevant information will be removed. I will also correct spelling or grammatical issues while preprocessing the dataset.

Data Segmentation: Once cleaned, I will categorize the data into three separate datasets based on the post timestamps—**before, during, and after** the tornado event.

Excel File Creation: I will merge clean data to excel files. The data will be divided into three parts according to the timeline.
Visualization: I will use different graphs and charts to portray recent situations of data.

5.2 Data Transformation

After cleaning and categorizing the data, I will transform it in such a way that I can use that data sets to train the LLM model.

Feature Engineering: I will create additional features like word count, post length, and engagement metrics (e.g., likes and comments) to assist in predictive analysis.

Base LLM training: I will take a base LLM model to train it with organized data. I will choose a model which is designed for social media data.

Fine Tune LLM: After the training, I will fine tune the model according to my data. It will help me to get better results for my dataset.

LLM Preparation: Once the fine tuning is complete, I will move on to train it with my new conversational data. I will categorize the text data as positive, negative, or neutral based on sentiment analysis and other behavioral indicators.

This structured process will ensure that the raw data is gradually refined, categorized, and analyzed to produce meaningful insights with the help of LLMs.

5.3 LLM-based Analysis

The core analysis will be conducted using fine-tuned LLMs to perform sentiment analysis and behavioral pattern detection.

LLM Model Selection: Models like Ts, RoBERTa, or BERT can be used to analyze sentiment and tone in each Facebook post. These models excel at understanding context and emotion in textual data.

Sentiment Analysis: I will analyze the sentiment on different time periods (before, during, and after). I will categorize this by positive, neutral and negative categories to understand which post needs to be focused. This will help in identifying emotional shifts during the event. I can use different methodology to analyze emotional shifts in this study. (Vinney, 2019)

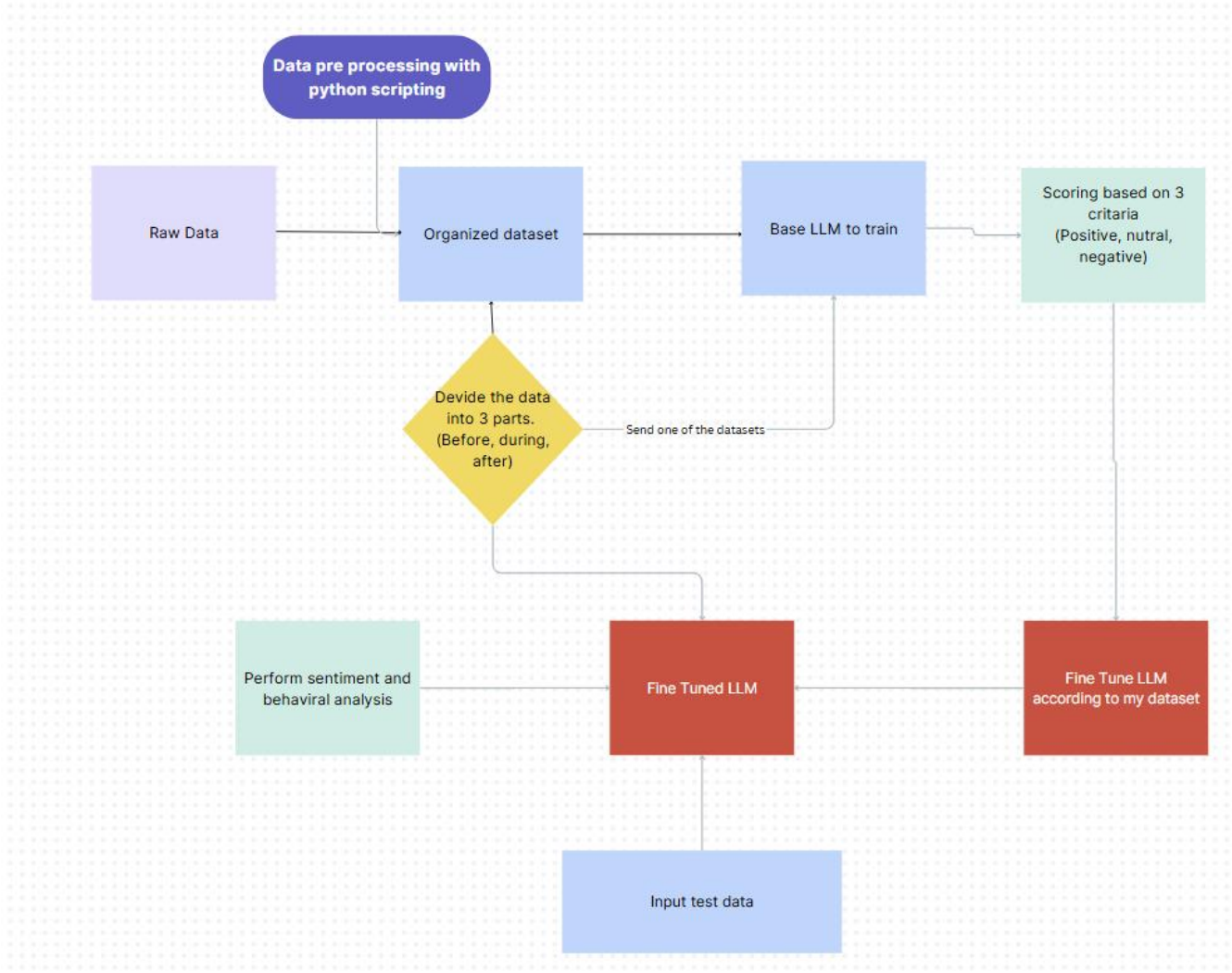
Topic Modeling: I will use LLMs to identify key discussion topics in each time-period to detect how the focus of user discussions evolves. (Assiri, 2024)

Behavioral Analysis: I will investigate how engagement (likes, shares, comments) changes across time periods, providing insights into community reactions and levels of engagement. We can find different patterns using user engagement on social media platforms during natural disasters. (Vinney, 2019)

6. Analysis Tools:

Different Python libraries will be used in different parts. Panda will be used for data handling and preprocessing. 'Transformers' will be used for the LLMs analysis. Nltk (natural language processing) will be used for the analysis.

7. Pipeline Visualization:



8. Expected Findings

The research is expected to reveal the following insights.

A distinct shift in sentiment from neutral/positive before the tornado to negative during the event. It might be followed by a gradual return to positivity after the event.

I will be able to find increased engagement levels during the tornado as users seek information and support. I will be able to Identify the key topics and behavioral patterns that occurred across posts, indicating public concerns and reactions during disaster times.

9. Conclusion

This thesis aims to show the power of LLMs in analyzing behavioral and emotional changes in social media communication during a natural disaster. By automating sentiment and topic analysis, we can gain

valuable insights about how people respond to natural crises. The findings of this study will be useful for emergency management agencies to better understand public reactions and tailor their responses.

10. References

* Vinney, C., Dill-Shackleford, K. E., Plante, C. N., & Bartsch, A. (2019, March 21). Development and Validation of a Measure of Popular Media Fan Identity and Its Relationship to Well-Being. Psychology of Popular Media Culture. Advance online publication. http://dx.doi.org/10.1037/ppm0000188

* Assiri, A. et al. (2024) ‘DeBERTa-GRU: Sentiment analysis for large language model’, Computers, Materials & Continua, 79(3), pp. 4219–4236. doi:10.32604/cmc.2024.050781.

* Miah, M.S.U., Kabir, M.M., Sarwar, T.B. et al. A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. Sci Rep 14, 9603 (2024). https://doi.org/10.1038/s41598-024-60210-7

* Hara, Y. (2015) ‘Behaviour analysis using tweet data and Geo-tag data in a natural disaster’, Transportation Research Procedia, 11, pp. 399–412. doi:10.1016/j.trpro.2015.12.033.