CS 584: Disaster Tweet Information Extraction Pipeline

Nitin Chaube
Stevens Institute of Technology
nchaube@stevens.edu

Krishna Sai Srinivas Venigalla
Stevens Institute of Technology
kvenigal@stevens.edu

Spring 2025

Abstract

This project presents an end-to-end pipeline to extract structured information from disaster-related tweets. Our system classifies tweets, categorizes crisis types, matches new reports with ongoing events, and extracts key details like location and casualties using GPT-4. This automated approach provides a fast and scalable way to monitor crises via social media, potentially aiding humanitarian response and situational awareness.

1 Introduction

Real-time crisis detection from social media is crucial for early response and situational awareness. X, a widely used platform for information dissemination, is often the first source to report natural disasters such as earthquakes, floods, or fires. However, most tweets are noisy, unstructured, and non-informative. The goal of this project is to design a pipeline that filters, classifies, and summarizes disaster-related tweets into structured formats that are easy to monitor and visualize.

2 Related Work

Past research includes crisis classification using logistic regression, CNNs, and LSTMs on the CrisisLex and CrisisNLP datasets. However, these models often fail to capture contextual semantics. Recent advances like BERT have shown promise in improving classification accuracy. Additionally, few pipelines integrate classification with retrieval (RAG + FAISS) and structured generation (GPT-4) for complete end-to-end systems. We aim to address this gap.

3 Methodology

Our pipeline includes eight major components:

1. **Tweet Input:** Tweets are ingested as raw text.
2. **Embedding Generation:** We use all-MiniLM-L6-v2 from Sentence-Transformers to encode semantic vectors.
3. **Binary Classification:** Fine-tuned BERT/BiLSTM classify tweets as informative or non-informative.
4. **Multiclass Classification:** If informative, tweets are categorized into classes like earthquake, flood, etc.
5. **Event Matching:** Embeddings are compared with existing summaries in a FAISS vector database. If matched, the tweet updates the event; else, it initiates a new one.
6. **Structured Extraction:** GPT-4 extracts structured data (location, casualties) and generates an event summary in JSON.
7. **JSON + Index Storage:** Updates metadata and FAISS index for future queries.
8. **Graph Visualization:** Events are visualized using `networkx` or `pyvis`.

4 Experimental Setup

4.1 Data

We use the MMD-Crisis V2 dataset, a publicly available multimodal dataset focused on disaster response. Our work uses only the text modality, comprising over 20,000 labeled tweets annotated with informativeness and crisis categories.

4.2 Evaluation Metrics

Binary classification performance is evaluated using Accuracy, ROC AUC, and F1 Score. Multiclass classification is evaluated with per-class accuracy and macro F1.

4.3 Comparison Methods

We compare BiLSTM with fine-tuned BERT models. We also evaluate performance differences with and without the semantic event matcher (FAISS).

5 Results

- Binary Classification: BERT achieved ROC AUC = 90%, Accuracy = 81.4%, F1 = 79.9%
- Multiclass Classification: Accuracy = 72%, Confusion matrix highlights difficulty between overlapping categories
- FAISS reduced redundancy and improved event clustering
- GPT-4 outputs were accurate and easily parsable into structured formats

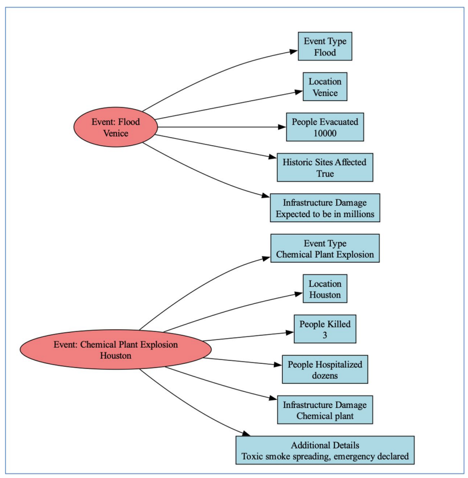
figure 

Figure 1: Enter Caption

6 Conclusion

We developed an end-to-end pipeline combining classification, retrieval, and structured information extraction for disaster tweet understanding. This approach is modular, scalable, and leverages the best of modern NLP: from BERT and FAISS to GPT-4.

7 Future Work

Future improvements include:

- Real-time tweet ingestion using streaming APIs
- Multilingual support for broader global applicability
- Integration with alert systems to provide actionable insights

References