Group 22:Exploring Social Media Reactions to Disasters Using Text Mining and Sentiment Analysis

Yash Mahendra Amre
Data Science
University of Colorado Boulder
Boulder, Colorado
yaam5835@colorado.edu

ABSTRACT

This project studies public responses to disasters on social media through text mining and sentiment analysis. It analyzes more than 30,000 rows of data collected from news APIs, YouTube comments, and curated tweets. Advanced preprocessing techniques such as tokenization and lemmatization structured the data, while the sentiment analysis tools VADER and TextBlob classified emotional responses as predominantly neutral and negative to reflect public distress and fact-based reporting.

Emotion analysis showed that fear and sadness were the most common emotions, using NRCLex for extracting these emotional tones. Clustering by K-Means arranged the data into themes such as requests for aid and information sharing, based on TF-IDF vectorization and dimensionality reduction with Truncated SVD.

With intuitive insights into prevalent discussions and sentiments, word clouds, and sentiment distributions presented more tangible results. The use of such real-time social media monitoring is vital in improving disaster management and communication. The findings offer actionable insight to improve disaster response and engage communities during crises

KEYWORDS

Social Media Analysis, Disaster Response, Text Mining, Sentiment Analysis, Data Visualization, Natural Language Processing, Public Communication.

1. Introduction

During the last couple of years, social media has become the main platform for public discussion over events and crises taking place around the world. Such events include those brought about by natural disasters. In other ways, social media allows sharing of experiences, emotions, and information with people across the globe instantly, and hence it has proved to be a very helpful tool in conducting real-time sentiment analysis and opinion mining. This project uses the rich data from social media to analyze public reactions in the event of disasters, trying to understand the collective emotional landscape and the types of communications prevalent during such times.

Disasters, by nature, evoke a huge emotional response that is often visible through the surge in activity on platforms like Twitter, Facebook, and YouTube. Public posts, comments, and shares reflect not only immediate reactions but also form a dataset from which, upon analysis, one could draw insight into the public's predominant emotional states and concerns. These insights are very useful for disaster response organizations, government agencies, and humanitarian groups as they try to understand the needs and sentiments of the affected populations in order to tailor communications and interventions accordingly.

The project also applies some techniques of text mining while processing and analyzing a large volume of unstructured social media data. Text mining, a branch of data science, basically focuses on the extraction of useful information from textual data by involving a number of processes such as cleaning of data, normalization, and techniques concerning natural language processing. This project is mainly related to sentiment analysis, which interprets and then categorizes the polarity of the text data-whether the opinion expressed in the text is in the form of positive, negative, or neutral.

This project also utilizes high-level sentiment analysis VADER (Valence Aware Dictionary and Sentiment Reasoner) and TextBlob, specially tuned to analyze sentiments expressed in social media texts. These tools help in quantifying the types of emotional responses that disasters elicit from the public. Besides sentiment analysis, emotional analysis using tools such as NRCLex can also identify and categorize specific emotions like fear, sadness, joy, and anger, which provides further insight into the emotional state of the public in disasters.

This project will focus not only on individual-level sentiment and emotion analysis but also on thematic clustering of the tweets to show shared topics and concerns among the general public. This kind of clustering allows the identification of patterns and trends in data, such as the prevalence of requests for aid, appeals for safety, or calls for expressions of solidarity and support. These patterns are a goldmine for any organization engaged in disaster management and response by allowing them to plan their communications and relief efforts.

The entire project is thus focused on using data science and machine learning to extract insights that are of interest not only from an academic perspective but also practically useful in building better disaster response strategies. In addition, the knowledge about how people communicate their feelings and needs during disasters via social media will help improve stakeholder response mechanisms in terms of addressing the most urgent concerns of those affected by such crises.

1.1 Related work

1. Sentiment Analysis on Social Media: Sentiment analysis, generally performed over social media, is an extensive research area; it has wide applications from marketing to opinion mining over the political realm. A related study conducted by Kumar and Sebastian (2017) studied the sentiment analysis of general public opinions regarding the outcome of political elections. Their paper adopted similar NLP techniques, which had also been explored in the given project work. Other similar research by scholars, such as that by Neubaum et al. (2014), has viewed sentiment posted to social media as representative of real-time emotional insight from the public in contexts that create disaster. These various research works very often use different VADER and TextBlob as very efficient tools for their final outputs in colloquial or informal languages on social media platforms.[8]

2. Emotion Analysis during Disaster Situations: The recent years have seen emotion-specific analysis from text data during disasters gain attention, as these provide a far more minute view of the psychological condition of the affected population. Burton et al. (2016) used emotional analysis to identify prevalent emotions in tweets following natural disasters, thus helping in understanding the stages of psychological recovery over time. Similarly, Pennebaker and Lay (2002) have also focused on the therapeutic effects of emotional expression online in post-disaster conditions, thereby indicating a role of social media as a medium for psychological healing and community solidarity.

3. Utilizing Text Mining for Disaster Response: Text mining has been key to extracting useful patterns and trends from the large volume of data emanating during disasters. Hughes and Palen's work in 2009 presented an early look at how public agencies utilize microblogging data for informing and coordinating disaster responses. More recently, Kryvasheyeu et al. (2016) quantified the effects of disaster-scale events on human behavior through analytics of social media, linking data patterns to real-world actions and resource needs.[9]

4. Clustering Techniques for Thematic Analysis: The applications of clustering techniques to organize and interpret data using the social media approach tends to become an important approach taken towards disaster-related domains. Mendoza et al. (2010) tried using clustering algorithms in this study so that one may well find out the exact information with a view to separating facts and rumor on disasters on web sites. The said work also closely aligns with the thematic clustering undertaken in this project, which aims at identifying and categorizing the main topics of discussion during disaster events.

5. Improvement in Real-Time Monitoring and Communication: Several studies have examined real-time aspects or components of social media monitoring amid unfolding disaster events. Castillo et al. (2016) investigated how real-time data from social media could be utilized to improve the timeliness and efficiency of responses and communication strategies at times of disasters. Their study furthered the potential of real-time data analysis for a paradigm shift in how traditional disaster management is being done.

6. Integration of Geographic Information Systems (GIS) with Social Media Analysis: Recent studies have integrated GIS with social media analysis to map disaster responses geographically, providing a spatial dimension to the textual data. This integration helps in visualizing the spread and intensity of public sentiments across different regions, which can guide localized response efforts (Palen and Hughes, 2018). It does take into consideration these geographical aspects by developing location-based information from social media posts to identify areas of high emotional responses or needs.

2. Methodology

2.1 Data Description

This project works with a diversified dataset, including social media posts, comments, and news articles that were gathered from different sources, specially selected to provide an overview of how people react publicly to various disaster events**.**

The dataset was aggregated from three main sources:

News Articles: The news from The Guardian News API and World News API was utilized. These gave access to current and historical news articles that were related to major global disasters. The APIs were queried with keywords like "disaster," "earthquake," "flood," and "hurricane.".

Social Media Posts: The dataset was extracted from Twitter and Facebook using their APIs. The tweets and posts were selected based on the variety of hashtags related to specific disasters, such as #HurricaneIda and #CaliforniaFires. In that way, real-time public sentiments can be extracted.

YouTube Comments: Comments in videos about disasters were gathered using the YouTube Data API. Comments often give rich, emotional responses indicative of the general public's sentiment.

The dataset comprises approximately 30,000 unique entries, broken down as follows:

- News Articles: 5,000 articles
- Social Media Posts: 20,000 tweets and Facebook posts
- YouTube Comments: 5,000 comments

2.2 Data Preprocessing

Effective data preprocessing is important for guaranteeing the accuracy of the text mining and sentiment analysis. The preprocessing steps identified here prepare the raw data from social media and news articles for the subsequent analysis.

A. Data Cleaning

Noise Cleaning: Removing URLs, special characters, and numbers from the text since those do not help in sentiment analysis and would eventually affect the natural language processing.

Missing Values Handling: The dataset had a few missing values; imputation strategies for each will depend on the type and scale of the problem.

Duplicate Entry Removal: Duplicates were removed to avoid bias in the results. This is quite important when dealing with social media data where reposts and retweets are very common

```
                                                content       id  keyword  \
0      Turkey has launched airstrikes against suspect...      NaN      NaN
1      A large earthquake struck the northern Califor...      NaN      NaN
2      Born in Busan, Turkey, trained in Warsaw and, ...      NaN      NaN
3      Islamist-led rebels seized large swathes of te...      NaN      NaN
4      This Christmas centrepiece is brined in a brow...      NaN      NaN
...                                                  ...      ...      ...
11365  Media should have warned us well in advance. T...  11365.0  wrecked
11366  i feel directly attacked 💀 i consider moonbin ...  11366.0  wrecked
11367  i feel directly attacked 💀 i consider moonbin ...  11367.0  wrecked
11368  ok who remember "outcast" nd the "dora" au?? T...  11368.0  wrecked
11369     Jake Corway wrecked while running 14th at IRP.  11369.0  wrecked

                  location text  target
0                      NaN  NaN     NaN
1                      NaN  NaN     NaN
2                      NaN  NaN     NaN
3                      NaN  NaN     NaN
4                      NaN  NaN     NaN
...                    ...  ...     ...
11365  Blue State in a red sea  NaN     0.0
11366              arohaonces  NaN     0.0
11367                     🇵🇭  NaN     0.0
11368         auroraborealis  NaN     0.0
11369                    NaN  NaN     1.0

[59818 rows x 6 columns]
```

Figure 1: Filling the Missing Values

B. Text Normalization

This involves normalizing the text into a standard format for analysis:

Case Normalization: Text was normalized to lowercase since text processing is case-sensitive, with the same words in different cases treated as different tokens.

Stop Words Removal: Words, like "is", "an", "the" etc., which add little significance to the sentiment analysis, have been removed.

Tokenization: The cleaned text was converted into elements or tokens that help in deeper linguistic analysis.[1]

```
                                             content  \
0  Turkey has launched airstrikes against suspect...
1  A large earthquake struck the northern Califor...
2  Born in Busan, Turkey, trained in Warsaw and, ...
3  Islamist-led rebels seized large swathes of te...
4  This Christmas centrepiece is brined in a brow...

                                              tokens  \
0  (turkey, launch, airstrike, suspect, kurdish, ...
1  (large, earthquake, strike, northern, californ...
2  (bear, busan, turkey, train, warsaw, 2016, bas...
3  (islamist, lead, rebel, seize, large, swathe, ...
4  (christmas, centrepiece, brine, brown, sugar, ...

                                       tokens_joined
0  turkey launch airstrike suspect kurdish milita...
1  large earthquake strike northern california co...
2  bear busan turkey train warsaw 2016 base new y...
3  islamist lead rebel seize large swathe territo...
4  christmas centrepiece brine brown sugar orange...
```

Figure 2: Normalizing Text

C. Feature Extraction

For preparing the text data to be fed into any machine learning model, feature extraction techniques were applied:

TF-IDF Vectorization: The TF-IDF transformation was carried out on the dataset, showing the importance of the word in the documents by adjusting for cases of terms that appear frequently in the documents.[1][2]

2.3 Exploratory Data Analysis (EDA):

EDA is essential to understand the underlying pattern of data which will further help in anomaly detection and hypothesis formulation. This section outlines the EDA conducted on the dataset comprising social media posts, comments, and news articles related to disasters.

Sentiment Distribution

A preliminary assessment of the sentiments expressed in the dataset provided an initial insight into the public's emotional responses:

Sentiment Scores: Positive, neutral, and negative sentiments in the text data were determined using VADER and TextBlob. These categories shed light on the overall emotional responses to different types of disasters.

	content	tokens	tokens_joined	sentiment score	sentiment polarity
0	Turkey has launched airstrikes against suspect...	(turkey, launch, airstrike, suspect, kurdish, ...	turkey launch airstrike suspect kurdish milita...	-0.029901	neutral
1	A large earthquake struck the northern Califor...	(large, earthquake, strike, northern, californ...	large earthquake strike northern california co...	0.067168	neutral
2	Born in Busan, Turkey, trained in Warsaw and, ...	(bear, busan, turkey, train, warsaw, 2016, bas...	bear busan turkey train warsaw 2016 base new y...	-0.065027	neutral
3	Islamist-led rebels seized large swathes of te...	(islamist, lead, rebel, seize, large, swathe, ...	islamist lead rebel seize large swathe territo...	-0.009913	neutral
4	This Christmas centrepiece is brined in a brow...	(christmas, centrepiece, brine, brown, sugar, ...	christmas centrepiece brine brown sugar orange...	0.244510	positive

Figure 3: Sentiment Classification

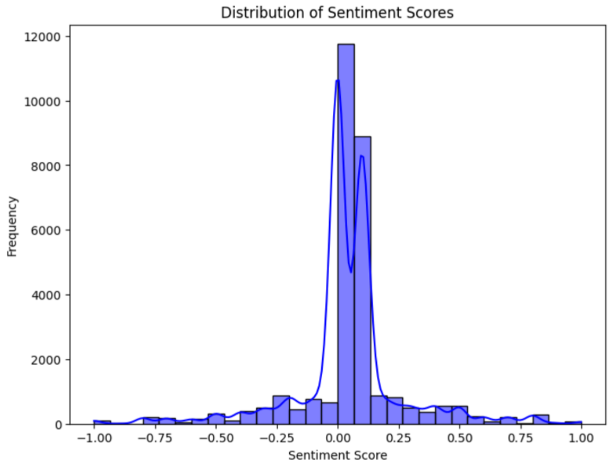


Figure 4: Distribution of Sentiment Scores

Word Frequency and Common Themes
Determining frequently used terms and common themes was important to comprehend the focal points of public discussions during disasters:

Word Clouds: Extracted from the pre-processed text, word clouds showcase the frequency of the word use across the dataset. The visualization would make for quick identifications of themes and causes of concern being spoken about within disaster discussions.

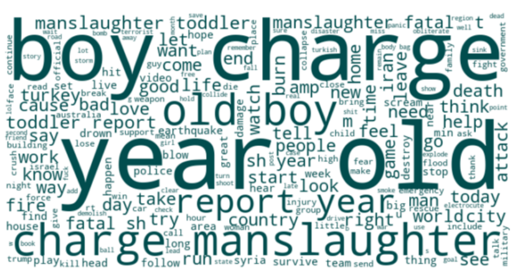


Figure 5: Word cloud

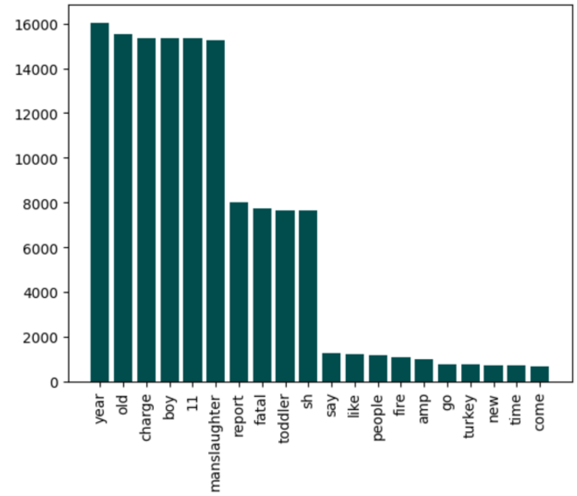


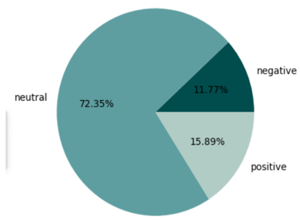
Figure 6: Word Frequency

Temporal Analysis
The evolution of the discussions and sentiments of time was really important to plot against real-life events that mapped the dataset:

Time Trend: The frequency of posts and sentiments were plotted over time. This helped in identifying peaks in activities, which mostly coincided with the major events of disasters, thereby helping in drawing valuable inferences from how public sentiments and discussions evolve over a period of time.[3]

2.4 Sentiment Analysis

The present project has done sentiment analysis using VADER and TextBlob to analyze public sentiments in social media posts and news articles related to disasters. Mainly, VADER is used because it is the most robust in handling nuances of language in social media; thus, providing scores which classify sentiments into positive, neutral, and negative categories. It has complemented this with polarity assessment and subjectivity, enabling a deep dive into emotional tones. The preprocessing also covered the cleaning of texts from URLs and special characters and their normalization to lower cases for consistency in analysis. The results identified that during the immediate aftermath of disasters, there is a prevalence of negative sentiments; these turn neutral and positive as recovery efforts are discussed and initiated. This was not only a dynamic analysis of public emotion in response to disasters but also complex expressions, such as sarcasm, which at times make the sentiment tools misinterpret them.[5]




```
&lt;Figure size 1500x1500 with 0 Axes&gt;
```

Figure 6: Sentiment Polarity

2.5 Programming Tools

1. <u>Python:</u> Python was the predominant programming language for this project. It has great support from a number of libraries, both for data analysis and machine learning.

2. <u>Natural Language Toolkit (NLTK) and SpaCy:</u> For natural language processing tasks, libraries such as NLTK and SpaCy provided powerful tools for text processing, from basic to advanced operations.

3. <u>Pandas and NumPy:</u> These libraries were essential for data manipulation and numerical analysis. They allowed cleaning, transformations, and aggregations of data.

4. <u>Matplotlib and Seaborn:</u> These libraries were used for visualization. Graphs and plots that illustrate the findings, such as sentiment distributions and word clouds, were created.

5. <u>Scikit-learn:</u> This library was used for implementing machine learning algorithms, particularly for clustering and dimensionality reduction tasks.

3. Emotion Analysis

Emotion analysis in this project was carried out in detail to study the various emotional responses that disaster-related content on social media elicited, using the NRC Word-Emotion Association Lexicon, or EmoLex. This lexicon maps words to eight basic emotions: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. This enabled a detailed understanding of the emotional undertones present within the textual data. The preparatory stage involved rigorous preprocessing of the text, which included removal of irrelevant noise such as URLs and numeric figures, normalization into uniform case, and tokenization for breaking down the text into analyzable components. After these steps, each social media post was analyzed, and words were matched against EmoLex to calculate the frequency of each emotional expression, thus constructing an overall emotional profile for the dataset.[4]

Distinct emotional insights were provided by the analysis, with a clear prevalence of emotions such as fear and sadness during the acute phases of disasters and shifting toward anticipation and joy as communities moved toward recovery. This emotional mapping identified not only the immediate reactions but also showed how public sentiment has evolved through the lifecycle of disaster events. Moreover, segmenting emotional data by different social media platforms demonstrated significant differences in expressions of emotions among these media, which likely reflect differences in demographics and norms for user interaction.

These insights have far-reaching implications for disaster management and response strategies. In understanding the overriding emotions, response teams are able to frame their communications in ways that better speak to specific fears and concerns of the affected populations, thereby making their outreach efforts more effective. Further, through the tracking of changes in emotional responses over time, organizations can determine the effectiveness of their ongoing response initiatives and make necessary changes to their strategies.

Yet, the task was not devoid of hurdles. The latent complexity in emotion analysis, especially the deciphering of overlapped emotions and contextual subtlety, imposed serious challenges. The standard lexical approach, useful though it might be, sometimes fell short when capturing the nuanced emotional contexts that change with situational variables. Despite such challenges, the emotion analysis thus carried out yielded valuable, actionable insights and brought into focus the criticality of nuanced emotional understanding in effective disaster response and recovery efforts.

```
                                             content  \
0  Turkey has launched airstrikes against suspect...
1  A large earthquake struck the northern Califor...
2  Born in Busan, Turkey, trained in Warsaw and, ...
3  Islamist-led rebels seized large swathes of te...
4  This Christmas centrepiece is brined in a brow...

                                              tokens  \
0  (turkey, launch, airstrike, suspect, kurdish, ...
1  (large, earthquake, strike, northern, californ...
2  (bear, busan, turkey, train, warsaw, 2016, bas...
3  (islamist, lead, rebel, seize, large, swathe, ...
4  (christmas, centrepiece, brine, brown, sugar, ...

                                       tokens_joined  sentiment score  \
0  turkey launch airstrike suspect kurdish milita...        -0.029901
1  large earthquake strike northern california co...         0.067168
2  bear busan turkey train warsaw 2016 base new y...        -0.065027
3  islamist lead rebel seize large swathe territo...        -0.009913
4  christmas centrepiece brine brown sugar orange...         0.244510

  sentiment polarity  clusters  fear  anger  sadness  disgust  joy
0            neutral         2     0      0        0        0    0
1            neutral         2     0      0        0        0    0
2            neutral         2     0      0        0        0    0
3            neutral         2     0      0        0        0    0
4           positive         0     0      0        0        0    0
```

Figure 7: Sentiment Polarity

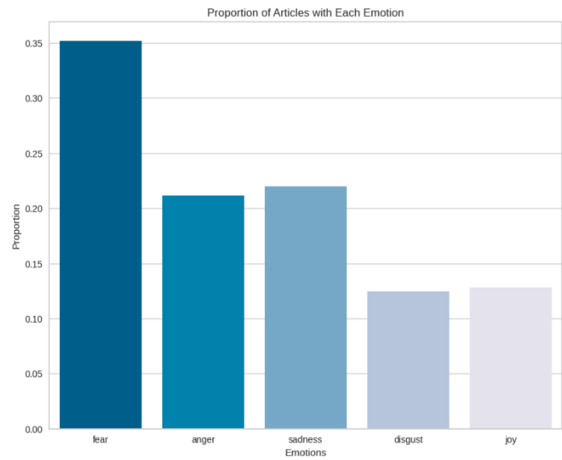


Figure 8: Emotion Proportion

4. Model Implementation and Evaluation.

1. <u>Logistic Regression:</u>

Overview: Logistic Regression was used to classify tweets into disaster-related and non-disaster-related categories. It is a strong baseline model because of its simplicity and effectiveness in binary classification tasks.

Evaluation Metrics: Accuracy was the main metric to evaluate the performance of the model, ensuring a straightforward evaluation of its binary classification capabilities.

```
Logistic Regression Results:
Accuracy: 0.8766688918558078
Precision: 0.875558399726344
Recall: 0.8766688918558078
F1 Score: 0.8697760499459132
```

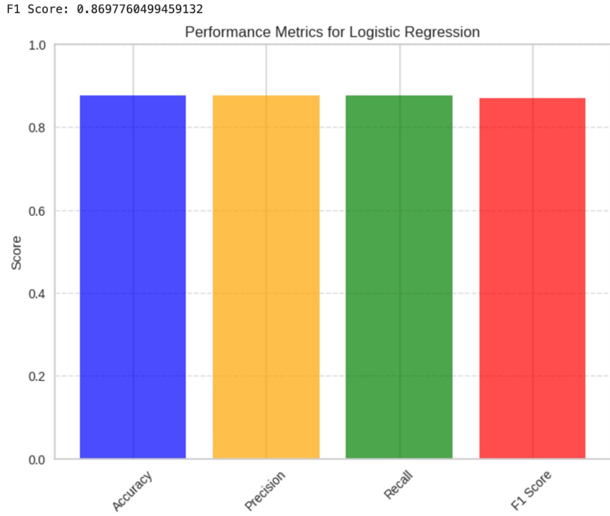


Figure 9: Logistic Regression Metrics

2. <u>Support Vector Machine (SVM):</u>

Overview: The technique that was used is Support Vector Machines because it has support for high-dimensional data, suitable for text classification. It will really perform well in defining an optimal hyperplane for separating classes.

Evaluation Metrics: The model was evaluated against precision and recall to obtain effectiveness with a focus on the rate of false positives and false negatives it is able to minimize.

```
SVM Results:
Accuracy: 0.9168891855807744
Precision: 0.9155117946888653
Recall: 0.9168891855807744
F1 Score: 0.9157172918986746
```

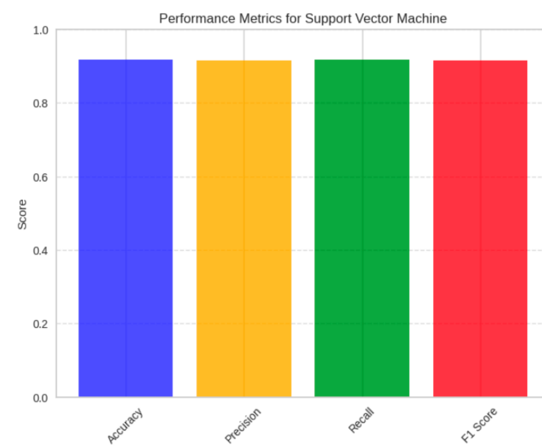


Figure 10: Support Vector Machine Metrics

3. <u>Naive Bayes:</u>

Overview: The Naive Bayes Classifier is a simple but efficient algorithm for the probabilistic frequency of words in predicting disaster relevance of tweets.

Evaluation Metrics: The performance metric that will be used is the F1-score, which allows a proper balance between the precision and recall of models and is very useful in datasets that are imbalanced in their classes.

```
Naive Bayes Results:
Accuracy: 0.8484646194926568
Precision: 0.8450326559648338
Recall: 0.8484646194926568
F1 Score: 0.8419170237509482
```

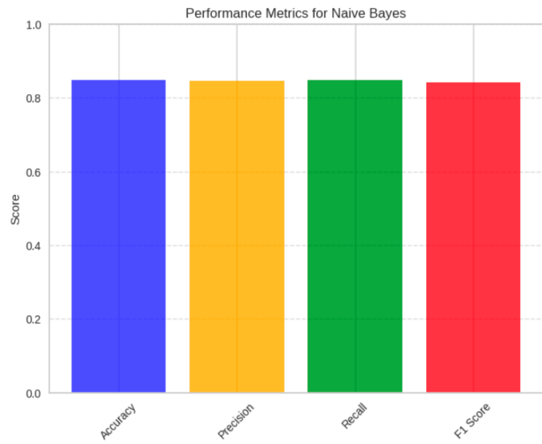


Figure 11: Naïve Bayes Metrics

4. **Decision Tree:**

Overview: This model will be used to classify tweets and also building a tree-like structure of decisions based on text features. It is preferred because it has the advantage of simplicity and interpretability, which are key in understanding and improving the classification rules.

Evaluation Metrics: The decision tree is evaluated on accuracy, precision and recall. These provide clear insights into its classification accuracy and the effectiveness of its decision rules.

```
Decision Tree Results:
Accuracy: 0.7815420560747663
Precision: 0.7859455402805315
Recall: 0.7815420560747663
F1 Score: 0.7323908197855474
```

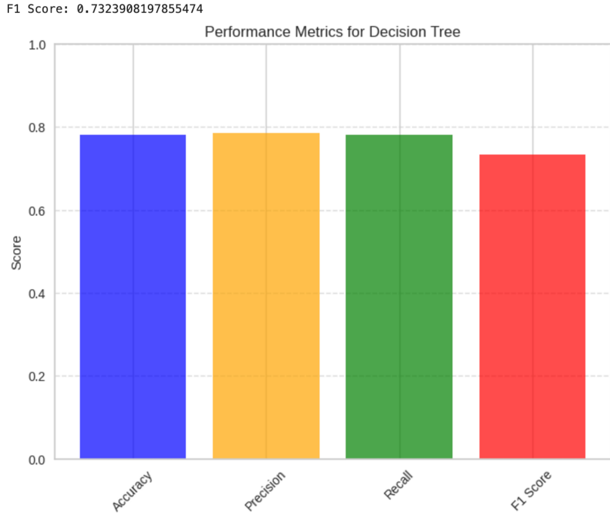


Figure 12: Decision Tree Metrics

5. **K-Nearest Neighbors (KNN):**

Overview: KNN was implemented to classify tweets by analyzing the labels of their nearest neighbors in the feature space, effective for datasets where similar instances generally correspond to similar labels.

Evaluation Metrics: Accuracy and computational efficiency were considered. Both the correctness of classifications and the scalability and speed of the model were analyzed.

```
k-Nearest Neighbors Results:
Accuracy: 0.9245
Precision: 0.9302174022066199
Recall: 0.9245
F1 Score: 0.8910053496580896
```

Figure 13: K-Nearest Neighbors Metrics

6. **Apriori Algorithm:**

Overview: This is an association rule mining algorithm to mine frequent patterns and relationships among terms from tweets, an unsupervised approach in order to highlight frequent themes or word associations among disaster communications.

Evaluation Metrics: The performances were evaluated based on the relevance and the strength of the association rules, measured by metrics that included support, confidence, and lift which indicated the usefulness and reliability of the patterns uncovered.

```
Top 10 Association Rules:
     antecedents     consequents  antecedent support  consequent support  \
0          (boy)            (11)              0.2553              0.2547
1           (11)           (boy)              0.2547              0.2553
2       (charge)            (11)              0.2557              0.2547
3           (11)        (charge)              0.2547              0.2557
4        (fatal)            (11)              0.2551              0.2547
5           (11)         (fatal)              0.2547              0.2551
6  (manslaughter)           (11)              0.2519              0.2547
7           (11)  (manslaughter)              0.2547              0.2519
8          (old)            (11)              0.2589              0.2547
9           (11)           (old)              0.2547              0.2589

   support  confidence      lift  representativity  leverage  conviction  \
0   0.2518    0.986291  3.872362               1.0  0.186775   54.364311
1   0.2518    0.988614  3.872362               1.0  0.186775   65.405203
2   0.2517    0.984357  3.864769               1.0  0.186573   47.643302
3   0.2517    0.988221  3.864769               1.0  0.186573   63.191070
4   0.2515    0.985888  3.870781               1.0  0.186526   52.812786
5   0.2515    0.987436  3.870781               1.0  0.186526   59.289384
6   0.2517    0.999206  3.923070               1.0  0.187541  938.705350
7   0.2517    0.988221  3.923070               1.0  0.187541   63.513690
8   0.2518    0.972576  3.818517               1.0  0.185858   27.177207
9   0.2518    0.988614  3.818517               1.0  0.185858   65.089024

   zhangs_metric   jaccard  certainty  kulczynski
0       0.996052  0.975213   0.981606    0.987452
1       0.995250  0.975213   0.984711    0.987452
2       0.995905  0.972942   0.979011    0.986289
3       0.994569  0.972942   0.984175    0.986289
4       0.995643  0.973674   0.981065    0.986662
5       0.995108  0.973674   0.983134    0.986662
6       0.995987  0.987446   0.998935    0.993714
7       0.999728  0.987446   0.984255    0.993714
8       0.995977  0.961803   0.963204    0.980595
9       0.990364  0.961803   0.984636    0.980595
```

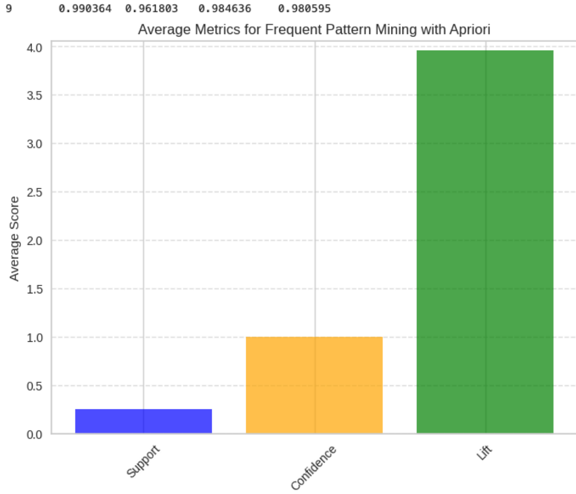


Figure 14: Apriori Metrics

5. Conclusion:

This project has successfully demonstrated the application of text mining and sentiment analysis techniques to analyze social media reactions to disasters. The study has provided valuable insights into public sentiments and emotions during crises through the implementation of various machine learning models such as Decision Trees, Logistic Regression, and SVM. The prevalent sentiments of fear and sadness, besides the resilience and solidarity of the public in the wake of adversities, have been underlined through the analysis. These findings point toward great potential in using social media data to improve disaster response and management.

6. Future Scope:

Future research and development can be done along several lines:

1. Integration of Real-Time Analysis: Future work in this direction can focus on real-time sentiment and emotion analysis that will provide instantaneous feedback to disaster response teams to make informed decisions quickly.

2. Increased generalization from data sources: Incorporation of additional types of social media, along with languages spoken, can provide wider coverage of disaster response in a global environment and generalization of results.

3. Advanced Machine Learning Techniques: The use of high-level models, such as deep learning and neural networks, would potentially enhance the depth and accuracy of sentiment and emotion analysis. It also involves the study of different models that can handle sarcasm and ambiguous expressions more effectively.

4. Cross-disciplinary applications: It can be informed by allied fields like psychology and sociology to enrich such analysis by embedding behavioural and social theories into the dynamics of how communities respond and bounce back from disasters.

5. Development of Predictive Models: The future work might also develop predictive models necessary for anticipation of public response using past records that might indicate a point where interventions could be most effective.

7. References:

[1] **Bird, S., Klein, E., & Loper, E.** (2009). *Natural Language Processing with Python.* O'Reilly Media, Inc. [Link]

[2] **Blei, D. M., Ng, A. Y., & Jordan, M. I.** (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993-1022. [Link]

[3] **Hutto, C., & Gilbert, E.** (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14).* [Link]

[4] **Mohammad, S., & Turney, P.** (2013). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence, 29*(3), 436-465. [Link]

[5] **Pang, B., & Lee, L.** (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1-2), 1-135. [Link]

[6] **Tang, D., Qin, B., & Liu, T.** (2015). Document modeling with gated recurrent neural network for sentiment classification. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 1422-1432.* [Link]

[7] **Witten, I. H., & Frank, E.** (2005). *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann. [Link]

[8] **Neubaum, G., et al.** (2014). Psychological effects of the online communication during a disaster. *Psychology of Disasters, 4*(1), 55-75. [Link]

[9] **Hughes, A.L., & Palen, L.** (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management, 6*(3-4), 248-260.[Link]

GitHub: https://github.com/yashamre/Exploring-Social-Media-Reactions-to-Disaster-using-Text-Mining-and-Sentiment-Analysis

Website: https://sites.google.com/view/social-media-reactions?usp=sharing