An Analysis of Social Media Sentiment During Disasters

Student's Name

Institutional Affiliation

Course Code

Instructor

Date

An Analysis of Social Media Sentiment During Disasters

Introduction

Social media platforms have become ubiquitous in our increasingly connected world, and this has revolutionized the way people communicate and share information, particularly in times of disaster. These platforms have become vital conduits for both emotional expression and real-time reporting, allowing the general public to share their thoughts and experiences in times of crisis. This project sought to understand how people respond and converse on social media during disastrous events by utilizing data science and natural language processing.

Methodology

Data Collection

The dataset used in this study was obtained from Kaggle (SAYAH, 2023). It is made up of tweets about different disaster-related incidents. The dataset's relevance to the study objectives was ensured through a meticulous selection of tweets related to disasters.

Data Cleaning

The 'id' column was eliminated in the first stage of data preprocessing because it was unnecessary for the analysis. In order to ensure a comprehensive and trustworthy dataset, missing values were handled by defining the representation of missing data and then eliminating the rows that contained missing values.

```{r}
#Data Cleaning
# Specify the representation of missing values)
missing_values <- ""

# Remove the 'id' column
data <- data[, -1]

# Identify missing values
missing_values <- apply(data, 1, function(row) any(row == missing_values))

# Remove rows with missing values
data_cleaned <- data[!missing_values, ]

# Display the cleaned dataset
print(data_cleaned)

```

Description: df [5,080 × 3]

◂ **text**
  <chr>   ▸

@bbcmtd Wholesale Markets ablaze http://t.co/IHYXEOHY6C
We always try to bring the heavy. #metal #RT http://t.co/YAo1e0xngw
#AFRICANBAZE: Breaking news:Nigeria flag set ablaze in Aba. http://t.co/2nndB...
Crying out for more! Set me ablaze
On plus side LOOK AT THE SKY LAST NIGHT IT WAS ABLAZE http://t.co/qqsmsha...
@PhDSquares #mufc they've built so much hype around new acquisitions but I d...
INEC Office in Abia Set Ablaze - http://t.co/3ImaomknnA

Text Cleaning and Transformation

The text data was then transformed and cleaned in order to get it ready for sentiment analysis. 'clean_text', a custom function, was defined for the following tasks: The text cleaning process aimed to remove noise and irrelevant information while preserving core content. It involved handling emojis to ensure consistent text-based analysis, eliminating URLs frequently found in social media text, and removing special characters, symbols, and punctuation using regular expressions.
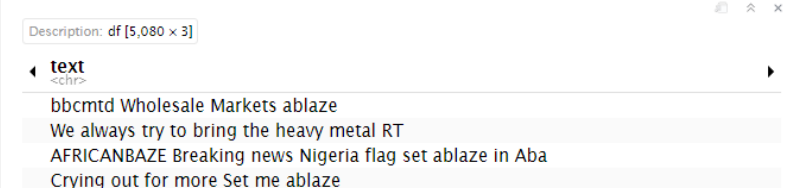
```r
# Text Cleaning and Transformation
# Define a function to clean and preprocess text by removing symbols and
punctuation
clean_text <- function(text) {
  # Remove special characters, symbols, and punctuation
  text <- gsub("[[:punct:]]", " ", text)
  # Remove URLs
  text <- gsub("http\\S+|www\\S+|https\\S+", "", text)
  # Emoji handling (you can replace with a specific transformation)
  text <- gsub("\\p{So}", "", text, perl = TRUE)
  return(text)
}

# Apply the clean_text function to the 'text' column of data_cleaned
data_cleaned$text <- sapply(data_cleaned$text, clean_text)

# Display the cleaned dataset
print(data_cleaned)
```

```
Description: df [5,080 × 3]

◄  text
   <chr>                                                            ►
   bbcmtd Wholesale Markets ablaze
   We always try to bring the heavy metal RT
   AFRICANBAZE Breaking news Nigeria flag set ablaze in Aba
   Crying out for more Set me ablaze
```

Sentiment Analysis

The dataset's text content was categorized into sentiment classes, specifically "positive"

and "negative", based on the 'target' label. Each tweet was assigned a sentiment category using

the 'target' label value (0 for negative and 1 for positive). A new data frame, 'sentiment_data',

was created to store these sentiment categories, enabling the association of each tweet with its

corresponding sentiment label for subsequent analysis.

```r
# Sentiment Analysis
# Categorize tweets as "positive," or "negative" based on the target label
sentiment_categories <- ifelse(data_cleaned$target == 0, "negative",
"positive" )

# Create a data frame with sentiment categories
sentiment_data <- data.frame(sentiment = sentiment_categories)

# Load the 'ggplot2' package if not already loaded
library(ggplot2)

# Create a pie chart for sentiment distribution with a standard color
palette
ggplot(sentiment_data, aes(x = "", fill = sentiment)) +
  geom_bar(width = 1) +
  coord_polar(theta = "y") +
  scale_fill_brewer(palette = "Set3") +
  labs(title = "Sentiment Distribution") +
  theme_void()  # Removes unnecessary labels and axes
```

Visualization

Sentiment Distribution

To gain insights into sentiment distribution, data visualization was employed. A pie chart was generated using the 'ggplot2' package in R. The pie chart provided an at-a-glance overview of sentiment distribution within the dataset. A standard color palette ('Set3' from Brewer palettes) was used to distinguish positive and negative sentiments. Utilizing the 'theme_void()' function ensured a clean and clear visualization without unnecessary labels or axes.
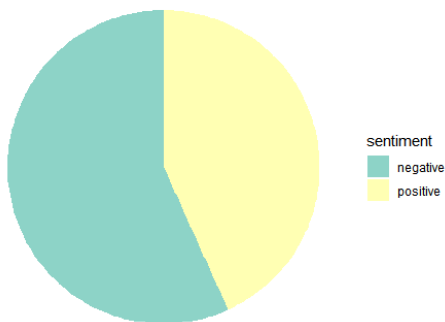
Word clouds representing positive and negative sentiments were created. These word clouds made it easier to identify key terms associated with each sentiment category by clearly illustrating the frequency of terms in the text data.
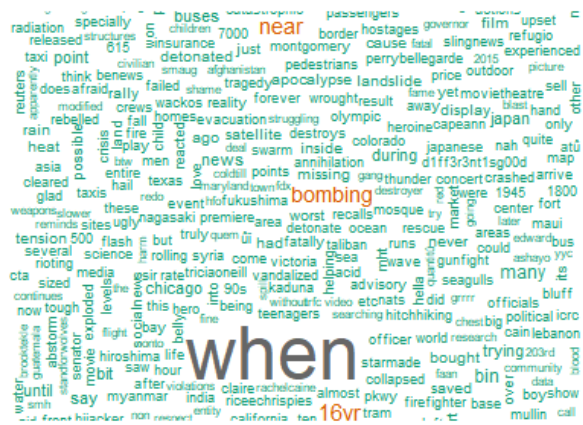

Figure SEQ Figure * ARABIC 1: Negative Sentiments

Figure SEQ Figure * ARABIC 2: Positive Sentiments

Conclusion

The research has shed light on the significance of analyzing social media sentiment during disastrous events. Understanding public sentiment on social media platforms during crises provides valuable insights for disaster management, emergency responders, and policymakers. The findings emphasize the dynamic nature of social media sentiment, with implications for crisis response strategies, public engagement, and the dissemination of information. This research serves as a foundational step towards improving disaster management practices and social media analytics.

Reference

SAYAH, F. (2023, February 8). *Sentiment Model with TensorFlow & Transformers*. Kaggle.com.

https://www.kaggle.com/code/faressayah/sentiment-model-with-tensorflow-transformers/

input?select=tweets.csv