

# Disaster Tweet Analyzer: NLP for Crisis Communication

## Week 1 Progress

### Project Goals

1. Analyzing disaster tweets.
2. Utilizing Natural Language Processing (NLP) techniques to classify tweets.
3. Developing a system to assist in crisis communication through social media analysis.

### Dataset

- Dataset: [Disaster Tweets Dataset \(https://www.kaggle.com/datasets/vstepanenko/disaster-tweets/data\)](https://www.kaggle.com/datasets/vstepanenko/disaster-tweets/data)
- Status: Successfully downloaded and verified.

### Initial Exploration of the Dataset

1. Loading the dataset and displaying the first few rows.
2. Analyzing the dataset.

```
# Import necessary libraries
import pandas as pd

# Load the dataset
file_path = 'disaster_tweets.csv' # Replace with your file path
tweets_df = pd.read_csv('/Users/karthik/Downloads/infosys/tweets.csv')

# Display the first few rows of the dataset
tweets_df.head(15)
```

	id	keyword	location	text	target
0	0	ablaze	NaN	Communal violence in Bhainsa, Telangana. "Ston...	1
1	1	ablaze	NaN	Telangana: Section 144 has been imposed in Bha...	1
2	2	ablaze	New York City	Arsonist sets cars ablaze at dealership https:...	1
3	3	ablaze	Morgantown, WV	Arsonist sets cars ablaze at dealership https:...	1

	id	keyword	location	text	target
4	4	ablaze	NaN	"Lord Jesus, your love brings freedom and pard...	0
5	5	ablaze	OC	If this child was Chinese, this tweet would ha...	0
6	6	ablaze	London, England	Several houses have been set ablaze in Ngemsib...	1
7	7	ablaze	Bharat	Asansol: A BJP office in Salanpur village was ...	1
8	8	ablaze	Accra, Ghana	National Security Minister, Kan Dapaah's side ...	0
9	9	ablaze	Searching	This creature who's soul is no longer clarent ...	0
10	10	ablaze	NaN	Images showing the havoc caused by the #Camero...	1
11	11	ablaze	NaN	Social media went bananas after Chuba Hubbard ...	0
12	12	ablaze	NaN	Hausa youths set Area Office of Apapa-Iganmu L...	1
13	13	ablaze	HYDERABAD	Under #MamataBanerjee political violence &...	1
14	14	ablaze	Reno, NV	AMEN! Set the whole system ablaze, man. https:...	0

## Basic Text Preprocessing Steps

- Handle missing values.
- Clean up the raw tweet text (remove links, special characters, etc.).

```

# Check for missing values
missing_values = tweets_df.isnull().sum()
missing_values

# Basic text cleanup function
import re

def clean_tweet(text):
    # Remove links, special characters, and extra spaces
    text = re.sub(r'http\S+|www\S+', '', text) # Remove URLs
    text = re.sub(r'^A-Za-z0-9+', ' ', text) # Remove special characters
    text = text.strip() # Remove leading and trailing spaces
    return text

# Apply the cleanup function to the tweet text
tweets_df['cleaned_text'] = tweets_df['text'].apply(clean_tweet)

# Display the cleaned text
tweets_df[['text', 'cleaned_text']].head(15)

```

	text	cleaned_text
0	Communal violence in Bhainsa, Telangana. "Ston...	Communal violence in Bhainsa Telangana Stones ...
1	Telangana: Section 144 has been imposed in Bha...	Telangana Section 144 has been imposed in Bhai...
2	Arsonist sets cars ablaze at dealership https:...	Arsonist sets cars ablaze at dealership
3	Arsonist sets cars ablaze at dealership https:...	Arsonist sets cars ablaze at dealership
4	"Lord Jesus, your love brings freedom and pard...	Lord Jesus your love brings freedom and pardon...
5	If this child was Chinese, this tweet would ha...	If this child was Chinese this tweet would hav...
6	Several houses have been set ablaze in Ngemsib...	Several houses have been set ablaze in Ngemsib...
7	Asansol: A BJP office in Salanpur village was ...	Asansol A BJP office in Salanpur village was s...
8	National Security Minister, Kan Dapaah's side ...	National Security Minister Kan Dapaah s side c...
9	This creature who's soul is no longer clarent ...	This creature who s soul is no longer clarent ...
10	Images showing the havoc caused by the #Camero...	Images showing the havoc caused by the Cameroo...

	<b>text</b>	<b>cleaned_text</b>
<b>11</b>	Social media went bananas after Chuba Hubbard ...	Social media went bananas after Chuba Hubbard ...
<b>12</b>	Hausa youths set Area Office of Apapa-Iganmu L...	Hausa youths set Area Office of Apapa-Iganmu L...
<b>13</b>	Under #MamataBanerjee political violence &...;	Under MamataBanerjee political violence amp va...
<b>14</b>	AMEN! Set the whole system ablaze, man. https:...	AMEN Set the whole system ablaze man

## Week 1 progress

- Dataset confirmation and verified.
- Initial exploration completed, and basic text preprocessing steps initiated.