Introduction to Disaster Management and NLP

Presented By
Sai Shailesh Gadde
Madhurachanna H B

Outline

1. Problem Statement
2. Challenges in Disaster Response
3. Data Collection
4. Techniques and Methodology
5. Visualizations (Frequencies and WordClouds)
6. Performance Measurements (Accuracy before and after pre-processing)
7. Integration with Gen AI (Improve human readability)
8. Future Work

Problem Statement:

In the context of disaster management, the ability to efficiently process and respond to disaster-related messages is critical for saving lives and mitigating damage. These messages, often gathered from social media, news reports, emergency hotlines, and other sources, are typically unstructured, diverse, and contain critical information that needs to be quickly categorized into relevant response categories (e.g., medical assistance, evacuation requests, infrastructure damage, etc.).

Challenges in Disaster Response

1 Timely Information Gathering

Quickly collecting and processing data from multiple sources to gain situational awareness during a disaster.

2 Coordinating Relief Efforts

Efficient coordination between emergency responders, government agencies, and non-profit organizations.

3 Resource Allocation

Ensuring critical supplies and aid are delivered to the right locations and people in a timely manner.

Data Collection

- Hugging Face

Data Collection

This dataset was acquired through Hugging Face and has a total of **40 classes**.

- This open-source dataset contains 30,000 messages from various disaster events, including the 2010 earthquakes in Haiti and Chile, the 2010 floods in Pakistan, and super-storm Sandy in the U.S.A. in 2012.

- It also includes news articles spanning many years and hundreds of different disasters. The data has been categorized into 36 different disaster response categories and stripped of any sensitive information.

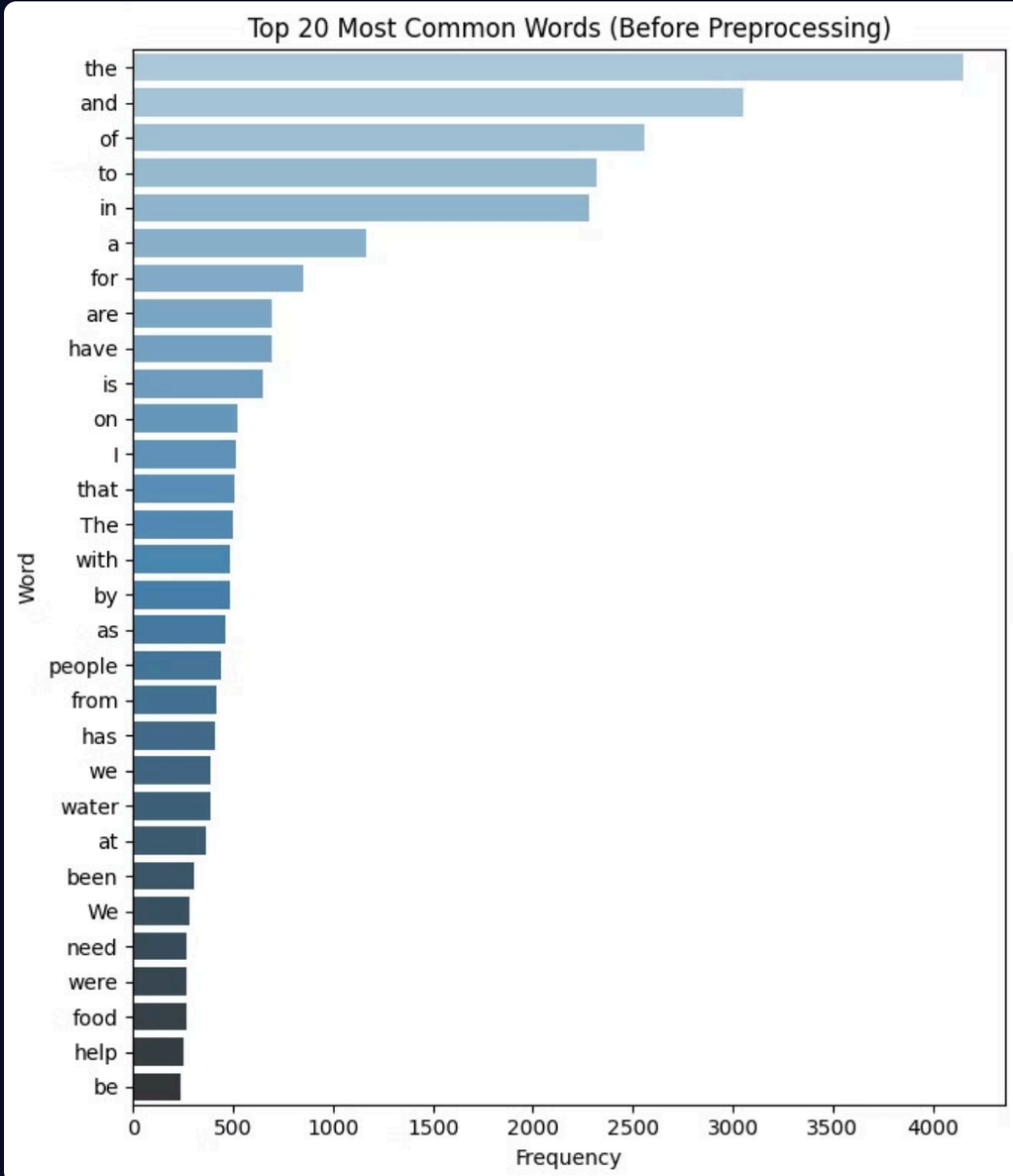
- These classes provide a comprehensive understanding of the specific context and sentiment of the messages.

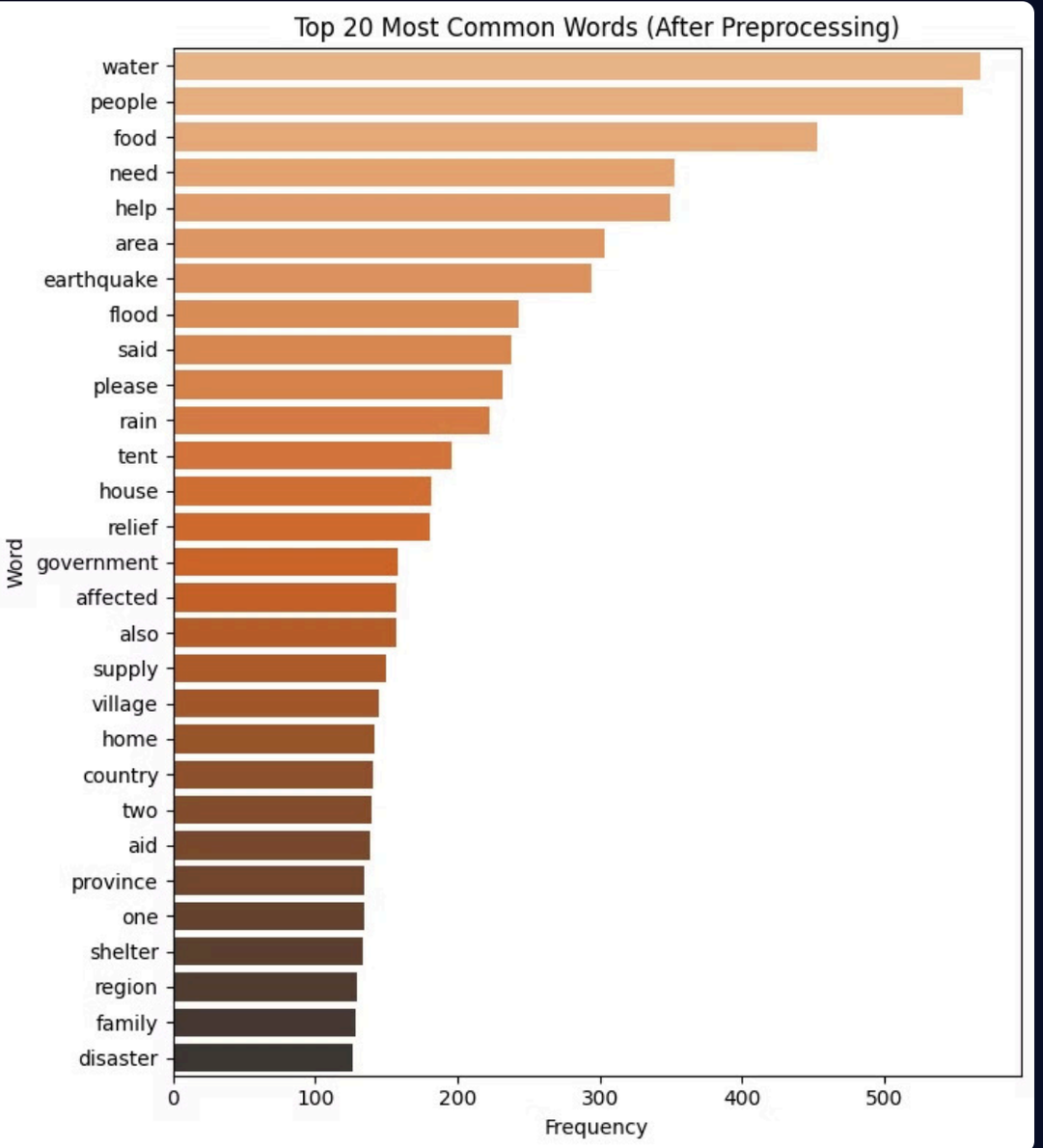
Data Field	Description	Values
related	Is the message disaster related?	1 = yes, 0 = no, 2 = maybe
PII	Does the message contain PII?	1 = yes, 0 = no
request	Does the message contain a request?	1 = yes, 0 = no
offer	Does the message contain an offer?	1 = yes, 0 = no
aid_related	Is the message aid related?	1 = yes, 0 = no
medical_help	Does the message concern medical help?	1 = yes, 0 = no
medical_products	Does the message concern medical products?	1 = yes, 0 = no
search_and_rescue	Does the message concern search and rescue?	1 = yes, 0 = no
security	Does the message concern security?	1 = yes, 0 = no
military	Does the message concern military?	1 = yes, 0 = no
child_alone	Does the message mention a child alone?	1 = yes, 0 = no
water	Does the message concern water?	1 = yes, 0 = no
food	Does the message concern food?	1 = yes, 0 = no
shelter	Does the message concern shelter?	1 = yes, 0 = no
clothing	Does the message concern clothing?	1 = yes, 0 = no
money	Does the message concern money?	1 = yes, 0 = no
missing_people	Does the message indicate missing people?	1 = yes, 0 = no
refugees	Does the message concern refugees?	1 = yes, 0 = no
death	Does the message imply death?	1 = yes, 0 = no
other_aid	Is there any other aid needed?	1 = yes, 0 = no
infrastructure_related	Does the message concern infrastructure?	1 = yes, 0 = no
transport	Does the message concern transport?	1 = yes, 0 = no
buildings	Does the message concern buildings?	1 = yes, 0 = no

Techniques and Methodology

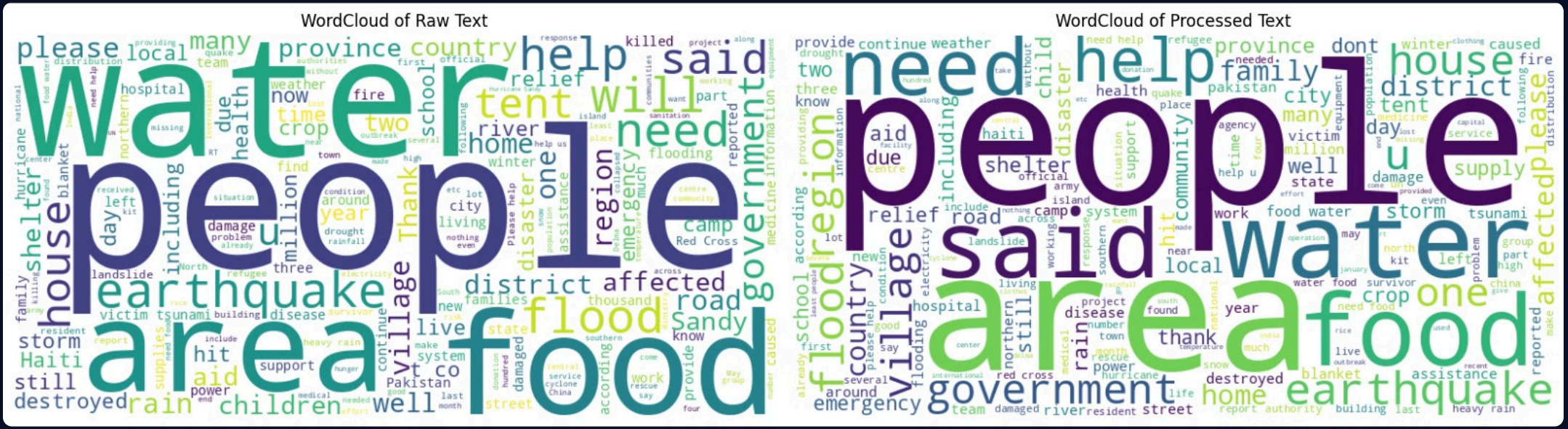
- Preprocessing:
  - Used regex patterns to remove HTML tags, mentions, hashtags, url's , numbers, and extra spaces.
  - Applied the NLTK library to eliminate stop words.
  - Lowercased all of the sentences.
  - Applied Lemmatization
- Addressing Label Skewness:
  - Reduced the count of highly skewed labels to prevent overfitting.
  - Normalized label distribution to enhance generalization.
- Model Training:
  - Employed **BertForSequenceClassification** from Hugging Face for multi-label classification.
- Generative AI Integration:
  - Utilized open-source generative AI models to generate explanations for the classified labels along with the message.

Most occurring words

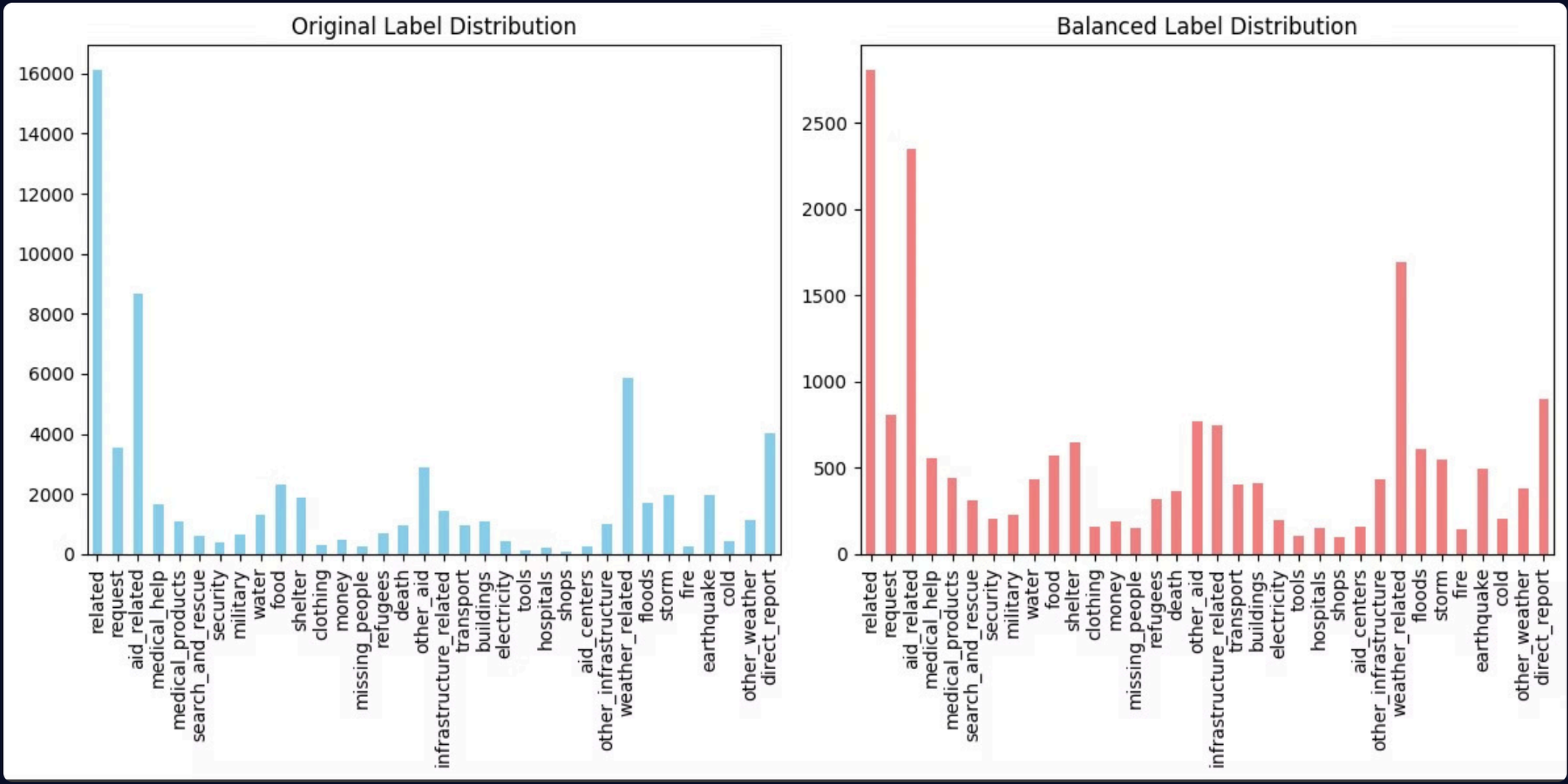




Word Cloud



Addressing Label Skewness



Metrics Comparison (Before Vs After Pre-Processing)

Classification Report:

	precision	recall	f1-score	support
related	1.00	1.00	1.00	558
request	0.81	0.57	0.67	150
aid_related	0.90	0.90	0.90	471
medical_help	0.86	0.18	0.29	102
medical_products	0.33	0.02	0.04	84
search_and_rescue	0.00	0.00	0.00	58
security	0.00	0.00	0.00	43
military	0.00	0.00	0.00	40
water	0.85	0.38	0.53	76
food	0.88	0.67	0.76	122
shelter	1.00	0.16	0.28	129
clothing	0.00	0.00	0.00	37
money	0.00	0.00	0.00	35
missing_people	0.00	0.00	0.00	30
refugees	0.00	0.00	0.00	69
death	0.00	0.00	0.00	75
other_aid	0.00	0.00	0.00	137
infrastructure_related	1.00	0.02	0.04	162
transport	0.00	0.00	0.00	92
buildings	1.00	0.01	0.03	79
electricity	0.00	0.00	0.00	38
tools	0.00	0.00	0.00	27
hospitals	0.00	0.00	0.00	28
shops	0.00	0.00	0.00	22
aid_centers	0.00	0.00	0.00	33
other_infrastructure	0.00	0.00	0.00	96
weather_related	0.93	0.78	0.85	343
floods	0.88	0.36	0.51	119
storm	0.89	0.41	0.56	116
fire	0.00	0.00	0.00	25
earthquake	0.96	0.57	0.71	88
cold	0.00	0.00	0.00	41
other_weather	0.00	0.00	0.00	87
direct_report	0.78	0.49	0.60	170
micro avg	0.92	0.45	0.61	3782
macro avg	0.38	0.19	0.23	3782
weighted avg	0.66	0.45	0.50	3782
samples avg	0.93	0.51	0.63	3782

Classification Report:

	precision	recall	f1-score	support
related	1.00	1.00	1.00	562
request	0.79	0.71	0.75	162
aid_related	0.91	0.88	0.90	461
medical_help	0.72	0.38	0.50	107
medical_products	0.58	0.25	0.35	73
search_and_rescue	0.00	0.00	0.00	59
security	0.00	0.00	0.00	34
military	0.95	0.56	0.70	36
water	0.90	0.73	0.81	88
food	0.86	0.80	0.83	123
shelter	0.83	0.63	0.72	125
clothing	1.00	0.18	0.31	33
money	0.85	0.28	0.42	40
missing_people	0.00	0.00	0.00	25
refugees	0.67	0.12	0.21	48
death	0.74	0.49	0.59	57
other_aid	0.48	0.10	0.17	160
infrastructure_related	0.54	0.28	0.37	146
transport	0.77	0.15	0.25	68
buildings	0.81	0.32	0.46	77
electricity	0.85	0.28	0.42	40
tools	0.00	0.00	0.00	25
hospitals	1.00	0.03	0.06	31
shops	0.00	0.00	0.00	16
aid_centers	0.00	0.00	0.00	28
other_infrastructure	0.43	0.11	0.18	90
weather_related	0.90	0.76	0.82	332
floods	0.93	0.49	0.64	117
storm	0.89	0.66	0.76	106
fire	1.00	0.12	0.22	32
earthquake	0.82	0.73	0.78	83
cold	0.85	0.29	0.43	38
other_weather	0.50	0.01	0.02	82
direct_report	0.78	0.63	0.70	196
micro avg	0.87	0.58	0.70	3700
macro avg	0.66	0.35	0.42	3700
weighted avg	0.78	0.58	0.64	3700
samples avg	0.88	0.64	0.70	3700

Model Outputs

```
Original Message: friend hello please dont forget u cope anymore eaten anything child take care
True Labels: related, request, aid_related, food, direct_report
Predicted Labels: related, request, aid_related, food, direct_report

Original Message: th grade teacher california class would like gather supply want make sure go someplace need gather blanket beany scarf glove hygiene product anyone interested flashlight battery well school supply also possibili
True Labels: related, aid_related, medical_products, shelter, clothing, other_aid, electricity
Predicted Labels: related, aid_related

Original Message: crackdown part malaysia largest blitz flush illegal immigrant three year
True Labels: related, aid_related, military, refugees, death, other_aid, direct_report
Predicted Labels: related, aid_related, military

Original Message: genesis ii pas santiago mag move ¬¨,àû elevation ¬¨,àû elevation ne min
True Labels: related, aid_related, other_aid, weather_related, earthquake, other_weather
Predicted Labels: related, weather_related

Original Message: member council reiterated serious concern rapidly deteriorating humanitarian situation country expressed deep concern increased terrorist threat north due presence among rebel member terrorist group alqaida isla
True Labels: related, aid_related, security, military, other_aid
Predicted Labels: related, aid_related, military
```

Integrating Gen-AI

```python
from transformers import AutoTokenizer, AutoModelForSeq2SeqLM, AutoModelForCausalLM

# Load the tokenizer and model
gen_tokenizer = AutoTokenizer.from_pretrained("Qwen/Qwen2.5-1.5B-Instruct")
gen_model = AutoModelForCausalLM.from_pretrained("Qwen/Qwen2.5-1.5B-Instruct")

# Define a function for generative output
def generate_actionable_text(classified_labels, message):
    # Convert labels back to their column names
    active_labels = [label_columns[i] for i, val in enumerate(classified_labels) if val == 1]
    if not active_labels:
        active_labels = ["No categories identified"]  # Handle case when no categories are detected

    # Craft the structured prompt
    prompt = (
        f"Message: {message}\n"
        f"Categories: {', '.join(active_labels)}\n\n"
        f"Context: Use the above message and identified categories to infer what the message is focusing on. "
    )
```

Messages from Gen-AI

```
Generating responses for sample test cases...

Original Message: british pulling helicopter used relief although floodswollen river posing fresh threat several community cut road rapidly repaired enable aid get
Detected Categories: related, aid_related, weather_related, floods
Generated Response: Message: british pulling helicopter used relief although floodswollen river posing fresh threat several community cut road rapidly repaired enable aid get
Categories: related, aid_related, weather_related, floods

Context: Use the above message and identified categories to infer what the message is focusing on. 1. Generate an explanation for the response.
2. Summarize the key points in a sentence or two.

Explanation:
The message discusses British forces using helicopters for disaster relief despite flooding from swollen rivers, with some communities still cutting roads due to damage. The aid has been able to reach these areas quickly, but there's concern about potential threats from the flooded rivers.

Key Points:
- British forces are using helicopters for relief efforts.
- Flooded rivers pose new threats
--------------------------------------------------

Original Message: need work hunger killing u gonaives
Detected Categories: related, request, aid_related, food, direct_report
Generated Response: Message: need work hunger killing u gonaives
Categories: related, request, aid_related, food, direct_report

Context: Use the above message and identified categories to infer what the message is focusing on.  Respond in English.

The message "need work hunger killing u gonaives" appears to be a call for help or assistance with employment issues. The context of needing work, coupled with the mention of hunger as something that could potentially kill someone, suggests a dire situation where individuals are facing significant challenges
--------------------------------------------------

Original Message: happen port au prince pot de paix
Detected Categories: related, request, aid_related, direct_report
Generated Response: Message: happen port au prince pot de paix
Categories: related, request, aid_related, direct_report

Context: Use the above message and identified categories to infer what the message is focusing on. 1. Message:
Happen Port-au-Prince Pot de Paix

2. Categories:

related - The event or topic is related to a specific place (Port-au-Prince) and has some connection with an organization or group (Pot de Paix).

request - There is a need for assistance or support from others.

aid_related - The context involves providing aid or assistance in some way, either directly or indirectly.

direct_report - The information comes from someone who experienced the event firsthand or witnessed it.
```

Future Work

Faster Response

Efficient Resource Allocation

Improving Accuracy