Unraveling the Yemeni Government-Houthi Conflict: Annotated Sentiment Analysis and Machine Learning Approach

By

GROUP 17

Victor Chiokwe (UB: 22029731)

Abdulrahman Aboluhom (21025487)

Ali Maqsood (22054277)

Muhammad Adil Khalil (23024857)

Oluwashina Atere (22019228)

Abonere Arubayi (23001049)

April 29th, 2024.

ABSTRACT

Yemen, situated in the Middle East, has been marred by a devastating civil war since 2014, leading to one of the world's worst humanitarian crises. This study presents an innovative approach to understanding the Yemeni crisis by analyzing sentiment and emotional expressions in YouTube comments. Leveraging machine learning models like BERT, it aims to predict sentiment and emotions accurately and provide actionable insights for policymakers and stakeholders.

Using both manual annotation and Large Language Model (LLM) methods, the study evaluates sentiment and emotional analysis of YouTube comments. Various machine learning models, including Logistic Regression, Support Vector Machine (SVM), and Recurrent Neural Network (RNN), are assessed for sentiment analysis and emotion prediction. The results show that the RNN model outperforms others, demonstrating high accuracy in sentiment classification and emotion prediction.

Furthermore, the study develops a BERT-based model for emotion prediction, achieving promising results. Notably, the BERT model achieved an accuracy of 40% in predicting emotions, highlighting its potential for understanding complex emotional expressions in online discourse. These findings contribute to the advancement of sentiment analysis research, offering valuable tools for understanding public opinion and discourse on complex socio-political issues like the Yemeni crisis.

Policy makers can utilize this information to formulate more informed policies and strategies for addressing the Yemeni crisis. By understanding public sentiment and emotional responses, policymakers can tailor humanitarian aid efforts to better meet the needs of the Yemeni population. Additionally, insights from sentiment analysis can inform diplomatic efforts to resolve the conflict and foster reconciliation. Moreover, by monitoring trends in public perception over time, policymakers can gauge the effectiveness of their interventions and adjust their strategies, accordingly, contributing to long-term stability and peace in Yemen. The study's methodology and results provide insights into prevailing attitudes, emotions, and trends in public perception, offering a deeper understanding of the multifaceted nature of the Yemeni crisis and its global impact.

Table of Content

1.0. INTRODUCTION

Yemen is a country located in the Middle East, at the southern tip of the Arabian Peninsula, shares borders with Saudi Arabia to the north and Oman to the northeast. Yemen split into North Yemen and South Yemen before reunifying in 1990, albeit internal tensions which have persisted ever since (Burrowes and Wenner 1999; BBC News 2011; Wikipedia 2024).

In recent years, Yemen has faced significant challenges, including political instability, economic struggles, and humanitarian crises. A civil war erupted in 2014, involving various factions and international intervention. The conflict caused widespread suffering, including famine and displacement, making Yemen one of the world's worst humanitarian crises. Efforts to resolve the conflict and alleviate the suffering of the Yemeni people continue, but the situation remains precarious (Al-Helali 2021).

The 1990s saw the rise of the Houthis, a Zaidi Shia movement born from marginalization in the rugged mountains of northern Yemen. Led by the charismatic scholar Husayn Badreddin al-Houthi, the movement emerged to address the grievances of the Zaidi community, who felt excluded by the Yemeni government. Fueled by poverty, corruption, and foreign interference, the Houthis rallied against a system perceived as unjust. Their discontent resonated with tribes, scholars, and youth alike, all burdened by economic inequality, political exclusion, and resentment towards foreign powers. A brutal government crackdown on the Houthis in Saada in 2004 sparked a full-blown insurgency. Al-Houthi's death did not extinguish the movement, but rather cemented his legacy and propelled the rebellion further with the help of the new leader, Abdal Malik Al-Houthi, the brother to Husayn (Wells 2015; Wintour 2019).

The political instability that followed Saleh's departure opened the door for various factions to vie for control. The Houthis, emboldened by victories in the north, seized the capital, Sana'a, in 2014, ousting President Hadi. This triggered a devastating civil war (BBC News 2011). Fearing a Shia-aligned Yemen potentially influenced by Iran, Saudi Arabia led a Sunni Arab coalition to intervene militarily. The result - a brutal conflict characterized by airstrikes, crippling blockades, and immense human suffering.

Determining responsibility for the chaos engulfing Yemen is a complex and contentious issue. While the Houthis have been criticized for their role in precipitating the crisis through armed rebellion and human rights abuses, including the recruitment of child soldiers, the Yemeni government has faced scrutiny for its corruption and ineffectiveness. Moreover, foreign actors, particularly Saudi Arabia, UAE and Iran, bear significant responsibility for exacerbating the conflict through military intervention and proxy support. Western powers, including the United States and the United Kingdom, have also been implicated due to their arms sales and political backing of the Saudi-led coalition (Wintour 2019; Haddad 2022; Nereim and Almosawa 2023).

Millions of civilians bear the brunt of the violence, displacement, and deprivation. The war has inflicted a devastating toll on Yemen's infrastructure, economy, and social fabric, pushing the country to the brink of collapse. (BBC News 2011; Robinson 2015; Al-Helali 2021).

This project presents an innovative approach to understanding the Yemeni crisis. We will utilize YouTube video comments, along with text and emotion annotation, employing both manual annotation and the Large Language Model (ChatGPT). We aim to develop a sophisticated machine learning model capable of accurately predicting sentiment and emotions, even when presented with entirely novel datasets distinct from those used in its training phase. Text and emotional annotation, covering Happiness, Fear, Sadness, Surprise, Disgust, and Anger, aid us in this endeavor. The use of the Large Language Model (LLM) allows us to process and analyze this data faster than the manual method. In addition, the Machine Learning Model approach will enable us to identify and predict the underlying drivers of these emotions and how they evolve over time.

Ultimately, our project seeks to make a meaningful contribution to resolving the Yemeni crisis and deterring similar conflicts from arising elsewhere in the world. Notably, our project stands out not only for its innovative approach in applying sentiment analysis and creating predictive models for YouTube comments in the context of the Yemeni/Houthi crises, but also for the potential broader application of our model in analyzing emotional sentiments across various related topics or with newly trained datasets on any subject.

In conclusion, Yemen's story is one of hardship, rebellion, and human suffering, with devastating consequences for millions of civilians. By leveraging technology and data-driven approaches, we can gain a deeper understanding of the conflict and work towards meaningful solutions.

1.1. Rationale for the Study:

Although numerous studies have investigated public sentiment and discourse on various social media platforms such as Twitter and Facebook, there has been a notable gap in sentiment analysis specifically focused on YouTube video reviews related to the Yemeni conflict. YouTube serves as a distinctive platform for public engagement, hosting a diverse range of user-generated content, including news clips, documentaries, personal vlogs, and commentary. This study aims to fill this gap by analyzing the sentiment expressed in comments and reviews posted on YouTube videos discussing the Yemeni crisis. Through this analysis, our goal is to unearth valuable insights into prevailing attitudes. We will conduct annotations using both manual and LLM methods to determine their respective efficacy, and ultimately, we aim to develop a machine learning model capable of accurately predicting both textual and emotional sentiment related to the topic at hand, as well as with any other dataset on any subject.

1.2. Objectives of the Study

1. Conduct sentiment and emotional analysis of YouTube video reviews on the Yemeni conflict, categorizing comments into positive, negative, and neutral sentiments, as well as emotions such as Happiness, Fear, Sadness, Surprise, Disgust, Anger and Neutral.
2. Carry out annotation to determine whether Large Language Model (LLM) or human annotators provide better sentiment and emotional analysis.
3. Develop machine learning models on annotated emotions and compare their performance to identify the best model for sentiment and emotional analysis.
4. Provide actionable insights for policymakers, researchers, and stakeholders regarding public sentiment, emotional responses, and discourse on the Yemeni crisis.

1.3. Significance of the Study

This study is pivotal for understanding public sentiment on the Yemeni conflict by analyzing sentiment and emotions in YouTube video reviews, offering insights into its multifaceted nature and global impact. Categorizing comments into positive, negative, and neutral sentiments, along with emotions like Happiness, Fear, Sadness, Surprise, Disgust, Anger, and Neutral deepens our understanding of viewers' attitudes. Comparing Large Language Model (LLM) and human annotators' performance and developing machine learning models for emotion analysis contributes to sentiment analysis research. Our BERT-based model excels at predicting sentiment and emotion analysis, offering valuable insights for informing policy decisions thereby promoting greater awareness and engagement with the Yemeni crisis.

2.0. LITERATURE REVIEW

Scholars have deeply studied the Yemeni crisis, providing insights into its causes and effects. Sheila Carapico's book, "Civil Society in Yemen" (1998), examines Yemen's society and politics, focusing on issues like economic inequality and corruption. She shows how these problems contributed to the Arab Spring protests. Carapico's research highlights the challenges Yemen faces in building a democratic government and dealing with tribal relationships. Her work helps us understand the 2011 uprising and its effects.

Researchers recently have increasingly explored the use of sentiment analysis, artificial intelligence (AI), and machine learning algorithms in understanding conflicts and crises globally. These techniques analyze copious amounts of textual data, such as social media posts and news articles, to understand public sentiment, identify key themes, and even predict future trends.

For instance, work by Hartmann et al. (2023) underscores the pivotal role of sentiment analysis in human communication and marketing endeavors, drawing insights from various data sources such as social media, news articles, and customer feedback.

According to Nandwani and Verma (2021) Social media platforms serve as crucial outlets for global emotional expression in today's digital age, utilizing various mediums such as text, images, audio, and video. However, managing the vast amount of unstructured data they generate is challenging.

Sentiment analysis has proven effective in forecasting unrest. For instance, a 2023 study by Oladele and Ayetiran utilized Twitter comments from Nigeria's prominent #EndSARS protest to forecast unrest likelihood, accurately predicting high protest probabilities, since most social unrest these days often originates from platforms like Twitter. This research attains high accuracy, with a 90% accuracy, 94% precision, 85% recall, and an 89% F1 score. Their model excels in predicting negative instances, notably influenced by unique words and training data volume. The study highlights SVM's efficacy in discerning sentiments in tweets, providing policymakers with an essential early warning system for potential unrest.

Looking forward, there is significant potential for further research in leveraging advanced computational methods to address complex geopolitical challenges.

3.0. METHODOLOGY

3.1.0 Data Collection: To gather data from YouTube, we used manual search and scraping methods, with the YouTube Data API. We carefully selected every major video related to Houthis from the past 10 years, sourcing them from reputable channels like BBC News, Al Jazeera English and Hindustan Times and TRT.

3.1.1. Videos Selection: Firstly, we compiled a list of prominent news and media channels known for their coverage of current events, global affairs, and relevant topics. This list included channels such as Al Jazeera English (13.1 million subscribers), BBC News (15.8 million subscribers), Hindustan Times (6.68 million subscribers), USMC (444,000 subscribers), VICE News (8.92 million subscribers), TRT World (6.26 million subscribers), CNN (16.1 million subscribers), and others.

We employed a set of keywords to ensure comprehensive coverage of relevant content. Our main keyword was "Houthis", and the secondary keywords were "Yemen conflict" and "Houthi insurgency". A combination of short-tail and long-tail keywords was used to search videos with diverse viewership. Short-tail keywords like "Yemen," "Houthis," "War," and "Iran" capture broad topics and appeal to a wide audience, while long-tail keywords such as "The Women Fighting to Protect Yemen." By using both types of keywords, we found all relevant videos relating to Houthis.

We selected videos from various times. Our oldest video is from the past 9 years as we wanted to consider the people's sentiments at the start, when people started becoming aware about the Houthis. We also added videos from 4 years ago to see the change in sentiments of people in the middle years.

Our oldest videos provided foundational insights into the Houthi movement. For instance, "Who are Yemen's Houthis?" by BBC News, with 33,000 views, served as a foundational inquiry of the group's origins and motivations. Transitioning to the middle years, videos like "Yemen's Houthis claim to capture Saudi city of Najran" by Al Jazeera English, uploaded four years ago, gained 1.1 million views and 1,291 comments. These videos captured pivotal moments in the conflict's evolution. Recent uploads like "Houthis: Yemen war's 'strategic turning point' explained" by Al Jazeera English, posted just over a month ago is very crucial video relating to Houthis. This offers insights into the shifting dynamics of the conflict, complementing the historical narrative.

3.1.2. Comments Scrapping: In the process of scraping YouTube comments, we utilized the YouTube Data API v3 along with Python's Google API Client library. Our methodology involved several key steps to gather comments from specific videos for further analysis. Firstly, we established the necessary parameters for accessing the YouTube Data API v3, including the API service name, version, and our developer API key obtained from the Google Cloud Console. This setup enabled us to interact with the YouTube API and retrieve comments data efficiently.

After that, we constructed requests to fetch comments from targeted YouTube videos. Upon executing the requests, we received responses containing the requested comments from the YouTube Data API. These responses were parsed to extract relevant information from each comment, such as the author's name, publication date, like count, and the comment text itself.

The extracted comments data was then stored into csv files. After that the csv file was read using Python and comments data was stored in Pandas Data Frame, for subsequent analysis. This is shown in Figure 1 and 2.

Loop for multiple Video IDs

```python
# Function to get comments for multiple videos (a list of YouTube video IDs.)
def get_comments_multiple(video_ids):
    df = pd.DataFrame()
    for video_id in video_ids:
        df2 = get_comments_single(video_id)
        df = pd.concat([df, df2])
    return df
```

Figure 1: Loop for multiple Videos

Retrieve comments

```python
# For multiple videos
# Video IDs here, replace with the right one here
video_ids = ['Kkh8L_me7Wk','iwvTqXHChtA','3PDBTSQ62_I','fHRyxo6gzQ8','pSYs0a1adXw','6omhtUzxHTg',
           'Hq82LnRNOkY','oYTnHIB01KA','WPa6HUxy11w','KEX66rTYfV8','WPa6HUxy11w','zaDnC7xtgw8',
           'dbU37lo5fV4','c_KGSO0MBcc','OQf7IprM72c','Vn5kBgbfIjk','nLRgdFP-s30','AfxzWywPUAk',
           'Hds-pqEIe44','WHsWo92mkt4','Spobr_Z7snE','b_FiIq3h30o','cJtkyYK8kT8','YJvuK1yYDZs',
           'e3KCOFCI4js','1_qod_2ZIxM']
```

```python
# Retrieve comments for multiple videos
df_comments = get_comments_multiple(video_ids)
```

```python
# Preprocess comments in DataFrame
df_comments_cleaned = preprocess_dataframe(df_comments)
```

Figure 2: |YouTube Video Scrapping

3.2.0 Data Annotation

We meticulously conducted our annotation process using Python and manual methods through collaborative efforts within the team, discerning positive, negative, and neutral sentiments. Additionally, we meticulously characterized emotional sentiments within the context of Houthi and Yemen crises video reviews, ensuring coherence and precision. Anticipating potential challenges, we outlined strategies to ensure consistency and accuracy throughout the annotation process.

3.3.0 Data Preprocessing: Our comprehensive data preprocessing strategy encompassed rigorous cleansing to eliminate extraneous information and standardize textual formats. We then carried out tokenization techniques, stopword removal, punctuation normalization, and the application of stemming or lemmatization to streamline textual data. All to make the data amenable to subsequent EDA and machine learning analyses, as illustrated in Figure 3.

Preprocessing

```
!pip install contractions -q
!pip install autocorrect -q

import pandas as pd
import string
import re
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import contractions
from autocorrect import Speller

# Download NLTK resources if not already downloaded
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')

# Lowercasing
df['CText'] = df['Text'].apply(lambda x: x.lower())

# Handling Contractions
df['CText'] = df['CText'].apply(lambda x: contractions.fix(x))

# Removing URLs
df['CText'] = df['CText'].apply(lambda x: re.sub(r'http\S+|www\S+|[^a-zA-Z\s]', '', x))

```

Figure 3: Preprocessing, from Libraries imported to removal of URLs

3.3.1. Removing Irrelevant Information: We carefully sifted through the data, discarding any unnecessary information, especially that is related to the author's privacy such as the Author details.

3.3.2. Lowercase Conversion: To ensure consistency in our analysis, we converted all comments to lowercase using the function: **df['CText'] = df['Text'].apply(lambda x: x.lower()).** This standardization process helps us maintain uniformity and coherence throughout our subsequent analyses.

3.3.3. Dealing with contraction: Contractions are abbreviated forms of words like "I'll" to "I will", "won't" to "will not", etc. These contractions needed to be expanded to their full forms for consistency and better analysis. By executing this code, all contractions within the 'CText' column were expanded, ensuring uniformity, and improving the quality of the text data for subsequent analyses.

3.3.4. Handling URLs, Punctuations and Non-Alphabetic Characters: For the quality and consistency of the text data and to facilitate accurate analysis and insights from the cleaned text data, we needed to carry out this process. This text cleaning process involved using regular expressions (regex) to remove URLs, punctuation, and non-alphabetic characters from the text data.

3.3.5. Taking Care of Special Character and Emojis: We did this to clean the text data in the 'CText' column by removing any non-word characters (excluding underscores) from each element, ensuring that only alphanumeric characters and spaces remained.

```python
# Removing Special Characters and Emojis
df['CText'] = df['CText'].apply(lambda x: re.sub(r'[^\w\s]|_+', '', x))

# Removing Numbers
df['CText'] = df['CText'].apply(lambda x: re.sub(r'\d+', '', x))

# Whitespace Removal
df['CText'] = df['CText'].apply(lambda x: ' '.join(x.split()))

# Tokenization
df['Tokens'] = df['CText'].apply(lambda x: word_tokenize(x))

# Removing Punctuation
df['Tokens'] = df['Tokens'].apply(lambda x: [word for word in x if word not in string.punctuation])

# Removing Stopwords
stop_words = set(stopwords.words('english'))
df['Tokens'] = df['Tokens'].apply(lambda x: [word for word in x if word.lower() not in stop_words])
```

Figure 4: Preprocessing, from removing special characters/emojis to removing stopwords

3.3.6. Removing Numbers: To further cleanse the text data in the 'CText' column, we eliminated all numerical characters in the text file, ensuring that only alphabetic and special characters (if any) remain.

3.3.7. Handling Whitespace: We cleaned the text data in the 'CText' column ensuring that words are separated by only one space, inorder to make the text data more consistent and easier to work with in subsequent analyses.

3.3.8. Tokenization: We tokenized the text data in the 'CText' column and stored the resulting tokens in a new column called 'Tokens'. Tokenization is a fundamental in any preprocessing step in natural language processing tasks because it breaks down the text data into smaller units (words or tokens) that can be further analyzed or processed.

3.3.9 Removal of Punctuations: To make our text much usable we removed punctuation tokens from the list of tokens stored in the 'Tokens' column. Removing punctuation tokens is usually beneficial for various NLP tasks, as punctuation often can interfere with analysis or models.

3.3.10. Stop Word Removal: We meticulously removed common stop words using the NLTK library using the function, **"df['Tokens'] = df['Tokens']. apply (lambda x: [word for word in x if word.lower() not in stop_words])**. By eliminating these insignificant words, we sharpened the accuracy of our sentiment analysis, allowing us to focus solely on the most meaningful textual content.

3.3.11. Lemmatization: Still Leveraging on diverse tools in NLTK Libraries, we used WordNetLemmatizer for this purpose. We systematically reduced words to their base forms or lemmas. This process ensures consistency in text representation, thereby optimizing the accuracy and interpretability of our analyses.

3.3.12. Spelling Correction: Typos and misspellings were diligently identified and rectified using the TextBlob library. This meticulous approach, which considered both context and intent, enhanced the integrity of our dataset, ensuring the reliability of our analyses.

```
# Lemmatization
lemmatizer = WordNetLemmatizer()
df['Tokens'] = df['Tokens'].apply(lambda x: [lemmatizer.lemmatize(word) for word in x])

# Finalizing Text by Join tokens back into a string
df['CText'] = df['Tokens'].apply(lambda x: ' '.join(x))

# Droping unnecessary columns
df = df.drop(columns=['Author', 'Public', 'Tokens'])

# Initialize the spell checker with English language
#spell = Speller(lang='en')
```

Figure 5: Preprocessing, from lemmatization to dropping unnecessary columns

3.4.13. Dataset Description

The features of the dataset are shown in table 1. They provide contextual information about each comment. The dataset contains 18,430 entries (comments) and 8 columns.

S/N	DATA FEATURE	DATA DESCRIPTION
1.	Author	The name of the person that left the comment
2.	Updated_at	The time the comment was published on the YouTube video
3.	Like_count	The number of likes the comment got
4.	text	The content of the comment
5.	Video_id	The ID of the video which the comment belongs to
6.	Public	Whether the comment is public or not
7.	Cleaned_text	The text after removing the emojis and punctuations
8.	Sentiment	This is the sentiment attached to each comment
9.	Emotion	The emotion attached to each comment

Table 1: Dataset description

3.5.0. Sentiment Analysis: We conducted our sentiment annotation exclusively using Python-based tools. Within We collaborated to identify and categorize sentiments into positive, negative, and neutral categories.

3.6.0. Emotion Annotation Guidelines: We developed detailed guidelines for annotating emotions in the YouTube comments. Happiness, Fear, Sadness, Anger, Disgust, Surprised, and Neutral emotion. Each emotion was defined along with specific criteria and keywords to assist annotators in their decision-making process. An example is provided in table 1.

Emotion	Description in Context	Example Comment
Happiness	Contentment or satisfaction with a positive development in the war, regardless of the party responsible.	"Relief to see a ceasefire finally holding. Maybe peace is possible."
Sadness	Disappointment, grief, or sorrow related to the war's human cost or actions by any party involved.	"Devastating news of another civilian airstrike. This war is a tragedy."
Surprise	Unexpected events or actions by any party involved in the conflict.	"Surprised by the Saudi coalition's sudden change in strategy."
Fear	Anxiety or apprehension about the war's future or actions by any party in the conflict.	"Worried about the US withdrawing support. Will the violence escalate again?"
Anger	Outrage or frustration directed towards any party involved in the war or their actions.	"Furious about the Houthi missile attacks targeting civilians. This needs to stop!" (Houthi Focus)
Disgust	Revulsion or disapproval of the war's brutality or actions by any party involved.	"Disgusted by the reports of human rights abuses on all sides. This conflict is barbaric."

Table 2: Table for emotion annotation guidelines

3.6.1. Manual Annotation: A team of human annotators from the team manually annotated a subset of the YouTube comments according to the provided guidelines. Annotators read each comment carefully and tagged it with the appropriate emotion(s) it conveyed. Annotators resolved any disagreements through discussion and consensus to ensure consistency and accuracy.

3.6.2. Language Model Annotation: Simultaneously, we employed GPT-3.5, a state-of-the-art Language Model, to automatically annotate emotions in the same subset of comments. Before annotation, we fine-tuned the LM on emotion annotation tasks using a dataset of annotated comments. This fine-tuning process aimed to enhance the LM's understanding and recognition of emotions in textual data. Once fine-tuned, the LM was applied to assign emotions to each comment automatically. The LM-generated annotations were then compared with the manually annotated ground truth for evaluation and analysis.

3.6.3. Inter-Annotator Agreement (IAA): Inter-Annotator Agreement (IAA) measures the level of agreement between human annotators or between human annotators and an automated system, such as an LM. In our study, we used Cohen's Kappa statistic to calculate IAA between manual annotation and LM annotation.

Cohen's Kappa agreement ranges from -1 to 1, where:

- Kappa = 1 indicates perfect agreement.
- Kappa = 0 indicates agreement equivalent to chance.
- Kappa < 0 indicates agreement worse than chance.

A higher Kappa value indicates a higher level of agreement between manual annotation and LM annotation. In Figure 6, the formula for obtaining the required results is presented.

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

where:

- κ represents the Cohen's kappa coefficient,
- P_o is the observed agreement between annotators,
- P_e is the expected agreement by chance.

Figure 6: Kappa’s Formula

3.7.0. Exploratory Data Analysis (EDA): In our data exploration, we focus on two main goals: understanding how feelings are spread out and finding any clear patterns or trends. To help us achieve this, we used some statistical methods and visualization tools, to understand how different emotions are represented and the details in the text.

3.8.0 Feature Engineering: We employed a smart approach to extract feature engineering. Our goal is to pick out the most key features that will help improve how accurately we classify information. We are careful to choose things that really matter and work well, making sure our feature engineering captures all the different feelings expressed in the video reviews.

3.9.0 Model Selection and Training: We carefully chose our models based on model performance. We outlined our process for training these models, which includes splitting the data into parts and fine-tuning parameters using cross-validation. Next, we developed a model using BERT capable of predicting emotions.

3.10.0 Model Evaluation: After building our models, we delved into gauging model performance across multiple dimensions, incorporating metrics such as accuracy, precision, recall, and F1-score.

3.11.0: Responsible AI practices

When annotating the YouTube comments, we carefully considered the ethics of researching sensitive topics like conflicts and crises. We adhered to the AI principles to ensure the process is fair by ensuring the video selection is representative of diverse viewpoints avoiding bias. The criteria for annotation were clearly documented to ensure transparency and accountability. The privacy of commenters was safeguarded by ensuring that their comments were anonymized. Maintaining human involvement was ensured to oversee the AI operations and make judgements that AI alone cannot make by incorporating diverse perspectives, particularly from those familiar with the regional context.

4.0 RESULTS AND DISCUSSION

4.1. Yearly Sentiment Trend Over the Years

The yearly and weekly sentiment trends reveal that the Houthi/Yemen crisis gained momentum from 2018, peaked in 2020, and reached its highest point in 2024. This significant increase in 2024 could be linked to the recent escalation in the Israel versus Palestinian crises. This information is represented in figures 7 and 8.

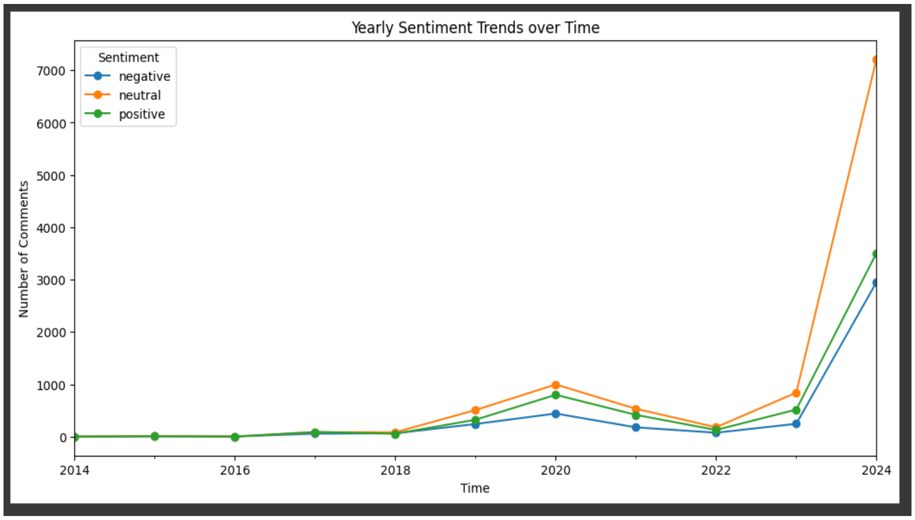

Figure 7: Yearly Sentiment Trends over time

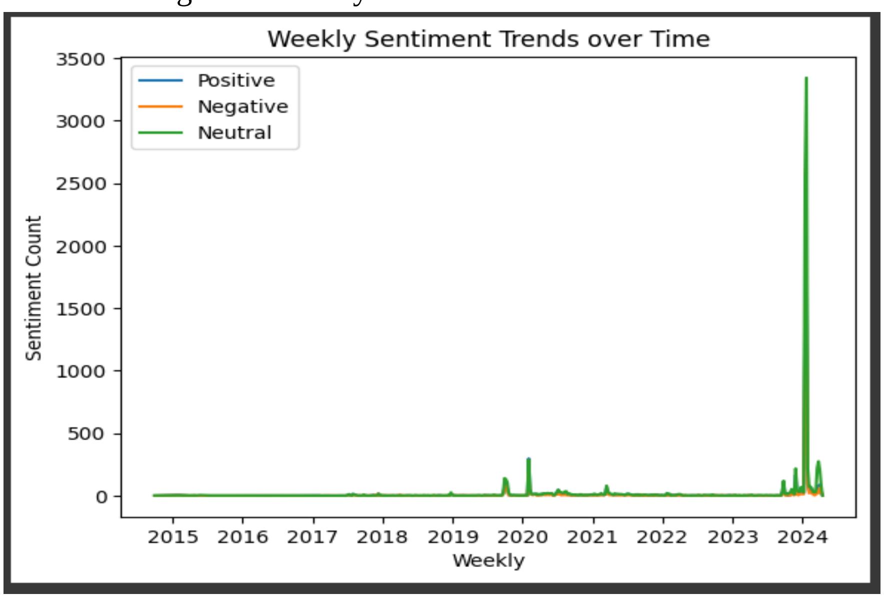

Figure 8: Weekly Sentiment Trends Over Time

Additionally, figure 9 demonstrates the distribution of sentiments across each year, it highlights the video ID with the largest number of comments and the year it was produced.

sentiment	negative	neutral	positive
updated_at			
2014-12-31 00:00:00+00:00	0	1	0
2015-12-31 00:00:00+00:00	5	12	9
2016-12-31 00:00:00+00:00	2	2	1
2017-12-31 00:00:00+00:00	57	85	89
2018-12-31 00:00:00+00:00	64	83	54
2019-12-31 00:00:00+00:00	241	508	323
2020-12-31 00:00:00+00:00	441	996	799
2021-12-31 00:00:00+00:00	178	529	416
2022-12-31 00:00:00+00:00	75	179	127
2023-12-31 00:00:00+00:00	244	840	517
2024-12-31 00:00:00+00:00	2947	7225	3510

Figure 9: Video ID With Comments on Them Showing Years

4.2. Sentiment Distribution Count

Figure 10 illustrates the distribution of sentiments based on the number of likes, with positive sentiments accounting for 51.59% of the total, followed by neutral sentiments at 30.57%, and negative sentiments at 17.84%. Similarly, in Figure 11, regarding comment counts, neutral sentiments were the highest at 50.88%, followed by positive comments at 28.43%, and negative comments at 20.69%. These findings indicate that most comments tend towards neutrality, expressing neither strong opposition nor support for the Houthis. Positive comments often convey strong opposition towards the crisis, while negative comments tend to support the Houthis, often containing strong language and hate speech against the West and its allies, including Saudi Arabia. Overall, most comments express a desire for the conflict to end soon.

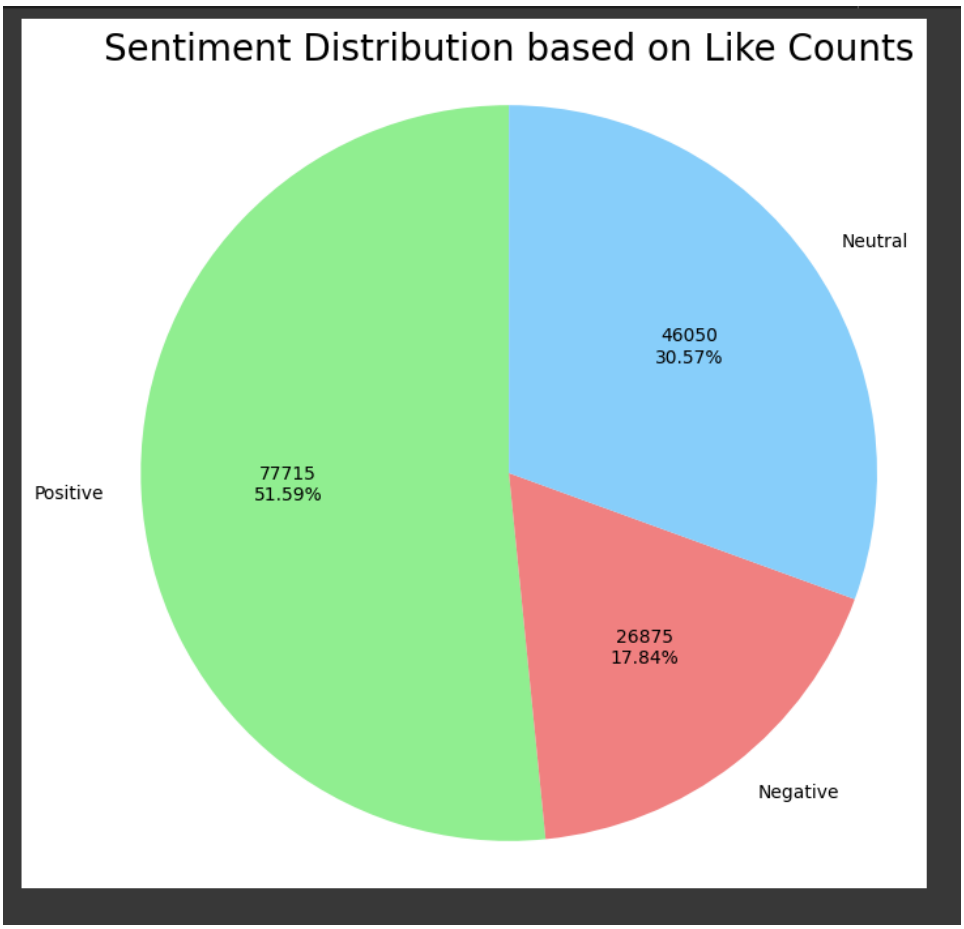

Figure 10: Sentiment distribution based on like counts

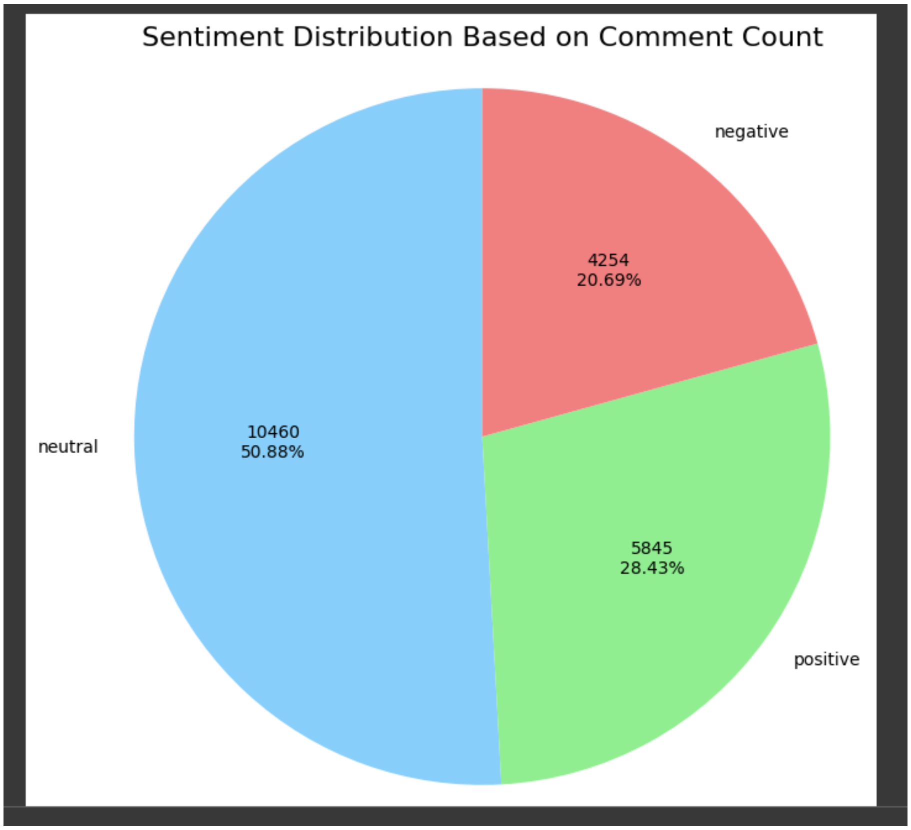

Figure 11: Sentiment distribution based on comment count

4.3. Sentiment Analysis Model

We tested sentiment analysis models like BERT, GPT, and TextBlob to understand how they interpret sentiments. They performed well in predicting positive, negative, and neutral sentiments. However, when it comes to understanding emotions in text, BERT shines. BERT learns from a wide variety of texts, allowing it to comprehend words in different contexts. Unlike TextBlob, which relies on fixed word lists, BERT's learning process is flexible, enabling it to understand words' meanings in various situations. Although GPT is good at generating human-like text, it lacks BERT's deep language understanding. In summary, BERT's extensive learning and contextual understanding make it the most effective tool for interpreting emotions in text. This is evident in Figures 12, 13 and 14.

	updated_at	like_count	text	video_id	comment	sentiment
0	2024-04-01T19:55:13Z	1	اللَّهُمَّ إِنِّي أَعُوذُ بِكَ مِنْ جَهْدِ ال...	Kkh8L_me7Wk		neutral
1	2024-04-01T16:18:04Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	israel making false claim	negative
2	2024-04-01T16:17:52Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	israel making false claim	negative
3	2024-04-01T16:17:41Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	israel making false claim	negative
4	2024-04-01T16:17:30Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	israel making false claim	negative
...	...	...	...	...	...	...
20547	2024-03-25T10:25:11Z	0	Saudi are two faced	1_qod_2ZIxM	saudi two faced	neutral
20548	2015-03-31T04:08:00Z	2	Houthi Leaders converted from Zaidi sect to 12...	1_qod_2ZIxM	houthi leader converted zaidi sect er shia	neutral
20549	2015-03-19T04:32:28Z	0	what's the music track?	1_qod_2ZIxM	whats music track	neutral
20550	2015-02-01T21:52:15Z	11	I support the Houthis and this video is just t...	1_qod_2ZIxM	support houthis video short give people idea f...	positive
20551	2014-09-24T08:39:05Z	1	Zaidis are named after Zayd ibn Ali(695-740), ...	1_qod_2ZIxM	zaidis named zayd ibn ali greatgrandson prophe...	neutral

20552 rows × 6 columns

Figure12: GPT predictions

	updated_at	like_count	text	video_id	cleaned_text	sentiment	sentiment_numerical
0	2024-04-01T19:55:13Z	1	اللَّهُمَّ إِنِّي أَعُوذُ بِكَ مِنْ جَهْدِ ال...	Kkh8L_me7Wk	NaN	neutral	1
1	2024-04-01T16:18:04Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	israel making false claim	negative	0
2	2024-04-01T16:17:52Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	israel making false claim	negative	0
3	2024-04-01T16:17:41Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	israel making false claim	negative	0
4	2024-04-01T16:17:30Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	israel making false claim	negative	0
...	...	...	...	...	...	...	...
20547	2024-03-25T10:25:11Z	0	Saudi are two faced	1_qod_2ZIxM	saudi two faced	neutral	1
20548	2015-03-31T04:08:00Z	2	Houthi Leaders converted from Zaidi sect to 12...	1_qod_2ZIxM	houthi leader converted zaidi sect er shia	neutral	1
20549	2015-03-19T04:32:28Z	0	what's the music track?	1_qod_2ZIxM	whats music track	neutral	1
20550	2015-02-01T21:52:15Z	11	I support the Houthis and this video is just t...	1_qod_2ZIxM	support houthis video short give people idea f...	positive	2
20551	2014-09-24T08:39:05Z	1	Zaidis are named after Zayd ibn Ali(695-740), ...	1_qod_2ZIxM	zaidis named zayd ibn ali greatgrandson prophe...	neutral	1

20552 rows × 7 columns

```
[ ] 1 # Save the updated DataFrame to a new CSV file
```

Figure 13: Textblob Predictions

	author	updated_at	like_count	text	video_id	sentiment	emotion
0	[redacted]	2024-04-01T19:55:13Z	1	اللَّهُمَّ إِنِّي أَعُوذُ بِكَ مِنْ جَهْدِ ال...	Kkh8L_me7Wk	negative	anger
1	[redacted]	2024-04-01T16:18:04Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	very negative	anger
2	[redacted]	2024-04-01T16:17:52Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	very negative	anger
3	[redacted]	2024-04-01T16:17:41Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	very negative	anger
4	[redacted]	2024-04-01T16:17:30Z	0	Israel making false claims again and again ‼	Kkh8L_me7Wk	very negative	anger
...	[redacted]	...	...	...	...	...	...
20547	[redacted]	2024-03-25T10:25:11Z	0	Saudi are two faced	1_qod_2ZIxM	very positive	disgust
20548	[redacted]	2015-03-31T04:08:00Z	2	Houthi Leaders converted from Zaidi sect to 12...	1_qod_2ZIxM	very positive	anger
20549	[redacted]	2015-03-19T04:32:28Z	0	what's the music track?	1_qod_2ZIxM	very positive	disgust
20550	[redacted]	2015-02-01T21:52:15Z	11	I support the Houthis and this video is just t...	1_qod_2ZIxM	very positive	anger
20551	[redacted]	2014-09-24T08:39:05Z	1	Zaidis are named after Zayd ibn Ali(695-740), ...	1_qod_2ZIxM	very positive	anger

20552 ro

Figure 14: BERT Predictions

4.4.0 Emotion Annotation Analysis

This section presents the analysis of the emotion annotations provided by two different annotators. The first column annotations were done by in-person annotators from within the team itself. The data was split among 4 people and rotated so that each person only annotated each row once.

4.4.1. Compilation of the Annotation

Two columns, labeled 'emotion_annotation_1' and 'emotion_annotation_2', contain the emotions as annotated by two different annotators (Manual and LLM respectively). The distribution of emotions in both annotations is depicted in figures 15 and 16. In Annotation 1 (Manual), represented by blue bars, Happiness had the highest frequency, followed by Anger and Disgust. Conversely, in Annotation 2 (LLM), depicted by green bars, the emotions are the same, but the distribution differs significantly. Anger had the highest frequency by a substantial margin, with a frequency of over 5,000, followed by Happiness. This discrepancy could stem from human bias or the LLM's inability to identify sarcasm or differentiate emotions accurately.

Emotion Annotation 1	**Emotion Annotation 2**
Happiness: 6,139	**Anger:** 7,282
Disgust: 3,330	**Happiness:** 5,853
Anger: 2,618	**Sadness:** 1,968
Sadness: 2,021	**Fear:** 1,296
Fear: 825	**Disgust:** 1,187
Surprise: 767	**Surprise:** 822
Neutral: 31	**Neutral**: 22

Figure 15: Distribution of Emotions in both Annotation

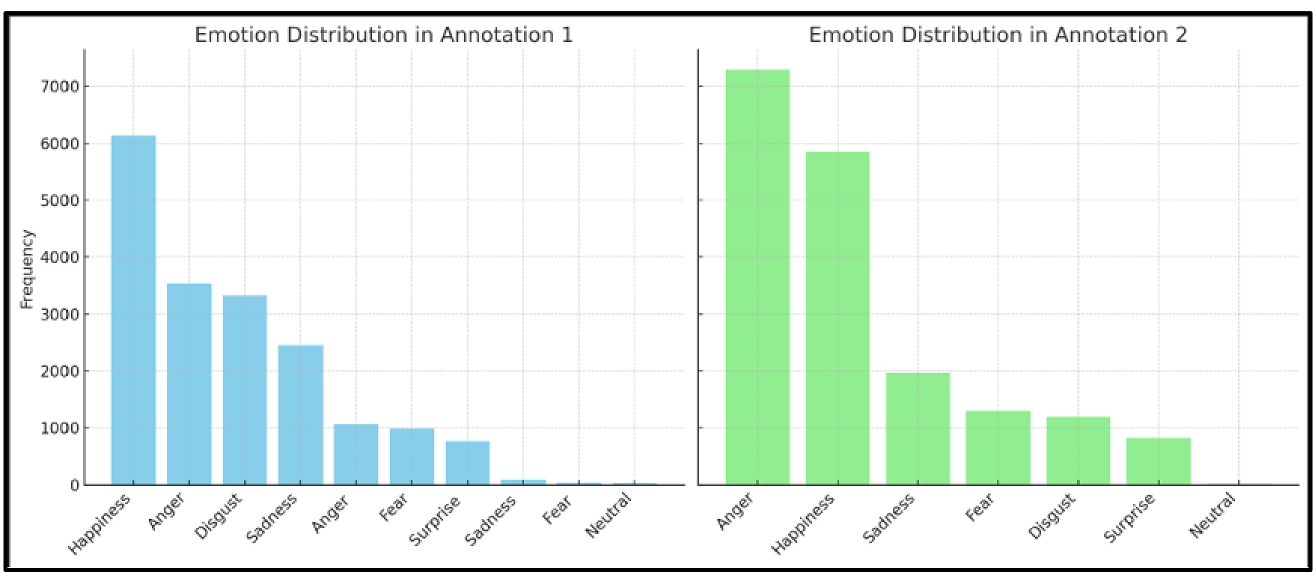

Figure 16: Emotion Distribution in Annotation 1 and Annotation 2

4.4.2. Consistency Between the Annotations

The heatmap below illustrates the correspondence of emotions between annotator 1 (Manual) and annotator 2 (LLM). For Happiness, Anger, and Disgust, both annotators agreed 3615, 2610, and 360 times, respectively. High values along the diagonal indicate strong agreement, suggesting consistent emotional assignments. This reliability makes the dataset suitable for training ML algorithms. Discrepancies outside the diagonal may result from human bias, lack of annotator consistency, or errors in the manual annotation process, highlighting areas for improvement.

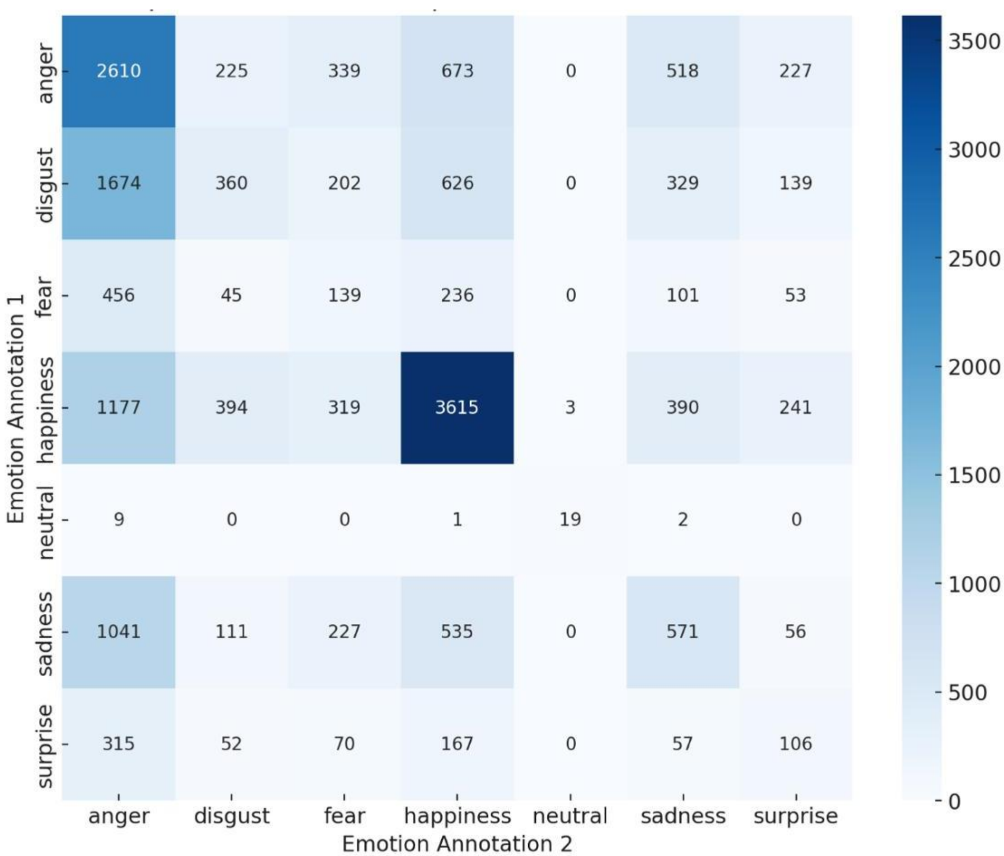

Figure 17: Heatmap of Emotion Overlap between Annotator 1 and Annotator 2

4.4.3. Inter Annotators Agreement (IAA)

In our analysis, we obtained a Kappa score of 0.66 between manual (human) and LLM (Chat GPT), indicating a significant level of agreement. This suggests considerable consistency between annotations by our team members and those generated by the language model, beyond chance. It demonstrates that the LLM's annotations align well with human judgments, making it a reliable tool for automated annotation tasks. Refer to Figure 18 for details.

```
# Check the data types of the columns
print("Data types before conversion:")
print(data[['emotion_annotation_1', 'emotion_annotation_2']].dtypes)

# Convert the columns to a consistent type (e.g., category or string)
data['emotion_annotation_1'] = data['emotion_annotation_1'].astype(str)
data['emotion_annotation_2'] = data['emotion_annotation_2'].astype(str)

# Remove any rows where annotations are missing
data.dropna(subset=['emotion_annotation_1', 'emotion_annotation_2'], inplace=True)
```

```
# Calculate Cohen's Kappa using the two columns
kappa = cohen_kappa_score(annotation_1, annotation_2)

# Print the Cohen's Kappa score to see the level of agreement
print("Cohen's Kappa Score:", kappa)
```

```
Cohen's Kappa Score: 0.6623992627195321
```

Figure18: Cohen’s Kappa Score

4.5 Machine Learning Models

4.5.1 Logistic Regression (LR)

The LR model's classification report indicates robust performance in predicting neutral sentiment, with high precision, recall, and F1-score. However, it is less accurate in identifying negative and positive sentiments, particularly in recall. For negative sentiment, the precision is 92%, recall 71%, and F1-score 80%. For neutral sentiment, precision is 85%, recall 99%, and F1-score 91%. For positive sentiment, precision is 90%, recall 81%, and F1-score 85%. Overall accuracy is 88%. While the model excels in identifying neutral sentiments, it struggles with negative and positive sentiments, especially in recall, as shown in figure 17.

Logistic Regression with TF-IDF

```python
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report

# Feature extraction with TF-IDF
tfidf_vectorizer = TfidfVectorizer(max_features=1000)
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

# Train Logistic Regression
log_reg = LogisticRegression(multi_class='ovr', solver='lbfgs')
log_reg.fit(X_train_tfidf, y_train)

# Evaluate Logistic Regression
y_pred_lr = log_reg.predict(X_test_tfidf)
print("Logistic Regression - Classification Report:")
print(classification_report(y_test, y_pred_lr))
```

```
Logistic Regression - Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.71      0.80       890
           1       0.85      0.99      0.91      2054
           2       0.90      0.81      0.85      1167

    accuracy                           0.88      4111
   macro avg       0.89      0.83      0.86      4111
weighted avg       0.88      0.88      0.87      4111
```

Figure 17: Logistic Regression Classification Report

To test our model further we carried out a confusion matrix as shown in figure 18 and we noticed that the model accurately predicted 632 negative, 2024 neutral, and 943 positive sentiments. However, it misclassified indicating areas for potential improvement. Check figure 18 for details.

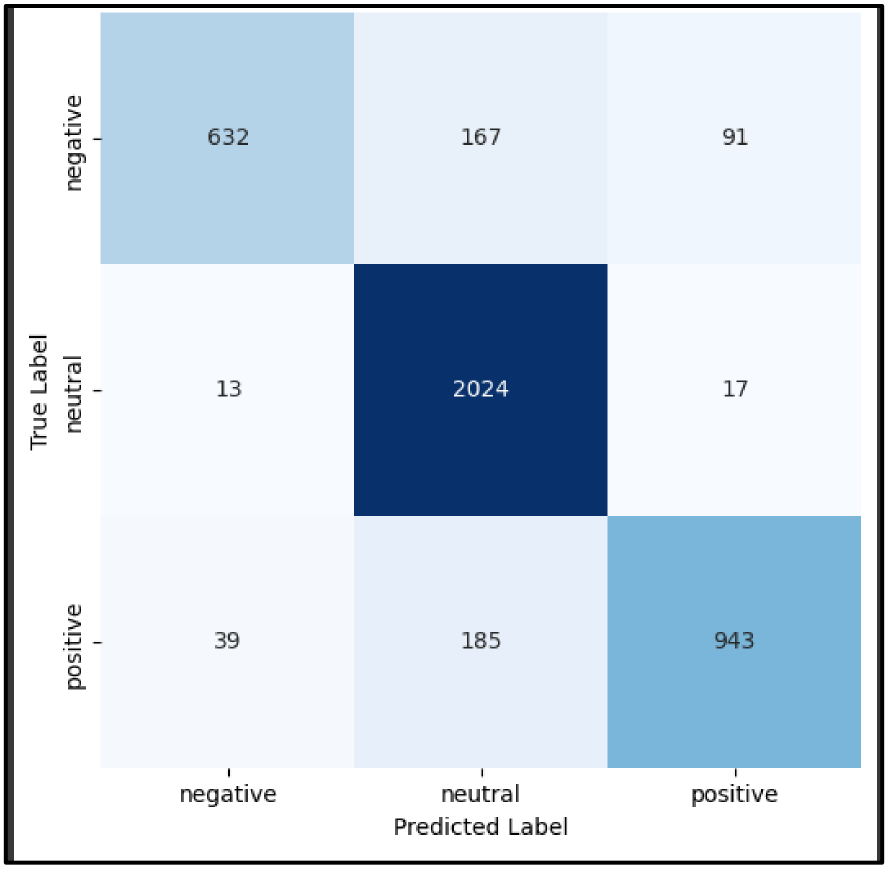

Figure 18: Confusion Matrix for Logistic Regression Classifier

4.3.2 Support Vector Machine

The SVM model's classification report shows robust performance, with high precision, recall, and F1-score for all sentiment classes. Negative sentiment (class 0) has 92% precision, 77% recall, and 84% F1-score. Neutral sentiment (class 1) achieves 88% precision, 99% recall, and 93% F1-score. Positive sentiment (class 2) reaches 91% precision, 83% recall, and 87% F1-score. The overall accuracy is 90%. However, the model exhibits slightly lower recall for negative and positive sentiments compared to neutral, suggesting potential for improvement as shown in figure 19.

Support Vector Machines (SVM) with TF-IDF

```
[ ]
# Train SVM
svm = SVC(kernel='linear')  # Use a linear kernel for text classification
svm.fit(X_train_tfidf, y_train)

# Evaluate SVM
y_pred_svm = svm.predict(X_test_tfidf)
print("SVM - Classification Report:")
print(classification_report(y_test, y_pred_svm))
```

```
SVM - Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.77      0.84       890
           1       0.88      0.99      0.93      2054
           2       0.91      0.83      0.87      1167

    accuracy                           0.90      4111
   macro avg       0.90      0.86      0.88      4111
weighted avg       0.90      0.90      0.89      4111
```

Figure19: SVM Classification report

The confusion matrix accurately predicts 686 negative, 2026 neutral, and 970 positive sentiments but misclassifies instances in all sentiment classes, indicating areas for improvement

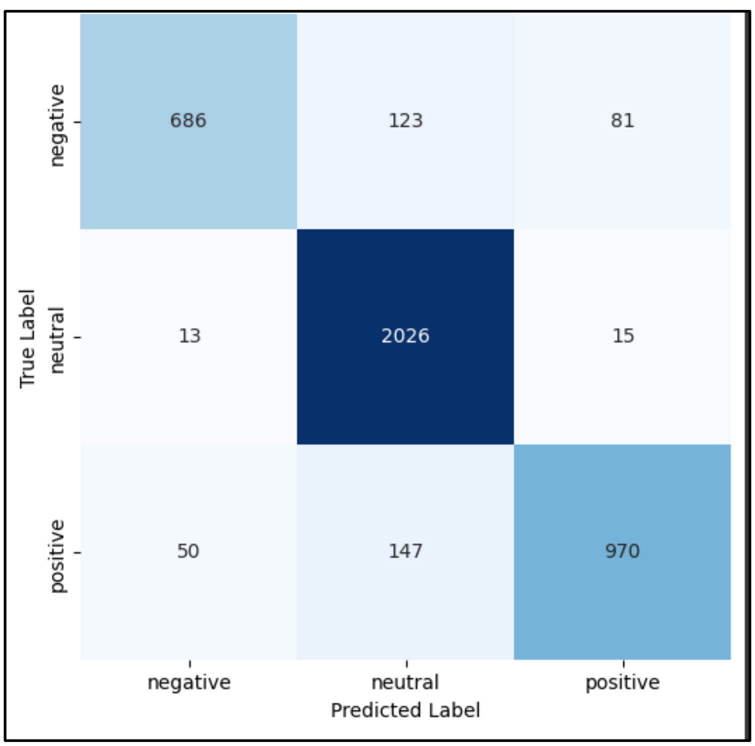

Figure 20: Confusion matrix for the Support Vector Classifier

4.3.3 Recurrent Neural Network (RNN)

The results are from the RNN trained over 5 epochs. Loss decreases significantly from 0.5620 to 0.0299, indicating improved model performance, while accuracy increases from 0.7566 to 0.9918. The classification report reveals high precision, recall, and F1-score for all sentiment classes, with 94% accuracy. Neutral sentiment (class 1) performs the best, followed closely by positive (class 2) and negative (class 0) sentiments.

RNNs

Recurrent Neural Networks (RNNs) with Word Embeddings

```python
# Tokenize the text data
tokenizer = Tokenizer(num_words=10000)  # Maximum number of words in the vocabulary
tokenizer.fit_on_texts(X_train)  # Fit the tokenizer on the training data

# Convert the text to sequences of integers
X_train_seq = tokenizer.texts_to_sequences(X_train)
X_test_seq = tokenizer.texts_to_sequences(X_test)

# Pad the sequences to ensure uniform length
max_sequence_length = 100  # Maximum length of sequences
X_train_pad = pad_sequences(X_train_seq, maxlen=max_sequence_length)
X_test_pad = pad_sequences(X_test_seq, maxlen=max_sequence_length)

# Build an RNN model
rnn_model = keras.Sequential([
    layers.Embedding(10000, 64, input_length=max_sequence_length),  # Embedding layer
    layers.LSTM(64),  # Long Short-Term Memory (LSTM) layer
    layers.Dense(3, activation='softmax')  # Output layer for multi-class classification
])

# Compile the model
rnn_model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])

# Train the RNN model
rnn_model.fit(X_train_pad, y_train, epochs=5, validation_data=(X_test_pad, y_test))  # A few epochs for illustration

# Evaluate the RNN model
y_pred_rnn = rnn_model.predict(X_test_pad).argmax(axis=1)  # Get the class predictions
print("RNN - Classification Report:")
print(classification_report(y_test, y_pred_rnn))
```

```
Epoch 1/5
514/514 [==============================] - 44s 77ms/step - loss: 0.5620 - accuracy: 0.7566 - val_loss: 0.2540 - val_accuracy: 0.9231
Epoch 2/5
514/514 [==============================] - 38s 74ms/step - loss: 0.1316 - accuracy: 0.9582 - val_loss: 0.1940 - val_accuracy: 0.9448
Epoch 3/5
514/514 [==============================] - 37s 72ms/step - loss: 0.0596 - accuracy: 0.9830 - val_loss: 0.2381 - val_accuracy: 0.9380
Epoch 4/5
514/514 [==============================] - 38s 74ms/step - loss: 0.0400 - accuracy: 0.9889 - val_loss: 0.2540 - val_accuracy: 0.9397
Epoch 5/5
514/514 [==============================] - 35s 69ms/step - loss: 0.0299 - accuracy: 0.9918 - val_loss: 0.2739 - val_accuracy: 0.9443
129/129 [==============================] - 3s 14ms/step
RNN - Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.90      0.91       890
           1       0.97      0.97      0.97      2054
           2       0.91      0.93      0.92      1167

    accuracy                           0.94      4111
   macro avg       0.94      0.93      0.93      4111
weighted avg       0.94      0.94      0.94      4111
```

Figure 21: RNN Model Classification Report

The confusion matrix displays an RNN model's sentiment analysis performance. It accurately classified 797 negative, 1989 neutral, and 1072 positive sentiments. However, misclassifications occurred suggesting areas for improvement. This is shown in figure 22 below.

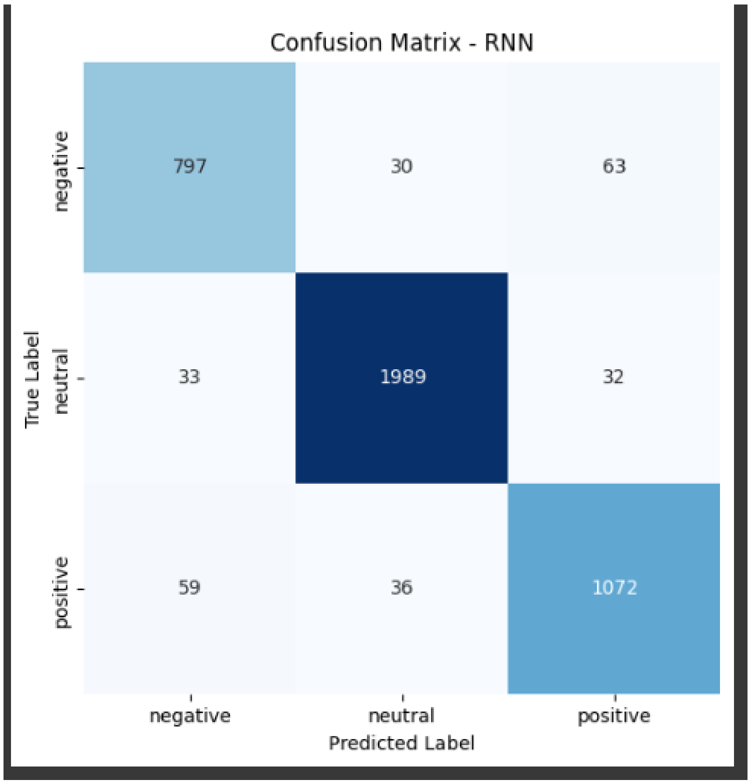

Figure 22: Confusion Matrix for the RNN model

4.3.4 Comparison of the Models

In evaluating the scoring metrics of all models, the RNN outperforms LR and SVM models for sentiment analysis. With a 94% accuracy, the RNN demonstrates high precision, recall, and F1-score across all sentiment classes. Notably, it has the fewest misclassifications in the confusion matrix, indicating superior accuracy in sentiment prediction. Although LR and SVM perform reasonably well, they exhibit slightly lower accuracy and more misclassifications. Hence, the RNN emerges as the most effective model for this sentiment analysis task. Nevertheless, factors like data quality and hyperparameter tuning can influence model performance.

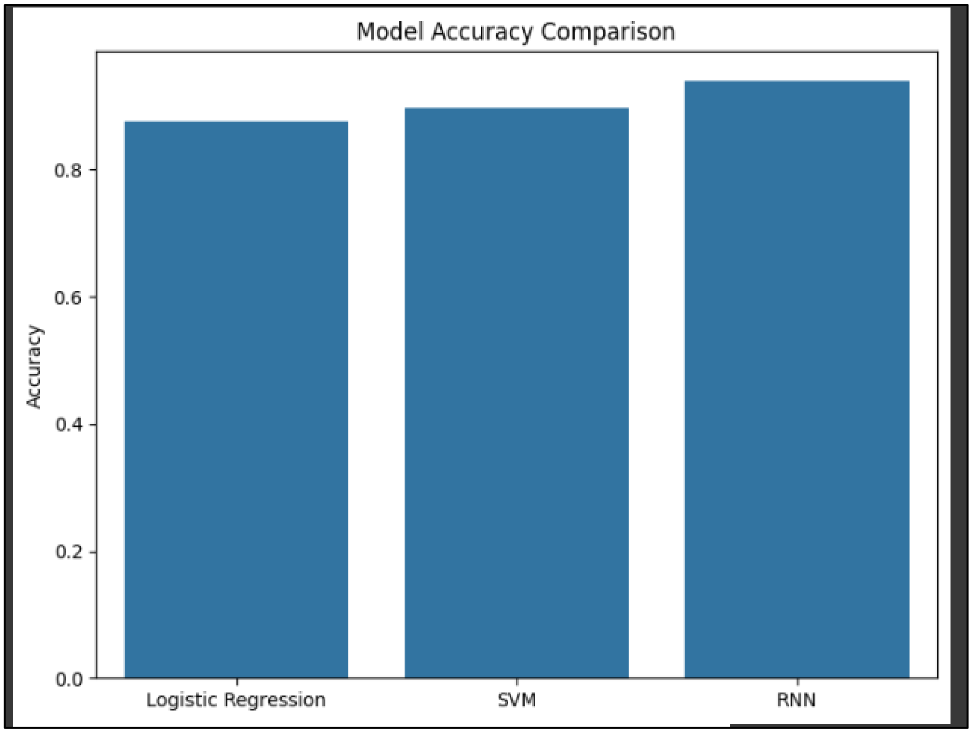

Figure 23: Model Accuracy Comparison

4.4.0. Predictive Model

We developed a BERT-based model for emotion prediction to annotate unlabeled data, aiming to discern complex emotional states expressed in comments. Using LLM annotated data for training, we found human annotators somewhat biased. LLM outperformed humans, achieving 40% prediction accuracy, which could improve with more data. The model underwent fine-tuning for improved accuracy and adaptability. Achieving 40% accuracy post-training was deemed reasonable. The fine-tuned model was saved for ongoing development. This is seen in

```python
# Model
model = BertForSequenceClassification.from_pretrained('bert-base-uncased', num_labels=len(label_encoder.classes_))
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
model.to(device)

# Optimizer
optimizer = AdamW(model.parameters(), lr=2e-5)

# Training loop
epochs = 5
for epoch in range(epochs):
    model.train()
    for batch in data_loader:
        input_ids = batch['input_ids'].to(device)
        attention_mask = batch['attention_mask'].to(device)
        labels = batch['labels'].to(device)

        model.zero_grad()
        outputs = model(input_ids, attention_mask, labels=labels)
        loss = outputs.loss
        loss.backward()
        optimizer.step()

    print(f"Epoch {epoch + 1} complete.")
# Save the model and tokenizer to your Google Drive
model.save_pretrained('/content/drive/MyDrive/AI and Data Science/Model')
tokenizer.save_pretrained('/content/drive/MyDrive/AI and Data Science/Model')

print(f"Model and tokenizer saved to {'/content/drive/MyDrive/AI and Data Science/Model'}")
```

Figure 24: Predictive Model Algorithm

CONCLUSION

The Yemeni crisis represents a complex and multifaceted challenge with devastating consequences for millions of people. This study has provided valuable insights into public sentiment and emotional responses towards the crisis through analysis of YouTube comments. By employing machine learning models like BERT and RNN, we were able to accurately predict sentiment and emotions expressed in online discourse.

Our findings reveal the depth of emotions and attitudes prevalent among the public, highlighting the urgency for effective intervention. The RNN model demonstrated superior performance in sentiment classification and emotion prediction, offering a nuanced understanding of the complexities of public perception.

Policy makers can leverage these insights to develop more targeted and effective strategies for addressing the Yemeni crisis. By understanding public sentiment, policymakers can tailor humanitarian aid efforts to better meet the needs of the Yemeni population and design diplomatic initiatives to foster reconciliation.

Furthermore, our study underscores the importance of leveraging technology and data-driven approaches in humanitarian and diplomatic efforts.

In conclusion, while the Yemeni crisis presents formidable challenges, our study demonstrates that by understanding public sentiment and emotions, policy makers can take proactive steps towards resolving the crisis and building a brighter future for Yemen and its people.

RECOMMENDATIONS

Based on our study of the Yemeni crisis, we offer the following recommendations:

Firstly, sentiment analysis can identify shifts in public sentiment towards reconciliation efforts. This allows policymakers to gauge the effectiveness of their initiatives and adjust them to better align with public sentiment.

Secondly, policymakers should tailor policies to directly address public concerns identified through sentiment analysis. By focusing on issues such as poverty, corruption, and political exclusion, policies can better meet the needs of the Yemeni people and foster trust and support.

Moreover, the use of the LLM for annotation proved efficient, offering speed and ease. However, it is important to mitigate human biases in annotation. Experts with domain knowledge and language proficiency should be involved in the process to ensure accuracy.

By implementing these recommendations, policymakers, humanitarian organizations, and stakeholders can enhance their understanding of the Yemeni crisis. This, in turn, can lead to more effective response strategies, contributing to a more stable and peaceful Yemen.

REFERENCES

Al-Helali, Hisham (2021) *No end to Yemen civil war on the horizon, senior UN official briefs Security Council*. https://news.un.org/en/story/2021/08/1098272 Accessed 28 March 2024.

BBC News (2011) *Yemen country profile*. https://www.bbc.co.uk/news/world-middle-east-14704852 Accessed 28 March 2024.

Burrowes, R. and Wenner, M.W. (1999) Yemen. *Encyclopedia Britannica* 26 July.

Carapico, S. (1998) *Civil Society in Yemen*. Cambridge University Press.

Haddad, M. (2022) Infographic: Yemen's war explained in maps and charts. *Al Jazeera* 9 February.

Hartmann, J., Heitmann, M., Siebert, C. and Schamp, C. (2023) More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing* 40 (1), 75–87.

Mohammad, S.M. (2021) Sentiment analysis. *Emotion Measurement* Elsevier. 323–379.

Nandwani, P. and Verma, R. (2021) A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining* 11 (1), 1–19.

Neogi, A.S., Garg, K.A., Mishra, R.K. and Dwivedi, Y.K. (2021) Sentiment analysis and classification of Indian farmers' protest using twitter data. *International Journal of Information Management Data Insights* 1 (2), 100019.

Nereim, V. and Almosawa, S. (2023) Saudi Arabia and Houthis Hold Peace Talks in Yemen: What to Know. *The New York Times* 10 April.

Oladele, T.M. and Ayetiran, E.F. (2023) Social Unrest Prediction Through Sentiment Analysis on Twitter Using Support Vector Machine: Experimental Study on Nigeria's #EndSARS. *Open Information Science* 7 (1), .

Robinson, K. (2015) *Yemen's Tragedy: War, Stalemate, and Suffering*. https://www.cfr.org/backgrounder/yemen-crisis Accessed 28 March 2024.

Schmitz, C. (2023) Project MUSE - Yemen in the Shadow of Transition: Pursuing Justice amid War by Stacey Philbrick Yadav (review). *The Middle East Journal* 77 (1), 119–120.

Wells, M. (2015) *The Houthis and the risks of internationalizing the Yemeni crisis*. https://theconversation.com/the-houthis-and-the-risks-of-internationalizing-the-yemeni-crisis-36729 Accessed 28 March 2024.

Wikipedia (2024) *Yemen*. https://en.wikipedia.org/wiki/Yemen Accessed 28 March 2024.

Wintour, P. (2019) Yemen civil war: the conflict explained. *The Guardian* 20 June.

APPENDICES

INDEXES OF ABBREVIATIONS USED WITHIN THE REPORT

BERT - Bidirectional Encoder Representations from Transformers

GPT - Generative Pre-trained Transformer

NLP – Natural Language Processing

LR – Logisitic regression

LLM- Large Language Model

SM – Support vector Machine

RNN – Recurrent Neutral Network

Name	Contribution
Victor Chiokwe **(UB: 22055845)**	Worked on the Introduction, Literature review and overall compilation of the report. Ensure the report was written in an excellent manner for submission. Also handled the emotion annotation manually.
Abdulrahman Aboluhom (21025487)	Handled the emotion annotation using LLM and the manual annotation. Inter- Annotator agreement.
Ali Maqsood (22054277)	Handled Video selection, comments extraction, data cleaning and preprocessing. Also contributed to sentiment analysis and emotion annotation and writing the methodology.
Muhammad Adil Khalil (23024857)	Worked on the technical aspect of the report, from scraping of the comments to building of the classifier models using the annotated dataset.
Oluwashina Atere (22019228)	Performed exploratory analysis and created visualization plots to showcase the trend of sentiments over time.
Abonere Arubayi (23001049)	Report and discussion aspect of the report and contributed to annotating the data using LLM.