A Natural Language Processing Approach to Analyzing Social Media Discourse on the Israel-Palestine Conflict

Tasfia Barshat
190042126
tasfiabarshat@iut-dhaka.edu

Azmayen Fayek Sabil
190042122
azmayenfayek@iut-dhaka.edu

ABSTRACT

Understanding public sentiment towards the Israel-Palestine conflict is crucial for policymakers, journalists, and peacebuilders. In this study, an approach combining Sentiment Analysis, Emotion Analysis, Topic Modeling and Discourse Analysis is used to accumulate the overall expression of public opinion as seen in social media news posts and comments. The study utilizes annotated data from the Israel-Gaza conflict of 2014 and a recently scraped dataset comprising news headlines and comments from social media and news portals. By applying Sentiment and Emotion Analysis to our dataset, not only the overall sentiment but also the sentiment expressed towards key aspects of the conflict, such as ceasefire agreements, civilian casualties, or diplomatic efforts, are revealed. This allows for a comprehensive analysis of the diverse opinions and viewpoints present in the discourse. This study combines Sentiment Analysis, Emotion Analysis, Topic Modeling and Discourse Analysis to better understand how public views are shaped on the Israel-Palestine conflict. Moreover, the methodology developed in this study provides a framework for analyzing public sentiment and discourse on other contentious issues in the digital age.

1. INTRODUCTION

Social media platforms have become significant channels for individuals to voice their opinions and share information in various communities. In today's digital age, people extensively use these platforms to express their thoughts and emotions freely. As the internet continues to gain widespread popularity, individuals spend considerable amounts of time on social media websites. Consequently, the widespread use of these platforms means that people have the potential to influence society, whether positively or negatively, through their opinions. Internet sources play a crucial role in shaping public perceptions on political matters, economic issues, and social dynamics. [1]

In recent years, the surge in internet usage has impacted public discourse, particularly evident on social media platforms like Twitter, which have become pivotal arenas for expressing sentiments and viewpoints. Amidst the ongoing Russia-Ukraine conflict, global opinions have expanded on Twitter, prompting novel research endeavors to discover the intricacies of public sentiment during geopolitical crises. One such study introduces an innovative approach that integrates ROBERTa, Aspect-based Sentiment Analysis, and Long Short-Term Memory (LSTM) networks for sentiment analysis. By scrutinizing a vast dataset of geotagged tweets related to the conflict, this hybrid model achieves remarkable accuracy, providing invaluable insights into public sentiment during chaotic geopolitical events.

Furthermore, another compelling study investigates the topic coverage and sentiment dynamics surrounding the Ebola health issue, both on Twitter and in traditional news publications. Through a comparative analysis, it explains the divergent representations and sentiment expressions across various media channels, shedding light on how health information is disseminated and perceived in different contexts. These studies collectively underscore the significance of sentiment analysis and discourse analysis in understanding complex societal phenomena. Additionally, studies explore hate speech patterns related to fat stigma on social media platforms, investigate critical discourse analysis of media representations of the Hizbullah-Israel conflict, and examine the language used to represent the Israeli-Palestinian conflict in influential media outlets. These studies help us better grasp how language, ideology, and power dynamics intersect to shape public discourse and perceptions across various fields.

2. LITERATURE REVIEW

2.1 Aspect based Sentiment and Emotion Analysis with ROBERTa, LSTM

In recent years, the rise in internet usage has significantly impacted public opinion on social media platforms, with Twitter being a prominent venue for expressing sentiments. Amidst the Russia-Ukraine conflict, global opinions have surfaced on Twitter. This study introduces a novel approach 1 that combines Roberta with Aspect-based Sentiment Analysis (ABSA) and Long Short-Term Memory (LSTM) for sentiment analysis. Utilizing a large dataset of geotagged tweets related to the conflict, the hybrid model achieves an impressive accuracy of 94.7%, surpassing existing techniques. This research provides valuable insights into public sentiment during geopolitical crises. [1].

2.2 Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news

Another study investigates the topic coverage and sentiment dynamics of Twitter and news publications concerning the Ebola health issue. Previous research in sentiment

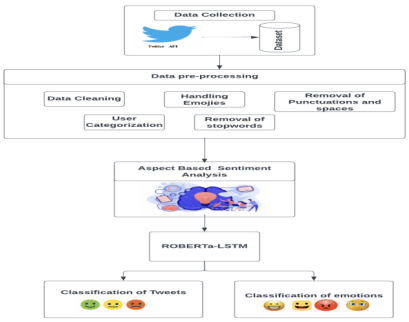

Figure 1. Overview of Proposed Model by [1]

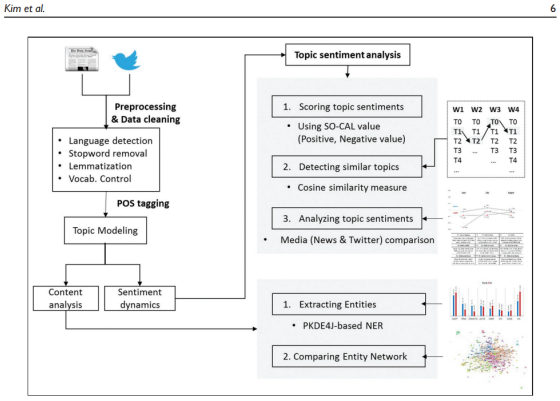

Figure 2. The overview of topic and sentiment analysis. by [2]

mining analysis and topic discovery has explored methodologies to understand public discourse on social media and traditional news outlets. Scholars have utilized sentiment analysis techniques to extract sentiments from textual data across various domains and employed topic modeling techniques like Latent Dirichlet Allocation (LDA) to identify key themes. While prior studies have examined public discourse on diseases such as Ebola, particularly on platforms like Twitter, few have directly compared the topic coverage and sentiment dynamics between social media and traditional news outlets on health issues. This study fills this gap by conducting a comparative analysis of Twitter and news publications' coverage and sentiments regarding the Ebola crisis 2. By examining differences in topic coverage and sentiment expression, the study contributes to understanding how health information is disseminated and perceived across different media channels. Overall, the literature on sentiment mining analysis and topic discovery informs the present study's research questions and methodology, providing a foundation for its findings within the broader research landscape. [2]

2.3 A critical discourse analysis of the selected Iranian and American printed media on the representations of Hizbullah-Israel war

This study investigates the relationship between language and ideology by comparing the coverage of the Hizbullah-Israel conflict in an Iranian newspaper and an American magazine. It utilizes Systemic Functional Linguistics and transitivity analysis, following M.A.K. Halliday's framework. The research demonstrates how news structures function as ideological apparatuses, using linguistic strategies like passivization and nominalization to obscure underlying meanings. Through separate analyses of the printed media, the study reveals distinct representations of the conflict, shedding light on the role of language in mediating ideological perspectives within news discourse. This is the methodology that was followed in this study, detailing the approach and techniques used to analyze the texts. [3]

1. **Transitivity System Analysis:** Following M.A.K. Halliday's framework, the study uses transitivity analysis to examine clauses in texts, focusing on processes and participants involved in representing events.

2. **Components of Transitivity:** Halliday identifies six kinds of processes and their associated participants within the transitivity system, clarifying who performs actions and who is affected or benefits from them.

3. **Purpose:** Transitivity analysis reveals how actions are performed, by whom, and on what, essential for understanding how different ideologies are conveyed through syntactic structures in clauses.

4. **Application:** The study applies transitivity analysis to Kayhan International and NEWSWEEK to evaluate how ideologies are expressed in their representations of warfare between Hizbullah and Israel.

5. **Analysis Process:** Data from each outlet are analyzed separately, considering different time points during the conflict. Main participants' roles and process percentages are quantified using SPSS software.

6. **Comparison and Interpretation:** Comparisons between the outlets' representations illustrate differences and similarities in the portrayal of the conflict. Findings are discussed, and data are reproduced in the appendix for reference.

2.4 Hate Speech Patterns in Social Media

This study explores hate speech patterns related to fat stigma on social media platforms by combining quantitative (sentiment analysis, emotion analysis, and topic modeling) and qualitative methods (discourse analysis). Sentiment analysis enables text mining of large amounts of user opinions (in some given context) to reveal overall attitudes and the emotional content of these expressions. It can facilitate the detection of hate content implanted in open-ended discussions by extracting the negative sentiments associated with sensitive topics. Furthermore, topic modeling,

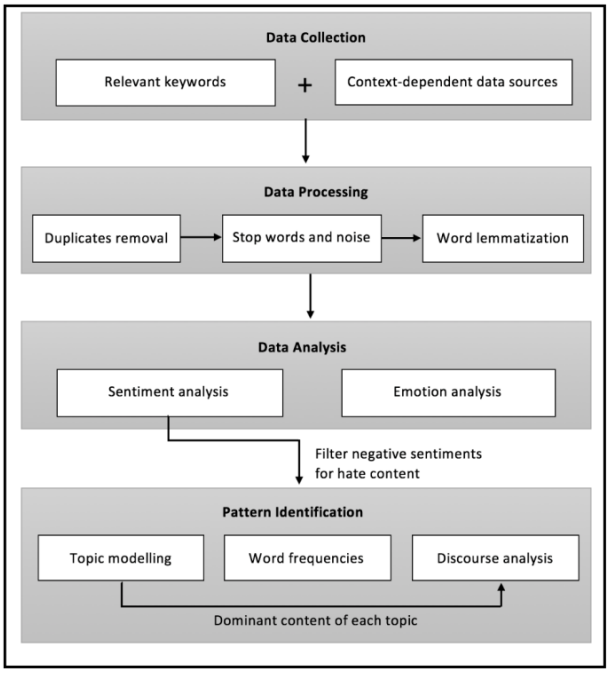

Figure 3. Proposed Methodology. by [4]

a machine learning method, helps find common themes in unstructured text data. This process allows for a systematic review of topics, sentiments, and emotions in social media conversations to reveal hate speech patterns. Additionally, discourse analysis adds a qualitative viewpoint to these quantitative approaches. Discourse analysis is a linguistic methodology that identifies social ideology on various topics through language analysis. It has been used to identify societal stances on topics such as racism, feminism or sexism. Therefore, a mixed-method analytical approach comprising machine learning techniques (e.g., sentiment analysis, emotion analysis, topic modeling) and discourse analysis, can help overcome current limitations in detecting hate speech (fat stigma) patterns to offer richer and more meaningful conversational insights.

The authors applied the framework to a corpus of texts related to gendered and weight-based data extracted from Twitter and Reddit. The framework helped detect different emotions expressed, word frequency patterns, and broader fat-based themes underlying hateful content online. Therefore, this study establishes a roadmap for mixed-method analyses, enhancing our understanding of hate speech patterns in social media, an area often overlooked in current information systems research. [4]

2.5 The Israeli-Palestinian Conflict in American, Arab, and British Media: Corpus-Based Critical Discourse Analysis

This study investigates the language used to represent the Israeli-Palestinian conflict in influential media outlets. Combining methods from Critical Discourse Analysis (CDA) and Corpus Linguistics (CL), the study analyzes discursive representations in American (CNN), Arab (Al-Jazeera

	Al-Jazeera	BBC	CNN
Number of words	1,681,254	711,787	315,192
Number of news reports	3,903	1704	640
Number of days collected	823	823	823
Average news articles per day	4.74	2.07	0.77
Average article length by words	431	418	492

Figure 4. General Statistics of the Israeli-Palestinian Conflict Study Corpora. by [5]

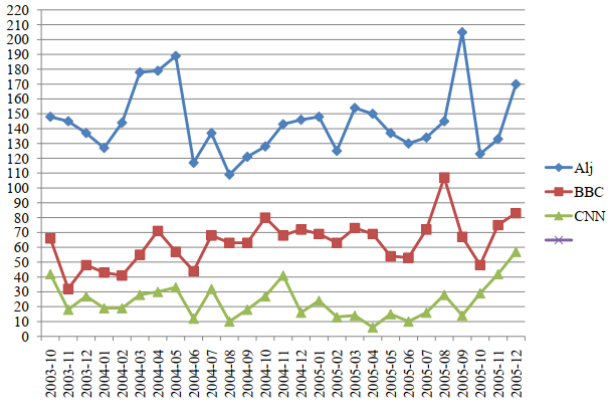

Figure 5. Number of news articles related to the Israeli-Palestinian conflict per month on AlJazeera, BBC, and CNN news websites. by [5]

Arabic), and British (BBC) media. CDA investigates how power and ideology are enacted through language in social and political contexts. CL employs computational techniques to analyze patterns of language use in large collections of natural texts. Key findings include topics emphasized or downplayed in coverage (such as terrorism, occupation, settlements, and the Israeli disengagement plan), as well as strategies employed by each news website to control positive or negative representations of conflict actors.

The study contributes to understanding how power and ideology are enacted through language in social and political contexts. It uses corpus-based techniques, including keyword analysis, concordancers, and collocation finders, to explore topics emphasized, downplayed, and strategies employed by news websites in representing the conflict. The corpus findings are interpreted using informative CDA frameworks, particularly Van Dijk's ideological square framework. This framework helps uncover underlying ideologies and power dynamics in media representations. [5]

	Al-Jazeera	BBC	CNN
Agents of Violence	(Israeli, occupation, security) forces, **Al-Aqsa**, Brigades, Martyrs, (Palestinian, Hamas) resistance, Al-Qassam	(Palestinian, Islamic, Hamas) militants, militant, army, Aqsa, Brigades, Martyrs, (suicide, Palestinian) bomber, troops, soldiers, military, bombers, soldier	IDF, terrorist, military, militants, Aqsa, forces, Qassam, militant, troops, soldiers, bomber, police, army, bombers, terrorists, soldier
Methods of Violence	firing, car, fidāʾīya, operation, (state, Israeli, what he called) terrorism, assassination	suicide, attack, attacks, raid, rocket, operation, bombing, rockets, nuclear, shot, violence, intifada, strikes, assassination, missile	attacks, attack, suicide, terror, fired, fire, rocket, rockets, (suicide, Palestinian) terrorism, bombing, blast, missile, helicopter, airstrike, targeted, bombings, mortar, nuclear, violence, weapons
Outcomes/ Victims Of Violence	was/were martyred, injury, death, martyr, wounded	killed, killing, fired, Gilad, burial, Iman	Killed, wounded, (Israeli) civilians, burial, killing, responsibility
Break From Violence	calm, truce	truce, ceasefire	cease (fire)

Figure 6. Key Keywords Referring to the Military Aspect of the Israeli-Palestinian Conflict. by [5]

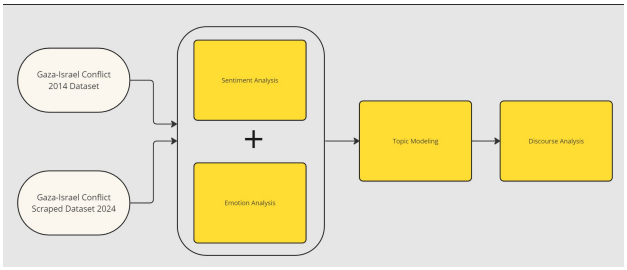

Figure 7. Proposed Methodology

3. RESEARCH QUESTIONS

The key research questions guiding the study, aimed at understanding public sentiment and discursive strategies surrounding the Israel Palestine conflict:

RQ01: How does public sentiment towards specific aspects of the Israel-Palestine conflict, vary in terms of social media posts and comments?

RQ02: What are the primary discursive strategies employed by social media users, news portals, and grassroots movements to influence public opinion on the Israel Palestine conflict, and how have these strategies evolved since 2014?

4. METHODOLOGY

Our proposed methodology involves collecting a contemporary dataset along with utilizing previous 2014 datasets. Both datasets are fed into the sentiment and emotion analysis, and the resulting output is subsequently used for topic modeling. Finally, the result of topic modeling goes through discourse analysis within socio-political contexts to reveal how power, ideology, and social structures shape communication dynamics. 7

عباس ،،، تكلمنا مع الجانب الأمريكي وطلبنا أن يوقفوا العمليات العسكرية من جانب إسرائيل، ونحن نحاول ان نقنع حركة حماس بوقف العمليات، ولكن للاسف لم ننجح
Abbas ,,, we spoke with the American side and asked to stop the military operations from the Israeli side. We are trying to convince Hamas to halt the operations. Unfortunately, we did not succeed

Figure 8. Arabic short posts resenting Breaking News

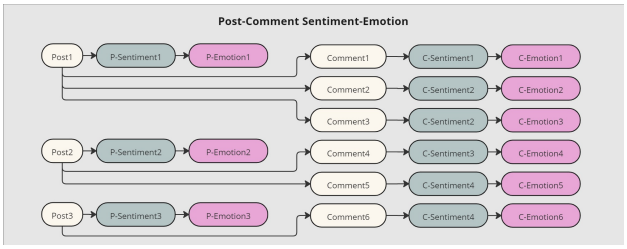

Figure 9. Sentiment and Emotion Analysis Output

4.1 Dataset Collection

We commence by acquiring two distinct datasets: the "2014 GAZA Attack" dataset, comprising historical data relevant to the Israel-Palestine conflict, and the "Web-scraped" dataset, obtained from currently social media platforms like twitter, reddit etc. 8 [6]

So the datasets utilized in this study comprise two primary sources:

1. **Social Media Reviews:** Arabic short posts representing Breaking News are collected from various social media platforms. These posts are selected based on their relevance to the Israel-Palestine conflict. Specifically, reviews are gathered from prominent Arabic news networks such as Al Jazeera and Al Arabiya, providing a diverse range of perspectives from reputable sources. [6]
   `https://shorturl.at/jtRX9`

2. **Web Scraping:** Additionally, data is obtained through web scraping of social media platforms like Reddit or Twitter. This process involves extracting relevant posts and comments related to the Israel-Palestine conflict from these platforms, offering a broader scope of user-generated content and discussions.

4.2 Sentiment Analysis and Emotion Analysis

We conduct sentiment analysis and emotion analysis on the collected datasets to understand the prevailing sentiments and emotional expressions related to the Israel-Palestine conflict. 9

4.2.1 Sentiment Analysis

Using exsiting sentiment analysis techniques, we analyze the sentiment polarity of the social media reviews and web-scraped content. This process involves categorizing each post or comment as positive, negative, or neutral based on the emotional tone conveyed.

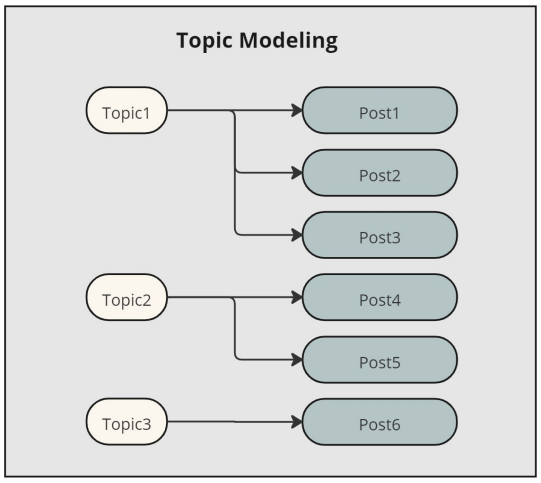

Figure 10. Topic Modeling Output

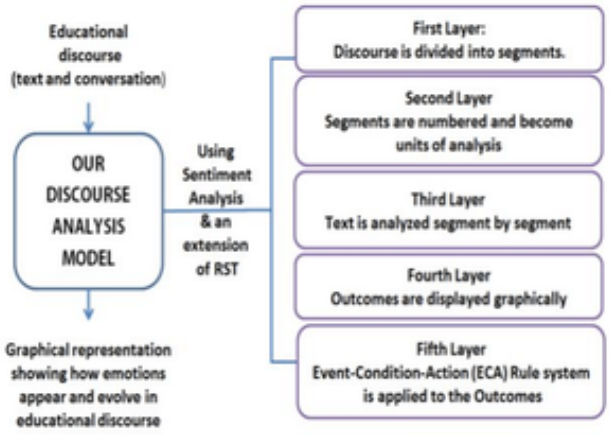

Figure 11. A layered discourse analysis approach with respect to emotion [7]

4.2.2 Emotion Analysis

Additionally, we use emotion analysis to explore the emotions present in the data. By detecting and categorizing emotions such as anger, joy, sadness, and fear, we gain insights into the emotional responses evoked by discussions surrounding the conflict.

These analyses enable us to capture not only the sentiment polarity but also the emotional intensity and diversity within the discourse, providing a holistic understanding of public sentiment towards the Israel-Palestine conflict.

4.3 Topic Modeling

The outputs from the sentiment analysis and emotion analysis serve as crucial inputs for the subsequent step of topic modeling. For topic modeling, we can use techniques such as Latent Dirichlet Allocation (LDA) or Non-negative Matrix Factorization (NMF) to uncover latent themes and topics within the data. By integrating sentiment and emotion insights into the topic modeling process, we ensure that the resulting topics capture not only the semantic content but also the emotional context of the discussions surrounding the Israel-Palestine conflict. 10

4.4 Discourse Analysis

From the topic modeling phase, we derive Topic and related Event/Comments, outlining the identified themes and events within the data.

We then integrate the outputs from sentiment and emotion analysis with the post and comments, generating the dataset that holds the sentiment and emotion of the posts and their respective comments. This integrated dataset provides a comprehensive view of sentiment and emotion expressed within the context of specific topics and events. 11 Finally, both the "Topic and Event/Comments" segments and the "Event/Comment-Sentiment and Emotion" dataset are utilized as inputs for Discourse Analysis. This final step involves a detailed examination of the socio-political context employed in the discourse surrounding the Israel-Palestine conflict, shedding light on the communicative strategies shaping public sentiment.

5. REFERENCES

[1] U. Sirisha and B. Chandana, "Aspect based sentiment and emotion analysis with roberta, lstm," *International Journal of Advanced Computer Science and Applications*, vol. 13, 01 2022.

[2] E. H.-J. Kim, Y. K. Jeong, Y. Kim, K. Y. Kang, and M. Song, "Topic-based content and sentiment analysis of ebola virus on twitter and in the news," *Journal of Information Science*, vol. 42, no. 6, pp. 763–781, 2016. [Online]. Available: https://doi.org/10.1177/0165551515608733

[3] M. Yaghoobi, "A critical discourse analysis of the selected iranian and american printed media on the representations of hizbullah-israel war," *Journal of Intercultural Communication*, vol. 9, pp. 1–15, 09 2009.

[4] V. Wanniarachchi, C. Scogings, T. Susnjak, and A. Mathrani, "Hate speech patterns in social media: A methodological framework and fat stigma investigation incorporating sentiment analysis, topic modelling and discourse analysis," *Australasian Journal of Information Systems*, vol. 27, 02 2023.

[5] M. A. Kandil, "The israeli-palestinian conflict in american, arab, and british media: Corpus-based critical discourse analysis," 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:140888339

[6] "Gaza 2014 dataset." [Online]. Available: https://shorturl.at/jtRX9

[7] M. Arguedas, T. Daradoumis, and F. Xhafa, "Towards an emotion labeling model to detect emotions in educational discourse," 07 2014.