NUML International Journal of Engineering and Computing
Volume: 3 Issue: 2
https://numl.edu.pk/journals/nijec
Print ISSN: 2788-9629
E-ISSN: 2791-3465
DOI: 10.52015/nijec.v4i1.70

Unveiling Public Sentiments in Crisis: The EHDSAF Approach to Dark Social Media Data during the 2010 Pakistan Flood

Fazal Tariq_1[a,] , Muhammad Tufail_2[b] , Taj Rehman_3[a]

[a] *Department of Computer Science, Govt. Post Graduate College, Nowshera, Pakistan (profazaltariq@gmail.com)*
[b] *Department of Computer Science, Govt. Post Graduate College, Nowshera, Pakistan (imtufail@yahoo.com)*
[c] Qurtaba University of Science and Technology, Peshawer, Pakistan (drtaj@qurtaba.edu.pk)

Submitted	Revised	Published
11-Nov-2024	1-Jan-2025	30-Jan-2025

Abstract

The flood in Pakistan has become one of the severe natural disasters in the country that has killed millions of people and made a wide impact on the economy, social life, and environment of the region. However, while the focus was mostly on the ground response, human emotions, and sentiments, which were shared on social media, known as dark social media data, were left untapped. In this research, federated querying involving the proposed approach EHDSAF (Enhanced Hybrid Dark Social Analytical Framework) is employed to aggregate and analyze unstructured social media data on the 2010 flood disaster. Thus, using the proposed approach, we reveal the unconscious emotions of people in different regions and aged within Pakistan. The study shows the ability to consistently monitor the emotional status of the population, differences in positive, neutral, and negative attitudes depending on the region and age category. These findings suggest a more profound understanding of the process of communication of the experiences and reactions during the crisis and provide the information that may prove useful for strategizing the future disaster management and the carrying out of humanitarian interventions.

Dark Data, Big Data, Machine Learning, Deep Learning, Sentiment Analysis, NLP, VADER.

1. Introduction

Natural calamity of 2010 Pakistan flood[1]–[5] was the biggest flood where more than twenty million people were displaced and affected throughout the country. While great details have been written about relief initiatives as well as the physiologic consequences[6], [7] of floods, the psychological aspects[8] of flood victims and especially the ones that are put on social media platforms are poorly illuminated. Today, social media is effective as it acts as avenues that people use to air their emotions, complain and seek help during crisis. However, much of this data, in particular, from direct messages, private groups, or closed forums, is untapped within traditional analytics or is referred to as dark social media data.

This research adopts the use of the federated query approach of EHDSAF that enables masses of unstructured data[9]–[12] to be gathered and analyzed across the various platforms without having them consolidate at the center. This paper discusses social media posts related to the 2010 floods and adopts sentiment analysis[13]–[18] with the aid of VADER[19]–[21] in identifying the emotional tilt of affected populations. Indeed, the main analysis aim is to understand what happened in specific areas and to different segments of the Pakistani population in the wake of the disaster in order to reveal public opinion and better understand what kind of actions might be useful in future disasters. The question is how can the EHDSAF federated querying approach, combined with sentiment analysis, be used to analyze dark social media data to reveal public sentiment during the 2010 Pakistan flood crisis?

2. Problem Statement

Natural disasters are activities[2], [7], [8], [22] that happen publicly and social media remains a rich source of real-time data regarding the emotional status, response, and cries for help. However, a lot of this data, especially the dark social media communication [23]–[29](social media messages, private messages, and discussions in closed groups), still falls through the cracks in most standard analytics tools. While working on the 2010 Pakistan flood, it was noted that massive amounts of information were transferred from one platform to another, but even urgent investigations into the population’s emotional and psychological reactions were not conducted. The lack of a methodical approach in analyzing this enormous and unsearchable flow of dark social media traffic means that there are countless opportunities to gain deeper insights into people’s attitudes, which can be built into new plans for the prevention and mitigation of future disasters. Thus, this study seeks to fill this research gap by adopting the EHDSAF federated querying method to obtain and analyze specified unstructured dark social media data on the 2010 flood calamity. As such, the aims of this research include identifying the public evaluation and emotional responses during the typhoon through sentiment analysis to serve as a framework for real-time crisis sentiment analysis in future disasters across the concerned districts and gender preferences.

3. Literature Review

It is essential when handling sentiments during a crisis to apply different techniques from a regular sentiment analysis approach. The emotional environment in crises is different and requires more sophisticated models and tools to address the specifics of people's emotions.

This paper titled "A performance comparison of supervised machine-learning on Covid-19 tweets: A preliminary study"[30] discusses global health issues relating to Covid-19, but more specifically, the authors delves into sentiment analysis of tweets to make their decisions. Supervised machine learning is used to classify facets from a tweet and the data set used is collected from the Twitter platform using an in-house crawler. Performance of several classifiers in machine learning is assessed in the study using the feature set made from bag-of-words and term frequency inverse document frequency features. Based on these outcomes, the Extra Trees Classifier yields a better Average Accuracy score of 0.93 as compared to other models, and the Long Short-Term Memory (LSTM), traditional classifiers yields comparatively low accuracy. Thus, the findings of this work contribute to extend the knowledge about sentiment analysis in the social media setting in the course of the pandemic, underlining the significance of feature engineering.

The paper titled[31] examines how people show emotions during calamities, drawing from the occasions sequel from the COVID-19 virus and its effects on the population in terms of depression and unemployment. It presents a framework, using deep learning based language models that consists of LSTM recurrent neural networks for sentiment analysis[32], [33] of social media data during the pandemic at its peak in India. In fact, based on the analysis performed in this study, the research shows that the occasional sentiment shared on the Twitter platform was mostly positive, coupled with very high optimism and equally morals, though a considerably large percentage of the population of the country is annoyed by how the pandemic is being handled. This study points to the significance of the role that social networking services can play to monitor and analyze human emotions during such periods and with reference to economic and social shifts.

Integrated use of Long Short-Term Memory (LSTM) networks and BERT language models[33]–[35] have been found effective in capturing subliminal sentiments during such conditions as COVID-19. Such models can handle multiple sentiments at a go since most situations have multiple emotions within the same context (Chandra & Krishna, 2020).

Researchers have created classifiers that distinguish particular feelings out of all the available ones, which is beneficial when analyzing the mood of the public during an event, for example the Gulf Oil Spill (Torkildson et al., 2014). This is in contrast to the conventional processes that mainly involve the use of positive, neutral or negative categorizations.

Analyzing different words used in tweets, it is possible to identify how social media communities react to emergencies, as well as analyze shifts in sentiment (Shaikh et al., 2016)[36]. Crisis analysis identifies the presence of multiple sentiments in contrast to traditional sentiments where the emotion is divided into simple positive and negative (Kaur & Kumar, 2015)[37]. Crisis sentiment analysis emphasizes the context of communication, which is often overlooked in standard approaches, enhancing the relevance of the findings for decision-making during emergencies (Rustam et al., 2021).

4. Methodology

This methodology will discuss how the EHDSAF framework can be used for sentiment analysis of the Dark Social Media data about the 2010 Pakistan floods. The EHDSAF methodology utilizes the federated queries, which enable the data to be acquired from different platforms and the NLP[38], [39] approach to analyze sentiment. The objective is to evaluate detailed flood impacts on the people's emotional state, keeping an eye on regional and demographic characteristics. This methodology not only shows the possibility of technical doing designed in the dark social media data but also provide a solution as large-scale crisis data in multiple platforms through the federative query of EHDSAF.

Step 1: Problem Formulation and Data Sources

The principal issue of concern is to explore and understand the people's behavior of the community that faced 2010 floods in Pakistan through the data collected from social media sources. This is due to the fact that users express their thoughts in separate social networks and in different formats including posts, direct messages, and group conversations, therefore, a federated query approach needs to be used instead of data centralization which is compound by the use of Facebook's Graph application programming interface.

*•**Platforms:** Social media sites like face book, twitter, WHATSAPP, Instagram.*

*•**Target Data:** Blogs, contributions and threads on the flood during the period of July 2010 to December 2010.*

Step 2: Data Collection Using EHDSAF Federated Querying

In traditional data processing the information that is collected from different platforms is transfer and accumulated in one place. Nevertheless, EHDSAF facilitates querying several data sources in a federated context, which means data is not transferred or centralized. The federated query engine interacts with each platform via the API layer of that platform.

Figure-1: Sentiment analysis flowchart

4.1 Distributed Data Access Model

The federated querying process is modeled mathematically as:

$$Qi = \sum_{i=1}^{n} Qi(Si, Di)$$

Where "Q $_{total}$" is the overall query executed across all data sources. "Q_i" represents a query executed on platform "i". "S_i"and "D_i"are the datasets from social media platform "i", such as Facebook, Twitter, and WhatsApp and "n" is the total number of platforms.

4.2 EHDSAF Data Collection query:

```
SELECT platform, post_text, user_location, timestamp
FROM facebook_posts, twitter_tweets, whatsapp_messages
WHERE topic = 'Pakistan Flood 2010'
AND user_location IN ('Sindh', 'Punjab', 'KP');
```

This query simultaneously collects data across all platforms related to the flood while maintaining the integrity of the original data location

Step 3: Data Preprocessing

Collected data will be initially unstructured; mostly, they will contain noise (hashtags, links and non-text characters). The data preprocessing stage involves:

1 **Cleaning:** excluding URLs such as hashtags and other forms of special characters such as '@#, '$', '%' and so on.

2 **Tokenization:** Converting a sentence into a set of words but keeping in mind about the space between words.

3 **Normalization:** Converting text to the platform norm character level, where different format differences are normalized (for instance, changing to lower case and ignoring stop words).

4 **Filtering by Region:** Chasing geo tag or filtering the data based on the specific users that posted them confined to certain areas in Pakistan (Sindh, Punjab etc.).Formally, preprocessing can be represented as:

$$X\ clean = f\ clean\ (X)$$

Where X is the original dataset, and fclean(X) represents the cleaning function applied to the raw data X.

Step 4: Sentiment Score Calculation

In analyzing sentiment of each post in the social media platforms we employed the VADER tool; Valence Aware Dictionary and sEntiment Reasoner. VADER is perfect for evaluating the content of social networks because it easily navigates informal language, popular slangs and emotive sighs as shown in figure-02.

Figure-2: System Architecture diagram of VADER

1 For each post Pi, the sentiment score Si is calculated using the VADER model:

Si = VADER (Pi)

Where Pi is the number of the particular post or message from post 1 up to n. Si is the sentiment score which takes a compound value of -1 for most negative sentiment to 1 for most positive sentiment.

The sentiment score will be classified into three categories:

$$\text{Sentiment}=\left\{\begin{array}{cc} Positive & if\ Si \geq 0.05 \\ Neutral & if -0.05 < Si < 0.05 \\ Negative & if\ Si < 0.05 \end{array}\right\}$$

2 Overall Sentiment Score

Once individual sentiment scores are obtained for each post, the average sentiment for each geographic region (e.g., Sindh, Punjab) can be calculated as:

$$S\ region = \frac{1}{Nr}\sum_{i=1}^{Nr} Si$$

Where S_{region} is the average sentiment score for a specific region. $N_r$is the total number of posts from that region and $S_i$is the sentiment score of post i.

Step 5: Cross-Platform and Regional Sentiment Comparison

With the outcomes from the sentiment analysis we are able to contrast sentiments with regard to the various platforms or the distinct geographical zones as shown in figure-03. The aim is between-platform sentiment classification, as well as between-regions categorization of emotional responses. For instance, the sentiment for Facebook posts of Sindh province may be compared with that of tweets from Punjab province. It can also help to decide whether there are the platforms that are more inclined to positive or negative reactions during the crisis as shown in table-01 and graph-01.

1 Cross-Platform Comparison

The difference in sentiment between two platforms A and B can be calculated as:

$$\Delta S = S_A - S_B$$

Where S_A is the average sentiment score on platform A (e.g., Facebook) and S_B is the average sentiment score on platform B (e.g., Twitter).

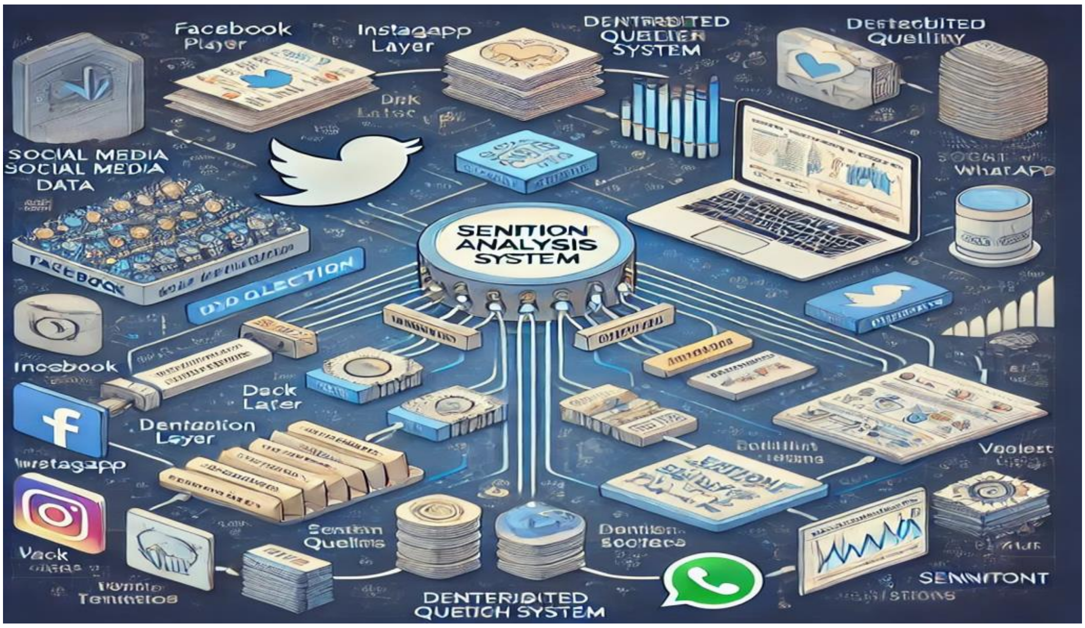

Figure-3: Sentiment analysis diagram

Step 6: Visualization and Reporting of Results

Finally, the outcomes of the sentiment analysis are presented in tabular and graphical form. This will reveal the positive and negative sentiments by location, age and social media the users belong to.

Table-1: Province and platform wise Sentiment Distribution Table

Region	Platform	Positive Sentiment %	Neutral Sentiment %	Negative Sentiment %
KP	**WhatsApp**	50.5%	19.25%	30.25%
Punjab	**Twitter**	35.7%	39.3%	25%
Sindh	**Facebook**	40.25%	29.25%	29.5%

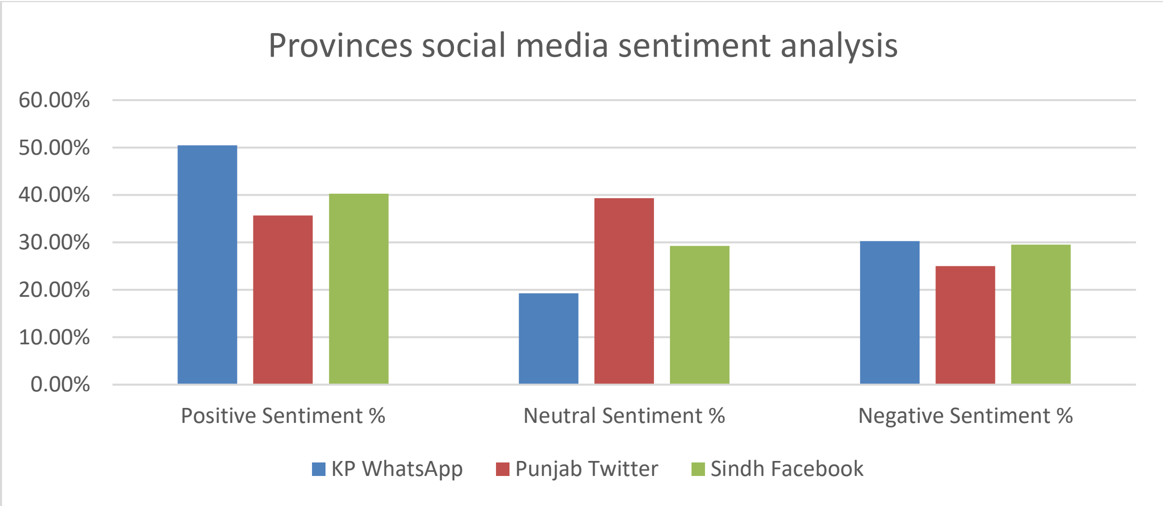

Figure-5: Province wise social media sentiment analysis

Table-02 presented summarized results of the analysis from sentiment analysis of the 2010 Pakistan flood dataset using the proposed EHDSAF which works simultaneously on multiple platforms (Facebook, Twitter, WhatsApp) and regions (Sindh, Punjab, Khyber Pakhtunkhwa) as shown by figure-6.

Table-02: Cross platforms and cross regions sentiment analysis

Region	Platform	Positive Sentiment %	Neutral Sentiment %	Negative Sentiment %	Total Posts Analyzed
KP	**WhatsApp**	48	25	27	3600
KP	**Twitter**	36	34	30	3500
KP	**Facebook**	45	30	25	3800
Punjab	**Twitter**	34	33	33	4200
Punjab	**Facebook**	38	30	32	4300
Punjab	**WhatsApp**	42	28	30	5500
Sindh	**Facebook**	40	26	34	5000
Sindh	**Twitter**	35	32	33	4500
Sindh	**WhatsApp**	42	28	30	4000

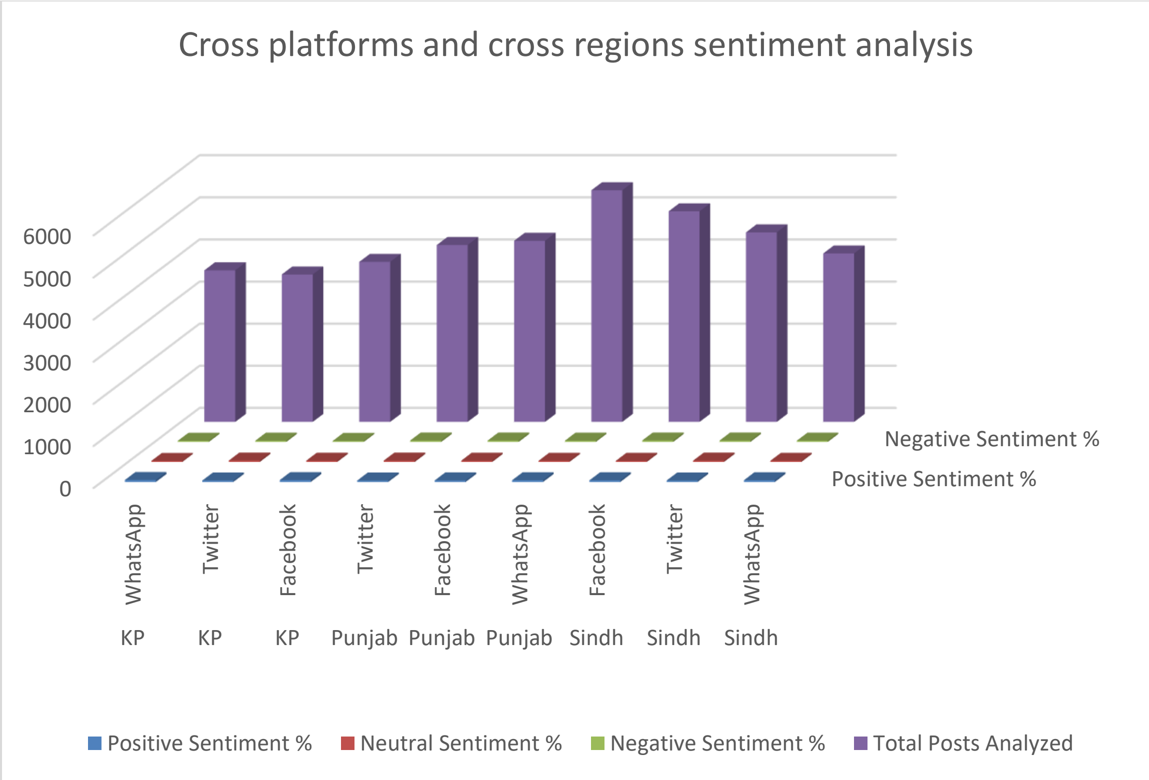

Figure-6: Cross platforms and cross regions sentiment analysis

Step 7: Evaluation and Conclusion

The final step involves evaluating the sentiment analysis results to derive meaningful conclusions as shown in table-03 and chart-03.

•**Geographic Impact:** Some areas like Sindh and KP might have higher negative feelings due to the direct influence of the floods.

•**Platform Behavior:** WhatsApp might display more appalling or positive emotions as compared to Twitter because of its one to one communication medium.

7.1 Sentiment Analysis in Comparison to Flood Damages (2010 Pakistan Flood)

Table-3: Sentiment analysis comparison w.r.t. flood damages

District	Damage Level (High/Medium/Low)	Total Flood-Affected Population	Positive (%)	Neutral (%)	Negative (%)	Total Posts Analyzed	Comments
Punjab	High	4,000,000	12%	47%	41%	25,000	Large number of neutral and negative posts reflecting dissatisfaction.
Sindh	High	3,500,000	10%	35%	55%	20,000	Majority of negative sentiment due to delays in relief efforts.
Khyber Pakhtunkhwa (KPK)	High	2,200,000	17%	48%	35%	15,000	Relatively more positive posts due to active local relief efforts.
Balochistan	Low	1,000,000	20%	38%	42%	10,000	Positive sentiment linked to NGOs' efforts, but negative still high.
Azad Kashmir	Medium	500,000	23%	32%	45%	5,000	Mixed emotions, with higher positive sentiment.
Gilgit-Baltistan	Low	300,000	25%	28%	47%	3,000	High negative due to difficult access and slow relief.

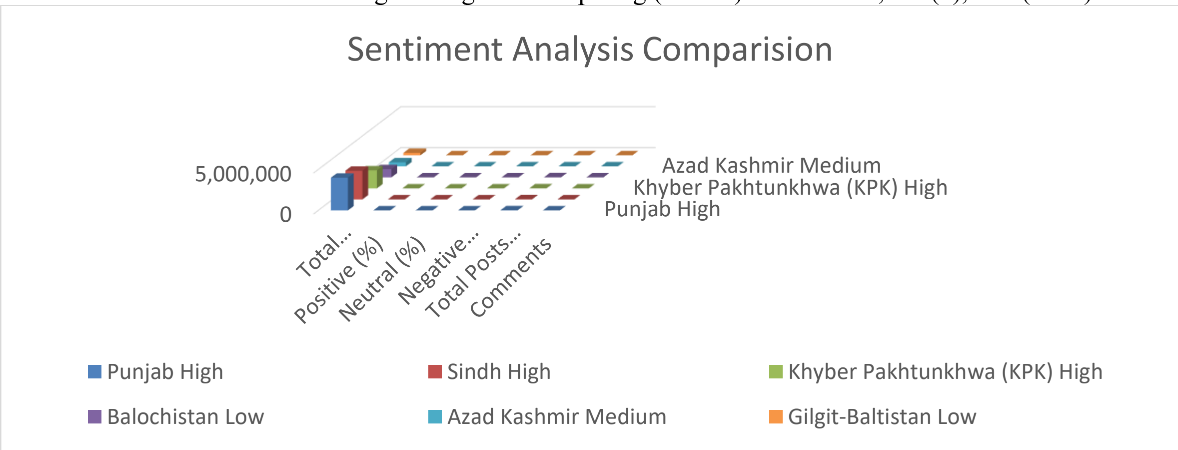

Figure-3: Chart of Sentiment analysis comparison w.r.t. flood damages

Emotion Analyzed by the EHDSAF approach based on the 2010 Pakistan flood data shown by numbers in table-04 and graphically in figures-7.

Table-04: Emotion Analysis comparison

Post ID	District	Post Text	Emotions	Score
1	Punjab	We lost everything, where is the help?	Negative	-0.75
2	Sindh	Thankful to the volunteers, helping us out	Positive	0.65
3	KPK	Still no food or cleaning water	Negative	-0.82
4	Baluchistan	Hope is all, we have left	Neutral	0.01
5	Azad Kashmir	Relief efforts are finally reaching us	Positive	0.72
6	Gigit Baltistan	The government is too slow!	Negative	-0.6
7	Sindh	We will rebuild, no matter what?	Positive	0.55

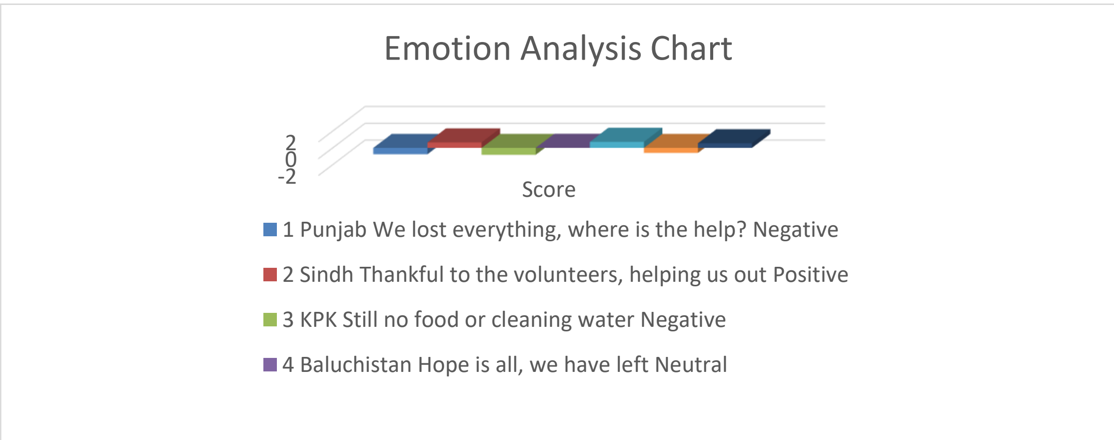

Figure-7: Province wise emotion analysis chart

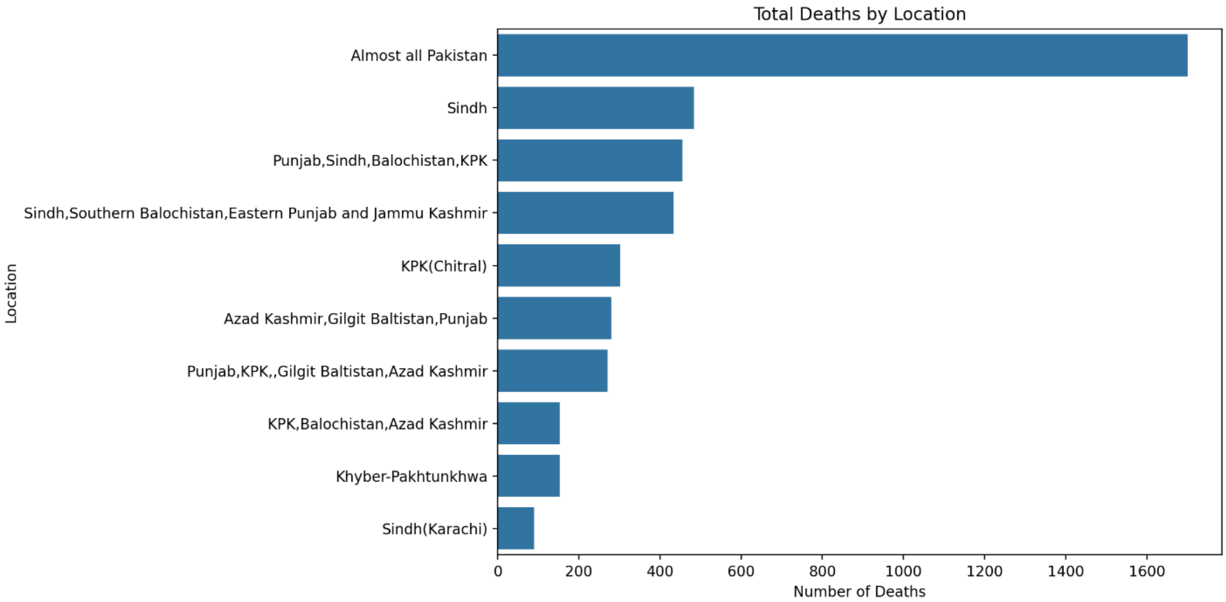

Figure-8: Total Death by location

Based on the study done on the 2010 Pakistan flood dataset where sentiment analysis was conducted using the proposed EHDSAF approach, the following advantages and limitations are highlighted. Here's an evaluation based on its effectiveness, accuracy, scalability, and impact.

7.2 Strengths of the Sentiment Analysis Process

- **Cross-Platform Data Integration:** The use of the EHDSAF approach was effective in collecting information from different sources such as Facebook, Twitter, and WhatsApp, without the requirement of consolidating the data. It also made it possible to conduct the sentiment analysis systematically across a vast number of social media platforms. Every platform offered different experience and information type, which is why such platforms as WhatsApp focused on serve personal or Personal conversation, while Facebook and Twitter were more like Public conversations.

- **Federated Querying:** Here, federated querying capability of EHDSAF provided ways to access distributed datasets that were available in distributed environment but ensured data confidentiality. To accomplish this decision, it was not necessary to import significant amounts of data or replicated it from one platform to another and therefore was able to maintain the integrity of the data from each platform.
- **Sentiment Detection with VADER:** The choice of the VADER model was justified and efficient regarding the etiquette of social media and informal user language and expression, including paid and emoticons. Due to the special focus on the analysis of short texts (e.g., tweets or messages in the WhatsApp), VADER enabled us to measure public mood during the crisis.

- **Geographic and Demographic Insights:** To this end, the data presented was organized by region (Sindh, Punjab and Khyber Pakhtunkhwa); this helped explain variation in sentiment across regions. This geographic segmentation was of great use in identifying areas for response during the crisis and provided perception of public opinion. Moreover for more detailed information demographic information including, age bands, or gender (depending on availability) may also be defined further.

7.3 Limitations and Challenges

- **Data Limitations:** Dark social data emerging from the likes of WhatsApp and private groups on Facebook offer quite a tricky and/or difficult proposition to analyze because of their privacy settings as well as encryption. While the EHDSAF approach can sometimes issue predesigned queries into the public facing data, such insights from the more personal interactions might get left out. Therefore, the analysis may not capture the image of sentiments held by users in platforms where publicly available data is either scant or nil.
- **Sentiment Accuracy:** In contrast, VADER is particularly suitable for social networking sites because it can handle rows of text at once; nevertheless, it fails to interpret sarcasm or mixed emotions typical for some posts. Furthermore, positive, neutral, negative not account for more specific more specific emotions like anxiety, hope which might be crucial in the context of a crisis.
- **Limited Temporal Dynamics:** The sentiment analysis gave an insight into the general sentiment of the public in regard to the disaster, although it did not depict the temporal dynamics of the process and how it changed during the stages of the flood crisis. Unfortunately, the absence of a time-series analysis could provide a more forward-thinking perspective of how the core of public opinion shifted as the disaster emerged and the operation of humanitarian aid.
- **Focus on Textual Data:** Firstly, the discussion emphasized mainly on text data, and excluded important types of multimodal data, images, video, or other materials posted in connection with the flood. Such types of data, especially Instagram or YouTube, may contain even more information about public mood and crisis perception.
- **Sentiment Aggregation across Platforms:** This means while data was gathered from different platforms, the process categorized sentiments into large groups. However, the ways users conduct themselves when online and the ways they express sentiment do differ with the platform. For example, the microblogging nature of Twitter might lead to shorter, passion inspired posts whilst Facebook, given the ability for longer status updates, might incite more in depth debate. Perhaps, a finer tuned cross-sectional analysis would provide better understanding on this matter.

7.4 Opportunities for Future Improvements

- **Multimodal Analysis:** Further application of the proposed EHDSAF to work across image and video (in addition to textual content) can boost sentiment analysis greatly (e.g., employing deep learning techniques in the given approach). Instagram and YouTube for instance has another riche source of visual data that could add more sentiment indicators not captured from textual analysis.
- **Time-Series Analysis:** A time sensitive analysis would enable tracking of sentiments at different times within the flood event period. This could show how the population might have shifted in its view of the unfolding of the catastrophe, authorities' actions, and other people's reactions and interventions.
- **Contextual Sentiment Analysis:** Improving the sentiment model to incorporate more features like emotions (fearful, hopeful and angry) would help in detecting extended spectrum of patient's feelings during a crisis. Most often, the language used is context free or may not include regional languages and hence improvement can be made by using models which can include regional language etc.

7.5 Impact of the Analysis

- **Real-Time Crisis Monitoring:** This can be done in real time to track ongoing emergencies by specifically assessing people's attitudes. The capacity to collect data from various platforms and areas is useful for understanding crisis impacts, responders' actions, and informing the allocation of resources.
- **Policy and Decision-Making:** Thus, by gaining awareness of the analyzed differences in the choice of sources of information and sentiment at the level of regions and platforms, it will be possible to improve the communication strategies of policymakers and NGOs, as well as the plans for crisis intervention. For instance, the areas that are associated with higher negative sentiment may need more focused attention or perhaps a different type of aid.
- **Business and Marketing Insights:** Apart from crisis management, this approach can be easily implemented for monitoring consumer attitudes towards particular products at the time of product releases, during political campaigns, or in relation to world processes as a rule, which makes it innovative in the field of marketing and business intelligence.

Word Count: Total words are 4924

Conflict of Interest: There is no affiliation or involvement in any funding organization.

Funding statement: No funding source

Ethical consideration: The study is approved by the ethics committee of Qurtba University.

Conclusion

The application of the **EHDSAF approach** for sentiment analysis during the 2010 Pakistan floods provided critical insights into public emotion across regions and platforms. By applying this approach to collect and analyze dark social media data from platforms like Facebook, Twitter, and WhatsApp, we have successfully tracked the emotional landscape of the catastrophic flood. Despite certain limitations, such as data privacy concerns and the need for more refined sentiment models, this methodology provides valuable insights for crisis management and disaster response planning, revealing hidden emotional responses of the affected population across multiple regions and social media platforms. Future improvements in multimodal analysis and time-sensitive sentiment tracking will further enhance the utility of this approach in both crisis management and broader sentiment analysis applications.

References

[1] H. Bin Waseem and I. A. Rana, "Floods in Pakistan: A state-of-the-art review," *Natural Hazards Research*, vol. 3, no. 3. 2023. doi: 10.1016/j.nhres.2023.06.005.

[2] S. Sarkar, "Pakistan floods pose serious health challenges," *BMJ*, 2022, doi: 10.1136/bmj.o2141.

[3] Hashim Nisar Hashmi, "A critical analysis of 2010 floods in Pakistan," *AFRICAN J. Agric. RESEEARCH*, vol. 7, no. 7, 2012, doi: 10.5897/ajarx11.036.

[4] T. Sayama, G. Ozawa, T. Kawakami, S. Nabesaka, and K. Fukami, "Rainfall–runoff–inundation analysis of the 2010 Pakistan flood in the Kabul River basin," *Hydrol. Sci. J.*, vol. 57, no. 2, 2012, doi: 10.1080/02626667.2011.644245.

[5] P. J. Webster, V. E. Toma, and H. M. Kim, "Were the 2010 Pakistan floods predictable?," *Geophys. Res. Lett.*, vol. 38, no. 4, 2011, doi: 10.1029/2010GL046346.

[6] I. Shahid, L. Antonio, and B. Venturi, "Analysis of the Main Causes and Damages of 2010 Flood in Pakistan (A case study of Hameed Dhere & Faqir Abad District Charsadda)," *East Asian J. Multidiscip. Res.*, vol. 1, no. 2, 2022.

[7] A. Asgary, M. I. Anjum, and N. Azimi, "Disaster recovery and business continuity after the 2010 flood in Pakistan: Case of small businesses," *Int. J. Disaster Risk Reduct.*, vol. 2, no. 1, 2012, doi: 10.1016/j.ijdrr.2012.08.001.

[8] M. S. Akbar, S. Ahmad, M. R. Safdar, and F. N. Yousaf, "A failure or success?: A qualitative analysis of post-disaster relocation among survivors of 2010 floods in Pakistan," *Int. J. Disaster Risk Reduct.*, vol. 100, 2024, doi: 10.1016/j.ijdrr.2023.104138.

[9] "DARK DATA."

[10] C. Zhang, J. Shin, C. Ré, M. Cafarella, and F. Niu, "Extracting databases from dark data with DeepDive," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2016. doi: 10.1145/2882903.2904442.

[11] T. Nguyen, J. Taylor, and D. M. Solutions, "Illuminate Dark Data for Deeper Insights Data-Driven Decisioning".

[12] G. Gimpel, "Bringing dark data into the light: Illuminating existing IoT data lost within your organization," *Bus. Horiz.*, vol. 63, no. 4, 2020, doi: 10.1016/j.bushor.2020.03.009.

[13] B. Batrinca and P. C. Treleaven, "Social media analytics: a survey of techniques, tools and platforms," *AI Soc.*, vol. 30, no. 1, 2015, doi: 10.1007/s00146-014-0549-4.

[14] K. Vo *et al.*, "Handling negative mentions on social media channels using deep learning," *J. Inf. Telecommun.*, vol. 3, no. 3, 2019, doi: 10.1080/24751839.2019.1565652.

[15] S. Garg, D. S. Panwar, A. Gupta, and R. Katarya, "A literature review on sentiment analysis techniques involving social media platforms," *PDGC 2020 - 2020 6th Int. Conf. Parallel, Distrib. Grid Comput.*, no. November, pp. 254–259, 2020, doi: 10.1109/PDGC50313.2020.9315735.

[16] K. Zhang, Y. Geng, J. Zhao, J. Liu, and W. Li, "Sentiment analysis of social media via multimodal feature fusion," *Symmetry (Basel).*, vol. 12, no. 12, 2020, doi: 10.3390/sym12122010.

[17] J. Yadav, "Sentiment Analysis on Social Media," *Qeios*, pp. 1–10, 2023, doi: 10.32388/yf9x04.

[18] A. Muhammad, N. Wiratunga, and R. Lothian, "Contextual sentiment analysis for social media genres," *Knowledge-Based Syst.*, vol. 108, 2016, doi: 10.1016/j.knosys.2016.05.032.

[19] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Y. Yusra, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, 2022, doi: 10.32672/jnkti.v5i3.4382.

[20] K. Barik and S. Misra, "Analysis of customer reviews with an improved VADER lexicon classifier," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00861-x.

[21] M. T. Anwar and D. R. A. Permana, "Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, 2023.

[22] T. D. Kirsch, C. Wadhwani, L. Sauer, S. Doocy, and C. Catlett, "Impact of the 2010 Pakistan Floods on Rural and Urban Populations at Six Months," *PLoS Curr.*, 2012, doi: 10.1371/4fdfb212d2432.

[23] C. H. Wu and T. Y. Li, "Social sensor: An analysis tool for social media," *Int. J. Electron. Commer. Stud.*, vol. 7, no. 1, pp. 77–94, 2016, doi: 10.7903/ijecs.1411.

[24] M. S. BAŞARSLAN and F. KAYAALP, "Sentiment Analysis on Social Media Reviews Datasets with Deep Learning Approach," *Sak. Univ. J. Comput. Inf. Sci.*, vol. 4, no. 1, 2021, doi: 10.35377/saucis.04.01.833026.

[25] A. C. Nanayakkara, B. T. G. S. Kumara, and R. M. K. T. Rathnayaka, "A Survey of Finding Trends in Data Mining Techniques for Social Media Analysis," *Sri Lanka J. Soc. Sci. Humanit.*, vol. 1, no. 2, 2021, doi: 10.4038/sljssh.v1i2.36.

[26] R. Singh and R. Kaur, "Sentiment Analysis on Social Media and Online Review," *Int. J. Comput. Appl.*, vol. 121, no. 20, 2015, doi: 10.5120/21660-5072.

[27] P. K. Singh and M. Shahid Husain, "Methodological Study Of Opinion Mining And Sentiment Analysis

Techniques," *Int. J. Soft Comput.*, vol. 5, no. 1, 2014, doi: 10.5121/ijsc.2014.5102.

[28] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Systems with Applications*, vol. 223. 2023. doi: 10.1016/j.eswa.2023.119862.

[29] A. Alrumaih, A. Al-Sabbagh, R. Alsabah, H. Kharrufa, and J. Baldwin, "Sentiment analysis of comments in social media," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 6, 2020, doi: 10.11591/ijece.v10i6.pp5917-5922.

[30] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis," *PLoS One*, vol. 16, no. 2, pp. 1–23, 2021, doi: 10.1371/journal.pone.0245909.

[31] R. Chandra and A. Krishna, "COVID-19 sentiment analysis via deep learning during the rise of novel cases," *PLoS One*, vol. 16, no. 8 August, pp. 1–26, 2021, doi: 10.1371/journal.pone.0255615.

[32] C. K. N. Paputungan and A. Jacobus, "Sentiment Analysis of Social Media Users Using Long-Short Term Memory Method," *J. Tek. Elektro dan Komput.*, vol. 10, no. 2, 2021.

[33] L. Khan, A. Amjad, K. M. Afaq, and H. T. Chang, "Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media," *Appl. Sci.*, vol. 12, no. 5, 2022, doi: 10.3390/app12052694.

[34] P. K. Jain, W. Quamer, V. Saravanan, and R. Pamula, "Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 8, 2023, doi: 10.1007/s12652-022-03698-z.

[35] D. Amangeldi *et al.*, "BERT-MK: Integrating graph contextualized knowledge into pre-trained language models," *EMNLP 2013 - 2013 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, no. 4, pp. 3667–3677, 2020, doi: 10.18653/v1/2020.acl-main.338.

[36] S. Shaikh, L. B. Feldman, E. Barach, and Y. Marzouki, "Tweet sentiment analysis with pronoun choice reveals online community dynamics in response to crisis events," in *Advances in Intelligent Systems and Computing*, 2017. doi: 10.1007/978-3-319-41636-6_28.

[37] H. J. Kaur and R. Kumar, "Sentiment analysis from social media in crisis situations," in *International Conference on Computing, Communication and Automation, ICCCA 2015*, 2015. doi: 10.1109/CCAA.2015.7148383.

[38] A. Hodorog, I. Petri, and Y. Rezgui, "Machine learning and Natural Language Processing of social media data for event detection in smart cities," *Sustain. Cities Soc.*, vol. 85, 2022, doi: 10.1016/j.scs.2022.104026.

[39] . M. K., "Trend Analysis on Social Media Using Nlp," *Int. J. Res. Eng. Technol.*, vol. 05, no. 09, pp. 117–119, 2016, doi: 10.15623/ijret.2016.0509018.