Machine Learning and Artificial Intelligence
J.-L. Kim (Ed.)

doi:10.3233/FAIA230763

Discovering Insights via Hybrid Thematic Analysis: A Case Study on Disaster Risk Reduction and Management for Legazpi City, Albay

Mideth ABISADO [a,1] Lany MACEDA [b] Ramon RODRIGUEZ [a]
Joseph Marvin IMPERIAL [a] Myron Darrel MONTEFALCON [a] Jay Rhald PADILLA [a]
Gizelle PONCE [a]

[a] *College of Computing and Information Technologies, National University, Sampaloc, Manila 1008, Philippines*
[b] *Bicol University, Legazpi, Philippines*

Abstract. Understanding the perceptions and experiences of a community regarding disasters is crucial in effectively planning and implementing disaster strategies. There are two known approaches to analyzing perceptions, qualitative and automated approaches for thematic analysis. This paper aims to investigate the strengths and limitations of the mentioned approaches. Thus, using both approaches, this study analyzed data from a focus group discussion about disaster risk reduction and management and climate change adaptation conducted in a typhoon-prone city in Legazpi City, Philippines. An inductive-deductive approach for the qualitative method while a language model-assisted approach for the automatic method of extracting prominent themes from the collected responses. The results show (dis) similarities regarding themes obtained from the two approaches, specifically the emphasis on concerns about the proper distribution of relief goods and donations, proper early monitoring of potentially powerful typhoons, and other forms of threat, including politically motivated ones. From these findings, we conclude the importance of incorporating a combined manual and automatic approach for thematic analysis of natural language.

Keywords. Qualitative Approach, NLP-based Approach, Thematic Analysis, DRMM-CCA, KeyBert Architecture

1. Introduction

The Philippines is highly vulnerable to natural disasters due to its location in the Pacific Ring of Fire, making it prone to frequent typhoons and earthquakes [1]. According to World Risk Index in the year 2022 [2], the Philippines has the highest disaster risk among 193 countries worldwide, scoring high in its exposure, vulnerability, susceptibility, lack

[1] Corresponding Author: Mideth B. Abisado, College of Computing and Information Technologies, National University, Philippines, Email: mbabisado@national-u.edu.ph

of response capacities, and lack of adaptive capacities in the face of disasters. For the Philippines to effectively respond to these disasters and mitigate their consequences, it is crucial to understand the needs and perspectives of those affected. Listening to citizens' perceptions in Disaster Risk Reduction and Management (DRRM) can help ensure that all perspectives and needs are considered in decision-making. The community has first-hand experience and knowledge of local risks and hazards. Moreover, engaging the citizens and getting their perspectives promotes inclusivity, improved planning, and increases their sense of ownership of the community, which can provide valuable insights into DRRM efforts.

The primary challenge for collecting and analyzing responses from the community is that it is time-consuming, labor-intensive, and prone to human errors [3]. With this, researchers have turned to use automated approaches via Natural Language Processing (NLP). There have been several applications of NLP-based analysis in disaster-related phenomena, including using NLP on social media analytics, as vast amounts of information is generated in social networking platforms during disasters. The studies [4]-[5] implemented NLP techniques to analyze disaster-related tweets from Twitter, understanding behavior and obtaining valuable insights from the citizens in real-time. Similarly, the work [6] performed sentiment analysis on disaster-related tweets, which enabled an understanding of the public's emotions and opinions. Aside from that, NLP-based approaches can be applied to cater to multilingual data [7] as information about disasters may be shared in multiple languages, allowing for a more comprehensive understanding of the event.

This paper aims to compare the effectiveness of the qualitative method and NLP-based approaches in analyzing responses from a focus group discussion conducted in a specific community about DRRM and CCA. Comparing these two approaches aims to shed light on uncovering important and unique themes as a tool for accurate analysis.

2. Review of related literature

Thematic analysis is an approach that uncovers the underlying meanings and perspectives that shape the experiences of individuals and communities. We divide the extensive literature on thematic analysis approaches in two with respect to their methodology: (a) conventional approaches using human interaction through interviews and focus group discussions (FGD) and (b) automated approaches using Natural Language Processing.

2.1. Conventional thematic analysis with human interaction

Several papers [8,9] have applied a qualitative approach to analyzing data on disaster-related topics in the Philippines. In the study of [8], the researchers examined the experiences of youth survivors of Typhoon Haiyan through an interview to gather each participant's own narrative of the event. The study found that the experience of a natural disaster can profoundly impact the self and identity of youth survivors, leading to changes in their physical and emotional well-being, relationships, and sense of self. Similarly, the study of [9] investigated the evacuation response of affected communities during the 2014 Mayon Volcano eruption in the Philippines. The authors used a qualitative method through interviews with the evacuated residents, local officials, and first respon-

ders to understand the factors that influenced the evacuation response and the challenges communities faced during the evacuation. The study found that evacuation response was influenced by community readiness, planning, communication, and transportation and highlighted the importance of these factors for effective evacuation. However, the use of the conventional qualitative approach for analysis also has some limitations as it can be time-consuming and subjective which could lead to potential bias [11].

2.2. Thematic analysis using natural language processing techniques

Using NLP techniques for thematic analysis has become increasingly popular in recent years. In the paper of [5], the researchers analyze the behavior of Filipino Twitter users using Latent Dirichlet Allocation (LDA) and Principal Component Analysis to extract the different topics discussed during a disaster. The study highlights that Filipino Twitter users shared various information related to the disaster, including updates on the disaster's progression, personal experiences, and calls for help. Similarly, in the works of [12] and [13], the authors performed NLP-based techniques like topic modeling and sentiment analysis on typhoons and earthquake-related tweets in the Philippines. Another NLP-based paper [14], utilizes 976 suggestions on how their village can help them better prepare for a disaster implemented computational methods, specifically topic modeling and word2vec, to assist in the analysis of qualitative data on disaster risk reduction suggestions. Based on the results, computational methods improved the efficiency and accuracy of the data analysis process. The study's findings show that the participants give importance to community preparedness for an emergency, helping the village in clean-up drives, and awareness through seminars and information dissemination.

2.3. Hybrid approaches

Combining the conventional qualitative thematic analysis with Natural Language Processing (NLP) techniques offer several advantages that merit exploration. By combining both approaches, researchers can leverage the strengths of each method to enhance the accuracy and efficiency of their analysis. Moreover, NLP techniques can help researchers identify themes and patterns that might not have been apparent in manual analysis, thus leading to more comprehensive and nuanced insights. Ultimately, integrating conventional and NLP-based approaches can provide a complete understanding of the data and facilitate the development of evidence-based strategies and interventions in disaster-related events.

2.4. Data collection through focus group discussions

A **Focus Group Discussion – Structured Learning Exercise (FGD-SLE)** was employed as a collection method as it provides a way to collect rich in-depth data from a group of people. Following the Risk-Based Approach in Community-Based DRRM, we have decided to utilize FGD-SLE to uncover new insights regarding the local grassroots community's knowledge and understanding of a complex issue such as DRRM-CCA.

This case study gathered DRRM-CCA perceptions and responses in the community of **Brgy. Padang, Legazpi City, Albay** in the Philippines. The chosen community is based on the endorsement of the Local Government Unit (LGU) as it is labeled as *village at risk* and the mentioned community has become infamous for the onslaught of Typhoon

Reming in the year 2006 [15] which caused a lot of casualties. Additionally, based on the Risk-based approach assessment, Brgy. Padang is both a high vulnerability index and capacity or response measures. Thus, this community is selected for the initial FGD-SLE to gather DRRM-CCA responses.

In the data collection process, it employed a purposive sampling method for the FGD-SLE. Wherein the participants of the FGD-SLE included sectoral and marginalized groups from Brgy. Padang, Legazpi City, Albay. A total of 16 participants included a diverse group of individuals such as women, youth, farmers, fisherfolks, persons with disability, the elderly, and local leaders of Brgy. Padang age ranges from 24 years old to 70 years old. The FGD-SLE was facilitated in a closed-door venue with activities like group workshops, group reporting, and breakout sessions. During the FGD-SLE, the responses were recorded using an audio recorder, with a signed waiver of data privacy and protection for each participant. A total of 4 hours of audio recording was collected. After the FGD-SLE, the audio data was validated, and the recording was transcribed before undergoing the analysis.

3. Experiment setup

This paper aims to provide an analysis comparing the qualitative thematic analysis approach with the NLP-based technique in getting themes and insights from the FGD-SLE data collected from the participants. This will help understand the strengths and limitations of both approaches empirically. The goal of the thematic analysis is to identify and understand the themes, patterns, or structures present throughout the FGD-SLE data that would help develop the questions for the DRRM-CCA component of the project.

3.1. Inductive-deductive approach

For the conventional qualitative approach, an **inductive-deductive analysis** was utilized, wherein the inductive phase was used to reveal the central themes. An inductive research method involves an analysis of texts and reasoning that stream from a set of empirical observations toward a systematic formulation of theories or conclusions [16]. This method has strengths drawn for qualitative research, whereby flexibility and the bottom-up, iterative process of discovering themes and patterns from raw data take priority over predetermined ideas and notions. We have decided to use an inductive approach as the study entails a more progressive qualitative, community-based, and participatory approach to research. Afterward, a deductive analysis was employed, to derive codes from the central themes of the specified community.

3.2. Language-model assisted keyword extraction

For the automated extraction of themes from the response data, we use **KeyBERT** [17], a simple neural language model-based architecture for keyword extraction from documents. KeyBERT leverages the use of text representations from a **Bidirectional Encoder Representations from Transformers (BERT)** language model called embeddings [18]. These embeddings contain rich linguistic information about the text such as semantic knowledge. We use this approach over statistical-based or static keyword extraction methods as it is a more applicable study since our goal is to obtain contextually-

similar terms that can represent the entire document. The process of KeyBERT keyword extraction starts with transforming a dataset $D = d_1, d_2, d_3 ... d_n$ into its corresponding document embeddings $H(d)$ by feeding it to a selected BERT model. Then, word-level representations $H(w)$ are also extracted for all unique terms of the document. Finally, each of the terms from the extracted vocabulary will be paired with the document via cosine similarity $\text{COS}(d_i, w_i)$, where the top words garnering the highest similarity scores are labeled as the most prominent keywords that can represent the context of the dataset.

For this study, since we are dealing with responses from the participants in transcribed in Tagalog, we use the robust version Tagalog BERT model called RoBERTa as the main language model of choice [19]. We set the parameters of KeyBERT to generate 10 potential keyword groups with each containing 5 candidate keywords. We then evaluate these groups to obtain the top five most semantically related to the document for analysis.

3.3. Evaluation metric

For any automatic analysis method conducted with real-world data, it is important to evaluate the results obtained from the experimental processes to ensure the reliability of the findings. In this work, we evaluate each of the 10 keyword groups generated from the KeyBERT experiment using the coherence-type metric **UMASS** [20]. The UMASS score is an intrinsic document co-occurrence based metric where (w_i, w_j) denotes the number of times both words w_i and w_j occurred in a document D normalized by the number of times w_i appears in the document. The formula is listed below:

$$\text{UMASS}(w_i, w_j) = log \frac{D(w_i, w_j) + 1}{D(w_i)} \tag{1}$$

4. Prominent disaster-related Themes

4.1. Themes from focus group discussions

Based on processing the maps and the group reports, significant themes were decoded from the FGD-SLE activity. Table 1 presents the main themes including **Perceptions of DRRM-CCA**, **LGU's Programs and Projects**, and **Community-Based Solutions and Recommendations**.

From the first theme, the participants shared personal reflections, indigenous knowledge, and perceived problems with DRRM-CCA in their community. From the reported risk-mapping activity, it is evident that the participants view typhoons as the most significant risk event, followed by volcanic eruptions at Mt. Mayon, the rainy season as medium risk, and earthquakes as low risk. Results from the activity suggest that the participants understand the local disaster risks in their village and that typhoons are perceived as the most significant hazard in their area. The second code highlights the perceived DRRM-CCA-related challenges, including trauma, financial difficulty, water scarcity, lack of sanitary toilets, stress, diseases, difficulty with household chores, and various health problems. Additionally, the participants reported a lack of proper knowledge and preparation for disasters, with only a basic understanding of the "drop, cover, and hold" method—a prominent must-do in the government's earthquake information

and education campaigns. The vulnerability map covers this theme, as the participants also reported the importance of social assistance and social safety nets (e.g., cash or crops, water and food resources, and medical support) during disaster recovery and rehabilitation.

The second theme relates to the FGD-SLE activity gauging the participants' understanding of the LGU's programs, projects, and policies in the community. It is evident that the community participants are highly aware of the LGU's programs and policies on DRRM-CCA. The participants can discuss scenarios and hazards where the LGU has prepared evacuation centers and early-warning equipment, divided responsibilities among staff, assisted senior citizens and traumatized evacuees, monitored health and food, and provided relief assistance to the affected constituents of Brgy. Padang. The method also captures the strengths and assets of the community through their positive perceptions of their LGU's programs and policies, including the character of leaders, tools, local expertise, leadership, and connections to other groups. According to the participants, the emergency local meeting and village staff play a crucial role in applying these strengths and ensuring the community's preparedness for disasters.

The last theme focuses on community-based solutions and disaster preparedness and resilience recommendations. This theme is encapsulated in the last group's recommendatory report on the community's actions to build resilience against disasters. The participants emphasized in their report the importance of individuals being self-sufficient and not overly dependent on the village LGU. To enhance resilience against disasters and climate change, the participants suggested educating the youth and increasing their early awareness and orientation to the impacts of disasters. They believe that confidence, courage, and faith in God are vital for the community to handle disasters effectively. According to them, with or without LGU support, the local people of Brgy. Padang must enhance their self-sufficiency and empower themselves to handle the various disasters in their way. They further noted that this is the only way for them to be sustainable in disaster prevention.

Table 1. Results from qualitative thematic analysis approach: inductive-deductive.

Theme	Code	Examples
Perceptions on DRMM-CCA	Personal Reflections and Knowledge on DRRM-CCA	The participants view typhoons as high risk, followed by a volcanic eruption at Mt. Mayon. They view the rainy season as medium risk and earthquakes as low risk.
	Perceived Problems on DRRM-CCA in the Community	Challenges reported by participants include trauma, financial difficulties, water scarcity, lack of sanitary toilets, stress, and various illnesses such as colds, coughs, and fevers.
LGU DRMM-CCA Programs and Policies	Awareness of LGU Programs and Policies on DRRM-CCA in the Community	The community exhibits strengths that enhance its crisis resilience, such as discussing scenarios, preparing evacuation centers and equipment, aiding vulnerable members, monitoring health and food, and providing relief assistance. The emergency CDRRMO local meeting and barangay staff help identify needs and damages, prepare centers and equipment, monitor the situation, check supplies, and distribute aid.
	Perception of Effectiveness of LGU DRRM-CCA Programs and Policies	Participants in Brgy. Padang reported ample disaster preparation programs from the LGU, including having food and non-food items as essential goods, health, and security protocols, and resources like medicine, flashlight, and gas.
Community-based Solutions and Recommendations	Community Initiatives and Recommended Actions	Participants in Brgy. Padang reported sufficient disaster supplies, including food, medicine, flashlights, and diesel. They emphasized the need for self-sufficiency, youth education, and faith in God to increase resilience and effective disaster response.

Table 2. Results from performing automatic keyword extraction using the KeyBERT architecture and Tagalog RoBERTa model.

Labelled Topics Groups	Keywords	UMASS
General Disaster Terms*	vulnerabilities, disasters, climate, point, cause	**-0.207**
Quarrying Concerns*	nagqquarry, disaster, pagpuputol, behavior, basura	**-0.769**
Description of Citizens	kabataang, marites, tambay, nakapagtapos, gurang	-14.072
Advisory Concerns	late, political, advisory, suspensions, advisories	-3.033
Vocational Education	beneficial, tesda, nakakatulong, kakatapos, kaalaman	-7.186
Suspension Concerns	nagbaha, nagsuspinde, klase, happy, bahay	-9.235
Preparation and Monitoring*	pagprepara, monitoring, typhoon, centers, bhw	**-0.069**
Donations*	mabibigay, diskarte, kailangan, pangangailangan, laging	**-0.496**
Volcanic Eruptions	forced, mayon, bulkan, evacuation, ashfall	-9.787
Local Politics*	nadadamay, barilan, pulitiko, lugar, violence	**0.009**

4.2. Themes from automatic Analysis

Table 2 presents the top 10 extracted thematic keywords using the KeyBERT method with the Tagalog RoBERTa model. From the results, we highlight five topics which can be considered prominent due to their relatively high UMASS score compared to the others. These topics are in the context of **general disaster terms, concerns on quarrying, preparation and monitoring, concerns on donations**, and **local politics**. Similar to related work [14], these themes were manually labeled based on the context provided by their supporting topic keywords as seen in the second column of Table 2.

From the prominent topics, we draw similarities and dissimilarities with the conventional qualitative approach presented in Table1. The common theme that appeared from both methods can be attributed to the context of **concerns on donations**. Associated words that make up this topic through the KeyBERT approach are *kailangan* or *need*, *magbibigay* or *to give*, and *pangangailangan* or *needs*. These words are contextually usable during times of *need* as seen in the second theme from the conventional qualitative approach which is LGU DRRM-CCA Programs and Policies. Interestingly, one word from the automatic method, *diskarte* or *the act of hustling or resourceful*, may be associated with the special attitude of Filipino people to exhaust all means to survive in times of disasters.

Another common theme from both methods is **preparation and monitoring**. As mandated, ensuring enough preparation of food stocks and accurate monitoring of any signals of incoming natural disasters such as heavy rains are under the responsibility of LGU. This concern is also highlighted in the second theme from the qualitative approach. Associated words formed through the KeyBERT method are *pagprepara* or *act of preparation*, *centers, monitoring*, and *typhoon*. Another prominent theme is associated with **quarrying concerns**, which can be linked to concerns of local communities towards illegal quarrying or mining activities and have been a long-time problem in the city[2]. Supporting thematic words building this context are *nagququarry* or *quarrying*, *disaster*, *pagpuputol* or *cutting*, *behavior*, and *basura* or *trash*. This particular concern can also be seen in the results from the qualitative approach which can be classified under the second code of the first theme which is Perceived Problems with DRRM-CCA in the Community.

[2]https://www.pna.gov.ph/articles/1122671

Lastly, using the automatic approach extracted an interesting topic which is **local politics**. This particular theme is associated with convincingly negative terms such as *nadadamay* or *affected*, *barilan* or *shooting*, and *violence*. This theme may suggest that even if there are other major concerns regarding natural disasters in the community, other threats to safety and security, including those that are politically motatived, can still cause worry and concerns to members of the community.

5. Conclusions and future Work

This study explored the perceptions and experiences of a typhoon-prone community in Legazpi City, Philippines, on disaster risk reduction and management (DRRM) and climate change adaptation (CCA). Our findings demonstrate the importance of understanding community perspectives to develop effective disaster management strategies. We used the conventional qualitative and NLP-based approaches to analyze the collected responses and identified key themes from both approaches. Comparing the results of the two approaches, we found strong similarities; both emphasized the need for proper distribution of relief goods and early monitoring of potentially powerful typhoons and other forms of threats, including politically motivated ones. This highlights the capability of using an NLP-based approach in data analysis as it can obtain similar insights from the conventional qualitative approach while uncovering underlying themes. On the other hand, some themes were only present in one of the approaches, indicating the importance of implementing both approaches, as it can provide a clearer understanding of the responses of the community.

Acknowledgments

We gratefully acknowledge the financial support the Philippine Commission on Higher Education (CHED) provided under Grant No. 2021-007. This funding has been instrumental in advancing our research endeavors and enabling us to make significant contributions to our field. We extend our sincere appreciation for CHED's continued commitment to the advancement of scientific knowledge and innovation.

References

[1] Bollettino V, Alcayna T, Enriquez K, Vinck P. Perceptions of disaster resilience and preparedness in the Philippines. Harvard University: Cambridge, MA, USA. 2018 Jun.
[2] Franziska A, Lotte K, Beáta P, Kristin BS, Sören S, Daniel W, Bündnis EH. World Risk Report 2022—Focus: Digitalization.
[3] Mason H, Patil DJ. Data Driven. Creating a data culture; 2015.
[4] Takahashi B, Tandoc Jr EC, Carmichael C. Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. Computers in human behavior. 2015 Sep 1;50:392-8.
[5] Lee Boaz J, Ybañez M, De Leon MM, Estuar MR. Understanding the behavior of Filipino Twitter users during disaster. GSTF Journal on Computing (JoC). 2013 Aug;3:1-6.
[6] Zhang T, Cheng C. Temporal and spatial evolution and influencing factors of public sentiment in natural disasters—a case study of typhoon haiyan. ISPRS International Journal of Geo-Information. 2021 May 5;10(5):299.

[7] Fuadvy MJ, Ibrahim R. Multilingual sentiment analysis on social media disaster data. In2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE) 2019 Oct 3 (Vol. 6, pp. 269-272). IEEE.
[8] Delante N. Natural disasters reconstruct the self: A thematic analysis of youth survivors' narratives about typhoon haiyan. Journal of Narrative and Language Studies. 2019 Dec 30;7(13):113-29.
[9] Martinez-Villegas MM, Solidum RU, Saludadez JA, Pidlaoan AC, Lamela RC. Moving for safety: a qualitative analysis of affected communities' evacuation response during the 2014 Mayon Volcano eruption. Journal of Applied Volcanology. 2021 Dec;10(1):1-24.
[10] Neuendorf KA. The content analysis guidebook. sage; 2017.
[11] Braun V, Clarke V. Using thematic analysis in psychology. Qualitative research in psychology. 2006 Jan 1;3(2):77-101.
[12] Maceda LL, Llovido JL, Palaoag TD. Corpus analysis of earthquake related tweets through topic modelling. International Journal of Machine Learning and Computing. 2017 Dec;7(6):194-7.
[13] Imperial JM, Orosco J, Mazo SM, Maceda L. Sentiment analysis of typhoon related tweets using standard and bidirectional recurrent neural networks. arXiv preprint arXiv:1908.01765. 2019 Aug 3.
[14] Gorro K, Ancheta JR, Capao K, Oco N, Roxas RE, Sabellano MJ, Nonnecke B, Mohanty S, Crittenden C, Goldberg K. Qualitative data analysis of disaster risk reduction suggestions assisted by topic modeling and word2vec. In2017 International Conference on Asian Language Processing (IALP) 2017 Dec 5 (pp. 293-297). IEEE.
[15] Orense RP, Ikeda M. Damage caused by typhoon-induced lahar flows from Mayon Volcano, Philippines. Soils and foundations. 2007 Dec 1;47(6):1123-32.
[16] Vanderstoep SW, Johnson DD. Research methods for everyday life: Blending qualitative and quantitative approaches. John Wiley & Sons; 2008 Dec 22.
[17] Grootendorst M. KeyBERT: Minimal keyword extraction with BERT. Zenodo. 2020.
[18] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018 Oct 11.
[19] Cruz JC, Cheng C. Improving large-scale language models and resources for filipino. arXiv preprint arXiv:2111.06053. 2021 Nov 11.
[20] Mimno D, Wallach H, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. InProceedings of the 2011 conference on empirical methods in natural language processing 2011 Jul (pp. 262-272).