Journal of Posthumanism
2025
Volume: 5, No: 6, pp. 3026-3038
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v5i6.2419

Indonesian Public Sentiment on Disasters: A Topic Modeling Study Using Twitter Data

Bayu Setia Ismawandani[1], Wahyu Wibowo[2], Mochammad Reza Habibi[3], Shuzlina Abdul Rahman[4], Harun Al Azies[5], Windiani[6]

Abstract

This study analyzes Indonesian public responses to natural disasters using Twitter data through Topic Modeling (LDA) and Sentiment Analysis. The process includes data collection, preprocessing, modeling, and sentiment classification. Tweets with the keyword "bencana" were scraped and prepared for analysis. LDA revealed 15 main topics covering disaster causes, impacts, responses, and resilience. Sentiments were categorized as positive, neutral, or negative. Positive sentiments dominated, especially regarding collaboration and response efforts. Neutral sentiments discussed resource distribution and awareness, while negative sentiments focused on dissatisfaction with disaster management and environmental issues. The study offers insights into public discourse on disasters and supports improved response strategies, awareness programs, and policy planning. This research uses NLP techniques to capture public concerns and highlights the need to strengthen community resilience in disaster situations.

Keywords: *Natural Disasters; Topic Modeling; Latent Dirichlet Allocation; Sentiment Analysis; Twitter Data Analysis.*

Introduction

Indonesia's susceptibility to natural disasters stems from its position at the convergence of three major tectonic plates (Erb et al., 2021). and its location within the Pacific Ring of Fire (Airlangga, 2024). This exposes the nation to frequent earthquakes, volcanic eruptions, and floods, with impacts extending beyond geology into social vulnerability shaped by socioeconomic disparities (Fuady et al., 2021). Understanding this interplay is crucial for disaster preparedness and resilience (Rahmafitria et al., 2021). Social media, particularly Twitter, plays a vital role in crisis communication by facilitating rapid information dissemination (Cladis, 2020). Analyzing public discourse on these platforms provides insights into societal resilience. Topic Modeling using Latent Dirichlet Allocation (LDA) helps identify key themes in disaster-related discussions (Zhou et al., 2021), while Sentiment Analysis categorizes public emotions, enhancing our understanding of community responses to disasters (Mendon et al., 2021).

Within the domain of Machine Learning and Natural Language Processing (NLP), Topic

[1] Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, Email: 10611910000071@mhs.its.ac.id
[2] Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, Email: wahyu_w@statistika.its.ac.id, (corresponding author)
[3] Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, Email: reza.habibi@its.ac.id
[4] School of Computing Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia, Email: shuzlina@fskm.uitm.edu.my
[5] Universitas Dian Nuswantoro, Semarang, Indonesia | Research Center for Quantum Computing and Materials Informatics, Email: harun.alazies@dsn.dinus.ac.id
[6] Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, Email: windi@mku.its.ac.id

Modeling serves as a powerful tool for automating the organization, comprehension, search, and summarization of vast electronic archives (Kang et al., 2020). Specifically, LDA, a prevalent topic modeling algorithm, operates by inferring latent themes inherent within documents and allocating the contributions of individual words to these topics(Ferner et al., 2020). Meanwhile, Sentiment Analysis, another fundamental NLP technique, is geared towards deciphering and categorizing the emotional nuances embedded within textual data (Poria et al., 2023). By employing these techniques synergistically, researchers can gain profound insights into the underlying themes and sentiments prevalent within extensive textual corpora, thereby facilitating a deeper understanding of complex phenomena such as societal responses to natural disasters.

Utilizing LDA on Twitter data gathered using the keyword "bencana" (disaster) unveils a spectrum of topics that engage public discourse during or following natural calamities. Subsequent Sentiment Analysis further classifies these topics into neutral, negative, or positive categories, thereby offering a nuanced understanding of community sentiment towards different facets of disaster-related discussions. By adopting this approach, researchers and policymakers can discern priority topics warranting immediate attention to mitigate vulnerability and bolster the resilience of Indonesian communities in the face of adversities. This methodological framework not only facilitates the identification of key areas for intervention but also fosters a more informed and targeted approach towards disaster preparedness, response, and recovery efforts.

The 2020 study by M. Choirul Rahmadan et al. applied LDA Topic Modeling to Twitter data, identifying nine discussion topics during Jakarta's flooding crisis (Choirul Rahmadan et al., 2020). Their findings provided valuable insights into public sentiment and concerns, aiding policymakers in refining disaster response strategies. While prior studies like Rahmadan et al. (2020) focused on specific events, our research expands the scope by integrating LDA and Sentiment Analysis to analyze disaster-related discourse on a national scale. This broader approach uncovers resilience mechanisms and key intervention areas across Indonesia, offering more profound insights into disaster preparedness and response.

Related Work

This study builds upon prior research on social vulnerability, disaster resilience, and public discourse. Bergstrand et al. (2015) analyzed socio-demographic factors and community dynamics in disaster resilience across the U.S., emphasizing the need for tailored emergency planning (Bergstrand et al., 2015). Expanding on this, Zou et al. (2018) examined Twitter engagement after Hurricane Sandy, revealing disparities in online participation and highlighting social media's role in damage assessment (Zou et al., 2018). Jacinto et al. (2020) operationalized social resilience in flood-affected communities, integrating text mining and expert surveys to identify key resilience dimensions, including individual, community, and governance factors (Jacinto et al., 2020). Rahmadan et al. (2020) applied LDA topic modeling to analyze Twitter discussions on Jakarta's flooding, identifying nine dominant topics that provided insights into public sentiment and information flow, supporting policymakers in disaster response strategies (Chilmi, 2021).

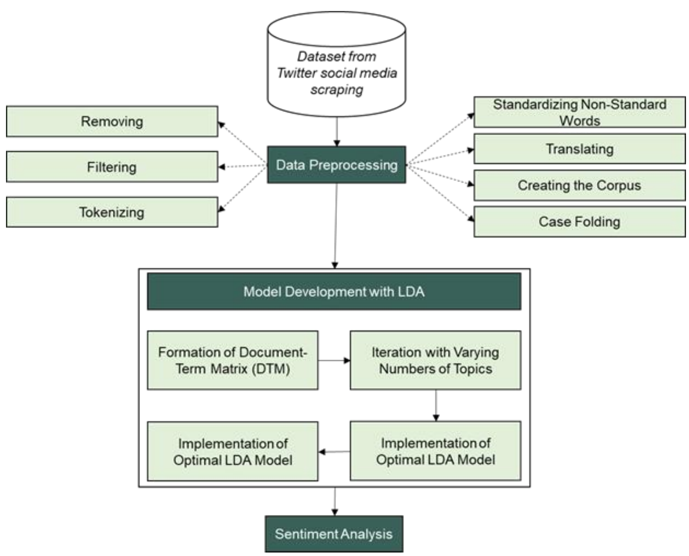
M. Luvian Chisni Chilmi (2021) analyzed public discussions on Twitter regarding the Omnibus Law Cipta Kerja in Indonesia, uncovering various issues of public concern, such as environmental impacts, socio-economic consequences, and waves of protests against the policy (Chilmi, 2021). This study successfully captured the spectrum of public opinions and sentiments by applying topic analysis, providing valuable insights for policymakers and academics in understanding social reactions to controversial regulations. Meanwhile, Rachma Awantina and Wahyu Wibowo (2022) study examined user-generated content on the Wattpad platform using Topic Modeling with the LDA method (Awantina & Wibowo, 2023). This research aimed to identify the main discussion topics among Wattpad users, revealing user preferences, experiences, and engagement patterns within the community. By analyzing recurring themes in user-generated content, the study offers valuable insights for content creators and platform developers to enhance user engagement strategies and improve the overall reading experience.

Material And Method

This study is structured into four key stages: data collection, preprocessing, model development, and sentiment analysis. Each stage is essential for systematically analyzing Indonesian public responses to disasters using Twitter data. The overall workflow of these stages is illustrated in Figure 1.

Data Collection

This study collects Twitter data using Python, systematically scraping 1,000–5,000 Indonesian-language tweets containing the keyword "bencana" within a specified timeframe [22]. The data focuses on real-time discussions and sentiments, forming the basis for Latent Dirichlet Allocation (LDA) topic modeling. This approach helps uncover prevalent themes and public concerns, providing insights to enhance disaster response strategies, public awareness campaigns, and policy development.



Methodological Framework for Analyzing Public Responses to Disasters Using Twitter Data.

Data Preprocessing

To enhance data quality, preprocessing begins with standardizing non-standard words using Kamus Besar Bahasa Indonesia (KBBI) (Rianto et al., 2020). Since LDA performs optimally in English, tweets are then translated using Google Translate (Hurtado Bodell et al., 2022). The text is converted into a structured corpus for further analysis (Ädel, 2020). Subsequent steps include case folding (lowercasing text) (Putra & Wijaya, 2022) (Kathuria et al., 2021) removing numbers, punctuation, and irrelevant elements (e.g., emojis, links, mentions, hashtags) (Abdul-Jabbar & George, 2017), and stopword filtering to eliminate common but insignificant words (Ladani & Desai, 2020). Finally, tokenization breaks the text into individual words, ensuring optimal conditions for topic modeling and sentiment analysis (Karthikeyan et al., 2020).

Model Development

The model development stage involves conducting topic modeling using LDA in Python. The process begins with constructing a Document-Term Matrix (DTM) from the preprocessed data, which captures word frequencies across documents as the foundation for analysis (Cozzolino & Ferraro, 2022) (Bafna & Saini, 2020). Next, the model undergoes iterative training with varying topic numbers, evaluating coherence scores in each iteration to determine the optimal topic count (Korencic et al., 2021). Once identified, the final LDA model is implemented using this optimal number of topics, ensuring meaningful topic extraction (Anowar et al., 2021; Kamaruddin et al., 2024; Öndin & Küçükdeniz, 2023). The results are then visualized with PyLDAvis, providing an interactive representation of topic distributions and relationships within the dataset (Onah & Pang, 2021) .

Sentiment Analysis on Extracted Topics

Following the topic modeling analysis with LDA, the subsequent step involves conducting sentiment analysis on the identified topics. Sentiment analysis aims to determine the prevailing sentiment or opinion associated with each topic(Alabdulkarim et al., 2024; Chakraborty et al., 2020; Wankhade et al., 2022). This process entails analyzing the sentiments expressed within the documents or discussions related to each topic, such as positive, negative, or neutral sentiments(Abualigah et al., 2020). By examining the sentiment distribution across topics, researchers can gain deeper insights into public perceptions, attitudes, and emotional responses toward various themes or issues discussed on social media platforms like Twitter.

Experimental Result

Model Development Result

Formation of Document-Term Matrix (DTM)

The Indonesian Twitter data collected using the keyword "bencana" undergoes LDA topic modeling to identify disaster-related discussions. The process starts with creating a Document-Term Matrix (DTM), converting unstructured text into a structured numerical format essential for LDA analysis. This matrix, generated using Python and shown in Table 1, captures term frequencies, allowing LDA to detect patterns. Once the text is structured, the analysis proceeds by determining the optimal number of topics for modeling.

Document-Term Matrix

No.	Term	Frequency
1	ability	1
2	afterlife	1
...	...	...
1.974.829	management	1
1.974.830	come	1

Defining the Range of Topic Numbers

Determining the range of candidate topic numbers for LDA topic modeling is crucial, as there is no definitive rule for selecting the optimal number of topics. LDA partitions text data into meaningful issues, making it essential to identify the range that best captures the dataset's structure. This process typically involves iterating through different topic ranges, starting broadly and gradually narrowing down. In this study, iterations were conducted with topic ranges of 1–50, 1–40, 1–30, 1–20, and 1–10 to identify the most coherent set of topics.

Iteration of Determining the Optimal Range of Topic Numbers

The iterations performed aim to obtain the truly optimal range of topics that remains efficient and has good interpretative capabilities. From the iterations, Table 2 will show the range of the optimal number of topics produced for each range, considering the coherence score.

Iteration of Optimal Topic Number Range

Range	Optimal Number of Topics	Coherence Score
1 to 50	20	0.487292
1 to 40	27	0.466890
1 to 30[a]	14	0.478454
1 to 20	16	0.473838
1 to 10	10	0.430744

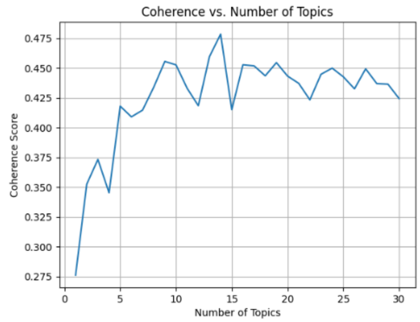
a. Range selected for balance between topic count and coherence.

The iteration results indicate that a moderate number of topics, precisely 14, yields a higher coherence score within the 1 to 30 topic range compared to the 1 to 40 and 1 to 20 ranges, which require more topics for coherence. Moreover, the coherence score for the 1 to 30 range is not significantly different from the 1 to 50 range, where 20 topics are needed to achieve a coherence score of 0.487292. The coherence scores for the 1 to 30 topic range are presented in Table 3.

Coherence Score for the Range 1 To 30

Number of Topics	Coherence Score
1	0,275969
2	0,35244
...	...
14	0,478454
15	0,41497
...	...
29	0,436394
30	0,424294

A line graph visualization of the coherence score for the entire range of topics is shown in Figure 2.



Coherence Score Visualization of Range 1 to 30.

Figure 2 illustrates that the highest coherence score in these observations occurs when the number of topics is 14, and for subsequent topics, the coherence score decreases. However, to ensure the optimal number of topics for the next analysis step, multiple observations will be conducted using the range of 1 to 30 topics.

Determining the Optimal Number of Topics

The determination of the optimal number of topics will be based on the coherence score results from multiple observations using the range of 1 to 30 topics. This iteration or repeated experiment is conducted to find a stable optimal number of topics. The iteration and repetition steps align with the theories presented in the foundational studies for this research, such as [10], [5], and [13]. Table 4 will show the results of this iteration.

Iteration of the Optimal Number of Topics

Observation	Optimal Number of Topics	Coherence Score
1	14	0.478454
2	13	0.462398
3	15	0.505317
4	14	0.493361
5	18	0.475374

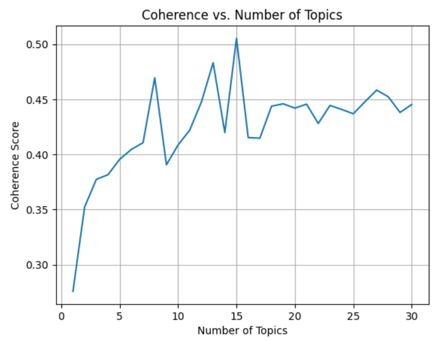
Based on the five iterations conducted, the third iteration produced the highest coherence score, reaching 0.5, with a total of 15 optimal topics selected. Therefore, this research will use the analysis results from this iteration. Table 5 will detail the coherence score results from this iteration.

Coherence Score for the Optimal Iteration

Number of Topics	Coherence Score
1	0,275969
2	0,35244
...	...
14	0,419938
16	0,415322
...	...

29	0,438104
30	0,445311

Figure 3 provides a visualization in the form of a line graph of the coherence score for the entire range of topics in this optimal iteration.

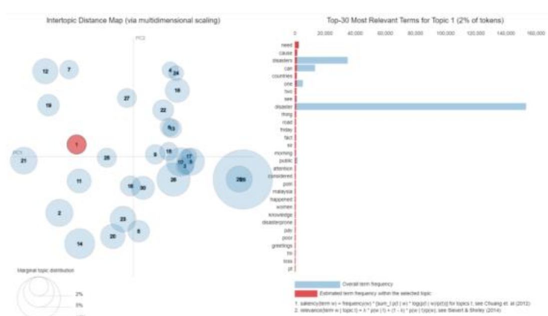


Coherence Score Visualization of Optimum Iteration.

Figure 3 shows the coherence score peaking at 15 topics before declining. However, further analysis determines 14 as the optimal number, balancing coherence and topic differentiation. Increasing topics beyond this does not enhance coherence significantly.

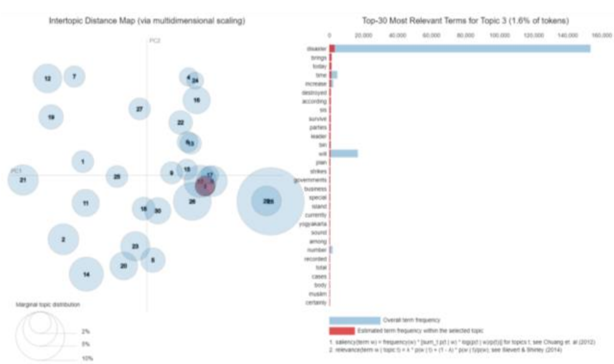
Topic Visualization

The optimal topic visualization, generated using the PyLDAvis library in Python, presents a bar chart illustrating the frequency of words within each topic. This visualization reveals diverse community perspectives on disasters, ranging from causes and impacts to humanitarian aid, preparedness, and emotional responses. A prominent theme emerging from the analysis is awareness of disaster causes and impacts.



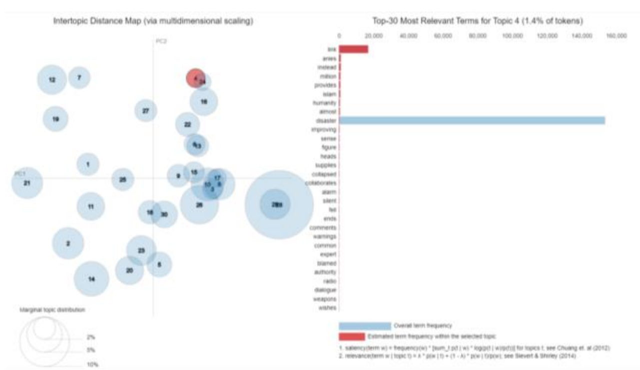
Topic 1 - Causes and Impacts of Natural Disasters.

As shown in Figure 4, discussions on this topic highlight key contributing factors, such as damaged roads ("road") and financial losses ("loss"). Additionally, terms like "attention," "considered," and "knowledge" suggest that public discourse emphasizes the importance of understanding disasters and mitigation strategies to minimize risks. In addition to discussing the causes of disasters, the community also highlights the aspects of time and survival strategies post-disaster. As seen in Figure 5, words like "time," "survive," "plan," and "business" reflect discussions about how people can plan recovery after facing a disaster.



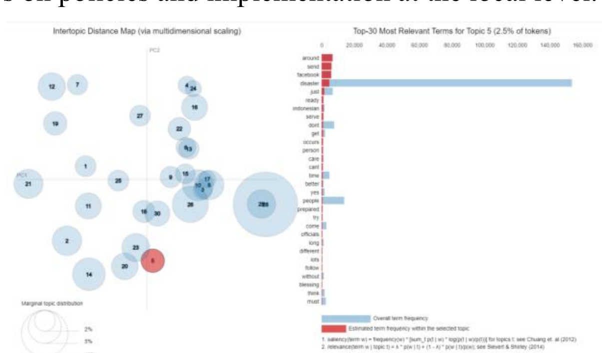
Topic 3 - Disaster Timing and Survival Strategie.

Social solidarity and humanitarian aid are crucial aspects often discussed in response to disasters. Figure 6 shows that the community actively discusses the distribution of aid, with words like "supplies," "collaborate," and "humanity" reflecting the coordination in delivering assistance to victims. Beyond direct aid, Figure 7 also demonstrates the role of social media in spreading information and raising awareness. Words like "Facebook," "send," "disaster," and "prepared" indicate that communities use digital platforms to share information and gather support during emergencies. The role of the government in disaster mitigation is also a key discussion point.

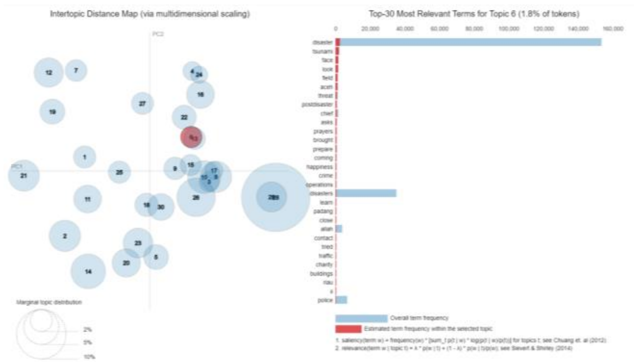


Topic 4 - Humanitarian Aid and Resource Distribution.

As seen in Figure 8, the community discusses how institutions like BPBD (Regional Disaster Management Agency) coordinate disaster management, including simulations at village and regional levels. Words like "management," "regional," "village," and "government" show that this discussion focuses on policies and implementation at the local level.

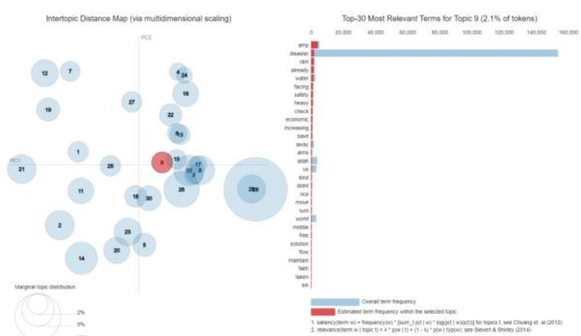


Topic 5 - Role of Social Media in Disaster Response.



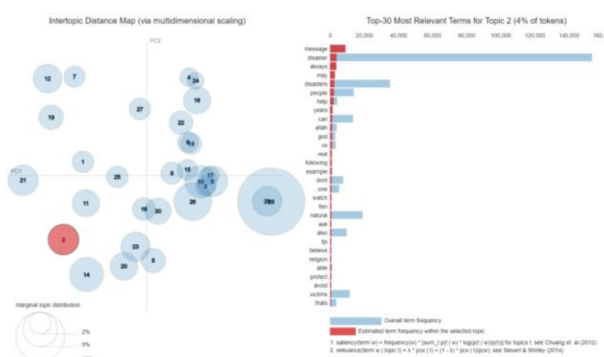
Topic 6 - Government and Regional Disaster Management.

Discussions on the environmental impact of disasters are also a significant concern, as seen in Figure 9. Words like "rain," "water," and "economic consequences" indicate that the community is discussing how extreme weather conditions affect clean water security and the regional economy.



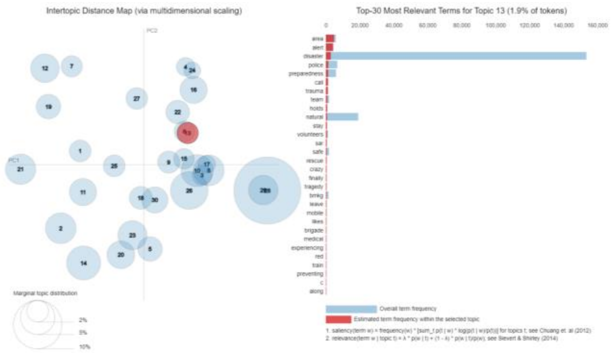
Topic 9 - Environmental and Economic Impact of Disasters.

Beyond technical aspects, the community also responds to disasters from emotional and religious perspectives. In Figure 10, it is evident that many discussions involve hope for divine assistance in facing disasters, reflected in words like "Allah," "God," and "religion."



Topic 2 - Religious Perspectives on Disasters.

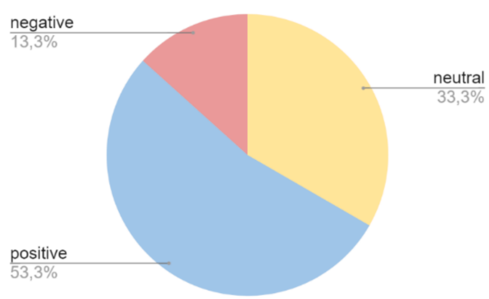
Moreover, psychological impacts are also a concern in community discussions. Figure 11 shows that words like "trauma," "safe," and "volunteers" indicate awareness of the importance of mental support for disaster victims.



Topic 13 - Psychological Impacts and Trauma from Disasters.

Sentiment Analysis Result

The sentiment analysis classifies fifteen disaster-related topics on Twitter into positive (53.3%), neutral, or negative, as shown in Figure 12. Positive discussions focus on preparedness, community efforts, and resilience, including hopes for divine intervention and reflections on the Aceh tsunami. Neutral topics cover social media use, disaster-prone areas, and victim trauma, emphasizing factual discussions. Negative sentiment appears in criticisms of local disaster management and environmental impacts, highlighting inefficiencies in aid distribution and ecological concerns.



Sentiment Distribution of Disaster-Related Topics on Twitter

Overall, the sentiment analysis highlights a mix of emotional responses from the public. While there is significant optimism and collective action in disaster recovery efforts, critical voices addressing shortcomings in disaster management are also present. This reflects the complexity of community engagement in disaster discourse, balancing hope, awareness, and accountability.

Conclusions

The Latent Dirichlet Allocation analysis identified 15 distinct topics in Indonesian Twitter discussions on natural disasters, covering a broad spectrum of issues. These topics ranged from causes and impacts of disasters, hopes for divine intervention, and survival strategies to resource distribution, disaster awareness via social media, and dissatisfaction with local disaster management. Specific events, such as the Aceh tsunami, collaborative response efforts, community preparedness, environmental concerns, trauma among victims, and drought affecting children, were also prominent. Sentiment analysis further categorized public responses into neutral, pessimistic, and optimistic sentiments. Over half of the topics reflected positive sentiments, particularly in discussions on collaborative efforts and effective response strategies. Neutral sentiments emerged concerning resource distribution and social media awareness, while negative sentiments were dominant in criticisms of local disaster management and environmental impacts. These findings offer valuable insights into Indonesian public sentiment and discourse on disasters. They provide a data-driven foundation for improving disaster response strategies, public awareness campaigns, and policymaking to better address the needs of affected communities.

Acknowledgment

The authors acknowledge and gratitude the financial support for this work provided by Institut Teknologi Sepuluh Nopember through International Partnership Research Grant based on Rector's Decree 1243/PKS/ITS/2024.

References

Abdul-Jabbar, S. S., & George, L. E. (2017). Fast Text Analysis Using Symbol Enumeration and Hashing Methodology. Iraqi Journal of Science, 345–354. https://www.ijs.uobaghdad.edu.iq/index.php/eijs/article/view/6169

Abualigah, L., Alfar, H. E., Shehab, M., & Hussein, A. M. A. (2020). Sentiment Analysis in Healthcare: A Brief Review. Studies in Computational Intelligence, 874, 129–141. https://doi.org/10.1007/978-3-030-34614-0_7

Ädel, A. (2020). Corpus Compilation. A Practical Handbook of Corpus Linguistics, 3–24. https://doi.org/10.1007/978-3-030-46216-1_1

Airlangga, G. (2024). ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR SEISMIC ANOMALY DETECTION IN INDONESIA: UNVEILING PATTERNS IN THE PACIFIC RING OF FIRE. Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika, 5(1), 37–48. https://doi.org/10.46306/LB.V5I1.489

Alabdulkarim, N. A., Haq, M. A., & Gyani, J. (2024). Exploring Sentiment Analysis on Social Media Texts. Engineering, Technology & Applied Science Research, 14(3), 14442–14450. https://doi.org/10.48084/ETASR.7238

Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Computer Science Review, 40, 100378. https://doi.org/10.1016/J.COSREV.2021.100378

Awantina, R., & Wibowo, W. (2023). Computational Linguistics Using Latent Dirichlet Allocation for Topic Modeling on Wattpad Review. https://doi.org/10.4108/EAI.1-11-2022.2326169

Bafna, P. B., & Saini, J. R. (2020). Marathi Document: Similarity Measurement using Semantics-based Dimension Reduction Technique. International Journal of Advanced Computer Science and Applications, 11(4), 138–143. https://doi.org/10.14569/IJACSA.2020.0110419

Bergstrand, K., Mayer, B., Brumback, B., & Zhang, Y. (2015). Assessing the Relationship Between Social Vulnerability and Community Resilience to Hazards. Social Indicators Research, 122(2), 391–409. https://doi.org/10.1007/S11205-014-0698-3/METRICS

Chakraborty, K., Bhattacharyya, S., & Bag, R. (2020). A Survey of Sentiment Analysis from Social Media Data. IEEE Transactions on Computational Social Systems, 7(2), 450–464. https://doi.org/10.1109/TCSS.2019.2956957

Chilmi, M. L. C. (2021). Latent dirichlet allocation lda untuk mengetahui topik pembicaraan warganet twitter tentang omnibus law. https://repository.uinjkt.ac.id/dspace/handle/123456789/56724

Choirul Rahmadan, M., Nizar Hidayanto, A., Swadani Ekasari, D., Purwandari, B., & Theresiawati. (2020). Sentiment Analysis and Topic Modelling Using the LDA Method related to the Flood Disaster in Jakarta on Twitter. Proceedings - 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020, 126–130. https://doi.org/10.1109/ICIMCIS51567.2020.9354320

Cladis, A. E. (2020). A shifting paradigm: An evaluation of the pervasive effects of digital technologies on language expression, creativity, critical thinking, political discourse, and interactive processes of human communications. E-Learning and Digital Media, 17(5), 341–364. https://doi.org/10.1177/2042753017752583/SUPPL_FILE/SUPPLEMENTAL_DOCUMENTATION_SHIFTING_PARADIGM_1.PDF

Cozzolino, I., & Ferraro, M. B. (2022). Document clustering. Wiley Interdisciplinary Reviews: Computational Statistics, 14(6), e1588. https://doi.org/10.1002/WICS.1588

Erb, M., Mucek, A. E., & Robinson, K. (2021). Exploring a social geology approach in eastern Indonesia: What are mining territories? The Extractive Industries and Society, 8(1), 89–103. https://doi.org/10.1016/J.EXIS.2020.09.005

Ferner, C., Havas, C., Birnbacher, E., Wegenkittl, S., & Resch, B. (2020). Automated Seeded Latent Dirichlet Allocation for Social Media Based Event Detection and Mapping. Information 2020, Vol. 11, Page 376, 11(8), 376. https://doi.org/10.3390/INFO11080376

Fuady, M., Munadi, R., & Fuady, M. A. K. (2021). Disaster mitigation in Indonesia: between plans and reality. IOP Conference Series: Materials Science and Engineering, 1087(1), 012011. https://doi.org/10.1088/1757-899X/1087/1/012011

Hurtado Bodell, M., Magnusson, M., & Mützel, S. (2022). From Documents to Data: A Framework for Total Corpus Quality. Socius, 8. https://doi.org/10.1177/23780231221135523/ASSET/IMAGES/LARGE/10.1177_23780231221135523-FIG6.JPEG

Jacinto, R., Reis, E., & Ferrão, J. (2020). Indicators for the assessment of social resilience in flood-affected communities – A text mining-based methodology. Science of The Total Environment, 744, 140973. https://doi.org/10.1016/J.SCITOTENV.2020.140973

Kamaruddin, S. S., Abdul-Rahman, S., & Wibowo, W. (2024). Understanding Malaysian Public Opinion on Suicide through Sentiment Analysis and Topic Modeling of Reddit Posts. Engineering, Technology & Applied Science Research, 14(6), 18055–18062. https://doi.org/10.48084/ETASR.8738

Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. Journal of Management Analytics, 7(2), 139–172. https://doi.org/10.1080/23270012.2020.1756939

Karthikeyan, S., Jotheeswaran, J., Balamurugan, B., & Chatterjee, J. M. (2020). Text Mining. Natural Language Processing in Artificial Intelligence, 167–210. https://doi.org/10.1201/9780367808495-7

Kathuria, A., Gupta, A., & Singla, R. K. (2021). A Review of Tools and Techniques for Preprocessing of Textual Data. Advances in Intelligent Systems and Computing, 1227, 407–422. https://doi.org/10.1007/978-981-15-6876-3_31

Korencic, D., Ristov, S., Repar, J., & Snajder, J. (2021). A Topic Coverage Approach to Evaluation of Topic Models. IEEE Access, 9, 123280–123312. https://doi.org/10.1109/ACCESS.2021.3109425

Kusumasari, B., & Prabowo, N. P. A. (2020). Scraping social media data for disaster communication: how the pattern of Twitter users affects disasters in Asia and the Pacific. Natural Hazards, 103(3), 3415–3435. https://doi.org/10.1007/S11069-020-04136-Z/METRICS

Ladani, D. J., & Desai, N. P. (2020). Stopword Identification and Removal Techniques on TC and IR applications: A Survey. 2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020, 466–472. https://doi.org/10.1109/ICACCS48705.2020.9074166

Mendon, S., Dutta, P., Behl, A., & Lessmann, S. (2021). A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters. Information Systems Frontiers, 23(5), 1145–1168. https://doi.org/10.1007/S10796-021-10107-X/METRICS

Onah, D., & Pang, E. (2021). MOOC DESIGN PRINCIPLES: TOPIC MODELLING-PYLDAVIS VISUALIZATION & SUMMARIZATION OF LEARNERS' ENGAGEMENT. EDULEARN21 Proceedings, 1, 1082–1091. https://doi.org/10.21125/EDULEARN.2021.0282

Öndin, S. Ö., & Küçükdeniz, T. (2023). latent Dirichlet allocation method-based nowcasting approach for prediction of silver price. Accounting, 9(3), 131–152. https://doi.org/10.5267/J.AC.2023.3.004

Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2023). Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research. IEEE Transactions on Affective Computing, 14(1), 108–132. https://doi.org/10.1109/TAFFC.2020.3038167

Rahmafitria, F., Sukmayadi, V., Suryadi, K., & Rosyidie, A. (2021). Disaster management in Indonesian tourist destinations: how institutional roles and community resilience are mediated. Worldwide Hospitality and Tourism Themes, 13(3), 324–339. https://doi.org/10.1108/WHATT-01-2021-0014/FULL/XML

Rianto, Benny Mutiara, A., Prasetyo Wibowo, E., & Insap Santosa, P. (2020). The crowdsourcing method to normalize "bahasa alay", a case of indonesian corpus. 2020 5th International Conference on Informatics and Computing, ICIC 2020. https://doi.org/10.1109/ICIC50835.2020.9288534

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review 2022 55:7, 55(7), 5731–5780. https://doi.org/10.1007/S10462-022-10144-1

Zhou, S., Kan, P., Huang, Q., & Silbernagel, J. (2021). A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura. Https://Doi.Org/10.1177/01655515211007724, 49(2), 465–479. https://doi.org/10.1177/01655515211007724

Zou, L., Lam, N. S. N., Cai, H., & Qiang, Y. (2018). Mining Twitter Data for Improved Understanding of Disaster Resilience. Annals of the American Association of Geographers, 108(5), 1422–1441. https://doi.org/10.1080/24694452.2017.1421897.