Jurnal Teknologi Informasi dan Multimedia
p-ISSN : 2715-2529
e-ISSN : 2684-9151
https://journal.sekawan-org.id/index.php/jtim

Utilizing Natural Language Processing for the Analysis of BMKG Decadal Atmospheric Dynamics Reports in 2025

Ardiansyah Ramadhan [1,2,*], Aaz Muhmmad Hafidz Azis[3]

[1] Computer Engineering, Telkom University, Indonesia
[2] Center of Excellence for Intelligent Sensing-IoT, Telkom University, Indonesia
[3] School of Computing, Telkom University, Indonesia

[*] Correspondence: ardiansyahramadhanar@telkomuniversity.ac.id

Citation: A. Ramadhan and A. M. H. Azis, "Utilizing Natural Language Processing for the Analysis of BMKG Decadal Atmospheric Dynamics Reports in 2025", Jurnal Teknologi Informasi Dan Multimedia, vol. 8, no. 1, pp. 185-196, 2026. https://doi.org/10.35746/jtim.v8i1.946

Received: 12-01-2026
Revised: 01-02-2026
Accepted: 09-02-2026

Abstract: The Meteorology, Climatology, and Geophysics Agency (BMKG) publish decennial reports that provide valuable insights into Indonesia's meteorological conditions and their temporal fluctuations. However, due to their narrative structure, conducting direct quantitative analysis is problematic. This study seeks to address this issue by using a transparent, repeatable natural language processing (NLP) method to identify temporal trends in climatic conditions favourable to the formation of acid rain. The collection contains 36 BMKG decadal atmospheric dynamics studies for 2025. The proposed approach entails gathering textual input, performing basic preprocessing (case normalization, character sanitization, space-based tokenization, and stop-word removal), and subsequently employing predefined keyword dictionaries for analysis. These dictionaries delineate weather conditions that either facilitate or inhibit the formation of acid rain. The scores for acid rain conditions are determined by the frequency of specific keywords, adjusted for the document's length. Subsequently, they are categorized into groups utilizing statistical thresholds derived from the mean and standard deviation of the adjusted scores. Non-parametric statistical tests are employed to examine temporal patterns with greater specificity. The findings indicate that normalized acid rain scores are elevated in the initial years of the decade, specifically 2025, before gradually declining until year-end. The Spearman rank correlation test reveals a statistically significant negative correlation between normalized scores and time ($\varrho = -0.494$, $p = 0.0022$). The Mann–Kendall test indicates a significant downward trend ($Z = -2.902$). These results demonstrate that the climatic conditions responsible for acid rain occurred only temporarily, rather than year-round. The core element of this work is a straightforward, lexicon-based NLP approach that is easily understood, replicable, and applicable, and can transform narrative atmospheric reports into structured quantitative metrics. This is beneficial for research on atmospheric dynamics and environmental analysis with official written data.

Keywords: Atmospheric dynamics, text analysis, natural language processing.

1. Introduction

Atmospheric dynamics are crucial for controlling changes in weather and climate, especially regarding precipitation processes and wet deposition mechanisms that lead to acid rain [1]-[3]. Human activities release sulphur dioxide and nitrogen oxides, which are the main causes of acid rain. Nevertheless, processes such as cloud formation, heavy rainfall, unstable air, and large-scale air mass movement have a big effect on how often and how bad these catastrophes are [4]. To understand how acid rain affects the ecosystem, needed to understand how the atmosphere works [5].

In Indonesia, systematic information on atmospheric conditions is regularly disseminated by the Meteorology, Climatology, and Geophysics Agency (BMKG) through decadal atmospheric dynamics reports [6]-[7]. These official bulletins summarize important weather events, including changes in temperature, humidity, wind patterns, atmospheric waves, and rainfall predictions. People often use them to monitor the weather, prepare for disasters, and make judgments across many areas [7]-[8]. A main goal is to ensure that most of the information in these reports is presented in a story-like, detailed way [7]. Because those are qualitative, those can't be used directly in quantitative analyses, time-series evaluations, or comparative environmental studies that need systematic indicators.

Previous research face considerable challenges in transforming narrative weather reports into structured data. There is a lot of information about the atmosphere, but we can't use it all because the methods we use to measure it aren't good enough. This is a very important part of learning about acid rain. People here often talk about how bad the rain is, how unpredictable the weather is, and how air masses generally move. These things can help us learn more about how wet deposition works, especially in places where there isn't much long-term chemical monitoring data.

Recent advances in Natural Language Processing (NLP) have enabled us to address this problem effectively [9]. They help extract organized data from unstructured text [9]. Researchers have utilized NLP techniques to discern trends, essential concepts, and patterns across time in narrative texts for policy analysis, environmental communication, and climate discourse research [10]-[11]. Researchers in atmospheric and environmental science have explored the potential of NLP for analysing climate reports, academic papers, and policy documents [9], [12]. NLP is rarely utilized in official meteorological reports, especially those released by national agencies in developing countries. This study still don't fully grasp how it could be used to explain how ecosystems work.

Researchers frequently utilize numerical modelling [13], remote sensing data [13], [14], and in situ observations [15]-[18] to examine acid rain and meteorological events. Most people think that official reports based on stories are only qualitative data. This creates a methodological gap since narrative reports include expert opinions and thorough evaluations of the topic that raw statistics may not fully capture. Systematically acquiring and measuring this information could improve traditional data-driven methods and environmental assessments, especially in situations where data are hard to obtain.

This study fills the gap by offering a simple, repeatable NLP-based approach to analyzing BMKG's 2025 records of atmospheric dynamics. The proposed method employs a lexicon-based approach to quantify meteorological concepts related to precipitation, atmospheric instability, humidity, and air mass dynamics. This makes it easier to read, check, and be honest. This study does not directly investigate the chemistry of acid rain. Instead, it focuses on understanding the weather conditions and long-term patterns that make wet deposition processes linked to acid rain more likely.

The main goal of this study is to show how a simple NLP framework links qualitative weather reports to quantitative environmental data. It changes how we view different environmental factors. Organizing official narrative reports makes it easier to use textual climate data in atmospheric and environmental studies. This shows that its goal differs from that of traditional research, which uses numbers or other measurements.

2. Materials and Methods

2.1. Data Source and Study Materials

This study utilizes textual data derived from official decadal atmospheric dynamics reports published by the Meteorology, Climatology, and Geophysics Agency (BMKG). A total of 36 reports covering the period from January [19]-[21] to December [22]-[24] 2025 were collected, corresponding to three decadal reports per month. Each report was treated as an independent text document representing atmospheric conditions for a specific ten-day period.

Table 1. Data sources of the BMKG Decadal Atmospheric Dynamics Reports for the Year 2025

No	Month	Decadal	Period (±)	Sources
1	January	1	1-10	[19]
2	January	2	11-20	[20]
3	January	3	21-30	[21]
…..	…..	…..	…..	…..
…..	…..	…..	…..	…..
34	December	1	1-10	[22]
35	December	2	11-20	[23]
36	December	3	21-30	[24]

The reports were obtained directly from the official BMKG website to ensure data authenticity and reliability. These documents contain narrative descriptions of large-scale and regional atmospheric phenomena, including wind circulation patterns, atmospheric wave activity, humidity conditions, rainfall analysis, and seasonal outlooks. Since the reports are publicly available and do not involve personal or sensitive data, no ethical approval was required for this study.

2.2. Text Extraction and Preprocessing

Text preprocessing was conducted to transform raw narrative text into a standardized and analyzable format [25]. The preprocessing pipeline included case folding to convert all characters to lowercase, removal of non-alphabetic characters (such as numbers, punctuation, and special symbols) using regular expressions, and whitespace-based tokenization to segment the cleaned text into individual word tokens. This simple tokenization strategy was selected to maintain methodological transparency and reproducibility. Subsequently, common Indonesian stopwords were removed using a predefined stopword list to reduce linguistic noise and retain semantically meaningful terms. The output of this process was a set of cleaned word tokens for each document, which served as the input for subsequent analysis [25].

2.3. Lexicon Construction and Scoring Strategy

A lexicon-based approach was employed to quantify atmospheric conditions described in the narrative reports [26]. The lexicon was constructed based on meteorological terminology commonly used in BMKG reports and on established concepts in atmospheric science and the acid rain literature. The lexicon was constructed based on meteorological terminology frequently used in BMKG reports, as well as established concepts in atmospheric science and acid rain literature. The keyword lexicon was organized into three main categories to represent different atmospheric tendencies related to acid rain conditions. Keywords supporting acid rain conditions include meteorological terms associated with enhanced rainfall, atmospheric instability, and wet deposition processes, including *hujan, lebat, konveksi, awan, MJO, ITCZ, konvergensi, angin, baratan, and deposisi*. Neutral keywords represent general or transitional atmospheric conditions that do not directly indicate the enhancement or suppression of rainfall, such as *normal, transisi, lokal, bervariasi, musiman, fluktuatif, and sebagian wilayah*. Keywords less supportive of acid rain conditions include terms describing atmospheric stability or dry conditions that inhibit precipitation and wet deposition, such as *kering, subsiden, stabil, tekanan, minim, and udara.*

For each document, a raw score (RS) was calculated to quantify the balance between supportive and less supportive atmospheric conditions [27]. As shown in Equation (1), the raw score was defined as the difference between the frequency of supportive keywords $(N_{support})$ and the frequency of less supportive keywords $(N_{nonsupport})$:

$$RS = N_{support} - N_{nonsupport}. \quad (1)$$

To account for variations in document length, the raw score was subsequently normalized by the total number of tokens in the document (N_{tokens}) [25]. The resulting normalized score (NS), as defined in Equation (2), represents the relative density of acid rain–related keywords within each document:

$$NS = \frac{RS}{N_{tokens}}. \quad (2)$$

This normalization ensured that the resulting score reflected relative keyword density rather than absolute frequency, thereby enabling consistent comparison across documents with different text lengths [26]-[28].

2.4. *Categorization of Atmospheric Conditions*

Normalized scores [29]-[30] were classified into three categories to represent the overall tendency of atmospheric conditions in each decadal period: supporting acid rain, neutral, and less supportive of acid rain. The classification was performed using statistically derived thresholds based on the mean and standard deviation of the normalized score distribution, allowing each category to reflect relative variations in atmospheric conditions across the study period.

First, the mean normalized score [29] (μ) was calculated across all documents as the central tendency of the dataset, as expressed in Equation (3):

$$\mu = \frac{1}{N}\sum_{i=1}^{N} N\,S_i \quad (3)$$

where NS_i represents the normalized score of the i-th document and N denotes the total number of documents. Subsequently, the standard deviation (σ), which quantifies the dispersion of normalized scores around the mean, was computed according to Equation (4):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(N\,S_i - \mu)^2} \quad (4)$$

Based on these descriptive statistics, categorization thresholds were defined using a data-driven rule-based approach [13]-[14], [31]. As summarized in Equation (5), documents with normalized scores exceeding one standard deviation above the mean $(NS > \mu + \sigma)$ were classified as supporting acid rain, whereas documents with scores lower than one standard deviation below the mean $(NS < \mu - \sigma)$ were classified as less supportive of acid rain. Scores falling within this interval were categorized as neutral:

$$\text{Category} = \begin{cases} \text{Supporting acid rain,} & NS > \mu + \sigma, \\ \text{Neutral,} & \mu - \sigma \leq NS \leq \mu + \sigma, \\ \text{Less supportive of acid rain,} & NS < \mu - \sigma. \end{cases} \quad (5)$$

This adaptive thresholding strategy avoids arbitrary cutoff values and ensures classification reflects the dataset's intrinsic distributional characteristics, enabling consistent and comparable categorization across all decadal periods.

2.5. Statistical Analysis

To evaluate the robustness and temporal characteristics of the extracted normalized scores, several non-parametric statistical tests were applied [32]. Non-parametric methods were selected because the normalized scores did not follow a normal distribution, as is commonly observed in text-based analytical outputs. All statistical analyses were conducted at a significance level of 0.05, and the tests were designed to assess distributional properties, monotonic relationships, and temporal trends in atmospheric condition indicators derived from narrative reports [32].

The Shapiro–Wilk test was first employed to assess the normality of the normalized score distribution and to justify the use of non-parametric statistical methods [33]. The Shapiro–Wilk test statistic W, as defined in Equation (6), compares the observed distribution of normalized scores with an expected normal distribution:

$$W = \frac{\left(\sum_{i=1}^{n} a_i \, x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad (6)$$

In this equation, $x_{(i)}$ represents the ordered sample values, $\bar{x}$ is the sample mean, and a_i are constants derived from the expected values of order statistics of a normal distribution. A statistically significant result ($p < 0.05$) indicates a deviation from normality, thereby supporting the application of non-parametric statistical approaches in subsequent analyses.

To investigate temporal behavior, the Spearman rank correlation test [34] and the Mann–Kendall trend test [35] were applied. Spearman correlation was used to evaluate the monotonic relationship between normalized scores and the chronological order of decadal periods. The Spearman correlation coefficient ρ, as expressed in Equation (7), is calculated based on the rank differences between paired observations [34]:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (7)$$

where d_i denotes the difference between paired ranks and n is the total number of observations.

In addition, the Mann–Kendall test was used to detect a statistically significant monotonic trend in the normalized scores over the study period. The Mann–Kendall test statistic, defined in Equation (8), is calculated as [35]:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sign}(x_j - x_i). \quad (8)$$

The combined use of correlation and trend analysis enables a comprehensive evaluation of temporal patterns in atmospheric conditions described in the narrative reports, allowing the direction and significance of long-term trends to be identified.

3. Results

3.1. Descriptive Data Explanation

As shown in Table 2, the descriptive statistics of text length indicate that the BMKG decadal reports are relatively consistent in size throughout 2025. The mean text length is approximately 4,877 words, with a moderate standard deviation of 605 words, suggesting limited variability across the documents. The range between the minimum and maximum values further confirms that all reports maintain comparable narrative depth, supporting the reliability of subsequent text-based analyses.

The distribution of token counts across the 36 BMKG decadal reports indicates a relatively consistent document length throughout the study period. After text preprocessing, the number of tokens per document ranges from approximately 450 to 730, with most documents clustering around mid-range values. This moderate variability suggests that the reports maintain comparable narrative depth and structural consistency across different months and decadal periods.

Table 2. Descriptive Statistics of Text Length in BMKG Decadal Reports (2025)

Statistic	Value
Count	36 documents
Mean	4,877 characters per documents
Standard Deviation	605 characters per documents
Minimum	4,342 characters per documents
25th Percentile (Q1)	4,472 characters per documents
Median (50th Percentile)	4,644 characters per documents
75th Percentile (Q3)	5,173 characters per documents
Maximum	7,215 characters per documents

The relatively uniform document length is important for ensuring the robustness of subsequent text-based analysis, particularly the normalization of keyword-based scores. Since no extreme disparities in token counts were observed, the normalized scores are unlikely to be biased by document length, thereby allowing fair comparison of atmospheric condition indicators across all decadal reports.

3.2. Keyword-Based Scoring

Table 3 summarizes the keyword-based scoring results for all 36 BMKG decadal atmospheric reports in 2025. Reports classified as supporting acid rain are characterized by relatively higher raw scores and normalized scores, indicating a greater frequency and density of meteorological keywords associated with rainfall enhancement and wet deposition processes. In contrast, reports categorized as neutral exhibit consistently lower raw and normalized scores, reflecting a more balanced representation of atmospheric conditions without a strong dominance of rainfall-related descriptors.

Table 3. Keyword-Based Scoring Results for BMKG Decadal Atmospheric Reports (2025)

Document	Number of Tokens	Raw Score	Normalized Score	Category
1	734	22	0.029973	Supporting Acid Rain
2	603	16	0.026534	Supporting Acid Rain
3	547	11	0.020110	Neutral
4	533	11	0.020638	Neutral
5	543	11	0.020258	Neutral
6	540	11	0.020370	Neutral
7	538	9	0.016729	Neutral

Document	Number of Tokens	Raw Score	Normalized Score	Category
8	456	7	0.015351	Neutral
9	484	8	0.016529	Neutral
10	602	17	0.028239	Supporting Acid Rain
11	583	16	0.027444	Supporting Acid Rain
12	568	12	0.021127	Neutral
13	455	6	0.013187	Neutral
14	458	6	0.013100	Neutral
15	461	6	0.013015	Neutral
16	467	6	0.012848	Neutral
17	471	6	0.012739	Neutral
18	459	6	0.013072	Neutral
19	451	6	0.013304	Neutral
20	456	6	0.013158	Neutral
21	473	6	0.012685	Neutral
22	491	6	0.012220	Neutral
23	482	6	0.012448	Neutral
24	473	6	0.012685	Neutral
25	499	6	0.012024	Neutral
26	458	7	0.015284	Neutral
27	475	7	0.014737	Neutral
28	480	7	0.014583	Neutral
29	499	7	0.014028	Neutral
30	490	7	0.014286	Neutral
31	512	7	0.013672	Neutral
32	469	7	0.014925	Neutral
33	484	7	0.014463	Neutral
34	490	7	0.014286	Neutral
35	495	7	0.014141	Neutral
36	481	7	0.014553	Neutral

The distribution of categories also reveals a temporal tendency, with supporting acid rain classifications occurring predominantly during the early part of the year. As the year progresses, both raw and normalized scores decrease and stabilize within a narrower range, resulting in a predominance of neutral classifications. This pattern suggests that atmospheric conditions conducive to acid rain were episodic and seasonally influenced rather than persistent throughout the study period, a trend clearly captured by the combined use of raw and normalized scoring metrics.

3.3. Normalized Acid Rain Score Trend

As shown in Figure 1, the normalized acid rain scores derived from BMKG decadal atmospheric reports in 2025 exhibit a distinct temporal pattern. Higher normalized scores are observed during the early decadal periods, indicating a greater relative presence of meteorological descriptors associated with enhanced rainfall, atmospheric instability, and wet deposition processes. These elevated values are consistent with Indonesia's climatological wet season, during which rainfall-related atmospheric dynamics are more prominently described in the narrative reports.

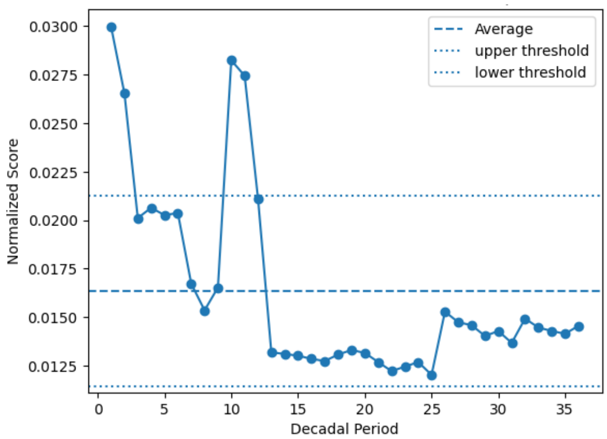

Figure 1. Integration of Normalized Scores and Statistical Thresholds in the Analysis of Acid Rain Conditions

As the decadal progresses, the normalized scores decline gradually and subsequently stabilize at lower values during the mid- to late-decadal periods. This pattern reflects a reduced emphasis on rainfall-supportive atmospheric conditions and a predominance of neutral atmospheric descriptions. Overall, the results indicate that atmospheric conditions conducive to acid rain were temporally concentrated rather than persistent throughout the year, demonstrating the effectiveness of the lexicon-based NLP approach in capturing seasonal variability within official atmospheric narratives.

3.4. Statistical Analysis

As summarized in Table 4, the Shapiro–Wilk test yields a very small p-value, indicating that the distribution of normalized acid rain scores deviates from normality and supporting the use of non-parametric statistical methods [33]. The Spearman rank correlation analysis reveals a statistically significant, moderate negative correlation between the normalized scores and the chronological order of decadal periods, suggesting a decreasing tendency in acid-rain–supportive atmospheric conditions over time [34].

Table 4. Results of Non-Parametric Statistical Tests on Normalized Acid Rain Scores

Statistical Test	Statistic / Parameter	Value
Shapiro–Wilk Test	p-value	2.51×10^{-6}
Spearman Rank Correlation	Correlation coefficient (ϱ)	−0.494
	p-value	0.0022
Mann–Kendall Trend Test	Trend	Decreasing
	Z-statistic	−2.902
	Kendall's Tau	−0.340
	Sen's slope	-1.78×10^{-4}

Furthermore, the Mann–Kendall trend test confirms a statistically significant decreasing trend in the normalized scores [35]. The negative Z-statistic and Kendall's Tau values indicate a consistent downward pattern, while the negative Sen's slope reflects a gradual rate of decrease across decadal periods. Overall, the results in Table 4 demonstrate that atmospheric conditions conducive to acid rain were more pronounced earlier in the year and diminished progressively throughout 2025.

4. Discussion

This research demonstrates that a lexicon-based NLP approach transforms descriptive atmospheric information into quantifiable indicators. This study keep track of how conditions related to acid rain change over time. The approach examines 36 BMKG atmospheric reports from 2025 and identifies patterns. For instance, acid rain is worse at the beginning of the year and becomes better as the year goes on. This is shown by lower scores and downward trends that follow Indonesia's weather cycle.

Based on these findings, it's crucial to note that they align with earlier studies on the atmosphere and environment, which show that seasonal rainfall patterns and large-scale air circulation are important to wet deposition processes in tropical regions. Prior research indicates that increased precipitation and fluctuating atmospheric conditions often increase the likelihood of pollutant removal and moist deposition, including acid rain events. The temporal pattern observed in this study—higher indicators during the wet season and lower indicators during the dry months—supports and extends what we already know. It also shows that narrative-based official reports can give us the same kinds of information as numerical or chemical measurements.

The predominance of neutral categories throughout most decadal intervals suggests that the weather conditions favorable to acid rain are ephemeral rather than perennial phenomena. This observation aligns with climatological research on Indonesia, which shows that considerable precipitation and atmospheric instability occur during specific seasonal periods. This study provides a robust, consistent framework for analyzing narrative reports of varying lengths and types using lexicon-based scoring and nonparametric statistical methods. This makes qualitative weather data more reliable and easier to understand.

From a practical point of view, these results have a significant impact on data managers and policy-makers. Meteorological organizations and environmental data managers can use the NLP-based framework to systematically convert many narrative reports into structured indicators. This method makes it easier to continuously monitor and compare them across different time periods, reducing the need for human qualitative analysis. These signs might help identify times when more in-depth analysis or focused sampling is needed, such as when chemical monitoring of precipitation isn't possible in a given area or at a given time.

The results imply to politicians and environmental regulators that narrative atmospheric reports can serve as early warning signs of conditions that could increase the likelihood of acid rain. Adding these indicators to decision-support systems can help people be ready for the seasons, communicate with the public, and decide what to do first to reduce the risks of wet deposition. The method works well in areas with limited data because it relies solely on official reports and requires no additional monitoring equipment.

Some problems need to be fixed, even though they are helpful. The study only looks at data from one year (2025), which makes it hard to look at how acid rain changes from year to year and over the long term. The lexicon-based approach relies on well-known keywords from meteorological terminology, which may not fully capture the subtleties of context or changes in language in official reports. The study investigates climatic conditions that promote acid rain instead of direct chemical evaluations, such as rainfall pH or pollutant concentrations.

Using this method across datasets over several years will make it more useful and reliable. Scientists who study the atmosphere and environment will add to the keyword lexicon to accurately describe how the atmosphere works. Combining rule-based natural language processing with adaptive, data-driven models will make it easier to find changes in the environment and in reporting. Improved methods will make the suggested framework necessary for monitoring operations, making policy decisions, and tracking real-time changes in acid rain.

5. Conclusions

This work shows that a lexicon-based Natural Language Processing (NLP) method can convert narrative atmospheric data into numerical values for evaluating the environment. The method showed strong seasonal patterns in acid rain conditions by examining 36 decadal atmospheric reports from BMKG in 2025. At the start of the year, conditions were right for acid rain, and this trend lasted until the end of the year, which was statistically significant.

Most decades were neutral, meaning that acid rain only fell at certain times of year, not year-round. Normalized scoring and nonparametric testing made it straightforward to compare accounts, regardless of story length. The methods were also clear and could be used repeatedly. These results show that narrative weather reports provide measurable data on how the weather changes with the seasons.

This study functions as a preliminary methodological assessment, rather than a replacement for traditional acid rain research. It offers a straightforward method to derive atmospheric context from qualitative narratives, particularly in the absence of substantial data or direct observations. The system needs to be better and tested more thoroughly, as a single-year dataset with a list of keywords demonstrates.

Future research should expand this methodology to include multi-year datasets, refine the vocabulary through expert input, and explore hybrid or adaptive NLP techniques to effectively capture contextual and linguistic diversity in official reports. If it goes in these directions, this framework might become a very important aspect of systems that help people make decisions and monitor the environment. It significantly connects qualitative atmospheric narratives with quantitative environmental assessment.

Acknowledgments: The authors would like to express their sincere gratitude to Telkom University for providing institutional support for this research. The authors also acknowledge the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) for making the decadal atmospheric dynamics reports publicly available, which enabled the data collection and analysis conducted in this study.

References

[1] A. Cardil *et al.*, “Climate teleconnections modulate global burned area,” *Nat. Commun.*, vol. 14, no. 1, Dec. 2023, https://doi.org/10.1038/s41467-023-36052-8.

[2] Z. Su, Y. Wang, Y. Zheng, Y. Wang, P. Li, and X. Zhang, “Acid deposition and meteorological factors together drive changes in vegetation cover in acid rain areas,” *Ecol. Indic.*, vol. 167, Oct. 2024, https://doi.org/10.1016/j.ecolind.2024.112720.

[3] X. Yang *et al.*, “Spring Rapid Temperature Variability in Southern China: Characteristics, Decadal Trend and Associated Climate Impacts on Crop Yield,” *International Journal of Climatology*, vol. 45, no. 9, p. e8880, Jul. 2025, https://doi.org/10.1002/joc.8880.

[4] P. Grennfelt, A. Engleryd, M. Forsius, Ø. Hov, H. Rodhe, and E. Cowling, “Acid rain and air pollution: 50 years of progress in environmental science and policy,” Apr. 01, 2020, *Springer*. https://doi.org/10.1007/s13280-019-01244-4.

[5] G. Stamm, A. Bhattacharjee, G. Basapuram, A. Dutta, and S. Duttagupta, “Molecular and Ionic Signatures in Rainwater: Unveiling Sources of Atmospheric Pollution,” *Environments - MDPI*, vol. 12, no. 10, Oct. 2025, https://doi.org/10.3390/environments12100351.

[6] Benny Hartanto, Ningrum Astriawati, Supartini, and Damar Kuncoro Yekti, “Pencarian dan Pemanfaatan Informasi Data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG),” *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 1, no. 5, pp. 553–564, Oct. 2022, https://doi.org/10.55123/insologi.v1i5.906.

[7] J. A. Siswanto, G. C. A. Sugiarto, A. Salma, S. G. J. Soesilo, and S. Hartono, “Optimizing Success: Analyzing Key Factors and Improving the Info BMKG Application,” in *2024 International Conference on Information Management and Technology (ICIMTech)*, 2024, pp. 512–517. https://doi.org/10.1109/ICIMTech63123.2024.10780786.

[8] D. Primanda and M. N. Burga, "Efektivitas Aplikasi Info BMKG Dalam Memberikan Informasi Cuaca Dan Bencana Terhadap Masyarakat Kota Tangerang Selatan," *Neo Politea*, vol. 3, no. 2, 2022. https://doi.org/10.53675/neopolitea.v3i2.1013

[9] S. C. Necula, F. Dumitriu, and V. Greavu-Şerban, "A Systematic Literature Review on Using Natural Language Processing in Software Requirements Engineering," Jun. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. https://doi.org/10.3390/electronics13112055.

[10] M. Cheriet, J.-F. Boucher, L. Gondim, A. Guimarães, and J.-M. F. Guest, "Accelerating the Socio-Ecological Transition Strategies and Innovations for Sustainable Development."

[11] E. A. d. Santos, D. G. B. de Souza, and C. E. S. da Silva, "What Matters in Hiring Professionals for Global Software Development? A SLR and NLP Criteria Clustering," *IEEE Trans. Eng. Manag.*, vol. 71, pp. 6291–6318, 2024, https://doi.org/10.1109/TEM.2023.3279769.

[12] J. O. Ojadi, E. C. Onukwulu, C. S. Odionu, and O. A. Owulade, "Natural Language Processing for Climate Change Policy Analysis and Public Sentiment Prediction: A Data-Driven Approach to Sustainable Decision-Making," 2023.

[13] A. Ramadhan, I. Chandra, A. Indrawati, D. A. Tanti, I. Mendonça, and M. Aritsugi, "Decision support system using the expert-fuzzy method in validating real-time acid rain measurements," *J. Reliab. Intell. Environ.*, vol. 11, no. 4, Dec. 2025, https://doi.org/10.1007/s40860-025-00258-y.

[14] A. Ramadhan, A. Indrawati, D. A. Tanti, W. Setyawati, and I. Chandra, "AI-Driven Imputation and Correlation Analysis of Environmental Parameters for Understanding pH Variability in Acid Rain," in *2025 IEEE International Conference on Networking, Intelligent Systems, and IoT (ICONS-IoT)*, 2025, pp. 75–81. https://doi.org/10.1109/ICONS-IoT65216.2025.11211132.

[15] D. Rievaldo *et al.*, "Effect of Rainfall on The Chemical Composition of Rainwater in Monitoring Acid Deposition In Greater Bandung," *INDONESIAN JOURNAL OF URBAN AND ENVIRONMENTAL TECHNOLOGY*, pp. 49–62, Feb. 2023, https://doi.org/10.25105/urbanenvirotech.v6i1.14051.

[16] A. Indrawati *et al.*, "Characteristics of Acid Deposition in Urban and Sub-Urban Area," in *Springer Proceedings in Physics*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 171–182. https://doi.org/10.1007/978-981-19-0308-3_13.

[17] R. Puji Lestari *et al.*, "Chemical Composition of Wet Deposition in Jakarta, Serpong, Bandung, Kototabang, and Maros During 2015-2019," *ECOLAB*, vol. 15, no. 2, pp. 89–100, 2021, https://doi.org/10.20886/jklh.2021.15.2.89-100.

[18] A. Indrawati *et al.*, "Spatiotemporal distribution in chemical composition of wet atmospheric deposition in Bandung Indonesia," *Environmental Science and Pollution Research*, Nov. 2024, https://doi.org/10.1007/s11356-024-35485-y.

[19] Badan Meteorologi Klimatologi dan Geofisika (BMKG), "Analisis Dinamika Atmosfer Dasarian I Januari 2025." https://www.bmkg.go.id/iklim/analisis-dinamika-atmosfer-dasarian-i-januari-2025

[20] Badan Meteorologi Klimatologi dan Geofisika (BMKG), "Analisis Dinamika Atmosfer Dasarian II Januari 2025." https://www.bmkg.go.id/iklim/analisis-dinamika-atmosfer-dasarian-ii-januari-2025

[21] Badan Meteorologi Klimatologi dan Geofisika (BMKG), "Analisis Dinamika Atmosfer Dasarian III Januari 2025." https://www.bmkg.go.id/iklim/analisis-dinamika-atmosfer-dasarian-iii-januari-2025

[22] Badan Meteorologi Klimatologi dan Geofisika (BMKG), "Analisis Dinamika Atmosfer Dasarian I Desember 2025." https://www.bmkg.go.id/iklim/dinamika-atmosfer/analisis-dinamika-atmosfer-dasarian-i-desember-2025

[23] Badan Meteorologi Klimatologi dan Geofisika (BMKG), "Analisis Dinamika Atmosfer Dasarian II Desember 2025." https://www.bmkg.go.id/iklim/dinamika-atmosfer/analisis-dinamika-atmosfer-dasarian-ii-desember-2025

[24] Badan Meteorologi Klimatologi dan Geofisika (BMKG), "Analisis Dinamika Atmosfer Dasarian III Desember 2025." https://www.bmkg.go.id/iklim/dinamika-atmosfer/analisis-dinamika-atmosfer-dasarian-iii-desember-2025

[25] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences (Switzerland)*, vol. 12, no. 17, Sep. 2022, https://doi.org/10.3390/app12178765.

[26] N. M. Sham and A. Mohamed, "Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches," *Sustainability (Switzerland)*, vol. 14, no. 8, Apr. 2022, https://doi.org/10.3390/su14084723.

[27] A. Prakash, S. Kumar Gupta, and M. Rawat, "Keyword Based Ranking of Web Pages by Normalizing Link Score," vol. 20, no. 6, pp. 4725–4729, 2022, https://doi.org/10.14704/nq.2022.20.6.

[28] L. Ren *et al.*, "Environmental conditions are the dominant factor influencing stability of terrestrial ecosystems on the Tibetan plateau," *Commun. Earth Environ.*, vol. 4, no. 1, Dec. 2023, https://doi.org/10.1038/s43247-023-00849-8.

[29] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *32nd International Conference on Machine Learning, ICML 2015*, International Machine Learning Society (IMLS), 2015, pp. 448–456. https://proceedings.mlr.press/v37/ioffe15.html
[30] I. Rodríguez, A. Ortiz, P. Caldevilla, S. Giganto, G. Búrdalo, and M. Fernández-Raga, “Comparison between the Effects of Normal Rain and Acid Rain on Calcareous Stones under Laboratory Simulation,” *Hydrology*, vol. 10, no. 4, Apr. 2023, https://doi.org/10.3390/hydrology10040079.
[31] A. Ramadhan *et al.*, “Central Tendency Data Real-Time Acid Rain Measurement to Evaluate Tool’s Performance Using Statistical Analysis,” vol. 14, no. 4, 2024. https://doi.org/10.18517/ijaseit.14.4.19273
[32] D. Li, H. Xie, and L. Xiong, “Temporal Change Analysis Based on Data Characteristics and Nonparametric Test,” *Water Resources Management*, vol. 28, no. 1, pp. 227–240, Jan. 2014, https://doi.org/10.1007/s11269-013-0481-2.
[33] E. González-Estrada and W. Cosmes, “Shapiro–Wilk test for skew normal distributions based on data transformations,” *J. Stat. Comput. Simul.*, vol. 89, no. 17, pp. 3258–3272, Nov. 2019, https://doi.org/10.1080/00949655.2019.1658763.
[34] J. H. Zar, “Spearman Rank Correlation: Overview,” in *Wiley StatsRef: Statistics Reference Online*, 2014. https://doi.org/10.1002/9781118445112.stat05964.
[35] K. H. Hamed and A. Ramachandra Rao, “A modified Mann-Kendall trend test for autocorrelated data,” *J. Hydrol. (Amst).*, vol. 204, no. 1, pp. 182–196, 1998, https://doi.org/10.1016/S0022-1694(97)00125-X.