**IJCCS (Indonesian Journal of Computing and Cybernetics Systems)**
Vol.17, No.2, April 2023, pp. 127~138
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.81397


# The Tweetology of New and Renewable Energy in Indonesia

**Ariana Yunita*[1], Sara Florensia Telaumbanua[2], Ade Irawan[3]**
[1,2,3]Department of Computer Science, Universitas Pertamina, Jakarta, Indonesia
e-mail: ***[1]ariana.yunita@universitaspertamina.ac.id**,
[2]105218048@student.universitaspertamina.ac.id, [3]adeirawan@universitaspertamina.ac.id

***Abstrak***

*Jumlah data tidak terstruktur meningkat setiap tahun, yang menjanjikan untuk mendapatkan wawasan. Twitter, platform yang menghasilkan data tidak terstruktur, saat ini menjadi salah satu platform media paling populer yang digunakan untuk melakukan penelitian tentang tren suatu topik. Studi ini mencoba menganalisis topik Energi Baru dan Terbarukan (EBT) di Indonesia. Tujuan dari penelitian ini adalah untuk mendapatkan wawasan tentang tren topik NRE selama sepuluh tahun terakhir dengan memodelkan topik yang dibahas di Twitter dan memeriksa distribusi lokasi pengguna yang memposting tweet tentang topik tersebut. Oleh karena itu, penelitian ini menggunakan analisis deskriptif, analisis geocoding, dan pemodelan topik. Hasil analisis deskriptif menunjukkan bahwa perkembangan EBT semakin pesat dalam beberapa tahun terakhir, khususnya pada tahun 2021. Analisis geocoding mengungkapkan bahwa sebaran penduduk yang melakukan kegiatan posting EBT didominasi oleh Provinsi DKI Jakarta. Pemodelan topik menghasilkan dua topik yang paling banyak dibicarakan oleh masyarakat Indonesia dalam kurun waktu 10 tahun. Kedua topik tersebut terkait dengan kebijakan pemerintah yang mendukung pengembangan EBT dan ketenagalistrikan yang menjadi fokus Indonesia dalam EBT. Studi ini menyoroti pentingnya menganalisis Tweetology NRE.*

***Kata kunci***— *Text mining, Energi Baru dan Terbarukan, Twitter, Geocoding analysis, Topic modeling*

***Abstract***

*The amount of unstructured data is increasing annually, which is promising for gaining insights. Twitter, a platform producing unstructured data, is currently one of the most popular media platforms used for conducting research on a topic's trend. This study attempts to analyze the topic of New and Renewable Energy (NRE) in Indonesia. The purpose of this study is to gain insights into the NRE topic trend over the last ten years by modeling the topics discussed on Twitter and examining the location distribution of users who post tweets about the topic. Accordingly, this study employed descriptive analysis, geocoding analysis, and topic modeling. The results of descriptive analysis show that the development of NRE has accelerated in recent years, particularly in 2021. Geocoding analysis reveals that the distribution of people who engage in NRE posting activities is dominated by DKI Jakarta province. Topic modeling yielding two topics that were discussed the most by Indonesians over a 10-year period. The two topics are related to government policies that support the development of NRE and electricity, which is Indonesia's focus in NRE. This study highlights the importance of analyzing the Tweetology of NRE.*

***Keywords—****Text mining, New and renewable energy, Twitter, Geocoding analysis, Topic modeling*

*Received January 17th,2023; Revised April 30th, 2023; Accepted April 30th, 2023*

## 1. INTRODUCTION

As the volume of data has exploded in recent years, big data is increasingly being elevated as a strategic tool for the future of science. The concept of "Everything is Data" leads to include any data, whether structured or unstructured [1]. The growth of unstructured data from year to year is getting faster, as much as 60-80% [2]. With the growing volume each year, unstructured data is currently used for various purposes, e.g., for regulatory, analytical, and decision-making purposes [3], [4]. It potentially powers analytics, machine learning, and business intelligence in the near future [5].

According to [6], at the start of 2022, Indonesia's internet penetration rate was 73.7 percent of the total population. In line with that, the use of social media also increased from year to year. In January 2022, Indonesia had 191.4 million social media users. One of the most often used media channels for trend analysis is Twitter [7]. According to data published in Twitter's advertising resources, Twitter had 18.45 million users in Indonesia in early 2022. It indicates that Twitter's ad reach in Indonesia at the time was equivalent to 6.6 percent of the total population.

As Twitter evolves into a microblogging platform, more data about prevalent themes, interests, dialogues, and social interactions are automatically recorded every second. Due to the open nature of tweets and the ease with which they may be accessed without being subject to privacy limitations, information extraction on Twitter is particularly popular among text mining researchers. Furthermore, collecting data from digital traces, such as Twitter, is considered to be a more efficient method of gathering relevant data because it reduces collection time, error risks, and incomplete data [8].

Affordable and clean energy is becoming the seventh Sustainable Development Goals, including accessing New and Renewable Energy (NRE) [9]. In Indonesia, NRE is a national issue. Since 1998, Indonesia has made clean energy a priority in its energy policy, and Indonesia has recognized the value of renewable energy. Despite Indonesia's abundance of renewable energy resources, the use of renewable energy is still not in accordance with the national target. Indonesia set a target of 23% NRE usage by 2025 [10], but it is currently at 12,8% [11].

This study reflects the tweetology [12] of NRE in Indonesia. Our main purposes are to gain the trend of NRE on Twitter, to determine the location of NRE-related Twitter posts, and then to reveal the popular topics amongst Twitter users regarding NRE within ten years. Therefore, three research questions guide this study:

1. What has been the trend in the number of Twitter data related to NRE over the last ten years?
2. Where is the location distribution of users who post tweets about NRE?
3. What topics have developed over the past ten years regarding NRE?

The rationale why Twitter is used in this study, the review of text mining and the review of previous works are described below.

### *1.1 Twitter for Conducting Academic Research*

With the growth of Twitter as a social media platform, it has become a popular forum for the general public to read the latest news, exchange ideas or opinions, and connect with people worldwide in real time. Twitter's features have helped to establish it as a popular social media platform for conducting academic research.

In general, there are five reasons why Twitter is suitable for conducting academic research. First, Twitter is one of the most popular social networks, with 319 million monthly active users from nearly every country [13]. Twitter users come from various backgrounds and professions, including governments, businesses, organizations, non-governmental organizations,

celebrities, athletes, journalists, academics, and the general public. Because approximately 42 percent of Twitter users use it every day and 24 percent use it at least once a week, Twitter provides a subset of almost all categories that researchers would be interested in [13].
Second, with 500 million daily tweets or messages, Twitter users generate a lot of data. Twitter is a social network with a large population, as evidenced by each message and the users who use it [13].

Third, obtaining data via Twitter is relatively simple because Twitter provides user data via an Application Programming Interface (API) that anyone with a Twitter account can access [13].

Fourth, with the API, it is easier to adjust the data received according to the type of research. Researchers can receive as much as one percent of all tweets per day via the streaming API or by selecting received tweets by keyword, location, username, or language used. Then with the REST API, researchers can download certain tweets, users' latest tweets, followers and followings, and information about user profiles [13].

Fifth, Twitter is a social network and also a very good data source when used to analyze networks and non-networks. Network and non-network are forms of communication or activity for Twitter users. However, the difference is that network activity is communication that is carried out online on Twitter. In contrast, non-network is user activity that is carried out offline or real life but remains relevant or related with the problem to be analyzed. This is because Twitter provides services in an accurate and structured manner as a liaison between users through its strong hashtags culture, making it easier to collect, sort, and expand searches during data collection; therefore, the data collected will be more accurate [14].

### *1.2 Text Mining*

According to Allahyari et al. [15], text mining was firstly introduced by Feldman et al. in 1995 with the term "knowledge discovery from text". The term refers to the process of extracting high-quality information from text. Approaches to text mining are attributed to traditional data mining and knowledge discovery methods. Mostly, text mining tasks are for classification, clustering, and extraction [15].

The idea for extracting topics in a massive body text, such as Twitter data, is to comprehend the underlying topics within the Tweets. Latent Dirichlet Allocation (LDA) is commonly used for extracting topics or topic modeling. It is interesting to note here that LDA algorithm is used for modeling topics in text, video, and image [16]. Regarding languages, LDA has been implemented in various languages, such as in Chinese [17] and Bahasa [18][19]. In addition, a more advanced technique for topic modeling in a multi-language environment has been proposed by Zosa and Granroth-Wilding [20].

### *1.3 Related Works*

In recent years, social media has been widely used in text mining research, owing to the growing number of Twitter users. Prior studies have shown that social media, especially Twitter data, is used extensively to understand any issue. For example, research conducted by Nurdeni et al. (2022), regarding the types of support for victims of natural disasters. They create a classification model by utilizing data from Twitter. In analyzing the data, the study used the Stanford NER method, which is a text-mining technique, to extract geographic information. The algorithm used is Nave Bayes, Support Vector Machine, Logistic Regression with OneVsRest, Binary Relevance, Power-set Label, and Classifier Chain. The research recommends providing an overview of the most effective methodology to extract data from citizens' Twitter accounts to determine the best course of action for the victims and visualize their position in real-time [3].

Differently, Louis et al. analyze Tweet data related Harvey hurricane using Latent Dirichlet Allocation (LDA). The analysis of the postings will provide a better understanding of the critical needs of people during a disaster, and we will be able to formulate data-driven policy recommendations for disaster management [4].

Another study also employs LDA to find the topic related to changes in consumer behavior during Covid-19 pandemic. They use social media data instead of the conventional method of obtaining data from interviews and questionnaire-based surveys. The study results in 6 topics related to food consumption [21]. In contrast with [3] and [4], Brzustewicz and Singh conducted the semantic analysis using Louvain Algorithm.

Other studies [22], [23] also used Twitter as their data source. Santis et al. [23] employ graphs for their analysis technique. Not only detecting topic, Santis et al. [9] were also tracking the topic. In contrast, Pons et al. [22] applied descriptive analysis, which is word of cloud in their study, and also conducted sentiment analysis with Naïve Bayes classifier.

This study will be different from other previous studies mentioned in the above paragraphs. This study employs descriptive analysis, geocoding analysis, and topic modelling analysis using LDA. This study also focuses on the NRE topic in Indonesia over the past ten years.

The structure of the paper is as follows. Section two will cover the relevant works. The third section is a description of the methods. In part four, the findings and discussions will be explained. The final section will provide the conclusion and recommendations for further study.

## 2. METHODS

In general, there were five research stages that underlying this study. The steps in this study were collecting data from Twitter, text preprocessing, descriptive analysis, geocoding analysis, and topic modelling analysis. The research stages were depicted in Figure 1. Below are the details of each research step.

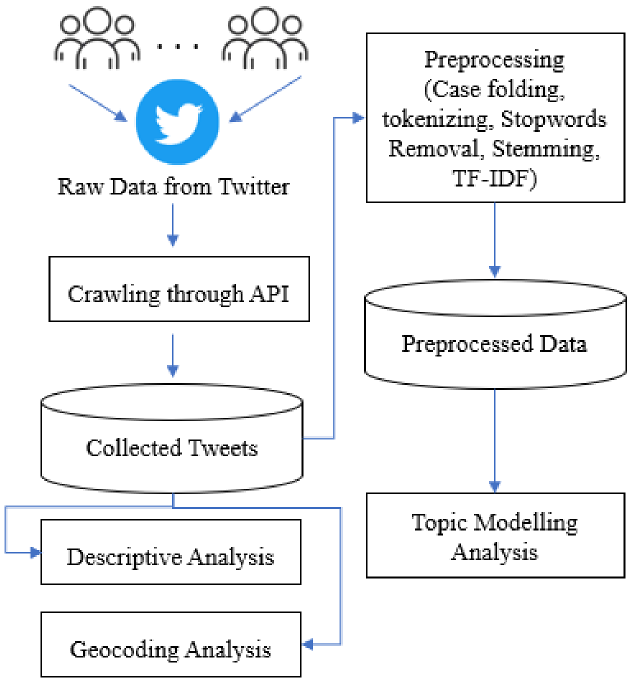


Figure 1 Research Stages

*2.1 Data Collection*

Data collection used an API provided by Twitter to retrieve collections of Tweet data by facilitating data crawling based on specific keywords. The first phase of the data collection process was registering a developer account on the Twitter developer platform. By registering with Twitter Developers, bearer tokens can be obtained that help in the authentication process. The data crawling process used the Python programming language to program using Tweepy as a Python-provided library to access the Twitter API. There were three keywords for crawling the Twitter data, as seen in Table 1.

Table 1 Keywords Used for Data Crawling

| Original Keywords | Translated |
|---|---|
| Energi Baru Terbarukan | New and Renewable Energy |
| Renewable Energy | - |
| EBT | Stands for “Energi Baru Terbarukan”, translated as “New and Renewable Energy” |

*2. 2 Text Preprocessing*

After the raw dataset was collected, the data was cleaned to prepare the data for the analysis phase. The preprocessing process transformed a collection of previously unstructured raw data into more structured and high-quality data.

*2.2.1 Case Folding*

This step converted all capital letters to lowercase. The goal is that the exact words are not detected to have different meanings because there are differences between capital and lowercase letters.

*2.2.2 Tokenizing*

The stages of tokenizing conducted in this study included removing the tab, new line, back slice, emoticon, mention, link, hashtag, and incomplete URL. Other processes were to remove numbers, punctuation, whitespace leading and trailing, multiple whitespaces, and single char.

The hashtag character in this study was omitted because it did not have spaces between words, whereas if the tweet keyword in this research was implemented in a hashtag, it would have a different meaning. Words in hashtags can certainly be separated, but it requires complexity in programming. To perform word separation on hashtags by the Python programming language, a library called WordSegment is needed. However, the WordSegment library can only process words in English, and hence its implementation will be challenging to do, so it becomes a limitation in this study.

*2.2.3 Stop words Removal*

In the text data set, most of the common words have no meaning and occur with sufficient frequency to affect the accuracy of the analysis results. In the stopword removal (filtering) stage, this research filtered the data frame using Indonesian stopwords obtained from the NLTK library, then checked whether there was a stopword list in the token or not. If there was a stopword, then the stopwords were removed from the data frame.

*2.2.4 Stemming*

In this study, the term was stemmed to get it back to its original form. This study employed the Sastrawi library's stemmer function during the stemming phase. However, this research added a library, termed Swifter, to help with the stemming process on the data frame because the Sastrawi library's function has shortcomings.

*2.2.5 TF-IDF*

The TF-IDF stage, the last step in this research's text preprocessing, was performed as a weighting that will be applied to document grouping. Below is the step-by-step of conducting Tweets preprocessing.

Table 2 Tweets Preprocessing

| Text Preprocessing | Results |
|---|---|
| Tweet as an input/ before preprocessing | Masih banyak daerah terpencil membutuhkan listrik…..<br>Ealah ini si bapak tiba2 ngomong EBT….<br>Lomba poster ilmiah "Energi Baru Terbarukan"…, |
| Case Folding | masih banyak daerah terpencil membutuhkan listrik…..<br>ealah ini si bapak tiba2 ngomong ebt….<br>lomba poster ilmiah "energi baru terbarukan"…. |
| Tokenization | masih, banyak, daerah, terpencil, membutuhkan, listrik,…..<br>ealah, ini, si, bapak, tiba, ngomong, ebt,….<br>lomba, poster, ilmiah, energi, baru, terbarukan,…. |
| Stop words removal | masih, banyak, daerah, terpencil, membutuhkan, listrik,…..<br>ealah, bapak, tiba, ngomong, ebt,….<br>lomba, poster, ilmiah, energi, baru, terbarukan,…. |
| Stemming | masih, banyak, daerah, pencil, butuh, listrik,…..<br>ealah, bapak, tiba, ngomong, ebt,….<br>lomba, poster, ilmiah, energi, baru, baru,…. |

*2.3 Data Analysis*

*2.3.1 Descriptive Analysis*

The descriptive analysis of this study describes the coverage of the number of tweets and likes on the tweet dataset that has been collected from each keyword from 2012 to 2022. In addition, Tweet posting platform sources, e.g., Android, Facebook, or others, were also analyzed.

*2.3.2 Geocoding Analysis*

Different location data can be converted into a small number of patterns to show position-mediated patterns and distance using geocoding techniques [12]. This technique aims to generate textual cues of universal geographic locations, such as longitude and latitude, for integration with the map. To conclude, Geocoding (spatial) analysis is a technique for analyzing trends, identifying patterns, and describing various geographical phenomena.

*2.3.3 Topic Modelling Analysis*

Topic modeling is a machine learning-based content analysis method that automatically finds the latent structure in a text collection by the number of big ones [24]. Topic modeling is an important process in text mining using techniques such as extracting and summarizing trending issues.

This study employed LDA [25] for topic modeling. The main inputs for developing the LDA topic model were a dictionary and a corpus. The coherence score is a metric used to assess topic modeling within LDA, which means that a good model produces topics with high topic coherence scores.

## 3. RESULTS AND DISCUSSION

### *3.1 Descriptive Analysis*

The development of tweets related to NRE over the past decade was felt to have increased rapidly and was widely discussed on Twitter, as shown in Figure 2. NRE has been widely discussed on Twitter since 2012 and increased from 2014 to 2016. In 2017, after looking at the raw data of tweets from Twitter users 2017, it was found that the number of tweets decreased and increased again in 2021. Overall, there was an increase in the popularity of tweeting NRE.

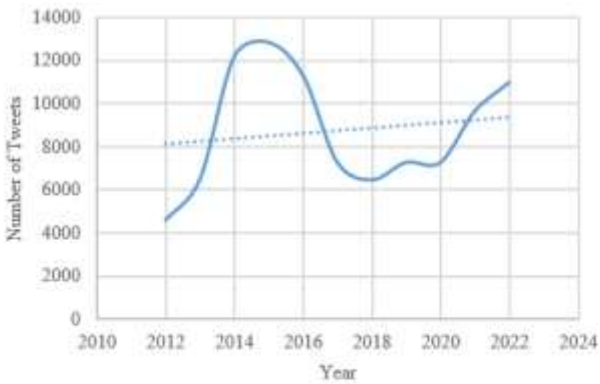


Figure 2 Total Number of Tweets Related to NRE, Each Year

The top 10 tweet sources used by Twitter users to post tweets for new renewable energy keywords using their PCs or smartphone devices were determined from the 40,173 tweets gathered. With a total of 8,545 tweets, the results indicated that Twitter for Android was the most popular media source for tweet-posting activities related to the keyword "new renewable energy". The second most popular media source for posting was twitterfeed. Another source, as seen in Figure 3, is Twitter for Blackberry. BlackBerry was one of the world's most well-known smartphone brands. At its peak in September 2011, there were 85 million BlackBerry regular customers worldwide. However, due to the growth of the Android and iOS platforms, BlackBerry's market dominance had shrunk. BlackBerry reported that its global subscriber count fell to 72 million in the first quarter of fiscal 2014, a decrease of 4 million subscribers from the previous quarter [26]. It might indicate that the Tweets from Twitter for Blackberry might be posted prior to 2012 and up to 2015.

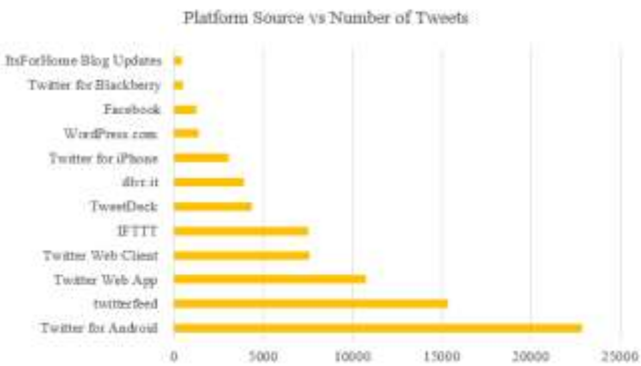


Figure 3 Number of Tweets from Each Platform Source

In the Descriptive Analysis, this study also analyzed the number of likes from the overall data on each keyword.

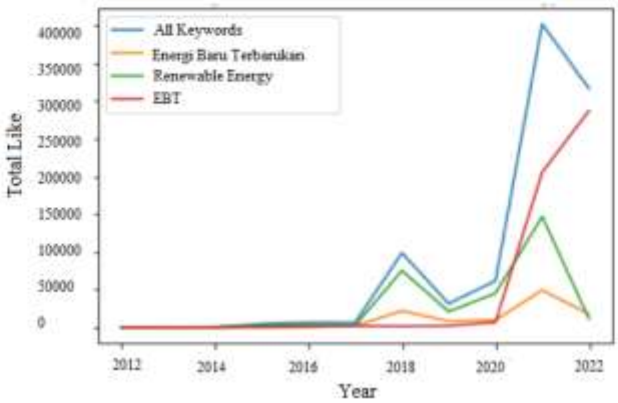


Figure 4 Total Like Each Year for Each Keyword

As explained in the previous section, there were three keywords used in this study: "Energi Baru Terbarukan", "Renewable Energy", and "EBT. As seen in Figure 4, there was a significant increase of total likes from year to year, especially in 2021. The total likes from 2012 until 2017 seems stagnant, then rose in 2018. Among the three keywords, the Twitter data keyword with the most total likes was "EBT". It seems that the keyword "EBT" might be the most popular among other keywords.

*3.2 Geocoding Analysis*

Based on the findings of the tweet data crawling, which produced a total of 109,705 tweets with locations and 52,590 of them in Indonesia, the results are shown in Figure 5. In order to perform geocoding analysis, this study made use of the Nominatim library, which employs Open Street Map data to identify the name, address, latitude, and longitude, as well as the nation, province, and city features. The findings indicated that, with a total of 20,823 tweets, the province of Jakarta had the most tweets about NRE in Indonesia.

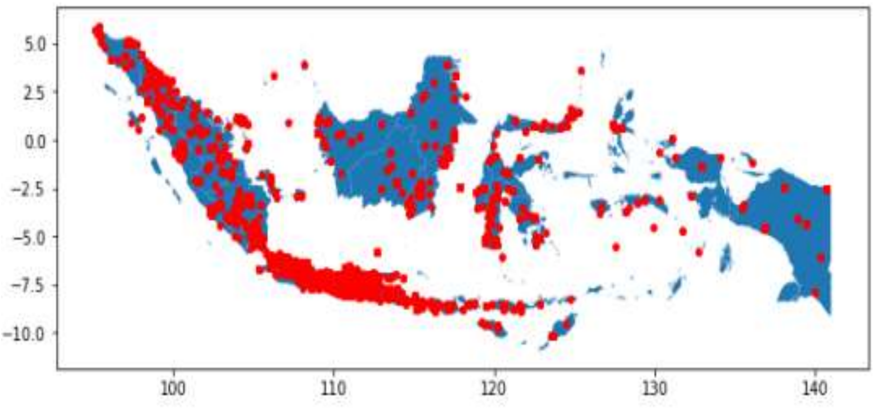


Figure 5 The Distribution of The Total Posting from Each Province Related to NRE over Ten Years

The province with the fewest posts was identified as South Papua, which only had a total of 4 tweets, making it the province with the fewest tweets about NRE. Using the principle "Everything is Data" [1], the data related to Indonesia digital divide were also collected. Jakarta has a low digital divide level [27], which might be why the province of Jakarta has the most tweets about NRE in Indonesia. In line with that, Papua has a high digital divide [27] that might also be similar to the number of tweets related to NRE.

*3.3 Topic Modelling Analysis Using LDA*

In determining the topic modeling using the LDA algorithm technique, this study used the Gensim package of Python programming language, which contains the gensim.model and gensim.corpora modules. The gensim.model module was applied to form the LDA model and perform the process of calculating the coherence model. On the other hand, the gensim.corpora module was implemented to create a dictionary through data text.

The LDA algorithm produced different topics each time the process was run. This happened because LDA uses randomness during initialization. Therefore, we needed to define the random state parameters when initializing the LDA model. In this study, random state = 1 was used as the solution to stop outcomes for different subjects in the LDA model. It is important to note here that the hyperparameter had not been optimized. In other words, fine-tuning hyperparameters was not conducted in this study.

The number of topic models was determined based on the graphical representation of the coherence score computation. The higher the coherence score, the easier it is to interpret the topic modeling results, so it is a benchmark for evaluating topic modeling by looking at the highest topic coherence scores. Based on the coherence score calculation results, the topic with the highest score was topic 2, with a coherence score of 0.539, as shown in Figure 6.

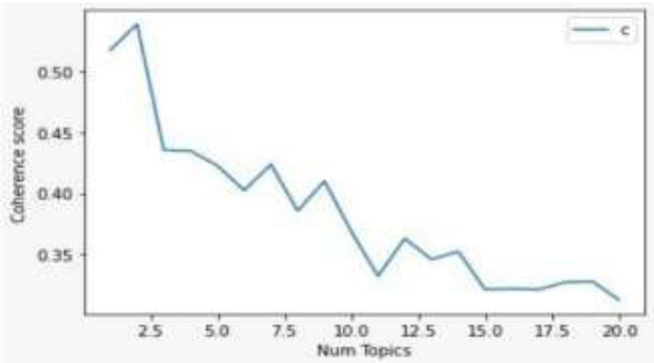


Figure 6 Number of Topics and Coherence Score

After the topic was generated, the next step is to visualize the results of the topic modeling using a library called PyLDAvis.

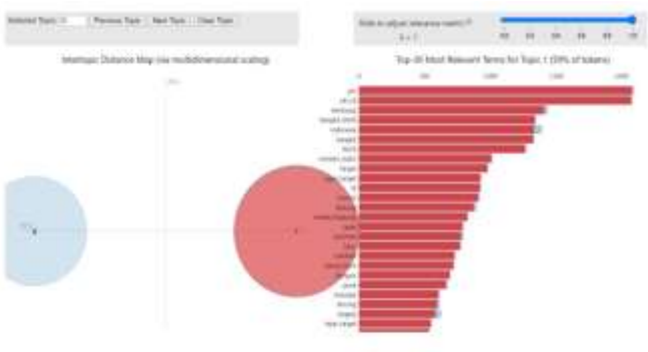

Figure 7 Visualization of Topic Modelling

Figure 7 shows the visualization of the topic modeling with Gensim. If a topic is selected, the circle will turn red, and the right panel will immediately adjust to display a list of the 30 terms most relevant to that topic. The entire frequency of terms in the corpus is shown by a bar chart in the right panel and will turn red when a topic is selected to indicate the occurrence of terms for that topic.

Using the same method as proposed by [28], the LDA topic is then labeled qualitatively. As seen in table II, the first topic was labeled as policy because the words in the first topic contains the terms “ruu” (ruu stands for rancangan undang undang, which is translated as a draft for establishing law or regulation). The second topic was labeled as electricity because it contains the term “pln” (pln stands for perusahaan listrik negara which is translated as the State Electricity Company). However, the labeling for the two topics is still debatable and needs evaluations through an expert judgment.

Table 3 Keywords Used for Data Crawling

| Topic | Words | Label |
|---|---|---|
| 1 | ruu, erick thohir, menteri bumn, tinggi indonesia, larang ekspor, erick, capai persen, kuat bidang, thohir, mobil listrik | Policy |
| 2 | pln, pln id, kembang, bangkit listrik, indonesia, bangkit, listrik, menteri esdm, target, capai target | Electricity |

For the first topic, "policy" has been the most popular topic related to NRE amongst citizens on Twitter for over ten years. The policies related to energy are formulated by The National Energy Council/Dewan Energi Nasional. There might be pros and contras amongst the draft for establishing laws or regulations related to NRE. One of the issues was the pros and cons in the community regarding the construction of nuclear power plants. Another issue was related to electricity sales at a rate below the national cost of supply.

The policy landscape in Indonesia has been rapidly changing. According to Maulidia et al. [29], Indonesia's energy policy lacks a long-term perspective, fails to account for the domestic fossil fuel supplies that are running out, and disregards factors like energy security and environmental sustainability. In addition, the economy plays a major role in how laws and policies are created. Maulidia et al. [29] also state that revoking feed-in-tariffs and other pro-renewable energy regulations and replacing them with regulations that have a weak basis to reflect the private sector's investment risks is seen as a policy uncertainty that deters investment in renewable energy in the country.

Related to second topic, the State Electricity Company or PLN is Indonesia's sole electricity provider and is in charge of electricity distribution to end users. In addition, the company owns the majority of Indonesia's generation assets as well as the entire national grid transmission network. As a result, the PLN has been accused of maintaining an integrated monopoly in the electricity sector, which impedes the development of efficient markets [29]. Due to energy inequality in Indonesia, the private sectors are urged to make investments in remote areas.

From the two topics that have been generated, it might indicate that NRE in Indonesia will accelerate in the following years. Electricity as a new alternative energy source, such as for electric vehicles, could be Indonesia's new hope. It is interesting to note here that other NRE terms have not been included in the topic modeling, such as wind energy, geothermal energy, hydropower, ocean energy, and others. It might indicate that most Indonesian society is unfamiliar with those topics.

## 4. CONCLUSIONS

Our findings indicate that there was a positive trend in the number of Twitter posting related to NRE. The distribution of people who engage in NRE-related posting activities is almost universally discussed, and DKI Jakarta is the province with the most tweets containing NRE-related discussions. During the last ten years, the community has widely discussed and analyzed topics related to NRE that are related to policies and electricity, which are the focus of Indonesia in NRE.

The significance of this study is in demonstrating the method for explaining tweetology in New and Renewable Energy. Revealing the tweetology also might be applied to other topics. The three types of analysis: descriptive, geocoding, and topic modeling analysis might be employed to other topics in Twitter.

## REFERENCES


[1] Deleted for Peer Review

[2] D. R.-J. G.-J. Rydning and others, "The digitization of the world from edge to core," *Framingham: International Data Corporation,* 2018. [Online]. Available: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf. [Accessed: 25-Dec-2021].

[3] Deleted for Peer Review

[4] L. Ngamassi, H. Shahriari, T. Ramakrishnan, and S. Rahman, "Text mining hurricane harvey tweet data: Lessons learned and policy recommendations," *Int. J. Disaster Risk Reduct.*, vol. 70, p. 102753, 2022.

[5] Z. A. Hasibuan, "Towards using universal big data in artificial intelligence research and development to gain meaningful insights and automation systems," in *2020 International Workshop on Big Data and Information Security, IWBIS 2020*, 2020, pp. 9–15.

[6] S. Kemp, "DIGITAL 2022: INDONESIA," 2022. [Online]. Available: https://datareportal.com/reports/digital-2022-indonesia. [Accessed: 23-Oct-2022].

[7] M. Ahlgren, "50+ Twitter statistics and facts for 2022." [Online]. Available: https://www.websiterating.com/research/twitter-statistics/. [Accessed: 16-Sep-2022].

[8] G. Fitzgerald and M. FitzGibbon, "A Comparative analysis of traditional and digital data collection methods in social research in LDCs-Case Studies Exploring Implications for Participation, Empowerment, and (mis) Understandings," *IFAC Proc. Vol.*, vol. 47, no. 3, pp. 11437–11443, 2014.

[9] United Nations Development Programme, "Do you know all 17 SDGs?" [Online]. Available: https://sdgs.un.org/goals. [Accessed: 23-Oct-2022].

[10] Direktorat Jenderal Kekayaan Negara (DJKN), "Bauran Energi Baru Terbarukan Ditargetkan 23 Persen di 2025," 2022. .

[11] V. N. Setiawan, "Capaian Bauran Energi Hijau Juni 2022 Baru 12,8%." [Online]. Available: https://www.cnbcindonesia.com/news/20220701174802-4-352304/capaian-bauran-energi-hijau-juni-2022-baru-128. [Accessed: 23-Oct-2022].

[12] M. Khalil, J. Wong, E. Er, M. Heitmann, and G. Belokrys, "Tweetology of Learning Analytics: What does Twitter tell us about the trends and development of the field?," in *LAK22: 12th International Learning Analytics and Knowledge Conference*, 2022, pp. 347–357.

[13] Z. C. Steinert-Threlkeld, *Twitter as data*. Cambridge University Press, 2018.

[14] W. Ahmed, P. A. Bath, and G. Demartini, "Using Twitter as a data source: An overview of ethical, legal, and methodological challenges," *ethics online Res.*, vol. 2, pp. 79–107, 2017.

[15] M. Allahyari *et al.*, "A brief survey of text mining: Classification, clustering and extraction techniques," in *Proceedings of KDD Bigdas*, 2017.

[16] U. Chauhan and A. Shah, "Topic modeling using latent Dirichlet allocation: A survey," *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–35, 2021.

[17] S. Che, D. Nan, P. Kamphuis, and J. H. Kim, "A comparative analysis of attention to

facial recognition payment between China and South Korea: a news analysis using Latent Dirichlet allocation," in *International Conference on Human-Computer Interaction*, 2021, pp. 75–82.

[18] E. Laoh, I. Surjandari, and L. R. Febirautami, "Indonesians' Song Lyrics Topic Modelling Using Latent Dirichlet Allocation," in *2018 5th International Conference on Information Science and Control Engineering (ICISCE)*, 2018, pp. 270–274.

[19] R. P. F. Afidh and Z. A. Hasibuan, "Indonesia's News Topic Discussion about Covid-19 Outbreak using Latent Dirichlet Allocation," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 2020, pp. 1–6.

[20] E. Zosa and M. Granroth-Wilding, "Multilingual dynamic topic model," *RANLP 2019-Natural Lang. Process. a Deep Learn. World*, 2019.

[21] P. Brzustewicz and A. Singh, "Sustainable Consumption in Consumer Behavior in the Time of COVID-19: Topic Modeling on Twitter Data Using LDA," *Energies*, vol. 14, no. 18, p. 5787, 2021.

[22] A. Pons, C. Vintrò, J. Rius, and J. Vilaplana, "Impact of Corporate Social Responsibility in mining industries," *Resour. Policy*, vol. 72, p. 102117, 2021.

[23] E. De Santis, A. Martino, and A. Rizzi, "An infoveillance system for detecting and tracking relevant topics from Italian tweets during the COVID-19 event," *Ieee Access*, vol. 8, pp. 132527–132538, 2020.

[24] B. Chae and E. Park, "Corporate social responsibility (CSR): A survey of topics and trends using Twitter data and topic modeling," *Sustainability*, vol. 10, no. 7, p. 2231, 2018.

[25] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.

[26] Z. Epstein, "BlackBerry lost 4 million subscribers in Q1 despite new launches," 2013. [Online]. Available: https://bgr.com/general/blackberry-subscribers-q1-2014/. [Accessed: 23-Oct-2022].

[27] N. Wilantika, D. I. Sensuse, S. B. Wibisono, P. L. Putro, and A. Damanik, "Grouping of provinces in Indonesia according to digital divide index," in *2018 6th International Conference on Information and Communication Technology, ICoICT 2018*, 2018, no. c, pp. 380–388.

[28] Deleted for Peer Review

[29] M. Maulidia, P. Dargusch, P. Ashworth, and F. Ardiansyah, "Rethinking renewable energy targets and electricity sector reform in Indonesia: A private sector perspective," *Renew. Sustain. Energy Rev.*, vol. 101, pp. 231–247, 2019.