Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 23, No. 1, November 2023, pp. 201∼214
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v23i1.3308

Detecting Disaster Trending Topics on Indonesian Tweets Using BNgram

Indra, Nur Aliza
Universitas Budi Luhur, Jakarta, Indonesia

Article Info

Article history:

Received August 21, 2023
Revised October 12, 2023
Accepted November 27, 2023

Keywords:

BNgram
Natural disaster
Trending topics
Twitter

ABSTRACT

People on social media share information about natural disasters happening around them, such as the details about the situation and where the disasters are occurring. This information is valuable for understanding real-time events, but it can be challenging to use because social media posts often have an informal style with slang words. This research aimed to detect trending topics as a way to monitor and summarize disaster-related data originating from social media, especially Twitter, into valuable information. The research method used was BNgram. The selection of BNgram for detecting trending topics was based on its proven ability to recall topics well, as shown in previous research. Some stages in detection were data preprocessing, named entity recognition, calculation using DF-IDF, and hierarchical clustering. The resulting trending topics were compared with the topics obtained using the Document pivot method as the basic method. This research showed that BNgram performs better in detecting trending natural disaster-based topics compared to the Document pivot. Overall, BNgram had a higher topic recall score, and its keyword precision and keyword recall values were slightly better. In conclusion, recognizing the significance of social media in disaster-related information can increase disaster response strategies and situational awareness. Based on the comparison, BNgram was proven to be a more effective method for extracting important information from social media during natural disasters.

Corresponding Author:

Nur Aliza, +62895391469971,
Faculty of Information Technology and Informatics Engineering Study Program,
Universitas Budi Luhur, Jakarta, Indonesia,
Email: 1811501426@student.budiluhur.ac.id

How to Cite:
N. Aliza and I. Indra, "Detecting Disaster Trending Topics on Indonesian Tweets Using BNgram", *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 23, No. 1, pp. 201-214, Nov, 2023.

1. INTRODUCTION

Being destructive and unexpected, natural disasters can harm various aspects of life, affecting survivors' well-being, claiming lives, and indirectly increasing unemployment and poverty rates [1]. Indonesia, located at the confluence of the Pacific Ring of Fire and the Alpide Belt, is vulnerable to natural disasters such as earthquakes, tsunamis, landslides, and volcanic eruptions.1 In 2022, a total of 3,544 disaster events were experienced by Indonesia based on data from Badan Nasional Penanggulangan Bencana (Indonesia's National Disaster Relief Agency).2 Although these disasters are difficult to avoid, measures can still be taken to mitigate their impact with the required role of all community individuals, aligning with UU No. 24 of 2007. This study aims to contribute to these mitigation endeavors by utilizing social media, which has become a research interest. Social media enables participatory collaboration and collective knowledge-sharing for public information and warnings during or before a disaster [2]. Twitter is one of the most widely studied social media platforms for disaster management. This is due to the real-time nature of Twitter, where people often post or communicate about events that happen in real-time [3, 4]. In times of disaster, the volume of posts or tweets on Twitter experiences a substantial increase [5]. This could prove useful in understanding how the disaster unfolds, with each user being a potential informant [6, 7]. In Indonesia, tweets on natural disasters have not been fully utilized because of the complexity of the disaster topics and the pervasive use of slang words in social media, which can be a challenging task, even though the tweets could help contribute invaluable information about disaster situations. Given the frequent emergence of natural disaster cases as trending topics on social media and the recognition of valuable insights within Indonesian tweets, in this study, we propose to detect natural disaster-related trending topics using BNgram.

Several studies related to natural disaster mitigation have been conducted. A previous study by [8] assessed tsunami vulnerability and risk in the Banyuwangi district. Using images extracted from the Landsat 8 satellite and optimizing the assessment results with machine learning models, their study evaluated and ranked the ten areas in Banyuwangi based on the vegetation index. Another study by [9] analyzed earthquake hazard predictions on Sumbawa Island. Due to the uncertain data parameters and causative factors of earthquakes, their study determined the relationship between predictor and response variables of earthquake hazards by utilizing the Multivariate Adaptive Regression Splines method. Research by [10] designed a geographic flood mitigation information system for the agricultural sector in East Java. The system could identify flood-prone areas and alert people who make a living in agriculture. A study by [11] created an e-alert application for earthquake disasters based on Android. Their application could convey mitigation information, such as the nearest evacuation sites and post-disaster directions. We utilize the BNgram method to detect natural disaster topics originating from Twitter. The BNgram is a trending topics detection method that was proposed by [12]. Their study compared the BNgram with topic modeling methods in detecting trending topics from 3 real-world events FA Cup, Super Tuesday Primaries, and US Elections datasets taken from Twitter. Their preprocessing techniques included tokenization, stemming, and aggregation. The results are that the BNgram method had better topic recall and maintained relatively good keyword precision and recall than the topic modeling methods. In Indonesia, research using the BNgram method was also carried out by [13]. Their research compared the effectiveness of BNgram and Document pivot methods in detecting political trends in Indonesia. Since the Indonesian language has a different structure and grammar compared to English, the researchers aimed to prove how well the BNgram method works with Indonesian tweets and uses stemming, stop word removal, and aggregation as their preprocessing techniques. The achieved results are that the BNgram method produced higher accuracy in detecting trending topics than the Document pivot method and could appropriately detect political trends in Indonesia.

This study differs from the four disaster-related studies in the scope and data source. We use Indonesian tweets as our data source, which can come from users anywhere in Indonesia. The types of disasters chosen for this study consist of earthquakes, floods, landslides, and eruptions. The outcomes are expected to be trending topics of natural disasters in Indonesia that are hoped to increase awareness regarding natural disasters as a form of mitigation. Additionally, the presented topics could likewise be useful for interested parties in making a decision. This research varies from the preceding two studies that utilized BNgram in terms of both the topic and the preprocessing methods of the dataset. Our dataset topic is natural disasters, whereas the topics of the previous studies are politics and sports. For our preprocessing techniques, we keep numbers because they contain important information in disaster and take into consideration replacing slang words, which was not considered in previous studies.

This study's aims and contributions can be summarized: First, we develop a desktop-based trending topics detection application, which uses Twitter data, utilizing the Python programming language to assist in processing information on disasters from social media. Second, we apply the BNgram method to detect trending topics on Indonesian tweets related to natural disasters, which has never been done before, and evaluate the method based on topic recall, keyword precision, and keyword recall metrics. Third, we compare the results of our evaluation with the ones obtained using the document pivot as a baseline trending topics detection method. Fourth, we present natural disaster-related trending topics that can be used as a source of information for local news. This is achieved by using trigrams, replacing slang words, and preserving numerical data as key parameters.

The rest of this paper is structured as follows. Section 2 describes in detail the research method of this study. Section 3 presents the results that are obtained and provides a discussion. Section 4 concludes the paper.

2. RESEARCH METHOD

The application of the BNgram involves several systematic sequential steps. The systematic stages implemented in this study include the stages of data collection, data preprocessing, named entity recognition (NER), n-gram extraction, the BNgram method, and evaluation of results. As an illustration, the image below visually represents the stages implemented in this study. Figure 1 represents the stages of this study from start to finish, which are also included in the built application as menus except for data collection.

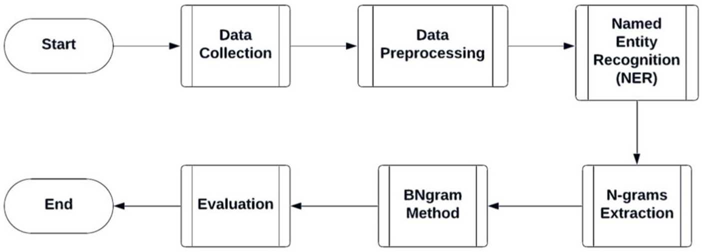

Figure 1. Illustration of research stages

2.1. Data Collection

We use data extracted from Twitter as research material by using the Python Tweepy library. Tweepy allows its users to access the Twitter API by using user keys and tokens to communicate and obtain Twitter data. For this study, we only chose 4 types of natural disasters: earthquakes, floods, landslides, and eruptions. The keywords used as queries consisted of 5, namely *bencana alam* (natural disaster), *#gempa* (#earthquake), *#banjir* (#flood), *#longsor* (#landslide), and *#erupsi* (#eruption). On Twitter, hashtags (#) are usually used to indicate the topic of a tweet; therefore, this study makes use of them. Successfully retrieved tweet data must contain at least 1 specified keyword.

2.2. Data Preprocessing

Data preprocessing is an important stage that prepares text data for analysis. The stages of data preprocessing are important, especially in research that uses data sourced from social media. Message content generated by social media users can be noisy and contain meaningless data [12]. Noisy data can affect the quality and accuracy of the analysis process. Data preprocessing is carried out to reduce the amount of noise by cleaning and transforming unstructured text data into clean data. The following data preprocessing procedure is applied in this study: First, all characters are converted to lowercase to standardize the text. Second, punctuation marks like colons, commas, and dashes occurring within words are replaced with periods to ensure uniformity when conveying the disaster information's date, time, and intensity.

Third, non-alphanumeric characters are removed, retaining numbers for earthquake and disaster data. Following these initial steps, tokenization is applied as the fourth step, segmenting the text into individual words. Fifth, periods at the end of tokens, links, and meaningless tokens are removed. Sixth, a dictionary replaces slang words with their formal forms, resulting from [14]. Seventh, stop words and common filler words are eliminated. In the eighth step, lemmatization simplifies word forms using the Nazief and Adriani Algorithm [15]. The tweets are aggregated every 30 minutes, which serves as the ninth and final step, to group related content by time intervals for further analysis [12, 13].

2.3. Named Entity Recognition (NER)

NER is a process in natural language processing (NLP) that identifies and categorizes named entities within text into predefined entity groups. A named entity refers to a term or expression that distinctly identifies a single item from a group of other items sharing comparable attributes. Examples of named entities generally obtained from NER are the person's name, location, organization, etc [16].

In this study, NER is used to identify entities from the tweet data collected for later to be given more weight at the BNgram stage. We utilize the Polyglot Python library, which results from research [17]. In their study, a massive multilingual NER model was designed. The NER model supports 40 languages, including Indonesian, Arabic, Japanese, Chinese, and others, and can identify the names of persons, locations, and organizations from the text. The training data used came from Wikipedia for each language. The training process utilized neural word embeddings (which produced distributed word representations) and Freebase attributes to build NER annotators automatically.

2.4. N-grams Extraction

N-grams are characterized as a series of consecutive words. Using Latin numeric prefixes, n-grams of size 1 are called unigrams, size 2 are called bigrams, size 3 are called trigrams, and so on. The value of n in n-grams determines the number of strings of words. For example, in a bigram, the sequence consists of two words that appear next to each other in the text. In a trigram, the sequence consists of three words. At this stage, clean data resulting from data preprocessing are extracted into n-grams. We use size 3 (trigram) so that the three words next to each other in the data are extracted. This consideration is made following the result of [13], which stated that using trigrams produced the highest topic recall achieved. This trigram extraction would continue until the last word of the data and would be repeated for all input data.

2.5. BNgram

The BNgram is a method that was proposed in research [12]. In their study, n-grams were considered in detecting trending topics because some terms (words) occur more frequently than others. Also, on Twitter, a large number of tweet messages are copies or retweets of previous messages, so important n-grams tend to have more frequency. With this use of n-grams, it naturally groups terms and can be thought of as offering the first level of term grouping. In addition, the research also proposed a new metric, namely $df - idf$, which is an extension of the metric $tf - idf$ with the introduction of time.

Referring to studies [12, 13], the BNgram method consists of 3 stages, as shown in Figure 2.

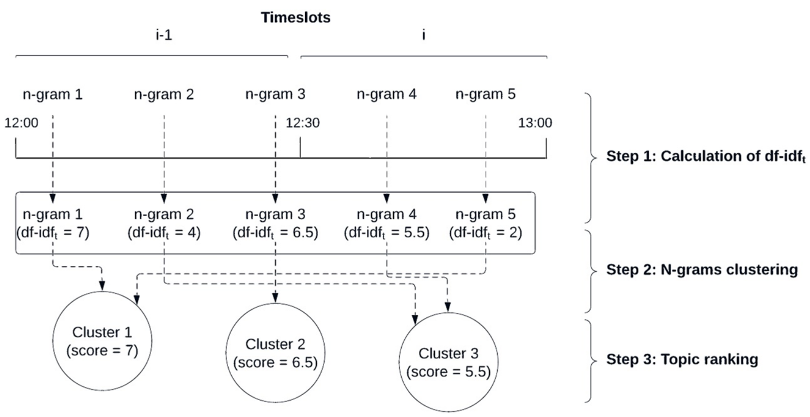

Figure 2. Illustration of the BNgram method

The BN-grams method has three steps, as shown in Figure 2, namely the calculation of $df - idf_t$, n-grams clustering, and topic ranking. In the first step, the n-grams extracted from the tweets are calculated for their $df - idf_t$ score. $df - idf_t$ is based on the frequency of occurrence of n-grams in several tweets in a certain timeslot compared to the logarithm of the frequency of occurrences of n-grams in several tweets in previous timeslots. The following is the formula of $df - idf_t$.

$$df - idf_t = \frac{df_i + 1}{log\left(\frac{\Sigma_{j=i}^{t} df_{i-j}}{t} + 1\right) + 1} \tag{1}$$

Where df_i is the frequency of n-grams appearing in several tweets in the timeslot i; df_{i-j} is the frequency of occurrences of n-grams in several tweets in the previous $i - j$ timeslot, and t is the total timeslots. The score boost is used to give more weight to important n-grams. In this study, n-grams containing named entities will be given a score of $boost$=1.5; besides that, $boost$=1.

In the second step, n-grams that have been calculated their $df - idf_t$, are clustered using Average Linkage Hierarchical Clustering. The clustering is defined using Equation (2). Combining several n-grams into clusters provides more diverse, detailed, factual, and complete information and can be the basis for the resulting trending topics.

$$d(IJ)K = \frac{\Sigma_{g1}\Sigma_{g2}d(g_1, g_2)}{N_{IJ}N_K} \tag{2}$$

Where $d(IJ)K$ is the distance between clusters IJ and K; N_{IJ} and N_K are the size or amount of data for each cluster; and $d(g_1, g_2)$ is the distance between n-grams g_1 and g_2, which is defined by the formula below.

$$d(g_1, g_2) = 1 - \frac{A}{min\{B, C\}} \tag{3}$$

Where $d(g_1, g_2)$ is the distance between n-grams g_1 and g_2; A is the number of tweets containing n-grams g_1 and g_2; B is the number of tweets that contain n-gram g_1; and C is the number of tweets that contain n-gram g_2. The n-grams g_1 and g_2 are combined if the distance is ≤ 0.5.

In the last step, each cluster represents a topic or event that occurs in social media. The clusters resulting from the previous stage are sorted based on their $df - idf_t$ scores. The clusters that contain n-grams with the highest $df - idf_t$ scores represent the most discussed topics.

2.6. Document-Pivot (Doc-p)

Document pivot is a topic detection and tracking (TDT) method used as a benchmark method in this study. The implementation of the method uses locality-sensitive hashing (LSH) for clustering the tweets. The method involves four stages, described as follows [12, 13]: clustering of tweets using LSH, eliminating clusters whose members are smaller than the threshold, calculating the score of each cluster, and ranking topic.

The first step is to cluster tweets using LSH. LSH is a method that relies on hashing each item into buckets, so the probability of collision is much higher for nearby items. The following are the steps of LSH: First, a dictionary that contains a unique glossary of collected tweets is created. Each dictionary entry includes an index term containing a single word within a sentence. Second, according to the index term in the dictionary, every collected tweet is converted into a bit array signature and is inserted into a collection of hash tables *S*.

The LSH uses *k* bits and *L* hash tables, and two documents are considered collided if only those two documents have the same bit array signature. A document is several tweets posted in a certain constant length of time. Third, collided tweets, which have the same bit array signature, are included in the same bucket in the hash tables collection *S*. Fourth, a cosine similarity is calculated on the tweets in *S*. In the last step, if the cosine similarity score exceeds 0.5, the tweets are included in the same cluster. When a new tweet arrives, it is hashed into a bucket and computed for cosine similarity to all other tweets processed. If the similarity is ≤ 0.5, then assign the tweet to the same cluster as its best match; otherwise, create a new cluster for the tweet.

The second step is eliminating clusters with members smaller than the threshold. The threshold defined in this study is 2. Clusters grouped from the previous stage would be eliminated if their members were under 2. The third step is to calculate the score of each cluster. Each cluster score is computed using Equation (4).

$$score_c = \sum_{i=1}^{|Docs_c|}\sum_{j=1}^{|Words_i|} exp(-p(w_{ij})) \tag{4}$$

Where w_{ij} is the number of terms j appearing in the document i in the cluster. The probability of the appearance of a single term $p(w_{ij})$ is estimated as defined by equation (5).

$$p(w|corpus) = \frac{N_w + \delta}{(\Sigma_u^n N_u) + \delta n} \tag{5}$$

Where N_w is the total appearances of term w in the corpus; N_u is the total occurrences of term u; n is the number of term types appearing in the corpus; and δ is the constant thaset to 0.5 to ensure that a new term that does not occur in the corpus is not assigned a probability of 0.

The fourth step is topic ranking. The clusters are sorted based on the scores computed in the previous stage, and the top clusters are considered trending topics. Trending topics are presented in the sentence that contains a set of keywords in each cluster.

2.7. Evaluation Method

Evaluation is conducted to measure the method used based on the selected metrics. We use 3 metrics to assess the relevance of trending topic results to data from Indonesian online news media. The three metrics are topic recall (TR), keyword precision (KP), and keyword recall (KR). The topic recall is the ratio between the results of trending topics relevant to topics in the ground truth (in this case, Indonesian online news media data) per all topics in the ground truth. Keyword precision is the ratio of consistent trending keywords to ground truth keywords per all trending topic keywords. Keyword recall is the ratio of consistent trending keywords to ground truth keywords per all ground truth keywords [13]. These three metrics are formulated in the equation(6, 7, 8).

$$topic\ recall\ (TR) = \frac{|GT \cap BT|}{|GT|} \tag{6}$$

$$keyword\ precision\ (KP) = \frac{|KGT \cap KBT|}{|KBT|} \tag{7}$$

$$keyword\ recall\ (KR) = \frac{|KGT \cap KBT|}{|KGT|} \tag{8}$$

Where GT is a set of topics from ground truth, BT is a set of trending topics, KGT is a set of keywords on ground truth, and KBT is a set of keywords on a trending topic.

3. RESULT AND ANALYSIS

3.1. Data Collection

The data collection stage was carried out from May 31 to June 06, 2023. Tweets containing one of the keywords *bencana alam* (natural disaster), *#gempa* (#earthquake), *#banjir* (#flood), *#longsor* (#landslide), and *#erupsi* (#eruption), were extracted to be used as research data. The selection of keywords was based on the Twitter hashtag feature, making it easier for Twitter users to find the topic of a tweet. Meanwhile, the keyword ”natural disaster” was chosen to find out public opinion or important information about natural disasters. In total, we collected 1,000 tweets. The collected data was converted into a data frame with the help of the Python Pandas library. Next, the data frame was exported to an Excel file. Table 1 is an example of Twitter data that was successfully extracted.

Table 1. Data Collection Sample

Content	Timestamp
#Gempa Mag:3.7 31-May-2023 19:04:02WIB Lok:5.80LS- 102.95BT (90 km Tenggara ENGGANO-BENGKULU) [#Earthquake Mag:3.7 31-May-2023 19:04:02WIB Loc:5.80LS- 102.95BT (90 km south-east of ENGGANO-BENGKULU)]	2023-05-31 19:07:49
Banjir dan Longsor Melanda 3 Kecamatan di Ambon https://t.co/PPLD6KztfR #ambon #banjir #Longsor [Floods and Landslides Hit 3 Districts in Ambon https://t.co/PPLD6KztfR #ambon #flood #Landslide]	2023-05-31 14:48:26
Aktivitas Gunung Merapi Meningkat: Sehari, 15 Kali Guguran Lava Pijar [Mount Merapi’s Activity Increases: 15 Times Incandescent Lava Falls a Day]	2023-06-02 13:44:28
potensi #banjir pesisir #rob https://t.co/jo43KjyNRd [potential #coastal flood #rob https://t.co/jo43KjyNRd]	2023-06-03 09:25:31
Wilayah Barat Daya Pidie Jaya Aceh Diguncang Gempa Bumi Pagi Ini, Penjelasan BMKG soal Gempa Terkini #Gempa #GempaBumi #GempaTerkini #Aceh #PidieJaya #BMKG [The Southwest Region of Pidie Jaya Aceh Was Rocked by an Earthquake This Morning, BMKG Explains the Latest Earthquake #Earthquake #Earthquake #LatestEmpa #Aceh #PidieJaya #BMKG]	2023-06-03 09:29:57

3.2. Preprocessing Results

The collected data underwent preprocessing at this stage. This was done to reduce the noise contained in the data and prepare the data for the remaining stages. There are 9 stages in preprocessing, but for clarity and to save space, we only show the first eight stages and choose only one example. Table 2 elaborates on the results of the preprocessing.

Table 2. Preprocessing Stages Example

Stage	Example
Raw data	#Gempa Mag:3.9, 31-May-2023 06:33:45 WIT, Lok:1.53 LU,127.50 BT (Laut, 52 km Utara JAILOLO-MALUT), Kedlmn:175 Km. ::BMKG-TNT https://t.co/KRGSfbggQP [#Earthquake Mag:3.9, 31-May-2023 06:33:45 WIT, Lok:1.53 North Latitude, 127.50 East Longitude (Sea, 52 km North of JAILOLO-MALUT), Depth: 175 Km. ::BMKG-TNT https://t .co/KRGSfbggQP]
Change all characters to lowercase	#gempa mag:3.9, 31-may-2023 06:33:45 wit, lok:1.53 lu,127.50 bt (laut, 52 km utara jailolo-malut), kedlmn:175 km. ::bmkg-tnt https://t.co/krgsfbggqp
Convert colon, comma, and dash to period	#gempa mag.3.9, 31.may.2023 06.33.45 wit, lok.1.53 lu.127.50 bt (laut, 52 km utara jailolo.malut), kedlmn.175 km. :.bmkg.tnt https://t.co/krgsfbggqp
Remove characters except for alphanumerics and periods	gempa mag.3.9 31.may.2023 06.33.45 wit lok.1.53 lu.127.50 bt laut 52 km utara jailolo.malut kedlmn.175 km. .bmkg.tnt httpst.cokrgsfbggqp
Tokenization	gempa, mag.3.9, 31.may.2023, 06.33.45, wit, lok.1.53, lu.127.50, bt, laut, 52, km, utara, jailolo.malut, kedlmn.175, km., .bmkg.tnt, httpst.cokrgsfbggqp
Remove period at the end of the token, links, and meaningless tokens	gempa, mag.3.9, 31.may.2023, 06.33.45, wit, lok.1.53, lu.127.50, bt, laut, 52, km, utara, jailolo.malut, kedlmn.175, km, .bmkg.tnt
Replace slang words	gempa, mag.3.9, 31.may.2023, 06.33.45, wit, lok.1.53, lu.127.50, buat, laut, 52, kilometer, utara, jailolo.malut, kedlmn.175, kilometer, .bmkg.tnt
Eliminate stop words	gempa, mag.3.9, 31.may.2023, 06.33.45, wit, lok.1.53, lu.127.50, laut, 52, utara, kilometer, jailolo.malut, kedlmn.175, kilometer, .bmkg.tnt
Lemmatization (resulting in clean data)	gempa, mag.3.9, 31.may.2023, 06.33.45, wit, lok.1.53, lu.127.50, laut, 52, utara, kilometer, jailolo.malut, kedlmn.175, kilometer, .bmkg.tnt

3.3. NER Results

At this stage, the clean data obtained from the previous stage were detected for named entities. As stated earlier, we use the NER model from [17]. Tweets that contain named entities are given a boost in the calculation of $df - idf_t$ stage, to prioritize the tweets. In doing so, the results of the trending topics would be more meaningful. Examples of named entities that have been detected using our application can be observed in Table 3. The NER model used the BIO technique for labeling the entities, which is indicated by the use of "I-," which means "inside." "I-LOC" and "I-PER" are the categories of entities successfully recognized from the tweets by the model; LOC indicates a location, while PER signifies a person.

Table 3. Examples of Detected Named Entities

Tweet	Named entities
Gempa M5,3 Guncang Mentawai, BMKG: Tak Berpotensi Tsunami Gempa Mentawai [M5.3 Earthquake Shakes the Mentawai, BMKG: No Potential Tsunami for the Mentawai Earthquake]	[I-LOC([Mentawai])]
Jokowi memimpin perjuangan melawan perubahan iklim bencana alam. NKRIMembangun [Jokowi leads the fight against climate change natural disasters. NKRIBuilding]	[I-PER([Jokowi])]

3.4. Implementation of BNgram

As we described earlier, implementing the BNgram involves three main steps, with one additional step: the extraction of n-grams. Clean data was extracted from their features through the NER process in the form of trigrams. One tweet can have many trigrams. Therefore, our 1000 tweets can increase exponentially. In our first experiment using our built application, we obtained a total of approximately 15,000 trigrams. Since the BNgram uses Hierarchical Clustering, where the clustering requires calculating the distance between all pairs of trigrams, the computational time required to handle our data increased. Our application got stuck while performing the clustering, so we had to accommodate it by splitting our 1000 tweets into several datasets. We split our tweets based on the date they were posted. Table 4 shows the details of the split datasets. The "date" column indicates the date the tweets in the dataset were posted, and the ”total tweets” column specifies the number of tweets in the dataset.

Table 4. Details of the Datasets

Dataset	Date	Total Tweets
D1	2023-05-31	93
D2	2023-06-01	147
D3	2023-06-02	173
D4	2023-06-03	168
D5	2023-06-04	172
D6	2023-06-05	205
D7	2023-06-06	42

We ran our application using the datasets sequentially. On average, we got approximately 1000 trigrams in the trigrams extraction step, though it can depend on the number and the length of tweets. D7 has the smallest number of trigrams, whereas D6 has the highest. After the feature extraction part, the acquired trigrams were computed based on Equation (1). The value of $df - idf_t$ represents the weight of a trigram in a timeslot by considering the occurrence of that trigram in the previous timeslots. The higher value indicates that the trigram is trending. However, if the trigram has frequently appeared in previous timeslots, its weight would be slightly reduced to detect new topics.

Next, trigram clustering was carried out using Equation (2), trigrams were grouped to form topic clusters. The clustering begins by making all the trigrams into their respective clusters. Clusters are formed by calculating the average distance between pairs of trigrams using Equation (3) for all existing trigrams. If the distance results are $\leq$ 0.5, then the two trigrams would be grouped into one cluster. Otherwise, there is no merging. This means that two trigrams must have a 50% occurrence of the same tweet to be considered as having the same topic [12].

Afterward, the clusters resulting from the clustering were ranked. This ranking refers to the weight $df - idf_t$ of each trigram contained in each cluster. The top-ranked clusters are considered trending topics. In our experiment using the split datasets, we obtained many clusters that have a higher $df - idf_t$ score than the rest. We noticed that we mostly got earthquake-based trending topics. This is because our data collection comprised mostly of earthquake topics, even though, at the time, we were not aware of these earthquakes happening in certain places. Some trending topics we got can be long and explain well. However, they can be redundant, whereas others (a small part of it) can be short and do not explain the topic well. Table 5 specifies a couple of the trending topics that we got from our experiment.

Table 5. Trending Topics Results

Dataset	Trending Topics	Sample Tweet
D1	gempa update mag.5.3 31.mei.23 07.58.38 wib lok.1.05 ls 98.40 pusat laut 167 baratlaut kep mentawai kedlmn.10 [earthquake update mag.5.3 31.mei.23 07.58.38 WIB lok.1.05 ls 98.40 sea center 167 northwest of Mentawai Islands depth.10]	[M5.3 Earthquake Shakes Mentawai, BMKG: No Tsunami Potential @BMKGinfo #Earthquake #MentawaiEarthquake] Gempa M5,3 Guncang Mentawai, BMKG: Tak Berpotensi Tsunami @infoBMKG #Gempa #GempaMentawai
D2	gempa darat timurlaut bukittinggi	#Gempa Mag:3.8 01-Jun-2023 16:29:32WIB Lok:0.29LS-100.39BT (2 km TimurLaut BUKITTINGGI-SUMBAR) Kedlmn:10 Km https://t.co/YXvCwscSTt #EARTHQUAKE #SUMBAR
	[land earthquake northeast of Bukittinggi]	[#Earthquake Mag:3.8 01-Jun-2023 16:29:32WIB Lok:0.29S-100.39E (2 km NEO-BUKITTINGGI-SUMBAR) Depth:10 Km https://t.co/YXvCwscSTt #EARTHQUAKE #SUMBAR]

3.5. Evaluation Results and Discussion

Evaluation in this study was carried out to measure the relevance (relationship) or similarity between trending topics results from the BNgram and the Doc-p to data from Indonesian online news media (Kompas.com, Tempo.co, antaranews.com, and liputan6.com), which are referred to as ground truth. The news data used as ground truth is news related to the topic of natural disasters, and its title was extracted. In addition, the data were selected around the period from which Twitter data was collected and came from official news media. For Doc-p implementation, we use programs developed by [12], in which we input our natural disaster datasets and obtain trending topics detected by the method. The preprocessing stages for Doc-p, consisting of tokenization, stemming, and aggregation, are the same as those designed by the authors. This could serve as a comparison between our preprocessing techniques and the techniques used in [12]. We use three metrics for the evaluation stage as defined in Equations (6)-(8) and manually measure each split dataset. Table 6 summarizes the results of our evaluation.

Table 6. Evaluation Results

Dataset	Topic Recall (%)		Keyword Precision (%)		Keyword Recall (%)	
	BNgram	Document pivot	BNgram	Document pivot	BNgram	Document pivot
D1	80	60	36	16	37	16
D2	80	50	27	24	32	19
D3	70	50	27	26	43	24
D4	60	50	31	24	57	26
D5	80	80	50	36	65	46
D6	80	60	37	30	55	32
D7	20	20	11	15	10	12

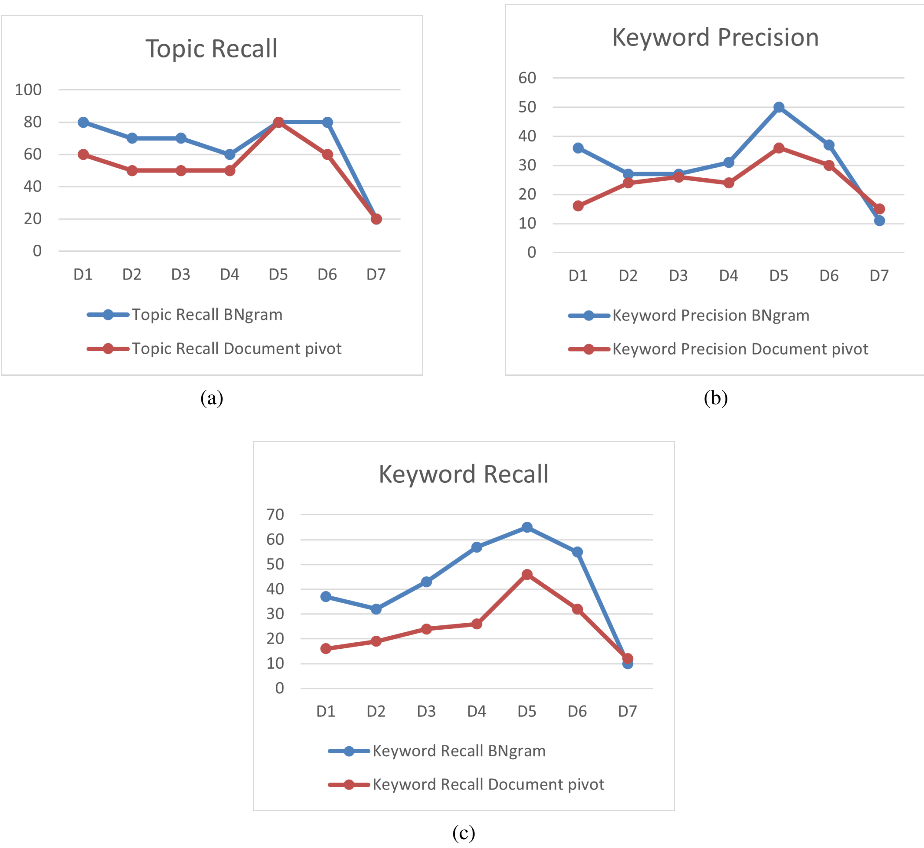

Figure 3. Evaluation results visualization (a) Topic Recall; (b) Keyword Precision; (c) Keyword Recall

Figure 3 visualizes the results of the two methods: topic recall and keyword precision. The x-axis represents the split datasets based on the time posted, and the y-axis explains the scores of evaluation metrics. Our evaluation shows that trending topics based on BNgram could better detect natural disasters that match ground truth compared to Doc-p. This can be seen in the topic recall results shown in Figure 3a. Aside from D5 and D7 scores that are even, the BNgram has higher scores on the remaining datasets than the Doc-p. The topic recall scores achieved by BNgram in D1, D2, D3, D4, and D6 are 80%, 70%, 70%, 60%, and 80%, respectively; where as the Doc-p scores are 60%, 50%, 50%, 50%, and 60%, respectively. This suggests that the BNgram could better identify and capture disaster trending topics, where one topic consists of a set of keywords that align with ground truth. This finding aligns with the results from [12, 13] and can be attributed to the BNgram using n-gram co-occurrences (instead of using uni-grams) and $df - idf_t$ topic ranking. Doc-p clusters documents based on similarities among the documents. Under conventional circumstances, the clustering of actual documents, using Doc-p would yield favorable results. However in social media, where there are noises in data and topics are not as clear, BNgram would perform better than Doc-p. In BNgram, the use of trigrams is thought to help capture and describe topics that are similar to ground truth. This is because trigrams contain three terms structure, which in Indonesian grammar represents subject, predicate, and object pattern, and therefore could explain the topic well [13]. The $df - idf_t$ is a score for burstiness detection that can significantly assist in determining the most rapidly emerging topics [12]. In detecting social media topics, burstiness topics could reflect subjects that are discussed the most and help monitor how those topics evolve.

The named entity recognition in the BNgram also contributed to detecting trending topics that match ground truth, where trigrams that contain proper nouns (persons, locations, and organizations) were boosted. As a result, the resulting trending topics from BNgram are more consistent with real-life news than trending topics produced from Doc-p, especially in the case of natural disasters where location names play a pivotal role in determining the affected areas. As for the topics that typically match ground truth, we found that the tweets that were posted from official government accounts, like *Badan Meteorologi dan Klimatologi Geofisika* (Indonesia's Meteorology, Climatology, and Geophysical Agency), contributed significantly. This is because the data used are valid and accountable, so people put more trust in those information. However, we noticed in our data collection that normal users also contributed by retweeting the tweets, consequently making them more available to a larger audience. Additionally, they give their perspective on natural disasters, which can, in turn, help describe the disasters in more detail in the detected trending topics.

In D7, both BNgram and Doc-p achieved very low in topic recall. This is caused by the small number of tweets contained in D7, which is only 42 tweets. This dataset did not provide much data or context for BNgram and Doc-p to detect trending topics. Hence, the score achieved is only 20%. Regarding the different results of topic recall between D1, D5, and D6, to D2 - D4 for BNgram, we assume they come from the contents variation of the datasets. A more varied dataset could hinder the BNgram from detecting trending topics relevant to news. This is because the topics in the datasets are mixed up and not focused. This reasoning is consistent with [13], which stated that datasets with complex topic distributions could result in the BNgram producing fewer topics consistent with local news. As for this study, because the keywords we chose to collect data are varied in disaster types, this could be the reason why our datasets are not focused. If we choose only one disaster type to detect trending topics, this potentially can lead to a higher topic recall for all of the datasets. Keyword precision and keyword recall results are low for both methods, as shown in Table 6, with the best scores obtained are 50% and 65%, respectively, for BNgram, while Doc-p is 36% and 46%, respectively. For Doc-p, this occurs because the method produced many keywords that did not match the keywords found in the ground truth. Some trending topics the method produces differ considerably (results not shown in this paper), consequently affecting the keyword precision and recall values. This could be linked to the inappropriate use of the preprocessing techniques, which we explain in the next paragraph. As for BNgram, the low keyword precision and recall scores are due to the word composition in trending topic sentences originating from BNgram being more detailed and numerous compared to keywords derived from the ground truth. Conversely, the ground truth relies on popular language with a more succinct word count in its composition. As a result, this leads to lower keyword precision and keyword recall values achieved using BNgram. Surprisingly, this discovery is not in line with [12], which said that the BNgram produced pretty clean keywords, which they found in their experiments. This suggests that our research techniques, specifically preprocessing, contributed to this change.

As mentioned, we maintained numbers in our data and considered removing slang words due to the frequent use of those in social media, which previous studies [12, 13] did not apply. Based on our observation, we note that preserving numbers gives us a clearer picture of the disasters that have occurred, such as date, time, and intensity of scale. The effect of preserving numbers could also be seen in the keyword precision and recall achieved by BNgram, which are higher than Doc-p, where the preprocessing did not include preserving numbers. This can be clearly observed in D5, where both methods achieved the same 80% topic recall, but BNgram attained greater keyword precision and recall scores than Doc-p, as seen in Figures 3b and 3c. However, maintaining numbers has some drawbacks, such as the BNgram producing large keywords, thus lowering keyword precision and keyword recall of the detected trending topics. The effect of removing slang words, we observe, is not significant. Some words even changed their meaning entirely because of the ambiguity of the slang words chosen by the users. The ambiguity occurred due to some slang words that have more than 1 formal form, e.g., "kt" could be *kita* (we) or *kata* (word or said) [14], which can lower the quality of the detected trending topics. This problem of converting inappropriate slang words could arise because incorrect dictionaries are used for natural disasters. To minimize the problem, we recommend using a relevant colloquial dictionary or even creating one.

Overall, we conclude that the BNgram better detects natural disaster-based trending topics than Doc-p. The majority of BNgram has higher topic recall scores and slightly higher keyword precision and recall values than the Doc-p. Preprocessing techniques such as preserving numbers and replacing slang words have some merits but have downsides if not utilized correctly.

3.6. Trending Topics Detection Application

In this study, we develop a trending topics detection application to aid in data processing and detecting disaster topics. Figure 4 depicts the user interface of our built application. Users can use this application to get important information or public opinion about natural disasters based on data from Twitter. The data must be imported first before it can be used, and the application currently can only process data in an Excel file. Subsequent stages are the same as explained above and must be followed sequentially. A word cloud visualization feature displays trending topics for easy comprehension. Hopefully, this application could contribute to mitigation efforts in Indonesia, and the resulting trending topics could serve as a source of information.

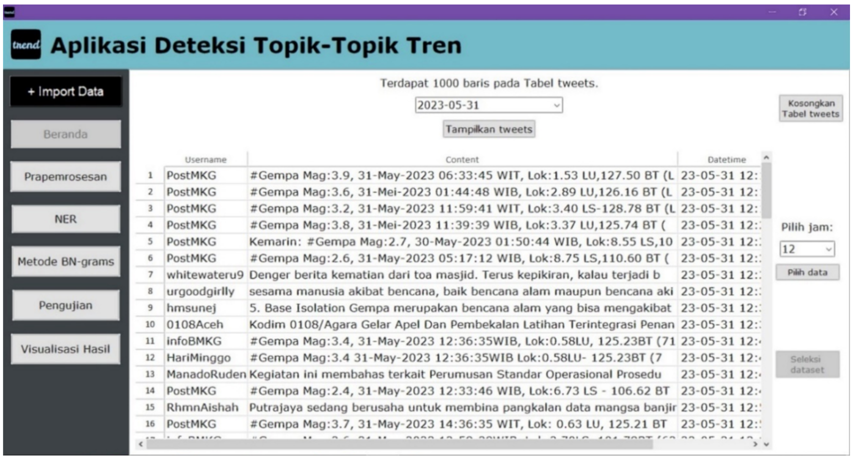

Figure 4. A screenshot of the built application

3.7. Comparison of Research Results

This study provides a comparison between its results and the results of previous studies that used BNgram. As shown in Table 7, the BNgram method typically has relatively small accuracy results. This is proven by the results of [13], which produced topic recall of 60%, 40%, and 60,5%. On the other hand, a study by [12] produced 76%, 50%, and 48,4%. This is the superior point of our research. BNgram produces a topic recall of 80% in the majority. This has higher topic recall even though the datasets are different and relatively few. The comparison is further detailed in Table 7.

Table 7. Comparison of Research Results

References	Novelty	Result (Previous Study)	Result (This Study)
[12]	Comparison: The result of a previous study shows that the highest topic recall was 76,9 %, 50%, and 48,4% in the FA Cup, Super Tuesday, and US Elections datasets, respectively, and using unigrams, bigrams, and trigrams in English tweets. The highest topic recall in the dataset with a narrow scope (the match itself). For df-idft, calculation was using NER with English tweets. Novelty: This research shows that the highest topic recall is 80% with natural disaster dataset D1-D7 and using trigrams in Indonesian tweets. The highest topic recall in the dataset also has a wide scope (Seven datasets with different days). This study also uses NER in the Indonesian language, which increases topic recall in the natural disaster dataset.	Dataset → Topic recall: FA Cup: 76,9% Super Tuesday: 50% US Elections: 48,4% NER: NER in English to detect persons, locations, and organizations for US Elections. N-grams contains a named entity boost=1,5;Otherwise=1;	Dataset → Topic recall: D1:80% D2:80% D3:70% D4:60% D5:80% D6:80% D7:20% D1-D7: Natural disaster dataset in Indonesian tweet NER: NER with the Indonesian language to detect persons, locations, and organizations related to natural disasters. n-grams that contains a named entity boost=1,5; Otherwise boost=1 (based previous research [12])
[13]	Comparison: The result of the previous study shows that the highest topic recall was 60%, 40%, and 60,5% in P1, P2, and P3 datasets, respectively, and using unigrams until sixgrams in Indonesian tweets related to Elections DKI Jakarta and politics in Indonesia. Trigrams have the highest topic recall. For df-idft, calculation was not using NER with Indonesian tweets. Novelty: This research shows that the highest topic recall is 80% with natural disaster dataset D1-D7 and also using trigrams in Indonesian tweets (based on previous research in [13]). The highest topic recall in the dataset has a wider scope and more distribution terms (Seven datasets with different days in D1-D7 related natural disasters). This study also uses NER in the Indonesian language, increasing topic recall in the natural disaster dataset.	Dataset → Topic recall: P1: 60%; P2: 40%; P3: 60,5% NER: NER is not used in this study because n-grams contain a named entity boost=1 for all terms.	Dataset → Topic recall: D1:80% D2:80% D3:70% D4:60% D5:80% D6:80% D7:20% D1-D7: Natural disaster dataset in Indonesian tweet. NER: NER with the Indonesian language to detect persons, locations, and organizations related to natural disasters. n-grams that contains a named entity boost=1,5; Otherwise boost=1 (based previous research [12])

4. CONCLUSION

In this paper, we proposed using BNgram to detect natural disaster trending topics and compared the results to Doc-p. From our experiments, we found that implementing BNgram could better detect trending topics following local news when compared to Doc-p. This is indicated by the topic recall scores that, in the majority, BNgram has higher values, where the volume and variety of the dataset play a significant role in achieving good topic recall value. The use of trigrams, $df - idf_t$, and named entity recognition also influenced the resulting trending topics produced by BNgram. For keyword precision and keyword recall scores, our experiments revealed that both methods achieved low scores in both aspects, with the best scores obtained at 50% and 65%, respectively, for BNgram, while Doc-p is 36% and 46%, respectively. This was caused by both methods' keywords failing to align with keywords present in local news, and some trending topics were not found in local news. However, overall, BNgram has slightly higher scores than Doc-p. The resulting trending topics of this study can be used as recommendation information for the government and as sources of news information for official news media outlets in the following days. Our application can offer a quicker means of obtaining information about natural disasters instead of going directly to the field, thus providing essential data for decision-making.

For future studies, we recommend improving preprocessing techniques, such as creating dictionaries for slang words and using a rule-based filter to remove unnecessary characters and words. Furthermore, the algorithm used for n-gram clustering in BNgram should be improved, or a more efficient clustering method should be considered to process more data.

5. ACKNOWLEDGEMENTS

We want to thank everyone who directly or indirectly helped us to finish this paper.

6. DECLARATIONS

AUTHOR CONTIBUTION

Two researchers conducted this study. Nur Aliza gathered and analyzed data, designed and programmed the application, and drafted the journal. Indra analyzed and evaluated data and provided guidance, constructive feedback, and final approval for the research journal.

FUNDING STATEMENT

This study did not receive grants from anywhere, including the public, commercial, or non-profit sectors.

COMPETING INTEREST

The authors declare that they have no competing financial interests or personal relationships that could have influenced the research in this paper.

REFERENCES

[1] R. Putri Pranandari, K. Shuzuran, and M. Ghafur Wibowo, "Pengaruh Indeks Risiko Bencana, Pdrb Dan Tingkat Pengangguran Terhadap Kemiskinan Di Wilayah Berdominasi Perkotaan Di Provinsi Jawa Barat Periode 2017-2020," *J-ESA (Jurnal Ekonomi Syariah)*, vol. 5, no. 1, pp. 14–34, 2022.

[2] B. Kusumasari and N. P. A. Prabowo, "Scraping social media data for disaster communication: how the pattern of Twitter users affects disasters in Asia and the Pacific," *Natural Hazards*, vol. 103, no. 3, pp. 3415–3435, 2020.

[3] R. Nugroho, C. Paris, S. Nepal, J. Yang, and W. Zhao, "A survey of recent methods on deriving topics from Twitter: algorithm to evaluation," *Knowledge and Information Systems*, vol. 62, no. 7, pp. 2485–2519, 2020.

[4] Y. Shi, T. Sayama, K. Takara, and K. Ohtake, "Detecting flood inundation information through Twitter: The 2015 Kinu River flood disaster in Japan," *Journal of Natural Disaster Science*, vol. 40, no. 1, pp. 1–13, 2019.

[5] M. Sreenivasulu and M. Sridevi, "Comparative study of statistical features to detect the target event during disaster," *Big Data Mining and Analytics*, vol. 3, no. 2, pp. 121–130, 2020.

[6] S. Mendon, P. Dutta, A. Behl, and S. Lessmann, "A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters," *Information Systems Frontiers*, vol. 23, no. 5, pp. 1145–1168, 2021.

[7] P. H. Barros, I. Cardoso-Pereira, H. Allende-Cid, O. A. Rosso, and H. S. Ramos, "Leveraging Phase Transition of Topics for Event Detection in Social Media," *IEEE Access*, vol. 8, pp. 70 505–70 518, 2020.

[8] G. C. a. Wibowo, S. Y. J. Prasetyo, and I. Sembiring, "Tsunami Vulnerability and Risk Assessment Using Machine Learning and Landsat 8," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 2, pp. 365–380, 2023.

[9] D. Priyanto, B. K. Triwijoyo, D. Jollyta, H. Hairani, N. Gusti, and A. Dasriani, "Data Mining Earthquake Prediction with Multivariate Adaptive Regression Splines and Peak Ground Acceleration," *Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, vol. 22, no. 3, pp. 583–592, 2023.

[10] M. R. Aprillya and U. Chasanah, "Geographic Information System Multi Attribute Utility Theory for Flood Mitigation in Agricultural Sector," *Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, vol. 22, no. 1, pp. 117–128, 2022.

[11] Apriani, S. J. Putra, I. Ismarmiaty, and N. G. A. Dasriani, "E-Alert Application In Facing Earthquake Disaster," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 187–194, 2020.

[12] L. M. Aiello, G. Petkos, C. Martin, D. Corney, S. Papadopoulos, R. Skraba, A. Goker, I. Kompatsiaris, and A. Jaimes, "Sensing trending topics in twitter," *IEEE Transactions on Multimedia*, vol. 15, no. 6, pp. 1268–1282, 2013.

[13] Indra, E. Winarko, and R. Pulungan, “Trending topics detection of Indonesian tweets using BN-grams and Doc-p,” *Journal of King Saud University - Computer and Information Sciences*, vol. 31, no. 2, pp. 266–274, 2018.

[14] N. Aliyah Salsabila, Y. Ardhito Winatmoko, A. Akbar Septiandri, and A. Jamal, “Colloquial Indonesian Lexicon,” in *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*. IEEE, 2018, pp. 226–229.

[15] M. Adriani, J. Asian, B. Nazief, S. M. Tahaghoghi, and H. E. Williams, “Stemming Indonesian,” *ACM Transactions on Asian Language Information Processing (ACM Trans. Asian Lang. Inf. Process.)*, vol. 6, no. 4, pp. 1–33, 2007.

[16] J. Li, A. Sun, J. Han, and C. Li, “A Survey on Deep Learning for Named Entity Recognition,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 50–70, 2022.

[17] R. Al-Rfou, V. Kulkarni, B. Perozzi, and S. Skiena, “POLYGLOT-NER: Massive Multilingual named entity recognition,” in *SIAM International Conference on Data Mining 2015, SDM 2015*, 2015, pp. 586–594.