Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 741-748
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3750

Clustering Content Types and User Motivation using DBSCAN on Twitter

Made Mita Wikantari, Yuliant Sibaroni, Aditya Firman Ihsan*

School of Computing, Informatics, Telkom University, Bandung, Indonesia
Email:[1]mitawikantari@student.telkomuniversity.ac.id,[2]yuliant@telkomuniversity.ac.id,[3,*]adityaihsan@telkomuniversity.ac.id
Correspondence Author Email: adityaihsan@telkomuniversity.ac.id
Submitted: 28/06/2023; Accepted: 21/08/2023; Published: 25/08/2023

Abstract−We are currently in an era full of information and communication technology. One of the communication media used is Twitter. Twitter is a microblogging service that is used by its users to express their thoughts on a topic called a tweet. Tweets that are posted can be either positive tweets or negative tweets. One of the topics that is currently being discussed by Twitter users is Anies Baswedan as a 2024 Indonesian Presidential Candidate. Many people have tweeted this but it is not known how many users support or reject Anies Baswedan to run as a 2024 Indonesian presidential candidate. To assist the analysis, use the method clustering namely algorithm (Density-Based Spatial Clustering of Application with Noise). DBSCAN has the advantage of being able to detect data that is not included in a cluster and will be considered noise. This can improve the accuracy of the grouping because the data in the cluster will be cleaner. The TF-IDF Vectorizer is used to make it easier for programs to manage data because it can turn sentences into vectors that can be processed by the algorithm. To determine the evaluation of the program, the silhouette score method will be used. The results of calculating the silhouette score show a value of 0.29 with the formation of 3 clusters. Then an analysis is carried out based on the top words from each cluster and it can be identified that cluster 0 has a positive category supporting Anies Baswedan to run for the 2024 Presidential Candidate and cluster 1 has a negative category that does not support Anies Baswedan not advancing for the 2024 Presidential Candidate.

Keywords: Twitter; Clustering; DBSCAN; TF-IDF Vectorizer; Silhouette Score

1. INTRODUCTION

The age of information and communication technologies is currently upon us. Advances in technology have provided information and communication resources that are broad than what humans already have. The need for information and communication is no less important than the need for human clothing and food[1].

One service that provides a source of information and communication is Twitter. Twitter is a micro-blogging service that is used by millions of users to convey ideas or opinions. Users can create, publish and exchange short messages called tweets. Twitter is available on various platforms such as applications on smartphones and websites [2] making it easier for Twitter users to access it anywhere and anytime.

With information and interactions carried out on Twitter, a topic of discussion will be formed indirectly with keywords related to that topic or commonly called hashtags. Currently, there are various kinds of topics on Twitter, one of which is the topic of politics. This topic is being widely discussed considering that the 2024 Indonesian presidential election will soon be held. Many candidates have been informed that they will run for president and one of them is Anies Baswedan. Anies Baswedan is one of the topics that is hotly discussed because of Anies' proven track record, which is proven by many pollsters that have noted that Anies' chances of winning the 2024 presidential election are relatively high [3]. One of the keywords in the topic of Anies as a 2024 presidential candidate is "#AniesPresiden2024". This keyword is often used because it relates to Anies' advance as president in 2024.

Previous research that became a reference for making this final project was research conducted by [4]. The author uses the DBSCAN method to classify text data with 2,184 text data. Several previous steps were carried out, namely cleaning, eliminating data duplication, stemming, and stopwords. Then classification was carried out with DBSCAN using different Eps and MinPts parameters. The Silhouette Index was used to evaluate, and the result was 0.413 with Eps 0.1 and MinPts 10 parameters. 31 clusters were formed with the highest frequency of occurrence of the word "kpu", followed by "firdaus", "kota", "pasang", and "ayat".

There is also research conducted by [5] on the topic of the influence of content and customer engagement in the context of social media. This study aims to determine the effect of information, entertainment, remuneration, and relational content on passive and active engagement behavior of social media users. The data used is data from 12 wine brands on Facebook for 12 months. Multi-variant Linear Regression Analysis was chosen as a method to investigate the effects of content on the behavior, contribution, and engagement of consumers. The results reveal the effect that rational attraction on social media has a superior effect in facilitating the active or passive engagement of social media users whereas emotional appeal facilitates passive rather than very active engagement behavior.

Another reference in [6] reveals the structural dimensions of consumers' motives for using Instagram and to explore the relationship between the identified motivations and the main attitudinal and behavioral intention variables by using a comprehensive survey on a total of 212 Instagram users. The study concluded that Instagram users have five main social and psychological motives: social interaction, archiving, self-expression, escapism, and peeking. Research conducted by [7] discusses the topic of the 2019 presidential election by

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 741-748
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3750

establishing a tweet grouping system to distinguish topics of discussion regarding the presidential election related to the two presidential candidate pairs. This grouping system uses the DBSCAN method combined with ontology-based concept weighting. This study uses ontology-based concept weighting to apply knowledge about the hierarchical structure of topics, so that each topic that is at the same hierarchical level has equality. In research [8] discussing the 2019 election on Twitter social media, people convey positive and negative comments and even tend to "black campaigns" and hoaxes before the election is held or when the election is in progress regarding the election being held, comments on Twitter at this time cannot be determined more precisely. positive or negative direction, therefore it is necessary to carry out a sentiment analysis to determine the tendency of public opinion towards elections. Based on research that has been done before, in this study using the DBSCAN method which has the advantage of being able to detect data that does not enter into clusters will be considered as Noise data. With this Noise data, clustering will be more optimal because the data included in the cluster is data that is cleaner and more compatible between one data and another. This is what makes this research different from previous related studies. In addition, the DBSCAN method uses the Euclidean distance formula to calculate the distance between data points and the Silhouette score to evaluate the program. The topic raised in this research is also different from previous related studies because it raises the topic of Anies Baswedan as a presidential candidate for RI 2024 which is being hotly discussed on the Twitter platform. This research also discusses the type of content and user motivation to find out how much the public supports and wants to overthrow Anies Baswedan as the 2024 Republic of Indonesia presidential candidate.

Based on research that has been done before, this research was made to detect whether the type of content is positive or negative and the user's motivation in writing a tweet on the topic. The topic chosen is Anies Baswedan as a candidate for President of Indonesia in 2024. The method that will be used is the DBSCAN Algorithm because DBSCAN has the advantage of being able to detect data that is not included in the cluster and will be considered as Noise data. With this Noise data, clustering will be more optimal because the data included in the cluster is data that is cleaner and more compatible between one data and another. This research will be carried out by collecting data, performing preprocessing, inputting it into the DBSCAN algorithm to determine the clusters formed, calculating the Silhouette Score to evaluate the program and performing cluster analysis to determine the type of content and user motivation in each cluster.

2. RESEARCH METHODOLOGY

2.1 Research Steps

In this study, eight stages will be carried out, namely Crawling data to get tweet data, pre-processing to produce clean data, TF-IDF process to convert sentences into vectors, DBSCAN Clustering to get clusters and Silhouette Score calculations, Data Visualization to display data in each cluster, Cluster Analysis to determine the type of user context and motivation. For more details, it can be seen in Figure 1.

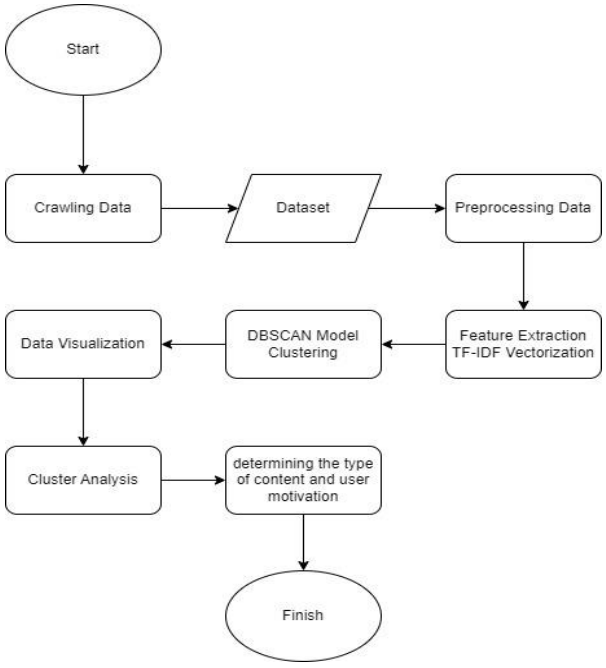

Figure 1. Research Flow

In the research flow, the first process is Crawling data to get the raw data to be used, after getting the dataset, the dataset will enter the pre-processing stage to be cleaned, then enter the TF-IDF weighting stage after

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 741-748
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3750

that it will be clustered using DBSCAN Clustering then on visualize it to be able to see the data for each cluster, after which it is analyzed to determine the type of content and motivation of the user.

2.2 Data Crawling

One way to get datasets is to do Data Crawling. Data crawling is an activity of retrieving data from a website or database [9]. One way to do data crawling is to use a python-language program combined with the snscrape library. To do crawling on twitter, use the Twitter API. Tweets were taken from about January 2022 to April 2023. The total data obtained was 49,519 tweets using the ni situ language using five hashtags or search terms, namely "AniesPresiden2024", "#AniesBaswedan2024", "#Anies2024", #AniesPresidenku", and # AniesPresident RI2024". The results of the crawling data will become the dataset used in this study. The results of crawling data can be seen in Table 1.

Table 1. Sample of Dataset

Username	Tweet
putra_kurniawan	Pak Anies Baswedan Kembali di sambut meriah dan penuh rasa persaudaraan oleh arek-arek Suroboyo semasa beliau berkunjung ke Surabaya #ItsTimeRestorasiIndonesia #NasdemNo5 #AniesPresidenku
JarnasABWBpn	Ingat! Anies dan Anis itu 2 orang yang berbeda lho… Jangan sampai salah sebut nama ya Gaesss #AniesPresidenRI2024
Fatrah_neo170	Anomali @aniesbaswedan, dirilis sejumlah Lembaga survey selalu dibawah GP dan PS, tapi setiap agenda jalan2x dimonitor dan dihadang dengan demo kecil-kecilan, oleh orang kecil yang dijanji fulus yang cukup buat makan sehari #AniesPresiden2024

2.3 Data Preprocessing

The dataset that has been crawled will be included in the preprocessing to get more optimal results. Six stages will be carried out in preprocessing, namely Cleaning Data, Case Folding Removal, Tokenizing, Data Normalization, Stop Word Removal, and Stemming. The preprocessing stages can be seen in Figure 2.

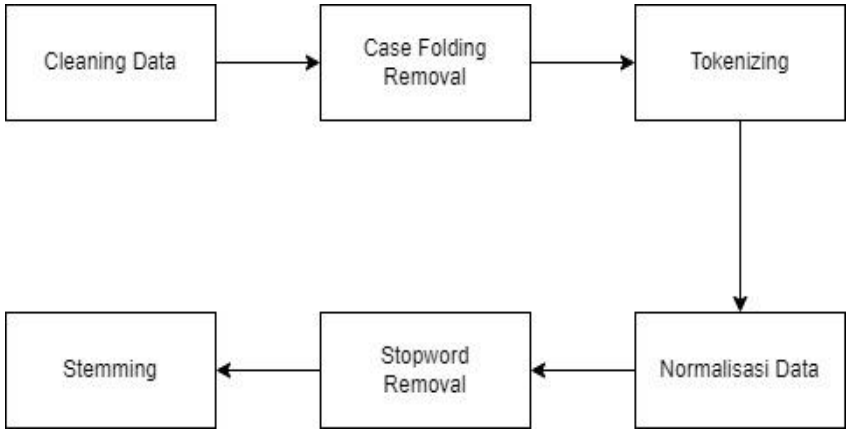

Figure 2. Data Preprocessing Steps

2.3.1 Data Cleaning

Data cleaning is a process used to remove urls, numbers, enter, tabs and symbols in sentences. This is done because when clustering the data is not used so it can be deleted.

2.3.2 Case Folding

Case folding is a process used to convert all capital letters into non-capital letters [10]. This aims to facilitate the process of Stop Word and Stemming in identifying words.

2.3.3 Tokenizing

In text mining, tokenizing is a procedure used to turn sentences into strings of words. Sentences with large dimensions will be divided into several smaller sentences and then separated into rows of words [11]. Some symbols that will be identified as delimiters are periods (.), commas (,), and spaces.

2.3.4 Data Normalization

Data normalization is the process of simplifying or changing a word into the standard form of that word according to the KKBI. The purpose of data normalization is to reduce word errors after the data crawling process [12].

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 741-748
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3750

2.3.5 Stop Word Removal

Stop word is a process carried out to eliminate words that lack or do not have the information contained in the sentence. This can improve the accuracy of the process because the data that is processed is data that has information value in the word. In Stopwords the missing words are called special words. In English, examples of words to be deleted are all, am, an, are, etc [13].

2.3.6 Stemming

Stemming is the process of mapping and decomposing various forms (variants) of a word into its basic form (stem). Stemming changes words that contain affixes to stems. The main goal of the stemming algorithm is to minimize grammatical forms and get meaningful terms from the morphological structure of the language [14]. In this study, a literature library will be used which contains Indonesian words as a reference in carrying out the Stemming process.

2.4 Feature Extraction

One of the methods used to determine the significance of words in a document is TF-IDF. The frequency of occurrence of a word in a document determines whether or not the word is significant [15]. In the TF-IDF method, feature extraction is used to determine the level of importance of words in a group of documents. To assist in the clustering process, datasets that are in the form of words must be converted into numbers so that they can be read by the program. The TF-IDF formula can be defined in the following equation (1) [16].

$$TFIDF(\mathrm{t},\mathrm{d}) = TF(\mathrm{t},\mathrm{d}) \times IDF(t) \quad (1)$$

After obtaining the TF-IDF results from the above formula, then proceed with the normalization process using the Euclidean method with equation (2).

$$v_{norm} = \frac{v}{\sqrt{v1^2 + v2^2 + .. + vn^2}} \quad (2)$$

2.5 DBSCAN Model Clustering

DBSCAN is a grouping technique or algorithm by creating regions based on related densities. DBSCAN has an advantage over other methods because this method can identify outliers and noise. Outliers or noise can be formed because items are not close to other objects [17]. Unlike the K-Means and K-Medoids algorithms, DBSCAN does not require defining the number of clusters to be formed because DBSCAN will identify disordered cluster structures using clustering techniques [18]. DBSCAN clustering has 2 parameters, namely MinPts and Epsilons. MinPts serves to determine the minimum data that can be used as initial data to determine the boundaries of a cluster and Epsilons are used to determine the distance between data in a cluster. DBSCAN uses the Euclidean Distance function to determine the distance between items. The Euclidean Distance formula can be seen in equation (3) [19]:

$$d_{ij} = \sqrt{\sum_a^p \left(x_{ia} - x_{ja}\right)^2, i = 1, \dots, n; j = 1, \dots, n.} \quad (3)$$

Where x_{ia} is the variable of object i ($i = 1, \dots, n; a = 1, \dots, p$) and d_{ij} is the Euclidean Distance value.

2.6 Sillhouette Score

The performance of the algorithms was assessed using the algorithm's silhouette score value. By measuring both intra-cluster cohesion and inter-cluster separation, Silhouette assists in determining whether allocating a data point to one cluster rather than another is the best course of action. The purpose of the cluster validation technique is to evaluate the cluster results, the results of this evaluation can be used to determine the number of clusters in the dataset. This technique provides a brief graphical representation of how well each object is located within its cluster. The Silhouette Score formula can be seen in equation (4) :

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

2.7 Data Visualization

Data visualization is used to display data with various methods, one of which is in the form of graphs or charts. One way to display visualization data is to use the Word Cloud library to collect words into an image and the Matploblib library to display the image so that it is easy to see [20].

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 741-748
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3750

2.8 Cluster Analysis

Cluster analysis was carried out to analyze the results of grouping data with the highest frequency of word occurrences in the cluster results. This stage is the last stage to find out the type of content and user motivation whether positive (support) or negative (bring down).

3. RESULT AND DISCUSSION

The dataset that has been collected by the data crawling process uses five different keywords or hashtags, namely#AniesPresiden2024, #AniesBaswedan2024, #Anies2024, #AniesPresidenku, and #AniesPresidenRI2024. And we have succeeded in obtaining data for a total of 49,519 Indonesian language tweet data. The dataset will enter into the pre-processing process to help clean up the data so that later the data entered into the clustering algorithm can produce more optimal clusters. Figure 3 is the result of the pre-processing process which is visualized in the form of an image using the word cloud library.

Figure 3. Data Visualization

After the preprocessing is done, the feature extraction process is carried out using TF-IDF to convert words into vectors so that it can make the program easier to carry out the clustering process. After that, the dataset will be entered into the DBSCAN Clustering algorithm program.

3.1 DBSCAN Clustering

In the clustering process using the DBSCAN Clustering algorithm, the clustering process will be carried out using different parameters to find the best results. In this research, an experiment will be carried out to find parameters using an Epsilons value of 0.01 and MinPts with a range from 1 to 10. To measure the evaluation value in the experiment, calculations are carried out using the silhouette score in each experiment. The results of the experiment can be seen in Table 2.

Table 2. DBSCAN Result

MinPts	Eps	Number of Clusters
1	0.1	4238
2	0.1	417
3	0.1	144
4	0.1	85
5	0.1	49
6	0.1	30
7	0.1	10
8	0.1	6
9	0.1	4
10	0.1	3

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 741-748
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3750

3.2 Silhouette Scoring

In this section, the clustering results from DBSCAN will be calculated with the values from the dataset. The calculations were carried out for a number of trials, and the results of calculating the silhouette score on the clustering results can be seen in Table 3.

Table 3. Silhouette Score Result

MinPts	Cluster	Score
1	4238	0.23
2	417	0.02
3	144	0.02
4	85	0.07
5	49	0.09
6	30	0.13
7	10	0.20
8	6	0.27
9	4	0.27
10	3	0.29

3.3 Cluster Analysis

The results in Table 3 show that the highest silhouette score is at MinPts 10 with the number of clusters formed being 3 clusters. These three clusters are labeled -1, 0, and 1. Labels with -1 are grouped as Noise. The highest frequency of words in each cluster can be seen in Table 4.

Table 4. Top Word Frequency

Label	Word	Count
-1	Anies	1950
	Nasdem	892
	Baswedan	600
	Partai	563
	Indonesia	389
0	Semangat	10
	Tumbuh	10
	Ubah	10
	Indonesia	10
1	Anies	10
	Bohong	10

From the table above it can be seen that on label 0 there are 4 words namely “Semangat”, “Tumbuh”, “Ubah”, and “Indonesia”. These words appear 10 times. On label 1 there are 2 words, namely “Anies” and “Bohong”. This result can be obtained because several users write tweets with sentences that are similar to one another, both from the same hashtags and the same content. The contents of label -1 can be categorized as Noise because the data cannot be clustered into 1 cluster because several other words have far-reaching values so they cannot be included in any cluster.

After all the steps are done, the final step is to determine the type of content and user motivation based on the type of tweets related to politics. To do this, it will take the words that appear the most in each cluster and then manually identify them by determining whether the type of content and motivation of users in each cluster is positive (supporting) or negative (bringing down). The cluster labeled -1 will be ignored because the cluster is Noise data which has very diverse information so it cannot be included in any cluster.

In the cluster labeled 0, the words that appear the most are “Semangat”, “Tumbuh”, “Ubah”, and “Indonesia” where the word indicates that this cluster has a positive type of content with user motivation, namely flattering Anies Baswedan as the 2024 presidential candidate. In the cluster labeled 1, the word that appears the most is “Anies”, and “Bohong”, this makes this cluster a negative type of content because the motivation of users to post tweets with these words is to destroy Anies Baswedan's image as the next 2024 presidential candidate. general elections. The grouping of content types and user motivation based on the words that appear the most can be seen in table 4 and table 5.

Table 5. Content Type

Top Word Frequency	Cluster	Conten Type
Semangat	0	Positive
Tumbuh		Positive
Ubah		Positive

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 741-748
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3750

Top Word Frequency	Cluster	Conten Type
Indonesia		Positive
Anies	1	Positive
Bohong		Negative

Table 6. User Motivation

Support	Bringing Down
Semangat Tumbuh Ubah	Bohong

4. CONCLUSION

The application of the DBSCAN Clustering method in determining the type of content and user motivation for a topic has proven to be optimal. This can be seen from the Silhouette Score, which is 0.29, which results in a total of 3 clusters. The clusters that are formed are Cluster -1, Cluster 0, and Cluster 1. After that, an analysis is performed on each cluster by looking at the words contained in the cluster. In connection with Cluster -1 contains Noise data so that the cluster can be ignored because the data is the result of data that does not enter into any cluster. In Cluster 0 you can see the many words that appear are “Semangat”, “Tumbuh”, “Ubah”, and “Indonesia” where these words are positive words which show that many people expect the spirit of growth and change in Indonesia with the advancement of Anies Baswedan as president of Indonesia 2024 In Cluster 1, it contains the words “Anies” and “Bohong” which show words of lack of trust so that they can be categorized as words against Anies Baswedan as President of Indonesia 2024. In the future, this research can be improved by improving the words in the Indonesian stopwords dictionary and can be combined with the Hyperparameter tuning method.

REFERENCES

[1] G. K. Jha and T. Ramakrishnudu, “User Behavior Pattern and Deeper Intention Analysis in Online Social Media,” in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 2019, pp. 1–5.
[2] A. C. Sari, R. Hartina, R. Awalia, H. Irianti, and N. Ainun, “Komunikasi dan media sosial,” *J. Messenger*, vol. 3, no. 2, p. 69, 2018.
[3] N. Hermawan, “Representasi Anies dan Ganjar pada Bursa Calon Presiden Indonesia 2024 dalam Berita Online Okezone. com,” *Syntax Lit. J. Ilm. Indones.*, vol. 6, no. 1, pp. 24–32, 2021.
[4] R. N. G. Indah *et al.*, “DBSCAN algorithm: twitter text clustering of trend topic pilkada pekanbaru,” in *Journal of physics: conference series*, 2019, vol. 1363, no. 1, p. 12001.
[5] R. Dolan, J. Conduit, C. Frethey-Bentham, J. Fahy, and S. Goodman, “Social media engagement behavior: A framework for engaging customers through social media content,” *Eur. J. Mark.*, vol. 53, no. 10, pp. 2213–2243, 2019.
[6] Y.-T. Huang and S.-F. Su, “Motives for Instagram use and topics of interest among young adults,” *Futur. internet*, vol. 10, no. 8, p. 77, 2018.
[7] A. Herdiani and I. Asror, “Klasterisasi Tweet Terkait Dengan Pemilihan Presiden 2019 Menggunakan Ontology-based Concept Weighting dan DBSCAN,” *eProceedings Eng.*, vol. 6, no. 2, 2019.
[8] I. Kurniawan and A. Susanto, “Implementasi Metode K-Means dan NaÃ ve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019,” *J. Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019.
[9] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, “Feature expansion for sentiment analysis in twitter,” in *2018 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2018, pp. 509–513.
[10] T. Mustaqim, K. Umam, and M. A. Muslim, “Twitter text mining for sentiment analysis on government’s response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm,” in *Journal of Physics: Conference Series*, 2020, vol. 1567, no. 3, p. 32024.
[11] A. Hassani, A. Iranmanesh, and N. Mansouri, “Text mining using nonnegative matrix factorization and latent semantic analysis,” *Neural Comput. Appl.*, vol. 33, pp. 13745–13766, 2021.
[12] A. M. Pravina, I. Cholisoddin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019.
[13] J. Kaur and P. K. Buttar, “A systematic review on stopword removal algorithms,” *Int. J. Futur. Revolut. Comput. Sci. Commun. Eng.*, vol. 4, no. 4, pp. 207–210, 2018.
[14] M. Haroon, “Comparative analysis of stemming algorithms for web text mining,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 10, no. 9, pp. 20–25, 2018.
[15] M. Z. Fauzi and A. Abdullah, “Clustering of Public Opinion on Natural Disasters in Indonesia Using

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 741-748
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3750

DBSCAN and K-Medoids Algorithms," in *Journal of Physics: Conference Series*, 2021, vol. 1783, no. 1, p. 12016.

[16] R. Novia, S. S. Prasetyowati, and Y. Sibaroni, "Identify User Behavior Based on The Type of Tweet on Twitter Platform Using Gaussian Mixture Model Clustering," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 502–506, 2022.

[17] S. F. Galán, "Comparative evaluation of region query strategies for DBSCAN clustering," *Inf. Sci. (Ny).*, vol. 502, pp. 76–90, 2019.

[18] D. Deng, "DBSCAN clustering algorithm based on density," in *2020 7th international forum on electrical engineering and automation (IFEEA)*, 2020, pp. 949–953.

[19] W. Gunawan, "Implementasi Algoritma DBScan dalam Pemngambilan Data Menggunakan Scatterplot," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, pp. 91–98, 2021.

[20] D. W. Laraswati and A. Fauzan, "Implementasi Metode Runtun Waktu dalam Pemodelan Total Harga Alat Kedokteran dan Kesehatan," *Jambura J. Probab. Stat.*, vol. 4, no. 1, pp. 30–38, 2023.