Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 749-756
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3751

Clustering Content Types and User Roles Based on Tweet Text Using K-Medoids Partitioning Based

Raisa Benaya[1], Yuliant Sibaroni[2], Aditya Firman Ihsan[3]

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

Abstract−In this modern era, the spread of information occurs rapidly through social media. One of the channels for disseminating information is through the Twitter platform. Many Twitter users respond to existing content with positive, negative and neutral responses. One of the hot content to respond to is political content. This content is currently being discussed considering the approaching election of the 2024 Presidential Candidate of the Republic of Indonesia. One of the candidate pairs discussed was Anies Baswedan. With so many responses from Twitter users, it will be difficult to track whether users support Anies Baswedan to run as a presidential candidate due to the large number of responses. This study aims to determine the response of twitter users to the advancement of Anies Baswedan as a presidential candidate. The method used in this study is the K-Medoids Partitioning-Based algorithm based on twitter user text. This algorithm was chosen because it is easy to implement considering the basis of K-Medoids development is the K-Means algorithm but the K-Medoids algorithm can overcome the shortcomings of the K-Means algorithm which is sensitive to outliners. The evaluation will be done using Silhouette Score which produces a value of 0.35 with the number of clusters is 2. Then an analysis of each cluster is carried out by looking at the words in the cluster. As a result, from the two clusters formed, both clusters contain positive content and show that Twitter users support Anies Baswedan to run as a 2024 presidential candidate.

Keywords: Twitter; Clustering; K-Medoids; TF-IDF Vectorizer; Silhouette Score

1. INTRODUCTION

Today, social media is used to obtain various information and also to interact with others. Even though it is in a different area or even country, interaction can still be done by utilizing this technology [1]. There is a lot of information that can be found on social media, both beneficial and detrimental. Users can engage in various activities such as sending messages, uploading status updates, sharing photos, adding friends, and more. Each user exhibits different behavior, which can be observed from the amount of activity they do on social media, one of the platforms is Twitter.

Twitter allows individuals to publish messages to express their interests, favorites, opinions, and sentiments on various topics and issues they encounter in everyday life [2]. Sometimes, users create messages or commonly called tweets that contain various types of information. Shared tweets typically consist of positive (praise) or negative (criticism) feedback regarding specific information circulating [3]. One of the topics discussed was political information. Tweets circulating in the political world today revolve around presidential candidates running for the 2024 presidential election.

This topic was chosen because the 2024 presidential election in Indonesia is approaching, and Anies Baswedan has emerged as one of the selected candidates. Anies is known as a leader who tends to practice a democratic leadership style. He always involves citizens in every change effort and is happy to receive input and aspirations from the community [4]. During his tenure as Governor of Jakarta, Anies successfully handled floods in Jakarta which occurred six times from the beginning of 2020 to the end of February 2020. His leadership and ability are appreciated, especially because he actively participates in flood management in the field and is considered a responsive leader in handling situations [5]. This caused sympathy and even empathy for his actions, making several political parties interested in holding Anies as a presidential candidate in the 2024 election, considering himself as a figure worthy of becoming President in 2024.

In previous research discussing social media user clustering using Hierarchical Clustering and Non-Hierarchical algorithms, social media user data was collected and processed using both algorithms. The results showed that this approach successfully grouped social media users based on the same pattern, which can help develop better marketing strategies on social media [6]. The clustering method is used to search and group data based on the similarity of characteristics between one data point and another data point [7]. Common grouping methods include K-Means, Hierarchical Clustering, and DBSCAN.

In 2020 Kharisma Jevi Shafira [8] conducted research on the use of K-Medoids clustering method to determine public opinion segmentation on Twitter. Tweets that contain public opinion through labeling, preprocessing, and clustering. The resulting clusters were evaluated using the Silhouette Coefficient, which resulted in a score of 0.19. This clustering method successfully determines public opinion with an accuracy rate of 80%.

In 2018, Jaka [9] conducted research on the use of K-Means algorithms used to group and analyze Twitter users' opinions about illicit alcohol cases. The results showed that the most optimally formed clusters were three

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 749-756
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3751

clusters based on a Dunn Index value of 0.8312. The case is still centered on the figure of oil dealers, officials, and victims.

In 2021, Mustakim [10] conducted research on grouping public opinion about natural disasters in Indonesia on Twitter by comparing the use of DBSCAN and K-Medoids algorithms. Grouping public opinion based on the similarity of topics and sentiments shows that these two algorithms have different advantages. The K-Medoids algorithm produces clearer and cohesive clusters, whereas the DBSCAN algorithm is better at identifying different public opinions than the majority, known as outliers.

In 2023 Syamsul Bahri [11] conducted a study focusing on grouping students at risk of dropping out of school using the K-Medoids method. In this study, students were grouped into clusters with similar characteristics to identify those who needed more attention to prevent dropping out. Cluster results from 389 data points show three attributes with different value ranges between clusters, namely GPA, Semester 1 GPA, and Scholarship Status. These three attributes characterize the differentiators between clusters.

Based on the explanation above, in this study we will conduct research on the hashtag "#AniesPresiden2024" as a keyword that appears on Twitter with various responses, such as positive (praise) or negative (criticism) tweets, which ultimately become trending topics. Based on this, we may collect user behavior data based on the type of content they respond to using the K-Medoids Partition-Based algorithm. The dataset will be taken from the process of crawling data with the keyword "#AniesPresiden2024", after the dataset is obtained it will continue with the preprocessing process to equalize the format and sentence structure in the dataset. This preprocessing stage is critical to improve prediction accuracy and reduce computational time on the system before implementing clustering methods [12]. Furthermore, the dataset will be entered into the K-Medoids algorithm for the clustering process. Silhouette Score is used to help the optimal cluster search process. When the cluster has been obtained, an analysis process will be carried out to determine the type of content and the role of Twitter users on the topic of Anies Bawsedan running as a 2024 presidential candidate.

2. RESEARCH METHODOLOGY

2.1 Research Stages

In this study, the flowchart in Figure 1 illustrates the stages of the content type and user role clustering process using K-Medoids. Some of the stages in this system include data crawling, preprocessing, TF-IDF, model building, and evaluation of results.

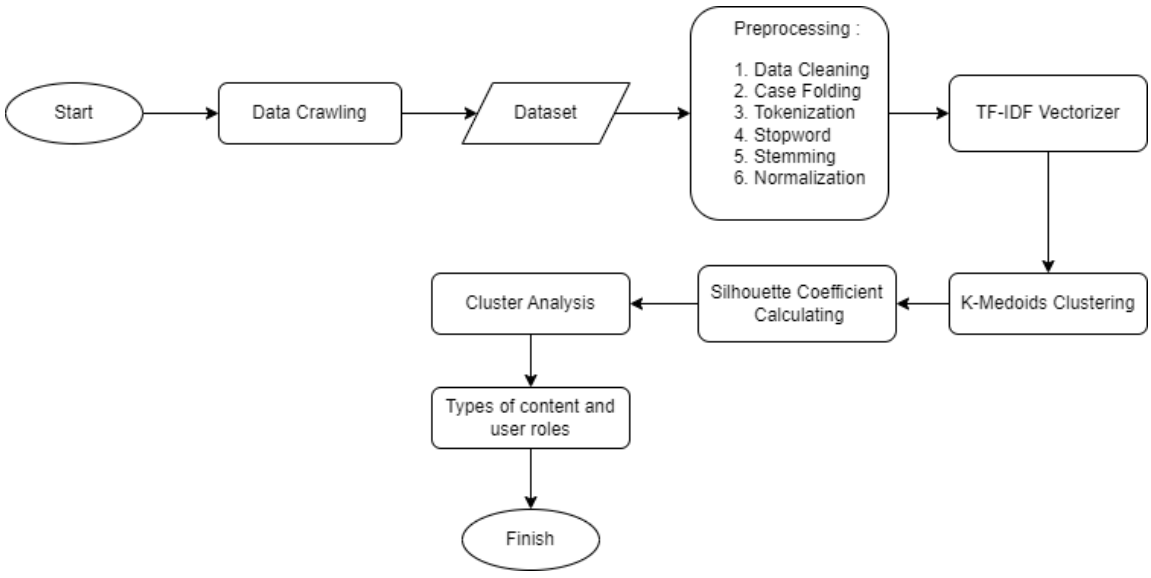

Figure 1. Research Flow

The design of this system starts from the stage of collecting datasets by crawling data on the Twitter platform. The data will go through the preprocessing stage where the data will be carried out data cleaning, case folding.tokenization, stopword removal, and stemming to equalize the format and sentence structure in the dataset. After that, silhouette score calculations are carried out to help the optimal cluster search process. When the cluster has been obtained, an analysis process will be carried out to determine the type of content and the role of Twitter users on the topic of Anies Bawsedan running as a 2024 presidential candidate.

2.2 Data Crawling

Crawling data is the process of retrieving data from a website using a specific programming language.[13]. Crawling data on Twitter refers to the procedure of fetching or acquiring data from the Twitter server using the

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 749-756
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3751

Twitter API, either in the form of user data or tweet data. This collected Twitter data serves as the basis for this research. The Twitter data is categorized into two sets: training data and testing data [14].

2.3 Data Preprocessing

Data preprocessing involves efficiently removing unnecessary attributes in a data set to improve clustering performance [11]. This stage has a significant influence on the quality of data resulting from data crawling. At this stage, the initially unstructured data is converted into structured data, which allows further analysis. All data will go through several stages, including:

2.3.1 Cleaning Data

This stage removes irrelevant attributes and removes null values in columns that are not needed for subsequent calculations [15]. Text that will be removed includes links or links, punctuation, unique symbols, and numbers that will be replaced with spaces.

2.3.2 Case Folding

This stage is carried out with the aim of facilitating text comparison in data processing. In this step, all text that was originally capital letters is converted to lowercase [16].

2.3.3 Tokenizing

This stage aims to examine the words in a sentence by breaking the text into tokens that include words, phrases, or other important things [17].

2.3.4 Stopword

In this step, the omission of words that have no informational value or do not have a negative or positive tendency is carried out. These words if left unchecked can interfere with the clustering process because the processed words are words that have information. Examples of stopword results are presented in Table 1.

Table 1. Stopword Result

Transformation Results
[yang, di, itu, tak, saja, mau,dulu, dengan, untuk, ini, pada, juga, ada, atau, saat, ia, adalah, dan, dari, akan, ke, lalu, secara, sampai, serta, Kembali, bila, terus, cukup, bahkan, perlu, maka, selalu, maupun, yaitu]

2.3.5 Stemming

The stemming stage is done to transform the word into its basic form. This process uses the help of a literary library as the basis for a dictionary of Indonesian basic words. Examples of stemming results are presented in Table 2.

Table 2. Stemming Result

The Original Word	Transformation Results
[penggunaan, menggunakan, digunakan]	guna
[pelajaran, mengajarkan, diajarkan]	ajar
[pemukulan, dipukul, memukul]	pukul
[kesamaan, disamakan, menyamakan]	sama
[menghitung, dihitung, penghitungan]	hitung

2.3.6 Normalization

At the normalization stage, non-standard or non-standard words are changed to standard words. Examples of normalization results are presented in Table 3.

Table 3. Normalization Result

The Original Word	Transformation Results
JATUH	jatuh
‘Mari makan…..’	‘mari makan’
k@sihan	kasihan
dgn	dengan
kupukupu	kupu-kupu

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 749-756
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3751

2.4 TF-IDF Vectorizer

After the preprocessing stage, the next step is to give weight to each word. One commonly used weighting method is the Term Frequency-Inverse Document Frequency (TF-IDF). This method involves assigning weight or value to each word in a text document. TF-IDF provides information about the importance of words in each document, measured through numerical values between 0 and 1 [18]. TF-IDF can be formulated in equation 1.

$$W_{ki} = tf_{ki} \times \log\left(\frac{N}{N_k}\right) \quad (1)$$

In the formula, Wki represents the weight of the word k in document i. tfki is the frequency of the word k in document i. N is the total number of documents used, and nk is the number of documents containing the word k.

2.5 K-Medoids Clustering

K-Medoids, also known as Partitioning Around Medoids (PAM) or K-Medoids, is a clustering algorithm that is a variation of the K-Means algorithm. The K-Medoids method uses objects in a set of objects as a representation of cluster data, where the distance between each pair of objects is calculated using the Euclidean distance in Equation 2.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(\boldsymbol{x_i} - \boldsymbol{y_i})} \quad (2)$$

The object chosen as a representation of a cluster is called a medoid. Cluster formation is done by calculating the similarity between the medoid and other non-medoid objects [19]. The weakness of K-Means that is prone to outliers is considered to be surmountable by K-Medoids. The medoid object is located in the center of the cluster, making it stronger against outliers. Clusters are formed by considering the proximity between medoid and non-medoid objects.

2.6 Silhouette Score Calculating

The Silhouette coefficient is a way to evaluate how good the quality of objects in a cluster is. The goal is to assess the extent to which objects correspond to their respective clusters [20]. The quality of the cluster can be evaluated using the silhouette coefficient in Equation 3.

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (3)$$

Description :
$d = distance$
$a(i) = lowest\ dissimimilarity$
$b(i) = lowest\ dissimilarity$
$s(i) = silhouette\ coefficient$
$x = value\ of\ the\ i-th\ object\ in\ the\ x-th\ variable$
$y = value\ of\ the\ i-th\ object\ in\ the\ y-th\ object$

The value obtained from the silhouette coefficient method ranges from -1 to 1. To find out whether the value is valid or not can be seen in Table 7 which presents the criteria for measuring the validity of the silhouette coefficient method.

Table 4. Rating Scale of Silhouette Score

Silhouette Coefficient	Interpretasi
0,7 < SC ≤ 1.0	Strong Structure
0.5 < SC ≤ 0.7	Medium Stucture
0.25 < SC ≤ 0.5	Weak Structure
SC ≤ 0.25	No Structure

2.7 Data Visualization

The use of data visualization aims to describe data from various dimensions in the form of graphs or diagrams. This is done with the aim of making it easier for other users to understand information related to the quantity, relationship, and relationship between the data presented. The types of data visualizations used are word cloud and matplotlib to display data visualizations using the Python programming language.

2.8 Cluster Analysis

This stage aims to see the results of data clustering that has been done in the previous stage. This stage examines the results of data clustering based on the frequency of words that appear most frequently in each cluster and is the final stage in determining the type of content and user roles.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 749-756
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3751

3. RESULT AND DISCUSSION

3.1 Analysis of Test Result

In this study, a total of 49,159 tweet data was collected using five keywords, namely #Anies2024, #AniesPresidenRI2024, #AniesPresidenku, #AniesPresiden2024, #AniesBaswedan2024. The data collection method uses crawling with the python programming language and SNSCrape library, so irrelevant words will not be processed further. Therefore, the preprocessing stage is important to clean and optimize the data to be processed. After this stage, proceed with the TF-IDF weighting stage, where data in the form of words is converted into numerical values by calculating the weights. Next, a clustering process will be carried out using K-Medoids. As a result of the visualization, the preprocessed data is displayed in the form of a word cloud as shown in Figure 2.

In this study, the dataset obtained came from tweets made by Twitter users using python-language data crawling and SNSCrape libraries. There are five keywords entered to collect data, namely #Anies2024, #AniesPresidenRI2024, #AniesPresidenku, #AniesPresiden2024, #AniesBaswedan2024. The crawling process is carried out by retrieving data from January 2022 to April 2023. The total data collected amounted to 49,159 Indonesian-language tweet data after the crawling process. After the dataset is obtained, the next process is to pre-process to clean and optimize the data to be processed. The result of crawling data can be seen in the table 5.

Table 5. Sample of Dataset

Username	Tweet
@PutraAlamsyahJ1	Seneng banget colong star sih pak!!! #aniesbaswedan #AniesPresidenRI2024 #AniesPresiden2024 #Pilpres2024
@ onomsalam2pas	#AniesPresiden2024 #kaltim # #indonesiaBerkeadilansosial untuk rakyat nkri
@the3brothers2	bang anies presidenkuuu...salam dari warga purwokertooo...jawa tenngahhh
@ SalehudinSaleh3	@DennyBrata1 @Candra4smara MAJU NEGERINYA SEJAHTERA RAKYATNYA #AniesPresiden2024
@Neng_Anyar	Kita bersyukur satu per satu dukungan resmi diputuskan untuk @aniesbaswedan akan hadirnya perubahan
@Cahyokt4	Yang begini begini nih harusnya ga terjadi,, masa iya ada proyek yang ga selesai dalam masa kepemimpinan satu periode.. #aniesbaswedan #aniesingksrjanji
@AsepSeptiyana_	Nah kan masyarakat juga tau pak mana yang diem-diem kampanye, apa lagi di kampus lagi haduhhh ada-ada saja ðŸ˜… #aniesbaswedan #aniespresiden2024
@mulyana0510	@GeiszChalifah Alhamdulillah..kesadaran bersama menuju perubahan...harus diatas ego kepentingan masing2 partai..semoga dimudahkan kemenangan koalisi perubahan #AniesPresiden2024

The results of pre-processing can be seen in figure 2 using the wordcloud and matplotlib libraries to help display data visualization.

Figure 2. Data Visualization

The next step is to convert sentences into vectors using TF-IDF weighting. The results of TF-IDF will be entered into the K-Medoids clustering algorithm program. To help the optimal cluster retrieval process, silhouette score assistance is used. The cluster with the highest silhouette score will be selected as the most optimal cluster.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 749-756
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3751

3.2 Coefficient Calculation

After carrying out the K-Medoids Clustering and Silhouette Score process, experimental results were obtained which can be seen in table 6.

Table 6. Results of the K-Medoids

No	Number of Cluster	Silhouette Score
1	2	0.35
2	3	0.23
3	4	0.12
4	5	0.13
5	6	0.12
6	7	0.13
7	8	0.14
8	9	0.15

Experiment

The experiment carried out was to test K values from a range of two to 9. This K value is considered a cluster. It can be seen that the most optimal value is K=2 with a Silhouette Score of 0.35. This grouped the interpretation that the structure has a weak value of structure. But this is better when compared to other clusters that have values below 0.25 which are considered to have no structure. Furthermore, with these results, an analysis of each cluster will be carried out.The following is displayed the frequency of occurrence of words that appear more than 20 words in clusters 0 and 1 are presented in table 7.

Table 7. Top Frequent Word

Label 0 / Cluster 1		Label 1 / Cluster 2	
Word	Frequency	Word	Frequency
Sehat	342	Anies	39
Bahas	162	Partai	38
RUU	159	Nasdem	36
Sambut	147	Baswedan	31
Tugas	87	Dukung	29

The highest word frequency in cluster 0 was the word 'sehat' which appeared 39 times, and the frequency of the word . The second highest is the word 'bahas' which amounts to 38 words. While in cluster 1, word frequency
the highest is the word 'Anies' with the second highest number of words and the second highest frequency is the word 'sambut'.

3.3 Cluster Analysis

The next step is to determine the content type and user role based on the tweets associated with the cluster that were obtained in the previous stage. To determine the type of content and user roles, manual analysis will be carried out using Top Frequent Words in each cluster.

Related to content types and user roles, in cluster 1 the most popping words are "Sehat", "Bahas", "RUU", "Sambut", and "Tugas" which can be seen in table 10. With the words "Sehat", "Sambut" and "Tugas" indicate that the user conveys positive words. In cluster 2, the most frequently appearing words are "Anies", "Partai", "Nasdem", "Baswedan", and "Dukung" which can be seen in table 11 showing the emergence of a positive word, namely "support" so that cluster 2 can be said to be a cluster with positive content. Based on the results above, it can be concluded that twitter users with all five keywords support Anies Baswedan to run for the 2024 presidential candidate. This can be proven by the emergence of positive words in each cluster. Also the grouping of content types and user roles based on the word that appear the most can be seen in table 8 and 9.

Table 8. Content Type

Top Frequent Word	Cluster	Content Type
Sehat	1	Positive
Bahas		Positive
RUU		Positive
Sambut		Positive
Tugas		Positive
Anies	2	Positive
Partai		Positive
Nasdem		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 749-756
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3751

Top Frequent Word	Cluster	Content Type
Baswedan		Positive
Dukung		Positive

Table 9. User Roles

Support	Bringing Down
Bahas	Positive
Sambut	Positive
Tugas	Positive
Dukung	Positive

4. CONCLUSION

The results of this study show that many Twitter users are tweeting the topic of Anies Baswedan running as a candidate for the 2024 President of Indonesia. This can be proven by taking five keywords, namely #Anies2024, #AniesPresidenRI2024, #AniesPresidenku, #AniesPresiden2024, #AniesBaswedan2024 and can generate a number of tweets as many as 49,159 tweet data in Indonesian. Furthermore, a clustering process was carried out using K-Medois and silhouette score to determine the cluster. The best result is the K value or the best number of clusters 2 with a value of 0.35. From each cluster, a manual analysis was carried out by looking at the highest number of word occurrences in each cluster with the results that cluster 1 and cluster 2 had positive content and supported Anies Baswedan to advance in the 2024 presidential candidate. This proves that the K-Medoids algorithm has worked well because each object in each cluster has good quality, where each object has been grouped based on a high degree of similarity. For future research, it is recommended to use other clustering methods such as the BIRCH Algorithm to achieve better cluster quality than previous studies.

REFERENCES

[1] Z. Ardi and S. A. Putri, "The analysis of the social media impact on the millennial generation behavior and social interactions," *Southeast Asian J. Technol. Sci.*, vol. 1, no. 2, pp. 70–77, 2020.

[2] A. S. M. Alharbi and E. de Doncker, "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information," *Cogn. Syst. Res.*, vol. 54, pp. 50–61, 2019, doi: 10.1016/j.cogsys.2018.10.001.

[3] S. N. Fikriyah and Y. Sibaroni, "Identify User Behavior based on Tweet Type on twitter Platform using Mean Shift Clustering," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1396, 2022, doi: 10.30865/mib.v6i3.4329.

[4] B. A. Prasetyo and U. M. Yogyakarta, "Analisis Gaya Kepemimpinan Anies Baswedan," no. June, pp. 0–17, 2021.

[5] K. Anies Baswedan and M. Putri Ramadhani, "KEPEMIMPINAN ANIES BASWEDAN," 2021.

[6] Z. Alamtaha, I. Djakaria, N. I. Yahya, J. Matematika, and F. Mipa, "Implementasi Algoritma Hierarchical Clustering dan Non-Hierarchical Clustering untuk Pengelompokkan Pengguna Media Sosial," *Estimasi J. Stat. Its Appl.*, vol. 4, no. 1, pp. 2721–379, 2023, doi: 10.20956/ejsa.vi.24830.

[7] A. BASTIAN, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *J. Sist. Inf.*, vol. 14, no. 1, pp. 28–34, 2018, doi: 10.21609/jsi.v14i1.566.

[8] K. J. S. Sepyanto, Y. H. Chrisnanto, and F. R. Umbara, "Sistem Segmentasi Program Talk Show Berdasarkan Media Sosial Twitter Menggunakan Metode K-Medoids Clustering," *Pros. SISFOTEK*, pp. 342–347, 2020.

[9] J. A. Pratama, N. Sunengsih, and M. Suherman, "Analisis Klaster Pada Dokumen Teks Opini Pengguna Twitter Terhadap Kasus Miras Oplosan Menggunakan Metode K-Means," *J. Stat.*, vol. 6, no. 1, pp. 49–55, 2018.

[10] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, "Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, 2021, doi: 10.1088/1742-6596/1783/1/012016.

[11] S. Bahri, D. Marisa Midyanti, and P. Korespondensi, "Penerapan Metode K-Medoids Untuk Pengelompokan Mahasiswa Berpotensi Drop Out Application of K-Medoids Method for Dropout Potential Student Grouping," vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.2023106643.

[12] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, "Effective text data preprocessing technique for sentiment analysis in social media data," *Proc. 2019 11th Int. Conf. Knowl. Syst. Eng. KSE 2019*, pp. 1–8, 2019, doi: 10.1109/KSE.2019.8919368.

[13] D. Atika, Styawati, and A. Ari Aldino, "Term Frequency-Inverse Document Frequency Support Vector Machine Untuk Analisis Sentimen Opini Masyarakat Terhadap Tekanan Mental Pada Media Sosial Twitter," *J. Teknol. dan Sist. Inf.*, vol. 3, no. 4, p. page-page, 2022.

[14] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, "Feature expansion for sentiment analysis in twitter,"

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 4, August 2023, Page 749-756
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i4.3751

Int. Conf. Electr. Eng. Comput. Sci. Informatics, vol. 2018-Octob, pp. 509–513, 2018, doi: 10.1109/EECSI.2018.8752851.

[15] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jipi.v6i2.2008.

[16] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019, doi: 10.30864/eksplora.v9i1.254.

[17] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 2, no. 2, pp. 91–97, 2021, doi: 10.37859/coscitech.v2i2.2851.

[18] S. Ramadhani, D. Azzahra, and T. Z, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 24–33, 2022, doi: 10.31849/digitalzone.v13i1.9292.

[19] N. Pulungan, S. Suhada, and D. Suhendro, "Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 329–334, 2019, doi: 10.30865/komik.v3i1.1609.

[20] S. Nurlaela, A. Primajaya, and T. N. Padilah, "Algoritma K-Medoids Untuk Clustering Penyakit Maag Di Kabupaten Karawang," *I N F O R M a T I K a*, vol. 12, no. 2, p. 56, 2020, doi: 10.36723/juri.v12i2.234.