SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2023
DOI : https://doi.org/10.33395/sinkron.v8i3.12616
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Analysis Content Type and Emotion of the Presidential Election Users Tweets using Agglomerative Hierarchical Clustering

Cika Carissa Sujadi[1)*], Yuliant Sibaroni[2)], Aditya Firman Ihsan[3)]
[1,2,3)]Telkom University, Indonesia
[1)]cikacarissa@student.telkomunivesity.ac.id, [2)] yuliant@telkomuniversity.ac.id,
[3)]adityaihsan@telkomuniversity.ac.id,

Submitted : Jun 29, 2023 | **Accepted** : Jul 6, 2023 | **Published** : Jul 7, 2023

Abstract: Over the past few years, social media has become essential for getting up-to-date information and interacting online. During presidential elections in Indonesia, Twitter has grown as a crucial platform for expressing opinions and sharing information. This study focuses on analyzing the content types and emotions of tweets related to Anies Baswedan, one of the presidential candidates. The results show a variety of discussions, including support, criticism, and discussion of policies for the 2024 presidential candidate. Clustering enables meaningful information extraction from vast Twitter data. Data were clustered using Agglomerative Hierarchical Clustering, which resulted in the identification of 10 clusters. With 4 clusters containing opinion content and 6 clusters containing information content. In addition, 6 clusters reflect excitement, 3 reflect expectations, and 1 reflect doubt. This research provides insights into the Twitter conversation around the 2024 presidential election, providing an understanding of content and emotions expressed by users.

Keywords: Twitter; Emotion; Content; Presidential Election; Agglomerative Hierarchical Clustering

INTRODUCTION

In recent years, social media has become an important platform for getting new information and also for exchanging information (Singh et al., 2020). The most frequently used social media platform is Twitter. Twitter is used as a place to discuss, comment, and provide information online. One of the topics often discussed on the Twitter platform is political issues. Therefore, Twitter has become a news source for public political discussions, which allows and provide text-based comments (Posegga & Jungherr, 2019). In addition, it is generally recognized that Twitter has great potential to investigate social and political behavior (Barberá et al., 2016).

Political issues are an interesting topic to discuss, especially with President Jokowi Dodo marking the end of his term in office. These elections can lead to pro and con opinions toward candidates trying to run for office, so there will be positive, neutral, and negative comments. Based on the research, positive comments have a longer reach than neutral and negative comments, but neutral and negative comments spread faster (Kušen et al., 2019). When making comments, social media users usually express their emotions related to the nominated candidate, such as sadness, surprise, fear, anger, happiness, and trust (Valle-Cruz et al., 2021). In addition, emotion processing is also an important factor in detecting user behavior, especially in the context of political movements.

The use of social media platforms is strongly influenced by the type of content supported by the platform (Singh et al., 2020). In the context of politics, political tweet text can be classified into two classes, namely opinion (subjective) and information (objective). Analyzing the difference can provide an understanding of the effect of using content on political issues (Chen et al., 2012). By analyzing the role and influence of these two types of content, we can understand more about how opinion and information play a role in political discussion on social media. In this context, Twitter is the main platform for voicing opinions and emotions related to being chosen candidate.

Unsupervised learning is performed to process data without having any target information or desired output. The main goal of unsupervised learning is to discover patterns, structures, or unseen information in data without labels or targets (Abdulhafedh, 2021). Agglomerative hierarchical clustering is an unsupervised learning method

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2023
DOI : https://doi.org/10.33395/sinkron.v8i3.12616
e-ISSN : 2541-2019
p-ISSN : 2541-044X

for clustering data [8]. This method is a bottom-up approach where at the earliest level the data points are treated as a single individual [9]. The major contribution of this method in twitter data is to iteratively divide and merge clusters to produce more organized categorized (Devika et al., 2018).

This study aims to classify the types of content and emotions conveyed in several hashtags about Anies Baswedan. The data used in this research consist of 12,500 Indonesian language data. The initial stage of data processing will be data preprocessing and then the use of TF-IDF Vectorizer to convert text into numerical representations that can be used in the analysis. Next, it will be clustered using the Agglomerative Hierarchical Clustering algorithm and using silhouette score to select the best cluster to perform content type and user emotion analysis.

LITERATURE REVIEW

Several studies have been conducted in the field of political behavior and characteristics of social media users. Research (NM, 2016) explores the influence of political behavior and the factors associated with social media use. Meanwhile, research (Sinha et al., 2020) proposed a method to characterize the behavior of corporate social media users. In the study, it was found that users in different groups have different topic interests but still have similarities within the group, known as homophilic. Research (FRHAN, 2017) also contributed with a new approach called Event WebClickviz, which utilizes agglomerative hierarchical clustering and TF-IDF feature extraction for visualization and analysis of detected event behavior. In research (Dirjen et al., 2017) Evang Mailoa analyzed tweet data that led to twitwar containing the hashtag "#4niesKingOfDrama" the results obtained were among the 10 main actors with the highest Degree Centrality value, there were several buzzer accounts. As a suggestion for further development, researchers suggest using a hierarchical clustering approach to see the topics discussed in each user group.

There have been several studies related to the clustering of Twitter users on political issues. One of them is research (Irawan et al., n.d.) which compares hierarchical clustering and K-Means methods to analyze the grouping of Twitter users based on their responses to political issues. Research (Zahrotun, 2015) conducted a clustering of the number of Trans Jogja passengers with the K-Means clustering method and Agglomerative Hierarchical Clustering. In addition, research (Yusup & Maharani, 2021) utilized Agglomerative Hierarchical Clustering to predict user personality based on social media content, with an accuracy rate of 20.1% and an average silhouette score of -0.23. Research (Belcastro et al., 2022) sought insight into the 2022 US presidential election with topic discovery, opinion mining and emotion analysis techniques on social media data. Research (Wulandari et al., 2023) conducted a comparison of K-Means and Agglomerative Nesting clustering methods with the results showing that agglomerative nesting is the best method for clustering for Twitter digital marketing with complete linkage with K = 2 and Silhouette Score 0.75442895. Research (Vijaya et al., 2019) has implemented agglomerative nesting on real time shopping data and compared the results of different metrics such as ward, single linkage, complete linkage and analyzed how the results vary. Research (Zahrotun et al., 2023) compared Analytic Hierarchical Clustering (AHC) with K-Medoids, showing that the AHC approach has a better Silhouette. Using this methodology, University X was able to cluster four departments with two achievements on schedule.

METHOD

This research was conducted by following the stages shown in the flowchart in Figure 1. The following is a flowchart of system design for clustering content types and emotional scope based on tweet text with Agglomerative Hierarchical Clustering.

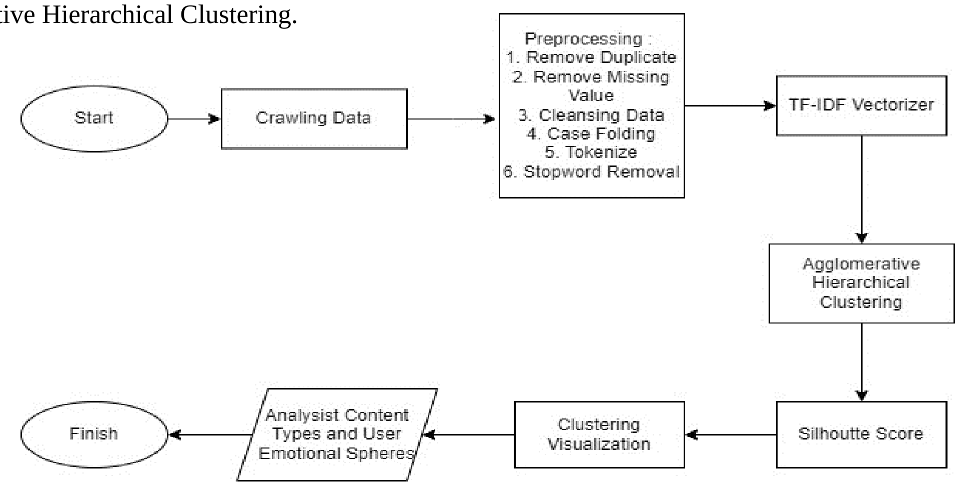

Figure 1. Flowchart System Design

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2023
DOI : https://doi.org/10.33395/sinkron.v8i3.12616
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In the research listed in Figure 1, several stages will be carried out, starting with collecting data using crawling techniques. Next, the collected data will go through a preprocessing stage, including data cleansing, case folding, tokenization, and stopword removal. After that, the TF-IDF Vectorizer method will be used to convert the text into a vector representation. Then, clustering is performed using the Agglomerative Hierarchical Clustering method, with evaluation using Silhouette Score. Visualization of the clustering is also done to gain a better understanding. Finally, the content type and user emotion scope of the clustering results were analyzed.

Crawling Data

To collect the data, this study used a Python module called snscrape. The data collected came from the Twitter platform using various hashtags related to the presidential candidate, Anies Baswedan. Each collected data consists of the Twitter user's username and the content of the related tweets.

Table 1. Results of data crawling

Username	Tweets
medcom_id	Bakal calon presiden (capres) Anies Baswedan ikut memimpin gerbong menjaring partai politik (parpol) lain untuk gabung di Koalisi Perubahan untuk Persatuan. Anies akan membuka ruang komunikasi dengan parpol lain. #AniesBaswedan https://t.co/ZKEFRjdB87

Preprocessing

Preprocessing is the first step in data processing that aims to clean and prepare raw data before further analysis. The goal is to remove distractions and errors in the data and ensure that the data is ready to be used properly. By doing preprocessing, the data will be organized in a more structured way, become cleaner, and be ready to be processed efficiently and accurately in the next stage of analysis. The following are the stages performed in Preprocessing:

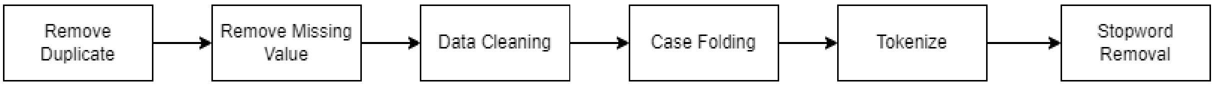

Figure 2. Preprocessing Step

A. Remove Duplicate
This stage is carried out with the aim of avoiding data incorrectness that causes results that are not maximized. The existence of duplicate data causes inconsistent results, so duplicate data must be cleaned.
B. Remove Missing Value
This process aims to remove missing values to ensure the cleanliness and quality of the data to be used in the analysis process to ensure the accuracy and validity of the analysis results to be carried out.
C. Data Cleaning
The purpose of this data cleansing process is to ensure the cleanliness and consistency of the data. During this stage, various actions are performed, such as eliminating punctuation, marks, removing number, link, emoticon. Any excessive spaces in the text are also addressed and removal as part of the cleansing process.
D. Case Folding
Case folding in preprocessing aims to convert the letters in the text into a uniform form, such as lowercase, to facilitate text analysis. This process helps to reduce variations in the form of the same word and prevents errors in matching words when processing text.
E. Tokenize
Tokenize in preprocessing aims to divide the text into smaller units called tokens. This process helps simplify text processing by separating words into separate entities so that it can make it easier to understand and manipulate text in a more structured and effective manner.
F. Stop word Removal
Stop word removal in preprocessing aims to remove common words that have no significant information value in the text. By removing stop words, the focus of analysis can be placed on more relevant and meaningful key words in the text, thus enabling more effective and accurate processing. The stop word dictionary utilized in this study was sourced from the Kaggle platform

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2023
DOI : https://doi.org/10.33395/sinkron.v8i3.12616
e-ISSN : 2541-2019
p-ISSN : 2541-044X

TF-IDF Vectorizer

TF-IDF (Term Frequency-Inverse Document Frequency) Vectorizer aims to convert text into a numerical representation that takes into account the frequency of the data in the document and the occurrence of the word in the entire document collection. Term Frequency (TF) is used to assess the importance of the occurrence of words in the text. The more often the word appears, the higher the weight or value given. Inverse Document Frequency (IDF) is used to balance the assessment of words based on how common or rare they are in the entire document. Words that appear frequently in many documents will have a lower IDF value. Here is the formula for TF-IDF (Nellie et al., n.d.).

$$W(_{i,n}) = TF(_{i,n}) * IDF(_{i,n}) \quad (1)$$

Description:

W = weight of i-th data
n = nth word of the keyword
i = i-th data

In this TF-IDF Vectorizer, N-Gram is added. The purpose is to increase the effectiveness and optimal results in the process of extracting words, as well as prevent errors in the calculation (Mardianti et al., 2018).

This makes it possible to notice important differences between words in documents that can help in clustering data based on text similarity.

Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering is a clustering method with a bottom-up approach. The procedure of the Agglomerative algorithm starts with n single clusters, then helps the sequence by combining clusters sequentially (Zhou et al., 2017). In this method, cluster merging is done based on Euclidean distance as a proximity measure. The formula used for Euclidean is (Purnamasari & Fidia Deny Tisna Amijaya, 2022):

$$d_{ij} = \sqrt{\sum_{k=1}^{p}(x_{ik} - x_{jk})^2} \quad (2)$$

Description :

p = number of variabels
d_{ij} = Euclidean distance between observation i and observation j
x_{ik} = the value of the i observation on k variabel
x_{jk} = the value of the j observation on the observation on the k variabel
i, j = 1,2,3,…n

In this clustering, ward linkage is also used which aims to minimize the total variance in the group. At each step, pairs of cluster pairs with the smallest distance are combined (Abdulhafedh, 2021).

Silhouette Score

Agglomerative Clustering uses the Silhouette Score method to evaluate cluster quality. This method calculates the average distance between an object and all other objects in the cluster and the minimum average distance between that object and all other clusters (Cahyo & Sudarmana, 2021). A Silhouette score close to 1 indicates a dense cluster, and the objects are far away from other clusters, indicating good cluster quality. If the silhouette score is close to -1, the cluster is not dense, and the objects are close to other clusters, indicating poor cluster quality. The following is the formula for silhouette score (Yusup & Maharani, 2021):

$$Sillhouette\ Score = \frac{1}{n}\sum_{i=1}^{n}\frac{(b_i - a_i)}{\max\,(a_i, b_{i)}} \quad (3)$$

Description :

n = amount of data
a = average distance
i = other data i in cluster
b = minimum average distance between other data in different clusters

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2023
DOI : https://doi.org/10.33395/sinkron.v8i3.12616
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Cluster Visualization

Cluster visualization is a graphical representation of the clustering results that makes it possible to observe the relationship between word elements associated with different clusters. Visualizations used include wordcloud, and scatter plot. Wordcloud is used to visualize the frequent words in each cluster, while scatter plot are used to visualize the relationship between two numerical variables in the form of dots on a graph.

Analysis Content Types and User Emotional Spheres

The analysis of content types and user emotions is carried out by identifying the characteristics of the content conveyed by users in the clusters formed. Based on the dominant words, an interpretation of the patterns of content and emotions in the cluster will be formed.

RESULT

In this section, seen in the included Table 2, there are test results using various cluster numbers. After analyzing the data, it is found that the highest value is found in cluster number 10 compared to the other cluster numbers. This shows that in the context of this research, cluster 10 provides the most optimized results compared to the other cluster values.

Table 2. Silhouette Score

No Cluster	Silhouette Score
3	0.0006003595013796891
5	0.0013466229985442167
7	0.002487323503842825
9	0.0034552870040058773
10	0.0035657088325235804
12	0.0025300973409010173

Centroids are represented as points in the plot that represent the center of the cluster. The position of each centroid reflects the center location of the clusters so as to gain a visual understanding of how clusters form and fit together.

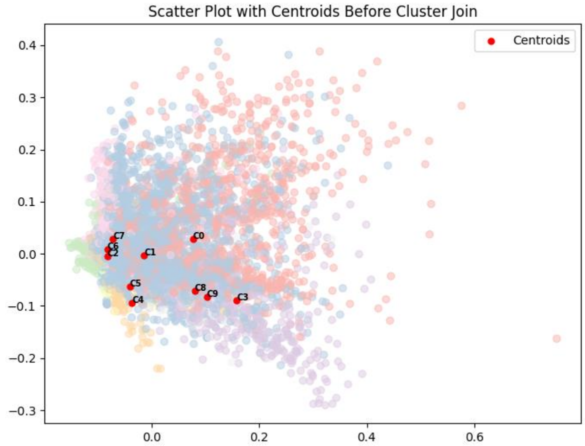

Figure 3. Scatter Plot with Centroid Before Cluster Join

The cluster results that have been obtained will be analyzed based on the frequency of occurrence of words in each cluster. With the frequency of occurrence of words, it is possible to identify the dominant words in each cluster. This can help in understanding the type of content and emotions that appear in each cluster. Here is a table for the frequency of data from each cluster :

Table 3. Frequency of Words

CLUSTER	KATA DAN KEMUNCULAN

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2023
DOI : https://doi.org/10.33395/sinkron.v8i3.12616
e-ISSN : 2541-2019
p-ISSN : 2541-044X

0	(anies, 1821), (baswedan, 983), (presiden, 349), (Indonesia, 174), (pemimpin, 172), (beliau, 154), (calon, 153), (lampung, 139), (capres, 134), (menyambut, 125), (warga, 119), (ri, 117), (sosok 116), (dukungan, 103), (relawan, 100), (perubahan, 99), (kedatangan, 94), (pilpres, 93), (antusias, 91), (kalimantan, 89), (prabowo, 85), (mendukung, 84), (pasar, 82), (disambut, 81), (banget, 79), (emang, 79), (pontianak, 79), (koalisi, 78), (selatan, 78), (hangat, 75)
1	(anies, 2498), (baswedan, 645), (indonesia, 464), (politik, 314), (jakarta, 268), (presiden, 251), (rakyat, 217), (perubahan, 198), (semoga, 190), (orang, 188), (banget, 181), (pemimpin, 177), (warga, 157), (dukung, 151), (pertamina, 147), (emang, 146), (plumpang, 139), (negara, 138), (allah, 129), (bangsa, 121), (mas, 116), (yaa, 116), (beliau, 111), (dki, 110), (kampanye, 108), (identitas, 104), (pemilu, 103), (depo, 102), (janji, 100), (ri, 99)
2	(pemilu, 268), (indonesia, 216), (semoga, 159), (umkm, 118), (ruu, 115), (bahu, 99), (surya, 92), (paloh, 92), (anies, 91), (anak, 89), (mantap, 84), (muda, 84), (meningkatkan, 78), (upaya, 74), (pemerintah, 73), (setuju, 73), (banget, 73), (konsisten, 72), (mendorong, 71), (bangsa, 71), (fraksi, 70), (maju, 68), (hukum, 64), (ekonomi, 64), (berjalan, 63), (pembangunan, 60), (sektor, 58), (kekerasan, 57), (mendukung, 55), (pprt, 55)
3	(anies, 363), (baswedan, 245), (menyambut, 103), (pontianak, 78), (kedatangan, 77), (kalimantan, 76), (antusias, 75), (lampung, 63), (selatan, 59), (sambutan, 57), (warga, 55), (disambut, 54), (hangat, 53), (meriah, 45), (presiden, 39), (kesultanan, 37), (banget, 30), (kunjungan, 29), (ri, 27), (beliau, 27), (menyapa, 25), (barat, 25), (berkunjung, 24), (hui, 24), (calon, 23), (way, 22), (indonesia, 21), (relawan, 20), (kalbar, 20), (sungai, 20)
4	(surya, 88), (paloh, 88), (prabowo, 26), (kalimantan, 21), (silaturahmi, 21), (kebangsaan, 16), (ahy, 16), (prananda, 15), (luhut, 15), (bacaleg", 12), (anies, 11), (kedatangan, 11), (tower, 10), (subianto, 10), (lumbung, 9), (suara, 9), (buka, 9), (balikpapan, 9), (menemui, 9), (meriah, 9), (pulau, 8), (kediaman, 8), (penyambutan, 7), (rumah, 7), (sambutan, 7), (kunjungan, 7), (cawapres, 7), (menargetkan, 6), (ketua, 6), (menyambut, 6)]
5	(depo, 69), (pertamina, 66), (plumpang, 63), (kebakaran, 40), (anies, 29), (relokasi, 21), (evaluasi, 14), (kilang, 13), (mendukung, 12), (baswedan, 10), (korban, 7), (peristiwa, 6), (unsur, 5), (kelalaian, 5), (tanggung, 5), (bertanggung, 5), (ahok, 5), (dukung, 5), (insiden, 5), (setuju, 5), (kemarin", 5), (keamanan, 4), (imb, 4), (warga, 4), (penyebab, 4), (dalang, 4), (semoga, 4), (pemukiman, 4), (terkait, 4), (bahan, 4)
6	(ruu, 113), (pprt, 55), (pekerja, 35), (rumah, 33), (tangga, 33), (disahkan, 29), (hukum, 19), (pengesahan, 19), (provinsi, 16), (mha, 16), (indonesia, 15), (perlindungan, 15), (dibutuhkan, 13), (semoga, 12), (saan, 11), (prt, 11), (alasan, 11), (jabar, 10), (landasan, 9), (bertujuan, 8), (mengisi, 8), (kekosongan, 8), (fraksi, 8), (pembahasan, 8), (menjaga, 8), (karna, 8), (adat, 8), (pemerintah, 8), (kesehatan, 7), (irma, 7)]
7	(indonesia, 183), (umkm, 118), (anak, 88), (semoga, 83), (muda, 83), (meningkatkan, 77), (upaya, 74), (konsisten, 71), (mendorong, 67), (bangsa, 65), (maju, 65), (pemerintah, 64), (ekonomi, 64), (pembangunan, 59), (sektor, 58), (banget, 55), (kekerasan, 54), (daerah, 52), (setuju, 52), (anies, 52), (beras, 51), (stunting, 49), (mantap, 49), (perempuan, 46), (generasi, 45), (nasional, 44), (digital, 44), (fraksi, 43), (literasi, 43), (negara, 43)]
8	(anies, 53), (alpukat, 52), (pohon, 46), (baswedan, 34), (lampung", 34), (penanaman, 31), (bibit, 30), (warga, 26), (menanam, 21), (rumah, 18), (desa, 16), (untoro, 15), (aligator, 14), (silaturahmi, 14), (berjenis, 13), (kebangsaan, 12), (milik, 11), (petani, 9), (trimurjo, 9), (halaman, 9), (bernama, 8), (kegiatan, 8), (rangkaian, 8), (dihalaman, 7), (beliau, 7), (saryano, 6), (kecamatan, 6), (provinsi, 6), (semoga, 5), (kunjungan, 5)

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2023
DOI : https://doi.org/10.33395/sinkron.v8i3.12616
e-ISSN : 2541-2019
p-ISSN : 2541-044X

9	(pasar, 59), (anies, 56), (natar, 49), (lampung, 32), (baswedan, 26), (pedagang, 24), (berdialog, 19), (warga, 17), (selatan, 14), (menyempatkan, 12), (berbelanja, 9), (menyapa, 9), (berkunjung, 8), (dialog, 7), (kunjungan, 7), (komoditas, 7), (kunjungannya, 5), (merakyat, 5), (terapung, 5), (langsung, 4), (pemimpin, 4), (calon, 4), (mengunjungi, 4), (antusias, 4), (bercengkrama, 4), (potret, 3), (banget, 3), (yaa, 3), (mantap, 3), (salah, 3)

DISCUSSIONS

Table 4. Results of Cluster Analysis

Cluster	Topics	Content	Emotions
0	Citizens' support and enthusiasm for Anies Baswedan as a leader in Indonesia.	Opinion	Excitement
1	The political changes expected by the public are related to Anies Baswedan's leadership in Jakarta.	Opinion	Expectations
2	Hope to improve the development sector and UMKM in Indonesia.	Opinion	Excitement
3	Anies Baswedan's visits to regions, such as Pontianak and Kalimantan, were greeted with enthusiasm by local residents.	Opinion	Excitement
4	Meetings and political activities involving Surya Paloh and Prabowo.	Information	Expectations
5	Issues surrounding Pertamina, including fire and relocation	Information	Doubt
6	Changes in the law (RUU) that received attention included some wording related to vacancies and worker protection.	Information	Excitement
7	Economic development, and tackling violence in Indonesia, as well as support for Anies Baswedan.	Information	Expectations
8	Tree planting activities, especially avocados, and Anies Baswedan's visit to Lampung to support these activities.	Information	Excitement
9	Anies Baswedan's visit to the market in Natar Lampung and interaction with traders and local residents	Information	Excitement

Based on the table above, it can be concluded that from the topics presented, the related content and the emotions contained are that the community at large provides strong support and enthusiasm for Anies Baswedan's leadership as a leader in Indonesia. People have hope and confidence that positive changes will occur in various sectors, including the development sector and UMKM. While there are some issues that raise doubts or uncertainties, such as issues surrounding Pertamina and changes in the law, overall, the analysis shows great support, excitement and hope for Anies Baswedan and the prospects for positive change in Indonesia. Thus, it can be concluded that the public is generally supportive of Anies Baswedan's leadership and hopeful for change for the better, although there are some issues that still raise doubts or uncertainties.

CONCLUSION

After conducting research using Hierarchical Clustering to categorize the content type and emotional scope of Twitter users during the 2024 Presidential Election, several conclusions can be drawn. Twitter users engaged in diverse discussions, including support, criticism, and policy debates related to the election. The optimal number of clusters obtained is 10 cluster with 4 clusters containing opinion content and 6 clusters containing informational content. Within these clusters, 6 clusters reflected excitement, indicating strong public support and enthusiasm for events involving Anies Baswedan, such as regional visits and tree-planting activities. 3 clusters represent expectations, indicating expectations for political change under Anies Baswedan's leadership, especially in sectors such as the economy, development, and violence reduction. Finally, 1 cluster reflects doubt, indicating uncertainty regarding issues such as Pertamina, fires, relocation, and changes in laws related to job

*name of corresponding author

SinkrOn JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 3, July 2023
DOI : https://doi.org/10.33395/sinkron.v8i3.12616
e-ISSN : 2541-2019
p-ISSN : 2541-044X

vacancies and worker protection. These findings are based on word frequency analysis. Future research may consider using Word2Vec to capture word meaning and relationships. In addition, expanding the scope of the study to include multiple candidates would provide a broader perspective in analyzing the content and emotions associated with the presidential election.

REFERENCES

Abdulhafedh, A. (2021). Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation. *Journal of City and Development*, *3*(1), 12–30. https://doi.org/10.12691/jcd-3-1-3

Belcastro, L., Branda, F., Cantini, R., Marozzo, F., Talia, D., & Trunfio, P. (2022). Analyzing voter behavior on social media during the 2020 US presidential election campaign. *Social Network Analysis and Mining*, *12*(1). https://doi.org/10.1007/s13278-022-00913-9

Cahyo, P. W., & Sudarmana, L. (2021). A Comparison of K-Means and Agglomerative Clustering for Users Segmentation based on Question Answerer Reputation in Brainly Platform. *Elinvo (Electronics, Informatics, and Vocational Education)*, *6*(2), 166–173. https://journal.uny.ac.id/index.php/elinvo/article/view/44486

Devika, R., Revathy, S., Sai Surriya Priyanka, U., & Subramaniya Swamy, V. (2018). Survey on clustering techniques in Twitter data. *Proceedings of the 2nd International Conference on Computing Methodologies and Communication, ICCMC 2018*, *Iccmc*, 1073–1077. https://doi.org/10.1109/ICCMC.2018.8487969

Dirjen, S. K., Riset, P., Pengembangan, D., Dikti, R., & Mailoa, E. (2017). Terakreditasi SINTA Peringkat 2 Analisis Node dengan Centrality dan Follower Rank pada Twitter. *Masa Berlaku Mulai*, *1*(3), 937–942.

FRHAN, A. J. (2017). Hierarchical Agglomerative Clustering Algorithm Based Real-Time Event Detection from Online Social Media Network. *Pdfs.Semanticscholar.Org*, *13*, 215–222. https://pdfs.semanticscholar.org/7dc8/c6bdd6fea850f4e99323d46a09d94e0a2cd9.pdf

Irawan, E., Mantoro, T., Ayu, M. A., Bhakti, M. A. C., Yogi, I. K., & Permana, T. (n.d.). *Analyzing Reactions on Political Issues in Social Media Using Hierarchical and K-Means Clustering Methods*.

Mardianti, S., Zidny, M., & Hidayatulloh, I. (2018). Ekstraksi tf-Idf n-gram dari komentar pelanggan produk smartphone pada website e-commerce. *Seminar Nasional Teknologi Informasi Dan Multimedia*, *6*(April), 79–84.

Nellie, V., Christanti, V., Novario, M., & Perdana, J. (n.d.). *IMPLEMENTASI METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING UNTUK SISTEM REKOMENDASI FILM*.

NM, E. (2016). A Typology of Voters: Creating Voters' Profiles via Clustering. *Journal of Political Sciences & Public Affairs*, *4*(2), 2–7. https://doi.org/10.4172/2332-0761.1000205

Purnamasari, I., & Fidia Deny Tisna Amijaya, dan. (2022). Perbandingan Hasil Analisis Cluster Dengan Menggunakan Metode Average Linkage Dan Metode Ward (Studi Kasus : Kemiskinan Di Provinsi Kalimantan Timur Tahun 2018) Comparison Of Cluster Analysis Results Using Average Linkage Method And Ward Method (Case Stud. *Jurnal EKSPONENSIAL*, *13*(1), 9–18.

Sinha, P., Dey, L., Mitra, P., & Thomas, D. (2020). A Hierarchical Clustering Algorithm for Characterizing Social Media Users. *The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020*, 353–362. https://doi.org/10.1145/3366424.3383296

Vijaya, V., Sharma, S., & Batra, N. (2019). Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering. *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, 568–573. https://doi.org/10.1109/COMITCon.2019.8862232

Wulandari, N. A., Pratiwi, H., & Handayani, S. S. (2023). *Perbandingan Metode K-means and Agglomerative Nesting untuk Clustering Data Digital Marketing di Twitter*. *2*, 189–194.

Yusup, A. H., & Maharani, W. (2021). *Pembangunan Model Prediksi Kepribadian Berdasarkan Tweet Dan Kategori Kepribadian Big Five Dengan Metode Agglomerative Hierarchical Clustering*. *1*(1), 2021.

Zahrotun, L. (2015). Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan Metode Clustering K-Means Dan Agglomerative Hierarchical Clustering (Ahc). *Jurnal Informatika*, *9*(1), 1039–1047. https://doi.org/10.26555/jifo.v9i1.a2045

Zahrotun, L., Linarti, U., Harli, B., Suandi, T., & Kurnia, H. (2023). *Comparison of K-Medoids Method and Analytical Hierarchy Clustering on Students ' Data Grouping*. *7*(June), 446–454.

Zhou, S., Xu, Z., & Liu, F. (2017). *Method for Determining the Optimal Number of Clusters Based on Agglomerative*. *28*(12), 3007–3017.

*name of corresponding author