YouTube Comment Clustering Using K-Means in A Case Study of The Indonesian New Capital City (IKN)

Sausan Hidayah Nova*[1], Afian Syafaadi Rizki[2], Dwi Agung Wibowo[3], M Najamudin Ridha[4], Cahya Karima[5], Nindy Permatasari[6]
Politeknik Negeri Tanah Laut, Jl. Ahmad Yani No.Km.06, Pemuda, Kec. Pelaihari, Kabupaten Tanah Laut, Kalimantan Selatan, (0512) 2021065
E-mail : sausannova@politala.ac.id[1], afianrizki@gmail.com[2], dwi@politala.ac.id[3], najamudin@politala.ac.id[4], cahyakarima@politala.ac.id[5], nindy@politala.ac.id[6]*
**Corresponding author*
Received 13 October 2025; Revised 6 November 2025; Accepted 7 November 2025

Abstract - The relocation of the capital city of the Republic of Indonesia from Jakarta to the Nusantara Capital City (IKN) is a critical topic for the public, as it is designated as a strategic national project. However, the lack of public participation may generate community concerns regarding its potential impact. This research involved extracting public opinion from YouTube comments to identify the community's desires, thereby providing policymakers with valuable information. Clustering the comments using the K-Means method successfully extracted public opinions from 27,063 comment data points. Among the key findings, a significant public concern is the potential for the construction project to be abandoned or stalled ("mangkrak"). Additionally, while the clustering results showed good cohesion, the cluster separation indicated a significant overlap in the data. This is further reflected by the average similarity score of 0.4234972.

Keywords – YouTube, Text Clustering, K-Means, Nusantara Capital City (IKN)

1. INTRODUCTION

The development of the Nusantara Capital City (IKN) in East Kalimantan is a national strategic project that has generated diverse responses from the public [1]. As a major issue involving political, social, economic, and environmental aspects, several key issues that emerged from previous research include the lack of public participation and the short timeframe for the formulation of legislation [2]. This has raised deep public concern about the humanitarian impact it will have [3], as well as concerns regarding social inequality and potential environmental damage [4].

The IKN is frequently discussed by Indonesian people in various media outlets, particularly social media. Through social media, people can freely express their opinions, whether in the form of support, criticism, or suggestions [5]. The large amount of distributed opinion data makes manual analysis inefficient, necessitating computational methods to process and analyze the data [6]. However, processing unstructured opinion data requires data mining and machine learning techniques to transform it into meaningful information. Therefore, several previous studies have used text processing techniques to extract information from online media [7], [8]. Research conducted by Ahmad Yusuf et al. analyzed the sentiment of 1,836 Twitter posts, resulting in six policy recommendations, including openness and transparency, mitigation of social impacts and evictions, environmental protection, efficient budget management, inclusive infrastructure development, and local economic development [9]. Other research related to IKN focuses on improving the performance of pre-processing and data classification techniques, namely that conducted by Hidayah using 2177 Twitter upload data and the Random Forest classification method [10]. Other methods that are often used are Naive Bayes and SVM [11], [12], [7].

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

Sentiment analysis is a method used to determine the tendency of public opinion on a particular issue, whether it is positive, negative, or neutral [8]. However, sentiment analysis using text classification methods is not always the right method for extracting information from text documents because classification techniques require documents or data with predetermined sentiment labels. Therefore, in this study, a data grouping approach using clustering techniques is emphasized to extract information from text documents. Data uploaded from social media was grouped using the K-Means method. K-Means is an unsupervised learning algorithm capable of grouping data into several clusters based on similar characteristics [13], [14]. In the context of sentiment analysis, K-Means can be used to group public opinion into specific categories without requiring labeled training data. The advantages of this algorithm are its simplicity, efficiency in handling large data, and the ability to reveal hidden patterns in public opinion data.

This study provides a new perspective on the application of clustering to analyze public opinion related to government policies. This study uses YouTube as the primary platform for public opinion formation in Indonesia. The use of YouTube comments specifically related to capital city issues has not been widely studied, so this study contributes to understanding the digital community's perceptions of the capital's development.

Thus, the application of the K-Means algorithm in cluster analysis related to the development of the National Capital City (IKN) is expected to provide a clearer picture of public perception. This research contributes to the development of social media analytics studies relevant to policymakers in Indonesia. The results of this analysis can serve as a reference for the government and policymakers in understanding public opinion and as a basis for decision-making that aligns with public aspirations.

2. RESEARCH METHOD

Research on sentiment analysis regarding the development of the Nusantara Capital City (IKN) using the K-Means clustering method was conducted through several systematic stages, as follows:

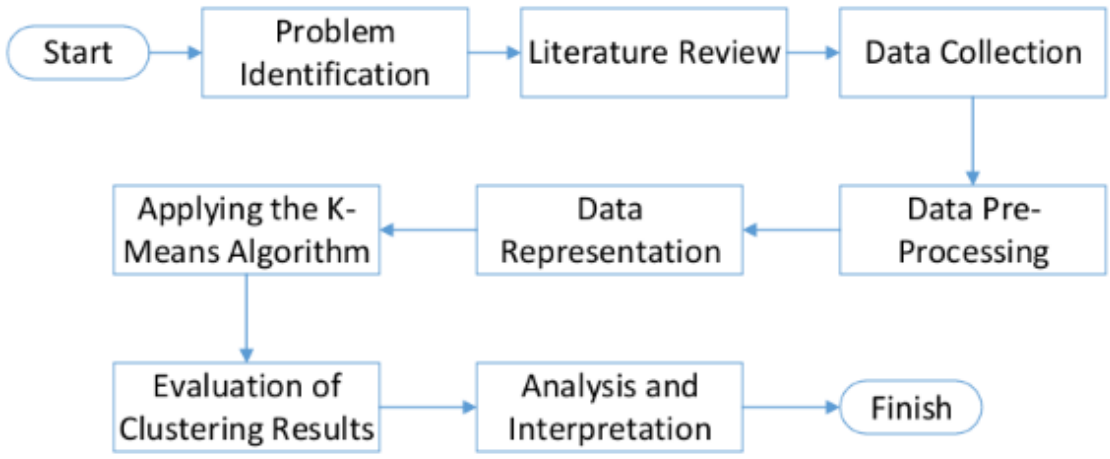

Figure 1. Flow and Research Stages

2.1. Problem Identification

The first stage is identifying the problem to be studied. The development of the Indonesian IKN has generated diverse public opinions, both supportive and anti-social. These opinions are widely disseminated on social media and online news portals. Therefore, a method

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

for analyzing public comments is needed to identify trends in public opinion regarding the development of the IKN.

2.2. Literature Review

Next, a literature review was conducted to review previous research related to text clustering analysis using clustering algorithms. Several text processing methods were chosen, including the text cleaning method, the Stemming method, and the Text Embedding method [15]. In this study, the K-Means algorithm was chosen as a text clustering method because it is capable of clustering text data quickly and efficiently, and is suitable for use with large amounts of data.

2.3. Data Collection

Data was collected from social media platforms, specifically YouTube. The data used in this research is 27063 comments data. The data collection process was carried out using official APIs or web scraping techniques. The videos were selected based on the keywords "IKN" or "Ibu Kota Nusantara". Other parameters considered were videos with the most views, the highest ratings, and the most recent release date.

2.4. Data Pre-Processing

The obtained raw data then undergoes a pre-processing stage to improve the quality of the text data [16], [17]. The steps taken include:

1) Crawling, which is collecting raw text data from YouTube videos, which is obtained from national news channels
2) Cleaning, which removes special characters, links, emojis, and unnecessary punctuation.
3) Case Folding, which converts all text to lowercase.
4) Tokenization, which is breaking down sentences or text documents into smaller units called tokens (usually individual words).
5) Stopword Removal, which removes common words that do not contribute to sentiment analysis, such as "yang", "dan", or "di".
6) Stemming, which converts words to their base form, using the literary library. This method is used to further reduce the dimension of data [18].

2.5. Data Representation

After pre-processing, the text is converted into a numerical representation using the sentence embedding method. The sentence embedding method used is the Universal Sentence Encoder with Deep Averaging Network. Universal Sentence Encoder is a powerful and widely used Natural Language Processing (NLP) model developed by Google. Its core function is to transform variable-length pieces of text, such as sentences, phrases, or short paragraphs, into fixed-length, high-dimensional numerical vectors. This method was chosen because of its speed and efficiency for large amounts of data [19].

2.6. Applying the K-Means Algorithm

The next step is applying the K-Means algorithm to cluster the data. The number of clusters is determined first. The K-Means algorithm works by determining the center point (centroid), then grouping the data into the closest clusters. This process is repeated iteratively until the centroid value stabilizes. K-Means uses a function shown in Equation 1.

$$WC = \sum_{i=1}^{k} \sum_{x \epsilon C_i} \|x - \mu_i\|^2 \quad (1)$$

W = squared distances between each data point
C = cluster i
k = number of clusters

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

x = embedding vector
μ_i = centroid of C

Since the data is pre-processed and transformed by Universal Sentence Encoder, the input to the algorithm is a matrix where each row is a dense, high-dimensional sentence embedding. The process is divided into two main stages: first is iterating to find the optimal k number using the Elbow method, and the final clustering. Both stages use K-Means clustering. Figure 2 shows the process used in this research.

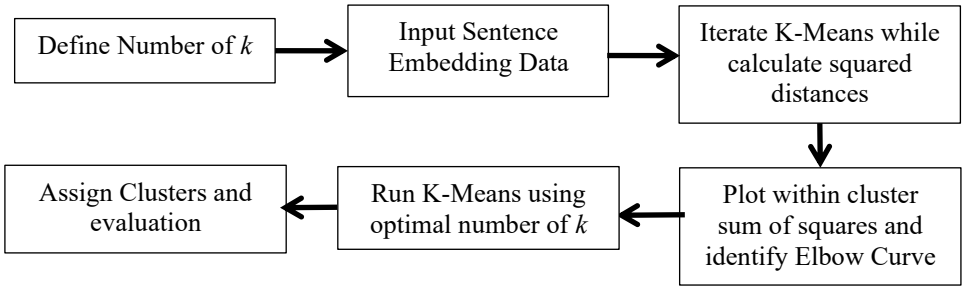

Figure 2. Steps of the K-Means Algorithm

2.7. Evaluation of Clustering Results

To assess the quality of clustering results, internal evaluation methods such as the Cosine Similarity Index are used [20], [21]. The Cosine Similarity is used to show how well each data point falls within its own cluster compared to other clusters. Since the Universal Sentence Encoder embeddings are typically approximately normalized, the denominator in the cosine similarity formula approaches 1. Therefore, the Inner Product of the two vectors is often used as an efficient and equivalent measure of similarity, as shown in Equation 2.

$$\text{Similarity}(A, B) \approx A \cdot B \quad (2)$$

The Cosine Similarity ranges from -1 to 1, including:

1) A value close to 1 indicates that the data fits within its cluster and is well separated from other clusters.
2) A value close to 0 indicates that the data is on the boundary between two clusters.
3) A value close to -1 indicates that the data is more similar to other clusters than to itself (a poor cluster).

Additionally, the researchers used a similarity matrix to assess the level of similarity between comments numerically and visualized the clustering results to assess the well-separated comment groups. To calculate the Similarity Matrix using the formula in Equation 3.

$$S = E \cdot E^T \quad (3)$$

If E is the embedding matrix where each row is a sentence vector, the similarity matrix S is computed as the dot product of the embedding matrix and its transpose. This operation yields an N times N matrix where N is the number of sentences.

The combination of these approaches was used to assess the success of the K-Means method in clustering YouTube comments based on similar content or discussion topics. Furthermore, the authors used a similarity matrix and visualized the clustering results to assess the success of the K-Means method in clustering comment text.

2.8. Analysis and Interpretation

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

Clustering results are then analyzed to determine trends in public opinion without determining sentiment. This analysis can be further deepened by examining the most frequently occurring words in each cluster. The interpretation results of each cluster will be used to formulate recommendations that can be used by the government in formulating policies.

3. RESULTS AND DISCUSSION

3.1. Problem Identification and Literature Review

An analysis of YouTube comments related to the development of the Nusantara Capital City (IKN) reveals a wide range of public opinions across various videos, particularly those discussing political, economic, and environmental issues. However, the sheer volume and unstructured nature of these comments pose a major challenge to efficiently understanding public perceptions. Initial data collection results indicate that YouTube comments often contain informal language, emoticons, and a mix of Indonesian and slang, complicating manual analysis.

This research focuses on applying the K-Means algorithm to group comment text into several thematic clusters that represent public opinion on the development of the new capital city. The clustering results are expected to assist the government, media, and researchers in understanding public opinion trends objectively and quickly.

3.2. Data Collection

In this study, 27063 comment uploads were successfully collected from 12 videos in June 2025. The attributes of the data collected include Author, Publish Date, Like Counts, and Comments.

3.3. Data Pre-Processing

Data preprocessing is used to reduce noise and data dimensionality. The methods used include cleansing, case folding, stop word removal, and finally, stemming. After stemming, empty comments and comments containing only one character are removed. The preprocessing results have successfully reduced the data size by 20%. Figure 3 below shows the data dimensionality reduction for each preprocessing method.

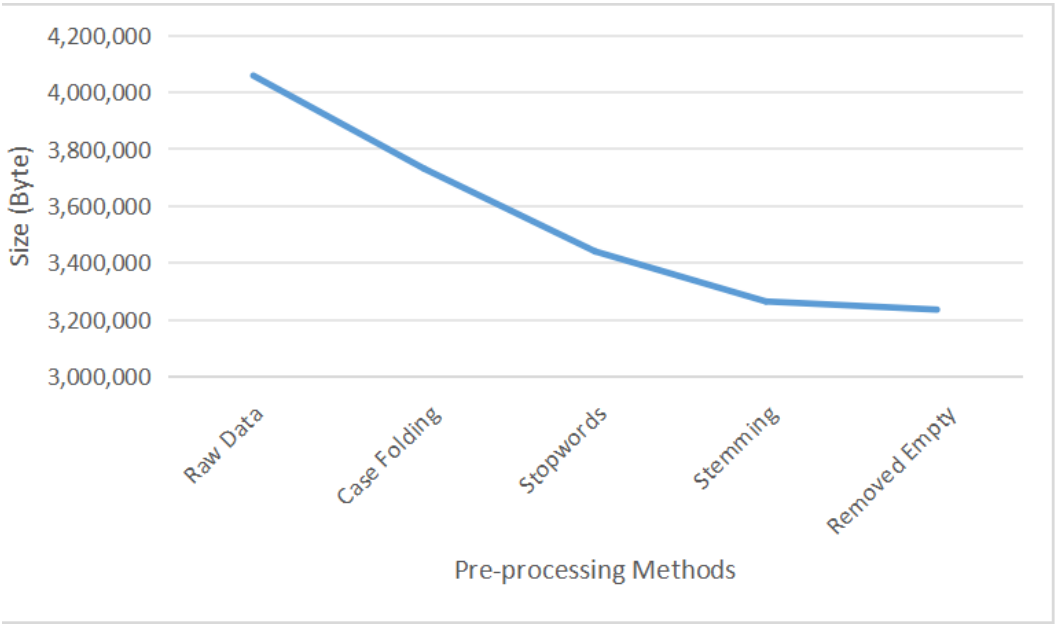

Figure 3. Data Size at Pre-Processing Steps

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

It can be concluded that the Case Folding and Stopwords methods contribute the most to data size reduction. This is because both processes involve the removal of words and characters. Stemming can reduce the data size, although not as much as Case Folding and Stopwords, as this stage only converts words to their bases without any word removal.

3.4. Sentence Embedding

Sentence Embedding is applied to data that has undergone preprocessing. At this stage, the data is successfully transformed into a vector of real numbers. Figure 4 below shows some of the results of sentence embedding using the Universal Sentence Encoder Library. The resulting embedding vector will be processed using the K-Means method to obtain its clusters.

```
Embedding for comment 1: [0.048547256737947464, -0.04365519434213638, -0.06206384301185608, 0.01150249782949686, -0.048517484217882156, 0.056807546
Embedding for comment 2: [0.007513972464948893, -0.023580534383654594, -0.06820865720510483, 0.01176538411527872, -0.06192219257354736, -0.04729676
Embedding for comment 3: [0.022273847833275795, -0.015585077926516533, 0.0759504958987236, 0.050845760852098465, -0.04657096043229103, 0.092526033
Embedding for comment 4: [0.095396026968956, -0.025834478437900543, 0.00972290150821209, 0.012414882890880108, 0.06293277442455292, -0.013178054243
Embedding for comment 5: [-0.0021627829410135746, -0.02602274902164936, -0.04631708562374115, 0.0367632657289505, 0.0629071369767189, -0.010080413
Embedding for comment 6: [-0.0024914084933698177, -0.007645791862159967, -0.013309292495250702, 0.04836001247167587, -0.052018895745277405, -0.012
Embedding for comment 7: [0.009611349552869797, -0.05221940577030182, -0.07494647055864334, -0.02480331063270569, 0.004439414944499731, -0.00566186
Embedding for comment 8: [0.045394141227006911, 0.012898515909910202, -0.06470436602830887, 0.03674280270934105, 0.064722679555541611, -0.00667998526
Embedding for comment 9: [-0.026660047471523285, -0.008884201757609844, -0.047404881566762924, -0.018724786111767292, -0.011997323483228683, -0.0452
Embedding for comment 10: [0.004320527892559767, -0.03344932943582535, -0.06529034674167633, 0.005697298329323353, -0.05010566860437393, 0.027280163
```

Figure 4. Comment Vector

3.5. K-Means Clustering

After sentence embedding, the text data is grouped into clusters using the K-Means method. The elbow method is used to determine the optimal number of clusters. This method relies on calculating the Within-Cluster Sum of Squares (Distortion), also known as the inertia, for a range of k values. WCSS measures the sum of squared distances between each point and its assigned cluster centroid. As k increases, the WCSS inherently decreases because the data points are partitioned into smaller, more numerous clusters, reducing the distance to their respective centroids.

The results are typically visualized by plotting the WCSS against k. The “optimal” k is identified at the point where the decrease in WCSS begins to slow down significantly, forming an “elbow” in the plot. Figure 5 below shows the number of k against the Sum of Squares (Distortion). Based on Figure 6, the number of clusters of 3 was chosen as the optimal k value because at that point the “elbow” shape is visible.

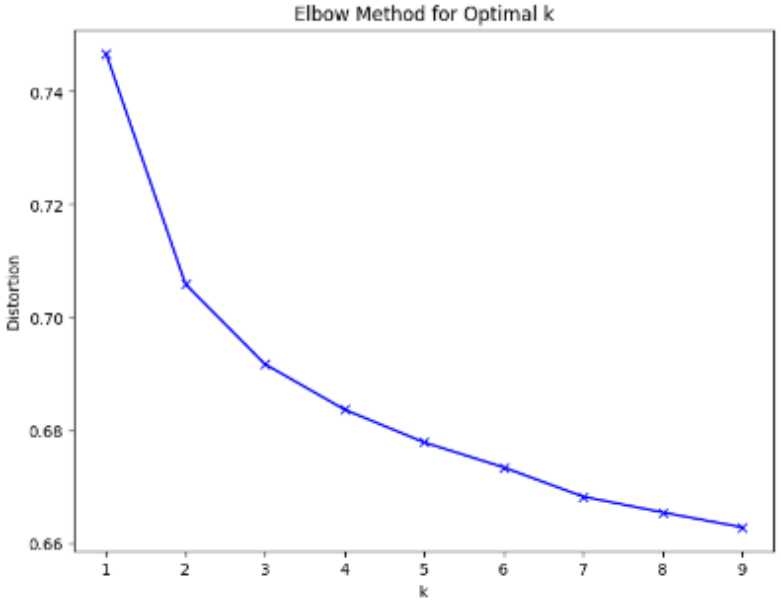

Figure 5. Elbow Method

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

The documents were then grouped into 3 clusters by running the K-Means method one last time. Next, the groups of sentences formed were labeled cluster 0, cluster 1, and cluster 2.

3.6. Evaluation of Clustering Results

To evaluate the performance of the K-Means method on clustering results, several methods were used, including calculating the average similarity, visualizing clustering results using scatter plots, and visualizing the similarity matrix.

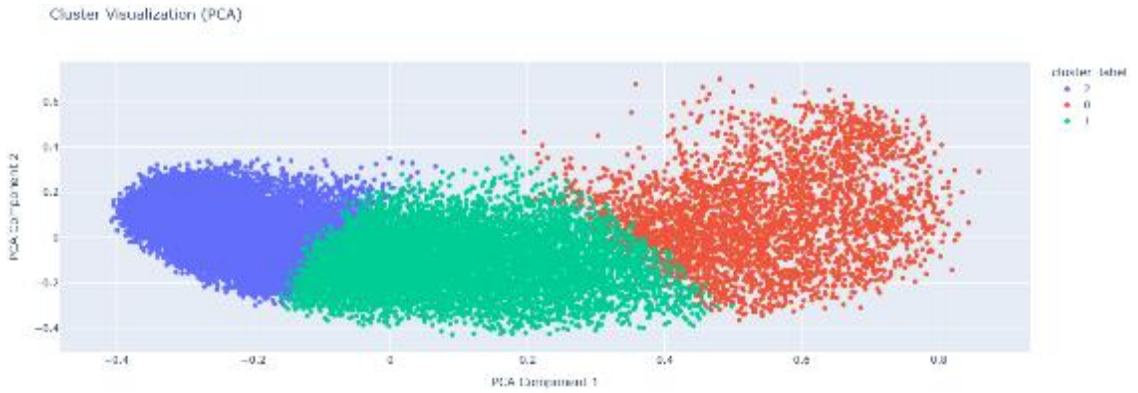

Figure 6. Scatter Plot 3 Clusters

Based on the figure, the performance of the K-Means algorithm can be summarized using the concepts of Separation and Cohesion. In terms of separation, a good clustering method can be seen from the clear boundaries between two clusters, as seen in Figure 6, where the boundaries of each cluster are clearly visible. However, the separation distance between the clusters is very close, indicating the method is less able to separate the clusters. This could be caused by the text data having high similarity.

Judging from the cohesion, it appears that Cluster 2 has a high level of cohesion. This means the cluster is very homogeneous or cohesive. All members of the cluster share very similar characteristics. Meanwhile, cluster 0 has scattered points, this could suggest the cluster is a good representation of a broad group, or that the clustering algorithm may not have found the most optimal grouping, or perhaps the data itself naturally forms a more diffuse group.

Figure 7 below shows the similarity matrix of the sentence embedding. It can be seen that the structure can be said to be weak, or there is a lot of overlap in the data. This is evident from the average similarity calculation of 0.4234972, indicating no directional relationship or unrelatedness in the data used.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

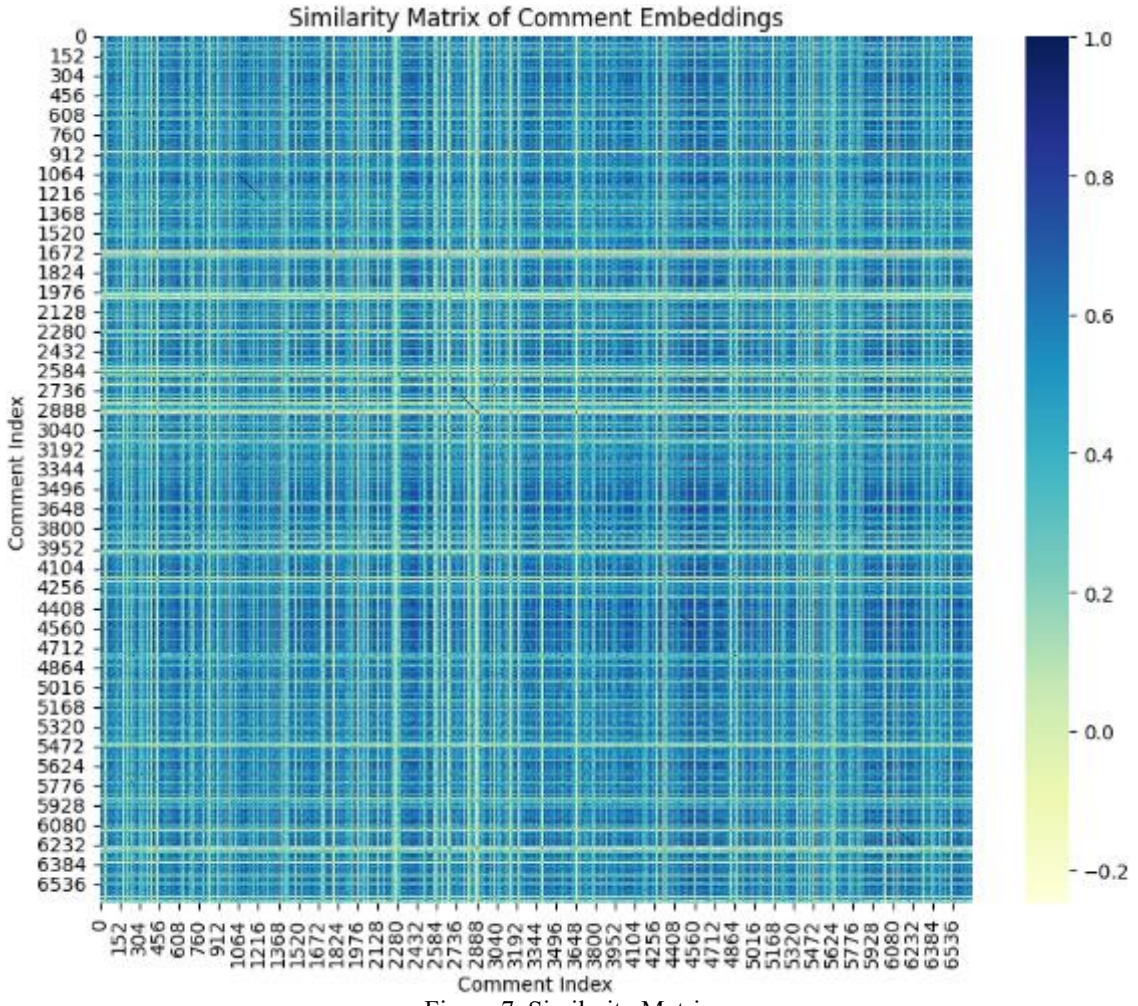

Figure 7. Similarity Matrix

3.7. Analysis and Interpretation

This section will discuss the interpretation of each cluster. There are three clusters interpreted: cluster 0, cluster 1, and cluster 2.

Figure 8. Word Cloud for Cluster 0

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

Cluster 0, as a whole, depicts a critical discussion about the future of the Nusantara Capital City Project, led by “Mulyono” (Jokowi). The discussion is driven by hopes (Mantap, Lanjut, Semoga) and criticism (Mangkrak), with the project’s success heavily dependent on investor support. This debate also allegedly involves the role of Buzzers, who attempt to shape public opinion.

Figure 9. Word Cloud for Cluster 1

Overall, Cluster 1 reflects the two-sided debate that frequently arises in the public sphere (especially on social media) regarding the Nusantara Capital City (IKN). One side conveys the message: “IKN (Keren Indonesia) harus Lanjut dan Mantap demi Rakyat”, which is balanced by a critical concern: “Apakah IKN di bawah Jokowi akan Mangkrak?” This cluster is not only about development, but also about the political future and legacy of the project.

Figure 10. Word Cloud for Cluster 2

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

Cluster 2 as a whole reflects deep public concern that the ongoing New Capital City (IKN) project is at risk of stalling, which would have significant consequences for Indonesia and its people, especially if it falls short of Jakarta's. The focus is on the urgency of the project's completion, and the responsibility for this is explicitly placed on Mr. Jokowi.

4. CONCLUSION

Based on the research findings, the comment documents obtained from YouTube regarding the IKN (Indonesia's new capital city) project have been successfully categorized into 3 clusters. From these three clusters, it can be concluded that the public desires clarity regarding the capital relocation project, especially in the event of a change in leadership. Furthermore, the risk of the project being abandoned ("mangkrak") would carry significant consequences for both Indonesia and its people, particularly in terms of financial loss and time. The history of other failed or unsuccessful projects may influence this concern. The government should reiterate its commitment to the IKN, which has already been done with the issuance of Presidential Regulation Number 79 of 2025. However, the project's continuation cannot be borne solely by the government but requires investor support, as implied in Cluster 0.

From the technical side of text clustering using K-Means, the Pre-Processing stage successfully reduced the data dimension by 20%. However, in the sentence embedding stage, the average similarity is 0.4234972, indicating a significant overlap within the data. This is reflected in the visualization of the cluster results, which are overlapping and closely spaced. Future research could compare whether other sentence embedding methods would yield the same pattern or if other clustering methods should be used. Alternative sentence embedding methods include TF-IDF, Word2Vec, and BERT, while other clustering methods include DBSCAN.

REFERENCES

[1] K. P. S. S. Hadiningrat, "Opportunities and Challenges in the Development of A Sustainable Capital City of Nusantara," *J. Lemhannas RI*, vol. 12, no. 1, hal. 53–64, 2024, doi: 10.55960/jlri.v12i1.534.

[2] E. Benia dan G. Nabilah, "Politik Hukum dalam Proses Pemindahan Ibu Kota Negara Melalui Pembentukan Undang-Undang Ibu Kota Negara (UU IKN)," *J. Huk. Lex Gen.*, vol. 3, no. 10, hal. 806–825, 2022, doi: 10.56370/jhlg.v3i10.323.

[3] A. Aningtyas, S. I. A. Wardhani, A. F. Lesmana, dan F. I. Hamida, "Pengambilan Keputusan Pembangunan Ibu Kota Negara (IKN): Analisis Perspektif Masyarakat di Luar Daerah Kalimantan Ditinjau dari Aspek Kemanusiaan," *Sanskara Ilmu Sos. dan Hum.*, vol. 2, no. 01, hal. 36–42, 2024, doi: 10.58812/sish.v2i01.487.

[4] M. A. Satryadin dan I. A. Aulawi, "Perubahan Kebijakan Pemberian Hak Atas Tanah di Ibu Kota Negara dalam Perspektif Peraturan Presiden Nomor 75 Tahun 2024," *J. Pertanah.*, vol. 15, no. July, hal. 30–43, 2025, doi: https://doi.org/10.53686/jp.v15i1.279 Perubahan.

[5] N. Wulandari, Y. Chayana, Rahmat, dan H. H. Handayani, "Sentiment Analysis of the Relocation of the National Capital (IKN) on Social Media X Using Naive Bayes and K-Nearst Neighbor (KNN) Methods," *J. Appl. Informatics Comput.*, vol. 9, no. 3, hal. 724–731, 2025, doi: 10.33395/sinkron.v9i2.14622.

[6] I. Syabri, R. Sutriadi, dan N. Ramadhany, "Exploring Public Sentiments Using Big Data on Superhub Spatial Development of Nusantara, the New Capital City of Indonesia," *J. Reg. City Plan.*, vol. 35, no. 1, hal. 44–68, 2024, doi: 10.5614/jpwk.2024.35.1.3.

[7] Y. R. Dewi, N. W. S. Saraswati, M. O. E. Monny, I. B. G. Sarasvananda, dan I. G. Andika, "Sentiment Analysis of the Relocation of the National Capital on Social Media X," *Sinkron*, vol. 9, no. 2, hal. 625–636, 2025, doi: 10.33395/sinkron.v9i2.14622.

[8] S. F. Huwaida, R. Kusumawati, dan B. Isnaini, "Analisis Sentimen Komentar YouTube

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905

terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes," *Jambura J. Informatics*, vol. 6, no. 1, hal. 26–39, 2024, doi: 10.37905/jji.v6i1.24718.

[9] A. Yusuf, A. Rizani, R. Fitri, K. N. P. Pamungkas, dan W. A. Saputra, "Sentimen Positif atau Negatif: Perspektif Masyarakat Terhadap Pemindahan Ibu Kota Negara," *J. Masy. Indones.*, vol. 50, no. 2, hal. 277–300, 2024, doi: 10.55981/jmi.2024.8842.

[10] A. K. Hidayah, Y. Erwadi, dan S. Handayani, "Analisis Sentimen Publik Terhadap Pemindahan Ibu Kota Negara di Twitter Menggunaka Metode Klasifikasi Random Forest dan Smote," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 5, hal. 9109–9114, 2025.

[11] S. D. Prasetyo, S. S. Hilabi, dan F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, no. 1, hal. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.

[12] A. Setiawan dan R. R. Suryono, "Analisis Sentimen Ibu Kota Nusantara Menggunakan Algoritma Support Vector Machine dan Naïve Bayes," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, hal. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.

[13] S. I. Safitri, C. Suhery, dan S. Bahri, "Implementasi Algoritma K-Means Untuk Clustering Sentimen Pada Opini Kualitas Pelayanan Jasa Penerbangan," *Coding J. Komput. dan Apl.*, vol. 9, no. 02, hal. 186, 2021, doi: 10.26418/coding.v9i02.47377.

[14] M. R. T. Ramdhani, M. R. P. Budika, M. V. Santoso, N. Alfira, dan N. Zahry, "Analisis Sentimen Terhadap Komentar Negatif (Hate Speech) di Twitter dengan Algoritma K-Means Clustering Menggunakan Rapidminer," *J. Inf. Technol. Informatics Eng.*, vol. 1, no. 1, hal. 57–61, 2025.

[15] A. Jabbar, S. Iqbal, M. I. Tamimy, A. Rehman, S. A. Bahaj, dan T. Saba, "An Analytical Analysis of Text Stemming Methodologies in Information Retrieval and Natural Language Processing Systems," *IEEE Access*, vol. 11, no. December, hal. 133681–133702, 2023, doi: 10.1109/ACCESS.2023.3332710.

[16] D. A. C. Rachman, R. Goejantoro, dan F. D. T. Amijaya, "Implementasi Text Mining Pengelompokkan Dokumen Skripsi Menggunakan Metode K-Means Clustering," *Eksponensial*, vol. 11, no. 2, hal. 167, 2021, doi: 10.30872/eksponensial.v11i2.660.

[17] A. N. Yusril, I. Larasati, dan Q. Aini, "Implementasi Text Mining Untuk Advertising dengan Menggunakan Metode K-Means Clustering Pada Data Tweets Gojek Indonesia," *Sistemasi*, vol. 9, no. 3, hal. 586, 2020, doi: 10.32520/stmsi.v9i3.924.

[18] A. S. Rizki, N. M. Aristi, M. N. Ridha, A. F. Zulfahri, dan D. A. Wibowo, "Implementation of The Indonesian Language Stemming Algorithm in Twitter Data Preprocessing. Case Study: Twitter Wargabanua and Instakalsel," *Fidel. J. Tek. Elektro*, vol. 5, no. 3, hal. 175–183, 2023, doi: 10.52005/fidelity.v5i3.170.

[19] M. Asgari-Chenaghlu, N. Nikzad-Khasmakhi, dan S. Minaee, "Covid-Transformer: Detecting COVID-19 Trending Topics on Twitter Using Universal Sentence Encoder," *arXiv*, 2020, doi: https://doi.org/10.48550/arXiv.2009.03947.

[20] T. Sumallika, V. Alekya, P. V. M. Raju, M. V. L. N. R. Rao, D. E. G. Shiney, dan M. V. Sudha, "Exploring Optimal Cluster Quality in Health Care Data (HCD): Comparative Analysis utilizing k-means Elbow and Silhouette Analysis," *Int. J. Chem. Biochem. Sci.*, vol. 25, no. 16, hal. 48–60, 2024, [Daring]. Tersedia pada: www.iscientific.org/Journal.html

[21] M. F. Fiqri, R. Muhammad, dan M. I. Ardimansyah, "Cluster Analysis of Emotions In Quranic Translation Using K-Means Clustering," *J. Softw. Eng. Inf. Commun. Technol.*, vol. 5, no. 2, hal. 123–134, 2024, doi: https://doi.org/10.17509/seict.v5i2.75942.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v24i4.14905