



# Implementation System for Cluster Analysis and Data Sentiment Using the K-Means Method to Determine the Most Discussed Topics

Wanra Tarigan<sup>1\*</sup>, Djoni<sup>2</sup>, Thamrin<sup>3</sup>, Lismardiana<sup>1</sup>

<sup>1</sup> Universitas Mandiri Bina Prestasi, Indonesia.

<sup>2</sup> Universitas Mikroskil, Indonesia.

<sup>3</sup> Institut Bisnis Informasi Teknologi dan Bisnis, Indonesia.

Received: January 17, 2026

Revised: February 04, 2026

Accepted: March 25, 2026

Published: March 31, 2026

Corresponding Author:

Wanra Tarigan

[wansibro@yahoo.com](mailto:wansibro@yahoo.com)

DOI: [10.29303/jppipa.v12i3.14838](https://doi.org/10.29303/jppipa.v12i3.14838)

 Open Access

© 2026 The Authors. This article is distributed under a (CC-BY License)



**Abstract:** The Community Activities Restrictions Enforcement (CARE) government rule, that is currently a matter of public concern, was implemented in 2021. Community Activities Restrictions Enforcement (CARE) highlights the community's pros and cons. Twitter and YouTube are social media that facilitate direct user expression. The material offered on social media platforms Twitter and YouTube is likewise quite diversified. Therefore, an automatic approach for topic detection is required, such as Mini Batch K-means Clustering, which facilitates user access to information. This study employs the Mini Batch method, which utilizes just a limited set of data for the clustering procedure. Based on testing with the Sum of Squared Error, this study's clustering results for tweet data including the phrase Community Activities Restrictions Enforcement (CARE) produced 12 cluster groups. The clustering results will be represented using Word Cloud, and the system will display the percentage of words based on Word Cloud.

**Keywords:** Clustering; Community Activities Restrictions Enforcement (CARE); Data sentiment; K-Means

## Introduction

The utilization of social media increases exponentially from year to year, in line with societal trends. Now, social media is a communal necessity for establishing relationships, a method of public communication for organizations and businesses, conducting commerce, and even enhancing the self-image of users. Twitter is one of the most popular social networking sites, particularly in Indonesia (Mujianto, 2019). Twitter is one of the social media developed by Jack Dorsey that is mostly used to transmit messages known as tweets. Twitter is the fifth most often used social media platform in 2020, with 59% of Indonesian individuals regularly using the platform. According to the available data, Twitter is an influential social media network among Indonesian users (Go et al., 2009; Habibi et al., 2016; Sulastri & Diartono, 2019; Widiastuti, 2017).

YouTube is the most popular social media platform in today's culture. It is anticipated that its popularity

would expand alongside the number of users. Prior to mid-2017, YouTube had 1.5 billion registered monthly users (logged-in monthly users). Youtube's rising popularity is a result of the site's video-sharing platform becoming more valuable to its users (Foster, 2020).

In 2021, the rule Community Activities Restrictions Enforcement (CARE) will be adopted, which is presently a source of anxiety for the public, notably on Twitter. Community Activities Restrictions Enforcement (CARE) is a policy created by the Government of Indonesia in early 2021 to fight COVID-19 (Krisdiyanto & Nurharyanto, 2021). Before Community Activities Restrictions Enforcement (CARE) was carried out in a number of Indonesian regions, the government had enforced extensive social restrictions. Community Activities Restrictions Enforcement (CARE) occurred in various places that became COVID-19 transmission hubs, notably on the islands of Java, Sumatra, and Bali (Ilyas, 2021; Kristhy et al., 2022; Nuraeny et al., 2021; Yahyadi & Latifah, 2022).

## How to Cite:

Tarigan, W., Djoni, Thamrin, & Lismardiana. (2026). Implementation System for Cluster Analysis and Data Sentiment Using the K-Means Method to Determine the Most Discussed Topics. *Jurnal Penelitian Pendidikan IPA*, 12(3), 584–592. <https://doi.org/10.29303/jppipa.v12i3.14838>

From the understanding that has been conveyed above, the research carried out is implementing a system for clusterization analysis and data sentiment using the K-means method to determine the most discussed topics (case study of Twitter and Youtube data about Community Activities Restrictions Enforcement (CARE)).

Prior research has been conducted on the analysis of clustering themes using K-Means clustering, such as the study "Proposed Clustering Grouping Models in Discussion of the 2015 Criminal Code Draft" (Supriyadi, 2015). The National Alliance for Criminal Code Reform provides another discussion model, referred to in this publication as a clustering or grouping-based discussion model. This strategy is thought to increase the efficacy and efficiency of Criminal Code Draft discussions. Because it is possible to identify which components are the primary causes of the problem.

In a further research conducted by the same author Lynda Rahmawati (Rahmawati et al., 2016), the clustering findings suggested the existence of comparable documents, which demonstrated document commonalities. However, the term use in the abstract is not exactly correct. Because processed words are insignificant, the usage of all words in a text might result in the inclusion of papers with various topics into the same cluster.

The definition of scientific clustering in data mining is the grouping of a number of data or objects into clusters (groups) so that each cluster has data that is as comparable to other clusters' data as feasible while being distinct from other clusters' data. In order to develop the best clusters, scientists continue to make numerous attempts to enhance the cluster model and compute the ideal number of clusters. We are familiar with two clustering approaches, namely hierarchical clustering and partitioning. Complete linkage clustering, single linkage clustering, average linkage clustering, and centroid linkage clustering comprise the hierarchical clustering approach. While the partitioning technique itself comprises of K-Means and Fuzzy K-Means, the partitions themselves are determined by fuzzy K-Means (Alfina et al., 2012).

Clustering is the process of grouping or categorizing items based on information gathered from data that describes the relationship between objects, using the maximization of similarities between members of one class and minimization of similarities between classes/clusters as its guiding principle (Rahmawati et al., 2016). Clustering organizes data into groups containing items with similar properties. Clustering is also known as unsupervised learning since no labels are

used during the clustering process and new labels might be applied after the clusters have been generated.

In general, the steps of text mining include tokenization, filtering, stemming, tagging, and analysis. Tokenizing is the process of isolating each individual word (token) from the input content. Filtering is a selection procedure for the tokenized words, which may be accomplished using stoplist or wordlist algorithms. The stoplist algorithm will eliminate words such as pronouns, adverbs, conjunctions, prepositions, and articles that are superfluous. In contrast, the wordlist algorithm will store the most significant words. The stemming procedure is then used to identify the root word of each word that has passed the filtering procedure (Rahmawati et al., 2016).

Text mining aims to extract valuable information from a collection of papers. The data source for text mining is therefore a collection of unstructured or at least semi-structured texts. Text grouping and classification are the specialized objectives of text mining.

Large volumes of unstructured data may be processed, organized/grouped, and analyzed via text mining. Text mining takes and develops several approaches from different disciplines, including Data Mining, Information Retrieval, Statistics and Mathematics, Machine Learning, Linguistics, Processing of natural materials, and Visualization, in order to provide answers (Nurhuda et al., 2016).

## Method

The implementation of the Community Activities Restrictions Enforcement (CARE) raises the pros and cons of the community. Twitter and Youtube are social media that aim to convey users' opinions directly. The contents of the information contained in social media Twitter and Youtube are also very diverse. Therefore, an automatic topic detection method is needed, such as the Mini Batch K-means Clustering method, which makes it easier for users to access information. In this study the approach used is the Mini Batch method which only uses a small group of data in the clustering process.

### *Mini Batch K-Means*

The Information System Framework method of clustering analysis of Twitter and Youtube data is shown in Fig 1.

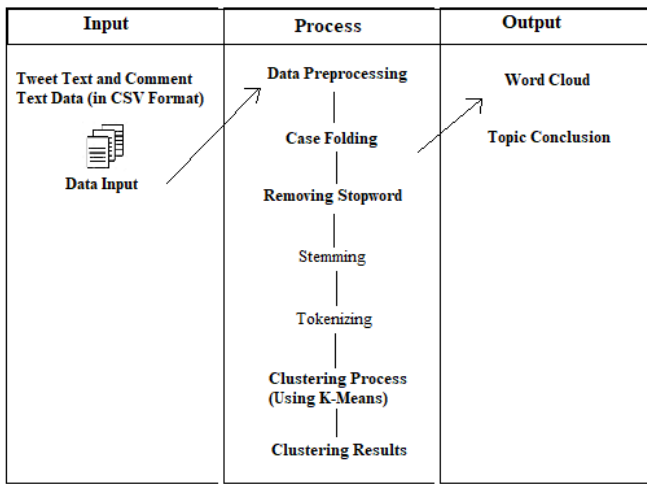


Figure 1. Information System Framework

The dataset included in this study consists of tweets or tweets on social media Twitter text and comment text data from the YouTube channel of the Indonesian President, which addresses Community Activities Restrictions Enforcement (CARE).

The first stage is to pre-process the dataset utilized; pre-processing is important for removing unneeded data properties from the dataset so that it can be used for subsequent processing. After preprocessing, clustering is the subsequent step. The clustering process is used to group the most frequently occurring words in the dataset.

The findings of this study are the most talked subjects based on the input data utilized in this study; the results are then represented using a word cloud derived from cluster results.

*Mini Batch K-Means*

The Mini Batch K-Means method is a variant of the K-Means algorithm that uses mini-batches to reduce computation time, while still trying to optimize the same objective function. This Mini Batch drastically reduces the number of computations required to integrate with local solutions. In contrast to other algorithms that reduce K-Means convergence time, Mini Batch K-Means produces results that are generally only slightly worse than standard algorithms.

This method seeks to minimize variations between data in a cluster and maximize variations with data in other clusters (Yudi Agusta, 2007). The K-Means method attempts to group existing data into several groups, with data in one group having the same characteristics as each other. An illustration of the K-Means algorithm can be seen in Fig 2.

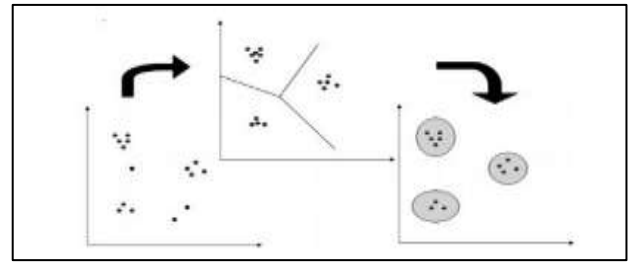


Figure 2. Illustration of the K-Means Algorithm (Safitri & Dewa, 2022)

Suppose a data set  $X = \{x_1, x_2, \dots, x_n\}$ , with  $x_n$  for  $n = 1, 2, \dots, N$  dimension  $D$ , which will be grouped into  $K$  clusters. Then  $b$  data will be taken randomly from set  $X$ , which is for example  $M$ , where variable  $b$  describes the size of the Mini Batch.

Furthermore, for the centroid update stage in the Mini Batch K-Means method, it will only use  $x_n \in M$  data. Using only a small portion of this data will obviously give lower results than using data in solving clustering problems. Therefore, the K-Means Mini Batch method uses the gradient descent method to update the centroid at each iteration. The use of the gradient descent method is intended so that the objective function used can quickly converge to the minimum point and produce a cluster model that is comparable to the K-Means method. The objective function to be optimized in this method is:

$$\text{Min } f(x, \mu) = \sum_{x_n \in M} \| \mu_k - x_n \|^2 \tag{1}$$

Furthermore, the K-Means Mini Batch Algorithm uses  $\eta$  which is one divided by the amount of data in the cluster. Because this method is iterative for every  $x_n \in M$ , the gradient descent formula used in the algorithm is :

$$\mu_{k_{i+1}} = \mu_{k_i} - \eta_k (\mu_{k_i} - x_n) \tag{2}$$

According to Yang et al. (2019) Mini Batch is a small data input sample that is randomly selected during each iteration, which significantly reduces computation time. In general, the Mini Batch K-Means algorithm can be described as follows: Determine the number of clusters; Randomly initialize  $\mu_k$  which is the initial centroid of each cluster; Take  $b$  data randomly, that is the data set  $M$ ; Calculating equation (2) to get the nearest centroid for each data  $x_n \in M$ ; Calculate  $\mu_k$  equal to one divided by the number of data in the cluster; Calculate the new centroid for each cluster using equation (1); Repeat steps e and f for each  $x_n \in M$ ; Evaluation of the convergence parameter  $\mu_k$ . If the convergence criteria have not been met, namely the condition when there is no movement of the data used or it reaches the maximum iteration value, then repeat steps c, d, e, f, and g using  $\mu_k$  and the

new data set  $M$ . The resulting output is the value of  $\mu_k$  as the centroid of each cluster.

*Word Cloud*

A word cloud is a collection of the most frequently occurring terms in the studied text data. This collection of words resembles a cloud of words, thus the name "word cloud" (Cooshna-Naik, 2022; Feng et al., 2022; Skeppstedt et al., 2024). The size of the letters in the word indicates the frequency with which the term is used. The more frequently a word occurs in the word cloud, the bigger its letters will be. This word cloud display is more engaging and faster at locating frequently occurring words, but contrary words cannot convey the frequency of terms in a text that is being evaluated (Adiyana & Hakim, 2015).

Understanding Python (programming language) is a high-level programming language that can execute a variety of multi-purpose instructions directly (interpretatively) using object-oriented programming techniques and employs dynamic semantics to make the syntax readable. Python is an easy-to-learn high-level programming language because it has automated memory management.

*Streamlit Framework*

Framework is a framework for developing websites. This framework was designed to help web developers write lines of code. Using a framework will make developing code more simple, quicker, and better organized.

The Streamlit framework is one of the available frameworks in the Python programming language. It may turn the program's look into a website version, making the program's results simpler for users to comprehend and utilize.

Streamlit is an open-source Python library. Streamlit facilitates the transformation of script data into interactive web applications. Streamlit is not only open source but also open sharing, making it easy to share.

**Result and Discussion**

*Data pre-processing*

At this step, each document's data is cleaned in accordance with the data pre-processing stage in order to generate data that is ready to enter the subsequent procedure. Following is an example of a remark before to data pre-processing and the results after case folding, filtering, deleting stopwords, and stemming have been performed.

Case Folding is often also referred to as lowercase. At this stage the process of uniforming the size of the letters in each document will be carried out where all

capital letters in the dataset will be converted to lowercase.

Filtering is done to remove all characters other than strings and to remove some of the characteristics of Youtube comments, namely in the form of punctuation marks, URL characters, emoticons, hashtags (#), username (@username), and white spaces (excess spaces). Removing Stopword or removing meaningless words in order to increase the accuracy value. Stemming is the process of extracting words from their affixes to get basic words. The result of the stemming process is called a stem. The application of the stemming process in each language varies depending on the morphology of each language. For this reason, even though the purpose of the stemming process is the same, namely getting basic words, a different algorithm is needed for each language. Tokenizing is the final stage in data pre-processing, namely tokenization using the N-terms feature where sentences are broken down according to the value of  $N=1$ .

*Model Implementation*

At the data pre-processing stage, a clustering process will be carried out to group data using the Mini Batch K-Means clustering algorithm with the Python programming language. In the model implementation process there are several stages as follows.

The library in question is a file containing Python computer code that is used to divide a huge program into smaller files, making it easier to maintain and organize. This library allows the code to be reused. The library used can be seen in Table 1.

```
In [16]: tfidf = TfidfVectorizer(
    min_df = 4,
    max_df = 0.95,
    max_features = 8000,
    stop_words = 'english')

tfidf.fit(df1['TextTweet'])
text = tfidf.transform(df1['TextTweet'])
```

**Figure 3.** The Process of Giving Word Weight

After getting a filtered or preprocessed list of words, TF-IDF with the N-terms feature is used to perform a word weighting process. As an example of using the word declaration for weighting words, the weight value for the word declaration is 1.577350 obtained using equation (1).

$$W = (1) \times \left( \ln \frac{(D+1)}{(DF+1)} \right) + 1$$

$$W = (1) \times \left( \ln \frac{(3+1)}{(1+1)} \right) + 1$$

$$W = (1) \times \left( \ln \frac{(4)}{(2)} \right) + 1$$

$$W = (1) \times (\ln(2)) + 1$$

$$W = (1) \times ((0.577350) + 1)$$

$$W = (1) \times (1.577350)$$

$$W = 1.577350$$

With the TF-IDF formula above, the implementation can be seen in Figure 3.

The data that has gone through the word weighting process will then be processed to cluster data on comments using the Mini Batch K-Means clustering algorithm with the Python programming language.

The process of grouping data in clustering commentary data. The implementation of the above algorithm flow can be seen in the Python instructions as shown in Fig 4.

**Table 1.** Used Library

Library Name	Comment Text
Case Folding	tolong ya kak, sampaikan inspirasi kami semua, jangan biarkan para elite politik hanya main2 saja... juga untuk next episode, mendingan usulkan saja pindah ibukota itu ke papua, soalnya di sana itu aman segalanya... di kalimantan ada kebakaran,
Filtering	tolong ya kak sampaikan inspirasi kami semua jangan biarkan para elite politik hanya main saja juga untuk next episode mendingan usulkan saja pindah ibukota itu ke papua soalnya di sana itu aman segalanya di kalimantan ada kebakaran
Removing StopWord	tolong ya kak sampai inspirasi kami semua jangan biar para elite politik hanya main saja juga untuk next episode mending usul saja pindah ibukota itu ke papua soal di sana itu aman segala di kalimantan ada bakar
Stemming	ya kak inspirasi semua biar elite politik main juga next episode mending usul pindah ibukota ke papua itu aman kalimantan bakar

```
In [18]: clusters = MiniBatchKMeans
(n_clusters=12, init_size=1024,
batch_size=2048, random_state=20)
.fit_predict(text)
```

**Figure 4.** Cluster formation

Fig 4 describes the formation of clusters using the Mini Batch K-Means clustering algorithm. In this method, there are two parameters to be considered, namely the batch size (batch\_size) and the centroid initialization parameter (random\_state). The batch size parameter was tested with a batch size larger than the number of clusters, namely (batch\_size=2048) and the size of the centroid initialization parameter, namely (random\_state=20). Determining the size of this batch also takes into account the size of each data.

Prior to the data grouping process, cluster optimal testing for comment data will be carried out using the Sum of Squared Error (SSE) method. This method is often used as a reference for related research in determining optimal clusters, by finding the squared value of the error difference from the data stored in the database with the test data. which can be seen in Figure 5.

In Figure 5 the data is tested, as an illustration given the value of the K iteration range for two, with the formation of 2 to 20 clusters with a batch size (batch\_size) = 2048, each K value counts the number of squared errors (SSE). The results of this stage describe a graph, the graph can be seen in Fig 6.

```
In [17]: def find_optimal_clusters(data, max_k):
iters = range(2, max_k+1, 2)

sse = []
for k in iters:
sse.append(MiniBatchKMeans(n_clusters=k,
init_size=1024,
batch_size=2048,
random_state=20).fit(data).inertia_)

print('Fit {} clusters'.format(k))

f, ax = plt.subplots(1, 1)
ax.plot(iters, sse, marker='o')
ax.set_xlabel('Cluster Centers')
ax.set_xticks(iters)
ax.set_xticklabels(iters)
ax.set_ylabel('SSE')
ax.set_title('SSE by Cluster Center Plot')

find_optimal_clusters(text, 20)
```

**Figure 5.** Optimal cluster testing

As can be observed in Fig 6(a) Graph of SSE values Data on YouTube comments and Fig 6(b) Graph of SSE values Data on Twitter comments containing the term "Community Activities Restrictions Enforcement (CARE)," the word is often used in online discussions. Fig 6(a) Graph of SSE values Data for YouTube comments had the highest fall in clusters from K=2 to K=12, but the best cluster is K=12 to K=14. And it can be differentiated in Fig 6(b) Graph of the SSE value of Twitter Tweet data containing the terms Community Activities Restrictions Enforcement (CARE) that K=10 and K=12 are the best cluster sizes (Wu et al., 2024).

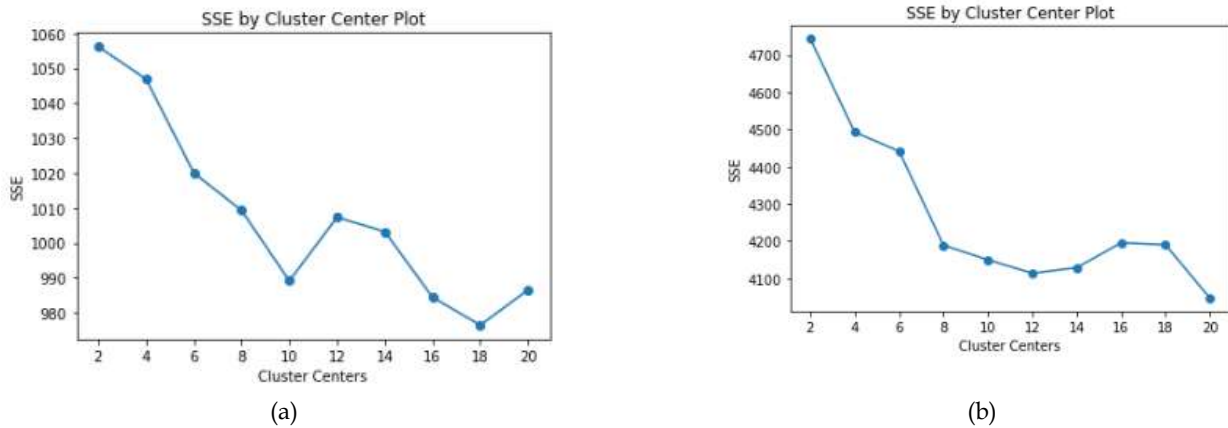


Figure 6. (a) SSE Cluster Centers Youtube Data (b) SSE Cluster Centers Data Twitter Fig. 6 Graph of SSE values

Table 2. Top keyword cluster results

Cluster	Youtube Top Keyword Data	Top Keyword Data Twitter
0	negara,kerja,bisa,makan,covid,pak,di,indonesia,ppkm,rakyat	gue,meng,alias,kayak,ppkm,plonga,kaya,planga,monyet,plongo
1	itu,jabat,jadi,jaga,jahat,utama,bawa,merah,sakit,positif	know,bgt,rsvm,gue,subdeng,covid,look,meme,udah,ppkm
2	ini,ppkm,rakyat,pak,kerja,perintah,gimana,kaya,sampe,mikir	atur,naik,indonesia,seluruh,jawa,panjang,perintah,ppkm,covid,level
3	covid,sangat,alaikum,assalamualaikum,indonesia,saya,jokowi,salam,presiden,hormat	sj,luhut,lapor,level,kak,masuk,covid,ppkm,ini,https
4	istri,itu,jabat,jadi,jaga,jahat,jajah,isolasi,youtube,surat	lancar,presidensi,jokowibapakinfrastruktur,dukung,bal,ppkm,presidensig,pulihbersama,recovertogetherrecoverstronger,Indonesia
5	presiden,beri,ku,kita,indonesia,semangat,bapak,pak,jokowi,sehat	tegak,disiplin,giat,hukum,titik,ramai,jajar,personel,polres,kukar
6	sengsara,jalan,klau,masjid,kenapa,mall,tutup,pasar,tradisional,buka	edar,waspada,merah,putih,covid,ppkm,vaksin,beritacovid,indonesiasehatindonesiahebat,endemicovid
7	kasih,ku,tuju,mundur,ya,presiden,jokowi,walaikum salam,love,pakde	gone,gonna,zenmai,ppkm,level,jgdekd,jgdekcmmioo,covid,kota,malang
8	youtube,presiden,sehat,semangat,cinta,tetap,panjang,umur,salam,walaikum	dont,diya,kiss,diyamin,think,soup,ppkm,butterfly,love,butterflysoup
9	president,kabar,dhe,rakyat,indonesia,like,mantap,halo,semangat,jokowi	spbu,mobilitas,pindah,camat,rangka,level,wilayah,giat,patroli,laksana
10	indonesia,info,terimakasih,ri,cinta,jokowi,bapak,semangat,sehat,presiden	golong,going,goes,gk,giving,gives,gitu,grabs,zenmai,ppkm
11	banyak,hancur,presiden,pak,bayar,harus,sy,gmna,jam,jual	did,face, didn, tell, frenchman, come, say, ve, way, spoilers

Table 2 shows the most prominent terms discovered by clustering Community Activities Restrictions Enforcement (CARE)-related tweets. Where each cluster generates a subject, 10 words are used to characterize each topic.

Word Cloud

In the following we can see in Fig 7 the pattern of word distribution in each cluster for topics related to Youtube comment data and for Twitter tweet data containing the word Community Activities Restrictions Enforcement (CARE) using the word cloud (Dumpawar et al., 2025; Shajari & Agarwal, 2025; Shevtsov et al., 2023).



- and Word Clouds as Strategies in Educational Research. *Journal of Learning for Development*, 9(1), 89–103. Retrieved from <https://eric.ed.gov/?id=EJ1344634>
- Dumpawar, A., Raut, A., Bais, A., Agrawal, R., Dhule, C., & Dhankar, P. K. (2025). Real-Time Sentiment Analysis of YouTube Comments Using NLP and TextBlob via YouTube Data API. *2025 International Conference on Engineering Innovations and Technologies (ICoEIT)*, 325–330. Retrieved from <https://ieeexplore.ieee.org/abstract/document/11211539>
- Feng, K. J. K., Gao, A., & Karras, J. S. (2022). Towards semantically aware word cloud shape generation. *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 1–5. <https://doi.org/10.1145/3526114.3558724>
- Foster, D. (2020). *Factors influencing the popularity of YouTube videos and users' decisions to watch them* [University of Wolverhampton]. <https://wlv.openrepository.com/items/d97d6e63-0fc4-4e05-8f3a-d003dbc4214b>
- Go, A., Huang, L., & Bhayani, R. (2009). Bhayani.: Twitter Sentiment Analysis (final project results). *J. Inform.* Retrieved from <https://nlp.stanford.edu/courses/cs224n/2009/fp/3.pdf>
- Habibi, R., Setyohadi, D. B., & Wati, E. (2016). Analisis sentimen pada Twitter mahasiswa menggunakan metode backpropagation. *Jurnal Informatika*, 12(1). <https://doi.org/10.21460/inf.2016.121.462>
- Ilyas, F. (2021). Analisis swot kebijakan pembatasan sosial berskala besar (psbb) dan pemberlakuan pembatasan kebijakan masyarakat (ppkm) terhadap dampak ekonomi di tengah upaya menekan laju pandemi covid-19. *Akrab Juara: Jurnal Ilmu-Ilmu Sosial*, 6(3), 190–198. <https://doi.org/10.58487/akrabjuara.v6i3.1559>
- Krisdiyanto, T., & Nurharyanto, E. M. O. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT*, 7(1). Retrieved from <https://shorturl.asia/YmVQI>
- Kristhy, M. E., Kristanto, K., Siswanto, E., Martono, A. B., & Nababan, R. M. (2022). Legal Politics of Regional Quarantine during the Covid-19 Pandemic with the Approach to Implementing Community Activities Restrictions (PPKM) Level 1-4. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, 5(3), 18308–18317. Retrieved from <https://www.bircu-journal.com/index.php/birci/article/view/5814>
- Mujianto, H. (2019). Pemanfaatan Youtube sebagai media ajar dalam meningkatkan minat dan motivasi belajar. *Jurnal Komunikasi Universitas Garut: Hasil Pemikiran Dan Penelitian*, 5(1), 135–159. <https://doi.org/10.10358/jk.v5i1.588>
- Nuraeny, R., Azizah, S. N., & Salam, A. N. (2021). Pengaruh Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Darurat (PPKM) Pengaruh Terhadap Ketahanan Keluarga Pedagang di Kebumen. *Jurnal Indonesia Sosial Teknologi: P--ISSN*, 2723, 6609. Retrieved from <https://shorturl.asia/HYmJT>
- Nurhuda, F., Sihwi, S. W., & Doewes, A. (2016). Analisis sentimen masyarakat terhadap calon Presiden Indonesia 2014 berdasarkan opini dari Twitter menggunakan metode Naive Bayes Classifier. *ITSmart: Jurnal Teknologi Dan Informasi*, 2(2), 35–42. <https://doi.org/10.20961/itsmart.v2i2.630>
- Rahmawati, L., Sihwi, S. W., & Suryani, E. (2016). Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus: Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret). *ITSMART: Jurnal Teknologi Dan Informasi*, 3(2), 66–73. Retrieved from <https://jurnal.uns.ac.id/itsmart/article/view/654>
- Safitri, L. A., & Dewa, C. B. (2022). Analisa kebijakan pemberlakuan PPKM pada masa pandemi Covid 19 terhadap industri skala kecil dan menengah makanan dan minuman. *Jurnal Manajemen DayaSaing*, 23(2), 97–107. Retrieved from <https://journals.ums.ac.id/dayasaing/article/view/16180>
- Shajari, S., & Agarwal, N. (2025). Safeguarding youtube discussions: a framework for detecting anomalous commenter and engagement behaviors. *Social Network Analysis and Mining*, 15(1), 54. <https://doi.org/10.1007/s13278-025-01470-7>
- Shevtsov, A., Oikonomidou, M., Antonakaki, D., Pratikakis, P., & Ioannidis, S. (2023). What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020. *Plos One*, 18(1), e0270542. <https://doi.org/10.1371/journal.pone.0270542>
- Skeppstedt, M., Ahlthorp, M., Kucher, K., & Lindström, M. (2024). From word clouds to Word Rain: Revisiting the classic word cloud to visualize climate change texts. *Information Visualization*, 23(3), 217–238. <https://doi.org/10.1177/14738716241236188>
- Sulastri, S., & Diartono, D. A. (2019). Analisa jejaring sosial twitter menggunakan klastering K-means dan hirarki agglomeratif. Retrieved from [https://www.unisbank.ac.id/ojs/index.php/sen-di\\_u/article/view/7400](https://www.unisbank.ac.id/ojs/index.php/sen-di_u/article/view/7400)
- Supriyadi, W. E. (2015). *Usulan Model Pengelompokan* (

- Clustering* ) dalam Pembahasan R KUHP 2015 Aliansi Nasional Reformasi KUHP', 2015. Retrieved from [https://reformasikuhp.org/usulan-model-pengelompokan-clustering-dalam-pembahasan-r-kuhp-2015\\_\\_trashed/](https://reformasikuhp.org/usulan-model-pengelompokan-clustering-dalam-pembahasan-r-kuhp-2015__trashed/)
- Widiastuti, T. W. (2017). Analisis elaboration likelihood model dalam pembentukan personal branding di Twitter. *Jurnal Aspikom*, 3(3), 588–603. <https://doi.org/10.24329/aspikom.v3i3.107>
- Wu, X., Chen, C., & Quan, L. (2024). Visual analysis and interactive interface design of students' abnormal behavior introducing clustering algorithm. *Technology and Health Care*, 32(6), 4947–4963. <https://doi.org/10.3233/THC-232054>
- Yahyadi, A., & Latifah, F. (2022). Analisis Sentimen Twitter Terhadap Kebijakan PPKM di Tengah Pandemi COVID-19 Menggunakan Mode LSTM. *Journal of Information System, Applied, Management, Accounting and Research*, 6(2), 464–470. <https://doi.org/10.52362/jisamar.v6i2.791>
- Yang, Z., Wang, C., Zhang, Z., & Li, J. (2019). Mini-batch algorithms with online step size. *Knowledge-Based Systems*, 165, 228–240. <https://doi.org/10.1016/j.knosys.2018.11.031>