INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS
ISSN(print): 2643-9840, ISSN(online): 2643-9875
Volume 05 Issue 12 December 2022
DOI: 10.47191/ijmra/v5-i12-17, Impact Factor: 6.261
Page No. 3423-3434

The Use of the K-Nearest Neighbor (KNN) Algorithm in Analyzing the Views of the Indonesian Society Regarding the Policy of Covid-19 Booster Vaccine

Yoannes Romando Sipayung[1], Agung Wibowo[2], Yoga Zaen Vebrian[3]
[1,2,3] Informatics Engineering, Universitas Ngudi Waluyo, Jalan Diponegoro No.186, Ngablak, Gedanganak, Kec. East Ungaran, Semarang Regency, Central Java 50512, Indonesia

ABSTRACT: At beginning of 2020, the world was shocked by the outbreak of the coronavirus from China. The coronavirus caused panic in China and killed thousands of Chinese people. Many people seemed to rely on social media for more information. As a result, social media platforms have become mediators between individuals and the rest of the world, even becoming the fastest-growing social applications. Among several well-known social media platforms, Twitter has gained special attention because users can easily broadcast information about their opinions on certain topics through public messages, called tweets. There was public anxiety about the Covid-19 vaccination. Public opinion and unrest are posted on Twitter. Based on this case, it is necessary to analyze the views or opinions of the public regarding the Covid-19 vaccination policy, especially regarding the Booster Vaccine type which can be categorized into positive or negative opinions circulating on Twitter. This research aims to analyze the views of the Indonesian people toward the Covid-19 Booster Vaccine using the K-Nearest Neighbor Algorithm (K-NN).

KEYWORDS: K-Nearest Neighbour, Twitter, Covid-19

I. INTRODUCTION

At the beginning of 2020, the world was shocked by the outbreak of the coronavirus from China. The coronavirus caused panic in China and killed thousands of Chinese residents. As another consequence, many small, medium and large companies had to temporarily close their businesses. In addition, thousands of food/beverage business establishments had to close (Bahtiar, 2021).

Based on these situations and conditions, WHO on March 11, 2020, declared Covid-19 a pandemic. Indonesia then followed WHO's steps by issuing Presidential Decree No.11 of 2020 concerning the Determination of Public Health Emergencies for Corona Virus Disease 2019 (Covid-19) and Presidential Decree Number 12 of 2020 concerning the Designation of Non-Natural Disaster for the Spread of Corona Virus Disease (Covid-19) as a National Disaster (Sipayung et al., 2020).

As the current situation is due to lockdowns in several parts of the world and the implementation of social distancing, the use of social media globally has increased. This is due to the success of connecting people from different geographical places and enabling them to exchange ideas. Besides, many people seem to rely on social media for more information. As a result, social media platforms have become mediators between individuals and the rest of the world, even becoming the fastest-growing social applications. Through this social media, people show different views, opinions, and emotions towards various events that have occurred as a result of the coronavirus pandemic (Yulita, Nugroho, et al., 2021).

Among the several well-known social media platforms, there is Twitter. Twitter is a type of microblogging social media that facilitates users to write and publish their activities and or opinions. Historically, with the presence and emergence of social media, Twitter provides a certain space or a maximum of 140 characters. Just like other social media, on Twitter, users can network with other users, disseminate information, promote the opinions/views of other users, discuss trending topics right away and become part of the issue by participating in tweeting using certain hashtags (Issn, n.d.).

Twitter has gained special attention because users can easily broadcast information about their opinions on certain topics through public messages, called tweets. In addition to information voluntarily offered by users, tweets can also store information related to the user's location and may contain links, emoticons, and hashtags that can help users better express their feelings and make them a valuable source of information (Yulita, Nugroho, et al., 2021). Due to the widespread use of social media, especially Twitter, people are more likely to express their opinions on social media.

The Use of the K-Nearest Neighbor (KNN) Algorithm in Analyzing the Views of the Indonesian Society Regarding the Policy of Covid-19 Booster Vaccine

Vaccination against COVID-19 generated new questions regarding the relatively short time required for vaccine development. As it is known, the vaccine development process usually takes a decade. The fastest previous vaccine development was four years in the case of the mumps vaccine. Even nearly forty years after the discovery of HIV, no effective vaccine has yet been developed. However, the time to develop a vaccine for COVID-19 was very short due to the state of emergency. As of December 18, 2020, the COVID-19 Vaccine Tracker website, hosted by the Milken Institute, indicated that 236 vaccines were in development, 38 were in clinical trials and 7 had reached regulatory decisions. However, on 8 December 2020 the first vaccine was administered in the UK (Yulita, Nugroho, et al., 2021).

There was public anxiety about the Covid-19 vaccination. Public opinion and unrest are written on Twitter. Based on this case, it is necessary to analyze the views or opinions of the public regarding the Covid-19 vaccination policy, especially regarding the Booster Vaccine type which can be categorized into positive or negative opinions circulating on Twitter.

Previous research was carried out under the title "*Analisis Sentimen Menggunakan K-Nearest Neighbour (K-NN) Terhadap New Normal Masa Covid-19 di Indonesia*" (Furqan et al., 2022). Sentiment analysis in this research was to predict comments or public opinions that tend to have positive or negative opinions. From the results of this research was concluded that the K-Nearest Neighbor (K-NN) classification using the 80% percentage split test was a more optimal test compared to other tests with 80% training data and 20% test data because it also had a high accuracy value.

The second previous research was research conducted by Winda Yulita, Eko Dwi Nugroho, and Muhammad Habib Algifari, with the title "*Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier*" (Yulita, Dwi Nugroho, et al., 2021).

The K-Nearest Neighbor algorithm is a method for classifying objects based on learning data that is closest to the object. KNN is a supervised learning algorithm where the results of the new query instance are classified based on the majority of the categories in the KNN algorithm. The class that appears the most will later become the class resulting from the classification (T et al., 2017).

Therefore, this research takes the title "The Use of the K-Nearest Neighbor (KNN) Algorithm in Analyzing Indonesian People's View Sentiments Regarding the Policy of Covid-19 Booster Vaccine ". Based on the description of this background, in this research, the formulation of the problem of formulated, it is: How to analyze the views of the Indonesian people towards the Covid-19 Booster Vaccine?

II. RESEARCH METHOD

The method used in this research was through several stages starting from data collection to the data testing stage through the previously created GUI. These stages were applied to a flow chart or what is often called a flowchart and this is be described as follows:

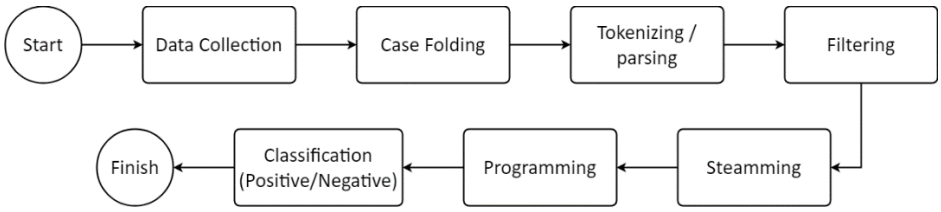

Figure 1. Research Flowchart

III. RESULTS AND DISCUSSION

A. Data Collection

The data were taken from Twitter social media, in which these comments were commonly referred to as "Tweet". The tweets submitted were a representation of the written language. Language in communication has a very huge impact. Language has an impact on human life. Subyantoro stated that language is basically integrated with human life. Humans convey ideas, concepts, thoughts, hopes, and desires through language. The use of language has various interests and functions of each. Language is used by humans for the benefit of education, culture, religion, and others (Permatasari & Subyantoro, 2020).

The data was stored in Ms. Excel table containing comment URL links, commenter account names, and the contents of comments taken since the vaccine was distributed in January 2022. The following is a sample data collection:

Table 1. Sample data collection

No.	Account's name	Fill the Comments (Tweet)
1	matahari_me	Tadi pagi mengantar Bapak ke sentra vaksin booster Unika Atma Jaya. Relawannya banyak dan ramah. Pesertanya juga banyak. Tapi semua tertib dan prosesnya berjalan dgn cepat 👏👍Terima kasih!
2	atiekkustia	Alhamdulillah, aku udah vaksin booster kemaren 18/1/2022 Temen² ayo vaksin yuukk… bantu pemerintah perangi Covid 19/ Omicron 😊👍
3	naevyblu	apakabar putusan MA yang menyatakan harus pakai vax halal min? oia lupa negara ini mah bebas aja ya ketika badan resmipun melanggar hukum. Selama penyuntik modal bilang lanjut, terabas aja 👏
4	Yuyuns152	Kasian masyarakat… Abis vaksin 1.. harus vaksin 2 biar tambah kuat imunnya, abis vaksin 2, 6 bulan imunnya turun, harus vaksin 3/boster… Ada ada saja, kebohongan itu bukan menyelesaikan masalah, tapi a

B. Case Folding

Case folding is the process of converting all letters in a document to lowercase. Only the letters "a" to "z" is accepted. Characters other than letters are omitted and considered delimiter (exceeding the limit) *(Yoannes Romando et al., 2019).* The following table shows samples of *case folding:*

Table 2. Case Folding Sample

No.	Account's name	Fill the Comments (Tweet)	Case Folding
1	matahari_me	Tadi pagi mengantar Bapak ke sentra vaksin booster Unika Atma Jaya. Relawannya banyak dan ramah. Pesertanya juga banyak. Tapi semua tertib dan prosesnya berjalan dgn cepat 👏👍Terima kasih!	tadi pagi mengantar bapak ke sentra vaksin booster unika atma jaya relawannya banyak dan ramah pesertanya juga banyak tapi semua tertib dan prosesnya berjalan dgn cepat terima kasih
2	atiekkustia	Alhamdulillah, aku udah vaksin booster kemaren 18/1/2022 Temen² ayo vaksin yuukk… bantu pemerintah perangi Covid 19/ Omicron 😊👍	alhamdulillah aku udah vaksin booster kemaren temen ayo vaksin yuukk bantu pemerintah perangi covid omicron
3	naevyblu	apakabar putusan MA yang menyatakan harus pakai vax halal min? oia lupa negara ini mah bebas aja ya ketika badan resmipun melanggar hukum. Selama penyuntik modal bilang lanjut, terabas aja 👏	apakabar putusan ma yang menyatakan harus pakai vax halal min oia lupa negara ini mah bebas aja ya ketika badan resmipun melanggar hukum selama penyuntik modal bilang lanjut terabas aja
4	Yuyuns152	Kasian masyarakat… Abis vaksin 1.. harus vaksin 2 biar tambah kuat imunnya, abis vaksin 2, 6 bulan imunnya turun, harus vaksin 3/boster… Ada ada saja, kebohongan itu bukan menyelesaikan masalah, tapi a	kasian masyarakat abis vaksin harus vaksin biar tambah kuat imunnya abis vaksin bulan imunnya turun harus vaksin boster ada ada saja kebohongan itu bukan menyelesaikan masalah tapi a

C. Tokenizing/parsing

Tokenization is the process of separating a row of words in a sentence, paragraph, or page into tokens of single word pieces or termed words that stand alone (Robinson, 2019). Tokenizing sample is shown in the following table:

Table 3. Tokenizing Sample

No.	Account's name	Case Folding	Tokenizing / parsing
1	matahari_me	tadi pagi mengantar bapak ke sentra vaksin booster unika atma jaya relawannya banyak dan ramah pesertanya juga banyak tapi semua tertib dan prosesnya berjalan dgn cepat terima kasih	tadi pagi mengantar bapak ke sentra vaksin booster unika atma jaya relawannya banyak dan ramah pesertanya juga banyak tapi semua tertib dan prosesnya berjalan dgn cepat terima kasih
2	atiekkustia	alhamdulillah aku udah vaksin booster kemaren temen ayo vaksin yuukk bantu pemerintah perangi covid omicron	mulai dr tgl mei kenaikannya diatas an hayo yg blm vaksin or booster bantu pemerintah dan

			org tercinta dikeluargamu atau saudara jg temanmu cegah covid ur healthy is your priority too
3	naevyblu	apakabar putusan ma yang menyatakan harus pakai vax halal min oia lupa negara ini mah bebas aja ya ketika badan resmipun melanggar hukum selama penyuntik modal bilang lanjut terabas aja	apakabar putusan ma yang menyatakan harus pakai vax halal min oia lupa negara ini mah bebas aja ya ketika badan resmipun melanggar hukum selama penyuntik modal bilang lanjut terabas aja
4	Yuyuns152	kasian masyarakat abis vaksin harus vaksin biar tambah kuat imunnya abis vaksin bulan imunnya turun harus vaksin boster ada ada saja kebohongan itu bukan menyelesaikan masalah tapi a	kasian masyarakat abis vaksin harus

			vaksin biar tambah kuat imunnya abis vaksin bulan imunnya turun harus vaksin boster ada ada saja kebohongan itu bukan menyelesaikan masalah tapi a

D. Filtering/Stopword Removal

Stopword Removal is a filtering process, selecting important words from the token results, namely what words are used to represent documents (Anwar et al., 2019). The filtering sample is shown below:

Table 4. Filtering Sample

No.	Account's name	Tokenizing / parsing
1	matahari_me	tadi pagi mengantar bapak ke sentra vaksin booster unika atma jaya relawannya banyak dan ramah pesertanya juga banyak tapi semua tertib

		dan prosesnya berjalan dgn cepat terima kasih
2	atiekkustia	mulai dr tgl mei kenaikannya diatas an hayo yg blm vaksin or booster bantu pemerintah dan org tercinta dikeluargamu atau saudara jg temanmu cegah covid ur healthy is your priority too
3	naevyblu	apakabar putusan ma yang menyatakan harus pakai vax halal min oia lupa

		negara ini mah bebas aja ya ketika badan resmipun melanggar hukum selama penyuntik modal bilang lanjut terabas aja
4	Yuyuns152	kasian masyarakat abis vaksin harus vaksin biar tambah kuat imunnya abis vaksin bulan imunnya turun harus vaksin boster ada ada saja kebohongan itu bukan menyelesaikan masalah tapi

E. Stemming

Stemming is the process of obtaining the root/stem or base word of a word in a sentence by separating each word from the base word and its affixes, both prefixes and suffixes (Wahyudi et al., 2017). The following table shows the stemming sample:

Table 5. Stemming Sample

No.	Account's name	Filtering	Stemming
1	matahari_me	pagi	pagi
		mengantar	antar
		bapak	bapak
		vaksin	vaksin
		booster	booster
		unika	unika
		atma	atma
		jaya	jaya
		relawannya	relawan
		banyak	banyak
		ramah	ramah
		pesertanya	peserta
		banyak	banyak
		tertib	tertib
		prosesnya	proses
		berjalan	jalan
		cepat	cepat
		terima	terima
		kasih	kasih
2	atiekkustia	alhamdulillah	alhamdulillah
		udah	udah
		vaksin	vaksin
		booster	booster
		kemaren	kemaren
		ayo	ayo
		vaksin	vaksin
		bantu	bantu
		perangi	perangi
		covid	covid
		omicron	omicron
3	naevyblu	apakabar	apakabar
		putusan	putusan
		ma	ma
		menyatakan	nyata
		pakai	pakai
		vax	vax
		halal	halal
		lupa	lupa
		negara	negara
		bebas	bebas
		badan	badan
		ketika	ketika
		resmipun	resmi
		melanggar	langgar
		hukum	hukum
		penyuntik	suntik
		modal	modal
		bilang	bilang

		lanjut terabas	lanjut terabas
4	Yuyuns152	kasian masyarakat harus vaksin biar kuat imunnya abis vaksin imunnya turun harus boster ada saja kebohongan bukan menyelesaikan masalah	kasian masyarakat harus vaksin biar kuat imun abis vaksin imun turun harus boster ada saja bohong bukan selesai masalah

F. Programming

In this research, the Matlab program was used. Matlab (Matrix Laboratory) is a software developed by Mathworks, Inc. by utilizing the matrix in its use (Atina, 2019). Besides Matlab, Microsoft Excel was also used as an implementing tool or a place to process data, as well as using the KNN algorithm as a decision-making system. When it entered the initial appearance of Matlab, it contained related tools and interfaces. Each program has its own uniqueness in using its tools, such as in making a GUI. Whereas in the Matlab program, to be able to bring up a GUI, the first thing to do was to type guide in the Command Window.

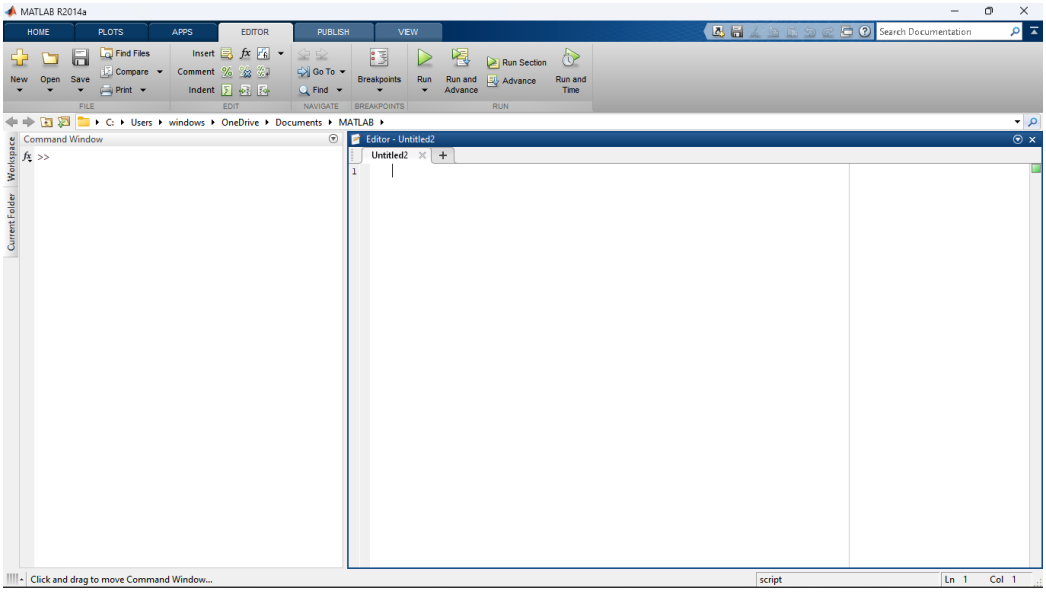

Figure 2. Display of Matlab Interface

Before inputting the KNN algorithm, there was 1 (one) file in .mat format or a Matlab built-in file that functioned as a variable, then 2 (two) Excel files with functions as test data and training data that had to be created. Thus, the program with the KNN algorithm could run properly. Variables with the .mat file format can be added by selecting the New Variable menu, then a table would appear that could be filled in so that later each point on the diagram obtained its own coordinates. These variables were obtained from the KNN formula which is explained as follows:

$$dis\ (x_1, x_2) = \sqrt{\sum_{i=0}^{n} (x_{1i} - x_{2i})^2}$$

Table 7. Test Data Sample

No.	Account's name	Fill the Comments (Tweet)	Cluster
1	upk_kemenkes	Semangat memutus rantai penyebaran dengan mengikuti vaksinasi Covid-19 dan disiplin protokol kesehatan. Salam sehat ❤️	1
2	GrangerKeren	Saya & keluarga tidak vaksin covid19, tetap sehat sampai sekarang. No Vaksin Covid, badan sehat dan bugar 💪.	2
3	black_granger23	Mari yang belum dapat vaksin, segera vaksin	1
4	matahari_me	Tadi pagi mengantar Bapak ke sentra vaksin booster Unika Atma Jaya. Relawannya banyak dan ramah. Pesertanya juga banyak. Tapi semua tertib dan prosesnya berjalan dgn cepat 👏👍Terima kasih!	1

G. Classification

The classification process was carried out using the KNN method as the algorithm. The results of the classification method were to provide a conclusion that the input comments contained negative or positive elements. It followed the test data as the key to its classification. In the process of testing the method, the results of the classification of comments on Twitter were as expected by researchers. This statement was proven by the image below after entering several comments from the platform and checking. The result obtained from the system was to notify the nature of the comment, whether it was negative or positive. The following shows some of the results.

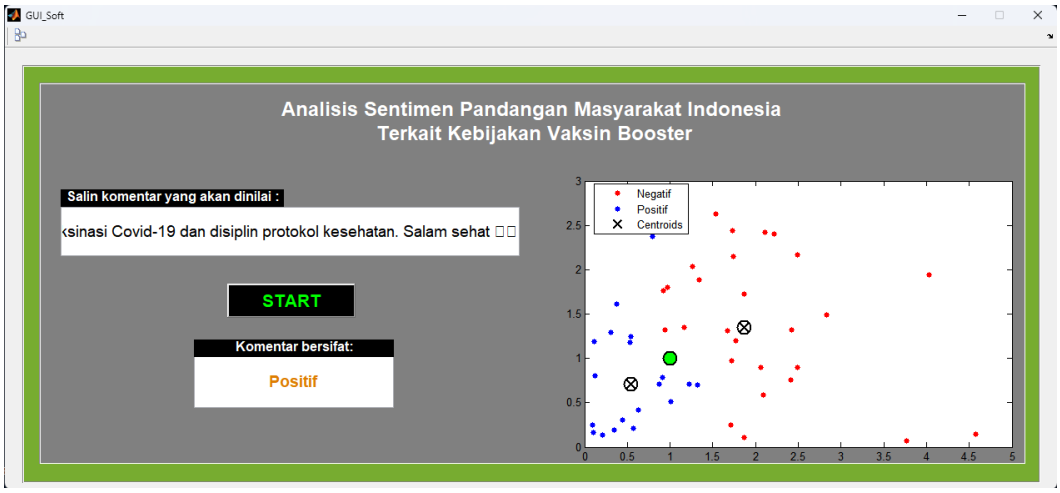

Figure 3. Display of Positive Comment Results

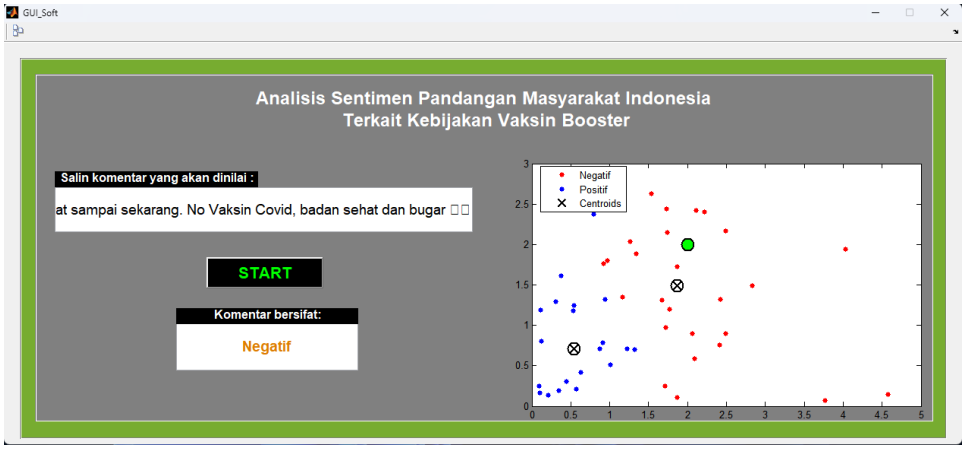

Figure 4. Display of Negative Comment Result

IV. CONCLUSIONS

According to the research results that have been analyzed with a total of 50 tweet comments, it is found that there are 37 samples with an accuracy of 74% comments on Twitter that are positive (support/agree with the Covid-19 booster vaccine policy). In addition, 13 samples with an accuracy of 26% comments on Twitter are negative (in this case it means less support for the Covid-19 booster vaccine policy). For future research, it is expected that they can analyze sentiment in a comment using English or other languages and add more data. therefore, the results obtained are even better and the level of accuracy can be higher.

ACKNOWLEDGMENT

We gratefully acknowledge that our research was financially supported by Universitas Ngudi Waluyo.

REFERENCES

1) Anwar, M. S., Subroto, I. M. I., & Mulyono, S. (2019). Sistem Pencarian E-Journal Menggunakan Metode Stopword Removal Dan Stemming. *Prosiding KONFERENSI ILMIAH MAHASISWA UNISSULA (KIMU) 2*, 58–70. http://lppm-unissula.com/jurnal.unissula.ac.id/index.php/kimueng/article/viewFile/8420/3887
2) Atina, A. (2019). Aplikasi Matlab pada Teknologi Pencitraan Medis. *Jurnal Penelitian Fisika Dan Terapannya (JUPITER)*, *1*(1), 28. https://doi.org/10.31851/jupiter.v1i1.3123
3) Bahtiar, R. A. (2021). Dampak Pandemi Covid-19 Terhadap Sektor Usaha Mikro, Kecil, dan Menengah Serta Solusinya. *Pusat Penelitian Badan Keahlian DPR RI, Bidang Ekonomi Dan Kebijakan Publik*, *XIII*(10), 19–24. https://berkas.dpr.go.id/puslit/files/info_singkat/Info Singkat-XIII-10-II-P3DI-Mei-2021-1982.pdf
4) Furqan, M., Sriani, S., & Sari, S. M. (2022). Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia. *Techno.Com*, *21*(1), 51–60. https://doi.org/10.33633/tc.v21i1.5446
5) Issn, J. H. (n.d.). *Pengaruh Penggunaan Media Sosial Twitter Terhadap Sikap Fanatisme Penggemar (Studi Pada Media Sosial Twitter @BTOBIndonesia Terhadap Sikap Fanatisme Penggemar)*. 120–133.
6) Permatasari, D. I., & Subyantoro, S. (2020). Ujaran Kebencian Facebook Tahun 2017-2019. *Jurnal Sastra Indonesia*, *9*(1), 62–70. https://doi.org/10.15294/jsi.v9i1.33020
7) Robinson, L. (2019). Implementasi Metode Generalized Vector Space Model Pada Aplikasi Information Retrieval untuk Pencarian Informasi Pada Kumpulan Dokumen Teknik Elektro Di UPT BPI LIPI. *Jurnal Ilmiah Komputer Dan Informatika*.
8) Sipayung, Y. R., Cahya, K., & Wibawa, S. (2020). The Role of Information Technology toward the Development of Online Learning Process in Covid-19 Pandemic Period in Indonesia. *PalArch's Journal of Archaeology of Egypt / Egyptology*, *17*(4), 948–958.
9) T, A. J., Yanosma, D., & Anggriani, K. (2017). Implementasi Metode K-Nearest Neighbor (Knn) Dan Simple Additive Weighting (Saw) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka. *Pseudocode*, *3*(2), 98–112. https://doi.org/10.33369/pseudocode.3.2.98-112
10) Wahyudi, D., Susyanto, T., & Nugroho, D. (2017). Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia. *Jurnal Ilmiah SINUS*, *15*(2), 49–56. https://doi.org/10.30646/sinus.v15i2.305
11) Yoannes Romando, Sulistyowati, R., & Wibisono, I. S. (2019). Identifikasi Komentar Negatif Berbahasa Indonesia Pada Instagram Dengan K-Means. *Multimatrix*, *II*(1), 6–8.
12) Yulita, W., Dwi Nugroho, E., Habib Algifari, M., Studi Teknik Informatika, P., Teknologi Sumatera, I., Terusan Ryacudu, J., Huwi, W., Agung, J., & Selatan, L. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining Dan Sistem Informasi*, *2*(2), 1–9. https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1344
13) Yulita, W., Nugroho, E. D., & Algifari, M. H. (2021). Sentiment Analysis on Public Opinion About the Covid-19 Vaccine Using the Naïve Bayes Classifier Algorithm. *Jdmsi*, *2*(2), 1–9.