JOURNAL OF DATA SCIENCE | Vol.2023:08
eISSN:2805-5160

Sentiment Analysis of Omicron COVID-19 Variant using Naïve Bayes Classifier and RapidMiner

Dayu Wijaya[1&2], Leon A. Abdillah[1,2*]

[1]Dept. of Information Systems, Universitas Bina Darma, Indonesia
[2]Research Group DeL-MOOC, Universitas Bina Darma, Indonesia

[*]**Email**: leon.abdillah@yahoo.com

Abstract

The Coronavirus or other designations is COVID-19 (Corona Virus Disease) appeared in November 2019 in Wuhan, China. Over time, the virus is no longer categorized as an outbreak but is categorized as a pandemic or has spread to almost all countries in the world, including Indonesia. The emergence of COVID-19 in Indonesia in February 2020 has resulted in many sectors experiencing losses, not only in health but also in the economic sector. Recently there was a new mutation to the COVID-19 Virus, namely Omicron. Omicron has been shown to be much more infectious than the other variants with an increased ability to evade vaccines and cause re-infection. This study aims to present a result of sentiment analysis on the new variant of the COVID-19 Virus, namely Omicron which is divided into three (three) classes: positive, negative, and neutral. Then, the comments will be manually labeled followed by classification using the Nave Bayes algorithm and RapidMiner software. This study's findings revealed that 84% of the community responded positively, 7% of the community responded Neutral and 9% of the community responded negatively. It can be concluded that the community responded positively to the issue of the latest variant of the COVID-19 Omicron virus because there is also the possibility that the contents of the latest Omicron COVID-19 virus may also be dangerous from the beginning of the emergence of the COVID-19 Virus in the world.

Keywords

COVID-19, Naïve Bayes, Omicron, Rapid Miner, Sentiment Analysis

Introduction

The trend of the development of information technology (IT) and its applications has brought a number of major changes (Abdillah, Mukti, et al., 2021). People's activities are made easier by the presence of a number of IT-based applications. For example, through the use of the Internet (Abdillah et al., 2023), many transactions can be made in paperless, electronic-based, and real-time modes. One of the emerging trends in internet is social media (Abdillah, 2022). The more widespread use of social media, such as making social media a source of information that is in

Submission: 3 August 2023; **Acceptance:** 24 August 2023

great demand by the Indonesian people (Guspita & Abdillah, 2021). To share opinions, many people use the social medical platform Twitter. This is because Twitter is a good resource for investigating public opinion (Bokaee Nezhad & Deihimi, 2022). Twitter (Kwak & Cho, 2018), via the Streaming application programming interface (API), provides an overview of its hundreds of millions of users and billions of tweets providing examples of all tweets that match a number of parameters from that API user. The use of Twitter is increasingly widespread with access via smartphones which allows people to write their tweets on a mobile (Abdillah & Kurniasti, 2022) basis.

Opinion mining, also known as sentiment analysis, is a computational examination of people's opinions, sentiments, and emotions conveyed in the form of text entities or attributes. (Aditya, 2015). Sentiment analysis is a science that is useful for analyzing one's opinion, one's sentiment, one's assessment, attitude, and emotions into written words. The primary goal of sentiment analysis (Ling et al., 2014) is to categorize the polarity of text in a document, sentence, or paragraph, or opinion sourced from online media. Polarity means whether a good, negative, or neutral feature of language in a document, sentence, or viewpoint, negative or neutral aspect. In this context, negative opinion refers to comments or opinions that might provoke hatred, shame, controversy, and arguments in cyberspace. Positive opinion is a positive statement or opinion that does not provoke hatred, embarrassment, controversy, or problems in cyberspace. While neutral is a word that does not contain negative or positive.

The method for analyzing the sentiment of the researcher uses the Naïve Bayes algorithm. Naïve Bayes is a probabilistic model machine learning technique (Wahyu Handani et al., 2019). Naïve Bayes is a simple algorithm with great text classification performance and accuracy (Routray et al., 2013).

RapidMiner is software that supports an easier way to load data equipped with a Graphical User Interface (GUI) Window for selecting data sets (Baharun et al., 2022). RapidMiner is used to examine data quantitatively and qualitatively in order to gain expected information and knowledge. (Uska et al., 2020).

At the end of 2019 and early 2020, the world began to be hit by an outbreak that became a global pandemic known as COVID-19. COVID-19 is caused by the coronavirus SARS-CoV-2, also known as the 2019 new coronavirus (2019-nCoV) (Lai et al., 2020), and was declared a global pandemic in March 2020 (WHO, 2020). The SARS-CoV-2 virus-caused the COVID-19 pandemic has had a tremendous impact on global health, economics, and social systems. The virus has undergone different changes over time, resulting in the creation of new varieties with possibly changing features. Recently, we were surprised again by the discovery of a novel COVID-19 viral variant, namely B.1.1.529 or Omicron, which was first discovered in South Africa on November 24, 2021. The Omicron variety has an extraordinary the total number of spike protein mutations, raising concerns about its transmissibility, pathogenicity, and ability to escape immunization. The Omicron variant is the latest variant which is currently a fully the majority are found in various countries. Thus, the World Health Organization (WHO) has classified it as a "variant of concern" (Table 1).

Understanding public sentiment toward the Omicron variety is critical in the face of this evolving dilemma. Social media platforms, news stories, and online forums are major sources of people's real-time ideas and emotions about current events. Sentiment analysis, a natural language processing technique, is crucial in assessing and comprehending this massive amount of textual data, assisting in gauging public views, anxieties, and reactions to the Omicron variation.

Table 1. COVID-19 Variants of Concern (VOC)

WHO Label & Pango Lineage*	Initial Documentation Sample	Appointment date
Alpha (B.1.1.7)	2020 Sept., United Kingdom	Dec. 18th, 2020
Beta (B.1.351)	2020 May, South Africa	Dec. 18th, 2020
Gamma (P.1)	2020 November, Brazil	Jan 11th, 2021
Delta (B.1.617.2)	2002 October, India	VOI: Apr. 4th, 2021 VOC: May 11th, 2021
Omicron* (B.1.1.529)	2021 November, Several Countries	VUM: Nov. 24th, 2021 VOC: Nov. 26th, 2021

Source : https://covid19.go.id/id

Data Mining can be used for several disciplines (Sabarmathi & Chinnaiyan, 2017), namely: 1) Database Technology, 2) Information Science, 3) Statistics, 4) Machine Learning, 5) Visualization, and 6) Others Disciplines. A number of studies have addressed the topic of sentiment analysis, such as: 1) The Machine Learning Method was used to analyze public sentiment against Joko Widodo in relation to the COVID-19 outbreak (Hikmawan et al., 2020), 2) An Examination of the Benefits and Drawbacks of Indonesian Society Sentiments on the COVID-19 Vaccine on Twitter (Rachman & Pramana, 2020), 3) Public Opinion Sentiment Analysis on Twitter Regarding COVID-19 Using the Naïve Bayes and KNN Methods (Syarifuddinn, 2020), 4) Sentiment Analysis of Public Opinion on COVID-19 Using the Naïve Bayes Classifier Algorithm (Yulita et al., 2021), and 5) Application of Artificial Neural Network Algorithm for Classification of Public Opinion Against COVID-19 (Saraswati et al., 2021).

None of these studies have reviewed sentiment related to COVID-19 for the Omicron variant. This study will analyze the sentiment towards the Omicron variant COVID-19 using the NBC method and RapidMiner data mining software. This paper examines sentiment analysis of the Omicron COVID-19 variant using the Naive Bayes Classifier (NBC) and the data science tool RapidMiner. The Naive Bayes Classifier is a well-known machine learning algorithm noted for its ease of use, efficacy, and speed when dealing with text classification tasks. RapidMiner, on the other hand, is a popular data science platform that allows for data pretreatment, modeling, and assessment, making it a great companion for our sentiment analysis work.

The remainder of this article is structured as follows: Section 2 describes the methods used, including data collection, preparation, and the construction of the Naive Bayes Classifier in RapidMiner. Section 3 shows the findings of the sentiment analysis and examines their consequences. Finally, Section 4 summarizes the study's contributions and makes recommendations for further research.

Methodology

Dataset

We began on a rigorous data-driven journey to elucidate the attitudes surrounding the Omicron version. Our research includes the collecting of textual data from the popular social media platform Twitter. Web scraping techniques and APIs were used to capture a corpus of debates, capturing

the digital population's collective voice. The data set was taken by crawling comments from the public about the development of the COVID-19 virus, especially the latest variant, Omicron, from social media Twitter. Furthermore, the data set will be analyzed to find out what the public opinion is about the development of the virus.

Method of Data Collection

Method of Data Collection is a way to get the data needed in carrying out a research (Abdillah, HS, et al., 2021). This research involves 3 (three) kinds of data collection methods, namely: 1) Observation. This procedure is carried out by immediately observing the development of social media about the latest virus, namely the Omicron COVID-19 Variant, 2) Literature Study. Searching and studying data from books or other references relevant to the drafting of research proposal reports to collect data. The book used by the author as a reference, while the methods used by the author in designing and developing can be seen in the bibliography, and 3) Crawl. This method is done by creating a script to retrieve comment data on tweets about COVID-19, omicron then later this data will be managed to determine public sentiment which is classified as Positive, Neutral and Negative sentiment.

Knowledge Discovery in Database (KDD)

This research method has a design of how the flow of this research system will run. In this study, researchers used the Knowledge Discovery in Database (KDD) method (Fayyad et al., 1996). In the following, the researchers describe the research based on the KDD stages as shown in figure 1.

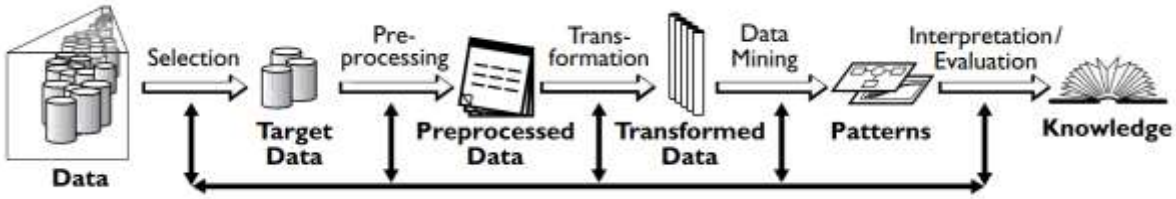


Figure 1. The Database Knowledge Discovery Process.

The first step is to select data. At this stage, the data that has been taken is managed using the Crawling technique on the Twitter API key. The advantage of this twitter is the availability of excellent API (Application Programming Interface) facilities, making it easier for users to obtain data from social media (Putra et al., 2020). The data used in this study is the data posted by Twitter users on the Twitter.com site. The results of the data crawling process will be carried out by a data labeling process to determine the classification of opinions or views from the tweets that have been crawled earlier. The labeling process is divided into 3 (three) classes, namely: 1) Positive class, 2) Negative class, and 3) Neutral class.

The second stage is Pre-processing. Preprocessing is the stage of the process to purge data from words or tweets that are not needed and words that have no meaning. This process (Hasibuan et al., 2023) is carried out according in relation to the data content of the data retrieval process or Twitter data crawling. The process of several steps of the preprocessing process has the following sequence: 1) Case Folding, 2) Cleaning Text, 3) Feature Normalization, 4) Tokenization, 5) Stop-word Removal, and 6) Stemming.

The third stage is Transformation. In this case, after all tweets are collected from the crawling results and the data results have gone through the preprocessing process, the next process is to create useful features to simplify the process of classifying the tweet data, usually this process is made with a feature extraction process, classification, and model testing. Term frequency-Inverse document frequency (TF-IDF) is employed in the extraction sentences by assigning values or weights to sentences. TF-IDF is used to examine the relevance of keywords to documents in the corpus (Qaiser & Ali, 2018). The TF and IDF values are multiplied to obtain the TF-IDF weight value (Widiastuti et al., 2017).

$$tf = 0.5 + 0.5\ x \frac{tf}{\max(tf)} \quad (1)$$

$$idft = \log\left(\frac{D}{dft}\right) \quad (2)$$

$$Wd.t = tfd.t\ x\ IDfd.t \quad (3)$$

The fourth stage is Data Mining. At this stage, apply the NBC algorithm or the classification search method from sentiment results that are already in the transformation stage into the form of Naïve Bayes Classification algorithm analysis using RapidMiner software. One of the better classification methods for implementing Bayesian learning is Naïve Bayes. Its capabilities is comparable to a decision tree algorithms and neural networks in some other applications, but the computational complexity is much less than other algorithms (Ning et al., 2019).

$$P(H|X) = \frac{\mathrm{P(X|H).P(H)}}{\mathrm{P(H)}} \quad (4)$$

The fifth stage is Interpretation/Evaluation. The researcher will at this point get results based on the info that has been collected or processed in the first stage of Data Collection to the Design stage, the final results regarding the level of accuracy, precision and confusion matrix.

RapidMiner

RapidMiner is a software platform that combines machine learning, text mining, predictive analytics, and business analytics are all examples of analytics. (J. Arunadevi et al., 2018). RapidMiner is commonly used in business, commercial applications, research, and education, and training needs because of its speed in prototyping and supporting all machine learning processes data preparation, visualization, validation, and optimization (Kori, 2017).

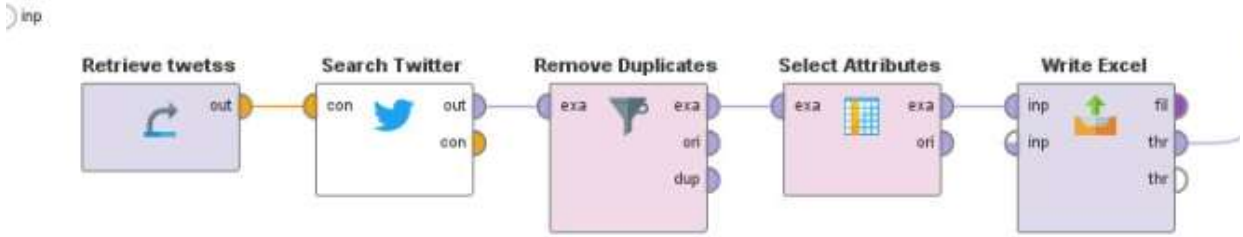
Results and Discussion

Crawled Data

The dataset uses crawling from community comments on Twitter with the keyword "Covid Omicron". The results obtained from the crawling process using RapidMiner software will be used in the preprocessing process. The results of the data crawling process (figure 2.a), using the Twitter

API Key, obtained data with a total of 1608 Tweet records, provided that the training data for tweet documents was 80% and the training data for tweet documents was 20%. After the crawling process by using RapidMiner, the results are as shown in figure 2b.

The results of the data crawling process will be carried out by a data labeling process to determine the classification of opinions or views from the tweets that have been crawled earlier. In this labeling process, it is divided into 3 (three) classes, namely: 1) Positive class, 2) Negative class, and 3) Neutral class.



(a)

Row No.	Id	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source
1	1569793353...	Sep 14, 2022...	Atila Iamarino	13769472	?	-1	pt	<a href="http...
2	1569857653...	Sep 14, 2022...	Alejandro Ma...	316273207	?	-1	es	<a href="http:/...
3	1569341165...	Sep 12, 2022...	EU Medicine...	116716199	?	-1	en	<a href="http...
4	1569944555...	Sep 14, 2022...	Daytheiyown ...	1448856908...	?	-1	en	<a href="http:/...
5	1569944288...	Sep 14, 2022...	Covid-19 / C...	1241102066...	?	-1	en	<a href="http...
6	1569944260...	Sep 14, 2022...	唯真社	1546712391...	?	-1	zh	<a href="http:/...
7	1569944135...	Sep 14, 2022...	Dr. Alon Altm...	17206784	?	-1	en	<a href="http:/...
8	1569944113...	Sep 14, 2022...	路见不平（费...	8829842213...	?	-1	zh	<a href="http:/...
9	1569944018...	Sep 14, 2022...	Matia Guadal...	9406821795...	?	-1	es	<a href="http:/...
10	1569943969...	Sep 14, 2022...	唯真社	1546712391...	?	-1	zh	<a href="http:/...
11	1569943961...	Sep 14, 2022...	唯真社	1546712391...	?	-1	zh	<a href="http:/...
12	1569943802...	Sep 14, 2022...	Hiram Rogelio	2648441282	?	-1	es	<a href="http:/...
13	1569943674...	Sep 14, 2022...	Marie	1520668762...	?	-1	fr	<a href="http:/...
14	1569943602...	Sep 14, 2022...	Annabelle Mu...	1541493893...	?	-1	und	<a href="http...

ExampleSet (1,000 examples, 1 special attribute, 11 regular attributes)

(b)

Figure 2. Twitter Crawled Data Results.

Pre-processing Results

Preprocessing is the stage of the process to clean data from words or tweets that are not needed and words that have no meaning. This process is carried out according to the data content of the data retrieval process or Twitter data crawling. The process of several steps of the preprocessing process has the following sequence: 1) Case Folding, 2) Cleaning Text, 3) Feature Normalization, 4) Tokenization, 5) Stop-word Removal, and 6) Stemming. The example of one of the tweets “Terima Kasih kepada staff medis yang membantu merawat pasien corona agar bisa

sembuh. COVID19 COVID19 virus corona :(#SaveIndonesia #Corona". At this stage case folding is done by utilizing the transform case operator on RapidMiner.

Table 2. Pre-processing Results

No	Phase	Before	After
1	Case Folding	Terima Kasih kepada staff medis yang membantu merawat pasien corona agar bisa sembuh. COVID19 COVID19 virus corona :(#SaveIndonesia #Corona	terima kasih kepada staff medis yang membantu merawat pasien corona agar bisa sembuh. covid19 covid19 virus corona :(#saveindonesia #corona
2	Cleaning Text	terima kasih kepada staff medis yang membantu merawat pasien corona agar bisa sembuh. covid19 covid19 virus corona :(#saveindonesia #corona	[terima] [kasih] [kepada] [staff] [medis] [yang] [membantu] [merawat] [pasien] [corona] [agar] [bisa] [sembuh][covid19] [covid19] [virus] [corona] [:][( ]
3	Feature Normalization	terima kasih kepada staff medis yang membantu merawat pasien corona agar bisa sembuh. covid19 covid19 virus corona :(#saveindonesia #corona	[terima] [kasih] [kepada] [staff] [medis] [yang] [membantu] [merawat] [pasien] [corona] [agar] [bisa] [sembuh][covid19] [covid19] [virus] [corona] [:][( ]
4	Stopword Removal & Tokenization	[terima] [kasih] [kepada] [staff] [medis] [yang] [membantu] [merawat] [pasien] [corona] [agar] [bisa] [sembuh][covid19] [covid19] [virus] [corona] [:][( ]	[staff] [medis] [membantu] [merawat] [pasien] [corona] [sembuh][covid19] [covid19] [virus] [corona]
5	Stemming	[staff] [medis] [membantu] [merawat] [pasien] [corona] [sembuh][covid19] [covid19] [virus] [corona]	[staff] [medis] [bantu] [rawat] [pasien] [corona] [sembuh][covid19] [covid19] [virus] [corona]

Naïve Bayes Analysis

The Nave Bayes classifier, a text classification stalwart, is at the heart of our effort to decipher the attitudes surrounding the Omicron COVID-19 variation. We used this algorithm to distill the deep emotional tapestry weaved inside the textual corpus, leveraging its simplicity and efficacy. The Nave Bayes classifier is based on the principles of Bayes' theorem and is based on the "naive" assumption of feature independence, which is generally true for text analysis, making it especially well-suited for our sentiment classification problem.

The Nave Bayes Classifier is used in the classification process algorithm it is split into two (two) processes, namely: 1) Training, and 2) Testing. At the training stage it is utilized to generate a sentiment analysis model that will subsequently be used as a reference for classification. sentiment with data testing. After previously carrying out the separation process between data for training and testing, then the data is processed again using the RapidMiner application with the operator part used in the classification process with Naïve Bayes.

The following are the stages of the process in predicting using the RapidMiner application: 1) Entering the Read Excel model into RapidMiner, 2) Connecting to the Split Data model in Read Excel to divide Testing and Terser data, 3) Adding the Naïve Bayes model to carry out the classification process, and 4) Finally use the Apply Model to apply the Naïve Bayes classification model. To add informational results such as error testing by adding a Performance model.

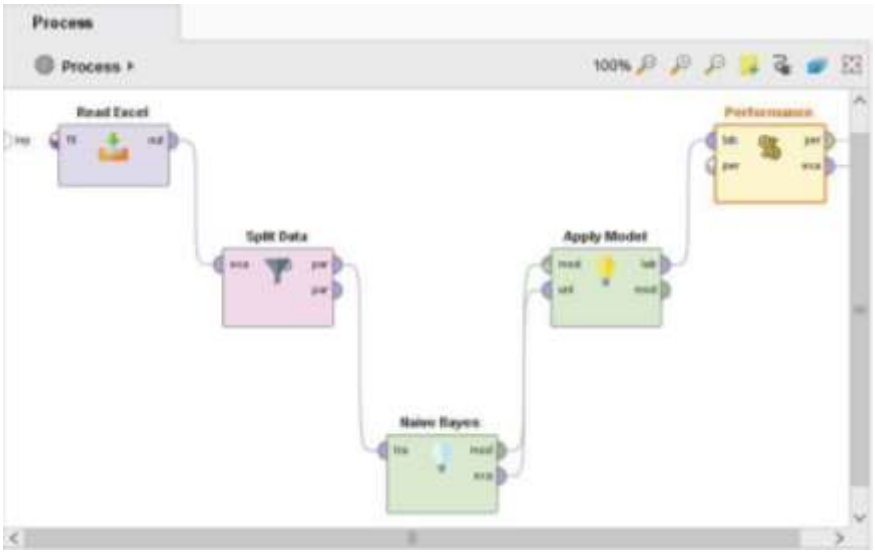


Figure 3. The Naïve Bayes Performance Model.

Discussions

The sentiment results obtained from the overall classification process in the form of a histogram can be seen in the graphic image. Based on the recapitulation of sentiment analysis regarding the development of the Omicron variant of the COVID-9 virus in society, a positive, negative and neutral sentiment group was obtained. The results obtained were that 84% of the people responded positively, 9% of the people responded neutrally and 7% of the people responded negatively.

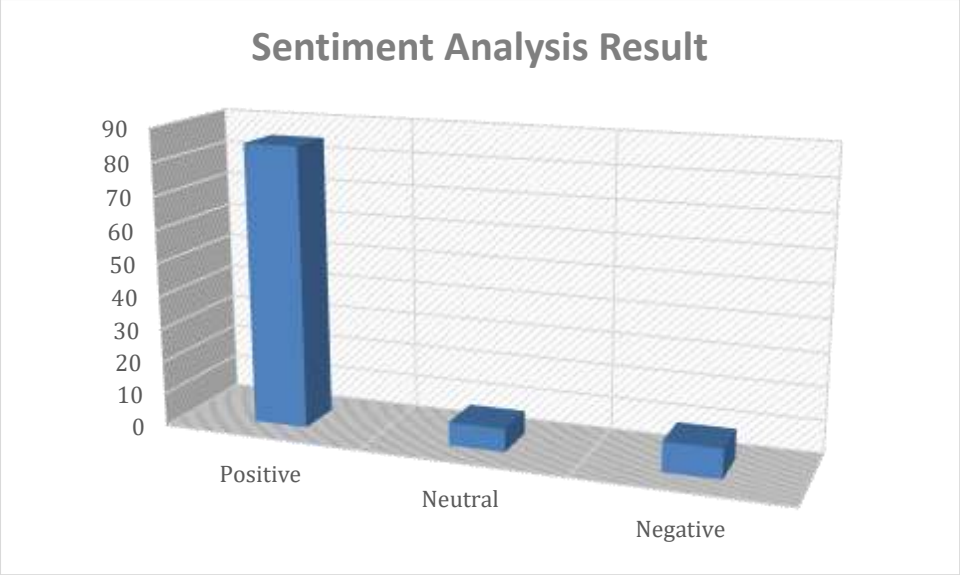


Figure 4. The Sentiment Analysis Results.

The Omricon variant of COVID-19 is indeed more contagious than other COVID-19 variants, but the severity tends to be less severe. This causes the public sentiment to be more positive in the class. In addition, as time goes by, people are getting more and more familiar with

COVID-19 so that it is no longer a scourge that is too scary. The positive class in sentiment analysis may reflect "positive sentiment," whereas the negative class may represent "negative sentiment." When applied to a binary classification issue, the Naive Bayes Classifier computes the conditional probability of the positive and negative classes given the input features. The projected class for the input data is then assigned to the class with the highest probability.

One of the core concepts in sentiment analysis is "positive classification." In this case, positive classification refers to finding text comment in Twitter that express a favorable opinion of the Omicron variation. Given the nature of the pandemic and the ever-changing public health context, identifying positive attitude might provide useful insights. Optimistic emotion could reflect an optimistic perspective, pleasant reactions to measures implemented, or even good news concerning vaccine efficacy. Finding favorable sentiment in the midst of a sea of data provides a more nuanced view of how this variant is seen.

One important part of our sentiment analysis is the concept of "negative classification." Negative classification in this case requires identifying and categorizing text data from social media posts that express a negative feeling toward the Omicron variety. This aspect of analysis provides a thorough view of the concerns, complaints, and anxieties that people have in response to the change.

Capturing the nuance of neutrality becomes an important part of sentiment analysis when people share their ideas, concerns, and facts. The process of identifying and categorizing written information, social media posts, and publications that reflect a neutral mood toward the Omicron variation is known as neutral classification. While positive and negative attitudes are frequently highlighted, knowing neutrality is also critical for a complete grasp of public image.

Conclusion and Suggestions

Based on the findings of sentiment analysis research on the COVID-19 variant of Omicron, it was found that 97% of the community responded positively, 0% of the community responded to Neutral and 13% of the community responded negatively. Also the issue of the latest Omicron COVID-19 virus is not dangerous from the beginning of the emergence of the COVID-19 virus in the world. For further research, other methods can be added to carry out sentiment analysis.

References

Abdillah, L. A. (2022). *Peranan Media Sosial Modern*. Bening Media Publishing.

Abdillah, L. A., HS, S., Muniarty, P., Nanda, I., Retnandari, S. D., Wulandari, Prasetyo, A. H., Sinambela, S., Mansur, Aulia, T. Z., Hamzah, A., Firmansyah, H., Andari, S., Rismadi, B., Purba, S., Gazi & Sina, I. (2021). *Metodologi Penelitian & Analisis Data Comprehensive*. Insania.

Abdillah, L. A. & Kurniasti, A. (2022). Mobile-based COVID-19 Vaccination Registration Application Prototype. *SinkrOn : Jurnal Dan Penelitian Teknik Informatika*, *7*(3), 2152–2159. https://doi.org/10.33395/sinkron.v7i3.11779

Abdillah, L. A., Mukti, Y. I., Puspita, D. & Suhartini. (2021). Indonesian Sharia Fintech Services and Social Media Usage. *Bulletin of Social Informatics Theory and Application*, *5*(2), 97–106. https://doi.org/10.31763/businta.v5i2.456

Abdillah, L. A., Putri, D. K. & Artavi, A. (2023). Enterprise Digital Payment Trends Survey Post COVID-19 Situation. *Scientific Journal of Informatics (SJI)*, *10*(2), 105–112. https://doi.org/10.15294/sji.v10i2.43068

Aditya, B. R. (2015). Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining. *JURNAL INFOTEL - Informatika Telekomunikasi Elektronika*, *7*(2), 93. https://doi.org/10.20895/infotel.v7i2.35

Baharun, N., Razi, N. F. M., Masrom, S., Yusri, N. A. M. & Rahman, A. S. A. (2022). Auto Modellingfor Machine Learning: A Comparison Implementation between Rapid Miner and Python. *International Journal of Emerging Technology and Advanced Engineering*, *12*(5), 15–27. https://doi.org/10.46338/ijetae0522_03

Bokaee Nezhad, Z. & Deihimi, M. A. (2022). Twitter sentiment analysis from Iran about COVID 19 vaccine. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, *16*(1), 1–7. https://doi.org/10.1016/j.dsx.2021.102367

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, *39*(11), 27–34. https://doi.org/10.1145/240455.240464

Guspita, Y. & Abdillah, L. A. (2021). Implementasi Naïve Bayes untuk Analisis Sentimen Publik terhadap Bentukan Kabinet Baru Indonesia 2019 pada Media Online. *The Third Bina Darma Conference on Computer Science (BDCCS2021)*, *3*(5), 981–986. https://conference.binadarma.ac.id/index.php/BDCCS/article/view/2878/1131

Hasibuan, F. F., Dar, M. H. & Yanris, G. J. (2023). Implementation of the Naïve Bayes Method to determine the Level of Consumer Satisfaction. *SinkrOn : Jurnal Dan Penelitian Teknik Informatika*, *8*(3), 1709–1723. https://doi.org/10.33395/sinkron.v8i2.12349

Hikmawan, S., Pardamean, A. & Khasanah, S. N. (2020). Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning. *Jurnal Kajian Ilmiah*, *20*(2), 167–176. https://doi.org/10.31599/jki.v20i2.117

J. Arunadevi, S. Ramya & Raja, M. R. (2018). A study of classification algorithms using Rapidminer. *International Journal of Pure and Applied Mathematics*, *119*(12), 15977–15988.

Kori, A. (2017). Comparative Study of Data Classifiers Using Rapidminer. *International Journal of Engineering Development and Research (IJEDR)*, *5*(2), 1041–1043. https://www.ijedr.org/download.php?file=IJEDR1702177.pdf

Kwak, J. A. & Cho, S. K. (2018). Analyzing public opinion with social media data during election periods: A selective literature review. *Asian Journal for Public Opinion Research*, *5*(4), 285–301. https://doi.org/10.15206/ajpor.2018.5.4.285

Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J. & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, *55*(3), 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924

Ling, J., Kencana, I. P. E. N. & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, *3*(3), 92. https://doi.org/10.24843/mtk.2014.v03.i03.p070

Ning, B., Junwei, W. & Feng, H. (2019). Spam message classification based on the naïve Bayes classification algorithm. *IAENG International Journal of Computer Science*, *46*(1), 46–53. http://www.iaeng.org/IJCS/issues_v46/issue_1/IJCS_46_1_05.pdf

Putra, M. W. A., Susanti, Erlin & Herwin. (2020). Analisis Sentimen Dompet Elektronik Pada

Twitter Menggunakan Metode Naïve Bayes Classifier. *IT Journal Research and Development*, *5*(1), 72–86. https://doi.org/10.25299/itjrd.2020.vol5(1).5159

Qaiser, S. & Ali, R. (2018). Text Mining : Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Application (IJCA)*, *181*(1), 25–29. https://www.ijcaonline.org/archives/volume181/number1/qaiser-2018-ijca-917395.pdf

Rachman, F. F. & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Health Information Management Journal*, *8*(2), 100–109. https://inohim.esaunggul.ac.id/index.php/INO/article/view/223/175

Routray, P., Swain, C. K. & Mishra, S. P. (2013). Sentiment Analysis in Social Networks. In F. A. Pozzi, E. Fersini, E. Messina & B. Liu (Eds.), *International Journal of Computer Application*. Morgan Kaufmann (Elsevier Inc.).

Sabarmathi, G. & Chinnaiyan, R. (2017). Reliable data mining tasks and techniques for industrial applications. *... Journal for Advanced Research in …*, *4*(7), 138–142. http://www.iaetsdjaras.org/gallery/20-jaras-339-december.pdf

Saraswati, E., Umaidah, Y. & Voutama, A. (2021). Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap COVID-19. *Generation Journal*, *5*(2), 109–118. https://doi.org/10.29407/gj.v5i2.16125

Syarifuddinn, M. (2020). Analisis Sentimen Opini Publik Mengenai COVID-19 Pada Twitter Menggunakan Metode Naïve Bayes dan KNNa. *INTI Nusa Mandiri*, *15*(1), 23–28. https://doi.org/10.33480/inti.v15i1.1347

Uska, M., Wirasasmita, R., Usuluddin, U. & Arianti, B. (2020). Evaluation of Rapidminer-Aplication in Data Mining Learning using PeRSIVA Model. *Edumatic: Jurnal Pendidikan Informatika*, *4*(2), 164–171. https://doi.org/10.29408/edumatic.v4i2.2688

Wahyu Handani, S., Intan Surya Saputra, D., Hasirun, Mega Arino, R. & Fiza Asyrofi Ramadhan, G. (2019). Sentiment analysis for go-jek on google play store. *Journal of Physics: Conference Series*, *1196*(1). https://doi.org/10.1088/1742-6596/1196/1/012032

WHO. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

Widiastuti, N. I., Rainarli, E. & Dewi, K. E. (2017). Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen. *Jurnal INFOTEL (Informatika, Telekomunikasi, Elektronika)*, *9*(4), 416–421. https://doi.org/10.20895/infotel.v9i4.312

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 (JDMSI)*, *2*(2), 1–9. https://jim.teknokrat.ac.id/index.php/JDMSI/article/download/1344/672