Sentiment Analysis on the Application of Tabungan Perumahan Rakyat (TAPERA) Policy in Indonesia

Mochamad Naufal Dzulfikar

Abstract— **Tapera is a program provided by the Indonesian government to provide poor people with decent housing, by charging all working classes with salaries above the Regional Minimum Wage of 3% for Tapera contributions. Even though it seems that this program aims to help poor people, this has given rise to a new polemic, which can be seen from some people who don't like this, as can be seen from several negative comments on social media, one of which is X or Twitter. This paper aims to determine whether public sentiment toward the TAPERA policy is positive or negative among Indonesians affected by it, regardless of whether the sentiment analysis results are predominantly positive or negative. This research is conducted by collecting data from Twitter and then processing it in such a way that, later will be checked for sentiment analysis with the Naive Bayes method. Out of 742 analyzed tweets (filtered from an initial 1,680), 651 expressed negative sentiment, while 91 were positive, indicating significant public dissatisfaction with the TAPERA policy.**

Keywords— **TAPERA, sentiment, data, Twitter, text mining**.

I. INTRODUCTION

Tapera is a program provided by the Indonesian government to provide poor people with decent housing, by charging all working classes with salaries above the Regional Minimum Wage of 3% for Tapera contributions [21]. Recently, news about Tapera has been widely discussed by the public, as seen on Google Trends, where on May 28, 2024, many people were looking for news related to Tapera [27].

Even though it seems that this program aims to help poor people, this has given rise to a new polemic, which can be seen from some people who don't like this, as can be seen from several negative comments on social media, one of which is X or Twitter. This created a polemic for them because this policy was burdensome for workers by cutting their salaries by 3% outside of income tax [21].

The writing of this report received high urgency due to conditions where the change of president and vice

president is only a matter of months away, making the new president think about whether this rule should be implemented or not. Apart from that, it is hoped that the use of social media analytics in this research can open the government's eyes to that in the government sector, the use of social media analytics is very necessary to find out the public's response and public opinion regarding the latest regulations issued by the government. Even though the Minister of Finance Sri Mulyani said that the TAPERA regulations were postponed, however, the policy will be implemented in 2027 [28], therefore, there is still urgency regarding writing this paper, because there is a possibility that this TAPERA regulation will cause problems for the community in the future.

The objective of this paper is to find out whether the sentiment of the TAPERA policy is positive or negative for Indonesian citizens affected by the policy, with the aim that the author can provide recommendations for the government also encourage the usage of Social Media Analytics on government sector in Indonesia, even the result of the sentiment analysis is majority positive or majority negative.

The systematic writing of this report will be divided into a literature review that discusses the basics that are the benchmark for this report, a research methodology that contains the research methods that the author will carry out, results, and a discussion that focuses on the results obtained and the reasons why the results can be achieved, and will end with a conclusion that focuses on the results of the research in a brief, concise, and clear manner, as well as the recommendations given.

II. LITERATURE REVIEW

A. Text Mining

Text Mining can be defined as mining data in the form of text where data sources are usually obtained from documents, and the goal is to find words that can represent the contents of the document so that it can analyze the relationship between documents (Harlian, 2006).

According to Harlian, there are 5 stages in text mining, namely Tokenizing, Filtering, Stemming, Tagging, and Analyzing.

Tokenizing is the stage of cutting the input string based on each word that composes it, where the data that was originally in the form of sentences, is converted into data per word. Tokenizing is the first step for many processing languages, for example, part of speech tagging, machine translation, spell checking, sentence boundary detection, information retrieval, and information extraction [11].

Filtering is the stage where the word data that has been obtained from the Tokenizing method will be sorted out so that the data needed can be taken perfectly. This method can be used by taking less important words or by keeping important words.

Stemming is the stage of taking root words from words that have been filtered before. Usually, the words that this method requires are words that have affixes. This method aims to prevent wasted time in analyzing sentiment with a large corpus because with stemming, the analysis system will work less, saving time for analysis [10].

Tagging is a stage where words that have been adjusted by the Stemming method will be searched for basic words. Usually, the words that are changed at this stage are words that come from English and use tenses.

The last method is Analyzing, which will determine the relationship between words between existing documents based on the results needed.


TEXT MINING INVOLVES A SERIES OF ACTIVITIES TO BE PERFORMED IN ORDER TO EFFICIENTLY MINE THE INFORMATION. THESE ACTIVITIES ARE:

GATHER — DATA ASSEMBLE FROM DIFFERENT RESOURCES
PREPROCESS — DATA PREPARATION AND TRANSFORMATION
INDEX — QUICK ACCESS AND SEARCH STORED DATA
MINING — ALGORITHM, INFERENCE AND INFORMATION EXTRACTION
ANALYSIS — USER ANALYSIS, NAVIGATION

Figure 1 Example of a Text Mining process diagram

B. Data Mining

Data Mining is another subdomain of Artificial Intelligence and can be defined as a process that aims to generate knowledge from data and fully present the findings to users. Generating knowledge in the context of Data Mining can be translated into the discovery of new and non-trivial patterns, relationships, and trends in data that are useful to users [8]. The process of Data Mining can be differentiated into descriptive, where knowledge is represented in the form of models that describe patterns and relationships in the data, and predictive, where knowledge is represented in the prediction of future conditions, trends, and relationships [4][6].

Most Data Mining applications are mostly related to quality management. There are very few Data Mining applications directly related to production complexity. However, other Data Mining applications in other areas of production management serve the purpose of managing production complexity very well [4].

C. Tapera (Tabungan Perumahan Rakyat)

Tapera is an abbreviation for *Tabungan Perumahan Rakyat* or Public Housing Savings in English. This policy is not new, because it has been established since 2016. The first legal basis for Tapera is Law No. 24 of 2016 concerning Public Housing Savings. Furthermore, Government Regulation (PP) No. 25 of 2020 was issued as a complement to the law, which was later replaced by PP No. 21 of 2024. This latest PP is what has become controversial in the current implementation of Tapera. The Tapera deduction is 3% of the worker's salary, with details of 2.5% borne by the worker and 0.5% borne by the employer. Meanwhile, for independent workers or freelancers, the 3% amount is fully borne by the worker [17].

Based on PP No. 21 of 2024 articles 5 and 7, Tapera participants are workers who are at least 20 years old or married. Other conditions including the minimum salary or income of the customers equal to UMR (Upah Minimum Regional, or Regional Minimum Wage in English) or more, have worked for a minimum of 12 months, and the customers don't have his/her own house yet.

D. Comparison with the Previous Research (Analysis of Government Policy Sentiment Regarding Vacation during the COVID-19 Pandemic Using the Bidirectional Encoder Representation from Transformers (BERT) by Intan Nurma Yulita et al [1])

This research focuses on analyzing public sentiment regarding regulations regarding holidays during the COVID-19 pandemic using Bidirectional Encoder Representation from Transformers (BERT), where the system checks contextual relationships between words in the text [2]. The results of this research provide enlightenment regarding the use of sentiment analysis, where the explanation of the research carried out is complete, and takes everything into account, especially the comparison of the BERT system with other sentiment analysis method systems, such as Naïve Bayes [1].

In this research, what differentiates the research conducted by the author from this research is the data collection method which will be based on one social media platform, namely X or Twitter, and the data collection method that will be used to collect the data.

E. Comparison with the Previous Research (Prediction of The Level of Public Trust in Government Policies in the 1st Quarter of The Covid-19 Pandemic using Sentiment Analysis by Zahra Nur Aziza et al [3])

Just like research by Nurma Yulita which focused on regulations during the COVID-19 pandemic, this research focuses on analyzing sentiment related to regulations issued by the government during the first quarter of the COVID-19 pandemic as well as measuring electability and the level of public trust in the government. Data collection was taken from Twitter and data calculations were carried out using the Naïve Bayes method [3].

In this research, what will differentiate this research from the previous research is the use of more data (whereas this previous report used only 100 sample

data) and the use of more diverse sentiment calculation methods.

III. RESEARCH METHODS

For the research method in this paper, the author will collect text data from Twitter social media with tweets where there are words "tapera" because these two words are very relevant to the topic described in the title of this report, where TAPERA policy is mentioned. For the period of tweet writing, it will be limited from the posting date of May 20, 2024, till the present day, to make sure that the author can get the maximum number of searches. The reason the author uses Twitter as a place for data collection is due to several factors, such as where Twitter focuses heavily on simple text (only 280 characters per tweet) compared to other social media such as Facebook and Instagram which causes Twitter data to be more meaningful for voicing opinions than data with lots of text or images [15], and because of the ease of accessing tweets with the help of OAuth (Open Authorization) which has been issued by Twitter [15][16].

For data retrieval, because Twitter began to change back features for data retrieval, such as a limited view rate limit and users who have not logged in cannot open Twitter posts, some applications that use the Twitter API are not running perfectly. Therefore, the author will use code from Helmi Satria, who happened to find a way to do data mining for academics using Twitter Harvest [20].

For data preprocessing, the author will use NLTK (Natural Language Toolkit) from Sastrawi [23]. Sastrawi Library is a stemmer library that is used to overcome the problem of changing words from words to basic words [22]. Sastrawi stemmer then applied an algorithm based on Nazief and Adriani, refined with the CS (Confix Stripping) algorithm, the ECS (Enhanced Confix Stripping) algorithm, further improved with the ECS Modification. It is hoped that by applying Literary Literature in the document preprocessing process, especially the stemming and stop word removal stages, it can make important words unimportant to increase the accuracy of sentiment analysis [25].

For sentiment analysis modeling, the Naïve Bayes model, along with SVM and Decision Tree Modifier will be used. Naïve Bayes is not a single algorithm but a collection of several algorithms, all of which have the same principal function, namely that each pair of classified features is independent of each other [24] . Meanwhile, the decision tree classifier (DTC) is a nonparametric supervised method used for classification. The objective of these methods is to create a model that estimates TARGET (dependent variable) values by learning simple decision rules from FEATURES (independent variables) [7]. And support vector machines are a learning system that uses a hypothesis space of linear functions in a high-dimensional feature space, trained with a learning algorithm from the optimization theory. SVM also attempts to minimize the upper bound on the generalization error based on the principle of structural risk minimization (SRM) [8]. With an average of 85 percent for the SVM method [25] and 78 percent for

Naive Bayes [24], it is hoped that the results of the sentiment analysis calculations this time will be accurate analysis calculations.
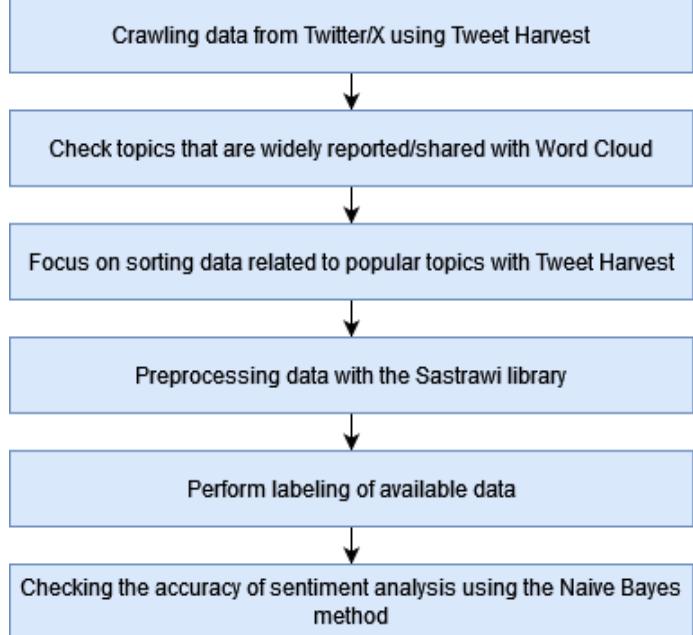
Figure 2 Research Methods to be Conducted

Figure 3 Code snippet used to retrieve Twitter data by Helmi Satria

Figure 4 Code Snippet used for preprocessing using Sastrawi on Python

IV. RESEARCH RESULT AND DISCUSSION

For the presentation of data from the research results, the author will present the results in two forms, namely bar charts from sentiment analysis results, word clouds and topic modeling from sentiment analysis tables. For this paper, the author will focus on sentiment analysis, not topic modeling. This research data is not divided into certain categories so that it only looks at the

sentiment and issue results for each sentiment, so the statistical tests do not seem necessary.

A. Research Results with Bag of Words

From 742 datasets that got from Twitter (after getting sortied, previously, 1680 data were obtained by this method, mostly because some of the tweets are not related to TAPERA policy at all, and talk about other discussions, also the author manually sorted through the data to identify and eliminate political biases and duplicates, ensuring a more accurate analysis of the information, by checking the datasets manually and deleted the data that is counted as duplicate or have a political bias, but still can't detect which writing is the bot or not, which is one of the weaknesses of this research), the author checks the words that often appear related to this issue through checking the bag of words. A bag of words is a text-based visual representation that displays the significance of words in terms of popularity and importance using different font sizes and colors [13]. Figure 5 shows the results of the bag of words for this study.
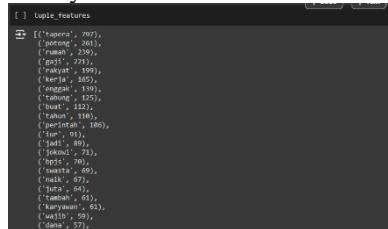
Figure 5 Bag of Words Code Snippets and Words with the Most Frequency

From the results of Figure 5 and Figure 6, it can be seen that the words 'tapera', 'potong', and 'rumah' are very dominant words from the collection of tweets that appear, according to the topic of discussion in this research. Some words show negative elements in the word cloud, such as the words 'tolak', and 'derita', which indicates that there is some negative sentiment towards this, although it is not known how many tweets show this.

In addition to using word clouds, the authors also checked Topic Modelling intending to know what topics are often discussed related to Scrum and Agile methods on Twitter social media, using BERTopic, which is a Topic Modelling method that is used to allow for easy interpretable topics whilst keeping important words in the topic descriptions [20]. Figures 7 and 8 show two topics discussed on data taken from Twitter Harvest, but one is omitted because the topic isn't relevant to the research (mostly Topic 1 talks about Gaza, which somehow has the same time frame as when TAPERA controversy happened).
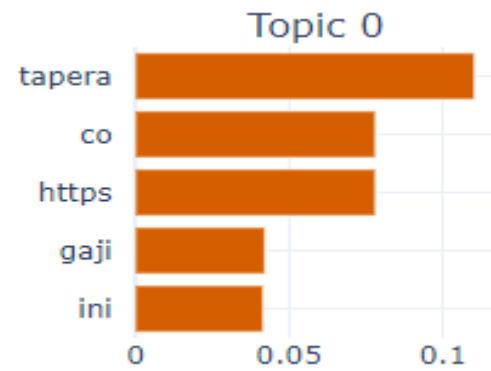

Figure 6 Results from Topic Modelling analysis from first Twitter data using BERTopic

Based on the results of Topic Modelling in Figure 6, the topics that are often discussed are related to 'tapera', and 'gaji', showing the main problem of TAPERA, which is the coercion of taking worker's salary for TAPERA dues. With the results of Topic Modelling, it can be concluded that the topics discussed when discussing TAPERA on Twitter are usually complaints about how much people need to spend their money to pay TAPERA dues.

B. Sentiment Analysis Results

For the sentiment analysis results, the author will show the results in a bar chart, that the more the bar is going up, the bigger the sentiment it is. Also, this research gave the results of the recall and precision of the two models used in this research, which is the Naïve Bayes Model and the SVM Model.
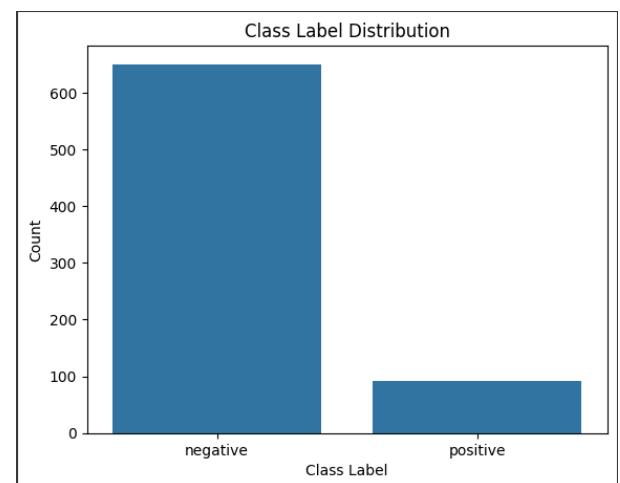

Figure 7 Comparison Between Positive Sentiment Count and Negative Sentiment Count based on Data Labelling

Table 1 Naive Bayes Model Recall and Precision During the Prediction Test

Target Class	Positive	Negative
Recall	0.25	0.93
Precision	0.36	0.88

Table 2 SVM Model Recall and Precision During the Prediction Test

Target Class	Positive	Negative

Recall	0.31	0.91
Precision	0.38	0.88

Table 3 Positive Sentiment Tweet Examples

No	Tweet Examples
1	@cimotyy Aku sangat mendukung bnget tabungan rumah tapera dari btn karena dengan itu semua orang yg ga memiliki gaji tetap bisa mempunyai kesempatan untuk memiliki rumah sendiri dari tabungan tapera ini #BTNSyariahUntukSemua
2	@twingkerbells bagus banget ada fasilitas Tabungan Rumah Tapera begini jadi semua bisa mudah ambil kredit rumah walau tanpa fix income #BTNSyariahUntukSemua
3	Bahagiakan keluarga dengan memiliki #RumahPertama menggunakan fasilitas pembiayaan #RumahTAPERA dari BP Tapera https://t.co/ihhVsZvf0b

Table 4 Negative Sentiment Tweet Examples

No	Tweet Examples
1	@hrdbacot Bgst baru aja PPh pake skema TER jadi makin gede kenanya sekarang ada lagi Tapera. Ujung2 nya middle class lagi yang kena upper class kapan dikenainnya cokk
2	Coba sekali-kali Pemerintah yg serius kakau mau bantuin rakyat jangan malah ngeprank terus. Dibalik aja.. Pemerintah wajib ngasih rumah ke semua rakyatnya yang belum punya rumah. Baru selanjutnya ada potongan Tapera 3% per bulan. @KasbiIndo https://t.co/Geyed5cTbz
3	Dipaksa ikut BPJS pelayanannya kek taik. Skrg dipaksa tapera bentuknya kek kandang babi kah? Rezim taik babi
4	Modus nya selalu membebani rakyat! Tapera jaman orba hanya PNS aja rezim ini nyasar karyawan swasta!

The results of Figure 7 show that 651 tweets have negative sentiments, and 91 tweets have positive sentiments, which shows how people are disappointed with the TAPERA policy, one of the reasons is because of their salaries, which most of the people are already low on it but need to be spent for TAPERA dues.

From the results shown in Table 1 and Table 2, there are significant results when looking at the comparison between precision and recall for positive sentiment and negative sentiment. With the accuracy between the SVM model and the Naive Bayes model being close together, the SVM accuracy condition is 81.5 percent, and the Naive Bayes accuracy is 82.8 percent, but the recall and precision for the two models between positive and negative sentiment are very different. This could be caused by too much negative data being taken, or you could say, too many negative responses from the public on Twitter regarding this problem.

Figure 8 Results from Word Cloud that Picking Up from the Negative Sentiment on TAPERA Policy

Figure 9 Results from Word Cloud that Picking Up from the Positive Sentiment on TAPERA Policy

Apart from trying to find out the sentiment analysis of the community regarding TAPERA, the author tries to find out what the problems are with TAPERA by checking the Word Cloud for existing negative sentiments and making estimates based on the results of the Word Cloud. The results from Figure 8 shows that there are words that often appear, namely 'tapera', 'people', 'salary', and 'cut'. The results of this Word Cloud show that the reason why TAPERA is rejected in the community is because they do not agree to be willing to have their salaries cut to pay for TAPERA, which of course will make it difficult for those with low salaries.

Based on the results from the word cloud, it can also be seen which words often appear from the positive sentiment that emerged from the results of Twitter data collection. The word cloud shows words such as 'tapera', 'savings', and 'people' which shows that there are some people who are still optimistic that the salary that will be deducted for Tapera costs will later be used as savings for housing, which can be enjoyed in the future coming.

The study reveals widespread public disapproval of TAPERA, not due to a lack of awareness but because

people believe it disproportionately burdens low-income workers, especially when combined with existing tax deductions. Not to mention the history of rampant corruption cases in Indonesia, where according to Transparency International, Indonesia is one of the most corrupt countries with an index score of 34 out of 100, where the lower the score, the more corrupt the country is [14], so it is not surprising that the public's trust in the government low, one example is when many people do not trust the government when it comes to handling COVID-19 in Indonesia [18].

Regarding the results obtained, several reasons are considered reasonable regarding the existence of positive sentiment from this case, one of which is the existence of political buzzers who support this case (in this case, fully supporting the TAPERA regulations). Buzzer is a term for people who are paid to become influencers for certain parties, be they religious organizations, or politicians and political parties. These buzzers usually form accounts in large numbers to get support for the party in question, such as tweeting about the good things the party in question is doing or defending themselves desperately if the party in question gets into trouble [12]. Apart from the buzzer, some support TAPERA, because of their high hopes for getting a liveable house.

Sentiment analysis plays a critical role in shaping future government policy regarding TAPERA by providing insights into public opinion and attitudes toward the program. Here's how it can impact policy decisions:

1. Understanding Public Perception: By analyzing sentiments expressed on social media platforms like Twitter, the government can gauge overall public approval or discontent with the TAPERA policy. A predominantly negative sentiment, as indicated by studies, suggests significant dissatisfaction, which could prompt officials to reconsider or revise the program.

2. Informed Decision-Making: If public sentiment is largely negative, the government may opt to delay or modify the policy to address concerns raised by citizens. Understanding specific aspects that are criticized can lead to targeted adjustments that better serve the population's needs, ensuring more effective support for low-income households.

3. Engagement and Communication: Sentiment analysis can guide the government in its communication strategies. If negative sentiments arise due to a lack of understanding or misinformation, officials can enhance their outreach efforts to educate the public about the program's benefits, thus reducing resistance.

4. Policy Feedback Loop: Ongoing sentiment analysis can create a feedback loop where the government not only implements policies but also continuously monitors public sentiment regarding those policies. This iterative process allows for real-time adjustments based on societal responses, ensuring that TAPERA remains relevant and beneficial.

5. Prioritizing Social Welfare: Strong negative sentiments could highlight the need to prioritize social welfare initiatives over burdensome policies. If the government recognizes that the benefits do not outweigh the costs for citizens, it may pivot to alternative solutions that still meet the goal of providing decent housing without placing an undue financial burden on workers.

6. Legitimacy and Public Trust: By actively considering public sentiment and responding appropriately to feedback, the government can enhance its legitimacy and build trust with the electorate. This proactive approach could lead to greater acceptance of future reforms and policies.

In summary, sentiment analysis serves as a valuable tool for the government to understand the needs and opinions of its citizens concerning the TAPERA policy. By leveraging this information, policymakers can implement more effective strategies that genuinely address the challenges faced by low-income groups while ensuring sustainable development in housing provision.

V. Conclusion

This study concludes that Indonesian public sentiment toward TAPERA is predominantly negative, though some positive views exist on social media, particularly Twitter. Because of this, the author recommends that the TAPERA policy needs to be re-evaluated to make sure that the policy will be better for both sides, which is the government side and the people's side. Also, the author reminds the government that the usage of social media analytics is a good way to maintain the opinion of the people heard by the government.

Although this research has been successfully carried out, there are still some shortcomings in this research, one of which is that the amount of data taken is quite limited (due to the changing rules of Twitter regarding data collection related to data mining, also with how low people in Twitter discussed this topic) which causes there are still things that have not been revealed in this research.

References

I. N. Yulita *et al*, "Analysis of Government Policy Sentiment Regarding Vacation during the COVID-19 Pandemic Using the Bidirectional Encoder Representation from Transformers (BERT)," *Data*, vol. 8, *(3),* pp. 46, 2023. Available: https://www.proquest.com/scholarly-journals/analysis-government-policy-sentiment-regarding/docview/2791602503/se-2. DOI: https://doi.org/10.3390/data8030046.

Yi Tay, Mostafa Dehghani, Dara Bahri, and Donald Metzler. 2022. Efficient Transformers: A Survey. ACM Comput. Surv. 55, 6, Article 109 (June 2023), 28 pages. https://doi.org/10.1145/3530811

Z. N. Aziza and D. Y. Kristiyanto, "Prediction of The Level of Public Trust in Government Policies in the 1st Quarter of The Covid-19 Pandemic using Sentiment Analysis," *E3S Web of*

Conferences, vol. 317, 2021. Available: https://www.proquest.com/conference-papers-proceedings/prediction-level-public-trust-government-policies/docview/2819262849/se-2. DOI: https://doi.org/10.1051/e3sconf/202131705013.

Witten, IH; Pal, CJ; Frank, E; Hall, MA. Data mining. Practical machine learning tools and techniques. Cambridge: Morgan Kaufmann; 2017.

Fayyad, U; Piatetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery in databases. AI magazine 1996;17:37-54.

G. Schuh et al, "Data Mining Definitions and Applications for the Management of Production Complexity," Procedia CIRP, vol. 81, pp. 874-879, 2019.

K. A. Silva, G. de Souza Rolim and de Oliveira Aparecido, Lucas Eduardo, "Forecasting El Niño and La Niña events using decision tree classifier," *Theoretical and Applied Climatology,* vol. 148, *(3-4),* pp. 1279-1288, 2022.

R. Ramesh *et al*, "Thermal error measurement and modeling in machine tools. Part II. Hybrid Bayesian Network—support vector machine model," *International Journal of Machine Tools & Manufacture,* vol. 43, *(4),* pp. 405-419, 2003.

Nurul Fitri Ramadhani, "Cek Fakta: benarkah Tapera jadi kontroversi hanya karena kurangnya sosialisasi?" *The Conversation*, Jun. 25, 2024. https://theconversation.com/cek-fakta-benarkah-tapera-jadi-kontroversi-hanya-karena-kurangnya-sosialisasi-233227 (accessed Aug. 22, 2024).

M. Haroon, "Comparative Analysis of Stemming Algorithms for Web Text Mining," International Journal of Modern Education and Computer Science, vol. 11, (9), pp. 20, 2018. Available: https://www.proquest.com/scholarly-journals/comparative-analysis-stemming-algorithms-web-text/docview/2114235714/se-2. DOI: https://doi.org/10.5815/ijmecs.2018.09.03.

Z. Rehman et al, "Morpheme Matching Based Text Tokenization for a Scarce Resourced Language," PLoS One, vol. 8, (8), 2013. Available: https://www.proquest.com/scholarly-journals/morpheme-matching-based-text-tokenization-scarce/docview/1431447106/se-2. DOI: https://doi.org/10.1371/journal.pone.0068178.

A. Seto, "Islamist Buzzers: Message Flooding, Offline Outreach, and Astroturfing," Austrian Journal of South-East Asian Studies, vol. 12, (2), pp. 187-208, 2019. Available: https://www.proquest.com/scholarly-journals/islamist-buzzers-message-flooding-offline/docview/2335163852/se-2. DOI: https://doi.org/10.14764/10.ASEAS-0021.

M. Chi et al, "Morphable Word Clouds for Time-Varying Text Data Visualization," IEEE Transactions on Visualization and Computer Graphics, vol. 21, (12), pp. 1415-1426, 2015.

Transparency International, "2023 Corruption Perceptions Index - Explore Indonesia's results," *Transparency.org*, Jan. 30, 2024. https://www.transparency.org/en/cpi/2023/index/idn

W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," Journal of Big Data, vol. 5, (1), pp. 1-10, 2018. Available: https://www.proquest.com/scholarly-journals/prediction-analysis-indonesia-presidential/docview/2158433012/se-2. DOI: https://doi.org/10.1186/s40537-018-0164-1.

Ravindran SK, Garg V. Mastering Social media mining with R: extract valuable data from your social media sites and make better business decisions using R. Birmingham: Packt Publisher; 2015

Maksum Rangkuti, "Apa Yang Dimaksud Dengan Tapera? Begini Penjelasannya," *Blog Info dan Berita*, Jun. 06, 2024. https://fahum.umsu.ac.id/blog/apa-yang-dimaksud-dengan-tapera-begini-penjelasannya/ (accessed Aug. 18, 2024).

Riefky *et al*, "Growing public distrust towards the Indonesian Government for lack of response to COVID-19 outbreak," *IOP Conference Series. Earth and Environmental Science,* vol. 716, *(1),* pp. 12072, 2021.

"Home - BERTopic," *maartengr.github.io*. https://maartengr.github.io/BERTopic/index.html

H. Satria, "Crawl data twitter menggunakan tweet Harvest - Juli 2023," Helmi Satria, https://helmisatria.com/blog/crawl-data-twitter-menggunakan-tweet-harvest (accessed Dec. 9, 2023).

"Tapera: Mengapa Kebijakan tabungan Perumahan Rakyat disebut 'Tidak Masuk Akal' Menyediakan Hunian rakyat Yang Terjangkau?", BBC News Indonesia, https://www.bbc.com/indonesia/articles/cyxxjdwk5z8o (accessed Aug. 16, 2024).

K. K. Purnamasari and I. S. Suwardi, "Rule-based Part of Speech Tagger for Indonesian Language," *IOP Conf Ser Mater Sci Eng*, vol. 407, no. 1, p. 12151, 2018, doi: 10.1088/1757-899X/407/1/012151.R

M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi," *IOP Conf Ser Mater Sci Eng*, vol. 874, no. 1, p. 12017, 2020, doi: 10.1088/1757-899X/874/1/012017.

Y. Ying, T. N. Mursitama, Shidarta, and Lohansen, "Effectiveness of the News Text Classification Test Using the Naïve Bayes' Classification Text Mining Method," *J Phys Conf Ser*, vol. 1764, no. 1, p. 12105, 2021, doi: 10.1088/1742-6596/1764/1/012105.

N. I. Nordin *et al*, "Enhancing COVID-19 Classification Accuracy with a Hybrid SVM-LR Model," *Bioengineering,* vol. 10, *(11),* pp. 1318, 2023. Available: https://www.proquest.com/scholarly-journals/enhancing-covid-19-classification-accuracy-with/docview/2892968688/se-2. DOI: https://doi.org/10.3390/bioengineering10111318.

Google Trends. "Tapera" Accessed August 16, 2024. https://trends.google.co.id/trends/explore?date=today%203-m&geo=ID&q=tapera&hl=id

V. S. Kumar and K. Vidhya, "Heart Plaque Detection with Improved Accuracy using Naive Bayes and comparing with Least Squares Support Vector Machine," *Cardiometry, (25),* pp. 1595-1599, 2022. Available: https://www.proquest.com/scholarly-journals/heart-plaque-detection-with-improved-accuracy/docview/2777085851/se-2. DOI: https://doi.org/10.18137/cardiometry.2022.25.15951599.Anggita Raissa, "Tapera Ditunda, Bagaimana Kelanjutannya?" *Deduktif.id*, 2024. https://deduktif.id/tapera-ditunda-bagaimana-kelanjutannya (accessed Sep. 13, 2024).