Sentiment and Emotional Analysis of The Public Housing Savings Program (TAPERA) using Orange Data Mining

Hawangga Dhiyaul Fadly [1]*, Handaru Jati [2]**
* Electronics and Informatics Engineering Education, Yogyakarta State University
** Electronics and Informatics Engineering Education, Yogyakarta State University
hawanggadhiyaul.2023@student.uny.ac.id [1], handaru@uny.ac.id [2]

Article Info

Article history:

Received 2025-03-17
Revised 2025-04-11
Accepted 2025-04-14

Keyword:

Sentiment,
Emotion,
VADER,
POMS,
TAPERA.

ABSTRACT

This study employs a text analysis methodology to assess public perception of the People's Housing Savings Program (TAPERA), by examining 3.078 tweets containing the keyword "tapera" using the Orange Data Mining application with two analytical approaches: the Valence Aware Dictionary and Sentiment Reasoner (VADER) for sentiment analysis and the Profile of Mood States (POMS) for emotional analysis. The sentiment analysis results indicate 1.481 tweets (48,2%) expressed negative sentiment, 830 tweets (27%) were neutral, and 767 tweets (24,8%) conveyed positive sentiment. These findings suggest that although there is a portion of positive responses toward the TAPERA policy, most of the public tends to express dissatisfaction or scepticism about the program. Furthermore, the emotional analysis identified depression as the most dominant emotion expressed by the public, appearing in 2.019 tweets (65,6%), followed by confusion (14,7%) and anger (9,6%). Positive emotions such as vigour and tension were recorded in significantly lower proportions, at 2,9% and 1,8%, respectively. These results illustrate that the public feels frustrated, confused, and anxious regarding the TAPERA policy, with minimal expressions of optimism or enthusiasm. This analysis highlights the need for a more transparent, educational, and data-driven communication approach to enhance public understanding, trust, and participation in the TAPERA policy. Therefore, the government must design more effective outreach strategies to address public concerns and ensure the successful implementation of this program.

I. INTRODUCTION

The Government of Indonesia established the Public Housing Savings Program (TAPERA) through Government Regulation No. 25 of 2020, which was later revised under Government Regulation No. 21 of 2024 [1]. This program was designed to provide access to homeownership for the public, particularly for low-income groups, through a savings-based mechanism. By utilizing this savings mechanism, TAPERA offers a pathway for the economically disadvantaged to acquire housing. The primary objective of this program is to address Indonesia's severe housing shortage. It is envisioned as a long-term solution to ensure that every citizen has access to adequate housing in the context of continuous population growth and the escalating demand for housing [2].

This regulation mandates formal and informal workers to allocate a portion of their income to a housing savings scheme under TAPERA. The funds collected are managed by the government and allocated to provide low-interest housing loans for eligible participants [3]. Participants can withdraw their savings and accrued returns at their employment or retirement. However, academics and practitioners continue to debate the implementation and effectiveness of the TAPERA policy. Concerns have been raised, particularly regarding its execution mechanisms, legal framework, and challenges in the dissemination and implementation processes at the grassroots level. Public understanding of the TAPERA concept and its benefits remains limited, which hampers participation in the program. This underscores the need for more comprehensive studies to evaluate the policy's reception and implementation at the community level [4]. For instance, while the public housing credit program meets affordability criteria in Aceh Province, it lacks sustainability aspects, such as access to public transportation and waste management [5]. This further highlights the need for in-depth studies to assess the public's local acceptance and implementation of TAPERA.

Analyzing TAPERA-related policy data and information is crucial in understanding the policy's impacts and barriers to implementation. One practical approach in this domain is text analysis [6], which enables researchers to systematically identify themes, patterns, and structures in policy documents. This analytical model has gained interest in the field of natural language processing, particularly for analyzing individual language and sentiment [7][8][9]. These methods are specifically developed to detect explicit and implicit sentiments embedded in individual expressions [10][11]. Text analysis has also proven effective in policy evaluation, such as analyzing public reactions to government policies to identify significant challenges and criticisms that policymakers can address [12].

This study employs text analysis to assess public opinion on the TAPERA policy, which is crucial in improving and developing future policies. This analysis utilizes text-based clustering and classification methods based on content similarity, allowing researchers to detect hidden patterns and trends in large-scale datasets [13][14]. This approach is particularly valuable in public policy analysis as it reveals prevailing attitudes and public impressions that conventional analytical tools may not capture. Clustering algorithms can be used to group data based on content similarity, thus facilitating the mapping of interrelated policy aspects [15][13].

Several other related studies have explored government policy analysis through public opinion text analysis. For instance, [16] analyzed public reactions to the cooking oil shortage in Indonesia. Their sentiment and polarity classification showed that public responses were reasonably balanced between positive and negative, with 28% being neutral. However, 69% of the public expressed depressive emotions, mainly due to discussions around cartels, high prices, scarcity, and the resulting impact on citizens. Studies on the relocation of the national capital and its associated risks were conducted by [17][18] revealing a wide range of public responses involving positive and negative emotions. These studies identified *surprise* as a dominant emotional response to the realization of the capital relocation. They also reported *joy* as an emotion arising from the public's hope for positive change despite ongoing concerns about infrastructure development's social and environmental impacts. Both studies emphasize the emergence of *fear* associated with potential adverse environmental effects, particularly in Kalimantan. A study on public emotional responses to the Omnibus Law was conducted by [19] which classified public emotions toward the policy. The most dominant emotion was *surprise*, indicating public shock over the law's enactment. The main topics discussed included "workers advancing with the omnibus law", "recovering the national economy together", "Job Creation Law", "millennials' time to do business", and "prosperous omnibus", signalling a need for serious attention from the government to ensure that the Omnibus Law is effective and aligns with the expectations of the people.

This study used the Valence Aware Dictionary and Sentiment Reasoner (VADER) algorithm to classify sentiment. This algorithm was employed to examine sentiments expressed in public tweets about TAPERA. VADER is particularly well-suited for analyzing social media data, as it can process informal language, emoticons, and slang commonly found in tweets [20]. By analyzing the emotional tone of tweets, VADER provides a numerical representation of expressed sentiments and classifies them as positive, negative, or neutral. This facilitates the identification of overall public sentiment and detecting trends and patterns in reactions to TAPERA.

Additionally, emotional analysis was conducted using the Profile of Mood States (POMS) method, which measures six primary emotional dimensions: tension, depression, anger, vigour, fatigue, and confusion [21]. The application of this model enables the assessment of emotional responses associated with the TAPERA policy discourse. The sentiment and polarity classification outcomes offer valuable insights into public attitudes toward TAPERA, highlighting areas of concern and potential improvements in policy communication and implementation. The combination of VADER sentiment analysis and POMS emotional analysis contributes to a comprehensive understanding of public opinion, supporting ongoing efforts to refine and enhance the TAPERA program.

II. METHODOLOGY

This study was conducted in several distinct phases, each with specific objectives to produce the necessary results. These phases include: (i) data collection, (ii) tweet content analysis and classification, (iii) data preprocessing, (iv) emotional analysis, and (v) polarity classification. The research methodology is illustrated in the Figure 1:

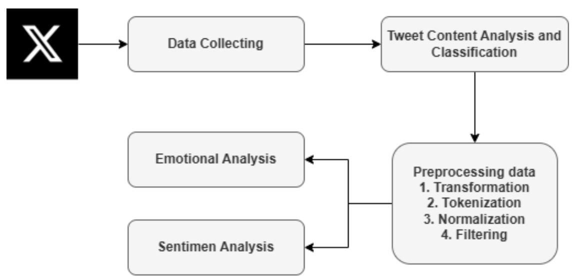

Figure 1. Research Methodology

A. Data Collection

Sentiment analysis involves collecting tweet data through the X platform's API. Several studies have demonstrated X's potential for disseminating information, gauging public opinion, and conveying emotions, making it a valuable tool for researchers and social media analysts [22][23]. Tweets were obtained via the X API using a crawling technique, with the keyword “TAPERA” employed to filter relevant content, ensuring that the data collected remained focused on the research topic.

B. Tweet Content Analysis and Classification

The main objective of this phase is to categorize the collected tweets based on the type of content (text, image, or video) and inherent characteristics. A naturalistic approach was applied in the content analysis, involving a random selection of tweets for detailed examination [24]. Prior to analysis, raw data underwent preprocessing to eliminate irrelevant elements.

C. Data Preprocessing

Data preprocessing is a crucial step before content analysis. At this stage, raw text data was cleaned by parsing hypertext markup language (HTML) content, removing uniform resource locators (URLs), converting uppercase letters to lowercase, and eliminating punctuation, mentions (@), hashtags (#), and other irrelevant symbols. The preprocessing process included several stages: transformation (converting text to lowercase), tokenization (splitting text into smaller units), filtering (removing irrelevant elements), and normalization (reducing words to their base form).

D. Sentiment Analysis

Sentiment analysis categorizes text based on its positive, negative, or neutral sentiment orientation. This is a crucial aspect as it helps identify the overall direction of sentiment conveyed in a statement [25]. The VADER algorithm was employed to convert words into numerical values based on their emotional intensity and assign sentiment scores to each text [20]. These scores reflect the overall tone of the text by calculating a compound score based on the weighted sum of each word. Positive sentiment is indicated by scores from 0.001 to 1, neutral sentiment ranges between -0.001 and 0.001, and negative sentiment falls between -0.001 and -1 [26]. This study used the Orange Data Mining application, which integrates VADER to compute compound scores and analyze sentiment, ensuring a robust and accurate assessment of the emotional tone across the tweets.

E. Emotional Analysis

Emotional analysis in this study employed the POMS algorithm, which calculates emotions based on an individual's mood over a given period. This method assesses six primary emotional dimensions: Tension, Depression, Anger, Vigor, Fatigue, and Confusion [21]. The study utilized the Orange Data Mining application, which includes a POMS algorithm feature to analyze emotional content within text data. This feature was used to measure the emotional tone of the tweets collected for this research, allowing the identification of the dominant emotional states expressed by users regarding TAPERA.

III. RESULT AND DISCUSSION

This study employed the Orange Data Mining application to perform sentiment and emotional analysis on text data. The research workflow is illustrated in Figure 2.

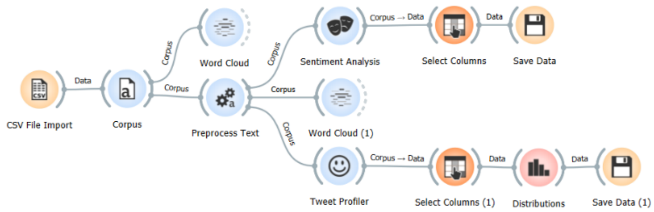

Figure 2. Research Workflow

A. Data Collection

This research collected 3,078 tweets containing the hashtag #tapera, which were used for further sentiment analysis. The collected tweets underwent various preprocessing steps before being analyzed, which were essential to extracting insights regarding public opinion and sentiment related to the #tapera topic.

B. Tweet Content Analysis and Classification

This study adopted a naturalistic approach to analyze the sentiments expressed in the tweets and to identify the formats (text, image, or video) used to convey those sentiments [24]. The initial phase involved randomly selecting a sample of 100 tweets to examine their sentiment characteristics. Each tweet was thoroughly analyzed to uncover subtle nuances and embedded emotions. For instance, a tweet like *“They already know the global market is in recession, instead of providing subsidies to local citizens to support domestic purchasing power, they introduced TAPERA”* was categorized as "emotion-evoking," as it conveyed feelings of frustration or concern. Another tweet, such as *“Welcome to a country where people work hard, yet their earnings are deducted for TAPERA while gamblers/families receive social assistance,”* was classified as "criticism," highlighting dissatisfaction with the system.

The categorization process identified seven distinct sentiment groups. This approach enabled the identification of core themes and topics in the tweet sample. Emotions and themes were organized into relevant categories using an inductive reasoning method. Consistent with previous studies [27][28], new sentiment categories were introduced to accommodate tweets that did not fit existing classifications. Manual analysis was expanded to the entire dataset of 3,078 tweets, enabling a more comprehensive understanding of user opinions, as outlined in Table 1. Additionally, the tweet data were translated into English to ensure greater sentiment and emotional analysis accuracy, as the algorithms used perform more effectively with English vocabulary models.

TABLE I
CLASSIFICATION OF TWEET SENTIMENT GROUPS

Category	Description
Emotion-evoking	Posts that evoke emotions such as happiness, sadness, positivity, or negativity regarding TAPERA
Asking Questions	Posts raising inquiries related to TAPERA
Criticism of Government	Posts criticizing TAPERA policy
Support for Government	Posts supporting government action regarding TAPERA
Disagreement	Posts expressing opposition or disagreement with TAPERA policy
Data and News	Posts containing statistical data and news related to TAPERA
Irrelevant	Posts not relevant or unrelated to the TAPERA topic

C. Data Preprocessing

The data preprocessing stage was crucial in preparing raw tweet data for analysis by converting it into a suitable format. This stage consisted of the following essential steps:

1) *Transformation:* All uppercase letters were converted to lowercase in this step to ensure uniformity. This process helped avoid treating words like "Tapera" and "tapera" as separate entities. Additionally, diacritical marks were removed from characters (e.g., converting "é" to "e") to simplify and standardize the text. HTML tags, often found in tweets when users share links or media, were parsed and removed to eliminate irrelevant or disruptive elements. URLs were also deleted to remove references to external websites that do not contribute to the textual content under analysis.

2) *Tokenization:* Tokenization involves breaking text into smaller units, typically words or phrases, for further analysis. At this stage, punctuation marks (commas, periods, and question marks) were removed, as they do not carry significant meaning for sentiment analysis. This step enabled a clearer understanding of the individual words without interference from unnecessary characters.

3) *Normalization:* Normalization aimed to reduce inflected words (with prefixes or suffixes) to their base or root forms. For example, words like *"running," "runner,"* or *"walked"* were reduced to the base form *"run."* This process helped determine each word's meaning, enhancing sentiment classification accuracy.

4) *Filtering:* This step involved removing elements such as hashtags (#), mentions (@), and non-essential symbols that may not directly contribute to the tweet's sentiment. While hashtags and mentions can be useful for identifying specific topics or individuals, they can interfere with analytical accuracy. Eliminating these elements ensured the analysis focused solely on the content and emotion expressed within the tweet itself. Table 2 presents samples of preprocessed data related to TAPERA, from the original text to preprocessed text.

TABLE II
TEXT PREPROCESSING RESULT

Original Text	Preprocessed Text
Is it right to use the word savings in the acronym tapera, aka public housing savings? https://t.co/yq144ut9nt #Opinion #Kompas59 #MultimediaMencerahkan	right use the word savings acronym tapera, aka public housing savings
@zy_zy_lestary @Yattie2023 @xquitavee @dika_maxelyou @adri_7i @HaramAsoy @arie_nurma01_24 Until you want to borrow money from the people under the pretext of tapera savings. Just go bankrupt and go bankrupt. Talking about queuing up investors. Eh.. The reality is, no one has come in yet.. Wild boar cheaters	want borrow money people pretext tapera savings bankrupt bankrupt talking queuing investor reality no one come yet wild boar cheater
Lazy Gamblers who lose at gambling will be given banson. Workers seeking halal sustenance will have their salaries cut in the name of tapera. Where is the logic?	lazy gambler lose gambling bans on worker seeking halal sustenance salary cut name tapera logic
@zy_zy_lestary I can't help but grin....I'm at a loss, most likely IKN will collapse due to lack of funds, so the UKT TAPERA will come out with something strange....Indonesia is not doing well right now.....	grin loss likely in collapse lack fund ukt tapera come strange indonesia not doing well right now
zy_zy_lestary yattie2023 xquitavee dika_maxelyou adri_7i haramasoy arie_nurma01_24 want borrow money people pretext tapera savings bankrupt bankrupt talking queuing investor reality no one come yet wild boar create	want borrow money people pretext tapera savings bankrupt bankrupt talking queuing investor reality no one come yet wild boar cheater

The visualization of this preprocessing stage utilized the word cloud widget, which generated a graphical representation of the most frequently occurring words in the dataset, highlighting key terms and their importance. The processed data was linked to the word cloud widget, allowing for an effective and efficient flow of information within the

Orange Data Mining software. This facilitated a more straightforward interpretation and analysis of word frequency and relevance within the sentiment analysis context. The workflow for preprocessing text and generating a word cloud in Orange Data Mining is shown in Figure 3 below.

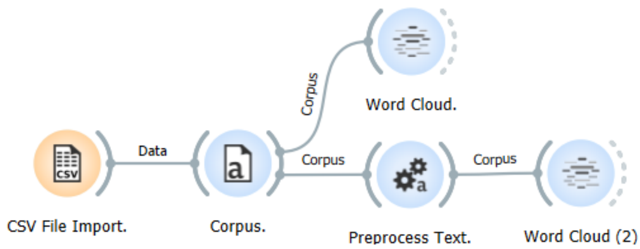

Figure 3. Preprocessing Text and Word Cloud Workflow

This study used the word cloud to represent tweets containing the hashtag #tapera before and after the preprocessing stage. The comparison of word cloud representations at these two stages highlights the impact of preprocessing on the textual data. Before preprocessing, the word cloud includes irrelevant elements such as hashtags, mentions, HTML tags, and URLs, which can obscure meaningful terms. However, these elements are removed after preprocessing, and the focus shifts solely to the keywords in the tweets. This comparison is illustrated in Figures 4 and 5, which visually depict the changes in the composition of the word clouds, offering a clearer understanding of the most discussed terms related to TAPERA.

Figure 4. Word Cloud before Preprocessing

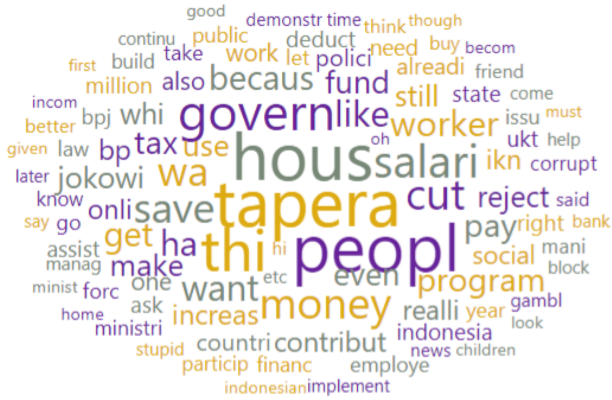

Figure 5. Word Cloud after Preprocessing

D. Sentiment Analysis

The sentiment analysis conducted in this study employed the VADER algorithm. Figure 6 illustrates the workflow of the polarity classification process, implemented using the Orange Data Mining application. It outlines the steps involved in processing and analyzing tweet data.

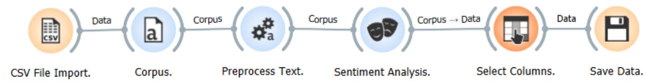

Figure 6. Sentiment Analysis Workflow

The sentiment analysis results on tweets related to the TAPERA program reveal that 1.481 tweets expressed negative sentiment, 830 were classified as neutral, and 767 conveyed positive sentiment. VADER was deemed particularly suitable for this analysis due to its optimization for social media text, as it effectively handles informal language, slang, and emoticons commonly found on platforms such as Twitter [29]. As such, VADER enabled a relevant interpretation of public sentiment regarding TAPERA, considering the characteristically concise and informal language typical of social media discourse.

These findings offer valuable insights into public perceptions of TAPERA by categorizing sentiments into three main types positive, neutral, and negative thereby reflecting public approval, indifference, or dissatisfaction with the program. The result shown in Figure 7 visually illustrates the sentiment distribution among tweets containing the hashtag #tapera, offering a clearer understanding of how public opinion has evolved and disseminated across the digital sphere. These results illustrate sentiment trends toward TAPERA and demonstrate VADER's effectiveness in analyzing sentiment within emotionally expressive and informal social media text.

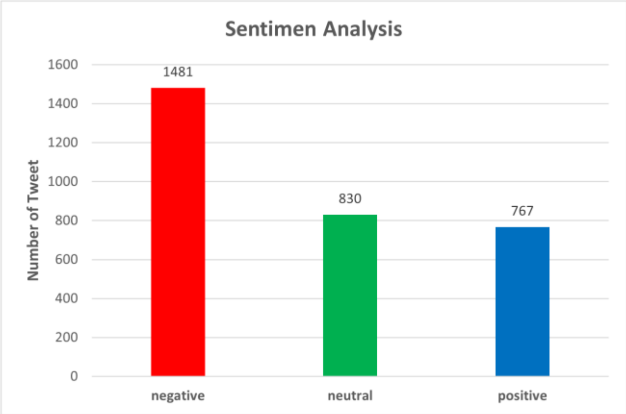

Figure 7. Sentiment Analysis Result

E. Emotional Analysis

The emotional analysis in this study employed the POMS method. The emotional analysis process involved connecting the Preprocess Text widget, which prepares the data for analysis, to the Tweet Profiler widget. This connection enabled the identification of emotional states present in the tweets based on six primary emotional dimensions. After processing the text data, the Select Column widget was used to filter relevant emotional dimensions, which were then visualized using a bar chart via the Distribution widget. The workflow outlining the steps in the emotional analysis process using Orange Data Mining software is shown in Figure 8.

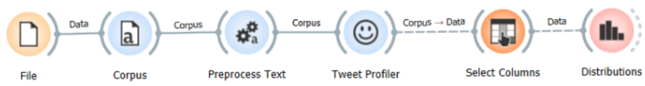

Figure 8. Emotional Analysis Workflow

The results of the emotional analysis of tweets related to TAPERA indicate that the most dominant emotion was depression, recorded in 2.019 tweets, followed by confusion in 452 tweets, anger in 297 tweets, fatigue in 167 tweets, vigour in 88 tweets, and least expressed emotion was tension, with only 55 tweets. These findings provide significant insights into the emotional states expressed by individuals on social media. The emotional analysis model used in this study aimed to offer a deeper understanding of how individuals convey emotions through their thoughts [10][30]. The results of this emotional analysis regarding the TAPERA policy reflect a range of emotional responses expressed by the public via tweets, offering a broader picture of societal perceptions toward the policy. Figure 9 presents a clear visual representation through a bar chart, depicting the emotional distribution within the #tapera tweet dataset, facilitating interpretation and comprehension of emerging emotional patterns, and providing more comprehensive information about the emotional impact of the policy on society.

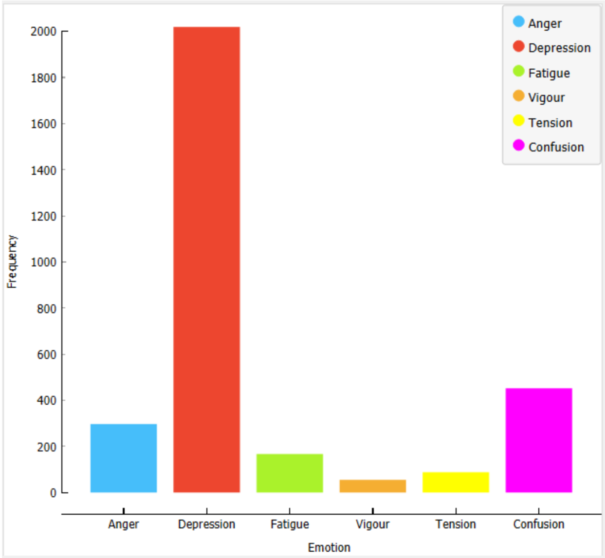

Figure 9. Emotional Analysis Result

The findings indicate a predominantly negative public sentiment and emotional response toward the TAPERA policy. This aligns with reports from credible news sources that suggest public dissatisfaction. For instance, a national media, *Kompas*.com reported that the Indonesian Consumer Protection Foundation (YLKI) criticized the comparison between TAPERA and BPJS Kesehatan, asserting that TAPERA contributions do not guarantee home ownership and raising concerns that public funds could be diverted toward the development of Indonesia's new capital [31]. A June 2024 poll conducted by *Litbang Kompas* revealed that 66% of respondents lacked confidence in TAPERA's potential to resolve housing issues, and 74,7% expressed scepticism regarding the government's ability to manage public investment funds effectively further amplifying the negative sentiment toward the program [32].

A separate study utilizing the IndoBERT Lite model on YouTube comments related to TAPERA reported that 65,6% of comments were negative, highlighting concerns about transparency, implementation, and communication. Positive sentiment accounted for only 17,9%, with neutral comments comprising 11,8% [33]. Additionally, media monitoring of platform X (formerly Twitter) revealed a spike in negative sentiment, peaking at 213 tweets between June 7–9, 2024. Mentions of TAPERA rose markedly, from 209 in May 26–28 to 359 in early June, signifying heightened public attention and concern [34]. The primary sources of criticism included the mandatory nature of program participation, perceived burdensome salary deductions (3%), lack of transparency in fund management, unclear long-term benefits, and insufficient public education or outreach. These concerns were especially pronounced among lower-income groups for whom the 3% deduction is particularly impactful [35]. The results of this study are consistent with previously documented public sentiments, indicating that most of the population holds a negative view of TAPERA, especially regarding mandatory enrollment, salary deductions, financial transparency, and the lack of adequate program socialization. Consequently, the government must improve communication strategies, enhance transparency, and engage in more robust outreach efforts to foster public acceptance and participation in the TAPERA initiative.

The sentiment and emotional analysis methods VADER and POMS proved highly suitable and effective despite alternative approaches such as TextBlob, Support Vector Machines (SVM), and Naïve Bayes. In this context, VADER demonstrated superiority in detecting sentiments in text containing emoticons, abbreviations, and colloquial expressions typical of social media communication [29]. Meanwhile, using POMS allowed for a more nuanced measurement of emotional states. Although POMS is traditionally used in clinical psychology, its application to social media text analysis has proven effective in assessing collective mood and mapping emotional expressions in everyday language [36][37]. Both methods offered distinct advantages in this study due to their ease of implementation, efficiency in processing informal text, and capability to swiftly generate relevant analyses of public sentiment and emotional expression toward the TAPERA policy.

Nevertheless, several limitations should be acknowledged. POMS is less flexible in handling informal language, sarcasm, and complex emotional expressions. Similarly, while VADER is fast and effective for short texts, it struggles with deeper emotional nuances and sarcasm and is constrained by a limited lexicon. Furthermore, reliance on tweets as the sole data source introduces bias, as it represents only a subset of the population active on social media and does not fully capture the broader public view. Factors such as geography, age, and educational background of social media users may influence how they engage with such policies, which are not comprehensively reflected in the current analysis. Therefore, future studies should consider integrating additional data sources such as direct surveys or interviews to provide a more holistic and representative understanding of public opinion.

IV. Conclusion

The sentiment and emotional analyses conducted on 3.078 tweets related to the TAPERA policy provide significant insights into public opinion and psychological responses. Using the VADER sentiment analysis algorithm, the data reveal that 1.481 tweets expressed negative sentiment, 830 tweets were neutral, and 767 tweets conveyed positive sentiment. These results indicate that although a segment of the public expresses optimism or support for the policy, the majority demonstrate dissatisfaction, scepticism, or a lack of engagement. The dominance of negative sentiment reflects a critical or doubtful stance towards the policy's effectiveness and implementation, while the prevalence of neutral sentiment suggests a broader ambiguity or detachment. The emotional analysis based on POMS shows that *depression* was the most prevalent emotion, identified in 2.019 tweets, followed by *confusion* in 452 tweets and *anger* in 296 tweets.

In contrast, positive emotions such as vigour and tension appeared far less frequently, recorded in only 89 and 55 tweets, respectively. This emotional profile underscores the extent of psychological distress, uncertainty, and disapproval expressed by the public regarding the TAPERA program. The predominance of negative sentiment and emotion highlights the need for the government to implement more effective, transparent, and citizen-focused communication strategies. These insights provide a deeper understanding of the public's emotional climate, primarily characterized by stress and apprehension about TAPERA's execution. The government needs to develop more effective communication and outreach strategies that directly address public concerns and promote the program's broader participation and long-term success. Clear, informative, and needs-based communication will be critical in reshaping public perception and fostering trust in the TAPERA policy.

Acknowledgement

The author would like to thank the master's program in Electronics and Informatics Engineering Education at Yogyakarta State University for its support of this research.

References

[1] Goverment of Indonesia, "Government Regulation of the Republic of Indonesia Number 28 Year 2012," 2024.

[2] G. Wijburg, "Tax-incentivized housing production and the affordability crisis: International lessons from the low-income housing tax credit program in the United States," *Hous. Stud.*, vol. 39, no. 7, pp. 1632–1657, Aug. 2024, doi: 10.1080/02673037.2022.2135172.

[3] G. U. Fayomi, E. K. Onyari, and S. E. Mini, "Evaluating social housing potential for low-income urban dwellers in Johannesburg," *Case Stud. Chem. Environ. Eng.*, vol. 9, 2024, doi: 10.1016/j.cscee.2024.100737.

[4] J. Ferri and F. Herranz-Baez, "Government expenditure and the housing puzzle: Unpacking mechanisms," *Econ. Model.*, vol. 140, no. November 2023, p. 106844, 2024, doi: 10.1016/j.econmod.2024.106844.

[5] S. Rassanjani, I. Rahmi, and S. Ulya, "Affordable and Sustainable Government-Subsidized Housing Development in Aceh Besar Regency," *Int. J. Sustain. Dev. Plan.*, vol. 18, no. 6, pp. 1893–1899, 2023, doi: 10.18280/ijsdp.180625.

[6] B. Ferrario and S. Stantcheva, "Eliciting People's First-Order Concerns: Text Analysis of Open-Ended Survey Questions," *AEA Pap. Proc.*, vol. 112, pp. 163–169, 2022, doi: 10.1257/pandp.20221071.

[7] R. Sarwar, A. T. Rutherford, S.-U. Hassan, T. Rakthanmanon, and S. Nutanong, "Native Language Identification of Fluent and Advanced Non-Native Writers," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 19, no. 4, Apr. 2020, doi: 10.1145/3383202.

[8] Q. Qiu and J. Qiu, "Encoding Sememe and Context Information into Sentence Representation for Implicit Sentiment Analysis," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 83, 2021, pp. 423–433.

[9] H. Wang and M. Hou, "Quantum-like implicit sentiment analysis with sememes knowledge," *Expert Syst. Appl.*, vol. 232, 2023, doi: 10.1016/j.eswa.2023.120720.

[10] L. Rao, "Sentiment Analysis of English Text with Multilevel Features," *Sci. Program.*, vol. 2022, 2022, doi: 10.1155/2022/7605125.

[11] D. Li, X. Shi, and M. Dai, "A Text Sentiment Classification Method Enhanced by Bi-GRU and Attention Mechanism," in *Lecture Notes in Electrical Engineering*, 2024, vol. 1125, pp. 176–186, doi: 10.1007/978-981-99-9239-3_18.

[12] Luth, Kismartini, D. Lituhayu, R. Maswati, and T. Baharuddin, "Public Response on Twitter: The Urgency of Government Policy on Electric Vehicles," *E3S Web Conf.*, vol. 440, pp. 1–8, 2023, doi: 10.1051/e3sconf/202344003023.

[13] Y. Ortakci, "Revolutionary text clustering: Investigating transfer learning capacity of SBERT models through pooling techniques," *Eng. Sci. Technol. an Int. J.*, vol. 55, no. April, p. 101730, 2024, doi: 10.1016/j.jestch.2024.101730.

[14] T. Zang and H. Gu, "State-Space Modeling of Housing Sentiment for Regressing Changes of Real Estate Prices Following Short-Term Control Policy in China," *Sustain.*, vol. 15, no. 16, 2023, doi: 10.3390/su151612660.

[15] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Naacl-Hlt 2019*, no. Mlm, pp. 4171–4186, 2018, [Online]. Available: https://aclanthology.org/N19-1423.pdf.

[16] Ismail, H. N. Rahmah, and R. Sulistiyowati, "Penggunaan Software Orange Data Mining Pada Implementasi Text Mining Dalam Analisis Sentimen Netizen Di Twitter Terhadap Kelangkaan Minyak Goreng," *Sigma-Mu*, vol. 14, no. 2, pp. 1–11, 2022, doi: 10.35313/sigmamu.v14i2.4667.

[17] H. Dhery, A. Assyam, and F. N. Hasan, "Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 341–349, 2023, doi: 10.30865/klik.v4i1.957.

[18] A. Munawaroh, R. Ridhoi, and R. Rudiman, "Sentiment Analysis Dengan Naïve Bayes Berbasis Orange Terhadap Resiko Pembangunan Ikn," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 587–592, 2024, doi: 10.36040/jati.v8i1.8454.

[19] R. A. raffaidy Wiguna and A. I. Rifai, "Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining," *J. Inf. Syst. Informatics*, vol. 3, no. 1, pp. 1–12, 2021, doi: 10.33557/journalisi.v3i1.78.

[20] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," *Eighth Int. AAAI Conf. Weblogs Soc. Media*, pp. 216–225, 2014, [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

[21] D. M. McNair, L. F. Droppleman, and M. Lorr, *Edits manual for the profile of mood states: POMS*. Edits, 1992.

[22] A. Kada, A. Chouikh, S. Mellouli, A. J. Prashad, S. E. Straus, and C. Fahim, "An exploration of Canadian government officials' COVID-19 messages and the public's reaction using social media data," *PLoS One*, vol. 17, no. 9 September, 2022, doi: 10.1371/journal.pone.0273153.

[23] A. R. Lubis and M. K. M. Nasution, "Twitter Data Analysis and Text Normalization in Collecting Standard Word," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 2, pp. 855–863, 2023, doi: 10.37385/jaets.v4i2.1991.

[24] S. Bashir *et al.*, "Twitter chirps for Syrian people: Sentiment analysis of tweets related to Syria Chemical Attack," *Int. J. Disaster Risk Reduct.*, vol. 62, no. May, p. 102397, 2021, doi: 10.1016/j.ijdrr.2021.102397.

[25] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets," *IEEE Access*, vol. 8, pp. 181074–181090, 2020, doi: 10.1109/ACCESS.2020.3027350.

[26] S. Elbagir and J. Yang, "Sentiment Analysis on Twitter with Python's Natural Language Toolkit and VADER Sentiment Analyzer," vol. 0958, pp. 63–80, 2020, doi: 10.1142/9789811215094_0005.

[27] S. Gul and S. Shueb, "Twitter as a Mourning Dove for the Blooming Buds: An Analysis of Twitter's Reaction to Peshawar School Attack," in *MILID Yearbook*, P. K. Jagtar Singh and E. Hamburger, Eds. Paris, France: United Nations Educational, Scientific and Cultural Organization, 2016, pp. 237–247.

[28] S. Gul *et al.*, "Twitter sentiments related to natural calamities," *Electron. Libr.*, vol. 36, no. 1, pp. 38–54, Jan. 2018, doi: 10.1108/EL-12-2015-0244.

[29] A. Tarango-García, S. O. Lugo-Reyes, and A. Alvarez-Cardona, "Sentiment analysis on allergen immunotherapy on Twitter using a natural language processing model," *Rev. Alerg. Mex.*, vol. 71, no. 1, pp. 8–11, 2024, doi: 10.29262/ram.v71i1.1263.

[30] A. Chug, S. Sudeep, P. A. Kumar, A. Sharma, D. Maganti, and E. Uchoi, "Machine learning-based emotional identification of social media text," in *Computational Methods in Science and Technology - Proceedings of the 4th International Conference on Computational Methods in Science and Technology, ICCMST 2024*, 2025, vol. 2, pp. 295–301, doi: 10.1201/9781003561651-42.

[31] KOMPAS.COM, "Tapera Tak Jamin Beri Rumah, Tak Bisa Disamakan dengan BPJS Kesehatan," 2024. https://nasional.kompas.com/read/2024/06/12/09001811/tapera-tak-jamin-beri-rumah-tak-bisa-disamakan-dengan-bpjs-kesehatan (accessed Apr. 10, 2025).

[32] KOMPAS.COM, "Jajak Pendapat Litbang "Kompas": Resistensi Publik terhadap Tapera Terbilang Tinggi," 2024. https://www.kompas.id/baca/riset/2024/06/11/jajak-pendapat-litbang-kompas-resistensi-publik-terhadap-tapera-terbilang-tinggi (accessed Apr. 10, 2025).

[33] M. P. Firdaus and D. Trisnawarman, "Public Sentiment Analysis of the Public Housing Savings Program Using the IndoBERT Lite Model on YouTube Comments Analisis Sentimen Publik Terhadap Program Tabungan Perumahan Rakyat Menggunakan Model IndoBERT Lite pada Komentar YouTube," vol. 5, no. January, pp. 359–368, 2025.

[34] M. P. Prameswari, R. Abidin, P. Studi, A. Syariah, and W. Pekalongan, "ANALISIS MEDIA MONITORING TERHADAP ISU TAPERA MELALUI MEDIA," vol. 4, no. 1, pp. 7–12, 2024.

[35] A. R. Lubis, S. Nailah, and Y. P. Utami, "Analisis Isu Pada Media Sosial (X) Terkait Program TAPERA Berdasarkan Peraturan Pemerintah," vol. 4, no. 3, pp. 395–405, 2025.

[36] K. Tamura, Y. Sano, and J. Shiozaki, "Quantifying Collective Emotions: Japan's Societal Trends Through Enhanced Sentiment Index Using POMS2 and SNS," *Proc. - 2024 IEEE Int. Conf. Big Data, BigData 2024*, pp. 3082–3087, 2024, doi: 10.1109/BigData62323.2024.10826099.

[37] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011, doi: 10.1016/j.jocs.2010.12.007.