J@TI JURNAL TEKNIK INDUSTRI UNIVERSITAS DIPONEGORO

Vol. 20, No. 1, Januari 2025

TEXTBLOB-BASED SENTIMENT ANALYSIS OF TABUNGAN PERUMAHAN RAKYAT (TAPERA) POLICY: A PUBLIC PERCEPTION STUDY

Muhammad Iqbal Faturohman*[1], Miftahol Arifin[2]

[1]*Program Studi Teknik Industri, Universitas Telkom*
[2]*Program Studi Teknik Logistik, Universitas Telkom*
Jl. DI Panjaitan No.128, Karangreja, Purwokerto Kidul, Kec. Purwokerto Sel., Kabupaten Banyumas, Jawa Tengah, Indonesia 53147

Abstract

TAPERA (Tabungan Perumahan Rakyat) has generated considerable debate in Indonesia. Intended to assist low-income communities with affordable housing, the program has faced criticism due to policy inconsistencies and mandatory monthly contributions, which some argue impose significant financial burdens on participants. Concerns about the transparency and accountability of fund management further complicate its reception. This study utilizes sentiment analysis with the TextBlob library within Google Colab to evaluate public opinions on TAPERA. The research involved collecting data from YouTube comments via the API, followed by data preprocessing through six stages: Data Cleaning, Stopword Removal, Tokenization, Normalization, Stemming, and Translation. Sentiments were then analyzed using TextBlob, with the final dataset comprising 2,228 comments. The sentiment analysis revealed that 76% of the comments were negative, 23% positive, and 1% neutral. The predominance of negative sentiment indicates widespread dissatisfaction, suggesting issues with TAPERA's effectiveness, coverage, and implementation. Negative feedback reflects broader concerns about transparency and communication. These findings highlight the need for continuous evaluation and transparent communication to enhance TAPERA's effectiveness and address public concerns. Sentiment analysis proves to be a valuable tool for gauging public perception and guiding policy refinements, ensuring that the program aligns more closely with community needs and expectations.

Keywords: *Sentiment Analysis; Natural Language Processing; Text Mining; TextBlob; TAPERA*

1. Introduction

Currently, TAPERA (Tabungan Perumahan Rakyat) has sparked considerable debate and contention in Indonesia. Initially launched to aid low-income communities in accessing affordable housing, the program has encountered numerous controversies since its inception (Ganesha Putra Henriko , Fahmi Erwin, 2020). TAPERA was conceived as a solution to Indonesia's housing deficit, particularly targeting economically disadvantaged individuals struggling to secure adequate housing. This is demonstrated by the housing backlog in Indonesia, which amounted to 13.5 million units in 2021, accompanied by an annual demand for 1 million additional homes to address the growing needs (Baidarus et al., 2023). The program aimed to foster higher rates of home ownership among this demographic (Soeprapto, 2021). However, TAPERA's journey has been marked by various policy adjustments, affecting both participant contributions and management protocols. These changes have sometimes been inconsistent, leading to confusion among current and prospective participants (Ganesha & Fahmi, 2020).

A key point of contention revolves around the mandatory monthly contributions required of participants. Critics argue that these payments can impose undue financial burdens on low-income earners, sometimes surpassing their economic capacities (Jayachandran, 2023; Lippman et al., 2023). Furthermore, concerns about transparency and accountability in fund management have been raised. Questions persist about how collected funds are managed, with worries about potential misallocation or inefficiency. Public opinion reflects these controversies, highlighting debates over TAPERA's efficacy in delivering meaningful benefits to its intended communities versus possibly exacerbating existing socio-economic challenges.

Although TAPERA was conceived as a solution to address the housing deficit in Indonesia, the program

*Corresponding Author

E-mail: faturohmaniqbal@ittelkom-pwt.ac.id

has not yet proven fully effective in mitigating the prevailing social and economic inequalities. The housing crisis in Indonesia is not solely attributed to infrastructural challenges, but is also a reflection of broader systemic issues, including unequal wealth distribution, limited access to economic opportunities, and low levels of social mobility. Additionally, the mandatory contribution scheme imposed on participants has been perceived as a financial burden for low-income groups, while concerns surrounding the transparency and accountability in the management of the program's funds have introduced further uncertainty regarding the overall success of TAPERA.

Given the complexities surrounding TAPERA as outlined above, the controversies underscore the challenges of implementing social programs in Indonesia. It emphasizes the necessity for ongoing evaluation, enhancement, and transparent dialogue among the government, policy experts, and the public to achieve desired outcomes without significant adverse effects. Sentiment analysis serves as a valuable tool in gauging public acceptance (Gudankwar et al., 2024; ChandraKala & Sindhu, 2012) and support for TAPERA policies. This analytical approach enables the government to assess public understanding and endorsement of the program. Positive sentiment indicates a level of trust and confidence in TAPERA's benefits, while negative sentiment signals apprehensions or discontent with specific program aspects. Therefore, sentiment analysis plays a pivotal role in helping the government grasp social dynamics and gather community feedback (Fakhruzzaman et al., 2024; Udayana et al., 2023) on TAPERA policies. This not only enhances program effectiveness but also ensures that implemented policies maximize benefits for those in need.

Contemporary research on social housing programs like TAPERA emphasizes the integration of advanced data analytics and stakeholder engagement to enhance program effectiveness. Cutting-edge approaches include leveraging big data and machine learning algorithms to predict housing needs and optimize resource allocation. For instance, predictive analytics can forecast the housing requirements of various socio-economic groups, enabling more tailored and effective interventions (Elgendy & Elragal, 2014). Additionally, participatory approaches, such as incorporating feedback through digital platforms and community consultations, have been shown to improve program responsiveness and transparency (De Filippi & Cocina, 2022). Innovations in sentiment analysis further allow for real-time monitoring of public opinion, enabling swift adjustments to policies based on emerging concerns and perceptions. These state-of-the-art methodologies provide valuable insights into both the effectiveness of social housing programs and the satisfaction of beneficiaries, guiding policymakers in making data-driven decisions that enhance program outcomes and stakeholder engagement.

2. Method

In this study, the methodology focuses on Natural Language Processing (NLP) and sentiment analysis to extract meaningful insights from textual data. The methods employed include text mining and sentiment analysis, which together serve to uncover hidden patterns and sentiments within the data (Lindgren et al., 2023; Thakur et al., 2024). Python was used for implementing these techniques, with research conducted within the Google Colab environment to streamline data processing and model development. The text mining process involves extracting relevant textual data, which is then subjected to sentiment analysis to classify comments based on their polarity. This sentiment analysis is crucial for understanding the public's opinions and emotions regarding the subject matter. Following the information extraction phase, the findings are visually represented through Word Clouds, a method that facilitates the visualization of word frequencies within the text. Larger font sizes in the word clouds indicate words with higher occurrence rates, allowing for an intuitive understanding of key topics or sentiments within the data (Feng et al., 2022; Lohmann et al., 2015). These visualizations support the study’s goal of identifying dominant themes and sentiments within the corpus of text, linking each methodological step to the overarching aim of understanding public sentiment.

To accomplish the research objectives, sentiment analysis is utilized to gauge public reception and endorsement of the TAPERA policy. The methodology begins with data collection, utilizing the YouTube API to crawl comments, followed by data preprocessing, sentiment analysis using TextBlob, and culminating in the visualization and conclusion of the results. The selection of TextBlob for sentiment analysis in this study was driven by several factors that align with the research's objectives and scope. TextBlob is a well-established, lightweight natural language processing (NLP) tool that provides an intuitive interface for sentiment analysis, rendering it particularly effective for efficiently processing large volumes of textual data. In contrast to more complex models such as BERT, which necessitate substantial computational resources and the use of large-scale datasets for training, TextBlob employs a simpler, rule-based methodology that offers sufficiently accurate sentiment classification for the aims of this study. Furthermore, TextBlob's straightforward implementation and processing speed make it a practical choice for large-scale sentiment analysis, particularly in contexts where real-time results are crucial. While models like VADER and more advanced machine learning algorithms can yield more nuanced sentiment interpretations, TextBlob strikes a favorable balance between accuracy, simplicity, and computational efficiency, positioning it as a particularly suitable tool for this research, which seeks to assess the general public’s sentiment toward TAPERA policies. These procedural steps are elucidated in the flowchart depicted in **Figure 1**.

Data Collection

In this study, data is gathered utilizing web crawling methodologies. These techniques automate the retrieval of comments from web pages, facilitating

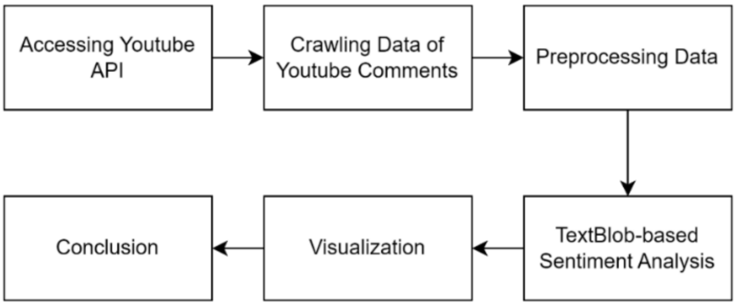

Figure 1. Research Procedural Steps

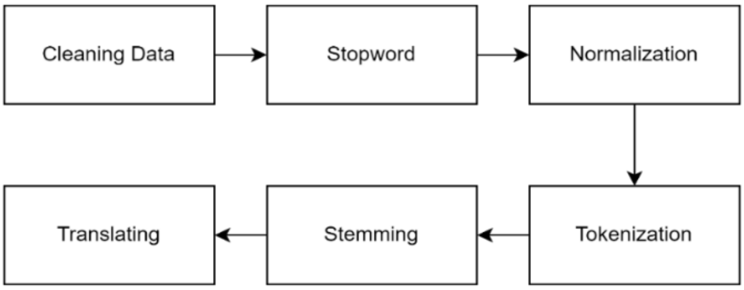

Figure 2. Research Flowchart Steps

subsequent analysis aimed at comprehending user sentiment regarding the product (Bharadwaj, 2023; Indrani P et al., 2024). The sampling process involves selecting YouTube videos with a high volume of comments that are relevant to the product or topic of interest. Criteria for inclusion include videos with significant engagement (e.g., a large number of views and comments) and those that attract diverse demographic groups in Indonesia, ensuring a broad representation of opinions. By focusing on YouTube comments, this study captures a wide range of sentiments from users of different ages, geographic locations, and socio-economic backgrounds, thus reflecting the diversity of the Indonesian population. Employing web crawling techniques enables the automated acquisition of data from these sources, eliminating the need for manual retrieval and ensuring the aggregation of a substantial and representative dataset, which can provide a comprehensive view of public sentiment.

Data pre-processing

Data preprocessing in this study is used to curate irrelevant textual elements for further processing. Preprocessing in this research consists of 6 stages: Data Cleaning, Stopword, Tokenization, Normalization, Stemming, and Translating. After data preprocessing, the amount of data that can be used for analysis is 2.228 data. The steps are as shown in the flowchart diagram in **Figure 2**.

Data Labeling

In this research, TextBlob is utilized for data labeling, offering an efficient and effective method for categorizing text data (Dallo, 2023). TextBlob is a comprehensive Python library designed for processing textual data, offering a range of tools and functionalities that facilitate the analysis of natural language (Choudhary & Kelley, 2023). It provides a straightforward API for common natural language processing (NLP) tasks, including part-of-speech tagging, noun phrase extraction, sentiment analysis, and translation (C. Wang et al., 2022). By leveraging underlying libraries such as NLTK (Natural Language Toolkit) and Pattern, TextBlob enables users to perform complex text processing with ease, making it a valuable tool for both beginners and advanced practitioners in the field of text analysis (N. T. Singh et al., 2023), (Goebel, 2019). Its user-friendly interface and support for text-based rules enhance its utility, making TextBlob a robust tool for managing and processing text data. This capability is crucial for conducting comprehensive analyses or developing advanced models in NLP applications.

Visualization

Word cloud visualization proves invaluable for illustrating sentiment analysis as it visually represents the most commonly appearing words in text, adjusting their sizes based on frequency (Lohmann et al., 2015). In sentiment analysis, word clouds are particularly useful for visually representing the most frequently occurring words associated with positive, negative, or neutral sentiments. TextBlob assigns a polarity score to each comment, with values ranging from -1 (most negative) to +1 (most positive). For this study, comments were classified based on their polarity scores: comments with scores above 0.1 were categorized as positive, those below -0.1 as negative, and those with scores between -0.1 and 0.1 were considered neutral. By applying these threshold values, we were able to identify the predominant sentiments in the data and generate word clouds that visually highlight the words most strongly associated with each sentiment category. This visualization simplifies comprehension of predominant sentiments within the

Table 1. The Result of Data Cleaning

No	Input	Output
1	Negara ini sangat kaya, tapi di korupsi terus. Tidak ada juga tindakan tegas untuk para koruptor,. Negara cuma menguras rakyat nya.	negara ini sangat kaya, tapi di korupsi terus. tidak ada juga tindakan tegas untuk para koruptor,. negara cuma menguras rakyat nya.
2	Gak ada sanksi tapi kok wajib, ini aja udah mencla mencle. Gimana nasib uang kita nih	gak ada sanksi tapi kok wajib, ini aja udah mencla mencle. gimana nasib uang kita nih

Table 2. The result of normalization

No	Input	Output
1	negara ini sangat kaya, tapi di korupsi terus. tidak ada juga tindakan tegas untuk para koruptor,. negara cuma menguras rakyat nya.	negara ini sangat kaya, tapi di korupsi terus. tidak ada juga tindakan tegas untuk para koruptor,. negara hanya menguras rakyat nya.
2	gak ada sanksi tapi kok wajib, ini aja udah mencla mencle. gimana nasib uang kita nih	tidak ada sanksi tapi wajib, ini saja sudah tidak dapat dipegang sama sekali. bagaimana nasib uang kita

Table 3. The Result of Stopword

No	Input	Output
1	negara ini sangat kaya, tapi di korupsi terus. tidak ada juga tindakan tegas untuk para koruptor,. negara hanya menguras rakyat nya.	negara sangat kaya, tapi korupsi terus. tidak ada tindakan tegas para koruptor, negara hanya menguras rakyat
2	tidak ada sanksi tapi wajib, ini saja sudah tidak dapat dipegang sama sekali. bagaimana nasib uang kita	tidak ada sanksi tapi wajib, saja tidak dapat dipegang sama sekali. bagaimana nasib uang kita

text and provides a concise overview of the dataset's overall sentiment landscape.

3. Result and Discussion

Crawling Data

The dataset utilized consists of user comments from YouTube's video titled "[FULL] Kebijakan Tapera, Siapa yang Cari Untung? | SATU MEJA" on the Kompas TV channel, conducted in Indonesian. The data was gathered using the crawling technique through the YouTube API and stored in CSV format. Specifically, for this study, the latest comment was collected on Wednesday, July 12, 2024. Following the data crawling process, a total of 2,268 comments were retrieved from the YouTube video

Data Pre-processing

There are 6 steps in the data pre-processing stage. The results of each stage are as follows.

Data Cleaning

Data cleaning in this study serves multiple objectives, including the removal of special characters, standardizing letter casing to lowercase, eliminating superfluous punctuation, and addressing non-standard text formats like HTML tags, particularly when data originates from web sources (Srivastava et al., 2022), (Kunilovskaya & Plum, 2021). **Table 1** present an illustration of the outcomes achieved through data cleaning as implemented in this research.

Normalization

Text normalization constitutes a critical preprocessing step in text mining, aimed at preparing raw text data for more advanced analytical procedures. By standardizing and cleaning textual data, text normalization enhances the accuracy and efficiency of subsequent analyses. This process involves converting text into a consistent and uniform format (Deviyani & Black, 2022) employing a range of techniques designed to minimize variations and inconsistencies within the text. Through normalization, issues such as spelling discrepancies, capitalization differences, and varying word forms are addressed (Deviyani & Black, 2022), (Zhang et al., 2022). An illustration of the outcomes achieved through the implementation of data normalization in this study is presented in **Table 2**.

Stopword

In text mining, the function of stopwords is to assist in the text preprocessing phase by eliminating common and less informative words (Fan et al., 2023), (Kucukyilmaz & Akin, 2024). By removing stopwords, the data becomes cleaner, with fewer irrelevant words, making it easier to analyze the remaining meaningful content (Kucukyilmaz & Akin, 2024). **Table 3** presents an illustration of the outcomes achieved through stopword as implemented in this research.

Tokenization

Raw text is often unstructured and can vary greatly in terms of format and content. Tokenization transforms raw text into a structured format, making it possible to extract features and apply various sentiment analysis techniques effectively

Table 4. The Result of Tokenization

No	Input	Output
1	negara sangat kaya, tapi korupsi terus. tidak ada tindakan tegas para koruptor, negara hanya menguras rakyat	[‘negara’, ‘sangat kaya’, ‘tapi’, ‘korupsi’, ‘terus’, ‘tidak’, ‘ada’, ‘tindakan’, ‘tegas’, ‘para’, ‘koruptor’, ‘negara’, ‘hanya’, ‘menguras’, ‘rakyat’]
2	tidak ada sanksi tapi wajib, saja tidak dapat dipegang sama sekali. bagaimana nasib uang kita	[‘tidak’, ‘ada’, ‘sanksi’, ‘tapi’, ‘wajib’, ‘saja’, ‘tidak’, ‘dapat’, ‘dipegang’, ‘sama sekali’, ‘bagaimana’, ‘nasib’, ‘uang’, ‘kita’]

Table 5. The Result of Stemming

No	Input	Output
1	[‘negara’, ‘sangat kaya’, ‘tapi’, ‘korupsi’, ‘terus’, ‘tidak’, ‘ada’, ‘tindakan’, ‘tegas’, ‘para’, ‘koruptor’, ‘negara’, ‘hanya’, ‘menguras’, ‘rakyat’]	negara sangat kaya tapi korupsi terus tidak ada tindak tegas para koruptor negara hanya kuras rakyat
2	[‘tidak’, ‘ada’, ‘sanksi’, ‘tapi’, ‘wajib’, ‘saja’, ‘tidak’, ‘dapat’, ‘dipegang’, ‘sama sekali’, ‘bagaimana’, ‘nasib’, ‘uang’, ‘kita’]	tidak ada sanksi tapi wajib tidak dapat pegang sama sekali bagaimana nasib uang kita

Table 6. The Result of Translating

No	Input	Output
1	negara sangat kaya tapi korupsi terus tidak ada tindak tegas para koruptor negara hanya kuras rakyat	the country is very rich but corruption continues there is no firm action against the corrupt state just drain the people
2	tidak ada sanksi tapi wajib tidak dapat pegang sama sekali bagaimana nasib uang kita	there are no sanctions but must not be able to hold at all what is the fate of our money

Table 7. The Result of Data Labeling

No	Comment	Label
1	the country is very rich but corruption continues there is no firm action against the corrupt state just drain the people	Negative
2	there are no sanctions but must not be able to hold at all what is the fate of our money	Negative

(Bagheri et al., 2023), (Yang, 2024). By breaking down raw text into smaller, manageable units—such as words, sub-words, or characters—tokenization facilitates the conversion of unstructured text into a structured format that can be systematically analyzed (Bagheri et al., 2023), (Yang, 2024). **Table 4** presents an illustration of the outcomes achieved through tokenization as implemented in this research.

Stemming

Stemming is a text normalization technique utilizing information retrieval to transform words into their canonical or root form (J. Singh & Gupta, 2017). The primary objective of stemming is to standardize the representation of words by systematically eliminating suffixes and prefixes, thereby ensuring that various morphological forms of a word are treated as a single, unified term (J. Singh & Gupta, 2017; S. Wang et al., 2024). This process enhances the efficacy of text analysis, indexing, and retrieval operations by amalgamating different word variants into a singular, representative form, thus facilitating more coherent and efficient data processing. The result of stemming is presented in **Table 5**.

Translating

Translating the text into a uniform language ensures that all data designated for labeling adheres to a consistent format (Aresta, 2018). This practice mitigates the potential for errors that may arise from linguistic discrepancies. Translation also addresses language ambiguities present in the original text, which is crucial because such ambiguities can complicate the labeling process if the text is not accurately translated (Aresta, 2018). In this study, the comments were translated into English. The outcomes of this translation are presented in **Table 6**.

Data Labeling

Data labeling is conducted to classify each comment into three categories: negative, positive, and neutral. In this research, the TextBlob library is employed for the classification of comments. TextBlob offers a robust API for performing various natural language processing tasks, including sentiment analysis, which entails evaluating the emotional tone conveyed in the text (Hasan et al., 2023). **Table 7** presents an illustration of the outcomes achieved through labeling using TextBlob library as implemented in this research.

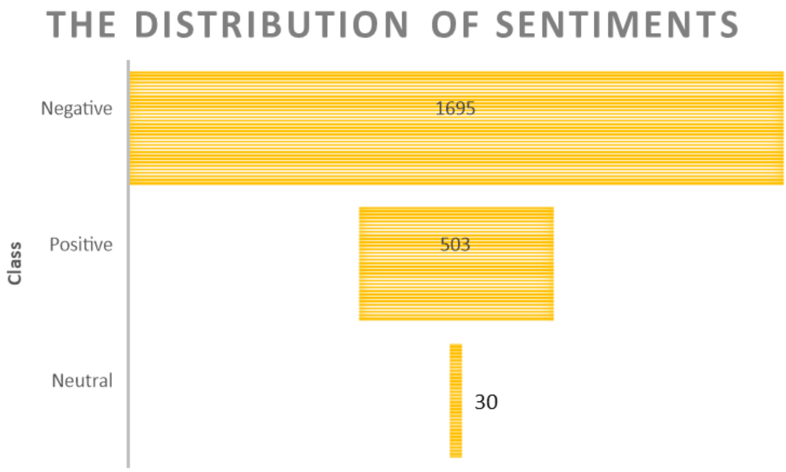

Figure 3. The Distribution of The Sentiments

Result

Upon completion of the sentiment labeling process, the next step involves aggregating the results to determine the distribution of the identified sentiments. This distribution is visually represented in Figure 2 below. The graph offers a clear depiction of the relative proportions of negative, positive, and neutral sentiments in the dataset, providing a foundation for further analysis of potential patterns or trends. The dataset includes a total of 2,228 comments, with the following sentiment distribution: 503 comments (23%) reflect a positive sentiment, 30 comments (1%) are categorized as neutral, and 1,695 comments (76%) exhibit a negative sentiment. This distribution highlights the reliability of the TextBlob library in classifying comments into distinct sentiment categories—positive, neutral, and negative.

The predominance of negative sentiment suggests significant dissatisfaction with the subject matter, which warrants a closer investigation into the underlying causes. The high volume of negative comments, many of which appear to be protests about the existence of TAPERA itself, indicates that the public's discontent is not only related to specific issues within the program but also to broader perceptions of its value and fairness. This overwhelming wave of negative sentiment may reflect frustrations over the perceived ineffectiveness or in equitability of the program. Numerous comments explicitly criticize the financial burden placed on low-income participants through mandatory contributions, with others questioning the transparency of fund management. The recurring theme of protest, particularly the sense that the program exacerbates existing socio-economic inequalities, suggests that TAPERA is viewed by many as failing to address the root causes of housing disparities in Indonesia.

Additionally, it is likely that these negative sentiments are compounded by a sense of disillusionment with government-led initiatives, especially if the public perceives TAPERA as a poorly executed or inequitable solution to a deeply ingrained housing crisis. The dissatisfaction, as seen through these protest-driven comments, may also reflect a broader discontent with the lack of effective social safety nets or alternative solutions for marginalized communities. To gain a deeper understanding of these concerns, a thematic analysis of the comments is crucial. This approach will identify specific aspects of TAPERA that provoke the most resistance and anger, allowing policymakers to better address these criticisms and reframe the program to meet public needs more effectively. By considering the feedback from these protests, the government can adapt TAPERA to reduce negative sentiment and increase public trust in its ability to contribute to solving Indonesia's housing crisis.

Wordcloud Visualization

To further explore the nature of the negative sentiment, a thematic analysis was conducted, focusing on the content of the comments. This analysis delves into the specific aspects of TAPERA that generate negative reactions, revealing that issues such as corruption, mismanagement of funds, and inefficiency were recurrently mentioned by participants. For example, terms like "corruptor," "corruption," "fund," "money," and "corrupt" were frequently used in conjunction with discussions about TAPERA, indicating that these concerns are central to the public's negative sentiment. The thematic analysis thus provides deeper insight into why negative sentiments dominate, identifying corruption and financial mismanagement as the most significant issues contributing to public dissatisfaction.

In addition to sentiment analysis, a Wordcloud visualization was employed to represent the most frequently occurring terms within the comments. This method allows for a quick identification of predominant topics discussed in the feedback. In Wordcloud, terms such as "TAPERA" and "people" appear prominently, underscoring their central role in the conversation. Furthermore, the repeated appearance of negative terms emphasizes the widespread dissatisfaction among commenters. This combined approach of sentiment and thematic analysis, alongside Wordcloud visualization, enables a more nuanced understanding of public sentiment, offering valuable insights into the areas of TAPERA that require attention and improvement.

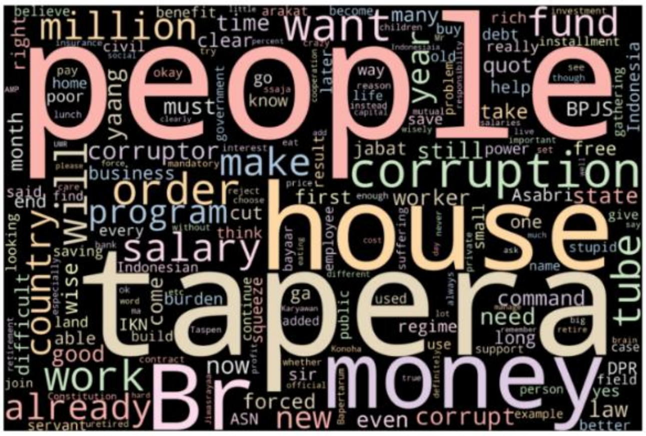

Figure 4. Research Wordcloud Visualization

Discussion

Sentiment analysis using TextBlob is an effective method for understanding public perceptions of TAPERA or similar policies (Kampatzis et al., 2024). TextBlob allows for nuanced assessment of emotional tones in textual data, such as public comments, reviews, or feedback. This analytical tool categorizes sentiments into predefined categories—positive, negative, or neutral—which provides a structured way to gauge the overall public opinion (Hasan et al., 2023). By leveraging TextBlob, we can systematically classify and analyze the sentiments expressed by individuals regarding TAPERA. This process not only helps in identifying general sentiment trends but also highlights specific areas of concern or approval among the public. For instance, if a significant portion of feedback is classified as negative, it may indicate widespread dissatisfaction with certain aspects of the program, such as its implementation or benefits. Conversely, a predominance of positive sentiments could suggest areas where the program is succeeding and meeting public expectations. TextBlob's ability to categorize and summarize sentiments efficiently provides valuable insights into public perception. These insights are crucial for policymakers and program administrators, as they can inform decision-making processes and guide policy refinement. Understanding the emotional landscape of public feedback enables stakeholders to address issues more effectively, make informed adjustments, and enhance the overall effectiveness of the policy. In summary, TextBlob offers a practical and powerful tool for evaluating public sentiment, which can lead to more responsive and well-informed policy adjustments.

In this research, sentiment analysis reveals a significant divergence in public opinion regarding TAPERA (Tabungan Perumahan Rakyat). Specifically, 76% of sentiments are classified as negative, 23% as positive, and 1% as neutral. This distribution provides a nuanced view of public reception and highlights critical areas that require further investigation and intervention. The predominant negative sentiment, representing 76% of the total, indicates a substantial level of dissatisfaction or criticism directed at TAPERA. Several factors may contribute to this negative feedback. A key issue appears to be perceptions of policy ineffectiveness. The public may view TAPERA as failing to meet its objectives due to inadequate coverage, insufficient benefits, or challenges in accessing its provisions. Furthermore, difficulties in policy implementation—such as bureaucratic obstacles, unclear application processes, and discrepancies between the policy's promises and its actual outcomes—could exacerbate negative sentiments. This suggests that the policy's execution may be falling short of public expectations, leading to frustration and disillusionment. Public perception and trust are also crucial in understanding these sentiment patterns. The widespread negativity might reflect a broader erosion of confidence in both TAPERA and the institutions managing it. Skepticism could stem from past experiences or a perceived gap between the policy's goals and it's real-world impact. Additionally, communication issues may play a significant role. If the benefits and mechanisms of TAPERA are not clearly communicated, misunderstandings and negative perceptions may arise. Ensuring that the policy's objectives and advantages are effectively conveyed is essential to improving public understanding and acceptance. Addressing these issues through targeted policy adjustments and enhanced communication strategies is vital for restoring trust and improving TAPERA's effectiveness.

Future research on sentiment analysis using TextBlob for evaluating public perceptions of TAPERA could benefit from exploring several avenues. Comparative studies with other sentiment analysis tools, such as VADER or advanced machine learning models, could reveal their relative effectiveness (Saha et al., 2023). Additionally, capturing more nuanced emotional tones and integrating qualitative methods like interviews could offer deeper insights into public sentiment (Humphreys et al., 2021). Longitudinal studies tracking sentiment over time and cross-platform analyses could provide a more comprehensive view (Liu et al., 2024). Investigating how sentiment trends correlate with policy changes and assessing the impact of communication strategies on public perception would further enhance understanding and guide more effective policy adjustments and public engagement (Liu et al., 2024).

4. Conclusion

TAPERA (Tabungan Perumahan Rakyat) has sparked significant debate in Indonesia. Intending to help low-income communities overcome housing challenges, the program has faced numerous issues since its inception. Key problems include mandatory monthly contributions, which critics argue place a heavy financial burden on low-income individuals, and concerns about transparency in fund management. Sentiment analysis reveals that 76% of public opinions about TAPERA are negative, with only 23% positive and 1% neutral. This negative sentiment suggests widespread dissatisfaction, likely due to perceptions of policy ineffectiveness, bureaucratic hurdles, and poor communication. The analysis highlights the need for a thorough review and improvement of TAPERA's policies. To address these concerns, it is essential for policymakers to enhance transparency, streamline program management, and communicate the benefits of TAPERA. By addressing these issues, TAPERA can improve its effectiveness in achieving its goals and providing meaningful benefits to Indonesia's low-income communities. Effective policy refinement and clear communication are crucial for restoring public trust and ensuring that TAPERA meets its intended objectives

5. References

Aresta, R. (2018). The Influence of Translation Techniques on the Accuracy and Acceptability of Translated Utterances that Flout the Maxim of Quality. *Jurnal Humaniora*, *30*(2), 176. https://doi.org/10.22146/jh.v30i2.33645

Bagheri, A., Giachanou, A., Mosteiro, P., & Verberne, S. (2023). *Natural Language Processing and Text Mining (Turning Unstructured Data into Structured) BT - Clinical Applications of Artificial Intelligence in Real-World Data* (F. W. Asselbergs, S. Denaxas, D. L. Oberski, & J. H. Moore (eds.); pp. 69–93). Springer International Publishing. https://doi.org/10.1007/978-3-031-36678-9_5

Baidarus, M., Febriano, D., Mubarok, D. A., & Ramadhani, M. A. (2023). Kajian Sistematis Kebijakan Skema Pembiayaan Kerja Sama Pemerintah Dengan Badan Usaha (Kpbu) Pada Sektor Perumahan Guna Mengatasi Backlog Di Indonesia. *Jurnal BPPK: Badan Pendidikan Dan Pelatihan Keuangan*, *16*(1), 1–13. https://doi.org/10.48108/jurnalbppk.v16i1.711

Bharadwaj, Lakshay (2023). Sentiment Analysis in Online Product Reviews: Mining Customer Opinions for Sentiment Classification. *International Journal For Multidisciplinary Research*, *5*(5), 1–34. https://doi.org/10.36948/ijfmr.2023.v05i05.6090

Choudhary, K., & Kelley, M. L. (2023). ChemNLP: A Natural Language-Processing-Based Library for Materials Chemistry Text Data. *Journal of Physical Chemistry C*, *127*(35), 17545–17555. https://doi.org/10.1021/acs.jpcc.3c03106

ChandraKala, S & Sindhu, C. (2012). Opinion Mining and Sentiment Classification: a Survey. *ICTACT Journal on Soft Computing*, *03*(01), 420–427. https://doi.org/10.21917/ijsc.2012.0065

Dallo, K. A. M. (2023). Natural language processing for business analytics. *Advances in Engineering Innovation*, *3*(1), None-None. https://doi.org/10.54254/2977-3903/3/2023038

De Filippi, F., & Cocina, G. G. (2022). *Digital Participatory Platforms: Conclusions and the Way Forward BT - Urban Regeneration and Community Empowerment Through ICTs: A Focus on Digital Participatory Platforms (DPPs)* (F. De Filippi & G. G. Cocina (eds.); pp. 121–138). Springer International Publishing. https://doi.org/10.1007/978-3-030-97755-9_5

Deviyani, A., & Black, A. W. (2022). *Text Normalization for Speech Systems for All Languages*. 20–25. https://doi.org/10.21437/s4sg.2022-5

Elgendy, N., & Elragal, A. (2014). Big data analytics: A literature review paper. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8557 LNAI*, 214–227. https://doi.org/10.1007/978-3-319-08976-8_16

Fakhruzzaman, M. N., Jannah, S. Z., Gunawan, S. W., Pratama, A. I., & Ardanty, D. A. (2024). IndoPolicyStats: sentiment analyzer for public policy issues. *Bulletin of Electrical Engineering and Informatics*, *13*(1), 482–489. https://doi.org/10.11591/eei.v13i1.5263

Fan, Y., Arora, C., & Treude, C. (2023). Stop Words for Processing Software Engineering Documents: Do they Matter? *Proceedings - 2023 IEEE/ACM 2nd International Workshop on Natural Language-Based Software Engineering, NLBSE 2023*, 40–47. https://doi.org/10.1109/NLBSE59153.2023.00016

Feng, K. J. K., Gao, A., & Karras, J. S. (2022). Towards Semantically Aware Word Cloud Shape Generation. *UIST 2022 Adjunct - Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. https://doi.org/10.1145/3526114.3558724

Ganesha Putra Henriko , Fahmi Erwin, T. K. (2020). *DI DKI JAKARTA Tabungan Perumahan Rakyat (Tapera) sesuai amanat Undang-Undang Republik Indonesia pembiayaan untuk memperoleh hunian . Namun , sejauhmana Tapera dapat diaplikasikan secara*. *3*(2), 321–332.

Goebel, R. (2019). PRICAI 2019 : Trends in Series Editors. In *Pricai* (Vol. 2, Issue 61672281). https://doi.org/10.1007/978-3-030-29894-4

Gudankwar, A., Pratik Milind Mendhe, Lukesh Namdeo Oghare, & Atharva Ramu Yemde. (2024). Sentiments of Public Opinion. *International Journal of Scientific Research in Computer Science, Engineering and*

Information Technology, *10*(2), 459–461. https://doi.org/10.32628/cseit2410239
Hasan, M., Ahmed, T., & Uddin, M. P. (2023). *Leveraging Textual Information for Social Media News Categorization and Sentiment Analysis*. https://doi.org/10.2139/ssrn.4425901
Humphreys, L., Lewis, N. A., Sender, K., & Won, A. S. (2021). Integrating Qualitative Methods and Open Science: Five Principles for More Trustworthy Research*. *Journal of Communication*, *71*, 855–874. https://doi.org/10.1093/joc/jqab026
Indrani P, Janaki P, Gayathri P, Chandrahasini P, Charitha Reddy, Apparao G, Rajeshwari. (2024). Product Review Sentiment Analysis. *International Journal For Multidisciplinary Research*, *6*(3), 1–8. https://doi.org/10.36948/ijfmr.2024.v06i03.20551
Jayachandran, S. (2023). The inherent trade-off between the environmental and anti-poverty goals of payments for ecosystem services. *Environmental Research Letters*, *18*(2). https://doi.org/10.1088/1748-9326/acb1a7
Kampatzis, A., Sidiropoulos, A., Diamantaras, K., & Ougiaroglou, S. (2024). Sentiment Dimensions and Intentions in Scientific Analysis: Multilevel Classification in Text and Citations. *Electronics (Switzerland)*, *13*(9). https://doi.org/10.3390/electronics13091753
Kucukyilmaz, T., & Akin, T. (2024). A Feature-based Approach on Automatic Stopword Detection. *Lecture Notes in Networks and Systems*, *825*, 51–67. https://doi.org/10.1007/978-3-031-47718-8_4
Kunilovskaya, M., & Plum, A. (2021). Text Preprocessing and its Implications in a Digital Humanities Project. *International Conference Recent Advances in Natural Language Processing, RANLP*, *2021-Septe*, 85–93. https://doi.org/10.26615/issn.2603-2821.2021_013
Lindgren, C. J., Wang, W., Upadhyay, S. K., & Kobayashi, V. B. (2023). Sentiment Analysis for Organizational Research. *Research in Occupational Stress and Well Being*, *21*(November 2023), 95–117. https://doi.org/10.1108/S1479-355520230000021006
Lippman, S. A., Libby, M. K., Nakphong, M. K., Arons, A., Balanoff, M., Mocello, A. R., Arnold, E. A., Shade, S. B., Qurashi, F., Downing, A., Moore, A., Dow, W. H., & Lightfoot, M. A. (2023). A guaranteed income intervention to improve the health and financial well-being of low-income black emerging adults: study protocol for the Black Economic Equity Movement randomized controlled crossover trial. *Frontiers in Public Health*, *11*(November). https://doi.org/10.3389/fpubh.2023.1271194
Liu, H., Tsang, S., Wood, A., & Tong, X. (2024). Longitudinal Sentiment Analysis with Conversation Textual Data. *Fudan Journal of the Humanities and Social Sciences*, *0123456789*. https://doi.org/10.1007/s40647-024-00417-0
Lohmann, S., Heimerl, F., Bopp, F., Burch, M., & Ertl, T. (2015). Concentri cloud: Word cloud visualization for multiple text documents. *Proceedings of the International Conference on Information Visualisation*, *2015-Septe*, 114–120. https://doi.org/10.1109/iV.2015.30
Saha, S., Showrov, M. I. H., Rahman, M. M., & Majumder, M. Z. H. (2023). *VADER vs. BERT: A Comparative Performance Analysis for Sentiment on Coronavirus Outbreak BT - Machine Intelligence and Emerging Technologies* (M. S. Satu, M. A. Moni, M. S. Kaiser, & M. S. Arefin (eds.); pp. 371–385). Springer Nature Switzerland.
Singh, J., & Gupta, V. (2017). A systematic review of text stemming techniques. *Artificial Intelligence Review*, *48*. https://doi.org/10.1007/s10462-016-9498-2
Singh, N. T., Mishra, A. M., Singh, A., & Chanu, A. D. (2023). Advancing Spelling Correction through Natural Language Processing and TextBlob: A Context-Aware Approach. *2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management, IC-RVITM 2023*, *March*, 1–6. https://doi.org/10.1109/IC-RVITM60032.2023.10435139
Soeprapto, D. D. (2021). Determining the key success factors in the organization of BP Tapera. *International Journal of Research in Business and Social Science (2147- 4478)*, *10*(6), 42–55. https://doi.org/10.20525/ijrbs.v10i6.1270
Srivastava, C., Pratishtha, ;, & Kaur, N. (2022). *An Overview on Data Cleaning on Real World Data*. https://doi.org/10.36227/techrxiv.21064039.v1
Thakur, M., Arora, A. M., Sandhu, Y. S., & Singh, G. (2024). Sentimental Journeys: Novel Insight into Textual Information. *Advancements in Communication and Systems*, 305–313. https://doi.org/10.56155/978-81-955020-7-3-26
Udayana, I. P. A. E. D., Indrawan, I. G. A., & Putra, I. P. D. G. A. (2023). Decision Support System for Sentiment Analysis of Youtube Comments on Government Policies. *Journal of Computer Networks, Architecture and High Performance Computing*, *5*(1), 27–37. https://doi.org/10.47709/cnahpc.v5i1.1999
Wang, C., Qiu, M., Zhang, T., Liu, T., Li, L., Wang, J., Wang, M., Huang, J., & Lin, W. (2022). EasyNLP: A Comprehensive and Easy-to-use Toolkit for Natural Language Processing. *EMNLP 2022 - 2022 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Demonstrations Session*, *2*, 22–29. https://doi.org/10.18653/v1/2022.emnlp-demos.3
Wang, S., Zhuang, S., & Zuccon, G. (2024). *Large*

Language Models Based Stemming for Information Retrieval: Promises, Pitfalls and Failures. 2492–2496. https://doi.org/10.1145/3626772.3657949
Yang, J. (2024). *Rethinking Tokenization: Crafting Better Tokenizers for Large Language Models*. https://doi.org/10.1075/ijchl.00023.yan
Zhang, H., Cheng, Y.-C., Kumar, S., Huang, W. R., Chen, M., & Mathews, R. (2022). *Capitalization Normalization for Language Modeling with an Accurate and Efficient Hierarchical RNN Model*. 2–6. http://arxiv.org/abs/2202.08171