Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Optimizing Twitter Sentiment Analysis on Tapera Policy Using SVM and PSO

Ahmad Al Kaafi[1)*], Suparni[2)], Hilda Rachmi[3)], Ahmad Maulana[4)], Ririn Nurtriani[5)]
[1,2,3,4,5)]Universitas Bina Sarana Informatika, Indonesia
[1)]ahmad.akf@bsi.ac.id, [2)]suparni.spn@bsi.ac.id, [3)]hilda.hlr@bsi.ac.id, [4)]17220894@bsi.ac.id,
[5)]12220398@bsi.ac.id

Submitted :Nov 9, 2024 | **Accepted** : Dec 3, 2024 | **Published** : Jan 9, 2025

Abstract: This study aims to analyse the sentiment of Twitter users towards the Public Housing Savings, known locally as Tapera (Tabungan Perumahan Rakyat) policy in Indonesia using the Support Vector Machine (SVM) algorithm optimised by Particle Swarm Optimization (PSO). In recent years, social media has emerged as a primary platform for individuals to express their views and opinions on public policies. The government programme, Tapera, which was designed to increase access to housing for the public, attracted considerable attention, with a range of responses, including both positive and negative sentiments. The methodology employed in this study comprised the collection of data from Twitter, the processing of text, and the application of SVM-based classification techniques, reinforced by PSO, with the objective of enhancing the accuracy and efficiency of the model. The results demonstrated that the PSO-optimised SVM model exhibited an accuracy of 85%, accompanied by an Area Under Curve (AUC) value of 0.84 and a ROC curve that indicated the model's notable capacity for differentiating between positive and negative sentiments. These findings indicate the existence of certain sentiment patterns that can be utilised for the evaluation and improvement of Tapera policies. In conclusion, this research is expected to provide a comprehensive picture of the public response to the Tapera policy and present an analytical model that can be applied to evaluate other policies. Further research is recommended to expand data coverage and develop algorithms to achieve more accurate results.

Keywords: Particle Swarm Optimization (PSO); Policy; Sentiment Analysis; Support Vector Machine (SVM), Tapera;

INTRODUCTION

The Policy on Public Housing Savings, known locally as Tapera (Tabungan Perumahan Rakyat) is one of the strategic initiatives launched by the Indonesian government with the noble intention of addressing the problem of home ownership for the community. One possible way to provide housing finance, especially for low-income people, is through the Public Housing Savings (Asril, Rifai, & Shebubakar, 2022).

The Tapera programme offers participants the opportunity to make periodic deposits over a specified period of time. These deposits can be utilized for housing finance and/or returned after participation ends, along with the resulting interest (Undang-Undang Republik Indonesia Nomor 4 Tahun 2016 Tentang Tabungan Perumahan Rakyat, n.d.). It is hoped that Tapera can provide a solution for Indonesian citizens to obtain housing finance (Putra, Fahmi, & Taruc, 2020).

The Tapera Law represents the government's commitment to ensuring the fulfillment of people's basic needs by providing housing as part of their rights (Tania, Novienco, & Sanjaya, 2021). Since the passing of the Tapera Law on 23 February 2016, a number of questions have been raised by the public, both in terms of formalities and materials. In 2020, amidst challenging economic, social, and health conditions, the government issued the PP Tapera as the implementing regulation for Tapera Law Number 4 of 2016 on 20 May 2020. Some members of the public felt that the implementation of PP Tapera at that time might not have been the most appropriate given the impact of the pandemic on purchasing power (Tania et al., 2021).

On 20 May 2024, President Joko Widodo issued a new regulation regarding contributions to the Public Housing Savings Programme (Tapera) through Government Regulation Number 21 of 2024 (Indonesia, n.d.). The contribution is set at 3 per cent of the worker's salary, with 0.5 per cent borne by the employer and 2.5 per cent by the worker, while self-employed workers bear the entire contribution. It would be greatly appreciated if employers

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

could kindly deposit Tapera contributions every month, no later than the 10th of the following month, into the Tapera Fund Account.

The paucity of information and socialisation received by the public has resulted in a range of responses. One issue is the lack of clarity regarding the management and investment of the programme's funds, which gives rise to concerns about its compatibility with the public interest. Furthermore, the absence of transparency regarding the benefits, requirements, and performance of the Tapera programme has instilled a sense of apprehension among the public regarding the programme's operational specifics. Nevertheless, there are a number of individuals who endorse the policy, as they believe it will assist workers, particularly those with low incomes, to gain access to affordable credit with fixed interest rates and flexible repayment terms. Despite the programme's laudable objectives, its implementation is not without challenges and intricate social dynamics. It is therefore crucial to gain insight into how the general public responds to this policy in order to identify potential avenues for improvement and optimise programme implementation.

It is of great importance to analyse the prevailing sentiment within the community with regard to the issues that are currently under discussion, as this can prove invaluable in the formulation of effective policy (Fatouros, Soldatos, Kouroumali, Makridis, & Kyriazis, 2023). The response of the community to this policy may be indicative of the level of acceptance, trust, and participation in the programme. In the absence of a comprehensive understanding, it may prove challenging to implement effective and targeted policy adjustments.

Sentiment analysis, otherwise known as opinion mining, is a field of textual data management that studies people's opinions, sentiments, evaluations, behaviours and emotions, which can be used as evaluation materials (Rachmi, Suparni, & Kaafi, 2021)(Kristiyanti, Kaafi, Purwaningsih, Nurelasari, & Nisa, 2023)(Sarirete, 2021)(Md Suhaimin et al., 2023). While some previous studies (Hendrastuty et al., 2021)(Kaafi et al., 2012)(Asro'i & Februariyanti, 2022)(Sasmito Aribowo, 2018)(Kurniawan & Waluyo, 2022)(Anggraini & Utami, 2021)(Supriyanto, Alita, & Isnain, 2023)(Arsi, Wahyudi, & Waluyo, 2021) have successfully applied sentiment analysis in the context of public policy and public services, research on the sentiment analysis of housing policies such as Tapera remains relatively limited.

LITERATURE REVIEW

Previous research has employed the Support Vector Machine (SVM) method and TF-IDF feature extraction to analyse public sentiment towards the lockdown policy implemented by the Jakarta government during the ongoing Coronavirus Disease 2019 (Covid-19) pandemic. Of the 1,377 Indonesian tweets analysed, 68.75% expressed a positive sentiment, while 31.25% conveyed a negative one. The SVM model demonstrated a 74% accuracy rate, 75% precision, 92% recall, and an 83% F1-score. The study concluded that the method is an effective means of classifying public sentiment and recommended the utilisation of optimisation algorithms, such as Firefly, to enhance the accuracy of the classification process (Rahman Isnain, Indra Sakti, Alita, & Satya Marga, 2021).

Another study employed the Naive Bayes classification method to analyse the sentiment surrounding the Indonesian government's 'New Normal' policy during the ongoing pandemic. This study employed Indonesian tweet data collected between 6 July 2020 and 25 July 2020. The application of TF-IDF and N-Gram feature extraction techniques yielded the highest accuracy of 84% when Naive Bayes was used with Trigram type N-Grams, with precision, recall, and F1-Score at 84%, 86%, and 85%, respectively. The results demonstrate that the Naive Bayes algorithm with N-Gram feature extraction exhibits robust performance in classifying public sentiment regarding the government's New Normal policy (Marga, 2022).

The following research project analyses public sentiment on Twitter regarding the handling of the Coronavirus Disease 2019 (Covid-19) in Indonesia using the Naive Bayes Classification method. The data set comprised 10,000 tweets published between 5 and 7 October 2020, with the content related to the implementation of the PSBB, the use of mandatory masks, and the imposition of curfews. The results of the analysis indicate that the majority of public responses are negative towards the handling of the Corona Virus Disease 2019 (COVID-19) pandemic, particularly with regard to the implementation of the Policy for Containment and Management of the Corona Virus Disease 2019 (PSBB), the use of mandatory masks, and the imposition of curfews. The Naive Bayes classification method demonstrated an accuracy rate of 87.34%, a sensitivity of 93.43%, and a specificity of 71.76%. Additionally, the visualisation of the classification results illustrated the frequency of the most frequently occurring words and a word cloud depicting the words with the highest frequency from the positive and negative sentiments (Naraswati et al., 2021).

The following section presents the findings of a sentiment analysis of public opinion regarding the Kampus Merdeka programme on the social media platform Twitter. The Naive Bayes Classifier algorithm was employed for this purpose. The data was obtained from the tweets and retweets of Twitter users between November 2021 and March 2022. The analytical process comprises the following steps: data collection, text preprocessing, TF-IDF calculation, Naive Bayes algorithm implementation, and k-fold cross-validation. The results of the analysis demonstrate that the classification of positive and negative sentiments exhibits an accuracy of 60%, with the k-fold cross-validation approach yielding the highest level of accuracy. The results are presented in the form of a

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

word cloud, which illustrates the key words associated with positive and negative sentiments. Although the accuracy results of the Naive Bayes Classifier algorithm are relatively low, evaluation with the k-fold cross-validation method is effective enough to yield maximum accuracy results (Febriyani & Februariyanti, 2023).

This research compares the performance of Random Forest and Support Vector Machine (SVM) models in flood classification in Jakarta , which is often affected by high rainfall phenomena and geographical factors that exacerbate the risk of inundation. Based on Jakarta flood data from 2016 to 2020 obtained from Kaggle, this study applies the stages of data acquisition, pre-processing, classification, and evaluation. Models were evaluated using accuracy, precision, recall, and F1-score, with Random Forest showing a more stable performance than SVM on various evaluation metrics. Random Forest's advantage comes from its ensemble approach which results in more consistent classification performance and is computationally efficient, especially on large datasets. SVM requires more careful parameter tuning to achieve optimal performance but proves more sensitive to data changes. These findings suggest that Random Forest is more reliable for flood classification under various data conditions. This research recommends adding additional variables such as drainage capacity, land use pattern, and elevation data to improve the accuracy and robustness of the model in predicting flood events(Purwati & Pristyanto, 2024).

METHOD

This research project is concerned with the examination of public opinion about Tapera through the utilisation of Twitter data. The data employed were Indonesian tweets, encompassing mentions, replies, likes, and retweets. In order to facilitate the collection and analysis of the data, the researcher implemented a series of methodological procedures. Figure 1 below will illustrate the methodological steps taken in the research framework:

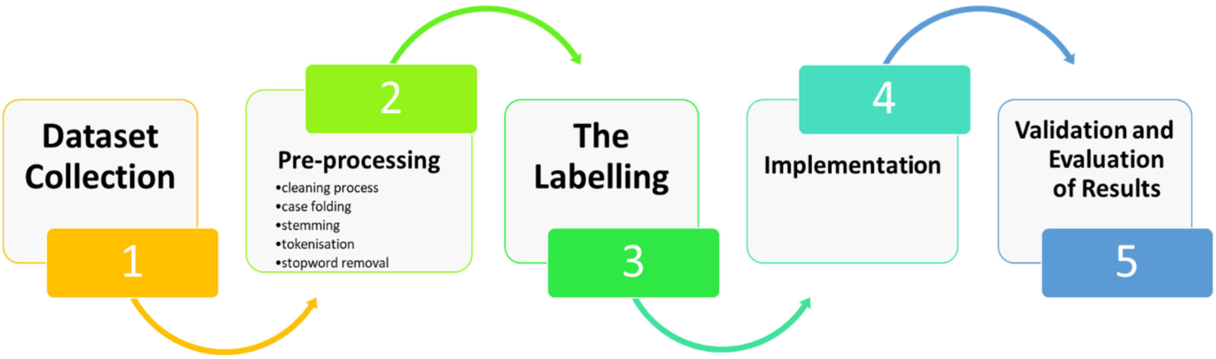

Fig. 1 The Research Framework

1. Dataset Collection: The data was collected using data crawling techniques. The researcher extracted 8,546 tweets from the beginning of June 2024 with the keywords 'Tapera' or 'TAPERA' in Indonesian on Twitter, then saved the data in a .csv file for analysis.
2. Pre-processing
   This process aims to structure the data into a more structured format, which includes:
   a. The cleaning process involved the following steps: The tweet text was cleaned by removing any superfluous characters, including HTML, URLs, emojis, mentions, hashtags, as well as slang words drawn from Indonesian vocabulary.
   b. The next stage of the process is case folding. The entire text is converted to standard format (lower case).
   c. Stemming is the process of reducing a word's lexical complexity by removing affixes, thereby identifying the base word. The function removes affixes in order to identify the base word.
   d. The process of tokenisation involves the division of text into discrete units, or tokens, which are then subjected to further analysis. The process of dividing the tweet text into specific units of meaning, known as "tokens."
   e. The process of stopword removal is as follows: The removal of superfluous words (stopwords), including "and," "or," and "then," is also a necessary step.
3. The labelling of each word that has been separated during the pre-processing stage is used to determine the polarity of the word. A list of Indonesian opinion words was employed to identify positive and negative words. Tweets with a greater number of positive words were assigned a value of +1, while those with a greater number of negative words were assigned a value of -1.
4. Implementation: The sentiment analysis process was performed using the Support Vector Machine (SVM) classification algorithm, with the application of Particle Swarm Optimisation to generate the most accurate patterns.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

5. Validation and Evaluation of Results Validation was conducted in 10 sections, with training and testing processes 10 times. The results and accuracy were assessed using confusion matrix and ROC curve to estimate the AUC results. The confusion matrix is used to evaluate the classification model, and the ROC curve evaluates the prediction results by displaying the classification performance based on the class distribution.

RESULT

The datasets comprising the Twitter conversation corpus pertaining to the keyword 'Tapera' in Indonesian were collated through the utilisation of the Python programming language, in conjunction with the Twitter API Authentication. The results of the data set retrieval are presented in Table 1.

Table 1. Dataset Collection

Created_At	Text	Tweet_Url	User_Id_Str	Username
Sat Jun 01 23:53:51 +0000 2024	Gotong royong nyumbang koruptor @BP_Tapera	https://x.com/G0rb4cev/status/1797053983397454081	1.61E+18	G0rb4cev
Sat Jun 01 23:53:16 +0000 2024	Yg ikut Tapera sekarang 50 THN nanti sdh koit semua. Hasil Tapera cuma bisa beli tanah satu meter.	https://x.com/MuhammadDjazul1/status/1797053834679996586	3.3E+09	MuhammadDjazul1
Sat Jun 01 23:53:09 +0000 2024	Lowongan pekerjaan baru dan banyak dibutuhkan: debt collector tapera	https://x.com/rubahkeIabu/status/1797053804774649922	2.69E+08	rubahkeIabu
Sat Jun 01 23:53:08 +0000 2024	@tanyakanrl murder karna nolak tapera	https://x.com/Kapikiran_/status/1797053802857861345	1.41E+18	Kapikiran_
Sat Jun 01 23:52:35 +0000 2024	@Aryprasetyo85 Aturan Tapera tdk cocok buat pekerja swasta apa lagi yg outsourcing Mau dapat kan kerja harus bayar 1 jt hingga 2 jt lalu dikotrak 1 thn atau 6 bln Setelah nabung 1 thn terus di PHK dan jadi pekerja serabutan Jika terjadi pada seribu. Puluhan ribu pekerja uang raib di Tapera https://t.co/IqklGitgOG	https://x.com/Manjada9547/status/1797053660767404446	1.77E+18	Manjada9547
…	…	…	…	…
Sat Jun 29 23:14:52 +0000 2024	hidihhh .. jam segini set. 6 udh ngelewatin banjiirr tapera ipar DKI jakarta anies jokowi elaelo kaesang vina menyala judol bansos gacoan	https://x.com/NurilHiLma2022/status/1807191031231946783	1.32E+18	NurilHiLma2022
Sat Jun 29 23:22:14	@ekky1995 kritik mah kritik aja bro tp kalo opini lo dicounter orang lain jangan kena mental trs menganggap mereka anti kritik. lagian gabut amat kritik anies yg skrg dah jd manusia biasa knp ga kritik	https://x.com/prollls/status/1807192886292975944	1.48E+18	prollls

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

+0000 2024	kominfo yg kerjanya ga becus? ga kritik tapera dan masalah lainnya yg lg rame itu?			
Sat Jun 29 23:30:00 +0000 2024	Meski Didemo Buruh Iuran Tapera Tetap Berlaku 2027 https://t.co/Ws1HwFC5OZ	https://x.com/okezonenews/status/180719 4838389207508	47 27 47 31	okezonenews
Sat Jun 29 23:40:55 +0000 2024	@Jumianto_RK Masih ada tapera mas untuk rakyat kecil hehe	https://x.com/LianJibran76553/status/18 07197586182529418	1.69 E+18	LianJibran76553
Sat Jun 29 23:56:15 +0000 2024	.. jam set. 6 pagi ngelewatin banjir moga hari ini dpt laba jualan â€ढ़ Kitabisa https://t.co/isucoOiCWJ Hilma Nuril Difabel yatimBerjualan. Panjang umur kebaikan. tapera ipar DKI jakarta anies jokowi elaelo kaesang vina menyala judol gacoan https://t.co/RlE58QhHpu	https://x.com/NurilHiLma2022/status/18 07201447601373272	1.32 E+18	NurilHiLma2022

From the entire dataset, comprising 8,546 tweet samples with the keyword 'Tapera' on the Twitter social media platform, a text preprocessing stage was conducted. The preprocessing stage comprises a number of steps, including cleaning, case folding, stemming, tokenisation and stopword removal. The objective of this pre-processing stage is to transform the raw data into a more structured and refined format. The outcomes of the pre-processing stage are presented in Table 2.

Table 2. The results of the pre-processing

Text	Text Cleaning	Text Case Folding	Text Stemming	Text Tokenizing	Text Stopword / Normalization	Text Filtering
Gotong royong nyumbang koruptor @BP_Tapera	Gotong royong nyumbang koruptor	gotong royong nyumbang koruptor tapera	[gotong, royong, nyumbang, koruptor, tapera]	[gotong, royong, nyumbang, koruptor, tapera]	[gotong, royong, nyumbang, koruptor, tapera]	[gotong, royong, nyumbang, koruptor, tapera]
Yg ikut Tapera sekarang 50 THN nanti sdh koit ...	Yg ikut Tapera sekarang 50 THN nanti sdh koit ...	yg ikut tapera sekarang thn nanti sdh koit sem...	[yg, ikut, tapera, sekarang, thn, nanti, s...	[yg, ikut, tapera, sekarang, thn, nanti, sdh, ...	[yang, ikut, tapera, sekarang, tahun, nanti, s...	[tapera, koit, hasil, tapera, beli, tanah, meter]
Lowongan pekerjaan baru dan banyak dibutuhkan:...	Lowongan pekerjaan baru dan banyak dibutuhkan:...	lowongan pekerjaan baru dan banyak dibutuhkan ...	[lowong, kerja, baru, dan, banyak, butuh, debt...	[lowong, kerja, baru, dan, banyak, dibut...	[lowong, kerja, baru, dan, banyak, dibut...	[lowong, kerja, butuh, debt, collector, tapera]
@tanyakanrl murder karna nolak tapera	murder karna nolak tapera	murder karna nolak tapera	[murder, karna, nolak, tapera]	[murder, karna, nolak, tapera]	[murder, karena, nolak, tapera]	[murder, nolak, tapera]
@Aryprasetyo85 Aturan Tapera tdk cocok buat pe...	Aturan Tapera tdk cocok buat pe...	aturan tapera tdk cocok buat pekerja	[atur, tapera, tdk, cocok, buat, kerja, swas...	[atur, tapera, tdk, cocok, buat,	[atur, tapera, tidak, cocok, buat, pekerja, ...	[atur, tapera, cocok, kerja, swasta, outsourci...

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

		swasta ap...		pekerja, sw...		

The Inset Lexicon method, which employs a positive dictionary, a negative dictionary and an Indonesian lexicon, represents a particular approach to dictionary-based sentiment analysis. This method is based on a list of words or phrases (lexicon) that have been classified as positive, neutral or negative, and which are used to determine the sentiment polarity in the text. Following the preprocessing stage, each word in the text is matched against the positive and negative dictionaries. In the event that a word is identified within the positive dictionary, it will be designated as positive. Conversely, if the word is found in the negative dictionary, it will be labelled as negative. Once all the words have been labelled, the sentiment score is calculated by subtracting the number of words classified as negative from the number of words classified as positive. For example, if the text contains a greater number of positive words, it is deemed to possess a positive sentiment; conversely, if it comprises a greater number of negative words, it is considered to have a negative sentiment. In the event that the calculated sentiment score is an even number, the text is deemed to evince a neutral sentiment. The classification of the text as a whole as positive, negative, or neutral will be based on the calculated sentiment score. The results of the pre-processing are presented in Table 3 and Figure 3.

Table 3. Sentiment Polarity

Text Filtering	Polarity Score	Indonesia Sentiment
[gotong, royong, nyumbang, koruptor, tapera]	0	Neutral
[tapera, koit, hasil, tapera, beli, tanah, meter]	-3	Negative
[lowong, kerja, butuh, debt, collector, tapera]	-3	Negative
[murder, nolak, tapera]	0	Neutral
[atur, tapera, cocok, kerja, swasta, outsourci...	3	Positive
...	...	...
[hidihhh, jam, gin, set, ngelewatin, banjiirr,...	-2	Negative
[kritik, mah, kritik, bro, opini, lo, dicounte...	-38	Negative
[demo, buruh, iur, tapera, laku]	-3	Negative
[rk, tapera, mas, rakyat]	0	Neutral
[jam, set, pagi, ngelewatin, banjir, moga, lab...	4	Positive

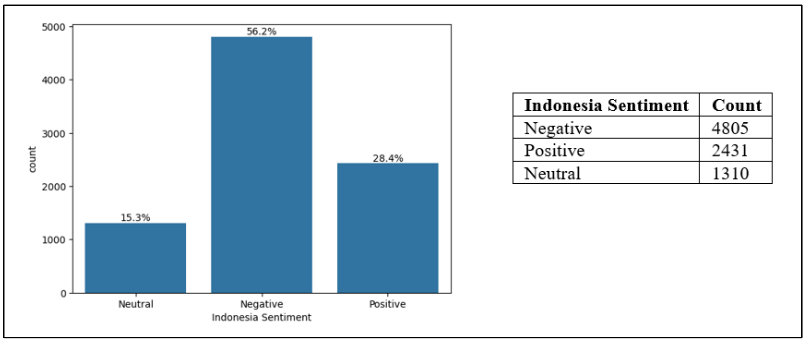

Fig 3. Sentiment Graph

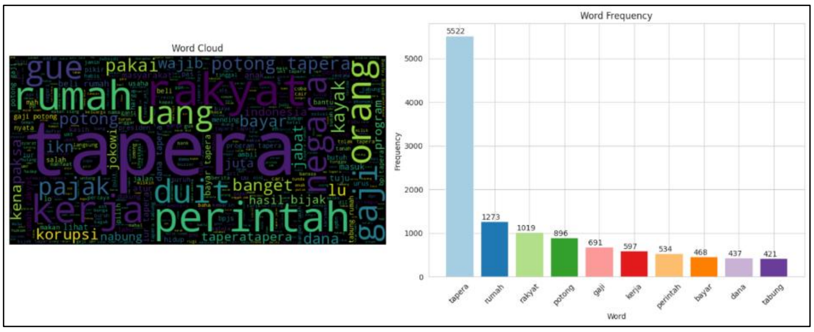

Fig 4. Word Cloud & Word Frequency

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X


Fig 5. Positive Word Cloud


Fig 6. Neutral Word Cloud


Fig 7. Negative Word Cloud

In the visualisation of Fig 4, the word cloud presents the most frequently occurring words within the dataset of tweets pertaining to the subject of 'Tapera'. The size of the word in the word cloud is indicative of its frequency of occurrence within the text, whereby larger words are associated with a higher frequency. Furthermore, the word frequency graph on the right illustrates the distribution of occurrences of the most frequently used words. The term "tapera" is the most frequently occurring word, with a total of 5,522 instances, followed by the word "people," which has 1,273 occurrences, and the word "government," which has 919 instances. This suggests that the subject of Tapera is closely associated with matters pertaining to the general public, the government, and other public policies. The visualisation offers an overview of the words that are most prevalent in Twitter conversations about Tapera. The high frequency of the words 'tapera', 'people', and 'government' suggests that this topic is closely related to public policies that have a significant impact on the general public.

Figure 5 presents the positive word cloud, which comprises solely those words that evince positive sentiment. The most salient words in the dataset, exhibiting positive sentiment, are "tapera," "house," and "people." This indicates that discussions concerning houses and the welfare of the general public are characterised by a positive sentiment, which may signify a sense of optimism or gratitude among some users. This visualisation highlights that positive conversations about Tapera frequently pertain to constructive matters, such as housing for the people and beneficial policies.

Figure 6 depicts the neutral word cloud, which reveals that words such as 'tax', 'deduct' and 'salary' appear with high frequency. This suggests that people have a neutral or ambiguous perception of the financial aspects of the programme. It is evident that the issue of payroll deductions for Tapera contributions is a significant consideration for many individuals.

Figure 7 depicts the negative word cloud, which encompasses terms with negative associations, including "cut," "government," "work," and "salary." The recurrence of these terms in the context of negative sentiment suggests that a subset of users may be discontented or dissatisfied with aspects of the Tapera policy, particularly those pertaining to salary reductions or other elements perceived as detrimental. The aforementioned negative word cloud indicates that the primary source of discontent with Tapera often pertains to the reduction of income or perceived inequities in its implementation. This offers valuable insights for policymakers seeking to comprehend the underlying reasons for public resistance or criticism.

Following the labelling process, a prediction analysis was conducted using the Support Vector Machine (SVM) algorithm with the application of Particle Swarm Optimization (PSO) to produce the optimal pattern with high accuracy. The algorithm was tested in 10 parts, with the training and testing process repeated 10 times. The results and accuracy were assessed using a confusion matrix and a receiver operating characteristic (ROC) curve to estimate the area under the curve (AUC) results. The confusion matrix is used to evaluate the classification model, and the ROC curve evaluates the prediction results by displaying the classification performance based on the class distribution.

Table 4. Classification Report SVM

	precision	**recall**	**f1-score**	**support**
Negative	0.88	0.93	0.91	507
Neutral	0.74	0.56	0.63	131
Positive	0.82	0.84	0.83	216
accuracy			0.85	854
macro avg	0.81	0.78	0.79	854
weighted avg	0.84	0.85	0.85	854

Table 4 illustrates the efficacy of the sentiment classification model with three distinct classes. The classification is divided into three categories: negative, neutral, and positive. The model demonstrates high precision in the negative (0.88) and positive (0.82) classes, but lower precision in the neutral class (0.74). The highest recall was observed for the negative class (0.93), while the neutral class recall was only 0.56, indicating that a considerable number of neutral instances were misclassified. The F1-score demonstrated a favourable

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

balance for the negative (0.91) and positive (0.83) classes, but a less optimal outcome for the neutral class (0.63). The overall accuracy of the model was 0.85, and the weighted average of precision and recall was also close to 0.85, indicating a commendable performance on classes with a larger support base. However, there is a clear necessity for improvement in the detection of neutral sentiment.

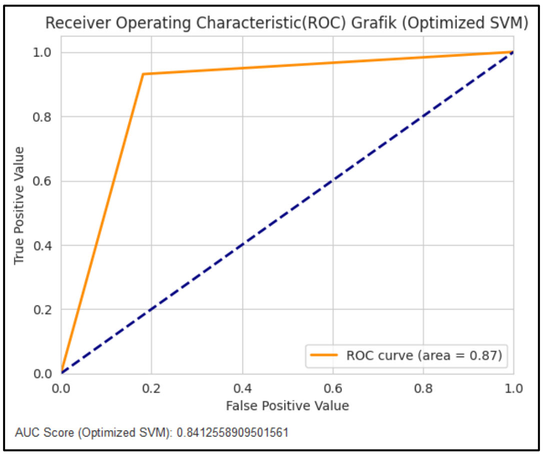

Fig 8. ROC Curve (Optimized SVM)

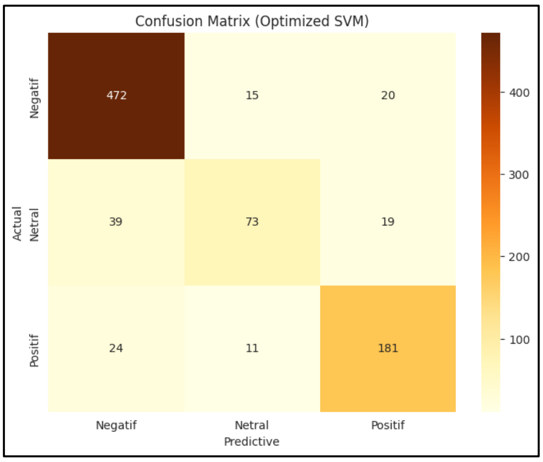

Fig 9. Confusion Matrix (Optimized SVM)

Fig 8. The receiver operating characteristic (ROC) curve is a tool used for the evaluation of classification models in terms of their performance. In this ROC curve, the x-axis depicts the false positive value, also known as sensitivity, while the y-axis represents the true positive value. The dotted diagonal line on the ROC curve serves as a baseline against which the performance of the model can be evaluated. This indicates how the model would perform if predictions were made randomly. The yellow ROC curve represents the performance of the tested model. A model that performs well in distinguishing between positive and negative classes will be situated closer to the upper left corner. The AUC (Area Under Curve) of 0.841 indicates that the model demonstrates a relatively high degree of accuracy in predicting the class. An AUC value approaching 1.0 indicates that the model exhibits excellent performance, demonstrating the capacity to minimise prediction errors in the detection of both false positives and false negatives.

Figure 9 presents a classification result that offers a more granular insight into the number of accurate and erroneous predictions made by the model. The matrix illustrates the model's predictions for the three sentiment classes: The three categories are negative, neutral, and positive. The rows of the confusion matrix represent the actual labels derived from the data set, while the columns represent the labels predicted by the model. A total of 472 instances with negative actual labels were correctly predicted as negative, while 29 instances with neutral actual labels were predicted as negative. In contrast, 73 neutral instances were correctly predicted as neutral. Of the 181 instances with positive actual labels, 157 were correctly predicted as positive, while only 24 were incorrectly predicted as negative and 11 as neutral.

From the confusion matrix, it can be seen that the model performs well in predicting the negative and positive classes, with relatively low prediction errors. However, there is some confusion in predicting the neutral class, with 29 instances incorrectly predicted as negative and 19 instances incorrectly predicted as positive. This indicates that the model has challenges in distinguishing neutral sentiment from other sentiments, which could be due to the similarity of words or phrases in neutral text with other classes.

DISCUSSIONS

The findings of this research demonstrate that the Support Vector Machine (SVM) method, when optimised with Particle Swarm Optimisation (PSO), is capable of enhancing the precision of sentiment analysis on Tapera policy on Twitter. The model demonstrates a high degree of accuracy, with a 85% success rate and an AUC (Area Under Curve) value of 0.84. This indicates a reliable capacity to distinguish between positive, negative, and neutral sentiments. The results of this analysis align with those of previous studies, such as that conducted by Arsi et al. (2021), which demonstrated that PSO optimisation can enhance the efficacy of the SVM algorithm in text classification. Nevertheless, the model's performance in the neutral sentiment category remains suboptimal (recall 0.56), indicating the difficulty in capturing the distinctive contextual patterns associated with neutral sentiment.

A comparison with related research demonstrates that lexicon-based approaches and traditional algorithms, such as Naive Bayes, frequently encounter limitations in their ability to process text with complex contextual nuances. The application of SVM optimised with metaheuristic algorithms such as PSO provides a more effective solution for improving classification accuracy.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

A limitation of this study is the data coverage, which is derived exclusively from Twitter and may not fully reflect the overall public opinion. Additionally, the model has not fully captured the nuances of ambiguous sentiment, particularly in the neutral category. To enhance future analysis results, advanced approaches, such as the integration of deep learning models, are necessary.

CONCLUSION

This study successfully demonstrated that the combination of SVM and PSO methods represents an effective approach to the analysis of public sentiment towards the Tapera policy on Twitter. The model achieved an accuracy of 85% and an AUC value of 0.84, indicating an excellent performance in classifying sentiment. This research contributes a sentiment analysis model that can be used to evaluate other public policies. However, further development is needed to improve the performance on the neutral sentiment category and expand the data coverage to reflect a wider range of public opinion.

REFERENCES

Anggraini, W. P., & Utami, M. S. (2021). Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Pekerja Di Indonesia. *Faktor Exacta*, *13*(4), 255. https://doi.org/10.30998/faktorexacta.v13i4.7964

Arsi, P., Wahyudi, R., & Waluyo, R. (2021). Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(2), 231–237. https://doi.org/10.29207/resti.v5i2.2698

Asril, A. A., Rifai, A., & Shebubakar, A. N. (2022). Penyelenggaraan Tabungan Perumahan Menurut Undang-Undang Nomor 4 Tahun 2016 Ditinjau Dari Perspektif Perlindungan Hukum. *Jurnal Magister Ilmu Hukum*, *7*(1), 1. https://doi.org/10.36722/jmih.v7i1.1185

Asro'i, A., & Februariyanti, H. (2022). Analisis Sentimen Pengguna Twitter Terhadap Perpanjangan Ppkm Menggunakan Metode K-Nearest Neighbor. *Jurnal Khatulistiwa Informatika*, *10*(1), 17–24. https://doi.org/10.31294/jki.v10i1.12624

Fatouros, G., Soldatos, J., Kouroumali, K., Makridis, G., & Kyriazis, D. (2023). Transforming sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications*, *14*(October), 100508. https://doi.org/10.1016/j.mlwa.2023.100508

Febriyani, E., & Februariyanti, H. (2023). Analisis sentimen terhadap program kampus merdeka menggunakan algoritma naive bayes classifier di twitter. *Jurnal Tekno Kompak*, *17*(1), 25–38. Retrieved from https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/2061

Hendrastuty, N., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N., Isnain, A. R., … Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, *6*(3), 150–155. Retrieved from http://situs.com

Indonesia, T. C. (n.d.). Besaran Iuran Tapera yang Potong Gaji Pekerja Setiap Tanggal 10. Retrieved June 6, 2024, from https://www.cnnindonesia.com/ekonomi/20240527141648-532-1102423/besaran-iuran-tapera-yang-potong-gaji-pekerja-setiap-tanggal-10

Kaafi, A. Al, Rachmi, H., Informasi, T., Bina, U., Informatika, S., Informasi, S., … Machine, S. V. (2012). Analisis Opini Masyarakat Terhadap Pemberlakuan ERP Di Jalan Ibu Kota Jakarta, (x), 1–11.

Kristiyanti, D. A., Kaafi, A. Al, Purwaningsih, E., Nurelasari, E., & Nisa, B. (2023). Deep learning for Twitter sentiment analysis about the pros and cons of Covid-19 vaccines in Indonesia. *AIP Conference Proceedings*, *2714*(February). https://doi.org/10.1063/5.0128686

Kurniawan, A., & Waluyo, S. (2022). Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter Application Of Naive Bayes Algorithm In Capital Movement Sentiment Analysis On Twitter. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*, (September), 455–461.

Marga, N. S. (2022). Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *2*(4), 453–463. https://doi.org/10.33365/jatika.v2i4.1602

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14227
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Md Suhaimin, M. S., Ahmad Hijazi, M. H., Moung, E. G., Nohuddin, P. N. E., Chua, S., & Coenen, F. (2023). Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions. *Journal of King Saud University - Computer and Information Sciences*, *35*(9), 101776. https://doi.org/10.1016/j.jksuci.2023.101776

Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R. (2021). Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *Sistemasi*, *10*(1), 222. https://doi.org/10.32520/stmsi.v10i1.1179

Purwati, S. E., & Pristyanto, Y. (2024). Model Random Forest and Support Vector Machine for Flood Classification in Indonesia, *8*(4), 2261–2268.

Putra, H. G., Fahmi, E., & Taruc, K. (2020). Tabungan Perumahan Rakyat (Tapera) Dan Penerapannya Di Dki Jakarta. *Jurnal Muara Sains, Teknologi, Kedokteran Dan Ilmu Kesehatan*, *3*(2), 321. https://doi.org/10.24912/jmstkik.v3i2.5630

Rachmi, H., Suparni, S., & Kaafi, A. Al. (2021). Analisis Sentimen Sistem Ganjil Genap Kota Bogor. *Jurnal ELTIKOM*, *5*(2), 92–99. https://doi.org/10.31961/eltikom.v5i2.429

Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jdmsi*, *2*(1), 31–37. Retrieved from https://t.co/NfhnfMjtXw

Sarirete, A. (2021). A Bibliometric Analysis of COVID-19 Vaccines and Sentiment Analysis. *Procedia Computer Science*, *194*, 280–287. https://doi.org/10.1016/j.procs.2021.10.083

Sasmito Aribowo, A. (2018). Analisis Sentimen Publik pada Program Kesehatan Masyarakat menggunakan Twitter Opinion Mining. *Seminar Nasional Informatika Medis (SNIMed)*, *0*(0), 17–23. Retrieved from https://journal.uii.ac.id/snimed/article/view/11877

Supriyanto, J., Alita, D., & Isnain, A. R. (2023). Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *4*(1), 74–80. https://doi.org/10.33365/jatika.v4i1.2468

Tania, N., Novienco, J., & Sanjaya, D. (2021). Kajian Teori Hukum Progresif Terhadap Implementasi Produk Tabungan Perumahan Rakyat. *Perspektif*, *26*(2), 73–87. https://doi.org/10.30742/perspektif.v26i2.800

Undang-Undang Republik Indonesia Nomor 4 Tahun 2016 Tentang Tabungan Perumahan Rakyat.

*name of corresponding author