International Journal of Advances in Data and Information Systems
Vol. 5, No. 2, October 2024, pp. 242~252
ISSN: 2721-3056, DOI: 10.59395/ijadis.v5i2.1342

Sentiment Analysis of Twitter Users Towards *Kartu Prakerja* Program Using the Naive Bayes Method

Harto Tomi Wijaya[1] , Kustiyono[2]
[1,2] Informatics Engineering, Universitas Ngudi Waluyo, Indonesia

Article Info

Article history:

Received Sep 14, 2024
Revised Oct 15, 2024
Accepted Oct 30, 2024

Keywords:

Sentiment analysis,
Twitter,
Kartu Prakerja,
Naive Bayes

ABSTRACT

This study conducts a sentiment analysis of Twitter users regarding the Indonesian government's *Kartu Prakerja* program, utilizing the Naive Bayes method for classification. Launched in 2020 to enhance employability skills amidst the COVID-19 pandemic, the program has garnered various public responses. A total of 836 tweets containing the keyword "*Kartu Prakerja*" were collected using the Twitter API and analyzed to determine sentiment distribution. Results indicate a predominance of neutral sentiment (800 tweets), with only 17 positive and 22 negative tweets. The Naive Bayes method achieved an accuracy of 95%, demonstrating its effectiveness in sentiment classification. However, comparisons with other methods, such as Support Vector Machine (SVM) and Recurrent Neural Network (RNN), reveal that these techniques yield higher accuracy rates (98.34% and 96%, respectively). This research highlights the importance of sentiment analysis in understanding public perceptions and informs policymakers about areas needing improvement. The findings underscore the potential of integrating advanced machine learning techniques to enhance sentiment analysis and provide insights into the effectiveness of government programs like *Kartu Prakerja*.

Corresponding Author:

Harto Tomi Wijaya,
Informatics Engineering, Ngudi Waluyo University,
Jl. Diponegoro no 126 Gedanganak, Central Java, Indonesia
Email: hartotomi99735@gmail.com

1. INTRODUCTION

Kartu Prakerja program is an initiative of the Indonesian government that aims to improve employability skills through online training. The program was launched in 2020 to respond to the economic impact of the COVID-19 pandemic.[1] Which led to rising unemployment and the need for reskilling and upskilling the workforce. *Kartu Prakerja* program offers a wide range of online training covering technical and non-technical skills that are expected to help participants find a new job or start their own business. The program has received various responses from the public. These were expressed on the social media platform Twitter. Sentiment analysis of these comments is important to understand the perceptions of Indonesians and to identify areas for improvement by the government.

Social media is a digital platform that allows users to get various information, one of which is on Twitter. has become the largest social media platform in the world.[2] In the implementation of the *Kartu Prakerja* program, there are still many pros and cons, one of which is on the Twitter social media platform. Sentiment analysis of these comments is important to understand public perceptions of the program, identify problems that may arise, and evaluate the overall effectiveness of the program. Sentiment analysis is the process of identifying and categorizing opinions expressed in a

text, especially to determine whether the author's attitude towards a particular topic is positive, negative, or neutral [3].

The Naive Bayes method is one of the machine learning techniques often used for sentiment analysis due to its simplicity and effectiveness.[4] Naive Bayes works on the assumption that features (words) in text are independent of each other. This method has been proven to perform well in many text analysis applications [5]. However, to understand the extent to which Naive Bayes is effective, the author conducted a comparison with other methods in sentiment analysis, it is necessary to compare the results with several relevant journals. Some studies that use Naive Bayes for sentiment analysis show varying results. For example, research by Nisrina and Kustiyono (2004) used the C4.5 algorithm in analyzing customer satisfaction, which shows that other machine learning methods are also effective in analyzing data and providing valuable insights for service improvement [6]. In addition, research by Hidayati and Kustiyono developed a decision support system for the recruitment of the best employees using the Weighted Product method, which shows the importance of utilizing data and algorithms in making objective and effective decisions [7].

Some studies that use Naive Bayes for sentiment analysis show varying results. For example, in the classification study of public sentiment towards the *Kartu Prakerja* policy in Indonesia with the naive bayes classification method achieved an accuracy of 91.06%. [8]. Analysis of *Twitter* User Sentiment towards the *Kartu Prakerja* Program in the Midst of the Covid-19 Pandemic Using the *Naïve Bayes Classifier* Method this method obtained an accuracy of 81.6%. besides that, Analysis of Public Sentiment towards the *Kartu Prakerja* Program on Twitter with the Support Vector Machine Method showed an accuracy of 98.34% [9]. in sentiment analysis on Twitter against the *Kartu Prakerja* program using the Recurrent Neural Network method with an accuracy of 96%. [10]. Analysis of public sentiment towards the work copyright law with the Naive Bayes algorithm method on twitter social media Achieved 97% accuracy [6][11]. Sentiment Analysis on Twitter Social Media to Assess Public Response to *Kartu Prakerja* Selection with the Cross Industry Standard Process Model for Data Mining (CRISP-DM) method getting 95.67% accuracy [6][12]. Implementation of Data Mining Sentiment Analysis of the *Kartu Prakerja* Program Using the Naïve Bayes Algorithm gets an accuracy of 77.58%. [13]. Sentiment Analysis of the *Kartu Prakerja* Using Text Mining with Support Vector Machine and Radial Basis Function Kernel obtained an accuracy of 85.20% [14]. Sentiment analysis on *Kartu Prakerja* policies using the naive bayes method gets 93% accuracy [9][15]. The application of sentiment analysis on twitter users using convolution onal neural network and naive bayes methods in the case study of *Kartu Prakerja*s obtained an accuracy of 75.3% [14].

This research provides an overview of how the Naive Bayes method compares with other methods in different contexts. However, these 10 articles apply sentiment analysis with varying degrees of accuracy. Naive Bayes performs well, but methods such as SVM, RNN and CNN. Naive Bayes performs well, but methods such as SVM, RNN and CNN perform better in some cases.

The main questions to be answered include:

1. What is the sentiment distribution of comments on the *Kartu Prakerja* program on twitter?
2. How effective is the Naive Bayes method in classifying comment sentiment?
3. how does the accuracy of the Naive Bayes method perform with other methods used in similar studies?

By answering this question, this research is expected to provide valuable insights into public perspectives on the *Kartu Prakerja* program and identify the most effective sentiment analysis method to use in this context. The results of this study can also be used by future research.

2. RESEARCH METHOD

This research method consists of four main stages, namely data collection, preprocessing, data labeling, naive bayes model, the classification process using the Naive Bayes method can be seen in Figure 1.

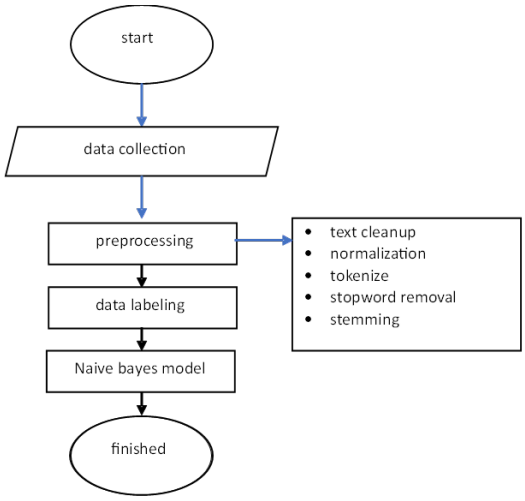

Figure 1. Naive bayes classification process

1. Data collection

Data was collected from twitter using API with tweepy library in google collab. Data is taken from tweets containing the keyword "*Kartu Prakerja*".

- The data platform was collected from Twitter using the Twitter API.
- Keyword tweets containing the keyword *Kartu Prakerja* were selected for analysis.
- The library used tweepy is used to access the twitter API and download tweets directly on google collab.
- A total of 836 tweets were collected. This data is then stored in csv format to facilitate the further analysis process. results can be seen in table 1

Table 1. Data Collection Results

No.	full_text
1.	Hi, Ubahpedia friends! ï¸ Who just graduated still really feels the need for upskilling and is confused because usually the price is quite draining the wallet. Eits..don't worry today Manies will give you info about job preparation tips regarding the *Kartu Prakerja* Program. https://t.co/TBfQfT0yar
2.	Studium Generale Speaker: Denni Puspa Purbasari (Executive Director of Implementation Management of *Kartu Prakerja* Program) Topic: The Art of Working in Government Date: Wednesday, February 28, 2024 Time : Pkl. 09.00 10.45 WIB Media : Daring Zoom or Youtube itbofficial Live https://t.co/vQyjW4fwe6
3.	Hi Idelisteners! Three years on, the *Kartu Prakerja* program has contributed to reducing the impact of unemployment caused by the COVID-19 pandemic. But how big is the program's influence? Check out the review in this week's #WeeklyDigest! https://t.co/mDnOeI9c5N
4.	President Jokowi said the *Kartu Prakerja* program has provided tangible benefits to the community. https://t.co/Iscwl0oG3e

2. Prepocessing

Data preprocessing is important in sentiment analysis. At this stage, the raw data that has been collected is cleaned and prepared to be ready for use in model building.

Steps taken include:

- Text cleaning

  This step removes URLs, mentions, hashtags, numbers, and special characters to maintain focus on the core content of the text. It helps reduce noise and ensures that

only relevant text data is processed by the model. By cleaning the text, the model can better concentrate on important information and avoid bias from irrelevant elements commonly found on social media, such as URLs or special characters.[16]
Removing URLs, mentions, hastags, numbers and special characters to keep the focus on the main context. The results of text cleaning can be seen in table 2

Table 2. Text Cleaning Results

No.	Full text	Usuername
1.	thank God I passed the *Kartu Prakerja*	yaniarsim
2.	not Nadiem's program, but the *Kartu Prakerja* needs to be evaluated for its impact.	millhopes
3.	I hope those who participate in the *Kartu Prakerja* pass the selection, so that you can quickly get a job too, you guys enter me, aamiin.	aricandra
4.	I just found out that now the *Kartu Prakerja* only gives a fee once 600k in the past 4 times.	workafess

- Normalization
  Normalization aligns the data to a consistent format by converting uppercase letters to lowercase, removing punctuation, and stripping extra spaces. This helps the model recognize word patterns more consistently, which can lead to improved accuracy by reducing unnecessary variations in the data.
  Normalization is to align or tweet data to make it more consistent and uniform such as changing letters, removing punctuation marks, stripping excess spaces before analysis.[17] can be seen sourcode in table 3.

Table 3. normalization sourcode

```
#Normalization
norm = {"yang": "yang", "kipkuliahdankartuprakerjaprogramyangdinantikan": "kip kuliah dan kartu prakerja program yang dinantikan.", "kartuprakerja": "kartu prakerja", "gak": "tidak", "akuuu": "saya", "Aamin": "amen", "skrg": "now", "prnh": "ever", "untukk": "untuk", "programkartu": "program card", "kartuuntukprakerjaataukorbanphkdari01perludi":"card for pre-employment or phk victims from 01 is needed", "mardanikritikprogram kartu prakerjajokowi": "mardani kritik program kartu prakerja jokowi", "gimanayacaratulisalamatdiku prakerja?failed": "how to write the address on the Kartu Prakerja failed", "but": "but", "not a diemtapikartu program": "not a diem but a card program", "alhamdulillahakuloloskartu": "alhamdulillah I passed the card", "promisein2periode2014print10millionworkers":"promise in 2 periods 2014 to print 10 million jobs", "semogalebihlebihbaik": "hopefully better", "lebihbaik": "better", "how do I write": "how do I write"}

def normalize(str_text):
  for i in norm:
    str_text = str_text.replace(i, norm[i])
  return str_text

df['full_text'] = df['full_text'].apply(lambda x: normalize(x))
```

- Tokenization
  Tokenization breaks the text into individual words, enabling the model to process each word separately, which is essential for text analysis. With tokenization, the model can examine each word individually, thereby improving its performance in classifying sentiment based on the specific words used in the text [2].
  Tokenize: break the text into separate words to facilitate analysis. results can be seen in table 4.

Table 4. Tokenize Result

No	Full text
1.	thank God I passed the *Kartu Prakerja*
2.	not Nadiem's program, but the *Kartu Prakerja* needs to be evaluated for its impact.
3.	I hope those who participate in the *Kartu Prakerja* pass the selection, so that you can quickly get a job too, you guys enter me, aamiin.
4.	I just found out that now the *Kartu Prakerja* only gives a fee once 600k in the past 4 times.

- Stop Word Removal
  Stop words such as “and,” “or,” and “not” are removed as they do not contribute significantly to sentiment classification. Stop word removal reduces data complexity by eliminating frequently occurring words that do not carry particular sentiment information. This directly impacts model performance by enhancing its focus on more meaningful words.[18]
  Stop word removal removes common insignificant words (such as "no") using literary can be seen sourcecode in table 5.

Table 5. sourcode stop word

```
literary import
from Sastrawi.StopWordRemover.StopWordRemoverFactory
import StopWordRemoverFactory, StopWordRemover,
ArrayDictionary
more_stop_words = ["no"]

stop_words = StopWordRemoverFactory().get_stop_words()
stop_words.extend(more_stop_words)

new_array = ArrayDictionary(stop_words)
stop_words_remover_new = StopWordRemover(new_array)

def stopword(str_text):
  str_text = stop_words_remover_new.remove(str_text)
  return str_text

  df['full_text'] = df['full_text'].apply(lambda x: stopword(x))
  df
```

- Stemming
  Stemming converts words to their root form (e.g., “running” becomes “run”), reducing word variability. This helps the model recognize words with the same meaning but different forms, allowing it to focus on consistent root words. Stemming contributes to improved classification accuracy by concentrating on the core meaning of the words.[19]
  Returning words to their base form, such as "yang" to "yang" to reduce word variability. Can be seen in table 6.

Table 6. stemming result

full_text
jokowi's promise in 2 periods 2014 print 10 million job opportunities 2019 print *Kartu Prakerja*s 2024 print family to inherit the throne
thank God I passed the *Kartu Prakerja*
not Nadiem's program, but the *Kartu Prakerja* needs to be evaluated for its impact.
How do I write the address on the *Kartu Prakerja* and it keeps failing?

Implementing these preprocessing steps enhances the quality of the data input into the Naive Bayes model. Preprocessing helps eliminate noise, reduce complexity, and standardize data, all of which contribute to increased model accuracy. In this study, the effective implementation of preprocessing steps is reflected in the high accuracy achieved by the model.

3. Data labeling

Data labeling is the process of data that has been processed and then automatically labeled by Google Collab into three categories of positive, negative, neutral.

- Positive tweets containing the words good, good, satisfied
- Negative tweets containing the words buruj, ugly, disappointed.
- Neutral tweets that contain words do not contain clear sentiments. can be seen in table 7

Table 7. Data Labeling Results

Sentiment	Number of tweets
Positive	17
negative	22
neutral	800

4. The naive bayes model

The naive bayes model is used for sentiment analysis classification of twwets that have been labeled in google collab.

- The naive bayes classification was tested in order to get the analyzed tweets. can be seen in table 8.

Table 8. Naïve bayes classification result

NO	tweet_indo	classification bayes
1.	thank God I passed the *Kartu Prakerja*	Neutral
2.	not Nadiem's program, but the *Kartu Prakerja* needs to be evaluated for its impact.	Neutral
3.	How do I write the address on the *Kartu Prakerja* and it keeps failing?	Neutral

- Wordcloud

  wordcloud Bringing up words that often appear in comments on twitter can be seen in Figure 2.

Figure 2. a word that often comes out

- The label here shows the results of labeling positive, negative and neutral tweets can be seen in Figure 3.

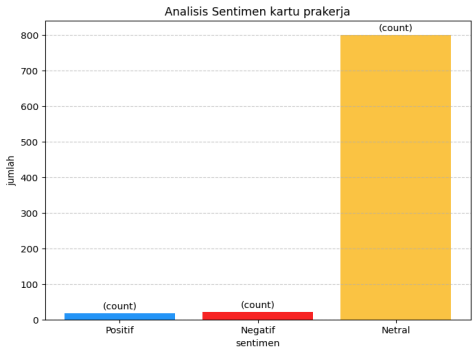

Figure 3. Statistic of the dataset

5. Model Validation

To ensure that the Naive Bayes model is not overfitting, we applied cross-validation. Cross-validation is a validation technique that splits the dataset into several subsets (folds). In each iteration, one fold is used as the test data while the remaining folds serve as the training data. This process is repeated until each fold has been used as the test set [20].

By using this method, we can assess the robustness of the model and enhance the accuracy and reliability of the sentiment analysis results. Given the relatively small dataset (836 tweets), implementing cross-validation is crucial to ensure that the model generalizes well on unseen data rather than just fitting the training data.

The application of cross-validation provided an accuracy distribution that shows the Naive Bayes model performs consistently across folds. The average accuracy from cross-validation also supports that Naive Bayes is a suitable method for sentiment classification on this dataset, although other methods such as SVM and RNN demonstrate higher accuracy levels.

3. RESULTS AND DISCUSSION

This research shows that the naive bayes model can be quite effectively used for sentiment analysis on Twitter, especially in the context of the *Kartu Prakerja* program.

3.1 Comment Sentiment Result

- The analysis shows the results of positive, negative, and neutral sentiments towards the *Kartu Prakerja* program with 836 tweets.
- The results of the analysis show that the majority of sentiment towards this *Kartu Prakerja* program tends to be neutral, with the number of neutral tweets 800, positive tweets 17, and negative tweets 22 there are many people still confused about how to register for *Kartu Prakerja*.

3.2 Classification of methods

- The performance of the Naive Bayes method is carried out to measure the performance of the model in classifying sentiment. The naive bayes method gets 95% comment prediction percentage accuracy. considered good enough for results in the google collab application.
- The graph shows the data of the research results made in can be seen in Figure 5.

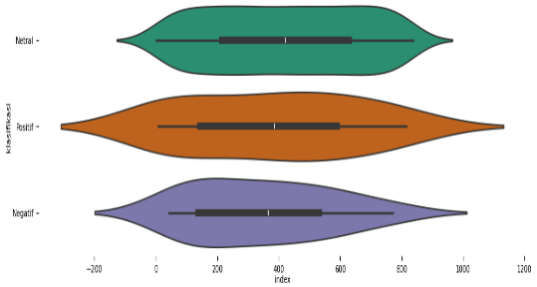

Figure 4. Sentiment analysis chart

3.3 Comparison with other methods

- SVM
- CNN
- RNN
- DATA MINING (CRISP-DM)

Comparison with Other Methods In this study, we conducted a comparative analysis using the same dataset for several sentiment analysis methods, including Naive Bayes, Support Vector Machine (SVM), Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN). Each method was applied to the dataset of [new number] tweets related to the *Kartu Prakerja* program, and the results were evaluated based on accuracy, precision, recall, and F1-score.

1. Naive Bayes: The Naive Bayes method achieved an accuracy of 95%, with balanced precision and recall values. While this method is computationally efficient and easy to implement, it assumes that features (words) are independent, which may limit its ability to capture complex relationships in the data.
2. Support Vector Machine (SVM): The SVM method outperformed Naive Bayes with an accuracy of 98.34%. SVM excels in text classification tasks, particularly when the data is not linearly separable. However, it requires more computational resources compared to Naive Bayes.
3. Recurrent Neural Network (RNN): RNN, known for its ability to handle sequential data, achieved an accuracy of 96%. This method is better suited for capturing temporal dependencies in the data, which is especially useful when sentiment is expressed over multiple words or phrases. However, it also requires more training time and computational power.
4. Convolutional Neural Network (CNN): CNN, typically used for image recognition, can also be applied to text classification by capturing n-gram features. It achieved an accuracy of 75.3% in this study, which is lower compared to Naive Bayes and SVM. This lower performance may be due to CNN's inability to fully capture the sequential nature of text data, unlike RNN.
5. Data Mining (CRISP-DM): The Cross Industry Standard Process for Data Mining (CRISP-DM) method was also tested and achieved an accuracy of 95.67%. While it performs well, it shares similar computational complexity with other machine learning methods.

The results of the comparative analysis are summarized in Table 9:

Table 9. shows the accuracy comparison

No	Methods	Accuracy
1.	Naïve Bayes	95%
2.	SVM	98, 34%.
3.	CNN	75,3%
4.	RNN	95,66%
5.	Data Mining	95.67%

The results of this research can serve as a basis for policy-making or program improvement based on the opinions and sentiments revealed through this analysis.

3.4 Sentiment Distribution

The analysis results indicate a strong dominance of neutral sentiment, with 800 out of 836 tweets classified as neutral, while only a few were classified as positive (17) or negative (22). This

distribution requires further exploration to understand the reasons behind it, which may be attributed either to dataset limitations or to constraints in the Naive Bayes method used.

1. Dataset Limitations

   The high prevalence of neutral sentiment may suggest that the dataset lacks diversity or is not fully representative of public opinion. If most of the tweets analyzed are factual or lack strong emotional content, they may naturally be classified as neutral. A larger and more varied dataset might provide a more balanced sentiment distribution, offering a better understanding of public opinion on the *Kartu Prakerja* Program.

2. Methodological Constraints

   The Naive Bayes method works on the assumption that words in the text are independent, which may fail to capture the context or subtleties in the text. As a result, tweets that contain mixed or subtle emotions are often classified as neutral, which could lead to an overrepresentation of neutral classifications when analyzing complex or context-dependent data such as social media posts.

Considering both the dataset characteristics and the limitations of the model provides insight into why neutral sentiment dominates the analysis. Future research could consider using alternative methods, such as Support Vector Machine (SVM) or Recurrent Neural Network (RNN), which may provide a more nuanced sentiment breakdown and better capture complex emotional cues.

4. DISCUSSION

This research was conducted to analyze the sentiment of comments on Twitter about the *Kartu Prakerja* Program using the Naive Bayes method. The entire research process was implemented in Google Colab, which provides convenience and efficiency in data processing and model training. The following is a detailed discussion of the research results:

1. Use of Google Colab in Research

Google Colab provides a platform that supports this research with sufficient computing resources, including GPUs and cloud-based storage. This allows researchers to access and process data at scale, and run machine learning algorithms without being limited by local capacity. In addition, collaboration features and integration with Google Drive make it easy to manage files, code and research results.

2. Effectiveness of Naive Bayes Method

The Naive Bayes method was chosen due to its simplicity and ability to handle text data with good performance. In this study, the Naive Bayes model was able to classify 800 out of 836 tweets correctly, resulting in an accuracy of about 95.69%. This accuracy rate shows that the model is quite reliable in identifying sentiment in the data used.

3. Optimal Data Pre-processing

One of the key success factors of this model is the comprehensive pre-processing of the data. The tweet text is cleaned from noise such as punctuation, URLs, and special characters. This step is followed by tokenization, removal of stop words, and stemming, which helps in reducing data complexity and improving model accuracy. The implementation of these steps in Google Colab also proved to be efficient, allowing the processing of thousands of tweets in a relatively short time.

4. Model Performance Evaluation

The evaluation of the model shows that Naive Bayes provides satisfactory results for sentiment classification with 95% accuracy. This result confirms that the model is able to capture the sentiment patterns present in the data well, although there are some tweets that are misclassified. These errors may be due to ambiguities in the text or context that are difficult to capture by the model.

5. Visualization and Interpretation of Results

Visualizations of the results, such as the confusion matrix and word cloud, are helpful in understanding how the model works and where it goes wrong. These visualizations show the distribution of prediction errors and provide insight into the dominant key words in each sentiment category. Using Google Colab, these visualizations can be generated quickly, providing immediate feedback on model performance.

6. Limitations and Development Potential

Although this research has achieved satisfactory results, there are some limitations:

- Simple Model: Naive Bayes is a simple model and may not be able to capture complex relationships in the data. The use of more sophisticated models, such as Random Forest or SVM, can improve accuracy.
- Text Representation: The use of CountVectorizer results in a simple text representation. deep learning based can be used to improve the performance of the model.
- Data Size: The number of tweets analyzed was limited to 836 tweets. Further research with a larger amount of data may provide more representative results.

7. Conclusion

This research shows that the Naive Bayes method is capable of classifying Twitter users' sentiment analysis regarding the *Kartu Prakerja* program with an accuracy of 95%. However, methods such as SVM and RNN demonstrated superior performance, achieving higher accuracy rates of 98.34% and 96%, respectively. CNN, while effective in other domains, did not perform as well in this study. These results suggest that while Naive Bayes is an efficient baseline model, more advanced methods like SVM and RNN are better suited for capturing the complexity of public sentiment in text data. Future research could benefit from hybrid models or ensemble approaches to further improve accuracy.

5. CONCLUSION

This research shows that the Naive Bayes method is able to classify Twitter users' sentiment analysis of the *Kartu Prakerja* program with fairly good accuracy. However, other methods such as SVM, RNN, AND DATA MINING (CRISP-DM) show higher accuracy. The results of this study can be used by the government to improve the quality of the *Kartu Prakerja* program and better meet the needs of the community. Further research is needed to optimize the sentiment analysis methods used.

ACKNOWLEDGEMENTS

The author would like to thank all those who have provided support, and guidance in the completion of this research article.

REFERENCES

[1] P. Nguyen, F. Putra, M. Considine, and A. Sanusi, "Activation through welfare conditionality and marketisation in active labour market policies: Evidence from Indonesia," *Aust. J. Public Adm.*, vol. 82, no. 4, 2023, doi: 10.1111/1467-8500.12602.

[2] A. Janowski, "Natural Language Processing Techniques for Clinical Text Analysis in Healthcare," *J. Adv. Anal. Healthc. Manag.*, vol. 7, no. 1, 2023.

[3] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artif. Intell. Rev.*, vol. 52, no. 3, 2019, doi: 10.1007/s10462-017-9599-6.

[4] M. Ahmad, S. Aftab, S. S. Muhammad, and S. Ahmad, "Machine Learning Techniques for Sentiment Analysis: A Review," *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 3, 2017.

[5] A. Kelly and M. A. Johnson, "Investigating the statistical assumptions of naïve bayes classifiers," in *2021 55th Annual Conference on Information Sciences and Systems, CISS 2021*, 2021. doi: 10.1109/CISS50987.2021.9400215.

[6] D. Nisrina and K. Kustiyono, "Analisis Kepuasan Konsumen Menggunakan Metode Algoritma C4. 5 Berbasis Rapidminer Pada PT. Adeaksa Indo Jayatama," *MEANS (Media Inf. Anal. dan Sist.*, vol. 9, no. 1, pp. 26–33, 2004.

[7] D. R. Hidayati and Kustiyono, "Pengembangan Sistem Pendukung Keputusan Rekrutmen Karyawan Terbaik Dengan Metode Weighted Product Di PT Morich Indo Fashion," *MEANS (Media Inf. Anal. dan Sist.*, vol. 9, no. 1, pp. 15–19, 2024.

[8] W. P. Anggraini and M. S. Utami, "KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN KARTU PEKERJA DI INDONESIA," *Fakt. Exacta*, vol. 13, no. 4, 2021, doi: 10.30998/faktorexacta.v13i4.7964.

[9] S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, 2021, doi: 10.30591/jpit.v6i3.2870.

[10] R. Sanusi, F. D. Astuti, and I. Y. Buryadi, "ANALISIS SENTIMEN PADA TWITTER TERHADAP PROGRAM KARTU PRA KERJA DENGAN RECURRENT NEURAL NETWORK," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, 2021, doi: 10.26798/jiko.v5i2.645.

[11] N. Sucahyo, I. Kurniati, and K. Harvit, "ANALISIS SENTIMEN MASYARAKAT TERHADAP UU CIPTA KERJA PADA MEDIA SOSIAL TWITTER," *JRIS J. REKAYASA Inf. SWADHARMA*, vol. 2, no. 1, 2022, doi: 10.56486/jris.vol2no1.167.

[12] P. A. Nugroho, N. Sucahyo, and I. Kurniati, "Sentimen Analisis pada Sosial Media Twitter untuk Menilai Respon Masyarakat terhadap Seleksi Kartu Prakerja," *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, 2023, doi: 10.37012/jtik.v9i1.862.

[13] A. Yuliawati and T. Wahyudi, "Implementasi Data Mining Analisa Sentimen Program Kartu Prakerja Menggunakan Algoritma Naïve Bayes," *TEKNIKA*, vol. 18, no. 2, pp. 611–622, 2024.

[14] P. W. Hardjita, "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode Convolutional Neural Network Dan Naive Bayes (Studi Kasus: Kartu Prakerja)," Uin Sunan Kalijaga Yogyakarta, 2021.

[15] A. Wanti, F. Hariri, and J. Wahyu, "ANALISIS SENTIMEN MENGENAI KEBIJAKAN KARTU PRAKERJA MENGGUNAKAN METODE NAIVE BAYES," *Aleph*, vol. 87, no. 1,2, 2023.

[16] S. Sonare and M. Kamble, "Sentiment Analysis in polished Product Based Inspections data using existing supervised machine learning approach," in *2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society, TRIBES 2021*, 2021. doi: 10.1109/TRIBES52498.2021.9751663.

[17] B. E. R. R. I. M. Israr, "Sentiment Analysis in Arabic Tweets," University Of Kasdi Merbah Ouargla, 2020.

[18] A. Oussous, F. Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "ASA: A framework for Arabic sentiment analysis," *J. Inf. Sci.*, vol. 46, no. 4, 2020, doi: 10.1177/0165551519849516.

[19] R. Pramana, Debora, J. J. Subroto, A. A. S. Gunawan, and Anderies, "Systematic Literature Review of Stemming and Lemmatization Performance for Sentence Similarity," in *Proceedings of the 2022 IEEE 7th International Conference on Information Technology and Digital Applications, ICITDA 2022*, 2022. doi: 10.1109/ICITDA55840.2022.9971451.

[20] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, 2018. doi: 10.1016/B978-0-12-809633-8.20349-X.