Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 2, September 2023, pp. 1006-1014
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

Sentiment Analysis of Indonesian Presidential Candidate 2024 on Facebook

Fardhan Saifullah Fattah[1], Chanifah Indah Ratnasari[2*]
[1,2]Informatics, Universitas Islam Indonesia, Indonesia
e-mail: fardhan.fattah@students.uii.ac.id[1], chanifah.indah@uii.ac.id[2]
(*) Corresponding Author

Abstract

The 2024 Indonesian Presidential Election has become a hot topic in daily life, including on Facebook. Everyone has the freedom to express or speak their opinions on the Indonesian presidential candidate for 2024. As a result, there are a variety of viewpoints, some of which are positive and some of which are negative. This research aims to classify positive and negative sentiments in each presidential candidate's Facebook post comment data. Several processes are involved in this sentiment analysis process, including data scraping, data preprocessing, data labeling, and data classification using the Support Vector Machine (SVM) approach. The data used are comments obtained from each presidential candidate's Facebook page between July 1 and July 31, 2023. Anies Baswedan, with a positive sentiment value of 88.20% and a negative sentiment value of 11.80%, is the presidential candidate with the highest positive sentiment and the lowest negative sentiment. The SVM approach with a linear kernel produced the best precision value for positive sentiment, namely 94%.

Keywords*: Sentiment Analysis, Support Vector Machine, 2024 Presidential Candidates, Facebook, SVM.*

Abstrak

Pemilihan Presiden Indonesia 2024 telah menjadi topik yang hangat diperbincangkan dalam kehidupan sehari-hari, salah satunya pada platform media sosial Facebook. Semua orang memiliki kebebasan untuk mengemukakan pendapat atau menyuarakan opini mereka mengenai calon Presiden Indonesia 2024. Hal ini menghasilkan beragam pendapat, pendapat tersebut ada yang mengandung sentimen positif ataupun negatif. Penelitian ini bertujuan untuk mengklasifikasi sentimen positif dan negatif pada data komentar postingan Facebook masing-masing calon presiden. Dalam proses analisis sentimen, terdapat beberapa langkah yang terdiri dari scraping data, preprocessing data, pelabelan data, dan pengklasifikasian data dengan menggunakan metode Support Vector Machine (SVM). Data yang digunakan adalah komentar yang diambil dari halaman Facebook tiap bakal calon presiden dari tanggal 1 sampai dengan 31 Juli 2023. Calon presiden yang mempunyai sentimen positif tertinggi dan sentimen negatif terendah adalah Anies Baswedan, dengan nilai sentimen positif sebesar 88,20% dan sentimen negatif sebesar 11,80%. Nilai presisi tertinggi pada sentiment positif didapatkan ketika menggunakan metode SVM dengan kernel Linear, yaitu 94%.

Kata kunci*: Analisis Sentimen, Support Vector Machine, Calon Presiden 2024, Facebook. SVM.*

1. INTRODUCTION

The main characteristic of a democratic country is a high level of public participation in general elections. This is due to the fact that public political participation is a critical component of democratic governance [1]. Even though the 2024 presidential election is still several months away, societal tensions regarding which candidate is worthy to lead Indonesia for the next five years are already becoming apparent [2]. Various parties have promoted several big names to run as presidential candidates in 2024, such as Prabowo Subianto, Ganjar

Pranowo, and Anies Baswedan. As time passes, information technology development has experienced rapid growth—technological progress results from applying and harmonizing existing knowledge with other knowledge. Many activities generally utilize information technology as part of the process [3]. Including politicians and political parties, all of whom also take advantage of social networks. Politicians use this platform for various purposes, including campaigning, public image development, and communicating ideas and opinions via specific pages [4]. One of the social media platforms often used by politicians is Facebook. Usually, Facebook provides information about users, shares general information, and conveys news. As a social media platform with many users, Facebook allows its users to express their opinions on various topics [5]. As a result, an accurate analytical approach is needed to manage text data to present information about Facebook users' views on political figures. This text-based collection of comments is collected and can be analyzed using the sentiment analysis method [6].

2. RESEARCH METHODS

The following is the sequence of research stages, as shown in Figure 1.

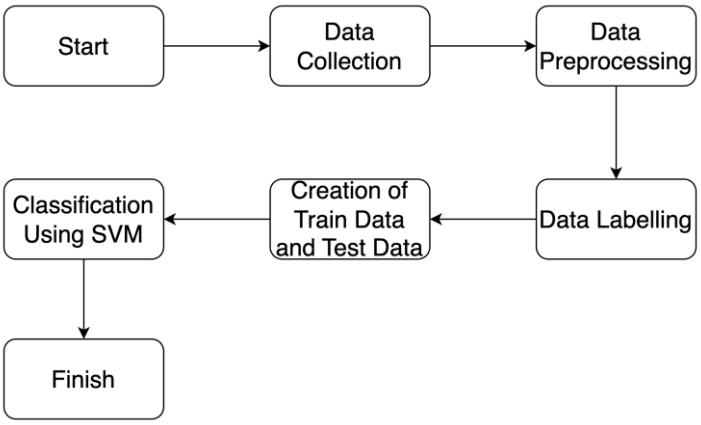

Figure 1. Research Stage

2.1. Data Collection

The first phase of this research was to collect comment data from the Facebook pages of each presidential candidate as an initial basis for analyzing public sentiment and views toward them.

2.2. Preprocessing

Following data collection, data preprocessing is the next phase. Data preprocessing is a method for cleaning and preparing raw data prior to its use in modeling. Data preprocessing steps include:

a. Case Folding

which converts existing uppercase letter data to lowercase letters [7].

b. Tokenizing

The process of breaking down or dividing text into smaller units or single chunks [8].

c. Normalization

The process of correcting misspelled terms in a text using a dictionary that has been compiled.

d. Stopword Removal

The procedure for removing unnecessary words.

2.3. Data labeling

Data labeling is labeling it as "positive" or "negative". The aim is to provide identification or labels that explain or describe the characteristics or attributes of the data.

2.4. Creation of Train Data and Test Data

After labeling is complete, the existing data will be divided into two parts: training and test data. Typically, data is separated in particular ratios. For example, 80% of the data is used for training, and 20% for testing, but these ratios can vary depending on the specific situation and needs in a particular use case.

2.5. Classification using Support Vector Machine (SVM)

Support Vector Machine (SVM) is a classification algorithm in machine learning that works by finding the best hyperplane to separate data into different classes. SVM can also be called Support Vector Networks [9].This hyperplane is chosen to have a maximum margin, namely the largest distance between the hyperplane and the closest data points representing different classes (Figure 2) [10].

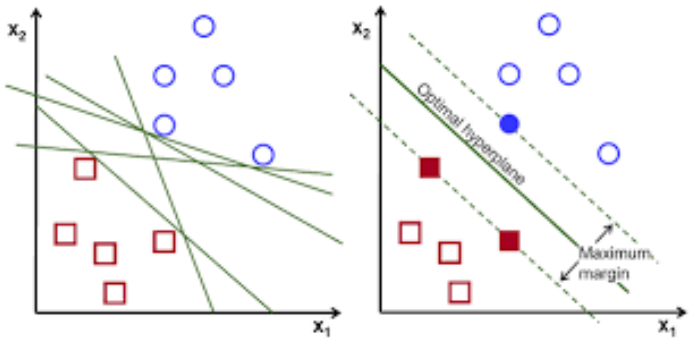

Figure 2. Support Vector Machine illustration

SVM classification is carried out using hyperplanes to separate classes. Each hyperplane has its characteristics, including direction (w), exact position in space or threshold value (b), and (xi) is an N-dimensional input vector or text indicating its class. As an illustration, an example of a training dataset that has been labeled shown in equation (1).

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 2, September 2023, pp. 1006-1014
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

$$(x_1, y_1),(x_2,y_2),\ldots\ldots,(x_k,y_k), \quad (1)$$

$X \in R^d$ where refers to the dimension of the vector; $y_i \in \{-1, +1\}$; i = 1, 2, … , k
By considering the decision function, which has the form f (x, w, b) = sgn ((w. x_i) + b), $w \in R^d$, $b \in R$.
The space between the hyperplane from the two classes is called the margin. The margin width is $\frac{1}{2}||w||$ and the core concept in the SVM algorithm is to achieve the maximum margin value by minimizing, which includes:

$$f(w, b) = \frac{1}{2}||w||^2 \quad (2)$$

This depends on:

$$wx_i + b >> 1, \text{ if } y_i = 1 \quad (3)$$
$$wx_i + b << 1, \text{ if } y_i = -1$$

The value is a constant or fixed value determined by the user, and ε is the margin error [11].

3. RESULTS AND DISCUSSION

3.1. Data Collection

In this study, data was collected using the scraping technique, specifically the Apify tool. The data retrieval process is simply copying the link to be retrieved to Apify, and then data retrieval will be carried out automatically. Figure 3 depicts instructions for extracting comments from the Facebook profiles of the 2024 Indonesian presidential candidates.

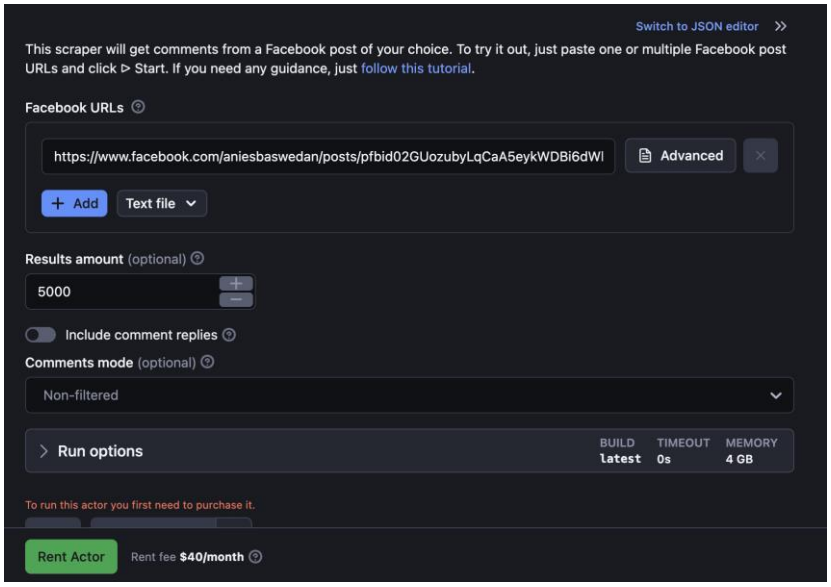

Figure 3. Data retrieval in Apify

The study's data was collected from the Facebook accounts of Anies Baswedan, Ganjar Pranowo, and Prabowo Subianto between July 1 and 31, 2023. Table 1 is a summary of the data collected from each candidate's Facebook page

for the 2024 Indonesian presidential election. The Ganjar Pranowo page contains 82 posts with over 21,000 comments. Anies Baswedan has only 32 posts with over 15,300 comments, while Prabowo Subianto has 42 posts with over 52,500 comments.

Table 1. Number of Posts and Comments

Name	Posts	Comments
Ganjar Pranowo	82	21,206
Anies Baswedan	32	15,333
Prabowo Subianto	42	52,528

3.2. Preprocessing

a. Case Folding

Case Folding is converts all comments on each presidential candidate's post from uppercase to lowercase. Table 2 displays examples of the outcomes of the process of case folding.

Table 2. Case Folding Process

Original Comments	Comennts after Case Folding
BAGUSSS PAK Ganjar Pranowo	bagusss pak ganjar pranowo
Semoga menjadi Pemimpin yang menyatukan	semoga menjadi pemimpin yang menyatukan
Prabowo selalu di hati	prabowo selalu di hati

b. Tokenizing

Tokenizing namely breaking down the sentences in the comment data into word-by-word form. This tokenizing step simplifies the transformation stage, allowing processing based on individual words rather than complete sentences, an example of the results of which is shown in Table 3.

Table 3. Tokenizing Process

Comments Before	Comments after Tokenizing
bagusss pak ganjar pranowo	"bagusss", "pak", "ganjar", "pranowo"
semoga menjadi pemimpin yang menyatukan	"semoga", "menjadi", "pemimpin", "yang", "menyatukan"
prabowo selalu di hati	"prabowo", "selalu", "di", "hati"

c. Normalization

The normalization process was done using existing dictionaries and updated with manual additions [12]. Below in Table 4 is an example of a list of words that will be normalized based on the corpus that has been created.

Table 4. Normalization Process

Sebelum	Sesudah
yg	yang
ttp	tetap
bpk	bapak
msyrkt	masyarakat
diem	diam

d. Stopword Removal

Conjunctions such as "yang," "dan," "di," "dari" were removed, as seen in Table 5.

Table 5. Stopword Removal Process

Comments Before	Comments after Removal Process
"bagusss", "pak", "ganjar", "pranowo"	"bagusss", "pak", "ganjar", "pranowo"
"semoga", "menjadi", "pemimpin", "yang", "menyatukan"	"semoga", "menjadi", "pemimpin", "menyatukan"
"prabowo", "selalu", "di", "hati"	"prabowo", "selalu", "hati"

3.3. Data labeling

The labeling comes after the data preprocessing stage. In this study, data is labeled as "positive" or "negative". Data can be categorized as positive or negative using classes from an Indonesian dictionary, which includes both positive and negative words. After identifying words with positive and negative meanings in a sentence, the next step is to calculate the sentence's weight value by summing the values of the opinion words. If the entire opinion value in a sentence is more than zero, the sentence's sentiment is considered positive. If the entire opinion value of the sentence is less than zero, the sentiment of the sentence is considered negative [13]. Table 6 displays the results of sentiment analysis for each presidential candidate.

Table 6. The Results of Sentiment Analysis

Sentiment	Ganjar Pranowo	Anies Baswedan	Prabowo Subianto
Positive	17,724 (83.58%)	13,524 (88.20%)	45,789 (87.17%)
Negative	3,482 (16.42%)	1,809 (11.80%)	6,739 (12.83%)

3.4. Creation of Train Data and Test Data

The data is separated into two parts: training data and testing data, with percentages assigned to each [14]. The data in this study is split into 80% training data and 20% test data.

3.5. Classification using Support Vector Machine (SVM)

In this step, sentiment is classified using four types of kernels in the Support Vector Machine algorithm after the data has been divided [15]. The four types of kernels are RBF, Linear, Polynomial, and Sigmoid. The accuracy results of the classification are presented in Table 7.

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 2, September 2023, pp. 1006-1014
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

Table 7. SVM Classification Results

Name	RBF	*Sigmoid*	*Linear*	*Polynomial*
Ganjar Pranowo	88%	81%	89%	85%
Anies Baswedan	89%	86%	90%	88%
Prabowo Subianto	91%	85%	92%	88%

3.6. Evaluation

Evaluation is carried out using the confusion matrix function. Table 8 - 10 provide the comparative results of the evaluation using the confusion matrix for each presidential candidate in the classification process utilizing four kernels' SVM algorithm.

Table 8. Comparison of Anies Baswedan's confusion matrix results

Kernel	Sentiment	*Precision (%)*	*Recall (%)*	*f1-Score(%)*
RBF	Negative	88	13	23
	Positive	89	100	92
Sigmoid	Negative	41	20	27
	Positive	89	96	93
Linear	Negative	65	51	57
	Positive	93	96	95
Polynomial	Negative	61	39	16
	Positive	88	99	93

Table 9. Comparison of Prabowo Subianto's confusion matrix results

Kernel	Sentiment	*Precision (%)*	*Recall (%)*	*f1-Score(%)*
RBF	Negative	79	38	51
	Positive	91	98	95
Sigmoid	Negative	43	31	36
	Positive	90	94	92
Linear	Negative	72	60	65
	Positive	94	96	95
Polynomial	Negative	56	15	24
	Positive	89	98	94

Table 10. Comparison of Ganjar Pranowo's confusion matrix results

Kernel	Sentiment	*Precision (%)*	*Recall (%)*	*f1-Score(%)*
RBF	Negative	77	37	50
	Positive	90	98	94
Sigmoid	Negative	35	31	33
	Positive	88	90	89
Linear	Negative	66	59	62
	Positive	93	94	94
Polynomial	Negative	59	10	17
	Positive	86	99	92

4. CONCLUSIONS

Based on the research, sentiment analysis is effectively used to determine public opinion on the 2024 Indonesian presidential candidates. Ganjar Pranowo made 82 Facebook posts with 21,206 comments, Prabowo Subianto made 42 posts

with 52,528 comments, and Anies Baswedan made 32 posts with 15,333 comments from July 1 to July 31, 2023. The highest positive sentiment and the lowest negative sentiment, from highest to lowest rank, are occupied by Anies Baswedan, Prabowo Subianto, and Ganjar Pranowo. In evaluating the performance of four Support Vector Machine kernel algorithms, it was found that the best precision value for positive sentiment was in the linear kernel, with a value between 93% and 94%. Meanwhile, the RBF kernel has the highest precision value for negative sentiment, ranging between 77% and 88%.

REFERENCES

[1] D. M. Liando, "Pemilu dan Partisipasi Politik Masyarakat (Studi Pada Pemilihan Anggota Legislatif Dan Pemilihan Presiden Dan Calon Wakil Presiden Di Kabupaten Minahasa Tahun 2014)," *J. LPPM Bid. EkoSosBudKum*, vol. 3, no. 2, pp. 14–28, 2016.

[2] I. Permata Sari, "Keberpihakan Media Dalam Pemilihan Presiden 2014," *J. Penelit. Komun.*, vol. 21, no. 1, pp. 73–86, 2018, doi: 10.20422/jpk.v21i1.492.

[3] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)," *Pros. SINTAK*, pp. 480–486, 2018.

[4] A. Rachmat C and Y. Lukito, "Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes," *J. Inform. dan Sist. Inf. Univ. Ciputra*, vol. 02, no. 02, pp. 26–34, 2016, [Online]. Available: https://github.com/sastrawi/sastrawi

[5] D. A. Ramadhan, Y. Nurhadryani, I. Hermadi, S. Srolwlfdo, and S. Fdqglgdwhv, "&dpsdljq $qdo\vlv ri 6rfldo 0hgld 8wlol]dwlrq lq -dnduwd /hjlvodwlyh (ohfwlrq," pp. 102–107, 2014.

[6] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019, doi: 10.30864/eksplora.v9i1.254.

[7] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.

[8] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[9] A. Rahmansyah, O. Dewi, P. Andini, T. Hastuti, P. Ningrum, and M. E. Suryana, "Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine," *Semin. Nas. Apl. Teknol. Inf.*, pp. 1–7, 2018.

[10] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.

[11] N. O. F. Elssied, O. Ibrahim, and A. H. Osman, "A novel feature selection based on one-way ANOVA F-test for e-mail spam classification," *Res. J. Appl. Sci. Eng. Technol.*, vol. 7, no. 3, pp. 625–638, 2014, doi: 10.19026/rjaset.7.299.

[12] S. Z. Pranida, "Analisis Sentimen Kepuasan Pelanggan Pada Jasa Eskpedisi Menggunakan Bilstm Dan Bigru," 2022, [Online]. Available: https://dspace.uii.ac.id/handle/123456789/40715%0Ahttps://dspace.uii.ac.id/bitstream/handle/123456789/40715/18523066.pdf?sequence=1

[13] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 2, no. 1, pp. 32–41, 2017, doi: 10.31284/j.integer.2017.v2i1.95.

[14] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279, 2019, doi: 10.26418/jp.v5i3.34368.

[15] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm," *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022.