Available online at http://icsejournal.com/index.php/JCSE
Journal of Computer Science an Engineering (JCSE)
e-ISSN 2721-0251
Vol. 5, No. 1, February 2024, pp. 1-11

Analyzing Public Trust in Presidential Election Surveys: A Study Using SVM and Logistic Regression on Social Media Comments

Marcel Afandi[1*], Khairunnisak Nur Isnaini[2]

[1,2] Department of informatics, Universitas Amikom Purwokerto, Purwokerto, Indonesia
[1]marcelafandi026@gmail.com*; [2]nisak@amikompurwokerto.ac.id
*corresponding author

ARTICLE INFO

Article History:
Received January 27, 2024
Revised February 15, 2024
Accepted February 29, 2024

Keywords:
Election Survey Results,
Sentiment Analysis,
Support Vector Machine,
Logistic Regression

Correspondence:
E-mail:
marcelafandi026@gmail.com

ABSTRACT

In the context of democracy in Indonesia, elections play a crucial role, and survey agencies often publish their results on social media. User responses, especially from voters, often express dissatisfaction, including distrust, insults, and negative comments, if the candidate they support receives low survey results. Therefore, this study aims to examine the level of public trust in the survey results of Presidential candidates in 2024 using the Support Vector Machine (SVM) and Logistic Regression algorithms. The study utilized data from 1778 Instagram comments and 985 Twitter tweets. The process involved problem identification, data collection, and system implementation, such as preprocessing, labeling, SMOTE, TF-IDF, data splitting, model classification, and evaluation. The results show that SVM with an 80% training data and 20% test data scenario provides high accuracy, namely 93.19% from Instagram and 91.19% from Twitter. Logistic Regression, with the highest accuracy of 89.79% from Instagram and 88.01% from Twitter in the same scenario. Sentiment analysis using SVM scenario one resulted in 195 positive comments and 216 negative comments. Logistic Regression scenario one shows 180 positive sentiments and 216 negative sentiments. From the classification results, it can be concluded that the level of public trust tends to be negative towards the survey results of the 2024 Presidential candidates, both using SVM and Logistic Regression.

1. Introduction

Elections are a crucial process in Indonesia's democratic system, allowing citizens to choose their representatives. However, this process is intertwined with power dynamics and complex interactions among political actors [1]. One of the most anticipated events in the upcoming general election is the selection of the President and Vice President [2]. Consequently, numerous feasibility surveys have been conducted by survey institutions to gauge the strength and electoral trends of the candidates. These surveys collect data on people's opinions, attitudes, and behaviors regarding the presidential candidates, which are then analyzed to produce survey results [3].

The results of these surveys often spark extensive discussions on social media, attracting significant public attention and serving as valuable research material [4]. User comments on these surveys sometimes question the neutrality of the survey institutes. For instance, a comment on Instagram reads, "Masih percaya survei-survei seperti ini, waduh berarti akal sehatnya tertutup. Lembaga survei itu untuk keuntungan, bayarannya mahal sekali. Sekali survei itu mahal sekali. Lembaga survei bukan lembaga sosial" [5]. In addition to expressing doubts about neutrality, users also express negative sentiments if the presidential candidate they support receives unfavorable survey results [6]. This public distrust has led to social media users becoming divided, which could potentially harm democratic harmony and the state [7].

However, advancements in information technology have made it possible to access and analyze large-scale data related to social media comments. Sentiment analysis, in particular, has emerged as a solution to help interpret people's reactions to survey results in a more nuanced manner [8]. Sentiment analysis involves determining the sentiment (positive or negative) of comments or

http://dx.doi.org/10.36596/jcse.v5i1.791

opinions expressed by users on social media [9]. Machine learning methods are often preferred for sentiment analysis due to their ability to handle complex tasks [10]. Algorithms such as Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, Random Forest, and K-Mean are commonly used for sentiment analysis [11].

Previous research by Singgalen (2022) [12] demonstrated that the SVM algorithm outperformed other algorithms, including NBC and DT, when using the SMOTE Upsampling operator. The SVM algorithm achieved a test result of 99.41% accuracy, compared to 98.08% for NBC and 94.40% for DT. Similarly, Pratama's research [13] comparing DT and Logistic Regression algorithms showed promising results, with DT reaching 88.7% accuracy and Logistic Regression achieving 88.89% accuracy. Given its good data classification capabilities, Logistic Regression was chosen for its high accuracy and precision.

This research aims to analyze the sentiment of social media users in response to survey institutions' surveys on Indonesian Presidential candidates for the 2024 Election. Data will be collected from Instagram and Twitter using web scraping techniques and then classified using the Support Vector Machine and Logistic Regression methods to determine positive and negative sentiments. A comparative analysis between SVM and Logistic Regression will be conducted to determine which method is superior in classification prediction, considering their accuracy and suitability for the research objectives [14].

2. Method

Research methodology is a systematic research stage that encompasses everything from research preparation to the collection of final results. This section explains the methods used in this research, detailing several stages: problem identification, data collection, system implementation, discussion of results, and implementation of findings. The stages of the research can be seen in Figure 1.

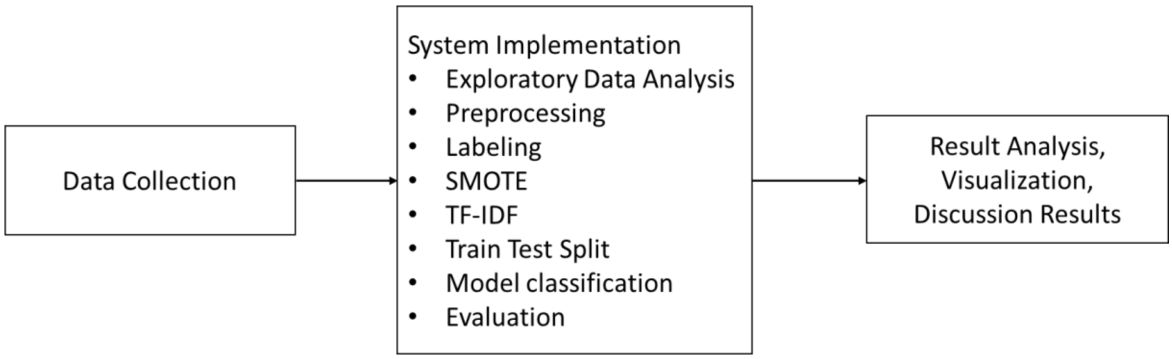

Figure 1. Diagram of Research Methodology

2.1 Data Collection

This step is conducted by researchers to gather the necessary data for completing the required research documents. The method employed in this research is Web Scraping, which enables the collection of structured data. Structured data refers to information that is easily readable or processed by a computer for inclusion in a database. Instead of relying on humans to read or process web pages, computers can quickly utilize the data in innovative and practical ways. With the right instructions, a machine can perform this task in less than a second. It can also repeat the process tirelessly and at any scale [15].

Web Scraping was used to gather results from social media users regarding survey outcomes on presidential candidates in the 2024 election. Data was obtained from Twitter (985 pieces) and Instagram (1778 pieces), collected between December 1, 2023, and December 15, 2023, using the Instant Data Scraper extension and the Twitter Harvest Library. The Instant Data Scraper extension functions by first searching for the Instagram post and opening the comment column, then clicking

http://dx.doi.org/10.36596/jcse.v5i1.791

through the user's comment replies one by one. The extension automatically retrieves all the comment data, which can then be saved in CSV format.

Table 1. Scraping Data Instagram

No	Data Instagram
1	iyain aja. media dia sendiri. apa kata hari tanoe. dia sedang menjilat
2	Servey di sini aja min.... yg pilih anis mana love nya?
3	Ini sebagai bahan mainin hasil pemilu, akan berdalih, hasil tidak jauh dari survey
4	Cuman Indonesia yang sudah tau hasil pemilu padahal belum di laksanakan dan masih lama
5	Ya terutama di jawa tengah..militan pdi..ada yg k prabowo karena tentara secara sby pak harto dari figur militer

Table 2. Scraping data Twitter

No	Data Twitter
1	Daridulu juga sering menang survei..sampai sujud sukur
2	Baru merasa menang hasil survei saja
3	Gini ya wirr kita semua tahu selain gimmick gemoy
4	Percaya Ama survei.. pilpres aja belum.. awokawok
5	kejujuran dan profesionalitas LS mmg hrs dijaga

2.2 System Implementation

Exploratory Data Analysis (EDA) involves analyzing the dataset initially by starting with deleting unused columns, and filling in missing values. Additionally, the Case Folding process is conducted, which involves converting all letters to lowercase, removing symbols, emojis, and repetitive words [16], and identifying and converting slang words into standard Indonesian words. These steps aim to clean the dataset to facilitate the preprocessing stage. The following are the steps of Case Folding: (i) Detect and read each line of data as a tweet to identify capital letters. (ii) Convert all capital letters into lowercase. For the cleaning stage, unnecessary characters such as punctuation, URLs, usernames, mentions, hashtags, emojis, and symbols in the dataset are removed, as they can significantly affect the classification results.

Preprocessing involves converting unstructured data into structured data. In this study, all comment data obtained during data collection is processed using the following process [17], as illustrated in Figure 2.

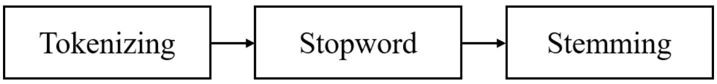

Figure 2. Preprocessing

The first step in preprocessing is Tokenizing, which separates words in a sentence to facilitate further text analysis. The words produced by this process are called tokens, and they are useful when extracting meaning from text [18]. For example, we can use tokens to detect nouns and verbs in a sentence or to identify the names of people mentioned. Additionally, the system can check each sentence separately, enabling detailed analysis. The second process, Stopword removal, eliminates words with low information from a text, allowing us to focus on important words only. This step removes meaningless or unclear words from each class, streamlining data processing and improving the speed of Natural Language Processing (NLP) [19]. The last step of preprocessing is the Stemming process, which cuts affixes or returns words to their base form. This grouping of words with the same root word optimizes data processing efficiency.

Labeling, this process involves categorizing the data into positive and negative classes. The data is processed automatically to calculate a point value, which determines the sentiment of each sentence. A lexicon vocabulary dictionary containing positive and negative words in Indonesian is used for

http://dx.doi.org/10.36596/jcse.v5i1.791

this purpose. Sentences with a score greater than 0 are classified as positive, while those with a score less than 0 are classified as negative.

Table 3. Labeling Instagram data using Lexicon

	Instagram	**Twitter**
Sentiment	***Result***	***Result***
Negative	1102	618
Positive	588	366
Total	1690	984

To categorize sentences as positive or negative, the system reads the data from the stemming process and matches it with the lexicon dictionary, which is the labeling method used in this study. The list of labels can be seen in Table 4 and 5.

Tabel 4. Instagram data Labeling sample

No	Labeling Instagram data	Score	Sentiment
1	garwo biar adil makmur sejahtera gemah ripah jawi insyaallah ganjar pranowo prabowo subianto	7	Positive
2	coba sindo bikin survei anies menang telak	-3	Negative
3	jalur asli turun cina jalur komunis baswedan presiden republik indonesia 2024 cerdas wawas	-5	Negative
4	gaspoll dukung ganjar mahfud	6	Positive
5	kebanyakan generasi jawa	-3	Negative
6	survey bayar	-2	Negative
7	tombol anis baswedan presiden 2024	3	Positive
8	malas pilih capres dukung banteng moncong putih capresnya	-12	Negative
9	anis harga mati sallam aminn	2	Positive
10	pasang ideal insyaalloh Langkah tiada haling aamiin	6	Positive

Tabel 5. Twitter data Labeling sample

No	Labeling Twitter data	Score	Sentiment
1	ganjar menang bilang akurat	1	Positive
2	tentu menang pres wapres hasil pengitungan suara hasil survei bayar sebar bohong	-4	Negative
3	tentu menang pres wapres hasil pengitungan suara hasil survei bayar sebar bohong	-4	Negative
4	percaya prabowo uang gemoy asa urut oleh elektabilitas	3	Positive
5	wkwkwk menghoax mending ambil survei litbang kompas 2022 menang ganjar bajer	-5	Negative
6	malu sebar bohong survei sebar release desember hahaha	-7	Negative
7	pokok hasil survei menang kandidat ente dukung ente bilang survive jujur salah sorry	1	Positive
8	kocak survei provinsi mending survei lembaga kredible menang ganjar semua	-8	Negative
9	prabowo menang survei sujud syukur ending kalah mulu	1	Positive
10	orang cerdas percaya survei percaya	4	Positive

SMOTE (Synthetic Minority Over-sampling Technique): Due to the greater number of negative sentiment labels in the labeling process, a data balancing feature is necessary to avoid overfitting. To address this, SMOTE (Synthetic Minority Over-sampling Technique) is chosen, which works by synthesizing new samples from the minority class. This process involves taking two samples from adjacent minority classes and combining them to create a new sample, considering the distance between samples. SMOTE is an oversampling technique commonly used to handle unbalanced data, where the proportion of minority classes differs significantly from that of majority classes, potentially biasing machine learning models towards the majority class.

http://dx.doi.org/10.36596/jcse.v5i1.791

TF-IDF: The next stage after preprocessing is to weight each word in the document using the Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF is a statistical technique used to measure the importance of a word in a document or corpus, and it is widely used in various fields, including sentiment analysis. During this process, each word in the document is weighted based on its frequency.

Train Test Split: Before classification, the dataset is divided into two parts: Training data and Test data. The training data is used to fit machine learning models, while the test data is used to evaluate the model fit. Several data division ratios are considered, such as 70:30, 80:20, 90:10, and 50:50, each with its own conditions. The 70:30 ratio provides a good balance between the size of the training data and testing data, leading to better accuracy in predicting new data. However, it requires longer training time. The 80:20 ratio offers a larger training data size, but the smaller test data size may reduce the prediction accuracy of the model. The 50:50 ratio requires less training time but may make it challenging for the model to learn and predict new data. Comparatively, the 70:30 and 80:20 ratios are preferred over others due to the small size of the dataset used in the study. A common approach is to use 70% or 80% of the data for training and the remaining 30% or 20% for testing, striking a balance between having enough data for training and evaluation.

Model Classification: In the classification stage, experiments are conducted on the data split into train and test data, using two scenarios: 70:30 and 80:20 division, and implementing the SVM and LR machine learning algorithms. The SVM algorithm is employed using the Python Library Scikit-Learn, with the researcher determining the parameters of the SVM model to be used. The classification process involves training and testing, based on the predetermined model parameters.

```
from sklearn.svm import SVC

svm = SVC(kernel='linear')
svm.fit(X_train,y_train)
```

Figure 3. Training data using SVM

The process results in a machine learning model using the SVM algorithm that can predict the label or sentiment on the test data. Subsequently, the classification stage is carried out using a predetermined algorithm, namely the Logistic Regression algorithm from the Python Library Scikit-Learn. The researcher selects the "penalty" parameter in the Logistic Regression model, which determines the type of regularization applied to the model. Regularization is a technique used to prevent overfitting and improve model generalization. In logistic regression, the penalty parameter can take on values such as "l1" or "l2", corresponding to different types of regularization, where the squared value of the coefficient is added as a penalty term to the loss function. This penalizes large coefficients and can help reduce multicollinearity among features. The code is shown in Figure 4.

```
# Algoritma Logistic Regression
from sklearn.linear_model import LogisticRegression

lr = LogisticRegression(penalty='l2')
lr.fit(X_train,y_train)
```

Figure 4. Training data using Logistic Regression

Evaluation: Evaluation is conducted after training the model for the classification of public opinion analysis on the 2024 presidential candidate survey results using the SVM and Logistic Regression algorithms. This evaluation aims to determine the scenarios and algorithms that perform best. The results are presented in the form of heatmaps, tables, and calculation steps based on the scenario testing results on a system developed with the Python programming language. The performance of the resulting model is assessed using a Confusion Matrix. This step involves calculating the system's

http://dx.doi.org/10.36596/jcse.v5i1.791

performance. After classifying the test data, calculations are performed to determine accuracy, precision, recall, and F1-score using the Confusion Matrix to measure the system's performance. This provides insights into how well the model predicts outcomes. Each column of the confusion matrix represents an instance of the predicted class.

Tabel 5. Confusion Matrix

		Predicted Class	
		Positive	***Negative***
True Class	Positive	*True Positive* (TP)	*False Negative* (FN)
	Negative	*False Positive* (FP)	*True Negative* (TN)

From Table 5, it can be interpreted as follows: True Negative (TN) indicates negative label data that is correctly predicted as a negative label, True Positive (TP) indicates positive label data that is correctly predicted as a positive label, False Negative (FN) indicates positive label data that is incorrectly predicted as a negative label, and False Positive (FP) indicates negative label data that is wrongly predicted as a positive label. After obtaining the values from the confusion matrix, the accuracy, precision, recall, and F1-score values can be calculated. The calculation formulas for precision, accuracy, recall, and F1-score are shown in the following equation:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} \quad (4)$$

Scenario 1: Confusion Matrix SVM and Logistic Regression data Instagram

For scenario one, a trial was conducted with a data sample of 80% training data and 20% test data using the SVM and Logistic Regression algorithms on the Instagram dataset with 441 test data, can be seen in figure 5 and table 6.

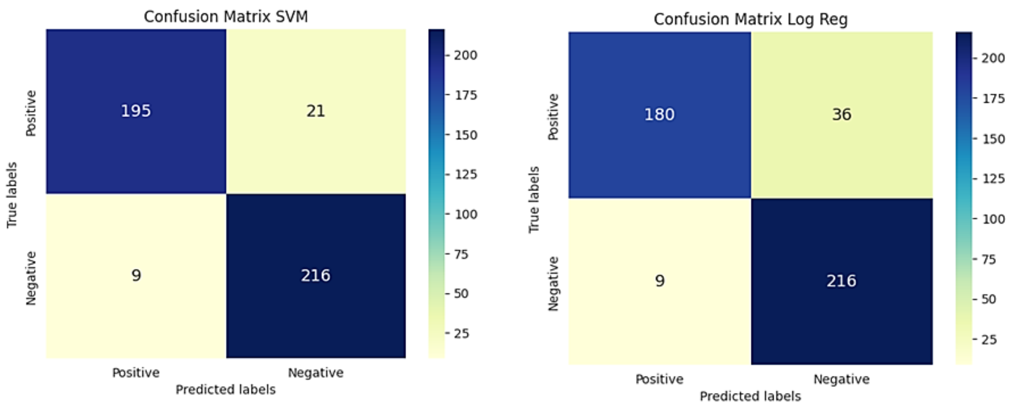

Figure 5. Scenario 1 Instagram data

Table 6. Confusion Matrix SVM Instagram data scenario 1

	SVM		**Logistic Regression**	
	Predicted Positif (1)	***Predicted Negatif (3)***	***Predicted Positif (1)***	***Predicted Negatif (3)***
Actual Positif (1)	195 (TP)	21 (FN)	180 (TP)	36 (FN)
Actual Negatif (3)	9 (FP)	216 (TN)	9 (FP)	N)

http://dx.doi.org/10.36596/jcse.v5i1.791

Scenario 2: Confusion Matrix SVM and Logistic Regression data Instagram

For the second scenario, a trial was conducted with a data sample of 70% training data and 30% test data using the SVM and Logistic Regression algorithms on the Instagram dataset with 662 test data, can be seen in figure 6 and table 7.

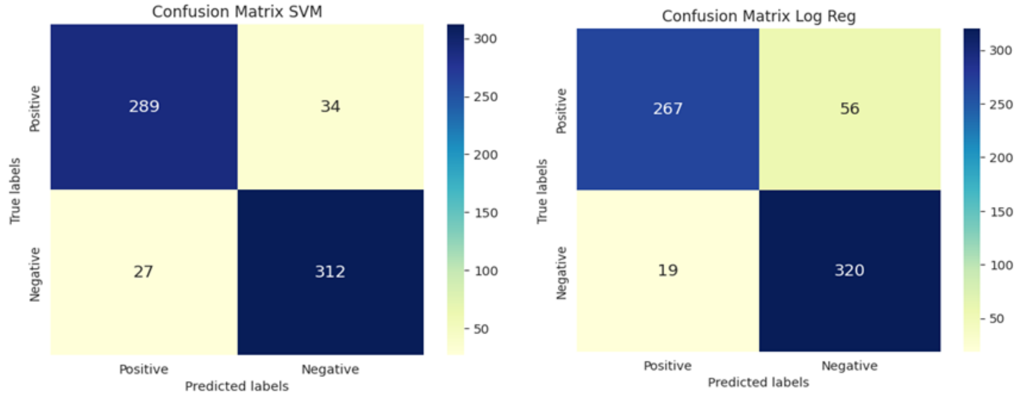

Figure 6. Scenario 2 Instagram data

Table 7. Confusion Matrix SVM Instagram data scenario 2

	SVM		Logistic Regression	
	Predicted Positif (1)	***Predicted Negatif (3)***	***Predicted Positif (1)***	***Predicted Negatif (3)***
Actual Positif (1)	289 (TP)	34 (FN)	267 (TP)	56 (FN)
Actual Negatif (3)	27 (FP)	312 (TN)	19 (FP)	320 (TN)

Scenario 1: Confusion Matrix SVM and Logistic Regression data Twitter

For scenario one, a trial was conducted with a data sample of 80% training data and 20% test data using the Logistic Regression and SVM algorithms on the Twitter dataset with 242 test data, can be seen in figure 7 and table 8.

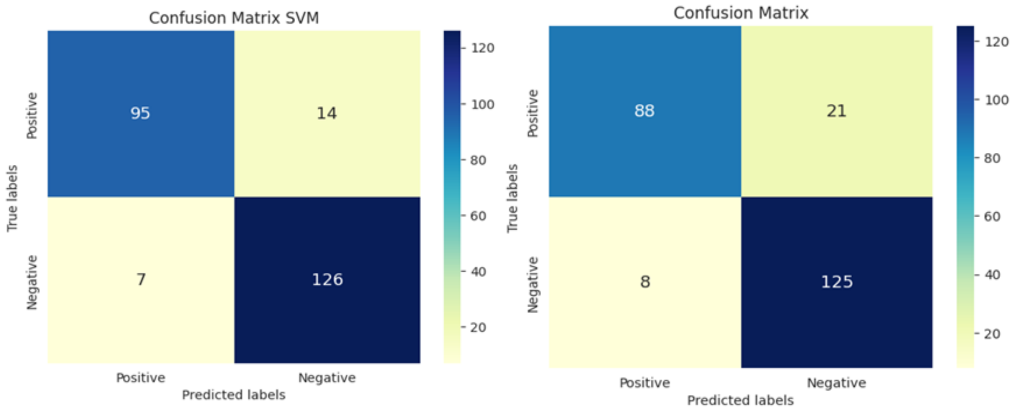

Figure 7. Scenario 1 Twitter data

Table 8. Confusion Matrix SVM Twitter data scenario 1

	SVM		Logistic Regression	
	Predicted Positif (1)	***Predicted Negatif (3)***	***Predicted Positif (1)***	***Predicted Negatif (3)***
Actual Positif (1)	95 (TP)	14 (FN)	88 (TP)	21 (FN)
Actual Negatif (3)	7 (FP)	126 (TN)	8 (FP)	125 (TN)

http://dx.doi.org/10.36596/jcse.v5i1.791

Scenario 2: Confusion Matrix SVM and Logistic Regression data Twitter

For scenario one, a trial was conducted with a data sample of 70% training data and 30% test data using the Logistic Regression and SVM algorithms on the Twitter dataset with 362 test data, can be seen in figure 8 and table 9.

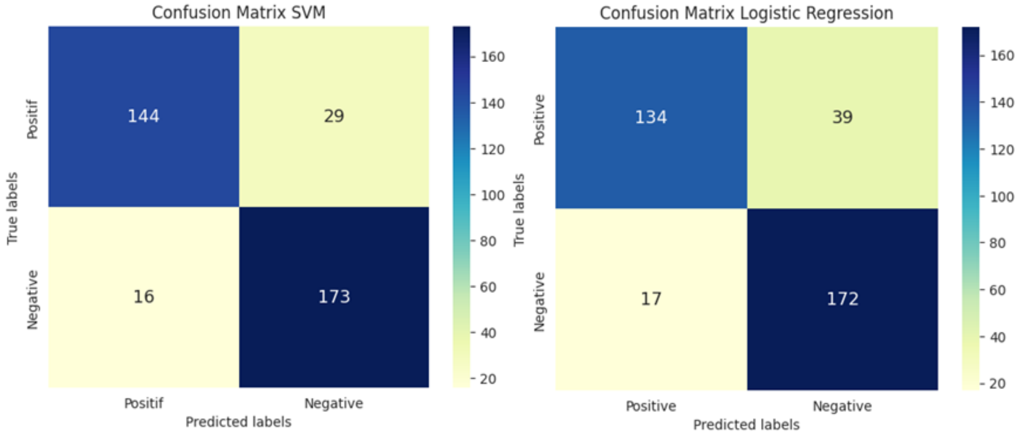

Figure 8. Scenario 2 Twitter data

Table 9. Confusion Matrix SVM dan Logistic Regression Twitter data scenario 2

	SVM		Logistic Regression	
	Predicted Positif (1)	***Predicted Negatif (3)***	***Predicted Positif (1)***	***Predicted Negatif (3)***
Actual Positif (1)	144 (TP)	29 (FN)	134 (TP)	39 (FN)
Actual Negatif (3)	16 (FP)	173 (TN)	17 (FP)	172 (TN)

3. Results and Discussion

3.1 Result Analysis

After evaluating the performance of the model against test data with different scenarios using confusion matrix, the results of the performance metrics of each model have been summarized in the following table 10.

Table 40. Classification result using SVM and Logistic Regression

		SVM		Logistic Regression	
		Scenario 1 (80% train & 20% test)	***Scenario 2 (70% train & 30% test)***	***Scenario 1 (80% train & 20% test)***	***Scenario 2 (70% train & 30% test)***
Instagram	Accuracy	93.19 %	90.79 %	89.79 %	88.67 %
	Precision	95.58 %	991.45 %	95.23 %	93.35 %
	Recall	90.27 %	89.47 %	83.33 &	82.66 %
	F1-Score	93.12 %	89.98 %	88.59 %	87.15 %
Twitter	Accuracy	91.19 %	87.56 %	88.01 %	84.53 %
	Precision	93.12 %	90.00 %	91.66 %	88.74 %
	Recall	87.15 %	83.23 %	80.73 %	77.45 %
	F1-Score	89.90 %	86.35 %	85.14 %	82.13 %

Based on Table 4, it can be concluded that the 80:20 data division yields higher Accuracy, Precision, Recall, and F1-Score results compared to Scenario 2, which divides the data into 70:30 from both Instagram and Twitter data. This occurs because the 80:20 data division provides more training data for learning classification algorithms and a relatively smaller test data compared to Scenario 2. Furthermore, the results of Accuracy, Precision, Recall, and F1-Score from Logistic Regression are still outperformed by the 80:20 data division, similar to the classification results of the Support Vector Machine (SVM) algorithm. Based on Table 4 and Table 5, it can be concluded that dividing the data into 80% training data and 20% test data demonstrates good performance in classifying public sentiment analysis of survey results on candidates for President and Vice President in the 2024 elections from social media, compared to the 70:30 data division.

http://dx.doi.org/10.36596/jcse.v5i1.791

3.2 Visualization

After reviewing the classification results of the test data from both scenarios for Instagram using the SVM and Logistic Regression algorithms, it was observed that the SVM algorithm outperformed Logistic Regression in terms of accuracy. The figure 9 and 10 illustrates the classification results, indicating that the SVM algorithm achieved higher accuracy results. Specifically, the highest accuracy for the SVM algorithm was attained in scenario one, which involved an 80% training data and 20% test data split.

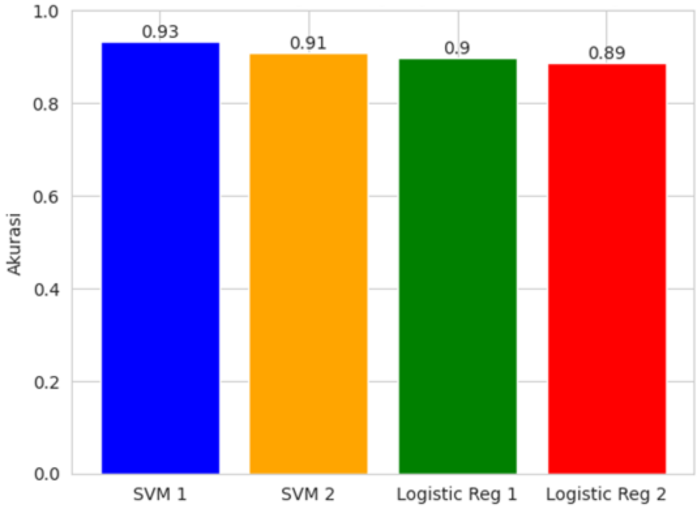

Figure 5. Accuracy Score scenario 1

In the classification using SVM and Logistic Regression algorithms with data from Twitter, the results vary significantly. Only SVM scenario 1 achieved an accuracy of over 90%, while the accuracy of Logistic Regression scenario 2 was the lowest.

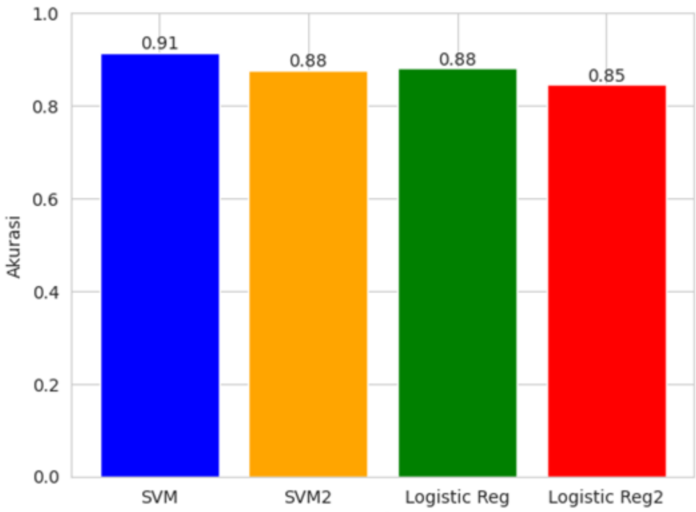

Figure 6. Accuracy score scenario 2

3.3 Discussion Results

After reviewing the classification results of the test data from Instagram and Twitter, it was found that the Support Vector Machine (SVM) algorithm with scenario one (80:20) achieved the highest accuracy. Subsequently, a comparison was conducted on the sentiment distribution from both algorithms to analyze the number of public reviews for each sentiment.

http://dx.doi.org/10.36596/jcse.v5i1.791

Table 51. Sentiment result using SVM and Logistic Regression.

		SVM		Logistic Regression	
		Scenario1 (80% train & 20% test)	***Scenario2 (70% train & 30% test)***	***Scenario1 (80% train & 20 test)***	***Scenario2 (70% train & 30 test)***
Instagram	Positive	195 Comment	289 Comment	180 Comment	267 Comment
	Negative	216 Comment	312 Comment	216 Comment	320 Comment
Twitter	Positive	95 Comment	144 Comment	89 Comment	134 Comment
	Negative	126 Comment	173 Comment	125 Comment	172 Comment

Based on Table 11, it can be concluded that both Instagram and Twitter data, in both scenarios 1 and 2, exhibit a higher prevalence of negative sentiment compared to other scenarios. The SVM algorithm particularly excels in classifying negative sentiments. Conversely, in scenarios 1 and 2 using LR, there is a notable imbalance in the number of sentiments between the two classes, despite the use of SMOTE for feature selection to address unbalanced data. Therefore, in both scenarios, the majority of sentiments are classified as Negative. These results suggest that, on average, survey results regarding the 2024 Presidential and Vice Presidential candidates, as depicted on social media, are predominantly negative. This demonstrates the successful application of sentiment analysis using the Support Vector Machine (SVM) and Logistic Regression algorithms to Instagram and Twitter datasets pertaining to the survey results on the 2024 Presidential and Vice Presidential candidates.

4. Conclusion

Based on the results of the discussion conducted by the researchers, it is concluded that the Support Vector Machine (SVM) and Logistic Regression algorithms have been successfully implemented in sentiment analysis of the level of public trust in the survey results of Presidential candidates in the 2024 election. The SVM classification performance is superior to Logistic Regression, especially in scenario 1 of SVM Instagram data, with an Accuracy level of 93.19%, Precision of 95.58%, Recall of 90.27%, and F1-Score of 93.12%. For datasets from Twitter, the highest accuracy value of the classification model is also found in the SVM scenario 1 algorithm, with an Accuracy value of 91.19%, Precision of 93.12%, Recall of 87.15%, and F1-Score of 89.90%. The classification results using the SVM algorithm and Logistic Regression indicate that sentiments towards the survey results of Presidential candidates in the 2024 election tend to be negative. Specifically, the sentiment analysis using the SVM algorithm in scenario 1 on Instagram data resulted in 216 comments classified as negative sentiment and 195 comments as positive sentiment out of 441 test data. On Twitter data, scenario 1 resulted in 95 comments classified as positive sentiment and 126 comments as negative sentiment out of 242 test data.

This research only considers Indonesian opinions, so it may be necessary to redevelop the data preprocessing process by incorporating foreign language translations and handling a larger amount of data. Additionally, using other classification algorithm methods such as Unsupervised Learning or employing undersampling techniques could be explored to determine the optimal performance in the classification process.

References

[1] Y. A. Aliano and M. J. Adon, "Percaturan Politik Genealogi Kekuasaan dalam Sistem Pemilu ' 2024 ' di Indonesia Menurut Etika Michel Foucault," *J. Filsafat Indones.*, vol. 6, no. 3, pp. 474–485, 2023.

[2] Komisi Pemilihan Umum, "Kilas Pemilu Tahun 2024," *Komisi Pemilihan Umum*. p. 1, 2023. [Online]. Available: https://www.kpu.go.id/page/read/1136/kilas-pemilu-tahun-2024

[3] Hidayah A., "Pentingnya Hasil Survei Jelang Pemilu Tahun 2024," *https://www.theindonesianinstitute.com/pentingnya-hasil-survei-jelang-pemilu-tahun-2024/*. p. 1, 2021. [Online]. Available: https://www.theindonesianinstitute.com/pentingnya-hasil-survei-jelang-pemilu-tahun-2024/

http://dx.doi.org/10.36596/jcse.v5i1.791

[4] M. R. Fais Sya' bani, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 265-273, Apr. 2022, doi: 10.30865/jurikom.v9i2.3989.
[5] SindoNews, "Survei Litbang Kompas: Ganjar Pranowo Capres Pilihan Gen Z," *SindoNews*. p. 1, 2023. [Online]. Available: https://nasional.sindonews.com/read/1219263/12/survei-litbang-kompas-ganjar-pranowo-capres-pilihan-gen-z-1696590636
[6] N. Pratiwi and P. Nola, "The Effect of Digital Literacy on the Psychology of Children and Adolescents," *J. iIlmiah Progr. Stud. Pendidik. Bhs. dan Sastra Indones.*, vol. 6, no. 1, pp. 11–24, 2019.
[7] K. Anwar, A. Syar'i, and F. Liadi, "Pemilihan presiden tahun 2019," 2019.
[8] H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine," *Inf. Syst. Dev.*, vol. 5, no. 2, pp. 7–13, 2020.
[9] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Di Twitter," *J. TEKNO KOMPAK*, vol. 17, no. 2, pp. 25–38, 2022.
[10] Y. Jianhua, W. U. Yang, L. U. Xin, Z. Yanyan, Q. I. N. Bing, and L. I. U. Ting, "Recent advances in deep learning based sentiment analysis," pp. 1–24, 2020.
[11] I. Muslim Karo Karo *et al.*, "Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1423–1430, 2023, doi: 10.47065/josh.v4i4.3943.
[12] Y. A. Singgalen, "Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1634–1646, 2022, doi: 10.47065/bits.v4i3.2766.
[13] Y. A. Pratama, F. Budiman, S. Winarno, and D. Kurniawan, "Analisis Optimasi Algoritma Decision Tree, Logistic Regression dan SVM Menggunakan Soft Voting," *J. Media Inform. Budidarma*, vol. 7, pp. 1908–1919, 2023, doi: 10.30865/mib.v7i4.6856.
[14] A. Jalu Narendra Kisma and C. Raras Ajeng Widiawati, "Analisis Aplikasi Di Playstore Berdasarkan Rating Dan Type Menggunakan Naive Bayes Dan Logistik Regresi," *Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 174–184, 2023, [Online]. Available: http://jurnal.mdp.ac.id
[15] S. Platform, D. Resources, and P. Pricing, "The Beginner ' s Guide To Web Scraping What is web scraping ?," pp. 1–16.
[16] N. A. Verdikha, R. Habid, and A. J. Latipah, "Analisis DistilBERT dengan Support Vector Machine (SVM) untuk Klasifikasi Ujaran Kebencian pada Sosial Media Twitter," *Metik J.*, pp. 101–110, 2023, doi: 10.47002/metik.v7i2.583.
[17] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.
[18] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.
[19] A. Salsabila, J. J. Sihombing, and R. I. Sitorus, "Implementasi Algoritma Support Vector Machine Untuk Analisis Sentimen Aplikasi OLX di Playstore," *J. Informatics Data Sci.*, vol. 1, no. 2, 2022, doi: 10.24114/j-ids.v1i2.42597.
[20] dan A. C. Ian Goodfellow, Yoshua Bengio, "Deep Learning," *Prmu*, pp. 1–10, 2016, [Online]. Available: www.deeplearningbook.org

http://dx.doi.org/10.36596/jcse.v5i1.791