Versi Online tersedia di : http://bsti.ubd.ac.id/e-jurnal

JURNAL TECH-E

| 2581-1916 (Online) |

Artikel

Sentiment Analysis of Comments on Instagram Posts of Indonesia's 2024 Presidential Candidates Using The Support Vector Machine Method

Alqis Rausanfita[1], Arip Ramadan[2], Muhammad Dzulfikar Fauzi[3], Qori Emalia Putri Mafidah[4], Emilia Ramona[5], Yudha Mahardika Putra[6]

[1,3,4,5,6]Informatika, Institut Teknologi Telkom Surabaya, Indonesia
[2]Sistem Informasi, Institut Teknologi Telkom Surabaya, Indonesia

SUBMISSION TRACK

Recieved: August 10, 2023
Final Revision: August 20, 2023
Available Online: August 28, 2023

KEYWORD

sentiment analysis, text mining, classification, support vector machine, comments on instagram post

KORESPONDENSI

E-mail: alqisrausanfita@ittelkom-sby.ac.id

A B S T R A C T

The rising number of Instagram user affecting higher number of comments appear on post especially Instagram accounts of Indonesia's 2024 presidential candidates that made it difficult to understand the public sentiment towards presidential candidate. Therefore, this research aims to classify Indonesian sentiment on Instagram comments of 2024 Indonesian presidential candidates using the Support Vector Machine method. This research has several stages that are crawling data, preprocessing data, TF-IDF Weighting, and Classification. The classified sentiment is divided into three classes, namely positive, negative, and neutral. The results shows that Sentiment Analysis of Comments on Instagram Posts of Indonesia's 2024 Presidential Candidates Using The Support Vector Machine Method has a good accuracy value of 89.41%. This results also obtain recall and precision values of 89% and 87% respectively. The high values of precision and recall mean that the accuracy of sentiment analysis is quite good and there is minimal discrepancy in the information obtained from the experiments that have been carried out.

INTRODUCTION

The development of information technology is accompanied by the development of social media, such as Instagram [1]. The number of Instagram users in Indonesia reaches 45 million per month, this data shows the largest users in Asia Pacific [2]. The number of Instagram users is mostly dominated by generation Z people [3]. The number of generation Z people who use Instagram makes the comment feature on Instagram widely used. Instagram users can comment on a post on an account on Instagram. On Instagram accounts with a high number of followers, the number of comments obtained also increases, such as on the Instagram accounts of Indonesia's 2024 Presidential Candidates who have an average of comment each post of 1,843 comments. The high number of comments makes it

difficult to understand the public sentiment towards presidential candidates. To overcome a lot of comment data, an automatic sentiment analysis method is needed [4].

Sentiment analysis has been studied using several classification methods, namely Naïve Bayes, Support Vector Machine, AdaBoost, Random Forest, Decision Tree, KNN and Logistic Regression [5]. Among these methods, the Support Vector Machine method shows the best accuracy results. Previous research compared the KNN method with the Support Vector Machine method in sentiment analysis of the United States president Donald Trump. This comparison aims to find out which algorithm has the best accuracy and fastest processing time. In this study, sentiment analysis using the svm method produced higher accuracy than the KNN method, which was 89.70%, while the KNN method got an accuracy result of 88.76% [6]. In addition, data in the form of Gojek user responses are classified into two sentiments, namely positive sentiment and negative sentiment using the Support Vector Machine (SVM) method. The results of sentiment classification from the results of labelling sentiment scoring data using the Support Vector Machine method on Gojek resulted in the best overall accuracy rate of 79.19% [7]. Another study used the SVM method for sentiment analysis of Instagram comment data regarding Covid-19 and obtained an accuracy value of 80.23% [8]. The high accuracy rate shows that the Support Vector Machine method is optimal for classifying the sentiment of each data.

On the other hand, research related to sentiment analysis of presidential candidates in Indonesia in 2024 with the Support Vector Machine method has not been conducted. The number of Instagram account that follows presidential candidates in Indonesia is reaching 5.9 million people that causing high number of comments posted every day. The high number of comments needs to be analyzed through sentiment analysis automatically to understand the public sentiment towards presidential candidates. Therefore, this research aims to classify Indonesian sentiment on Instagram comments of 2024 Indonesian presidential candidates using the Support Vector Machine method. The classified sentiment is divided into three classes, namely neutral, positive, and negative. By doing sentiment analysis, it is hoped that each 2024 Indonesian presidential candidate can find out the sentiment of the community in each Instagram post comment. The next section of this paper will be dedicated to the description of related work (Section 2). The methodology of the research are reported in Section 3. The results of analysis setiment in Section 4. In Section 5, we describe the analysis of the result, followed by a conclusion in Section 6.

I. LITERATURE REVIEW

Some studies have used some classification methods such as support vector machine decision trees logistic regression etc. Compares Naïve Bayes Support Vector Machine and Logistic Regression methods for various aspects of analyzing Instagram data to detect sentiment. Sentiment analysis deals with opinion divided into positive negative and neutral. This study shows that machine learning methods such as Naive Bayes have the highest accuracy [9].

In another study, the decision tree method and Support Vector Machine with the resulting features from TF-IDF were used for sentiment analysis of comments on Instagram related to the performance of the Indonesian Football Association (PSSI) [4].

The Support Vector Machine method is used to analyze reviews from users of the movie streaming application in Play store. In previous research, it was found that the Support Vector Machine algorithm has a fairly high accuracy rate for classifying positive and negative responses from users of the film streaming application on Google Play store [10].

Another study aims to analyze the sentiment in each sentence from twitter users towards Indonesian football using K-Nearest Neighbor. The data used in the study is twitter data related to football in Indonesia taken

from the official PSSI twitter account. The data was collected and then several stages were carried out, namely preprocessing and word weighting using Term Frequency-Inverse Document Frequency (TF-IDF). Optimal accuracy results are obtained at a value of k = 23, which is 79.99% [11].

In other research, comparing the Support Vector Machine method and the naïve bayes method in sentiment analysis of Covid-19 information. The Support Vector Machine method produces a better accuracy value than using the Support Vector Machine method, which is 80.23%, while the Naïve Bayes method produces an accuracy value of 78.02% [8].

In addition to Instagram, the researchers also analyzed sentiment data on other social media, such as Twitter. One of them is analyzing the sentiment of the Indonesian people towards the Pre-Employment Card Program about the government's breakthrough efforts to overcome unemployment and victims of labor layoffs. The twitter data obtained was analyzed using the Support Vector Machine Method. The study divided the results of the Support Vector Machine classification into three classes, namely neutral as much as 98.34%, negative class as much as 0.99%, and positive class as much as 0.66% [12].

Another study analyzed twitter sentiment data on public opinion about this campus, University of Muhammadiyah Malang (UMM). This research compared two methods, namely Naïve Bayes and Support Vector Machine. The results of the comparison of the two methods show that classification using the Naïve Bayes method gets better accuracy results than Support Vector Machine with an accuracy of 73.65%, while classification using the Support Vector Machine method only produces an accuracy of 70.20% [13].

II. METHODS

In this research there are several stages shown in **Figure 1.**

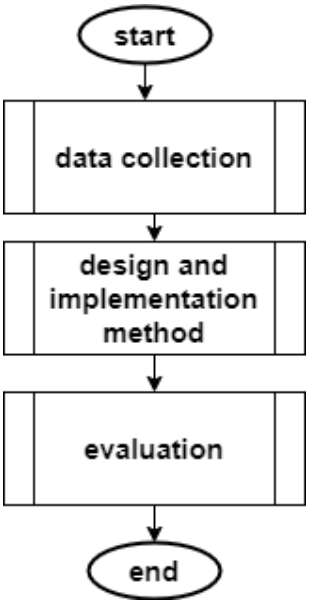

Figure 1. Flowchart of the study

Collection data

In this study, Instagram data was collected using the phantom buster application. Data crawling was carried out by searching for Indonesian-language comment data on the account posts of the 2024 Indonesian presidential candidates, namely: Prabowo, Anies Baswedan, and Ganjar Pranowo. The amount of data generated during crawling was 8,287 data. The data obtained has several attributes, but only a few attributes are used, namely comments, number of comment likes, and number of Instagram comment replies.

Design and implementation method

In this study there are several stages that can be seen in **Figure 2.**

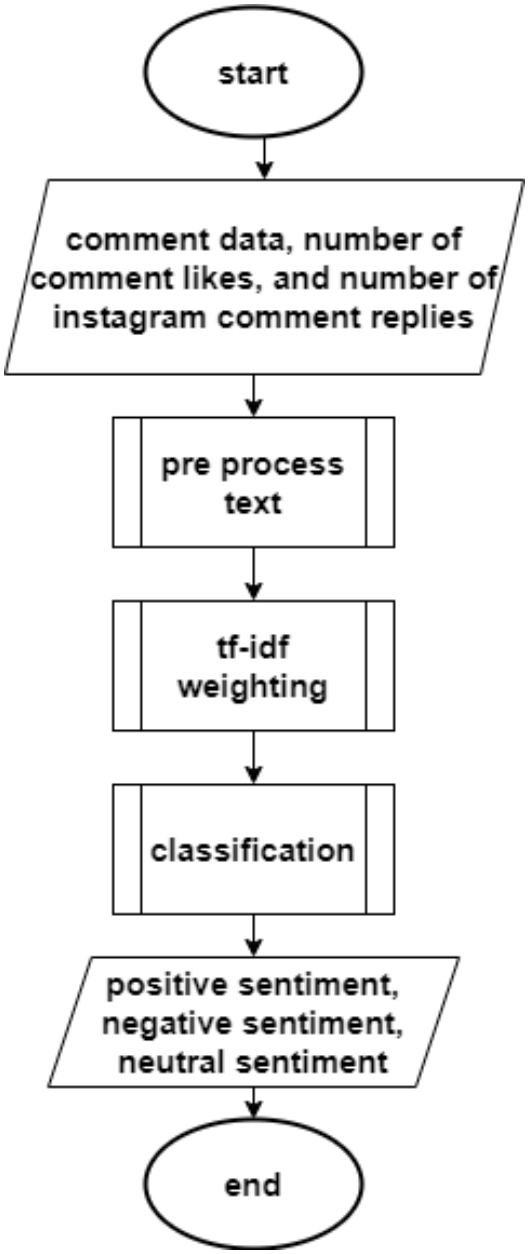

Figure 2. Design and implementation method flowchart

Instagram post comment data obtained will be pre-processed data. The pre-process stage aims to clean the data before entering the next stage [14]. This research pre process has several stages that are carried out, namely: lower casing, removing usernames, removing numbers, removing punctuations, removing leading & trailing whitespace, removing multiple whitespace into single whitespace, removing stop words, lemmatization, and tokenizing. Lower casing is a process where all characters will be converted into all lowercase letters. Remove stop words needs to be done so that all unimportant words become noise. All usernames, punctuation marks, numbers, and whitespace must be removed in order to produce a unique bag of words model. Lemmatization is the process of normalizing words to find the basic form of the word based on its lemma form [15]. In the tokenization process all sentences will be broken down into terms.

Term Frequency – Inverse Document Frequency Weighting

At this stage the net data is weighted using the TF-IDF method. TF-IDF (Term Frequency-Inverse Document Frequency) weighting is the process of converting text data to numerical data to assign a weight to each word or feature [11]. In this study Instagram comment data is considered as a document. The frequency of occurrence of words in each document is calculated. The frequency of occurrence of a word in a document indicates how important the word is in the document [16]. In addition to calculating the word frequency of each document this study also calculates the frequency of the document. Document frequency is the number of documents that contain the word. The resulting document frequency is used to calculate the inverse document frequency (idft). After getting the value of term frequency and inverse document frequency (idft) TF-IDF weight is calculated using the formula in Equation (1).

$$W_{t,d} = tf_{t,d} \; X \; idf_t \qquad (2)$$

where:
$W_{t,d}$ is tf-idf weighting
$tf_{t,d}$ is the frequency of occurrence of words
idf_t is inverse document frequency

Support Vector Machine Classification

The data for each comment in the form of non-textual weighting is added as a new feature to the textual weighting result data. Then the data is classified using the Support Vector Machine method. The support vector machine algorithm is a very popular supervised learning algorithm. The Support

Vector Machine algorithm is designed for regression and classification problems. The SVM algorithm aims to form an appropriate optimal decision boundary (called a hyperplane) which divides the n-dimensional space into different classes making it easier to assign different points to the appropriate class. In the support vector machine algorithm extreme vector points called support vectors are selected to help create the corresponding hyperplane [17].

In the Support Vector Machine method, there is a parameter C that serves to control the optimization between margin and misclassification. The larger the C parameter, the greater the penalty for misclassification. Support Vector Machine is basically a linear classifier [18]. However, Support Vector Machine can be developed into a nonlinear classifier. Basically, the Support Vector Machine learning process to find the support vector only depends on the dot product of the data in the feature space. The dot product calculation can be replaced with a kernel function. The concept of kernel trick in higher dimensional space can handle nonlinear classifier. In Support Vector Machine, there are several types of kernels, namely linear, sigmoid, polynomial, and RBF. The development of the Support Vector Machine method, the more cases that are solved using the Support Vector Machine method very quickly and accurately, one of which is in the case of data classification [10].

Evaluation

The classification results obtained must be evaluated for performance using the Confusion Matrix method. Confusion matrix is an algorithm performance evaluation central role shall be played by the matrix that represents an aggregation of predicted and actual class instances [19]. There are four indicators calculated based on the confusion matrix, namely: Accuracy, Precision, Recall, and F-measure. The confusion matrix is shown in **Table 1**.

Table 1 Confusion matrix

		Predicted Class	
		True	**False**
Actual Class	**True**	TP	FN
	False	FP	TN

Source: [20]

Accuracy is the value of the correct record accuracy of the test prediction results compared to the actual value. If the accuracy result is 100%, it indicates that the predicted condition is true according to the original. Accuracy can be calculated using the formula in Equation $Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$ (3)

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

Precision or confidence is the ratio between predicted positive items that are also positive in reality to all predicted positive items. Precision can be calculated using the formula in Equation $Precision = \frac{TP}{TP+FP}$ (4)

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall or Sensitivity is the ratio between predicted positive items that are also positive in actual to all actual positive items. The formula for obtaining the recall value is shown in Equation $Recall = \frac{TP}{TP+FN}$ (5).

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F-measure or F-Score is a measurement that assesses the reciprocity between precision and recall (weighted harmonic mean between precision and recall). The value of F-measure is obtained by calculating using the formula in Eq. $F\text{-}measure = \frac{2x precision x\, recall}{precision+recall}$ (6).

$$F - measure = \frac{2x precision x\, recall}{precision+recall} \quad (6)$$

III. RESULT

This chapter contains the results obtained. The data obtained in crawling is 8,287 data. 8,287 data obtained from crawling are still found in languages other than standard Indonesian, so it is necessary to translate them into Indonesian manually first before further processing. After the data is correct, each data is manually labelled into three types of labels, namely positive, negative, and neutral.

The manual labelling process resulted in 7,364 positive data, 437 neutral data, and 486 negative data. examples of negatively labelled data are "Ngomong Indonesia aja bikin pusim nggak ngerti,, apa lagi ngomong Inggris" and "Jangan pilih pemimpin yang seperti ini, hanya pintar menata kata, hasil kerja tidak ada yang nyata, semua dibohongi". Examples of data in the form of positively labelled sentences, are "Mantap sekali, Bapak @aniesbaswedan. Ini baru mantap kalau jadi presiden berikunya, presiden 2024" and "Pengalaman Pak Anies luar biasa, intelektual yang tak perlu diragukan lagi. Semoga Allah permudahkan urusan Pak Anies untuk menjadi RI-1. Aamiin". While examples of data labelled neutral are: "Desain no 3 @taka_dad15" and "yang mau beli kaos slide 7 bisa dm saya".

Preprocessing result

The data obtained is then carried out data cleaning. several stages are carried out, namely: lower casing, remove username, remove numbers, remove punctuations such as hyphens, question marks, exclamation marks, underscores, parentheses, hash marks, symbols &, and etc. In addition, the preprocessing stage also removes several types of whitespace, such as remove leading & trailing whitespace, remove multiple whitespace into single whitespace. The results will then be compared with the list contained in the stop word. If the term is found in the stop word list, it will be removed. The results obtained are then processed to be converted into basic words or also called lemmatization, for example the word "berbicara" if changed to the base word it will change to "bicara". The last stage is tokenizing, breaking a sentence into phrases. The result of the pre-process is shown in **Table *2***.

Table 2. Pre-process result data

No	Comment on Instagram Posts	Label
1	['ngomong', 'indonesia', 'bikin', 'pusing', 'erti', 'ngomong', 'inggris']	Negatif
2	['pintar', 'berkatakata', 'kerja nol komen fakta realita hati sehat']	Negatif
3	['moga', 'prabowoganjar', 'menang']	Positif
4	['moga', 'allah', 'swt', 'tuntun', 'bapak', 'anies', 'meridhoinya', 'besar', 'harap', 'warga', 'makassar', 'bapak', 'presiden']	Positif
5	['desain', 'no']	Netral
6	['beli', 'kaos', 'slide', 'dm']	Netral

Data Ratio Composition Testing Results

There are two types of data that are input to the TF-IDF weighting process, namely test data and training data, so it is necessary to know the optimal comparison ratio. At this stage, the comparison of training data and test data is tested. Testing is done using 5 compositions of data division ratios, the goal is to find out at what ratio of training data and test data the model can perform well. the composition of the division of training data and test data is 90:10, 80:20, 70:30, 60:40 and 50:50. This research tests using the TF-IDF weighting method and the Support Vector Machine method. When the results are out, the accuracy, precision, recall values of each ratio composition are calculated. The following results are obtained on the performance of Support Vector Machine (SVM) on each ratio shown in **Table *3***.

Table 3. Accuracy results rata ratio acquisition testing

Ratio	Accuracy Value
90:10	88.54 %
80:20	88.30%
70:30	89.10%

60:40	89.47%
50:50	89.33%

Term Frequency Inverse Document Frequency Weighting Result

After data cleaning is done in the preprocessing process, the data will be weighted using Term Frequency – Inverse Document Frequency (TF-IDF) weighting algorithm. TF-IDF weighting is built using a module from the Python programming language, Scikit-learn or can also be called Sklearn. TF-IDF calculation is done twice, the first process calculates the weight of the training data word with the input data which is the training. After that, it continues to calculate the weight of the test data word with the input data which is the test. The comparison between training data and test data is 60:40, where 60% is training data and the remaining 40% is test data. The training data used is 4,972 data, while the amount of test data is 3,315 data. TF-IDF weighting on training data produces a 5,800 x 6,772 dimension vector. TF-IDF weighting on test data produces a vector with a dimension of 2,487 x 6,772.

Classification Results Using The Support Vector Machine (SVM) Method

After the TF-IDF weighting results are obtained, the classification process is then carried out using the Support Vector Machine method which is built using a module from the Python programming language, Scikit-learn. In this study, two tests were carried out, namely comparing four support vector machine kernels and finding the optimal value of the C and gamma parameters.

The first test compares the results of four Support Vector Machine kernels namely linear, sigmoid, polynomial, and RBF. Then the test calculates the accuracy value of the sentiment classification. the accuracy value is shown in **Table *4***

Table 4. Kernel testing accuracy results

No	Kernel	Accuracy Value
1	RBF	89.47%
2	Sigmoid	89.32%
3	Linear	89.35%
4	Polynomial	89.26%

The next test aims to determine the optimal Support Vector Machine parameter values for the sentiment analysis process. In the RBF kernel, there are two parameters tested, namely the C value with parameter values of 0.1, 1, 10, 100, and 1000. The other parameter is gamma with tested parameter values of 1, 0.1, 0.01, 0.001, and 0.0001. The most optimal results are obtained when the C parameter is 10 and the gamma parameter is 0.1.

Based on these two tests, the sentiment analysis classification model of Instagram comments on the 2024 Indonesian presidential candidates was built using the best kernel, namely RBF with a C parameter of 10 and a gamma parameter of 0.1. To measure the performance of the model built, the model is evaluated with the confusion matrix algorithm shown in **Table *5***.

Table 5. Confusion matrix

		Predicted Class		
		Negative	**Netral**	**Positif**
Actual Class	**Negative**	37	23	145
	Netral	9	50	109
	Positif	37	28	2877

Based on the confusion matrix in **Table *5***, the precision, recall, accuracy, and F-Measure values can be calculated. The accuracy value obtained by the system is 89.41%, the recall value is 89%, the precision value is 87%, and the F-1 Score value is 89.41%.

IV. DISCUSSION

This section contains analysis of the testing process of the system built using the Support Vector Machine method. The testing process carried out are data ratio composition testing and Support Vector Machine classification method parameter testing.

Testing the ratio composition of test data and training data is done to find out whether the composition of the data ratio has an influence on accuracy results. There are 5 types of ratios tried, namely 90:10, 80:20, 70:30, 60:40, and 50:50. **Table *3*** shows that the lowest accuracy value is obtained when using the composition of the ratio of training data to test data of 80:20, which is 88.30%, while the highest accuracy value is obtained when using the composition of the ratio of training data to test data of 60:40, which is 89.47%. This shows that training data with a ratio of 60:40 produces a model that better represents the test data than using training data with a ratio of 80:20.

The next test is testing the parameters of the Support Vector Machine classification method. The Support Vector Machine method has several parameters that can be set in order to get the best classification results, namely the kernel function, C parameter value, and Gamma parameter value. By using the ratio of training data: test data of 60:40, the best accuracy is obtained when using the rbf kernel which is 89.47 &, while the lowest accuracy value is obtained when using the polynomial kernel function. This means that this data is more suitable to use the rbf kernel function than the polynomial kernel function. In addition to testing to get the best kernel function parameters, this study also tested to get the best C parameter [7]. C parameter values tested were 0.1, 1, 10, 100, and 1000. The most optimal C parameter value is 10. C parameter testing is accompanied by Gamma parameter testing. The parameter values tested are 1, 0.1, 0.01, 0.001, and 0.0001. gamma parameters that produce the most optimal value when gamma is 0.1.

The best part of the support vector machine method is the parameter values that represent the opinion of Instagram comments for the Indonesian presidential candidate 2024. How to create test data 60-40 kernel functions are used. the value of the parameter rbf c. 10 and the value of the gamma parameter is 0.1. The classification system by setting the optimal parameters produces a confusion matrix shown in **Table *5***.

Based on **Table *5***, it can be seen the accuracy of the system in predicting the sentiment of Instagram comments on the 2024 Indonesian presidential candidates. The data that is actually labelled positive and predicted positive is 2,877 data, while the amount of data the system mis predicts, which should be positive but is considered negative there are 37 data and 28 data that are predicted to be neutral. In the data that is actually negative and predicted negative there are 37 data, while the amount of data the system mis predicts, which should be negative but is considered positive there are 245 data and the predicted neutral is 23 data. In the data that is actually neutral and predicted correctly, namely labelled neutral, there are 50 data, while the amount of data the system mis predicts, which should be neutral but is considered positive is 109 data and the predicted negative is 9 data.

In general, the system's ability to classify the sentiment of Instagram comments on Indonesian presidential candidates 2024 using the Support Vector Machine method is generally quite good with an accuracy value of 89.41%. This is in line with the recall and precision values obtained which are quite high, namely 89% and 87%. The results of recall and precision are reinforced by the high harmonic mean value of precision and recall, which is 89.41%.

The high values of precision and recall mean that the accuracy of sentiment analysis is quite good and there is minimal discrepancy in the information obtained from the experiments that have been carried out. However, the high performance when classifying in general is only due to the accuracy of classification of one class label. The highest accuracy value is obtained in the positive labelled class which is 97.79%, the classification in the neutral class produces a low accuracy of 29.76%, and the lowest is obtained in the data with the negative label class which is 18.05%. The low accuracy value in both label classes could be due to overfitting as a result of the uneven distribution of data shown in **Figure 3**.

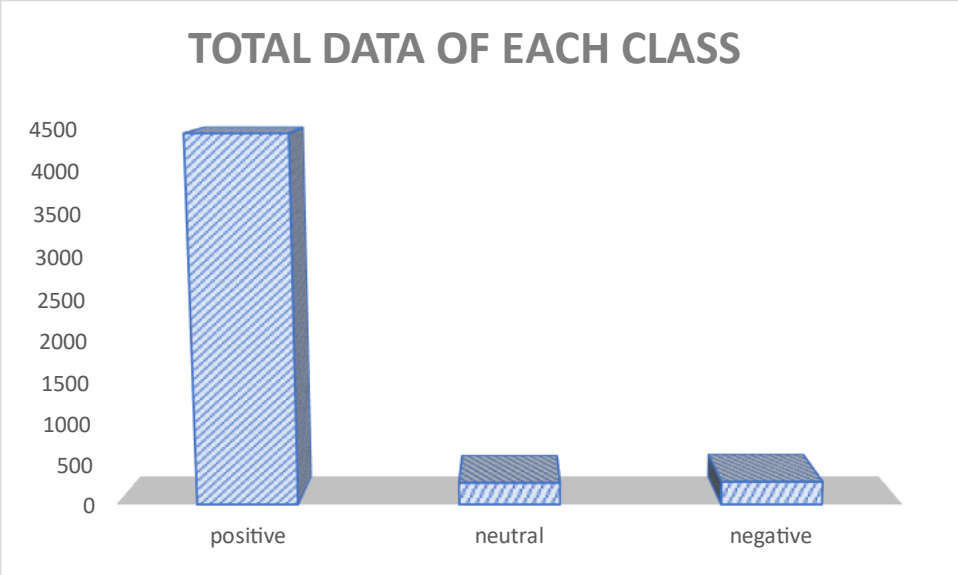

Figure 3 Total data of each class

Figure 3 shows that out of 4,972 training data, the most data is data labelled as positive class, which is as much as 4,422 data, while data labelled as neutral is only 269 data, and the least data is in the class labelled as negative, which is only 281 data.

V. CONCLUSION

Based on the results of the research, sentiment analysis of Instagram comments on Indonesian presidential candidates 2024 using the Support Vector Machine method gets a good accuracy value of 89.41%. However, if the research is seen in more detail again, good accuracy is only produced when the data is labelled with a positive class of 97.79%, while the data labelled with a neutral class only produces an accuracy of 29.76%, and the lowest is in data with a negative label class which only gets an accuracy of 18.05%.

REFERENCES

[1] M. Rachmansyah and L. P. Supratman, "Peran Media Instagram Dalam Memasarkan Produk Fashion Dollies the Role of Instagram in Marketing Fashion Product Dollies," *J. Stud. Komun. Dan Media*, vol. 24, no. 1, pp. 73–90, 2020.

[2] M. Fadly and A. Wantoro, "Model Sistem Informasi Manajemen Hubungan Pelanggan Dengan Kombinasi Pengelolaan Digital Asset Untuk Meningkatkan Jumlah Pelanggan," *Pros. Semin. Nas.*, pp. 46–55, 2019, [Online]. Available: https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/1749.

[3] R. Khrishananto and M. A. Adriansyah, "Pengaruh Intensitas Penggunaan Media Sosial Instagram dan Konformitas Terhadap Perilaku Konsumtif di Kalangan Generasi Z," *Psikoborneo J. Ilm. Psikol.*, vol. 9, no. 2, p. 323, 2021, doi: 10.30872/psikoborneo.v9i2.5973.

[4] M. F. Asshiddiqi and K. M. Lhaksmana, "Perbandingan Metode Decision Tree Dan Support Vector Machine Untuk Analisis Sentimen Pada Instagram Mengenai Kinerja Pssi," *eProceedings Eng.*, vol. 7, no. 3, pp. 9936–9948, 2020, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/14217.

[5] S. Redjeki and S. Widyarto, "Comparison of Seven Machine Learning Algorithms in the Classification of Public Opinion," *Tech-E*, vol. 5, no. 2, pp. 143–149, 2022, doi: 10.31253/te.v5i1.1046.

[6] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.

[7] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisi Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, Aug. 2020, doi: 10.14710/j.gauss.v9i3.28932.

[8] Ratino, N. Hafidz, S. Anggraeni, and W. Gata, "Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes," *J. Penelit. Ilmu dan Teknol. Komput.*, vol. 12, no. 2, pp. 1–11, 2020, [Online]. Available: https://jurnal.polsri.ac.id/index.php/jupiter/article/view/2388.

[9] Ezenwobodo and S. Samuel, "International Journal of Research Publication and Reviews," *Int. J. Res. Publ. Rev.*, vol. 04, no. 01, pp. 1806–1812, 2022, doi: 10.55248/gengpi.2023.4149.

[10] M. F. Al-shufi and A. Erfina, "Sentimen Analisis Mengenai Aplikasi Streaming Film Menggunakan Algoritma Support Vector Machine Di Play Store," *Sismatik*, pp. 156–162, 2021.

[11] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.

[12] N. Hendrastuty *et al.*, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, [Online]. Available: http://situs.com.

[13] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.

[14] A. P. Widyassari, E. Noersasongko, A. Syukur, and Affandy, "The 7-Phases Preprocessing Based On Extractive Text Summarization," in *2022 7th International Conference on Informatics and Computing, ICIC 2022*, Dec. 2022, pp. 1–8, doi:

10.1109/ICIC56845.2022.10006998.

[15] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 123–131, 2021, doi: 10.29207/resti.v5i1.2587.

[16] V. Amrizal, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.

[17] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," *Decis. Anal. J.*, vol. 3, no. May, p. 100071, 2022, doi: 10.1016/j.dajour.2022.100071.

[18] X. Wu and V. Kumar, *The Top Ten Algorithms in Data Mining*. Chapman and Hall/CRC, 2009.

[19] I. Markoulidakis, G. Kopsiaftis, I. Rallis, and I. Georgoulas, "Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem," *ACM Int. Conf. Proceeding Ser.*, no. Cx, pp. 412–419, 2021, doi: 10.1145/3453892.3461323.

[20] M. Desai and M. A. Mehta, "Techniques for sentiment analysis of Twitter data: A comprehensive survey," in *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, 2017, pp. 149–154, doi: 10.1109/CCAA.2016.7813707.