JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i2.5058

ENHANCING SENTIMENT ANALYSIS ACCURACY IN DELIVERY SERVICES THROUGH SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE

Setia Sri Anggraeni[1]; Septi Andryana[2]*

Department of Information Technology[1,2]
Universitas Nasional[1,2]
https://unas.ac.id/[1,2]
tiaanggraeni2@gmail.com[1], septi.andryana@civitas.unas.ac.id[2]*

(*) Corresponding Author

Abstract—*Paxel is one of the delivery services that use the application. On Google Play, there are more than 10 thousand users leaving reviews. From this review data, a sentiment analysis was then carried out to determine the level of user satisfaction with Paxel's services. The methods used in this study are Random Forest (RF) and Support Vector Machine (SVM), as well as applying Synthetic Minority Oversampling Technique (SMOTE) to overcome data imbalance. The results showed that the method testing by dividing the data into two, namely training data and testing data by 80:20, stated that by applying the SMOTE, a higher accuracy value was obtained, where the accuracy of the RF method reached 91%, and the SVM method reached 87%. The level of user satisfaction with Paxel services tends to be neutral. This can be seen in the classification of the RF method with F1-Score values for the Positive class 89%, Neutral class 93%, and Negative class 92%.*

Keywords: *random forest, sentiment analysis, smote, SVM.*

Intisari—*Paxel merupakan salah satu jasa pengiriman yang menggunakan aplikasi. Pada Google Play tercatat lebih dari 10 ribu pengguna memberikan ulasan. Dari data ulasan ini kemudian dilakukan analisis sentimen untuk mengetahui bagaimana tingkat kepuasan pengguna terhadap layanan Paxel. Metode yang digunakan dalam penelitian ini adalah Random Forest (RF) dan Support Vector Machine (SVM), serta menerapkan Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan data. Hasilnya menunjukkan bahwa pengujian metode dengan membagi data menjadi dua yaitu data training dan data testing sebesar 80:20, menyatakan bahwa dengan menerapkan SMOTE didapatkan nilai akurasi yang lebih tinggi, dimana akurasi metode RF mencapai 91%, dan metode SVM mencapai 87%. Tingkat kepuasan pengguna terhadap layanan Paxel cenderung netral. Hal ini dapat dilihat pada klasifikasi metode RF dengan nilai F1-Score untuk class Positive 89%, class Neutral 93% dan class Negative 92%.*

Kata Kunci: *random forest, analisis sentimen, smote, SVM.*

INTRODUCTION

Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) was surveyed in 2023 and stated that internet users in Indonesia reached 79.19%. Otoritas Jasa Keuangan (OJK) also noted that 88.1% of internet users had made online purchases. Moreover, since COVID-19 and the implementation of Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), people have been shopping more often, since 2020 there have been 12 million new users shopping online [1].

The increase in people's style of buying and selling transactions online has also greatly influenced the increase in the number of goods that must be sent by delivery service providers. With so many packages that must be sent, this also affects the quality of the delivery service provider [2]. For this reason, delivery service providers compete to provide the best service, one of which is creating applications to make it easier for users in the process of ordering package deliveries.

Paxel is a delivery service that uses an application. On Google Play, it is recorded that it has reached more than one million downloads and has received more than 10,000 reviews.

Sentiment analysis is a field of Natural Language Processing (NLP) that analyzes an opinion or information in text form [3]. Starting from unstructured data, then transformed into more structured data to produce opinion classes, such as positive class, negative class, or neutral class [4].

In recent years, many fields have used sentiment analysis, starting from market research [5][6], customer feedback [7], education [8], finance

[9], brand monitoring, customer service, and many more.

C. Villavicencio, in his sentiment analysis, used data from Twitter to find out how the Philippines people reacted to the COVID-19 vaccine. The sentiments were trained using the Naïve Bayes method to classify the English and Filipino languages into positive, neutral, and negative classes. It produced an accuracy value of 81.77%, where 83% of the Philippines people were enthusiastic and responded positively to the vaccine [10].

Hermanto, in his research on user sentiment analysis for the Gojek and Grab applications, used the Naïve Bayes and Support Vector Machine method combined with the Smote technique to produce higher accuracy values. For the Gojek application, the highest accuracy result was 81.09%, and the AUC value = 0.922. Also, for the Grab application, the accuracy value of 73.20% and AUC value = 0.848 [11].

D. Sudigyo, uses text mining to create a pipeline that is expected to help researchers accelerate the development of stunting supplementation intervention research in Indonesia [12]. W. L. M. Leurs also used text-mining to look for characteristics of text written by nurses related to "falls", the results stated that nurses' notes could help to find words that were often used by patients who had experienced falls, making it easier to identify people. elderly people are at high risk of falls [13].

So, from Paxel's review data taken from Google Play, sentiment analysis will be carried out to find out the level of user satisfaction with Paxel services. Also, to prove that SMOTE can increase accuracy values, this research will use a method that is different from previous research, which is Random Forest (RF) and Support Vector Machine (SVM) methods.

MATERIALS AND METHODS

The flow of research carried out in this research is shown in Figure 1.

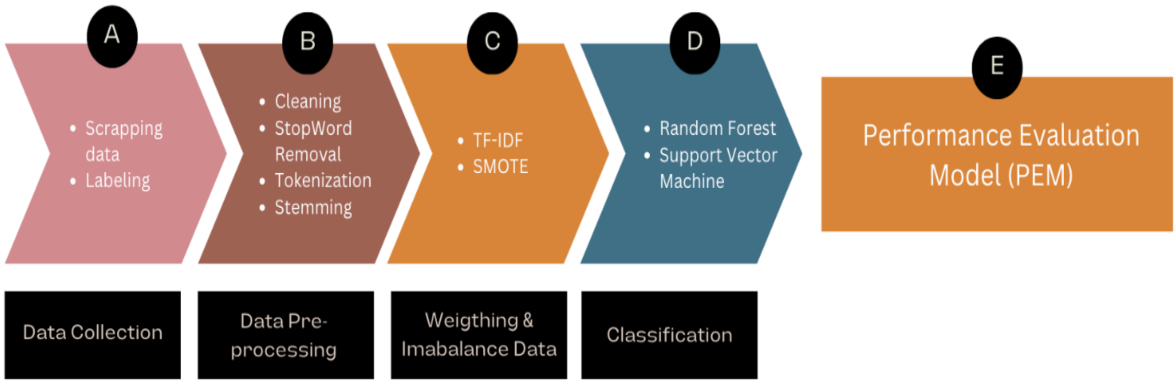

Figure 1. Research flow

Figure 1 shows the flow of this research, starting from data collection, pre-processing, and word weighting, followed by classification using the RF [14], SVM [15] method, and model testing to obtain accuracy values.

A. Data Collection

Data was taken by scraping Paxel reviews on Google Play using Google Play Scraper version 1.2.4 and data was obtained for 12,012 reviews since 2017.

Then, the data is grouped based on the review score, where a score of 1-2 is included in the Negative Class category, 3 is included in the Neutral Class and 4-5 is included in the Positive Class.

B. Data Pre-processing

In this stage, the process of cleaning unnecessary and meaningless data is carried out. Starting from cleaning means deleting hashtags, punctuation, URLs, extra whitespace, and next characters, and changing the text to lowercase. The next step is StopWords Removal where words that have no meaning will be removed such as "di", "dan", "ke", etc. Next, tokenization is carried out, which means changing the sentence into parts of words where white spaces become separators which are then called tokens. The final step in the data pre-processing stage is stemming, which means changing words into their basic form, such as "membantu" which becomes "bantu". Since the

VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i2.5058

data is in Bahasa, for the stemming process use the Sastrawi library.

C. Weighting & Imbalance Data

The word weighting technique used in this research is the Term Frequency-Inverse Document Frequency (TF-IDF) which is responsible for determining the frequency of a word that appears in the data [16][17]. Imbalanced data causes misclassification in the minority class [18], so data balancing is needed. In this research, the Synthetic Minority Over-sampling Technique (SMOTE) method was used to overcome unbalanced data by re-sampling the minority class, so that the proportion of data becomes more balanced.

D. Classification

Once the data is clean, it can be classified using the RF and SVM methods. The classification process requires test data and training data, for this they need to be declared. In this study, 80:20 training data and test data were used which were taken randomly.

E. Performance Evaluation Model (PEM)

After the classification is complete, testing can be carried out on each method used. The results show accuracy, precision, recall, F1-Score, and confusion matrix values.

RESULTS AND DISCUSSION

The data used in this research were 12,012 Paxel reviews from Google Play since 2017. After the data is cleaned, label them into positive, negative, and neutral classes. Here is a word cloud that can be displayed.

Figure 2. WordCloud of positive class

Figure 2 is a WordCloud from the positive class where the most frequently mentioned words are *"paxel", "kirim", "bantu", "bagus"*, and *"cepat".*

Figure 3. WordCloud of negative class

Figure 3 is a WordCloud from the negative class where the most frequently mentioned words are *"aplikasi", "kirim", "paket", "kurir"*, and *"kecewa"*.

Figure 4. WordCloud of neutral class

Figure 4 is a WordCloud from the neutral class where the most frequently mentioned words are *"paxel", "kirim", "aplikasi", "update"*, and *"paket".*

During the classification process, the data is divided into two, training data and testing data by 80:20. Testing was carried out with a comparison between applying the SMOTE and without applying the SMOTE.

Table 1. RF accuracy value without the SMOTE

	Precision	Recall	F1-Score	Support
Negatif	0.72	0.62	0.67	498
Neutral	0.33	0.03	0.06	120
Positif	0.87	0.96	0.91	1785
Accuracy			0.84	2403
Macro avg	0.64	0.54	0.55	2403
Weighted avg	0.81	0.84	0.82	2403

As shown in Tabel 1, the accuracy for the RF method without applying the SMOTE is 84%.

VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v9i2.5058

Table 2. RF accuracy value with the SMOTE

	Precision	Recall	F1-Score	Support
Negatif	0.91	0.92	0.92	1843
Neutral	0.93	0.93	0.93	1801
Positif	0.90	0.89	0.89	1722
Accuracy			0.91	5366
Macro avg	0.91	0.91	0.91	5366
Weighted avg	0.91	0.91	0.91	5366

In Table 2 by combining the RF with the SMOTE, the accuracy reached 91%.

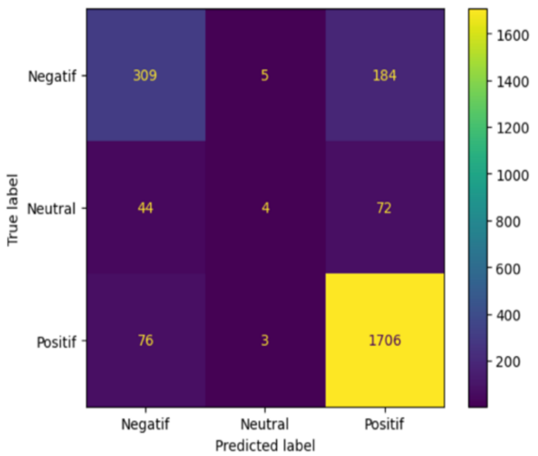

Figure 5 Confusion matrix RF method without the SMOTE

Figure 5 is the confusion matrix of the RF method without the SMOTE. It shows that for the positive class, the True Positive is 1706, False Positive is 79. For the neutral class, the True Neutral is 4, False Neutral is 116 and for the negative class, the True Negative is 309, False Negative 189.

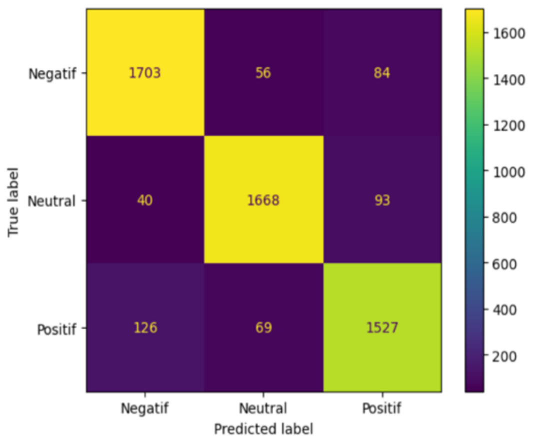

Figure 6. Confusion matrix RF method with SMOTE

Figure 6 is the confusion matrix of the RF method by combining the SMOTE. It shows that for the positive class, the True Positive is 1527, False Positive is 195. For the neutral class, the True Neutral is 1668, False Neutral is 133 and for the negative class the True Negative is 1703, False Negative is 493.

Table 3. SVM accuracy value without SMOTE

	Precision	Recall	F1-Score	Support
Negatif	0.75	0.67	0.70	498
Neutral	0.00	0.00	0.00	120
Positif	0.88	0.96	0.92	1785
Accuracy			0.85	2403
Macro avg	0.54	0.54	0.84	2403
Weighted avg	0.81	0.85	0.83	2403

Table 3 The accuracy for the SVM method without applying the SMOTE is 85%.

Table 4. SVM accuracy value with the SMOTE

	Precision	Recall	F1-Score	Support
Negatif	0.96	0.78	0.86	1843
Neutral	0.77	0.98	0.86	1801
Positif	0.95	0.86	0.90	1722
Accuracy			0.87	5366
Macro avg	0.89	0.87	0.87	5366
Weighted avg	0.89	0.87	0.87	5366

As shown in Table 4. by combining the SVM with the SMOTE, the accuracy reached 87%.

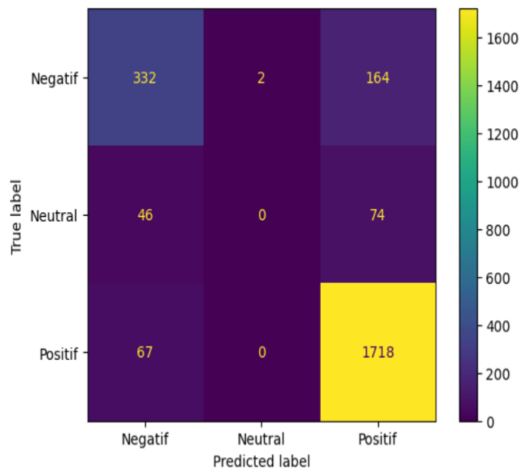

Figure 7. Confusion matrix SVM method without the SMOTE

Figure 7 is the confusion matrix of the SVM method without the SMOTE. It shows that for the positive class, the True Positive is 1718, False Positive is 67. For the neutral class, the True Neutral is 0, the False Neutral is 120, and for the negative class, the True Negative is 332, False Negative 166.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)
VOL. 9. NO. 2 FEBRUARY 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i2.5058

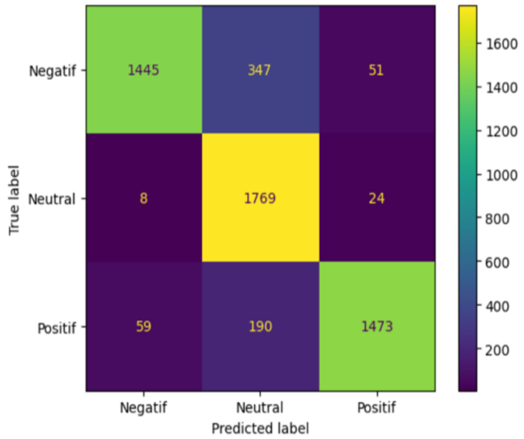

Figure 8. Confusion matrix SVM method with the SMOTE

Figure 8 is the confusion matrix of the SVM method by combining the SMOTE. It shows that for the positive class, the True Positive is 1473, False Positive is 249. For the neutral class, the True Neutral is 1769, False Neutral is 32, and for the negative class the True Negative is 1445, False Negative is 398.

Table 5. Method testing results without the SMOTE

Method		Positive	Neutral	Negative
RM	Precision	0.87	0.33	0.72
	Recall	0.96	0.03	0.62
	F1-Score	0.91	0.06	0.67
SVM	Precision	0.88	0.00	0.75
	Recall	0.96	0.00	0.67
	F1-Score	0.92	0.00	0.70

In Table 5, the results of the testing method without applying the SMOTE show that the RF methods on the F1-Score value for the Positive class are 91%, the Neutral class is 6% and the Negative class is 67%. For the Support Vector Machine method, the F1-Score value for the Positive class is 92%, the Neutral class is 0% and the Negative class is 70%.

Table 6. Method testing results with SMOTE

Method		Positive	Neutral	Negative
RM	Precision	0.90	0.93	0.91
	Recall	0.89	0.93	0.92
	F1-Score	0.89	0.93	0.92
SVM	Precision	0.95	0.77	0.96
	Recall	0.86	0.98	0.78
	F1-Score	0.90	0.86	0.86

In Table 6, the results of the testing method using the SMOTE show that the RF methods on the F1-Score value for the Positive class are 89%, the Neutral class is 93% and the Negative class is 92%. For the SVM method, the F1-Score value for the Positive class is 89%, the Neutral class is 93% and the Negative class is 92%.

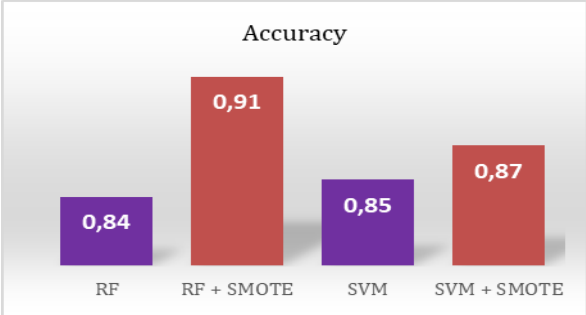

Figure 9. SMOTE implementation comparison

As shown in Figure 9, we can see the differences when applying the SMOTE and without applying the SMOTE. The RF method testing without applying the SMOTE obtained an accuracy value of 84%, while by applying the SMOTE the accuracy value reached 91%. Likewise, with SVM testing without the SMOTE the accuracy value was 85% while applying the SMOTE it reached 87%.

CONCLUSION

Sentiment analysis using 12,012 Paxel review data from Google Play using the RF and SVM methods as well as applying the SMOTE has been carried out. The results show that by applying the SMOTE a higher accuracy value is obtained, where the accuracy of the RF method reaches 91% and the accuracy of the SVM method is 87%. From the results of testing the RF method, it was also found that users were neutral with Paxel's services, which can be seen from the F1-Score value in the Positive class of 89%, the Neutral class of 93%, and the Negative class of 92%. However, the accuracy results may be different if classification is carried out with different data or methods. For future research, carry out the classification using the same or different method and with different data to validate the accuracy value.

REFERENCE

[1] R. Nariswari, T. S. Prakoso, N. Hafiz, and H. Pudjihastuti, "Biplot analysis: a study of the change of customer behavior on e-commerce," *Procedia Comput Sci*, vol. 216, pp. 524–530, 2023, doi: 10.1016/j.procs.2022.12.165.

[2] M. Kaplan, S. Hotle, and K. Heaslip, "How has COVID-19 impacted customer perceptions and demand for delivery services: An exploratory analysis," *Transp Policy (Oxf)*, vol.

134, pp. 217–230, Apr. 2023, doi: 10.1016/j.tranpol.2023.02.020.

[3] F. J. Damanik and D. B. Setyohadi, "Analysis of public sentiment about COVID-19 in Indonesia on Twitter using multinomial naive Bayes and support vector machine," in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing Ltd, Apr. 2021. doi: 10.1088/1755-1315/704/1/012027.

[4] W. F. Satrya, R. Aprilliyani, and E. H. Yossy, "Sentiment analysis of Indonesian police chief using multi-level ensemble model," *Procedia Comput Sci*, vol. 216, pp. 620–629, 2023, doi: 10.1016/j.procs.2022.12.177.

[5] E. Hirata and T. Matsuda, "Examining logistics developments in post-pandemic Japan through sentiment analysis of Twitter data," *Asian Transport Studies*, vol. 9, Jan. 2023, doi: 10.1016/j.eastsj.2023.100110.

[6] S. Andryana, B. Rahman, A. Gunaryati, "Predicting Market Segments from Twitter Data Using Arima Time Series Analysis" *JITK*, vol. 9, August. 2023, doi: DOI: 10.33480/jitk.v9i1.4275.

[7] K. M. Hasib, M. A. Habib, N. A. Towhid, and M. I. H. Showrov, "A Novel Deep Learning based Sentiment Analysis of Twitter Data for US Airline Service," in *2021 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD 2021 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Feb. 2021, pp. 450–455. doi: 10.1109/ICICT4SD50815.2021.9396879.

[8] I. Spada, F. Chiarello, S. Barandoni, G. Ruggi, A. Martini, G. Fantoni, "Are universities ready to deliver digital skills and competences? A text mining-based case study of marketing courses in Italy", Technological Forecasting and Social Change, Volume 182, 2022, 121869, ISSN 0040-1625, doi:10.1016/j.techfore.2022.121869.

[9] Y. Hai Hoang, V. Minh Ngo, N. Bich Vu, "Central bank digital currency: A systematic literature review using text mining approach", Research in International Business and Finance, Volume 64, 2023, 101889, ISSN 0275-5319,doi:10.1016/j.ribaf.2023.101889.

[10] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Information (Switzerland)*, vol. 12, no. 5, May 2021, doi: 10.3390/info12050204.

[11] Hermanto, A. Y. Kuntoro, T. Asra, E. B. Pratama, L. Effendi, and R. Ocanitra, "Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012102.

[12] D. Sudigyo, A. A. Hidayat, R. Nirwantono, R. Rahutomo, J. P. Trinugroho, and B. Pardamean, "Literature study of stunting supplementation in Indonesian utilizing text mining approach," *Procedia Comput Sci*, vol. 216, pp. 722–729, 2023, doi: 10.1016/j.procs.2022.12.189.

[13] W. L. M. Leurs, L. A. S. Lammers, W. N. Compagner, M. Groeneveld, E. H. H. M. Korsten, and C. M. J. van der Linden, "Text mining in nursing notes for text characteristics associated with in-hospital falls in older adults: A case-control pilot study," *Aging Health Res*, vol. 2, no. 2, p. 100078, Jun. 2022, doi: 10.1016/j.ahr.2022.100078.

[14] H. Malik, E. M. Shakshuki, and A. U. H. Yasar, "Approximating Viewership of Streaming T.V Programs Using Social Media Sentiment Analysis," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 94–101. doi: 10.1016/j.procs.2021.12.215.

[15] D. Wang and Y. Zhao, "Using News to Predict Investor Sentiment: Based on SVM Model," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 191–199. doi: 10.1016/j.procs.2020.06.074.

[16] M. Demircan, A. Seller, F. Abut, and M. F. Akay, "Developing Turkish sentiment analysis models using machine learning and e-commerce data," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 202–207, Jun. 2021, doi: 10.1016/j.ijcce.2021.11.003.

[17] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst Appl*, vol. 212, Feb. 2023, doi: 10.1016/j.eswa.2022.118715.

[18] N. Putu Yulika Trisna Wijayanti1, E. N Kencana, I Wayan Sumarjaya, "SMOTE: Potensi dan Kekurangannya Pada Survei", *E-Jurnal Matematika Vol. 10(4), November 2021*, DOI:https://doi.org/10.24843/MTK.2021.v10.i04.p348.