

Implementation of Ensemble Machine Learning Classifier and Synthetic Minority Oversampling Technique for Sentiment Analysis of Sustainable Development Goals in Indonesia

Acep Irham Gufroni^a, Irani Hoeronis^b, Andi Nur Rachman^{a,*}, Cecep Muhamad Sidik Ramdani^a,
Nur Fajar^b, Heni Sulastri^a

^a Department of Information System Universitas Siliwangi, Tawang, Tasikmalaya, 46115, Indonesia

^b Department of Informatics Universitas Siliwangi, Tawang, Tasikmalaya, 46115, Indonesia

Corresponding author: *andy.rachman@unsil.ac.id

Abstract— As part of the Sustainable Development Goals (SDGs), governments worldwide have committed to improving people's lives to improve the quality of life for all, including the 17 such goals that were agreed upon in 2015 to benefit the human race as a whole. It would be interesting to see how society responds to the SDGs after approximately half of them have been achieved. This public response was analyzed in terms of sentiment. Within the total number of internet users in Indonesia, there are 18.45 million Twitter users. The platform enables anyone to write about anything they are experiencing in their lives, such as what is happening in their environment, what is happening in their education system, what is happening in the food industry, how people feel, and many more. The platform enables anyone to write about anything they are experiencing in their lives, such as what is happening in their environment, what is happening in their education system, what is happening in the food industry, how people feel, and many more. To model the data collected, the researchers used Ensemble Machine Learning Classifiers (EMLC) to model the data by using a machine learning classifier that uses machine learning techniques. The best model in this study is EMLC-Stacking with a data splitting of 80:20 and using SMOTE, which obtains an accuracy of 91%. This accuracy results from a 5% increase compared to when not using SMOTE. From 15,698 tweets, this research found that 47% were positive sentiments, 28% were negative sentiments, and 25% were neutral sentiments. The results that we measured offer hope that there will be a positive trend in the journey of the SDGs until 2030 if these findings are true.

Keywords— Data Science; machine learning; SDGs; sentiment analysis; SMOTE.

Manuscript received 6 Jul. 2023; revised 23 Oct. 2023; accepted 27 Nov. 2023. Date of publication 31 May. 2024.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

The agenda for the Sustainable Development Goals (SDGs), which the United Nations General Assembly endorsed in September 2015 as a new universal development agenda [1], has superseded the framework for the Millennium Development Goals (MDGs), which was agreed upon in the year 2000. Figure 1 illustrates the seventeen Sustainable Development Goals set for the world. SDGs have Universal, Integrative, and Inclusive principles to ensure that no one is left behind or is called No One Left Behind.



Fig. 1 Sustainable Development Goals

SDGs are global and national commitments to improve the welfare of society, covering 17 goals [2] as follows:

- No Poverty.
- Zero Hunger.
- Good Health and Well-Being.
- Quality Education.
- Gender Equality.
- Clean Water and Sanitation.
- Affordable and Clean Energy.
- Decent Work and Economic Growth.
- Industry, Innovation, and Infrastructure.
- Reduced Inequalities.
- Sustainable.
- Cities and Communities.
- Responsible Consumption and Production.
- Climate Action.
- Life below Water.
- Life on Land.
- Peace, Justice, and Strong Institutions.
- Partnership for The Goals.

The Fundamental 5P of SDGs can be seen in Figure 2.

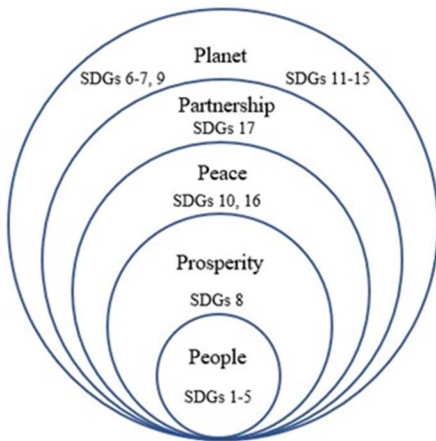


Fig. 2 Fundamental 5P of SDGs

The SDGs contain five fundamental principles (Figure 2) that strike a balance between economic, social, and environmental dimensions: 1) People, 2) Planet, 3) Prosperity, 4) Peace, and 5) Partnership. These five fundamental principles are known as the 5P [1] and encompass 17 Goals and 169 Goals that cannot be separated, connected, or integrated to improve the quality of human existence.

In recent years, social media has evolved as a virtual place that allows users to express their worries about the environment and health in addition to other problems of public interest such as transportation [3], health [4], [5], education [6], [7], communication [7], and economics [8][9]. The usage of social media to carry out day-to-day activities is highly prevalent all over the world. Social media platforms are an essential source of information regarding the communication and viewpoints of internet users [10].

Twitter users can produce and express their ideas directly with one another through the usage of tweets and tweets. Users can also engage with one another through the platform, which may be done by following other users, viewing other users' accounts, and using the same hashtags to discuss topics of interest to both parties. Following that, Twitter was utilized more frequently for research objectives[11], including the investigation of public opinion[3]. Twitter's growth statistics

are favorable, and the number of active users is growing every month [12], in contrast to the growth statistics of other social media platforms such as Facebook.

As shown in Figure 3, people can write about anything happening in their lives on Twitter, which is one reason why there are 18.45 million users of Twitter among all internet users in Indonesia. Therefore, when researching sentiment analysis, Twitter is a significant social network.

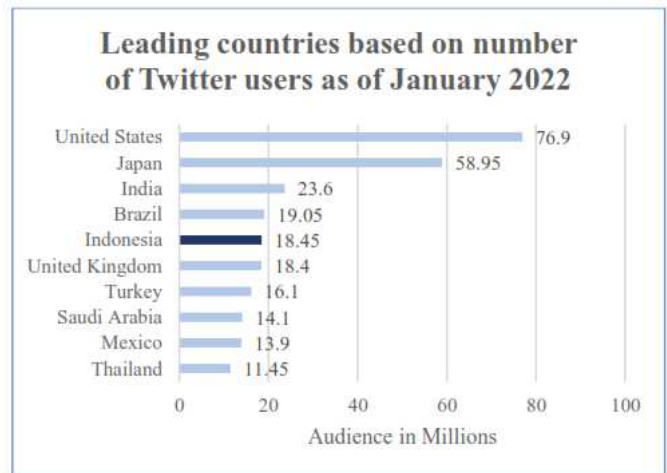


Fig. 3 Country Ranking by Twitter Users

The originality of this study is demonstrated by sentiment analysis (identifying positive, neutral, and negative connotations) of all 17 goals that make up the Sustainable Development Goals (SDGs). RapidMiner and the Python programming language will be the technologies utilized for this project. The Foundational Methodology for Data Science (FMDS), in general, was used throughout the process of carrying out this research. As a result, when we finish our study, we have an idea of the precision of the utilized models, the impact of using SMOTE, and the total number of positive, neutral, and negative reactions to the Sustainable Development Goals in Indonesia.

II. MATERIALS AND METHOD

The field of study known as sentiment analysis examines individuals' thoughts, feelings, and perspectives through written language [13]. In sentiment analysis, the first step is to classify the text into sentences or documents, and the second step is to decide if the opinions represented in those phrases or documents are positive, negative, or neutral. Emotions such as happiness, sadness, or rage can also be gleaned via sentiment analysis. Expression of sentiment refers to the focus of a specific issue; hence, a remark regarding one subject may have a different meaning than a statement regarding another subject.

Based on literature studies, most of the previous studies were only related to one SDGs point. SDGs 1 [14], [15]; SDGs 2 [16], [17]; SDGs 3 [4], [18], [19],[20], [21], [22]; SDGs 4 [6], [23], [24], [25], [26]; SDGs 5 [27], [28]; SDGs 7 [29], [30]; SDGs 8 [3], [8], [31], [32]; SDGs 9 [33]; SDGs 10 [34], [35]; SDGs 11. [36]; SDGs 12 [37], [38]; SDGs 13 [39], [40]; SDGs 14 [41], [42]; SDGs 16 [43]; and SDGs 17[44].

Flores et al. [24] used SMOTE to get higher accuracy, concluding that using SMOTE increased accuracy in using the

K-12 program dataset in the Philippines. Jayapermana et al. [20] used the Ensemble Machine Learning Classifier to research the COVID-19 vaccine. This resulted in the conclusion that using the stacking ensemble classifier obtained higher accuracy than using only a single algorithm. Based on related research, this study showed the effect of using SMOTE and the Ensemble Machine Learning Classifier on the accuracy of sentiment analysis of Twitter data related to the overall SDGs in Indonesia. This research uses the Foundational Methodology for Data Science[45]. This methodology is a development of the CRISP-DM framework. Figure 4 is an FMDS scheme.

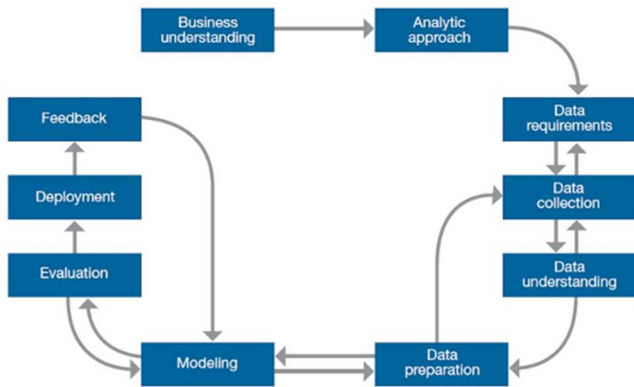


Fig. 4 Foundational Methodology for Data Science

A. Business Understanding

Every project or research begins with a business understanding that forms the basis for an effective solution to a business problem. This phase aims to answer the question, "What problem are you trying to find a solution to?" In this phase, the research begins by defining the problem and determining the objectives and scope of the study.

B. Analytical Approach

After the business problem is clearly understood, this research is continued by determining the approach to solving the problem. This phase aims to answer the question, "How can researchers use existing data to solve problems?" This stage requires posing the problem in the context of machine learning. This research used an approach that implemented the Ensemble Machine Learning Classifier and SMOTE to analyze sentiment toward SDGs in Indonesia.

C. Data Requirements

The analytical approach chosen determines the data requirements. This stage aims to answer the question, "What data is needed to solve your problem?" In particular, the analytical method requires specific content, format, and data representation guided by domain knowledge. The data needed for this research is from Indonesian-language tweets regarding SDGs originating from Twitter users in Indonesia.

D. Data Collection

At this stage, the research began to collect data. This stage aims to answer the question, "Where does the data come from, and how do you get it?" The data used in this study is a collection of Indonesian language tweets regarding the 17 SDGs on 10-16 July 2022. The data was obtained from

Twitter using the Twitter API (Application Programming Interface) via RapidMiner.

E. Data Understanding

After data collection is reviewed, determine whether it is by business and data needs. This stage aims to answer the question, "Does the data obtained meet the needs to solve the problem?" At this stage, the researchers ensured that the data obtained was as needed and understood its contents and initial information.

F. Data Preparation

This stage includes all activities to construct a dataset in the modeling stage. The Research Phase aims to answer the question, "What should the researchers do to prepare the data so they can continue modeling using existing data?" This phase has two stages (Figure 5): the first is to clean the data, and the second is to label the data in the modeling stage.

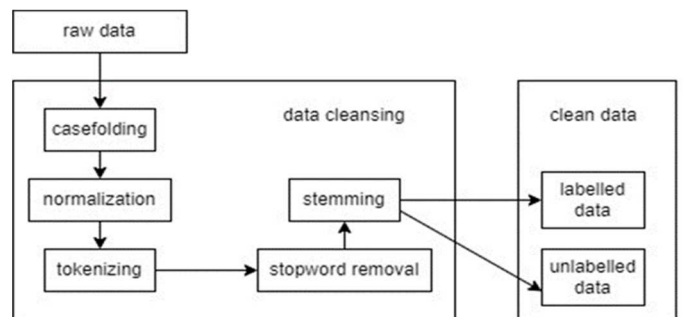


Fig. 5 Data Preparation Scheme

G. Modelling

The modeling phase focuses on developing a predictive model using a predetermined analytical approach. At this stage, the researchers conducted several experiments and combinations of algorithms with their respective parameters to find the best model for the next phase. This phase aims to answer the question, "How can the data be visualized to get the required solution?"

This study compared the performance of the Ensemble Machine Learning Classifier - Stacking and Voting. EMLC Stacking used the Naïve Bayes (NB) algorithm, Support Vector Machine (SVM), and k-nearest Neighbors (k-NN) as first learners and the k-NN algorithm as metal learners. Meanwhile, EMLC Voting used the NB, SVM, and k-NN algorithms as the learner. The split data used a ratio between 70:30 and 80:20 because, empirically, it is the best range for splitting data [56]. Research also examines algorithm performance and the effect of using SMOTE on existing datasets. Table 1 shows the model scheme used in this study.

TABLE I
MODELS SCHEME

Models	Without SMOTE		With SMOTE	
	70:30	80:20	70:30	80:20
EMLC- Stacking	A	B	C	D
EMLC-Voting	E	F	G	H

H. Evaluation

At this stage, the researchers evaluated the quality of the model based on the accuracy obtained at the previous stage. This phase aims to answer the question, "Does the model

obtained really answer the initial question or does it need to be readjusted?" At this stage, the researchers evaluated the best model using the confusion matrix with several assessments, including accuracy, precision, recall, and f1-score obtained from the best confusion matrix model. The following is the formula used for model evaluation.

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

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

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

I. Deployment

At this stage, the researchers implement the best model obtained in the previous stage for data that does not yet have a label (data that has passed the data preparation phase). At this stage, it has output as prediction results for all data processed in the previous stage.

J. Feedback

At this stage, the researchers conclude the previous stages and provide feedback based on the word cloud of each SDG. The word cloud here was taken based on the SDGs results in the data preparation stage. Based on the word cloud, researchers can provide feedback on topics being discussed by internet users in Indonesia.

III. RESULTS AND DISCUSSION

The research data collection took place on July 16, 2022, which means that the tweets obtained were between July 10-16, 2022, or up until a week before the data was taken from Twitter. Researchers obtained data from as many as 30,381 tweets by using RapidMiner and Twitter's application programming interface (API). This information was found by searching with the terms listed in Table 2 (which may be found below).

TABLE II
SEARCHING KEYWORD FOR SDGS

SDG	Keyword (in Bahasa Indonesia)	Translated Keyword (in English)
1	<i>Kemiskinan</i>	Poverty
2	<i>Kelaparan, Pertanian</i>	Hunger, Agriculture
3	<i>Kesehatan, Kesejahteraan</i>	Health, Well-being
4	<i>Pendidikan, Keahlian, Kampus Merdeka</i>	Education, Skills, Independent Campus
5	<i>Kesetaraan Gender, Hak Yang Sama, Pemberdayaan Perempuan</i>	Gender Equality, Equal Rights, Women's Empowerment
6	<i>Sanitasi, Air Bersih</i>	Sanitation, Clean Water
7	<i>Energi Bersih, Akses Energi, Energi Terbarukan</i>	Clean Energy, Energy Access, Renewable Energy
8	<i>Pertumbuhan Ekonomi, Lapangan Kerja, Pekerjaan Layak</i>	Economic Growth, Employment, Decent Work
9	<i>Infrastruktur, Industrialisasi, Inovasi</i>	Infrastructure, Industrialization, Innovation

10	<i>Kesenjangan, Diskriminasi, Transmigrasi</i>	Inequality, Discrimination, Transmigration
11	<i>Pemukiman, Warisan Budaya, Transportasi Umum</i>	Settlements, Cultural Heritage, Public Transportation
12	<i>Produksi, Konsumsi</i>	Production, Consumption
13	<i>Perubahan Iklim, Pencemaran Lingkungan, Rumah Kaca</i>	Climate Change, Environmental Pollution, Greenhouses
14	<i>Ekosistem Laut, Maritim, Lautan</i>	Marine, Maritime, Ocean Ecosystems
15	<i>Keanekaragaman Hayati, Kehutanan</i>	Biodiversity, Forestry
16	<i>Masyarakat Damai, Keadilan</i>	Society of Peace, Justice
17	<i>Kemitraan, Globalisasi, Kerja Sama</i>	Partnership, Globalization, Cooperation

Data understanding is obtained as a file in CSV (comma-separated values) format. This dataset has three columns: ID, Text, and SDGs, which are explained in Table 3 below. In the next stage, the Text attribute was used to clean the data.

TABLE III
DATASET ATTRIBUTE

No.	Attribute	Description
1	Id	Contains a different Unique ID for each tweet.
2	Text	It contains tweets and sentiments from Twitter users.
3	SDGs	Contains the SDGs found on Twitter.

Data preparation was processed so that it was ready for the modeling stage. This stage uses 2 tools: Jupiter Notebook (for cleansing, case folding, normalization, tokenizing, stop word removal, and stemming) and Microsoft Excel for manual labeling. Data is removed by eliminating duplication after the text preprocessing process. After these two processes, the data is considered clean and ready to be used for the next stage. However, before the modeling stage, the clean data is divided into 2, unlabeled and labeled data. Labeled data is obtained by manually labeling Sentiments, while unlabeled data was used at the deployment stage to make predictions based on the best model. Table 4 is an example of the results of the stages of data preparation.

TABLE IV
EXAMPLE RESULT OF THE DATA PREPARATION STAGES

No.	Raw Data	Data Preparation Result	Label
1	@collegemenfess bersyukur udah di kasih privilege sama Allah, aku bisa lanjut pendidikan walau ga suka sm jurusan yg skrng dan inget ortu pastinya	syukur kasih privilege allah didik suka jurusan orang tua pasti	Positive
2	@jjaeyei INFO TRANSMIGRASI KE MARS DONG	info transmigrasi mars	Neutral

No.	Raw Data	Data Preparation Result	Label
3	@wishicouldx Astaga, kesenjangan sosial ini sangat nampak???	astaga senjang sosial tampak	Negative

The amount of tweet data obtained after this process also decreased from the initial amount, which was 30,381 to 15,698 tweet data. Figure 6 is a graph comparing data preparation and data collection results.

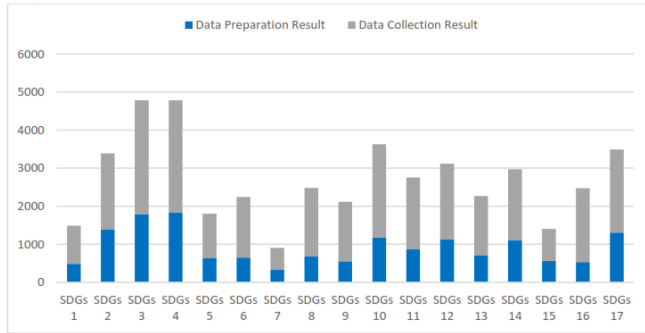


Fig. 6 Comparison of Data Preparation and Data Collection Results

The model carried out in this study is in Table 1 in the research methodology section. Table 5 contains a summary of the comparison of the accuracy of the models carried out in this study. Based on the modeling, it was found that Model D (EMLC-Stacking, SMOTE, 80:20) has the highest accuracy of 91%. This accuracy increases by 5% when compared to when not using SMOTE.

TABLE V
COMPARISON OF MODEL ACCURACY

Models	Without SMOTE		With SMOTE	
	70:30	80:20	70:30	80:20
EMLC- Stacking	81,21%	86%	83,22%	91%
EMLC-Voting	81,21%	86%	79,87%	90%

In the previous stage, it was concluded that EMLC stacking using SMOTE and 80:20 data division resulted in the highest accuracy compared to when not using SMOTE. Table 6 below displays the confusion matrix obtained for Model D.

TABLE VI
CONFUSION MATRIX FOR MODEL D

	T. Pos.	T. Neu.	T. Neg.	Precision
Pred. Positive	39	1	1	95.12%
Pred. Neutral	2	16	2	80.00%
Pred. Negative	1	2	36	92.31%
Recall	92.86%	84.21%	92.31%	

With the caption T.Pos is True Positive, T.Neu is True Neutral and T.Neg is True Negative. Next Pred. is an abbreviation for Predictions. By using formula (1), the researchers get the accuracy of the best model based on the correctly predicted sentiment. The accuracy obtained is 91% with the following calculation description.

$$Accuracy = \frac{Correct Prediction}{Total Number of Dataset} \quad (5)$$

$$Accuracy = \frac{91}{100} \quad (6)$$

By using formulas (2), (3), (4) the results of recall, precision, and f1-score for each sentiment derived from the best model are obtained. Table 6 below is a complete summary of the evaluation of the best models used for the next stage which includes precision, recall and f1-scores for the positive, neutral and negative sentiments that have been obtained from Model D.

TABLE VII
SUMMARY EVALUATION OF MODEL D

	Precision	Recall	F1-Score
Positive	95.12%	92.86%	94.15%
Neutral	80.00%	84.21%	82.32%
Negative	92.31%	92.31%	91.89%

Using the best model, the next step is to make predictions for the 15,698 unlabeled tweets. Figure 7 shows the predicted ratio of Indonesian public sentiment towards the SDGs. Most tweets have positive sentiments of 47%, followed by negative and neutral sentiments of 28% and 25%. This indicates that the positive opinions tweeted by the people of Indonesia are more than negative or neutral. This is because the majority of the 17 SDGs have keywords that have positive sentiments. The keywords (Table 2) used to search for tweets influence the number of positive sentiments.

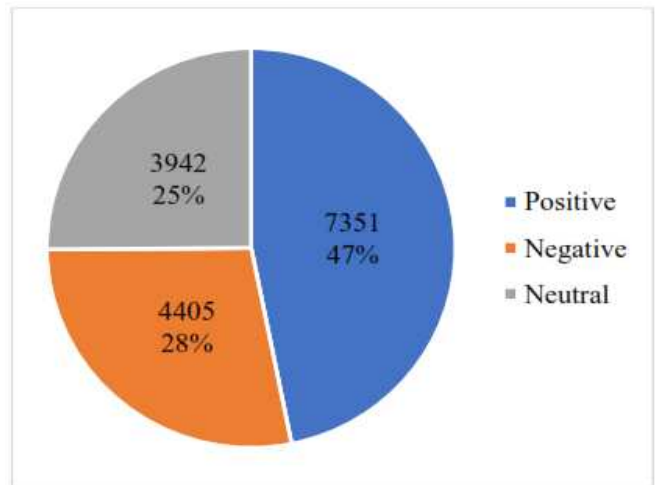


Fig. 7 Pie Chart of SDGs Sentiment in Indonesia

Based on Figure 8, the researchers obtained knowledge for each SDGs that came from tweets from internet users in Indonesia. SDGs 1, SDGs 2, SDGs 10, and SDGs 13 have a higher negative sentiment than other sentiments, which indicates that the sentiment of internet users in Indonesia regarding the SDGs is primarily negative. SDGs 3, SDGs 4, SDGs 5, SDGs 6, SDGs 7, SDGs 8, SDGs 9, SDGs 12, SDGs 14, SDGs 15, SDGs 16, SDGs 17 have a higher positive sentiment than other sentiments indicating that user sentiment the internet in Indonesia regarding the SDGs has the majority of positive sentiments. SDG 11 has a neutral sentiment that is higher than other sentiments, which indicates that the sentiment of internet users in Indonesia regarding the SDGs is mostly neither positive nor negative.

Based on the previous stages, researchers can find out that the majority of internet user sentiment in Indonesia towards the SDGs is positive. However, there are several SDGs where the majority have negative and neutral sentiments. The following is a collection of word clouds for each sentiment, positive sentiment is shown in Figure 9, neutral sentiment is shown in Figure 10, and negative sentiment is shown in Figure 11. Each sentiment has a word that occurs frequently respectively.

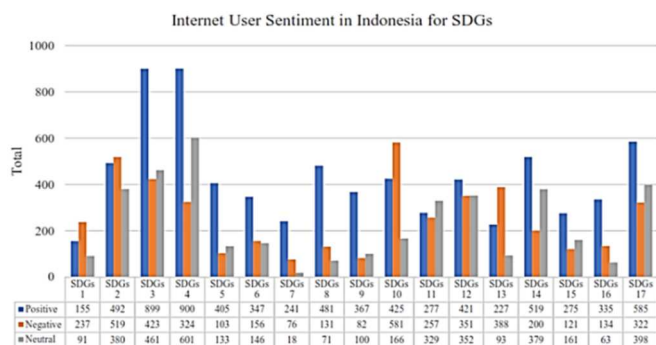


Fig. 8 Graph of SDGs Sentiment in Indonesia



Fig. 9 Word cloud of Positive Sentiment Tweets about SDGs in Indonesia



Fig. 10 Word cloud of Neutral Sentiment Tweets about SDGs in Indonesia

Based on Figure 9, we can see that the 5 words that often appear for positive sentiment tweets are "Indonesia", "kerja (work)", "masyarakat (community)", "Menteri (Minister)" and "sehat (healthy)". These words are related to SDGs 3 and SDGs 8, which, in general, based on Figure 8, are the SDGs with the majority of positive sentiment. Based on figure 10, we can see that the 5 words that often appear for neutral sentiment tweets are "kerja (work)", "ikan (fish)", "laut (sea)",

"sehat (healthy)" and "kampus (campus)". These words are related to SDGs 3, SDGs 4, SDGs 8, and SDGs 14 which, in general, based on Figure 8, are SDGs with most positive sentiments, but there are also many tweets with neutral sentiments on these SDGs. Based on Figure 11, we can see that the 5 words that often appear for negative sentiment tweets are "orang (people)", "kerja (work)", "lapar (hungry)", "Indonesia" and "miskin (poor)". These words are related to SDGs 1, SDGs 2 which in general, based on Figure 8, are the SDGs with the majority of negative sentiment.



Fig. 11 Word cloud of Negative Sentiment Tweets about SDGs in Indonesia

IV. CONCLUSIONS

This research's findings indicate that the Ensemble Machine Learning Classifier (EMLC) stacking kind is superior to the other available options. In this investigation, the best model that uses EMLC-Stacking achieves an accuracy of 91% when SMOTE and 80:20 data division are utilized. Compared to the prior model, which had an accuracy of 86% without employing SMOTE, the accuracy can rise by as much as 5% when utilizing SMOTE. The model's performance is dependent not only on the use of algorithms and SMOTE but also on the distribution of data for both the training data and the test data. In this study, the optimal data sharing consisted of 80% for training data and 20% for test data.

The sentiment of internet users in Indonesia regarding the SDGs has been predicted to fall into one of three categories, with the following results: 47% positive sentiment, 28% negative emotion, and 25% neutral sentiment. The percentage of people who feel positively about SDG 7 is the highest, while the percentage of people who think negatively about SDG 13 is the highest. The percentage of people who feel neutral about SDG 11 is the highest.

In further research at the modeling stage, it is recommended that Python be used. For further study, it is recommended to collect datasets for a whole month or a longer time (performing the data collection process every 7 days). So, the dataset used aims to describe the sentiments of internet users in Indonesia towards the SDGs in a longer time. Future research can try to change parameters or use a different combination of algorithms to get higher accuracy than this research.

Future studies are advised to examine and further analyze each SDG so that it is hoped that it can aid the government in devising policies based on sensitivity to problems in each SDG sector, which are currently being discussed by internet users in Indonesia. It is hoped that it can help the government formulate policies based on sensitivity to problems in each

SDG sector, which are currently being discussed by internet users in Indonesia. This will help increase the level of living of people in Indonesia to accomplish the aims of Sustainable Development (SDGs) by the year 2030.

REFERENCES

- [1] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and Communications Technologies for Sustainable Development Goals: State-of-the-Art, Needs and Perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2389–2406, 2018, doi:10.1109/comst.2018.2812301.
- [2] United Nations, "Transforming our world: the 2030 Agenda for Sustainable Development Transforming our world: the 2030 Agenda for Sustainable Development Preamble," 2015.
- [3] A. Rahmatulloh, R. N. Shofa, I. Darmawan, and Ardiansah, "Sentiment Analysis of Ojek Online User Satisfaction Based on the Naive Bayes and Net Brand Reputation Method," 2021 9th International Conference on Information and Communication Technology (ICoICT), Aug. 2021, doi:10.1109/icoict52021.2021.9527466.
- [4] 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 Naive Bayes," *Information*, vol. 12, no. 5, p. 204, May 2021, doi: 10.3390/info12050204.
- [5] S. S. Aljameel et al., "A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, p. 218, Dec. 2020, doi: 10.3390/ijerph18010218.
- [6] T. E. Tarigan, R. C. Buwono, and S. Redjeki, "Extraction Opinion of Social Media in Higher Education Using Sentiment Analysis," vol. 2, no. 1, 2019, [Online]. Available: <http://jurnal.kdi.or.id/index.php/bt>
- [7] T. Pawar, P. Kalra, and D. Mehrotra, "Relevance Feedback on Mobile Data Using RapidMiner," in *4th International Conference on Applied and Theoretical Computing and Communication Technology iCATccT 2018*, 2018, pp. 166–169.
- [8] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes." [Online]. Available: <http://dlvr.it/Qb83n8pic.twitter.com/8MucIMhUMO>,
- [9] S. Warjiyono, S. Aji, R. Fandhilah, N. Hidayatun, H. Faqih, and T. Liesnaningsih, "The Sentiment Analysis of Fintech Users Using Support Vector Machine and Particle Swarm Optimization Method." [Online]. Available: <https://play.google.com/store/apps/details?id=ovo.id&show>
- [10] M. Bertolotti and P. Catellani, "Effects of message framing in policy communication on climate change," *European Journal of Social Psychology*, vol. 44, no. 5, pp. 474–486, Jul. 2014, doi:10.1002/ejsp.2033.
- [11] L. K. John, D. Mochon, O. Emrich, and J. Schwartz, "What's the Value of a Like? Social Media Endorsements Don't Work the Way You Might Think," 2017.
- [12] A. Aggarwal, S. Kumar, K. Bhargava, and P. Kumaraguru, "The follower count fallacy," *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, Apr. 2018, doi:10.1145/3167132.3167318.
- [13] A. Mittal and S. Patidar, "Sentiment Analysis on Twitter Data," *Proceedings of the 2019 7th International Conference on Computer and Communications Management*, Jul. 2019, doi:10.1145/3348445.3348466.
- [14] S. Redjeki, S. Widarto, and A. Suwasto, "Deep Learning Approach for Identification of Poverty Through Sentiment Analysis," in *International Multidisciplinary Postgraduate Virtual Conference 2020 (IMPC20)*, Elsevier Inc., Jun. 2020, pp. 59–64, doi:10.1016/j.bdr.2015.01.006.
- [15] S. Zivanovic, J. Martinez, and J. Verplanke, "Capturing and mapping quality of life using Twitter data," *GeoJournal*, vol. 85, no. 1, pp. 237–255, Dec. 2018, doi: 10.1007/s10708-018-9960-6.
- [16] C. wen Shen, T. Ha Luong, and T. Pham, "Exploration of Social Media Opinions on Innovation for Sustainable Development Goals by Topic Modeling and Sentiment Analysis." [Online]. Available: <https://scholars.ncu.edu.tw/en/publications/exploration-of-social-media-opinions-on-innovation-for-sustainabl>
- [17] A. Hooda, "Sentiment Analysis of Recent Tweets for Agriculture from BRICS Countries," 2018, doi: 10.13140/RG.2.2.17830.68166.
- [18] S. P. Kristanto, J. A. Prasetyo, and P. Edwin, "Naive Bayes Classifier on Twitter Sentiment Analysis BPJS of HEALTH," in *2019 2nd International Conference of Computer and Informatics Engineering (IC2IE)*, 2019.
- [19] D. A. Nurdeni, I. Budi, and A. B. Santoso, "Sentiment Analysis on Covid19 Vaccines in Indonesia: From The Perspective of Sinovac and Pfizer," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIconCIT), Apr. 2021, doi:10.1109/eiconcit50028.2021.9431852.
- [20] R. Jayapermana, A. Aradea, and N. I. Kurniati, "Implementation of Stacking Ensemble Classifier for Multi-class Classification of COVID-19 Vaccines Topics on Twitter," *Scientific Journal of Informatics*, vol. 9, no. 1, pp. 8–15, May 2022, doi:10.15294/sji.v9i1.31648.
- [21] P. N. Venkit and S. Wilson, "Identification of Bias Against People with Disabilities in Sentiment Analysis and Toxicity Detection Models," Nov. 2021, [Online]. Available: <http://arxiv.org/abs/2111.13259>
- [22] E. V. Paz and M. C. Puga, "Sentiment Analysis in Children with Neurodevelopmental Disorders in an Ingroup/Outgroup Setting," 2019. [Online]. Available: <https://www.researchgate.net/publication/336170312>
- [23] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. Septia Nugraha, and W. Gata, "Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine," 2020.
- [24] A. C. Flores, R. I. Icy, and C. F. Peña, "An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set," in *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, 2018.
- [25] R. Baragash and H. Aldowah, "Sentiment analysis in higher education: a systematic mapping review," *Journal of Physics: Conference Series*, vol. 1860, no. 1, p. 012002, Mar. 2021, doi: 10.1088/1742-6596/1860/1/012002.
- [26] X. Yu, S. Wu, W. Chen, and M. Huang, "Sentiment Analysis of Public Opinions on the Higher Education Expansion Policy in China," *SAGE Open*, vol. 11, no. 3, p. 215824402110407, Jul. 2021, doi:10.1177/21582440211040778.
- [27] J. Oyasor, M. Raborife, and P. Ranchod, *Sentiment Analysis as an Indicator to Evaluate Gender disparity on Sexual Violence Tweets in South Africa*.
- [28] E. Alawneh, M. Al-Fawa'reh, M. T. Jafar, and M. A. Fayoumi, "Sentiment Analysis-Based Sexual Harassment Detection Using Machine Learning Techniques," 2021 International Symposium on Electronics and Smart Devices (ISESD), Jun. 2021, doi:10.1109/isesd53023.2021.9501725.
- [29] R. Ibar-Alonso, R. Quiroga-Garcia, and M. Arenas-Parra, "Opinion Mining of Green Energy Sentiment: A Russia-Ukraine Conflict Analysis," *Mathematics*, vol. 10, no. 14, p. 2532, Jul. 2022, doi:10.3390/math10142532.
- [30] K. Muludi, M. S. Akbar, D. A. Shofiana, and A. Syarif, "Sentiment Analysis Of Energy Independence Tweets Using Simple Recurrent Neural Network," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 4, p. 339, Oct. 2021, doi:10.22146/ijccs.66016.
- [31] E. Sasmaz and F. B. Tek, "Tweet Sentiment Analysis for Cryptocurrencies," 2021 6th International Conference on Computer Science and Engineering (UBMK), Sep. 2021, doi:10.1109/ubmk52708.2021.9558914.
- [32] W. Bourequat and H. Mourad, "Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine," *International Journal of Advances in Data and Information Systems*, vol. 2, no. 1, pp. 36–44, Apr. 2021, doi: 10.25008/ijadis.v2i1.1216.
- [33] M. V. M. Buladaco, "Sentiments Analysis On Public Land Transport Infrastructure in Davao Region using Machine Learning Algorithms," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1, pp. 685–690, Feb. 2020, doi:10.30534/ijatce/2020/97912020.
- [34] E. Lee, F. Rustam, P. B. Washington, F. E. Barakaz, W. Aljedaani, and I. Ashraf, "Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model," *IEEE Access*, vol. 10, pp. 9717–9728, 2022, doi:10.1109/access.2022.3144266.
- [35] S. Kumar, M. Gahalawat, P. P. Roy, D. P. Dogra, and B.-G. Kim, "Exploring Impact of Age and Gender on Sentiment Analysis Using Machine Learning," *Electronics*, vol. 9, no. 2, p. 374, Feb. 2020, doi:10.3390/electronics9020374.

- [36] M. Paolanti *et al.*, “Deep Convolutional Neural Networks for Sentiment Analysis of Cultural Heritage,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Aug. 2019, pp. 871–878. doi: 10.5194/isprs-archives-XLII-2-W15-871-2019.
- [37] P. Sasikala, S. P. #1, L. Mary, and I. Sheela, “Sentiment Analysis of Online Food Reviews using Customer Ratings,” *Article in International Journal of Pure and Applied Mathematics*, 2018, doi:10.5281/zenodo.1249390.
- [38] A. Rasool, R. Tao, K. Marjan, and T. Naveed, “Twitter Sentiment Analysis: A Case Study for Apparel Brands,” *Journal of Physics: Conference Series*, vol. 1176, p. 022015, Mar. 2019, doi:10.1088/1742-6596/1176/2/022015.
- [39] N. Mohamad Sham and A. Mohamed, “Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches,” *Sustainability*, vol. 14, no. 8, p. 4723, Apr. 2022, doi:10.3390/su14084723.
- [40] N. Mucha, “Sentiment Analysis of Global Warming Using Twitter Data.”
- [41] S. Ohtani, “How is People’s Awareness of ‘Biodiversity’ Measured ?Using Sentiment Analysis and LDA Topic Modeling in the Twitter Discourse Space from 2010 to 2020,” Oct. 2021, doi:10.21203/rs.3.rs-922908/v2.
- [42] D. Papachristos, N. Nikitakos, and M. Lambrou, “A Neuroscience Approach in User Satisfaction Evaluation in Maritime Education,” *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 7, no. 3, pp. 319–326, 2013, doi: 10.12716/1001.07.03.01.
- [43] M. Wongkar and A. Angdresey, “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter,” 2019 Fourth International Conference on Informatics and Computing (ICIC), Oct. 2019, doi: 10.1109/icic47613.2019.8985884.
- [44] M. Merle, G. Reese, and S. Drews, “#Globalcitizen: An Explorative Twitter Analysis of Global Identity and Sustainability Communication,” *Sustainability*, vol. 11, no. 12, p. 3472, Jun. 2019, doi: 10.3390/su11123472.
- [45] F. Foroughi and P. Luksch, “Data Science Methodology for Cybersecurity Projects,” *Computer Science & Information Technology*, Feb. 2018, doi: 10.5121/csit.2018.80401.