BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1439–1448
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1439-1448

COMPARISON OF RANDOM FOREST AND NAÏVE BAYES CLASSIFIER METHODS IN SENTIMENT ANALYSIS ON CLIMATE CHANGE ISSUE

Fatkhurokhman Fauzi [1*], Wiwik Setiayani [2], Tiani Wahyu Utami[3], Eko Yuliyanto[4], Iis Widya Harmoko[5]

[1,2,3]Department of Statistics, Faculty of Mathematics and Natural Science, Muhammadiyah Semarang University,
[4] Department of Chemistry Education, Faculty of Mathematics and Natural Science, Muhammadiyah Semarang University
Semarang, 50273, Indonesia

[5]Meteorology Climatology and Geophysics
Semarang, 50145, Central Java, Indonesia

*Corresponding author's e-mail: * **fatkhurokhmanf@unimus.ac.id***

ABSTRACT

The last decade was recorded as a decade with a bad record on the issue of disasters in the world due to climate change. Measuring public opinion is one of the steps to mitigate the impact of climate change. Twitter is a popular social media for channeling opinions. Twitter provides a great source of data for understanding public opinion and the perceived risk of an issue. In recent decades, when discussing climate change, there are those who agree and those who oppose it. Sentiment analysis is a branch of learning in the realm of text mining that is used as a solution to see opinions on a problem, one of which is climate change. In this study, we will try to analyze opinions on climate change issues using the Random Forest and Naïve Bayes classifier methods. Data were obtained from Twitter for the period January 2022-June 2022. The training data used in this research is 80%:20%. There are slightly more positive statements than negative ones. The results obtained with the Naïve Bayes classifier method are an accuracy of 76.25%, an F-1 score of 78%, and a recall of 80%. While the results of the random forest method are 70.6% accuracy, 69% F-1 score, and 63% recall. The Nive Bayes method is better than the Random Forest method for classifying climate change opinions with an accuracy of 76.25%.

Article History:

Received: 20th February 2023
Revised: 1st August 2023
Accepted: 9th August 2023

Keywords:

Climate change;
Naïve bayes classifier;
Random forest;
Sentiment analysis;
Text mining;
Twitter

How to cite this article:

F. Fauzi, W. Setiayani, T. W. Utami, E. Yulianto and I. W. Harmoko., "COMPARISON OF RANDOM FOREST AND NAÏVE BAYES CLASSIFIER METHODS IN SENTIMENT ANALYSIS ON CLIMATE CHANGE ISSUE," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1439-1448, September, 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The last decade has been recorded as a decade with a series of bad records on disaster issues in the world. According to the publication of the Institute for Essential Services Reform, it shows several world disasters such as landslides and devastating floods in China, flash floods in Pakistan and toxic smog in Russia [1], [2]. Natural disasters that occurred in Pakistan, China and Russia are in accordance with the predictions of the Intergovernmental Panel on Climate Change in 2007 as the impact of climate change [3]. Climate change is a change in the physical conditions of the earth's atmosphere, including temperature and distribution of rainfall which has a wide impact on various sectors of human life [4].

When discussing climate change, there are those who defend, and there are those who oppose human-caused global warming. This scenario shows a 50-50 split between those who support and oppose the perceived climate change [5]. In fact, there are still arguments or opinions that do not believe in climate change. Therefore, it is necessary to analyze public opinion as one of the measures to mitigate the impacts of climate change. Twitter is a popular social media for channeling opinions. Twitter provides a great source of data for understanding public opinion and risk perceptions about an issue [6]. With so many Twitter users, it will be an opportunity to be used as research material on climate change in Indonesia. The number of opinions will be extracted by extracting information to see the sentiments of twitter users towards climate change that is happening in Indonesia. The public opinion needs to be classified into the form of positive or negative opinion using sentiment analysis.

Sentiment analysis is a branch of learning in the domain of text mining which studies the analysis of an opinion, sentiment, attitude, evaluation which is poured into textual form [7]. In conducting sentiment analysis, of course, requires a method to classify data. Several classification methods can be applied such as naïve bayes classifier, support vector machine, decision tree and random forest [8].

One of the best methods used for text data classification is the Naïve Bayes method [9]. The basis of the Naïve Bayes method is Bayesian theory [10]. The Naïve Bayes method can aim at the opportunities it catches. Classification methods similar to Naïve Bayes include Decision Trees and Neural Networks. A comparison between the Naïve Bayes, Decision Tree, and Neural Network methods has been carried out by Xhemali et al [11] . The results obtained are that the Naïve Bayes method is still better than the Decision Tree, and Neural Network methods.

Random Forest is one of the Decision Tree methods. The Decision Tree classifies a sample data whose class is unknown. Overfitting cases can be avoided by using the Decision Tree method. Random forest classification has very good results in various cases [12]. In the study by Supriyadi et al [13], comparing the Support Vector Machine (SVM) with the Random Forest gave the result that the Random Forest method had good performance.

Research on the issue of climate change has been studied by Lydiri et al [14]. This study uses the Convolutional Neural Network (CNN) method. Research generates input in increasing the impact of climate change on a country. The accuracy produced by the Convolutional Neural Network method is quite good.

Subsequent research by Dahal et al [15] in 2019. This research contains an analysis of public opinion in the United States regarding the issue of climate change in the world. Iceland is the country with the most voices on climate change via Twitter. People's negative statements regarding climate change dominate.

Literacy about the problem of climate change will determine the country's public policy. Indonesia is a country that is vulnerable to the impact of climate change. Literacy is needed regarding climate change. The ability of community literacy can be seen from the way of opinions on social media. The absence of sentiment analysis related to the issue of climate change in Indonesia is the background of this research. In addition, the two best methods based on the above research will be compared.

2. RESEARCH METHODS

The data analysis steps used in this study are (1) Collecting tweet data using the Twitter API, (2) Pre-processing data, (3) Visualization, (4) Word weighting with TF- IDF, (5) Sharing training and testing data, (6) Classifying data using random forest and naïve bayes classifier methods, (7) Comparing the performance results of random forest and naïve bayes classifier.

2.1 Study and Data

The research conducted is sentiment analysis on twitter data regarding climate change or climate change. Some of the keywords used include climate change, climate change, climate crisis, global warming and global warming. Tweet data was collected from January 2022-June 2022. The data analyzed were data that had gone through a preprocessing process, totaling 1600 documents with a ratio of 80%: 20%.

Table 1. The Preprocessing

Raw Data	Cleaning and Case Folding	Tokenizing	Stemming and Filtering	Finish
b'RT @ICRC_id: Kita semua punya peran untuk memerangi perubahan iklim.\n\n#ClimateActionNow \n\nhttps://t.co/uNDskdfXtC'	kita semua punya peran untuk memerangi perubahan iklim	['kita', 'semua', 'punya', 'peran', 'untuk', 'memerangi', 'perubahan', 'iklim']	['peran', 'perang', 'ubah', 'iklim']	peran perang ubah iklim

Based on **Table 1**, the pre-processing stages and the results from each stage are provided. The first column is the original document before entering the pre-processing process. Furthermore, the document goes through the first pre-processing process, namely cleaning and case folding. After that proceed to the tokenizing process to the stemming and filtering processes. The end of the pre-processing stage will produce a new document as in the finish column.

2.2 Sentiment Analysis

Sentiment analysis is a field of study to analyze opinions, views, evaluations, judgments, attitudes, and emotions on the aspects expressed through texts. The main purpose of sentiment analysis is to measure the perspective, sentiment, evaluation, attitudes and emotions of the speaker or writer based on the computational treatment of subjectivity in a text. **[16]**.

Text mining has the goal of finding words in a set of documents and analyzing the relationship between words in the document **[17]**. The steps that can be taken in doing text mining are:

1. Text preprocessing
2. Visualization using word cloud.

The formula for the TF-IDF weighting can be describe as following:

$$W_{ad} = TF_{ad}\ x\ \log\frac{D}{Df_a} \tag{1}$$

Where:
W_{ad} = TF-IDF weighting of term d in document a
TF_{ad} = The frequency of term d occurs in document a
D = The total number of documents in collection
Df_a = The number of documents that contain the term a

Word cloud is a system that creates visualization of words by emphasizing the frequency of occurrence of related words in written discourse **[17]**. In general, word cloud is a visual representation of text data, usually used to describe data on a site. The word cloud display can be distinguished according to the sentiment label category used in sentiment analysis classification research.

2.3 Random Forest

Random forest is a learning method using a decision tree as the base classifier by applying the bootstrap aggregating (bagging) method, and random feature selection. The algorithm in building a random forest on a data cluster consisting of n observations and consisting of p explanatory variables (predictors), the following stages**[18]**:

a. Bootstrapping: The training dataset is randomly sampled with replacement, so each decision tree is given a different subset of data to learn from.

b. Tree Formation: For each sampled dataset, a decision tree is formed using a randomly selected subset of features. This helps avoid the creation of trees that are too correlated.

c. Voting or Averaging: Once all decision trees are formed, they provide their respective predictions. In classification, the prediction results from each tree are tallied, and the class with the majority vote becomes the final prediction. In regression, the prediction values from each tree are averaged.

$$Entropy\ (S) = \sum_{i=1}^{c} -p_i log_2 p_i \quad (2)$$

Where

S is data set

c is the number of classes, the class is obtained from the label on the dependent variable. In this case there are two levels, namely positive and negative.

p_i is the probability of class *i* frequency in the dataset. The probability is obtained from the number of labels (positive/negative) divided by the amount of data.

$$Entropy(predictor) = P(S_i) x Entropy\ (S_i) \quad (3)$$

Where $P(S_i)$ is the probability (S_i) and Entropy (S_i) is the entropy for the sample that has a value of i.

$$Gain\ (A) = Entropy\ (S) - Entropy\ (predictor) \quad (4)$$

Where $Entropy\ (S)$ is the entropy of the target.

The highest gain value will be the root node. Furthermore, the entropy and gain are calculated again from the remaining variables to produce entropy = 0 which is the end of the node (the end of the branch).

2.3 Naïve Bayes classifier

Naïve Bayes Classifier is a classification with probability and statistics proposed by British scientist Thomas Bayes, which predicts future opportunities based on previous experience **[17]**. The attributes used in the naïve bayes classifier are training data, prior probabilities, and posterior probabilities. In general, the adjusted Bayes theorem probabilities are formulated as follows:

$$P(C|F_1, F_2, \dots, F_n) = \frac{P(C).P(F_1, F_2, \dots, F_n|C)}{P(F_1, F_2, \dots, F_n)} \quad (5)$$

Where C represents class and $F_1, F_2, \dots, F_n$ is a necessary characteristic as a basis for classification. $P(C|F_1, F_2, \dots, F_n)$ in the formula is a posterior probability, namely the probability of entering a certain characteristic sample in class C. While $P(C)$ is a prior probability, which is the probability that class C will appear before the sample is entered. $P(F_1, F_2, \dots, F_n|C)$ is the probability of the appearance of class C sample characters. $P(F_1, F_2, \dots, F_n)$ referred to as the probability of the emergence of a sample character globally or called evidence. Here are the basic steps of the Naive Bayes algorithm:

a. Collect Training Data: First, you need to have a dataset that contains properly labeled data samples. This dataset will be used to train the Naive Bayes model.

b. Calculate Class Probabilities: Calculate the probabilities of each class in the training dataset. This can be done by counting how many times each class appears in the dataset divided by the total number of samples.

c. Calculate Feature Probabilities: For each attribute (feature) in the dataset, calculate its probability of occurrence in each class. This involves counting how many times the attribute appears in each class divided by the total number of samples in that class.

d. Calculate Posterior Probabilities: After obtaining class probabilities and feature probabilities, you can calculate the posterior probabilities (the probability of a class given the features) for each class.

2.4 Evaluation Model using Confusion Matrix

Confusion matrix is one method that can be used to measure the performance of a classification method. For a binary classification problem can be represented by table below:

Table 2. Confusion Matrix

Predited Class	Actual Class	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Based on the confusion matrix above, it can be obtained the value of accuracy, precision, recall, and the value of the F-1 score which is used as a model evaluation. The following Equations are used:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

The accuracy value describes how accurately the system can classify the data correctly.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

Precision is the proportion of the number of documents found and considered relevant for the needs of an information.

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

Recall is the proportion of the number that can be recovered in the search process.

$$F - 1\ score = \frac{2\ x\ (Recall\ \times Precision)}{(Recall + Precision)} \tag{9}$$

F1-Score is the mean of precision and recall.

3. RESULTS AND DISCUSSION

This sentiment analysis discusses the climate change that occurred in Indonesia where the sentiment is divided into two classes, namely the positive class and the negative class. Positive class contains positive sentences such as suggestions, information and support for government programs on climate change in Indonesia. While the negative class contains negative statements. The data was successfully categorized into positive and negative sentiments as shown in the figure below.

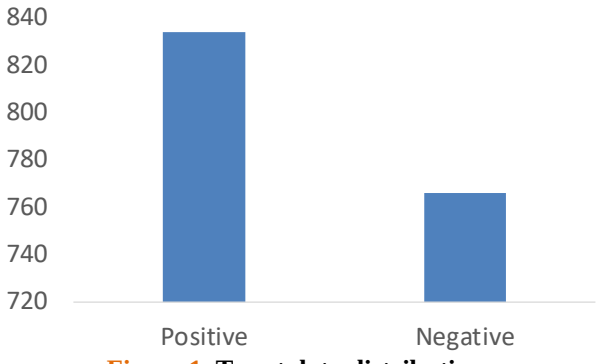

Figure 1. Tweet data distributions

Based on Figure 1, there are more tweets categorized as positive than tweets categorized as negative. A total of 834 data or 52% are categorized as positive data and 766 data or 48% are categorized as negative data. To see a picture of words that often appear in tweets, both words that are in positive and negative categories can be seen on word cloud.

Figure 2. Positive class word cloud

Based on **Figure 2**, the word “ubah iklim” in English is climate change. That word is often appears in positive category tweets related to the topic of climate change and is followed by other words. Climate change is a call from tweeters to the general public and the government for each individual to try to protect the earth from climate change.

Figure 3. Negative class word cloud

Based on **Figure 3**, it can be seen that the word "panas global" is a word that often appears in negative category tweets related to the topic of climate change and is followed by other words. “Panas global” is the basic word of global warming. Global warming is a process of increasing the average temperature of the atmosphere, oceans, and land of the earth.

The term weighting is used in order to generate the classification model. The Term Frequency-Inverse Document Frequency (TF-IDF) weighting can be seen in the table below.

Table 3. Weight of Each Document

Documents	Weight
1	2.838012
2	2.495462
3	2.431002
⋮	⋮
1280	2.192915

The weight of each document indicates the level of relevance (suitability) between the document and the term. The value of the weight on the document is directly proportional to the level of document similarity to the term being searched for. Therefore, the weight of the documents listed in **Table 3** which has the highest level of similarity to all terms is owned by document 5 and the lowest level of relevance is owned by document 5.

Furthermore, to classify sentiments using the random forest method by utilizing tuning parameters which are carried out computationally using the scikit-learn of functions in python. Based on tuning parameters, there are 50 random forest classifications used.

Table 4. Confusion Matrix of Random Forest

Predicted Value	Actual Value	
	Positive	Negative
Positive	105	31
Negative	61	121

From **Table 4**, it can be seen that, the data entered in True Positive or predicted to be exactly positive were 105 data and False Positive or predicted positive were actually negative as many as 31 data. There are 121 true negatives or those predicted to be exactly negative and 63 false negatives or predicted negatives actually positive.

Furthermore, calcification with the Naïve Bayes method. The first step is to calculate the probability of each class.

$$P(v_{positive}) = \frac{n_{positive}}{n_{documents}} = \frac{1}{2} = 0.5$$

$$P(v_{negative}) = \frac{n_{negative}}{n_{documents}} = \frac{1}{2} = 0.5$$

After getting the probabilities of positive and negative sentiments, then the probability of each word appearing in each category is calculated. An example of a word whose probability will be calculated is "babat" using **Equation (5)**.

$$P(a_{babat}|v_{positive}) = \frac{n_{babat} + 1}{n_{positive} + n_{total}} = \frac{0+1}{9+14} = 0.04348$$

$$P(a_{babat}|v_{negative}) = \frac{n_{babat}}{n_{negative} + n_{total}} = \frac{1+1}{7+14} = 0.09524$$

The following is the result of calculating the probability of each word appearing in each category (**Table 5**).

Table 5. Calculating The Probability of Word

Keywords	Positive		Negative	
	Term Frequency	Probability	Term Frequency	Probability
Babat	0	0.04348	1	0.09524
Global	1	0.08696	1	0.09524
Gundul	0	0.04348	1	0.09524
Habis	0	0.04348	1	0.09524
Hadap	1	0.08696	0	0.04762
hutan	0	0.04348	1	0.09524
ikn	1	0.08696	0	0.04762
indonesia	1	0.08696	0	0.04762
kalimantan	0	0.04348	1	0.09524
ksp	1	0.08696	0	0.04762
panas	1	0.08696	1	0.09524
pindah	1	0.08696	0	0.04762
serius	1	0.08696	0	0.04762
wujud	1	0.08696	0	0.04762

The highest probability value between the positive class and the negative class is used as the class keyword. The next step is to find the highest trustworthiness of the tested tweets. For example, there are keywords “panas”, “global”, “hutan” and “gundul”. Using the probability values in **Table 5**, the calculation of the conditional probability values for these keywords is as follows.

$$\begin{aligned} P(v_{positive}) \prod P(a_i|v_{positive}) &= (0.5)\ (P(panas|positive) \times P(global|positive) \times \\ &\quad P(hutan|positive) \times P(gundul|positive)) \\ &= (0.5) \times 0.08696 \times 0.08696 \times 0.04348 \times 0.04348 \\ &= 0.00000715 \end{aligned}$$

$$P(v_{negative}) \prod P(a_i|v_{negative}) = (0.5)\ (P(panas|negative) \times P(global|negative) \times P(hutan|negative) \times P(gundul|negative))$$
$$= (0{,}5) \times 0{,}09524 \times 0.09524 \times 0.09524 \times 0.09524$$
$$= 0.00004114$$

Based on the calculation above, the highest probability value is the negative class probability of 0.00004114. Thus, the keyword "global heat of deforestation" is included in the negative class. The same process is also carried out for all data testing. The results of the classification using the Naïve Bayes method are written in the confusion matrix.

Table 6. Confusion Matrix of Naïve Bayes Classifier

Predicted Value	Actual Value	
	Positive	Negative
Positive	135	43
Negative	33	109

From Table 6 above, it can be seen that the data entered in True Positive or which were predicted to be exactly positive were 135 data and False Positive or predicted positive were actually negative as many as 43 data. True Negatives or those predicted to be exactly negative were 109 data and False Negatives or predicted negatives were actually positive as many as 33.

Table 7. Comparison of Random Forest and Naïve Bayes Classifier Methods

Methods	Acuration	F-1 Score	Recall
Random Forest	70.6%	69%	63%
Naïve Bayes	76.25%	78%	80%

After analyzing each method, the best algorithm for classifying sentiment on climate change issues is the naïve bayes classifier. Based on Table 8, the Naïve Bayes classifier method succeeded in classifying 320 data with an accuracy of 76.25%. The Naïve Bayes classifier has greater accuracy than the Random Forest method which has an accuracy of 70.6%. In addition, the value of F-1 Score and recall of Naïve Bayes classifier is greater than the random forest method. The F-1 score for the Naïve Bayes classifier is 78% and the recall value is 80%. Meanwhile, the F-1 score for the Random Forest method is 69% and the recall value is 63%.

4. CONCLUSIONS

The issue of climate change is an issue that is often discussed in various countries. Most Indonesians support action to prevent climate change. However, quite a few people do not believe it. Classification is important for mapping and predicting one's opinion on climate change. This study used two classification methods, namely random forest and naive Bayes classifier, and then compared.

The results of the accuracy of the naïve Bayes classifier is 76.25%, the F-1 score is 78%, and the recall is 80%, which is better than the random forest method with an accuracy of 70.6%, the F-1 score is 69%, and the recall is 63%. Expanding the data collection time span and multiple classification methods can be used to improve performance. The limitations of this study are the short data collection, which is six months. Therefore, to improve performance, expand the data collection period.

ACKNOWLEDGMENT

Acknowledgements to the Ministry of Education, Culture, Research, and Technology (KEMENDIKBUDRISTEK) and the Ministry of Finance (KEMENKEU) for financial support in this research.

REFERENCES

[1] Y. Chen, J. Li, and A. Chen, "Does high risk mean high loss: Evidence from flood disaster in southern China," *Science of The Total Environment*, vol. 785, p. 147127, 2021, doi: https://doi.org/10.1016/j.scitotenv.2021.147127.

[2] B. A. Revich, "Toward the assessment of the Russian population mortality risk factors and the feasibility of their reduction: Comments on the World Bank Report 'dying Too Young,'" *Stud Russ Econ Dev*, vol. 17, no. 6, pp. 643–654, Dec. 2006, doi: 10.1134/S1075700706060104.

[3] S. Solomon, *Climate change 2007 : the physical science basis : contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2007.

[4] Y. Suryadi, D. N. Sugianto, and Hadiyanto, "Climate Change in Indonesia (Case Study: Medan, Palembang, Semarang)," *E3S Web of Conferences*, vol. 31, pp. 3–8, 2018, doi: 10.1051/e3sconf/20183109017.

[5] Supari, F. Tangang, L. Juneng, and E. Aldrian, "Observed changes in extreme temperature and precipitation over Indonesia," *International Journal of Climatology*, vol. 37, no. 4, pp. 1979–1997, 2017, doi: https://doi.org/10.1002/joc.4829.

[6] B. Dahal, S. A. P. Kumar, and Z. Li, "Topic modeling and sentiment analysis of global climate change tweets," *Soc Netw Anal Min*, vol. 9, no. 1, p. 24, 2019, doi: 10.1007/s13278-019-0568-8.

[7] B. Zhang Lei and Liu, "Sentiment Analysis and Opinion Mining," in *Encyclopedia of Machine Learning and Data Mining*, G. I. Sammut Claude and Webb, Ed., Boston, MA: Springer US, 2017, pp. 1152–1161. doi: 10.1007/978-1-4899-7687-1_907.

[8] Bahrawi, "SENTIMENT ANALYSIS USING RANDOM FOREST ALGORITHM-ONLINE SOCIAL MEDIA BASED."

[9] P. Liu, H. han Zhao, J. yu Teng, Y. yan Yang, Y. feng Liu, and Z. wei Zhu, "Parallel naive Bayes algorithm for large-scale Chinese text classification based on spark," *J Cent South Univ*, vol. 26, no. 1, pp. 1–12, Jan. 2019, doi: 10.1007/s11771-019-3978-x.

[10] H. Annur, "KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES," *ILKOM Jurnal Ilmiah*, vol. 10, no. 2, pp. 160–165, 2018.

[11] C. J. Hinde, R. Stone, D. Xhemali, C. J. Hinde, and R. G. Stone, "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *IJCSI International Journal of Computer Science Issues*, vol. 4, no. 1, 2009, [Online]. Available: https://hdl.handle.net/2134/5394.

[12] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, vol. 1, no. 1, pp. 27–31, 2018.

[13] R. Supriyadi, W. Gata, N. Maulidah, A. Fauzi, I. Komputer, and S. Nusa Mandiri Jalan Margonda Raya No, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *JURNAL ILMIAH EKONOMI DAN BISNIS*, vol. 13, no. 2, pp. 67–75, 2020, [Online]. Available: http://journal.stekom.ac.id/index.php/E-Bisnis■page67

[14] M. Lydiri, Y. el Habouz, and H. Zougagh, "Sentiment Analysis Decision System for Tracking Climate Change Opinion in Twitter," in *Business Intelligence*, M. Fakir, M. Baslam, and R. el Ayachi, Eds., Cham: Springer International Publishing, 2022, pp. 188–196.

[15] B. Dahal, S. A. P. Kumar, and Z. Li, "Topic modeling and sentiment analysis of global climate change tweets," *Soc Netw Anal Min*, vol. 9, no. 1, Dec. 2019, doi: 10.1007/s13278-019-0568-8.

[16] A. Sabika, "SENTIRESEARCH: LEXICON-BASED WEB APPLICATION FOR INDONESIAN SENTIMENT ANALYSIS," 2020.

[17] P. M. Dikky and A. Indriani, "IMPLEMENTASI TEXT MINING DALAM KLASIFIKASI BUKU DENGAN METODE NAÏVE BAYES CLASSIFIER STUDI KASUS PADA PERPUSTAKAAN STMIK PPKIA TARAKANITA RAHMAWATI," in *Seminar Nasional Inovasi dan Tren*, 2014, pp. 243–247.

[18] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, 2012, doi: https://doi.org/10.1016/j.isprsjprs.2011.11.002.