International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 6, June - 2022, Pages: 104-110

Sentiment Analysis of Sustainable Development Goals on Twitter with Classifying Decision Tree C5.0 and Classification and Regression Tree

Siswanto[1*], Muhammad Yusran[1], Sapriadi Rasyid[1], Evi Sagita[1], Rahmah Ningsih Dwika Julia[1]

[1]Departement of Statistics, Hasanuddin University, Makassar 90245, Indonesia
*siswanto@unhas.ac.id

***Abstract**: Sustainable Development Goals (SDGs) as one of the programs carried out by the United Nations with the main objective of reducing inequality, ending poverty and protecting the environment. Sentiment analysis is used to process and further analyze public opinion on a topic. This study aims to find out how many neutral, positive and negative sentiments are for the SDGs topic on Twitter and to build a classification model of public sentiment towards SDGs using Decision Tree C5.0 and Classification and Regression Tree (CART). The data used in this study are tweets on the topic of SDGs in English from May 17, 2022 to May 24, 2022. The results showed that there were 3,956 tweets with neutral sentiment, 5,228 tweets with positive sentiment and 816 tweets with negative sentiment. The Decision Tree C5.0 model has an accuracy, precision and recall value of 94.54%, 94.54% and 99.43%, respectively, while the CART model has an accuracy, precision and recall value of 92.97%, 99.42% and 92.93%, respectively. The best model for classifying sentiment on the topic of SDGs is the Decision Tree C5.0 model.*

Keywords— Sentiment Analysis, Sustainable Development Goals, Twitter, Decision Tree C5.0, Classification and Regression Tree.

1. INTRODUCTION

Technology has become one of the aspects that play an important role in obtaining information or knowledge needs. Technological developments make it easier for someone to access public opinion. One form of rapid technological development is the presence of social media. Social media is one of the public services to convey or obtain news, opinions and comments in short messages. Twitter is one of the most popular social media in Indonesia [1]. According to the Statista report as of January 2022, twitter users in Indonesia reached 18.45 million users, making Indonesia the 5th largest twitter user in the world [2]. Sustainable Development Goals (SDGs) is one of the topics on Twitter.

SDGs is one of the programs carried out by the United Nations as an international agenda in the context of the welfare of the world community. The main goals of the SDGs are reducing inequality, ending poverty and protecting the environment [3]. However, the SDGs have generated many opinions, both pro and contra on the topic of the SDGs. To process and further analyze opinions on the topic of SDGs, sentiment analysis can be used.

Sentiment analysis is a method of finding the percentage level of negative sentiment and positive sentiment on a topic [4]. Sentiment analysis is one way to see the tendency of public opinion towards a problem or topic [5]. The main purpose of sentiment analysis is to classify texts that are negative, neutral or positive based on the results of sentiment [6]. There are several classification methods that can be used in sentiment analysis, including Decision Tree C5.0 and Classification and Regression Tree (CART) [7].

Decision Tree C5.0 has the advantage that it is able to distinguish data based on attributes that provide information and divide attributes according to the greatest information gain value [8]. The application of Decision Tree C5.0 plays an important role in decision making because the decisions taken are more objective [9]. Meanwhile, CART is an innovative tree construction technique in determining the magnitude of the effect of predictor variables on outcome variables. The advantage of CART is that the resulting tree is clear and easy to interpret [10].

Research conducted by Wirdhaningsih, et al. (2001) applied Decision Tree C5.0 to forecast forex with time frame M15 for EUR/USD currency with buy accuracy of 84.49%, sell accuracy of 83.69% [11]. The CART method has been used by Zimmerman, et al. (2016) to predict influenza in primary care patients. The results of this study provide a negative predictive value (NPV) of 94% for the development sample, a sensitivity value of 84%, a specificity of 48% and a positive predictive value (PPV) of 23% [12].

Based on this explanation, a research will be conducted on sentiment analysis of the Sustainable Development Goals on Twitter by classifying the Decision Tree C5.0 and Classification and Regression Tree to identify the tendency of public opinion towards the SDGs on Twitter. This study aims to find out how many neutral, positive and negative sentiments are for the SDGs topic on Twitter and to build a classification model of public sentiment towards the SDGs using Decision Tree C5.0 and Classification and Regression Tree. It is hoped that this research can help the United Nations and the government to implement the SDGs program.

2. LITERATURE REVIEW

2.1 Text Mining

Text mining is a process of mining textual data derived from text and has the aim of finding words that can visualize the contents of the text so that correlations can be analyzed between one text and another. Therefore, in this process the data source is used in the form of an unstructured collection of documents with the identification and exploration of interesting patterns [13]. Text mining as a form of variation of data mining has uses for classifying. The technique seeks to explore interesting patterns sourced from large text data sets. [14].

2.2 Sentiment Analysis

A Sentiment analysis has the aim of determining public perception or subjectivity regarding a topic of discussion, event, or problem [15]. This analysis is applied to determine the public's tendency towards a topic or problem, whether it tends to be positive or negative towards a problem through the process of reviewing, extracting and processing textual data in order to produce sentiment information in a sentence.

Identification of market opinion regarding an object of goods is one form of using sentiment analysis. Sentiment analysis has a great influence and benefit on the rapid developments in the field of research and applications related to sentiment analysis [5]. Data mining on sentiment analysis is carried out to examine, extract, and process data in the form of text on a subject, such as phenomena, products, services, individuals, and certain topics that can include review texts, forums, tweets, or blogs by identifying sentiments, and making sentiment classification model using available opinions [16].

2.3 Twitter

Twitter is one of several social media that has a fairly high popularity, including in Indonesia. The social media with logo is the blue larry bird was founded by Jack Dorsey in 2006 [17]. The main function of Twitter is to send messages called tweets. The uniqueness of twitter lies in the maximum size of 140 characters in the tweet [18]. Twitter is quite attractive to its users as a medium of communication and information is positive and or negative in the form of opinions, questions or comments [19].

2.4 Application Programming Interface Twitter

Application Programming Interface (API) is a way used by programs on a computer to communicate with one another to request and present the required information. This is done by allowing the program to call an address associated with a certain type of information needed, the address is called an endpoint [20]. Twitter API to make it easier for other developers to access the information contained on Twitter [18]. The Twitter API has several functions to carry out their respective tasks, so that software developers simply call the required functions in the software that is made.

In addition, the Twitter API uses a REST (Representational State Transfer) architecture, therefore the Twitter API can be used in various data formats such as XML and JSON. The Twitter API consists of the Twitter Search API and the Twitter Streaming API. The difference lies in the search function where the Twitter Search API focuses on the past while the Twitter Streaming API focuses on the future [21].

Twitter provides the possibility to build software that integrates with Twitter by providing access to several services using API, one example being companies that can respond to user feedback on Twitter. The twitter API allows users to openly access twitter data that the user allows to be accessed by the public. Twitter also allows APIs to allow software developers to manage non-public Twitter user information and share information with developers who are authorized to obtain it [20].

2.5 Decision Tree C5.0

Decision Tree C5.0 is the development from Iterative Dichotomiser (ID3) and C4.5 which was developed by Ross Quinlan in 1987. Decision Tree C5.0 can provide solutions on continuous and discrete attributes. The selection of attributes in the model will be processed using information gain. The attribute that will be chosen as the root for the next node is the attribute that has the highest gain value [22].

Decision Tree C5.0 is used to categorize data that produces decision tree models and decision rules. These rules are used to assist in choosing the right decision. Decision Tree C5.0 is also used to extract data where correlations can be found between certain variables and the target variable [23]. C5.0 produces a tree with various branches per node [24]. The steps for Decision Tree C5.0 are as follows [25]:

1. 1. Calculating the target variable based on the variable used for the predictor for each occurrence of events in each predictor.
2. Calculate the total entropy including the entropy for each predictor variable. As a note on each branching tree not all attributes will be used. Thus, entropy is needed as a means of controlling the attributes used to build the tree. The formula for calculating entropy is:
$$\mathrm{Entropy}(S) = -\sum_{i=1}^{n} p_i \times \log_2 p_i \tag{1}$$
with:
S : case set
n : number of partitions of S
p_i : proportion of S_i to S
3. Calculating information gain for all variables. The highest information gain value on the variable will act as the root tree. The formula for calculating information gain is:
$$\mathrm{Information\ Gain}(S, A) = \mathrm{Entropy}(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times \mathrm{Entropy}(S_i) \tag{2}$$
with:
S : case set
A : variable
n : number of partitions of variable A
$|S_i|$: number of case on ith partition
$|S|$: number of case on S
4. Calculate the gain ration with the formula:
$$\mathrm{Gain\ Ratio} = \frac{\mathrm{InformationGain}(S,A)}{\sum_{i=1}^{n} \mathrm{Entropy}(S_i)} \tag{3}$$

5. Form the first node if the variable has the highest gain ratio.
6. Repeating the calculation until all data is included in the same class, and the selected variable is no longer used in the repetition

2.6 Classification and Regression Tree

Classification and Regression Tree (CART) is a nonparametric statistical method developed for classification analysis, both for continuous and categorical response variables. CART produces a regression tree if the response variable is continuous and produces a classification tree if the response variable is categorical. CART has the aim of obtaining an accurate data set as a predictor of a classification [26]. The CART process goes through three stages, namely [27]:

1. Formation of a classification tree. This stage begins with determining predictors and thresholds to be explored for separating each node.
2. Pruning the classification tree. Trees formed in the previous process are very large, which can cause overfitting, but if they are too small they will become underfitting. Therefore, to determine the appropriate tree size, pruning will be carried out with minimum cost complexity.
3. Determination of the optimum classification tree. This determination can be made by using a cross validation estimate.

2.7 Confusion Matrix

Confusion matrix is a table that shows the amount of test data that has been categorized correctly or incorrectly by a classification model [28]. The confusion matrix can be used to measure the classification performance of a model. The following is a Confusion matrix 2 class model [29].

Table 1: Confusion Matrix

Prediction	Observation	
	1	2
1	True Positive	False Positive
2	False Negative	True Negative

The meaning of the confusion matrix entries above are:

a. True Positive (TP), data categorized as class 1 which comes from the number of class 1 data.
b. True Negative (TN), data that is categorized as class 2 which comes from the number of class 2 data.
c. False Positive (FP), data categorized as class 2 which comes from the number of class 1 data.
d. False Negative (FN) data that is categorized as class 1 which comes from the number of class 2 data.

By using the four representations of the classification results, it can be calculated

Accuracy, namely the comparison between the data that is categorized as true with the overall data obtained with the formula:

$$\text{Accuracy } = \frac{\text{TP+TN}}{\text{TP+TN+FP+IN}} \times 100\% \qquad (4)$$

Precision, namely the amount of data that is categorized as positive and true divided by the total data that is categorized as positive obtained by the formula:

$$\text{Precision } = \frac{\text{TP}}{\text{TP+FP}} \times 100\% \qquad (5)$$

Recall, which is the amount of data categorized as positive that is categorized correctly divided by data that is categorized as positive and true and negative and false obtained by the equation model:

$$\text{Recall } = \frac{\text{TP}}{\text{TP+FN}} \times 100\% \qquad (6)$$

2.8 Sustainable Development Goals

Sustainable Development Goals (SDGs) are global action plans that are agreed by world leaders. The aim is to improve the development of sustainable community economic welfare, development that maintains the sustainability of community social life, and development that maintains environmental quality [30]. The SDGs were officially ratified by the United Nations General Assembly in September 2015, lasting from 2016 to 2030 and officially becoming the goals of the United Nations. The SDGs set a vision, principle and commitment for a more just and sustainable world to end poverty, reduce inequality and protect the environment [31].

3. RESEARCH METHOD

Data used in this study are tweets on the topic of SDGs in English language which were obtained by text mining method through twitter API access. Tweets access requires an access code in the form of an access token, access token secret, consumer key, and consumer secret. This data consists of tweets from 17 May 2022 to 24 May 2022. The data analysis process in this study was carried out using R 4.1.2 software.

The steps in this research are:

1. Collecting data using the text mining method on Twitter with the keyword "SDGs". The number of tweets collected is limited to 10,000 tweets.
2. Perform data preprocessing to convert data that is still in the form of raw data into data that is ready for analysis. This preprocessing data stage consists of cleaning (process of deleting certain components contained in tweets, namely URLs, usernames, RT/retweets, HTML characters, hashtags, punctuation marks, numbers, etc.), removing stopwords (a list of general words that have functions but have no meaning, such as pronouns and conjunctions), stemming (decomposition of words into basic word forms) and case folding (making all letters in the data is lowercase).

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 6, June - 2022, Pages: 104-110

3. Labeling the sentiment on each tweets. This process generates sentiments in the form of neutral, positive and negative sentiments for each tweets. Data that has a neutral sentiment will be deleted.
4. Change the words in the tweets into predictor attributes. The words used are set to have a minimum frequency of 25.
5. Divide the data into training data and testing data with a ratio of 80:20.
6. Create a sentiment classification model using Decision Tree C5.0 and CART using training data. The steps of Decision Tree C5.0 are as follows:
   a. Calculates the target variable based on the variable used for the predictor for each occurrence of events in each predictor.
   b. Calculate the total entropy for each predictor variable using equation (1).
   c. Calculating the information gain for all variables using equation (2).
   d. Calculate the gain ratio with equation (3).
   e. Form the first node by using the variable that has the highest gain ratio.
   f. Repeat the calculation until all data is included in the same class, and the selected variable is no longer used in the iteration.

   The steps of CART are as follows
   a. Formation of a classification tree by determining the predictor and threshold to be used as a separator for each node.
   b. Pruning the classification tree with minimum cost complexity.
   c. Determination of the optimum classification tree by using a cross validation estimate.
7. Measure the performance of the classification model on the test data using a confusion matrix.
8. Comparing the two sentiment classification models and selecting the best classification model.

4. RESULT AND DISCUSSION

4.1 Data Collection

The data collection process resulted in 10,000 tweets of data on the topic of SDGs. Some of the tweets obtained based on the results of data collection carried out through twitter API access are shown in Figure 1.

```
[[1]]
[1] "michaeldacosta: RT @michaeldacosta: How to keep tax at Bay...    cc Josep M.
 Pique Juan Manu@JosepMPiqueostiola #fiscal #policy - #Davos , would be very in
t…"

[[2]]
[1] "GreenLibraries: RT @QNLib: We were pleased to hold our very first 'Libraries
 Lead' forum! We look forward to initiating programs to support the UN's Sustai…"

[[3]]
[1] "EnelFoundation: Recently the Siena International School on Sustainable Devel
opment held a webinar on the importance of the #SDGs in… https://t.co/Wb7CNBKhwK"

[[4]]
[1] "ATD_Europe: RT @AminaJMohammed: From the COVID-19 recovery to the climate cr
isis, the global community must step up their investments for prevention…"

[[5]]
[1] "rahul05ranjan: RT @Eli_Krumova: I gave +Kred to @GirlsinTech on @Kred. #infl
uence #WomeninTech #WomenInSTEM #WomeninBusiness #WomenInCrypto #STEM #EdTech…"

[[6]]
[1] "@OtSci: RT @Eli_Krumova: I gave +Kred to @GirlsinTech on @Kred. #influence #
womenintech #WomenInSTEM #WomeninBusiness #WomenInCrypto #STEM #EdTech…"
```

Fig. 1. Data Collection Results

4.2 Data Preprocessing

This stage consists of cleaning, removing stop words, stemming, and case folding. The results of the data preprocessing stages are shown in Figure 2.

```
[1] "   how to keep tax at bay      cc josep m  pique juan manu   fiscal  policy
   davos   would be very int "
[2] "   we were pleased to hold our very first  libraries lead  forum  we look fo
rward to initiating programs to support the un s sustai "
[3] "recently the siena international school on sustainable development held a we
binar on the importance of the  sdgs in    "
[4] "   from the covid  recovery to the climate crisis  the global community must
 step up their investments for prevention  "
[5] "   i gave  kred to   on     influence  womenintech  womeninstem  womeninbusi
ness  womenincrypto  stem  edtech "
[6] "   i gave  kred to   on     influence  womenintech  womeninstem  womeninbusi
ness  womenincrypto  stem  edtech "
```

Fig. 2. Data Preprocessing Results

4.3 Sentiment Labeling

The sentiment labeling stage for each tweet produces sentiment in the form of neutral, positive and sentiment for each tweet. Based on the data processing, there were 3,956 tweets with neutral sentiment, 5,228 tweets with positive sentiment and 816 tweets with negative sentiment. so that 6,044 tweets remained after the tweets with neutral sentiment were deleted. The summary of sentiment is shown in Table 2.

Table 2: Sentiment Labeling Summary

Labeling	Amount of Data	Percentage
Neutral	3.956	39,56%
Positive	5.228	52,28%
Negative	816	8,16%
Total	10.000	100%

4.4 Predictor Attribute Creation

This stage is done by changing the words in the tweets into predictor attributes. The words used are set to have 25 minimum frequency. Thus, the number of words used as attributes is 403 words. Some of the words with the highest frequency are shown in Table 3.

Table 3: Frequency of Occurrence of Words in Tweets

Word	Frequency On Tweets
Sdgs	1345

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 6, June - 2022, Pages: 104-110

Amp	678
Just	434
Making	403
Giving	374

4.5 Dividing of Training Data and Testing Data

The data is divided into training data and test data with a ratio of 80:20. Training data is used to provide training to the model, while test data is used to determine the performance of the formed model. The training data consisted of 4836 data and the test data consisted of 1208 data.

4.6 Classification Model and Performance Measurement

4.6.1 Decision Tree C5.0 Model

The Decision Tree C5.0 model that was formed is visualized through the tree diagram shown in Figure 3.

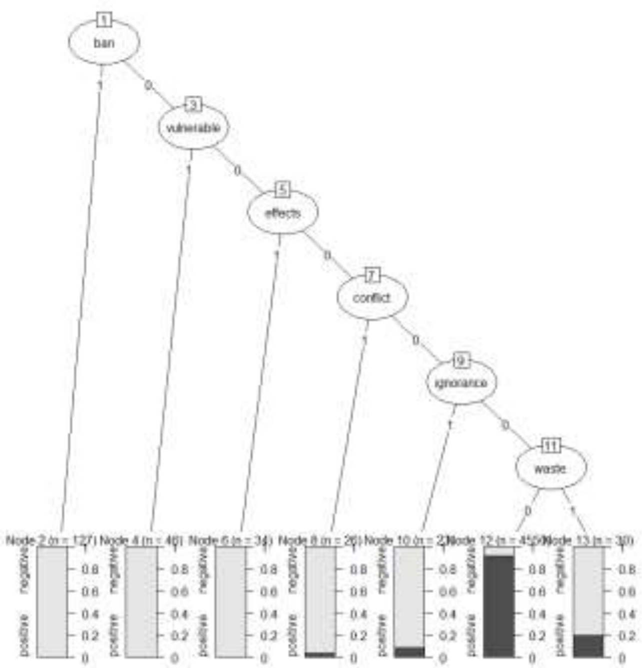

Fig. 3. Decision Tree C5.0 Model

Based on Figure 3, the most influential attribute in the case of SDGs is "ban". In addition, the highest percentage level of use of attribute modeling is shown in Table 4.

Table 4: Percentage of Attribute Usage

Attribute	Percentage
Ban	100,00%
Vulnerable	97,37%
Effects	96,42
Conflict	95,72%
Ignorance	94,67%
Waste	94,19%

The next step is to measure the classification performance using test data. The Confusion Matrix for the Decision Tree C5.0 Model is shown in Table 4 with the accuracy, precision and recall values of 94.54%, 94.54% and 99.43%, respectively.

Table 5: Confusion Matrix of Decision Tree C5.0 Model

Prediction	Observation	
	Positive	Negative
Positive	1039	60
Negative	6	103

4.6.2 Classification and Regression Tree Model

The formed CART model is shown in Figure 4. The important attributes in this model are ban, change, climate, development, risk, waste and ignorance.

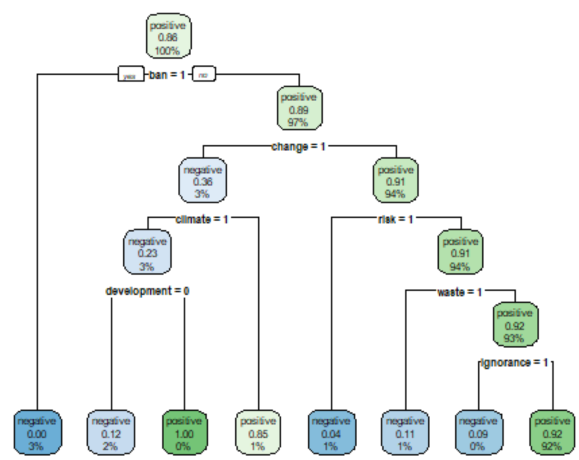

Fig. 4. CART Model

Based on Figure 4, the most influential attribute in the case of SDGs is "tire".

The next step is to measure the classification performance using test data. The confusion matrix for the CART model is shown in Table 6 with the accuracy, precision and recall values of 92.97%, 99.42% and 92.93%, respectively.

Table 6: Confusion Matrix of CART Model

Prediction	Observation	
	Positive	Negative
Positive	1039	6
Negative	79	84

4.6.3 Comparison of Decision Tree C5.0 and CART

Based on the classification results in the Decision Tree C5.0 Model and the CART Model, the comparison between the two models is shown in Table 7.

Table 7: Comparison Model

Model	Accuracy	Precision	Recall

Decision Tree C5.0	94,54%	94,54%	99,43%
CART	92,97%	99,42%	92,93%

It was found that the Decision Tree C5.0 Model was better based on the values of accuracy, precision and recall while the CART Model was better based on precision. Therefore, the best sentiment classification model for the SDGs topic was obtained, namely the Decision Tree Model C5.0.

5. CONCLUSION

Based on the research that has been done, it is concluded that there are 3,956 tweets with neutral sentiment, 5,228 tweets with positive sentiment and 816 tweets with negative sentiment. The classification using Decision Tree C5.0 Model has accuracy, precision and recall values of 94.54%, 94.54% and 99.43%, respectively, with the most influential attributes on the model are tires, vulnerable, effects, conflict, ignorance, and waste. The CART model has accuracy, precision and recall values of 92.97%, 99.42% and 92.93%, respectively. The important attributes of the model are tires, change, climate, development, risk, waste and ignorance. The comparison value of the two models gives the result that the best classification model is the Decision Tree C5.0 Model.

6. REFERENCES

[1] Tresnawati, Y. (2017). Sentiment Analysis of Twitter Using Agglomerative Hierarchical Clustering. *Undergraduate Thesis*, Yogyakarta: Sanata Dharma University.

[2] Annur, C.M. (2022). Pengguna Twitter Indonesia Masuk Daftar Terbanyak di Dunia. URL: https://databoks.katadata.co.id. Accessed on: 28 May 2022.

[3] Ulfah, N. (2022). Sustainable Development Goals (SDGs): Kenali Definisi dan 17 Tujuannya Di Sini!. URL: https://www.ekrut.com/media/sdgs-adalah. Accessed on : 28 May 2022

[4] Pramana, S., et al. (2018). Data Mining dengan R Konsep serta Implementasi, Bogor: In Media.

[5] Rozi, I. F., Pramono, S. H., and Dahlan, E. A. (2013). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS*, 6(1), pp. 37-43.

[6] Isnain, A.R., Supriyanto, J. and Kharisma, M.P. (2021). Implementation of K-Nearest Neighbor (K-NN) Alghorithm for Public Sentiment Analysis of Online Learning. *Indonesian Journal of Computing and Cybernetics Systems*, 15(2), pp. 121-130.

[7] Mulyani, S. (2022). Perbandingan Kinerja Metode Mesin Vektor Pendukung Tanpa dan dengan Fitur Berbasis Leksikson Pada Analisis sentimen Vaksinasi Covid-19 di Indonesia. *Undergraduate Thesis*, Makassar: Hasanuddin University.

[8] Utomo, D.P., et al. (2020). Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0. *Jurnal Media Informatika Budidarma*, 4(4), pp. 994-1006.

[9] Kurniawan, I. and Saputra, R.A. (2017). Penerapan Algoritma C5.0. pada Sistem Pendukung Keputusan Kelayakan Penerimaan Beras Masyarakat Miskin. *Jurnal Informatika*, 4(2), pp. 236-240.

[10] Barlin, et al. 2013. Classification and Regression Tree (CART) Analysis of Endometrial Carcinoma: Seeing the Forest for the Tree. *Gynecol Oncol*, 130(3), pp. 452-456.

[11] Wirdhaningsih, K.P., Ratnawati, D.E. and Marji. (2021). Penerapan Algoritma Decision Tree C5.0 Untuk Peramalan Forex. *Undergraduate Thesis*, Malang: Brawijaya University.

[12] Zimmerman, R.K., et al. (2016). Classification and Regression Tree (CART) Analysis to Predict Influenza in Primary Care Patient. *BioMed Central*, 16(503), pp. 1-11.

[13] Putri, E. K., and Setiadi, T. (2014). Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes. *Jurnal Sarjana Teknik Informatika*, 2(3), pp. 73-83.

[14] Ratniasih, N. L., Sudarma, M., and Gunantara, N. (2017). Penerapan Text Mining dalam Spam Filtering untuk Aplikasi Chat. *Majalah Ilmiah Teknologi Elektro*, 16(3), pp. 13-18.

[15] Rachman, F. F., and Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Indonesian of Health Information Management Journal (INOHIM)*, 8(2), pp. 100-109.

[16] Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., and Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(1), pp. 157-163.

[17] Krisdiyanto, T. (2021). Analisis sentimen opini masyarakat Indonesia terhadap kebijakan PPKM pada media sosial Twitter menggunakan naïve bayes clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 7(1), pp. 32-37.

[18] Sembodo, J. E., Setiawan, E. B., and Baizal, Z. A. (2016). Data Crawling Otomatis pada Twitter. *Indonesian Symposium on Computing*, pp. 11-16.

[19] Nurjanah, W. E., Perdana, R. S., and Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), pp. 1750-1757.

[20] Putri, D. D., Nama, G. F., and Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika dan Teknik Elektro Terapan*, 10(1), pp. 34-40.

[21] Rustiana, D., and Rahayu, N. (2017). Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 8(1), pp. 113-120.

[22] Zamasi, N., Hasibuan, N. A., and Suginam, S. (2021). Penerapan Algoritma C5.0 Dalam Analisa Data Potensi Pertanian dan Perternakan (Studi Kasus: Unit Pembinaan Perlindungan Tanaman (UPPT) Biru-Biru). *Journal of Informatics Management and Information Technology*, 1(2), pp. 85-90.

[23] Benediktus, N., and Oetama, R. S. (2020). The Decision Tree C5. 0 Classification Algorithm For Predicting Student Academic Performance. *Ultimatics: Jurnal Teknik Informatika*, 12(1), pp. 14-19.

[24] Sari, N. (2021). Penerapan Algoritma C5.0 Dalam Pencarian Pola Hasil Audit Laporan Keuangan Perusahaan Publik (Studi Kasus: PT. Bursa Efek Indonesia). *Pelita Informatika: Informasi dan Informatika*, 9(4), pp. 250-256.

[25] Dumham, M.H. (2003). Data Mining Introductory and Advanced Topics. London: Pearson Education.

[26] Karisma, R. D. L. N., and Otok, B. W. (2017, July). Model Machine Learning CART Diabetes Melitus. *Prosiding Seminar Nasional Integrasi Matematika Dan Nilai-Nilai Islami*, 1(1), pp. 485-491.

[27] Sumartini, S. H. and Purnami, S. W. Penggunaan Metode Classification and Regression Tree (CART) untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya. *Jurnal Sains dan Seni ITS*, 4(2), pp. 211-216.

[28] Normawati, D. and Prayogi, S.A. 2021. Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks pada Twitter. *Jurnal Sains Komputer dan Informatika*. 5(2), pp. 697-711.

[29] Azmi, Z., and Dahria, M. (2013). Decision Tree Berbasis Algoritma Untuk Pengambilan Keputusan. *Jurnal Ilmiah SAINTIKOM*, 12(3), pp. 157-164.

[30] Sustainable Development Goals (SDGs). *Apa itu SDGs*. URL: https://www.sdg2030indonesia.org/page/8-apa-itu. Accessed on 2 June 2022.

[31] Morton, S., Pencheon, D., and Squires, N. (2017). Sustainable Development Goals (SDGs) and their implementation a national global framework for health, development and equity needs a systems approach at every level. *British Medical Bulletin*, 124, pp. 1-10.