ISSN 2085-4552

Public Sentiment Analysis on the Transition from Analog to Digital Television Using the Random Forest Classifier Algorithm

Elfajar Bintang Samudera[1], Alexander Waworuntu[2], Ester Lumba[3]
[1,2] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]elfajar.bintang@student.umn.ac.id, [2]alex.wawo@umn.ac.id
[3] Department of Informatics, Bunda Mulia University, Jakarta, Indonesia
[3]l0178@lecturer.ubm.ac.id

Accepted 3 June 2024
Approved 26 June 2024

***Abstract*— Television is one of the most popular media for entertainment and information. Analog television is the most widely used type among the public. However, with technological advancements, analog television is becoming obsolete and is being replaced by digital television, which offers better performance. On November 2, 2022, the Government officially mandated the transition from analog to digital broadcasting. This Analog Switch Off program has elicited various pro and con opinions among the public. X, a widely used social media platform, facilitates rapid communication and information dissemination among users. This study aims to classify public sentiment regarding the Analog Switch Off policy as either positive or negative. The classification model used is the Random Forest algorithm, with the Lexicon Inset for data labeling, Count Vectorizer and TF-IDF Vectorizer for data vectorization and weighting, and various train-test data splits. The study achieved the best classification performance using the Count Vectorizer method, with an 80%:20% train-test data ratio, yielding an accuracy of 88%, precision of 88%, recall of 88%, and an F1-score of 88%.**

***Index Terms*— Analog Television; Digital Television; Sentiment; Random Forest; X**.

I. Introduction

Television is a widely favored medium for entertainment and information. Its audiovisual nature enables it to offer various forms of entertainment such as movies, music, reality shows, sports, soap operas, and celebrity-related programs. In Indonesia, television is the most popular medium and serves as the primary promotional tool for industries to market products and services [1].

One of the oldest types of television among the public is analog television. Analog television uses analog signals to transmit images and sound, broadcasting a range of voltages and signal frequencies [1]. It requires a signal receiver called an antenna. The initial development of analog television involved the use of a disc with specific patterns for scanning images, known as mechanical television. Analog television programs are broadcasted by various national stations and are available for free.

Digital television, on the other hand, operates using digital modulation, where the carrier wave's properties are modified into bits (0 or 1) [2]. Digital television offers much higher resolution compared to analog television and provides more efficient use of the radio frequency spectrum. Digital television reception is also versatile, as it can receive broadcasts from regular transmission stations, regular antennas, cable television, and satellite television [3].

On November 2, 2022, the government mandated the migration from analog to digital television broadcasting. The cessation of analog television broadcasting is part of the digital transformation in the broadcasting sector, in accordance with the mandate of Law No. 11 of 2020 (Undang-undang No. 11 Tahun 2020 tentang Cipta Kerja). The implementation of this decision considers various benefits, such as reducing broadcast traffic congestion, optimizing spectrum usage, more efficient utilization of frequency resources, and providing more frequency space to accelerate internet access in Indonesia [1].

The implementation of the Analog Switch Off program across various regions has elicited a range of public opinions, both supportive and opposing. Director General Usman Kansong stated that transitioning to digital television offers benefits such as clearer pictures, better sound quality, and more advanced technology [1]. Conversely, Nailul Huda, a Digital Economy Analyst from the Institute for Development of Economics and Finance (INDEF), pointed out that switching to digital television could negatively impact industries by causing a loss of market share, affecting advertisements that would not reach those still using analog television [1]. Given these diverse sentiments, sentiment analysis plays a crucial role in this research. Sentiment analysis is the process of processing data to track public responses to a specific topic on the internet. With the advancement of information technology parallel to the broadcasting sector, platforms like X are used for public responses regarding the transition from analog to digital television.

Sentiment analysis requires a method, and one frequently used is the Random Forest Classifier. The Random Forest Classifier is employed for classification processes. It works by constructing multiple decision trees and combining them to produce stable and accurate predictions [4]. The advantages of the Random Forest Classifier include its ability to handle noise and its suitability for classifying large datasets [4].

Several previous studies relevant to this research include Aisyah Nurul Izza et al.'s analysis of tourist attraction sentiments in South Sulawesi Province based on visitor reviews using the Random Forest Classifier method, which achieved an accuracy of 82% [5]. Hana Chyntis Morama et al. analyzed aspect-based sentiments regarding Hotel Tenrem Yogyakarta reviews using the Random Forest Classifier algorithm, achieving an accuracy and F1-score of 90% [6]. Evita Fitri et al. analyzed sentiments towards the Ruangguru application using the naive bayes, random forest, and support vector machine algorithms, achieving an accuracy of 97.16% and an AUC value of 0.996, with the best accuracy and performance obtained using the Random Forest algorithm [7]. Therefore, this research will analyze public sentiment regarding the transition from analog to digital television using data from X and the Random Forest Classifier algorithm. The aim of this research is to understand public sentiment towards the transition from analog to digital television and to determine the accuracy, precision, recall, and F1-score values produced by using the Random Forest Classifier algorithm on X data.

II. Theoretical Framework

A. Sentiment Analysis

Sentiment analysis is the process of determining the sentiment of a text and categorizing it as either positive or negative [8]. It is often equated with opinion mining [9] because it focuses on opinions that express either positive or negative sentiments. Textual data such as product reviews, services, phenomena, and individuals can be the subject of sentiment analysis research [10].

B. Analog Switch Off

Analog Switch Off (ASO) is a policy that mandates the migration from analog to digital television broadcasting [3]. This policy aligns with technological and informational advancements in Indonesia. Government Regulation No. 46 of 2021 concerning Post, Telecommunications, and Broadcasting, in Article 72, paragraph 8, states that the migration from analog to digital terrestrial broadcasting must be completed no later than two years after its enactment. Consequently, the ASO policy was to be implemented by November 2, 2022 [2].

C. X

X is a social media platform that allows users to write and publish short, concise texts expressing their opinions [11]. In this research, the snscrape library will be used. Snscrape is a scraping tool for social networking services (SNS). This library can scrape data such as users, user profiles, hashtags, searches, and posts without using the X API.

D. Text Preprocessing

Text preprocessing is a crucial step aimed at preventing significant degradation in the performance of analysis [12]. Generally, text preprocessing is divided into several stages, such as data cleaning, data labeling, case folding, stemming, and tokenizing [13]. The details of these stages are as follows:

1) Data Cleaning: In this stage, the data is cleaned by removing characters such as symbols. Additionally, emojis, URLs, and usernames are also removed. This step aims to reduce and avoid disruptions in the classification results [14].
2) Data Labeling: In this stage, a process is conducted to label the tweets. This is done using existing libraries to classify the data into positive or negative sentiments.
3) Case Folding: This stage involves converting text into a uniform format, specifically by converting all text to lowercase [14].
4) Stemming: Stemming is the process of reducing words to their base form, known as the stem. This process helps in reducing the vocabulary size that the NLP model needs to process, thereby enhancing the model's efficiency [14].
5) Tokenizing: In this stage, sentences are broken down into individual words, known as tokens [14].

E. Decision Tree

The Decision Tree is a popular classification method due to its ease of interpretation by humans. It employs a tree or hierarchical structure to create a predictive model [15]. The concept behind this algorithm is to transform data into a decision tree and decision rules. It simplifies complex decision-making processes into a more straightforward form.

The Decision Tree is named so because the rules formed resemble a tree structure. The data within the decision tree are interpreted in a table format with attributes and records. The Decision Tree has nodes representing attributes, where each branch represents the result of a test, and leaf nodes represent the classes [16]. The construction of a decision tree involves three types of nodes: the root node, internal nodes, and leaf nodes. The root node is the top node with no input and more than one output. Internal nodes are branching nodes with one input and at least two outputs. Leaf nodes are the terminal nodes with one input and no output.

The decision tree construction starts by calculating the entropy to determine the impurity of the attributes and the information gain [17].The formula for calculating entropy is: (1)

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

where ***pi*** represents the proportion of samples in subset ***S*** with the ***i***-th attribute value. Information gain is a metric used in the segmentation process [18]. The formula for calculating information gain is: (2)

$$\text{Gain } (S, A) = \sum_{V \in V(A)} \frac{|S_V|}{|S|} \text{Entropy } (S_V) \quad (2)$$

where ***V(A)*** is the range of attribute ***A***, and S_v is the subset of ***S*** that has the same value as attribute ***V***.

F. Random Forest Classifier

The Random Forest Classifier is an algorithm used for classifying large datasets. It combines multiple decision trees into a single model, enhancing accuracy with an increasing number of trees [5]. Each tree in the Random Forest model contains a collection of random variables, and the aggregation of these trees leads to improved accuracy.

The Random Forest algorithm combines each decision tree model into one comprehensive model. The number of trees used significantly impacts the accuracy, precision, and recall of the Random Forest model. The selection of trees from the decision tree model begins with calculating the entropy value to find the best trees for use in the Random Forest model [19].

The steps involved in the Random Forest algorithm begin with selecting random samples from the provided dataset. For each selected sample, a decision tree is constructed, and predictions are obtained from these decision trees. A voting process is then conducted for each prediction, where the most frequent value (mode) is used for classification problems, and the average value (mean) is used for regression problems. Finally, the algorithm determines the final prediction by selecting the prediction with the majority vote.

G. Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) is a technique used to assign a weight or value to a word within a document [20]. TF-IDF can be applied to various tasks, such as token extraction from articles, ranking determination, and calculating the similarity between documents. Term Frequency (TF) indicates how often a word appears in a document [20], while Inverse Document Frequency (IDF) signifies the importance of a word [20]. The formula for calculating TF-IDF is as follows [21]:

$$TF - \text{IDF}_{t,d} = \text{TF}_{t,d} \times iDF_t \quad (3)$$

$$IDF_t = \log \frac{N}{DF_t} \quad (4)$$

$$TF_{t,d} = \frac{n_{i,j}}{\sum_k n_{i,j}} \quad (5)$$

where:

- $\boldsymbol{TF\text{-}IDF_{t,d}}$ represents the weight of a word (***t***) in a document (***d***).
- $\boldsymbol{TF_{t,d}}$ is the Term Frequency.
- $\boldsymbol{IDF_t}$ is the Inverse Document Frequency.
- ***DFt*** is the number of documents containing the word.
- $\boldsymbol{n_{i,j}}$ is the total occurrences of the term in one document.
- $\boldsymbol{\sum_k n_{i,j}}$ is the total number of words in one document.
- ***N*** is the total number of documents.
- ***t*** denotes the words being calculated.
- ***d*** is the document weight.

H. Hyperparameter Tuning

Hyperparameter tuning is the process of finding the optimal combination of parameters for a machine learning model [22]. The goal of this process is to identify the best hyperparameter combination to enhance performance and reduce the risk of overfitting and underfitting [22]. In this study, the following parameters will be optimized using the random search method:

1) **n_estimators:** This parameter refers to the number of decision trees to be created by the algorithm. A higher value of n_estimators results in more trees being formed [22].
2) **min_samples_split:** This parameter indicates the minimum number of samples required to split an internal node in a tree [22].
3) **min_samples_leaf:** This parameter specifies the minimum number of samples that must be present in a leaf node [22].
4) **max_depth:** This parameter denotes the maximum depth of each decision tree within the ensemble method [22].

5) **max_features:** This parameter controls the number of features randomly selected to build each decision tree [22].

6) **bootstrap:** This parameter manages the use of bootstrapping when constructing each decision tree. If set to True, each tree is built using bootstrap samples. If set to False, each tree is constructed using the original training dataset [22].

I. Confusion Matrix

The Confusion Matrix is a method used to calculate the accuracy of a classification model. It is represented as a table that shows the number of correct and incorrect classifications for the test data [10]. Several metrics can be derived from the Confusion Matrix to evaluate a classification model, including accuracy, recall, precision, and F1-score [4].

Accuracy represents the percentage of correctly classified tuples in the test data. It is calculated using the formula:

$$\text{accuracy} = \frac{TP+T}{TP+TN+FP+F} \quad (6)$$

Recall indicates the rate at which positive tuples are correctly identified, measuring how well the model identifies true positives. It is calculated using the formula:

$$\text{recall} = \frac{TP}{TP+FN} \quad (7)$$

Precision represents the percentage of tuples labeled as positive that are actually positive. It is calculated using the formula:

$$\text{precision} = \frac{TP}{TP+FP} \quad (8)$$

The F1-score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is calculated using the formula:

$$f1 - \text{score} = \frac{2x \text{ Precision } x \text{ Recall}}{\text{Recall + Precision}} \quad (9)$$

III. RESEARCH METHODOLOGY

This study involves several methodologies and processes, as outlined below:

1) **Literature Review:** The literature review process begins with studying relevant literature and theories pertinent to this research. The theories covered include sentiment analysis, analog switch off, X, text preprocessing, decision tree, random forest, Term Frequency-Inverse Document Frequency (TF-IDF), and confusion matrix.

2) **Data Collection:** In this stage, data collection involves gathering tweets from Indonesian-speaking users about their opinions on the transition from analog to digital television. The collected data will form a dataset that will be processed in subsequent stages. Data collection is conducted using the snscrape library. Figure 1 illustrates the data collection process in a flowchart format.

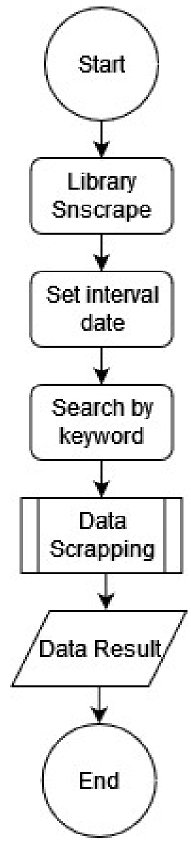

Fig. 1. Data Scraping Flowchart

3) **Data Processing:** This stage, also known as text preprocessing, involves several critical steps. First, data cleaning is performed to remove non-alphabet characters, symbols, emoticons, URLs, and usernames to prevent disruptions in classification results. Next, casefolding converts all words and sentences to lowercase, ensuring a uniform text format and eliminating any capitalized words or sentences. Tokenizing then segments the text into smaller parts, or tokens. In the stopword removal step, commonly used words that have little impact on the sentence's meaning, such as "and," "or," and "which," are removed. Normalizing involves

identifying and replacing incorrect word forms with their correct versions according to the Kamus Besar Bahasa Indonesia (KBBI), ensuring all spellings and abbreviations match KBBI standards. Finally, stemming processes words with affixes to convert them to their root forms by removing the affixes. Figure 2 illustrates these data processing steps in a flowchart format.

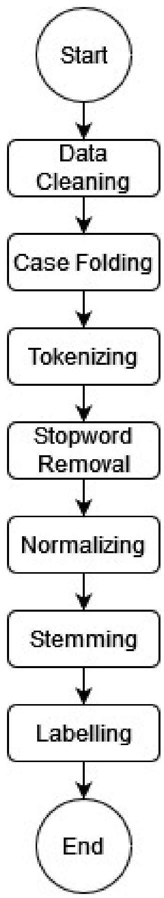

Fig. 2. Preprocessing Flowchart

4) **Applying TF-IDF Features:** TF-IDF is used for various tasks such as token extraction from articles, ranking, and calculating document similarity. Term Frequency (TF) indicates how often a word appears in a document [20], while Inverse Document Frequency (IDF) indicates the importance of a word [20]. Figure 3 illustrates the process of applying TF-IDF features in a flowchart format.

GAMBAR

In addition to TF-IDF vectorization, this research also employs the Count Vectorizer method. The difference between these methods lies in their output and process. Count Vectorizer uses the concept of term frequency alone, where the weight of each word appearing in a document is calculated, resulting in a bag-of-words model [23]. Unlike TF-IDF vectorization, Count Vectorizer does not calculate the IDF score for each feature, thus not assessing the importance of a feature within a document.

5) **Implementation of the Random Forest Classifier Algorithm:** In this stage, a model is created using data that has undergone the text preprocessing steps. The output from text preprocessing is processed using the Random Forest algorithm. The Random Forest model is trained to provide classification predictions for positive and negative sentiments.

6) **Evaluation Testing:** This stage involves testing the model by measuring the values of accuracy, recall, precision, and F1-score. These measurements are obtained by evaluating the results of the Random Forest model using a confusion matrix.

IV. Results and Discussion

A. Testing Scenario

In the model testing phase using Random Forest, various scenarios were implemented. These scenarios included the use of Count Vectorizer and TF-IDF Vectorizer, testing the model on different dataset sizes split with test sizes of 20%, 30%, 40%, and 50%, and employing hyperparameter tuning with the random search method to identify optimal parameters. The various scenarios implemented in the machine learning model are presented in Table I. After multiple processes and testing scenarios, the evaluation results for the best model performance for each vectorization method are shown in Table II.

B. Discussion

Based on the evaluation results in the table above, the model that produced the best performance was in the first scenario, where the training and testing data split ratio was 80%:20%, and the text data vectorization method used was Count Vectorizer. This model achieved an accuracy of 88%, with identical values for recall, precision, and F1-score. As the dataset's split ratio for training and testing data was adjusted, the accuracy significantly decreased, while the other metrics also declined, though less dramatically. The lowest performance for the Count Vectorizer method occurred in the eighth scenario, with a dataset split ratio of 50%:50% and hyperparameter tuning using random search. Despite the tuning, the results were not as good as the previous models due to the random nature of parameter selection, which did not consider inter-parameter relationships.

In the scenarios using the TF-IDF Vectorizer method, the best model performance was obtained in the tenth scenario. Here, the training and testing data split ratio was 80%:20%, with fine-tuning involving

600 trees, a minimum of 10 samples required for splitting, a minimum of 1 sample per leaf, automatic feature selection for splitting, a maximum tree depth of 90, and each tree built using the entire training dataset without bootstrap sampling. This model achieved an accuracy of 87.75%, with recall, precision, and F1-score values of 88%. The worst performance for the TF-IDF Vectorizer method occurred in the fifteenth scenario, with a training and testing data split ratio of 50%:50%. Despite using two different text data vectorization methods, the worst performance occurred with a split ratio of 50%:50% for training and testing data.

TABLE I. TABLE OF MACHINE LEARNING SCENARIO TEST

No.	Scenario
1	CountVectorizer()
	test_size 20%
2	CountVectorizer()
	test size 20% + Hyperparameter Tuning
3	CountVectorizer()
	test size 30%
4	CountVectorizer()
	test size 30% + Hyperparameter Tuning
5	CountVectorizer()
	test size 40%
6	CountVectorizer()
	test size 40% + Hyperparameter Tuning
7	CountVectorizer()
	test size 50%
8	CountVectorizer()
	test size 50% + Hyperparameter Tuning
9	TFIDFVectorizer()
	test size 20%
10	TFIDFVectorizer()
	test size 20% + Hyperparameter Tuning
11	TFIDFVectorizer()
	test size 30%
12	TFIDFVectorizer()
	test size 30% + Hyperparameter Tuning
13	TFIDFVectorizer()
	test size 40%
14	TFIDFVectorizer()
	test size 40% + Hyperparameter Tuning
15	TFIDFVectorizer()
	test size 50%
16	TFIDFVectorizer()
	test size 50% + Hyperparameter Tuning

TABLE II. TABLE OF TEST RESULT WITH HIGHEST AND LOWEST PERFORMANCE FOR EACH VECTORIZATION METHOD

Model	S	A	P	R	F1
CountVectorizer 80%:20%	1	88%	88%	88%	88%
CountVectorizer 50%:50%	8	80.68%	81%	81%	81%
TFIDF80%:20% Hyperparameter Tuning	10	87.75%	88%	88%	88%
TF-IDF 50%:50%	15	81.78%	82%	82%	82%

Legend:
S: Scenario; A: Accuracy; P: Precision; R: Recall; F1: F1 Score

V. CONCLUSION

Based on the research conducted on sentiment analysis using the Random Forest algorithm, several conclusions can be drawn. The sentiment analysis of public opinion regarding the transition from analog to digital television using the Random Forest algorithm was successfully implemented. This study utilized Count Vectorizer and TF-IDF methods for word weighting, and Random Search for hyperparameter tuning. The machine learning model that achieved the highest accuracy, precision, recall, and F1-score was the one using Count Vectorizer for word weighting, with a train-test split ratio of 80%:20%, without undergoing hyperparameter tuning. The results obtained were 88.00% accuracy, 88% precision, 88% recall, and 88% F1-score. Model performance significantly decreased, particularly in terms of accuracy, with an increase in test data size. The lowest performance was observed in the model using the Count Vectorizer method, with a train-test split ratio of 50%:50% that had undergone hyperparameter tuning. The results were 80.60% accuracy, 81% precision, 81% recall, and 81% F1-score.

REFERENCES

[1] K. Alfarizi, "Siaran digital indonesia — gugus tugas migrasi siaran televisi analog ke digital," Nov. 2022.

[2] A. D. Gultom, "Digitalisasi penyiaran televisi di indonesia," Buletin Pos dan Telekomunikasi, vol. 16, pp. 91–100, Dec. 2018. [Online]. Available: https://bpostel.kominfo.go.id/index.php/bpostel/article/view/160202

[3] M. Mubarok and M. D. Adnjani, "Kesiapan industri tv lokal di jawa tengah menuju migrasi penyiaran dari analog ke digital," Communicare: Journal of Communication Studies, vol. 7, no. 1, pp. 18–32, 2020.

[4] M. Y. Aldean, P. Paradise, and N. A. S. Nugraha, "Analisis sentimen masyarakat terhadap vaksinasi covid-19 di twitter menggunakan metode random forest classifier (studi kasus: Vaksin sinovac)," Journal of Informatics Information System Software Engineering and Applications (INISTA), vol. 4, pp. 64–72, Jun. 2022. [Online]. Available: https://journal.ittelkom-pwt.ac.id/index.php/inista/article/view/575

[5] A. N. Izza, D. E. Ratnawati, W. Hayuhardhika, N. Putra, and P. Korespondensi, "Analisis sentimen objek wisata di provinsi sulawesi selatan berdasarkan ulasan pengunjung menggunakan metode random forest classifier," Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi, vol. 3, pp. 97–105, Oct. 2022. [Online]. Available: https://just-si.ub.ac.id/index.php/just-si/article/view/77

[6] H. C. Morama, D. E. Ratnawati, and I. Arwani, "Analisis sentimen berbasis aspek terhadap ulasan hotel tentrem yogyakarta menggunakan algoritma random forest classifier," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2548, p. 964X, 2022.

[7] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis sentimen terhadap aplikasi ruangguru menggunakan algoritma naive bayes, random forest dan support vector machine," Jurnal Transformatika, vol. 18, pp. 71–80, Jul. 2020. [Online]. Available: https://journals.usm.ac.id/index.php/transformatika/article/view/2317

[8] A. Perdana, A. Hermawan, and D. Avianto, "Analisis sentimen terhadap isu penundaan pemilu di twitter menggunakan naive bayes classifier," Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 11, pp. 195–200, Jul. 2022. [Online].

Available: http://jurnal.atmaluhur.ac.id/index.php/sisfokom/article/view/1412

[9] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, 2012.

[10] A. M. Zuhdi, E. Utami, and S. Raharjo, "Analisis sentiment twitter terhadap capres indonesia 2019 dengan metode k-nn," Jurnal Informa: Jurnal Penelitian dan Pengabdian Masyarakat, vol. 5, pp. 1–7, Aug. 2019. [Online]. Available: http://www.poltekindonusa.ac.id/SUBDOMAIN/informa/index.php/informa/article/view/73

[11] A. K. Fauziyyah and D. H. Gautama, "Analisis sentimen pandemi covid19 pada streaming twitter dengan text mining python," Jurnal Ilmiah SINUS, vol. 18, pp. 31–42, Jul. 2020. [Online]. Available: https://p3m.sinus.ac.id/jurnal/index.php/ejurnal SINUS/article/view/491

[12] M. Syarifuddinn, "Analisis sentimen opini publik terhadap efek psbb pada twitter dengan algoritma decision tree, knn, dan naive bayes," INTI Nusa Mandiri, vol. 15, pp. 87–94, Aug. 2020. [Online]. Available: http://ejournal.nusamandiri.ac.id/index.php/inti/article/view/1433

[13] R. T. Vulandari, "Data mining: Teori dan aplikasi rapidminer," 2017.

[14] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan algoritma naive bayes untuk analisis sentimen review data twitter bmkg nasional," Jurnal Tekno Kompak, vol. 15, pp. 131–145, Feb. 2021. [Online]. Available: https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/744

[15] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi dengan pohon keputusan berbasis algoritme c4.5," in PRISMA, Prosiding Seminar Nasional Matematika, vol. 3, 2020, pp. 64–71.

[16] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Perbandingan algoritma random forest, naive bayes, dan support vector machine pada analisis sentimen twitter mengenai opini masyarakat terhadap penghapusan tenaga honorer," Seminar Nasional Official Statistics, vol. 2022, pp. 563–572, Nov. 2022. [Online]. Available: https://prosiding.stis.ac.id/index.php/semnasoffstat/article/view/1410

[17] B. B. Baskoro, I. Susanto, S. Khomsah, P. Informatika, P. S. Data, I. D. T. T. P. J. Panjaitan, and J. Tengah, "Analisis sentimen pelanggan hotel di purwokerto menggunakan metode random forest dan tf-idf (studi kasus: Ulasan pelanggan pada situs tripadvisor)," Journal of Informatics Information System Software Engineering and Applications (INISTA), vol. 3, pp. 21–29, Jun. 2021. [Online]. Available: https://journal.ittelkom-pwt.ac.id/index.php/inista/article/view/218

[18] B. T. Jijo and A. M. Abdulazeez, "Classification based on decision tree algorithm for machine learning," vol. 02, pp. 20–28, 2021. [Online]. Available: www.jastt.org

[19] M. Huljanah, Z. Rustam, S. Utama, and T. Siswantining, "Feature selection using random forest classifier for predicting prostate cancer," IOP Conference Series: Materials Science and Engineering, vol. 546, p. 052031, Jun. 2019. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052031 https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052031/meta

[20] Imamah and F. H. Rachman, "Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regression," 2020 Information Technology International Seminar (ITIS), pp. 238–242, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9320958

[21] M. Guia, R. R. Silva, and J. Bernardino, "Comparison of Naive Bayes, Support Vector Machine, Decision Trees and Random Forest on Sentiment Analysis," 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), pp. 525–531, 2019.

[22] R. Ahuja, K. Vats, C. Pahuja, T. Ahuja, and C. Gupta, "Pragmatic analysis of classification techniques based on hyper-parameter tuning for sentiment analysis," EasyChair, Tech. Rep., 2020.

[23] H. P. Doloksaribu and Y. T. Samuel, "Komparasi algoritma data mining untuk analisis sentimen aplikasi pedulilindungi," Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika, vol. 16, no. 1, pp. 1–11, 2022.