khazanah informatika

Jurnal Ilmu Komputer dan Informatika

# Optimization of Sentiment Analysis of Government Regulation in Lieu of Law on Job Creation Using KNN, Random Forest, and PSO

**Tupari[1], Sri Lestari*[2], Yan Aditiya Pratama[3], Suhendro[4]**

[1, 2,4] Faculty of Computer Science, Master of Informatics Engineering
Institute of Informatics and Business Darmajaya,
Bandar Lampung
[3]Faculty of Economics and Business, Manajemen
Institute of Informatics and Business Darmajaya
Bandar Lampung
*korespondensi: [2]srilestari@darmajaya.ac.id

**Abstract-** Twitter is one of the social media used by the public to convey their views regarding the government's policy of issuing a Government Regulations in Lieu of Laws (*Bahasa: Peraturan Pemerintah Pengganti Undang-undang (Perpu)*). The public's pros and cons of this policy are material for sentiment analysis. The purpose of this study was to analyze Twitter users' opinions regarding the Job Creation Perpu using the K-Nearest Neighbors (KNN), Random Forest (RF), and Particle Swarm Optimization (PSO) methods. The data was 3.128 tweets from Twitter social media users regarding the Government Regulation in Lieu of Law on Job Creation. Based on 3.128 data, 1.599 sentiments were positive, 1.473 sentiments were negative and 53 sentiments were neutral. The results showed that PSO feature optimized Twitter social media sentiment analysis against this regulation. KNN and RF algorithms for sentiment analysis was carried out before and after optimization with PSO. Experimental results using RapidMiner 9.10 showed that PSO feature succeeded in increasing classification accuracy in both algorithms. Before optimization, the KNN accuracy value reached 80.40%, then increased significantly to 85.23% after optimization with PSO was applied. Meanwhile, Random Forest accuracy value before optimization was 77.21% and increased to 80.53% after PSO was applied. This result indicated that the PSO-based KNN algorithm had better performance in conducting sentiment analysis of the Government Regulation in Lieu of Law on Job Creation on Twitter compared to the Random Forest algorithm in the context of this study. It concluded that Random Forest algorithm based on PSO is the best classifier for sentiment analysis and a potential and effective algorithm for classifying and analyzing sentiment on the same topic.

**Kata Kunci**: *Sentiment Analysis, Government Regulation in Lieu of Law on Job Creation, K-Nearest Neighbors, Random Forest, Particle Swarm Optimization*

Article info: *submitted November 16, 2023, revised April 7, 2025, accepted July 23, 2025*

## 1. Introduction

The Government Regulations in Lieu of Laws (*Bahasa: Peraturan Pemerintah Pengganti Undang-undang (Perpu)*) Number 2 of 2022 concerning Job Creation was issued to address the problem of unemployment and increase economic competitiveness. The Job Creation Perpu provides rules related to the provision of manpower, taxation, licensing, and protection of workers' rights. The purpose of Government Regulations in Lieu of Laws on Job Creation is to encourage the country's economic recovery and economic growth that is inclusive and sustainable. In addition, the Government Regulations in Lieu of Laws on Job Creation can also reduce investment barriers, namely licensing issues and increase labor efficiency.

The Government Regulations in Lieu of Laws on Job Creation has become controversial because there are articles that are considered detrimental to workers and laborers. There are those who support it and some who oppose it, which is reflected in the discussions and responses from the public on social media such as Twitter. Some people believe that the Government Regulations in Lieu of Laws on Job Creation is detrimental to workers, contract employees, and so on. A number of demonstrations have also taken place as a form of rejection of this Government Regulations in Lieu of Laws, including a demo announced via the Twitter account @tempodotco. However, there are also members of the public who welcome the Government Regulations in Lieu of Laws on Job Creation, as seen from the support provided by Gajah Duduk's Twitter Account.

Social media is an easy means for people to express opinions and demonstrate. Indonesian people can choose social media in Indonesia which has a variety of functions to the characteristics of each of these social media. For example, Twitter and Facebook,

both of which have advantages in conveying information or opinions, Twitter with the trending topic feature and Facebook with its posting feed [1].

On social media Twitter users are free to issue opinions or opinions [2]. Users can tweet up to 140 characters which makes it unique. Twitter is one of the social media platforms that is widely used by the public to convey expressions regarding the Government Regulations in Lieu of Laws on Job Creation. Therefore, it is important to carry out a sentiment analysis of the Government Regulations in Lieu of Laws on Job Creation discussed on the Twitter social media so that it becomes meaningful information. Unstructured data is processed using certain techniques and methods so that it becomes useful news for users.

The selection of social media Twitter for data mining due to three main points namely: a) Twitter API has a good design and is easy to access, b) Twitter data is available in a convenient format for analysis, and c) Twitter's policy for data is relatively liberal compared to other APIs. In addition, Twitter has the most data compared to other social media, so it can help in terms of the accuracy of applying the sentiment analysis method [3].

Twitter sentiment analysis has received increasing attention in recent years [4]. Sentiment analysis has been widely discussed since 2002 [5]. Sentiment analysis is the process of understanding, extracting, and processing opinions from textual data automatically to obtain positive and negative sentiment information in an opinion sentence [6]. Sentiment analysis is an important field that allows us to understand the general attitude of users about certain topics [7].

Sentiment analysis in this study is the process of understanding and classifying emotions (positive, negative, and neutral) contained in writing using text analysis techniques used to determine public opinion on various topics [8], including this Government Regulations in Lieu of Laws on Job Creation. Sentiment is the personal opinion of internet users to provide an assessment of something. With so many opinions expressed, of course it will be difficult to determine sentiment on the topics discussed.

Sentiment analysis plays a role in classifying text polarity based on sentiment [8]. This is consistent with patterns in data mining, namely the study of knowledge extraction and pattern recognition in data. The purpose of this sentiment analysis is to provide insight. The purpose of sentiment analysis is to provide insight regarding how the Indonesian people respond to government policies related to employment and the economy regulated by the Government Regulations in Lieu of Laws on Job Creation. The results of sentiment analysis can provide valuable information and contributions to the government in improving or improving policies that have been taken previously, so that they can strengthen existing policies or even take more appropriate steps for further policies.

This study sees this problem as interesting to study in order to find out the number of pros and cons of sentiment issued by people who express their opinions via Twitter. Moreover, this study wants to classify the sentiments issued by the public towards the Government Regulations in Lieu of Laws on Job Creation so that it becomes meaningful information using KNN and RF algorithm based on PSO.

The state of the art in this study is the analysis of sentiment towards Government Regulations in Lieu of Laws on Job Creation on social media Twitter using KNN and RF algorithm based on PSO. This study aims to analyze public sentiment towards the Government Regulations in Lieu of Laws on Job Creation and predict whether a tweet contains positive, negative, or neutral sentiments towards the regulation.

Based the previous study, the novelty of this study proposes that KNN and RF algorithm based on PSO which have not been widely used in sentiment analysis on social media. This algorithm is able to optimize the parameters of the KNN and RF algorithms so that it can increase accuracy in sentiment analysis. PSO techniques for optimization include increasing the attribute weight of all attributes or variables used, selecting attributes and feature selection [9] [10].

This study proposes the use of a combination of KNN and Random Forest (RF) algorithms optimized using Particle Swarm Optimization (PSO) for sentiment analysis of the Government Regulation in Lieu of Law (Perppu) on Job Creation. The selection of this approach is based on the unique characteristics of each algorithm and the need to achieve optimal sentiment classification accuracy.

KNN is chosen for its simplicity and its ability to classify data based on feature proximity. In the context of sentiment analysis, KNN classifies a new tweet by identifying the 'k' nearest tweets in the training dataset and assigning the majority sentiment of those neighbors [11]. One of the main advantages of KNN lies in its non-parametric nature, meaning that it does not assume any specific data distribution, making it flexible for various types of complex text data [12]. Additionally, KNN has been proven effective in various text classification tasks [13] [12]. However, the performance of KNN heavily depends on the proper selection of the value of 'k' and can be affected by irrelevant features or high-dimensional data, which are common challenges in text data.

Meanwhile, Random Forest (RF) is an ensemble learning algorithm that builds a large number of decision trees and combines their predictions [14]. The advantage of RF lies in its ability to handle high-dimensional data, which is particularly relevant for text data with numerous features (e.g., extracted text features) [1] [15]. RF is known to be highly robust against overfitting due to its use of bagging (bootstrap aggregating) and random feature selection at each tree, making it a strong choice for diverse datasets [16]. This algorithm can also provide insights into feature importance, which helps in identifying the most influential words or phrases in sentiment classification [17].

RF was first introduced by Breiman to solve regression, unsupervised learning, and classification problems and has been successfully applied in many domains including geotechnical engineering [18], which further demonstrates its versatility.

Nevertheless, achieving optimal RF performance often requires careful tuning of hyperparameters such as the number of trees and maximum depth, and this is often done using optimization algorithms [18].

To overcome the limitations in determining optimal parameters in KNN and RF, this study integrates PSO. PSO is a

population-based optimization algorithm inspired by the social behavior of bird flocks or fish schools [19] [20] [15]. In this context, PSO acts as an automatic hyperparameter tuning mechanism for KNN and RF. Previous studies have shown that the integration of PSO - RF significantly improves classification performance across various domains, including image classification and chronic disease diagnosis [18].

The novelty of this study states in the object studied, namely the Government Regulations in Lieu of Laws on Job Creation. It meant that it is a regulation just published at the end of 2022 and it has been a controversial topic among the public. This topic has not been widely studied by the previous study. Thus, it hoped that this study can make a new contribution in sentiment analysis on social media and help the government understand public sentiment towards public policies.

The integration of PSO with KNN and RF represents a significant novelty in this study. Many sentiment analysis studies have utilized KNN or RF independently or with manual optimization. However, the application of PSO to automatically optimize both algorithms in the context of sentiment analysis on the Omnibus Law (Bahasa: Perpu Cipta Kerja) is a relatively unexplored approach. This method is expected to produce a more accurate and robust sentiment classification model, offering a novel contribution to the effort of understanding public opinion on government policies through social media.

## 2. Method

The result of study [21] stated that The RF algorithm produces an accuracy value of 98% and a Kappa value of 0.96 in classifying plant health levels. It concluded that the model developed and the algorithm used were considered effective in classifying plant health levels.

This study used several stages as shown in Figure 1 below.

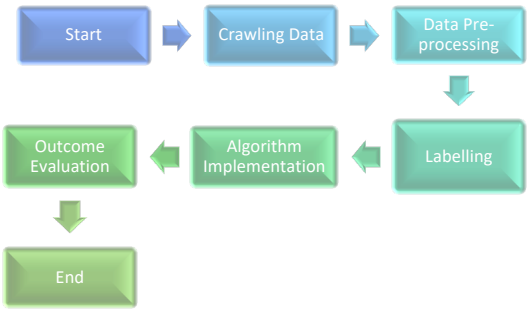



**Figure 1. Research Stages**

Based on Figure 1, it was explained as follows.

2.1. Crawling Data

The data in this study was tweets discussing the Government Regulations in Lieu of Laws on Job Creation on Twitter. The data collected includes uploads, comments or responses related to these regulations. Data collection was carried out using the Twitter API with keywords related to the Government Regulations in Lieu of Laws on Job Creation. The tools used are RapidMiner version 9.10.

2.2. Data Pre-Processing

Crawling opinion data from the Twitter platform must avoid standard words, dictionary words, or phrases in the local language, and must not remove these opinions. In sentiment analysis, non-standard words affect the calculation of data analysis [22] Therefore, data preprocessing techniques was needed. Data preprocessing was a series of techniques and processes used to modify and prepare raw data before it was entered into a data analysis model or algorithm. The purpose of data preprocessing was to clean, integrate, transform, reduce dimensions, and convert raw data into more suitable and easier format to process by data analysis algorithms.

In the classification of news for the data type in the form of text, there was several types of processes carried out, including case folding, removing punctuation, tokenization, lemmatization, and stop word removal [23] [3]. This step was done to clean the data from noise and prepare it for sentiment analysis. The steps in data were included [22] [24]: 1) data cleaning, 2) case folding, 3) normalization, 4) filtering, 5) stemming, and 6) tokenizing. Data cleaning was a step to remove data that is irrelevant, duplicate, or contains errors such as strange characters, invalid symbols or numbers [22].

Case folding was changing all letters to lowercase or uppercase so it did not affect the analysis due to differences in capitalization. Normalization was changing words or phrases that had different forms but it had the same meaning into the same form. For example, the words "menulis" and "menuliskan" are changed to "menulis". Filtering was removing irrelevant words or considered as "stop words" such as prepositions, conjunctions, or other common words that had no significant meaning in the analysis. Meanwhile, stemming was changing words into basic words or base words. For example, the words "menulis", "menulisi", and "menuliskan" are changed to "tulis". At the end, Tokenizing was separating text into words or tokens to facilitate analysis..

2.3. Labelling

The labeling stage was giving positive, negative, or neutral labels to each text data based on the sentiments contained in the text. This stage was done by reading each tweet and determining whether the sentiment in the tweet was positive (contains praise), negative (contains criticism), or neutral (does not contain strong sentiments).

This labeling stage was important for building datasets to be used in machine learning, because the model would not be able to learn and produce accurate predictions without the right labels. The dataset was created by dividing the processed tweets into training data and testing data. Training data was used to train the model and data testing was used to test the model's performance.

After the dataset was ready, the next step was to extract features from the dataset in the form of a numeric vector so that it was used by the KNN and RF algorithms. The features were extracted the words that appeared most frequently in the dataset. Feature extraction was performed using the TF-IDF method to convert text into a numerical representation by using algorithm.

2.4. Implementation of Algorithm

Before the model was implemented using the PSO-based KNN and RF algorithms, model training was carried out so that

the results were optimal. The KNN algorithm was used to classify sentiments based on nearest neighbors, while RF combined several decision trees.

This step was carried out to optimize the parameters of the KNN and RF algorithms using the PSO algorithm. The KNN algorithm was a classification method based on the concept of the shortest distance between new data and training data [22]. In this algorithm, we found the k closest neighbors of the test data and look for the shortest distance between the test data and each training data [25]. Imputation with KNN did not require the establishment of a forecasting model for each data criterion that had missing value data. The weakness of imputation with KNN was used to looking for observations that best match observations that had missing values. Then, imputation with KNN was able to search for all training data or datasets [26].

KNN was able to classify based on training data to be seen by the closest distance of a data based on the value of k. The distance in question was calculated using the Euclidean distance equation. The Euclidean distance equation served to calculate the distance to interpret the closeness of the distance between two objects [23].

RF classification was carried out through the formation of a tree by conducting training on the sample data to be owned [27]. The Random Forest Algorithm was a classification method that combines several decision trees [28] with training on the given data. The classification process in RF was done by randomly splitting the data into several decision trees [29]. This algorithm was used to produce the final output of the identification system. In this study, the prediction result was obtained by looking at the majority voting results from each class that had been labeled in the previous data.

PSO was a computational evolutionary technique that uses a random particle population to search [30]. This particle population moved in the search space to find the best solution. The PSO technique was used to perform optimization in several ways, such as increasing attribute weights or selecting attributes on the variables used in the analysis [31].

In this study, the use of Particle Swarm Optimization (PSO) for the KNN and RF methods was carried out with finding optimal parameters to improve the accuracy and performance of the two methods simultaneously. PSO was used to perform parameter optimization in both methods simultaneously, so that it found the best parameters that provide optimal results for sentiment analysis.

### 2.5. *Outcome Evaluation*

After sentiment analysis, the result of this study was evaluated using several evaluation metrics such as accuracy and kappa. The kappa statistic was a measure corrected by chance and it required to value agreement beyond what happened by chance [32]. The result of this evaluation was used to assess how good the algorithm and to conduct a sentiment analysis of the Government Regulations in Lieu of Laws on Job Creation on the Twitter social media platform. Positive (negative) examples that are correctly classified by the classifier are called true positives (true negatives), false negatives (false positives) are positive (negative) examples that are misclassified [33].

Kappa value performance [34], was classified into five groups as shown in table 1 below.

Table 1. Kappa Value Classification

| No | Kappa Value | Classification |
|---|---|---|
| 1 | 0.81 – 1.00 | *Very Good* |
| 2 | 0.61 – 0.80 | *Good* |
| 3 | 0.41 – 0.60 | *Moderate* |
| 4 | 0.21 – 0.40 | *Fair* |
| 5 | 0.00 – 0.20 | *Poor* |

Interpretation of the results to identify public sentiment and opinion about the Government Regulations in Lieu of Laws on Job Creation on Twitter was carried out to draw the conclusions. Furthermore, this study presented important findings and relevant implications to serve as input for the government and related parties in making policies related to these regulations.

## 3. Result

### 3.1. Data Collection

The data was collected using the RapidMiner software in stages. The data collected was Twitter data from December 2022 to June 2023 regarding public opinion on the Government Regulations in Lieu of Laws on Job Creation. Using RapidMiner, this study obtained the information from Twitter using the "Search Twitter" operator and store this information using the "Write csv" operator in RapidMiner.

The data collection model with RapidMiner was shown in Figure 2.

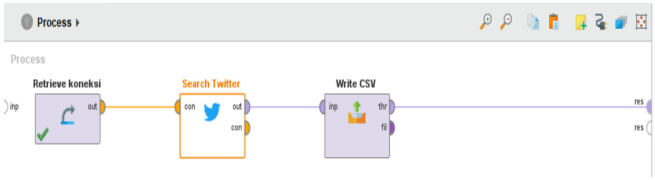



**Figure 2. Process of Twitter Data Crawling**

In Figure 2, it obtained 10.248 data tweets. Preceding the classification modeling process, data pre-processing was carried out included case folding, tokenization, stopword removal, stemming, and word weighting using the TF-IDF method using the RapidMiner software.

### 3.2. Data Processing

The data carried out a data cleaning process. The purpose of this action was to extract and clean the data in Figure 3.

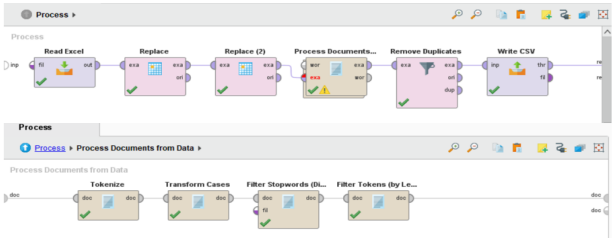



**Figure 3. Process of Data Processing**

In Figure 3, the data collected was processed to obtain data on words in the term frequency. The results from each pre-processing stage in table 2.

Table 2. Example of Preprocessing Results

| Stages | Results |
|---|---|
| *Data cleaning* | **Before**<br>*Dengan PERPPU Cipta Kerja, Indonesia hanya akan terus mengundang investasi yang mengeksploitasi sumber daya alam dan tenaga asing*<br>**After**<br>*Dengan PERPPU Cipta Kerja Indonesia hanya akan terus mengundang investasi yang mengeksploitasi sumber daya alam dan tenaga asing* |
| *Case Folding* | **Before**<br>*Dengan PERPPU Cipta Kerja Indonesia hanya akan terus mengundang investasi yang mengeksploitasi sumber daya alam dan tenaga asing*<br>**After**<br>*dengan perppu cipta kerja indonesia hanya akan terus mengundang investasi yang mengeksploitasi sumber daya alam dan tenaga asing* |
| *Normalizes* | **Before**<br>*dengan perppu cipta kerja indonesia hanya akan terus mengundang investasi yang mengeksploitasi sumber daya alam dan tenaga asing*<br>**After**<br>*dengan perppu cipta kerja indonesia hanya akan terus mengundang investasi yang mengeksploitasi sumber daya alam dan tenaga asing* |
| *Stopword Removal* | **Before**<br>*dengan perppu cipta kerja indonesia hanya akan terus mengundang investasi yang mengeksploitasi sumber daya alam dan tenaga asing*<br>**After**<br>*perppu cipta kerja indonesia terus undang investasi eksploitasi sumber daya alam tenaga asing* |
| *Stemming* | **Before**<br>*perppu cipta kerja indonesia hanya akan terus mengundang investasi yang mengeksploitasi sumber daya alam tenaga asing*<br>**After**<br>*perppu cipta kerja indonesia hanya terus undang investasi eksploitasi sumber daya alam tenaga asing* |
| *Tokenizing* | **Before**<br>*perppu cipta kerja indonesia hanya terus undang investasi eksploitasi sumber daya alam tenaga asing*<br>**After**<br>*'perppu', 'cipta', 'kerja', 'indonesia', 'hanya', 'terus', 'undang', 'investasi', 'eksploitasi', 'sumber', 'daya', 'alam', 'tenaga', 'asing',* |

There were 3.128 clean data to be obtained after preprocessing for further process. The data was divided into 2.189 training data and 939 test data, or 70% training data and 30% test data.

3.3. Labelling

After obtaining the cleaned data, this study processed the data by labeling it and adding attributes to determine the sentiment value of the data. Labels were negative, positive, or neutral according to the existing tweets. The following was a model for sentiment analysis training data.

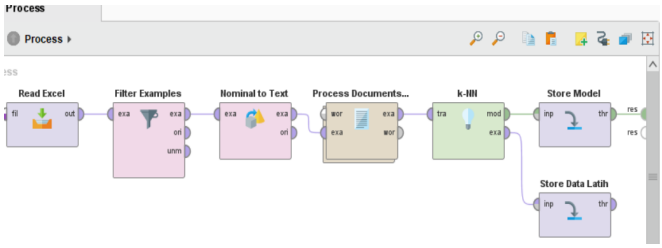
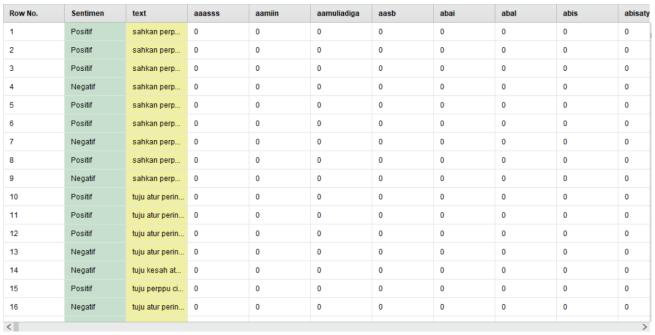



**Figure 4. Model of Training Data Processing**

The results were positive, negative and neutral sentiment values as follows.



| Row No. | Sentimen | text | aaasss | aamiin | aamuliadiga | aasb | abai | abal | abis | abisaty |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Positif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | Positif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | Positif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | Negatif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | Positif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | Positif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | Negatif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | Positif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | Negatif | sahkan perp... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | Positif | tuju atur perin... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | Positif | tuju atur perin... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | Positif | tuju atur perin... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | Negatif | tuju atur perin... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | Negatif | tuju kesah at... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | Positif | tuju perppu ci... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | Negatif | tuju atur perin... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 5. Result of Sentiment Value**

From the training data, the data was labeled using the test data model as follows.

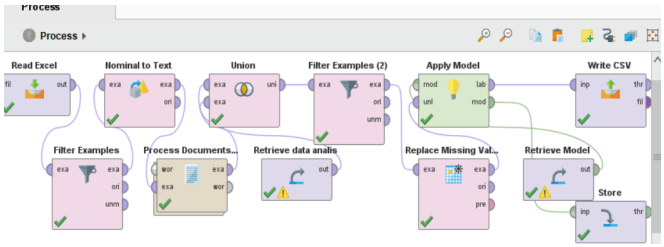



**Figure 6. Model of Data Processing Testing**

The results showed 3.128 examples, 6 special attributes, and 3.394 regular attributes as shown in Figure 7 below.

Open in Turbo Prep Auto Model Filter (3,128 / 3,128 examples): all

| Row No. | Sentimen | prediction(S... | confidence(... | confidence(... | confidence(... | aamiin | aamuliadiga | aasb |
|---|---|---|---|---|---|---|---|---|
| 1 | Negatif | Positif | 0.021 | 0.483 | 0.495 | 0 | 0 | 0 |
| 2 | Positif | Positif | 0.018 | 0.475 | 0.508 | 0 | 0 | 0 |
| 3 | Positif | Positif | 0.018 | 0.434 | 0.548 | 0 | 0 | 0 |
| 4 | Positif | Positif | 0.018 | 0.436 | 0.546 | 0 | 0 | 0 |
| 5 | Negatif | Positif | 0.018 | 0.480 | 0.502 | 0 | 0 | 0 |
| 6 | Positif | Positif | 0.018 | 0.395 | 0.587 | 0 | 0 | 0 |
| 7 | Negatif | Negatif | 0.020 | 0.533 | 0.447 | 0 | 0 | 0 |
| 8 | Negatif | Negatif | 0.019 | 0.524 | 0.457 | 0 | 0 | 0 |
| 9 | Positif | Positif | 0.018 | 0.419 | 0.562 | 0 | 0 | 0 |
| 10 | Positif | Positif | 0.018 | 0.395 | 0.587 | 0 | 0 | 0 |
| 11 | Positif | Positif | 0.018 | 0.395 | 0.587 | 0 | 0 | 0 |

ExampleSet (3,128 examples, 5 special attributes, 3,394 regular attributes)

**Figure 7. Test Results**

## 3.4. Implementation of Algorithm

The algorithm implementation model was as shown in Figure 8 below.

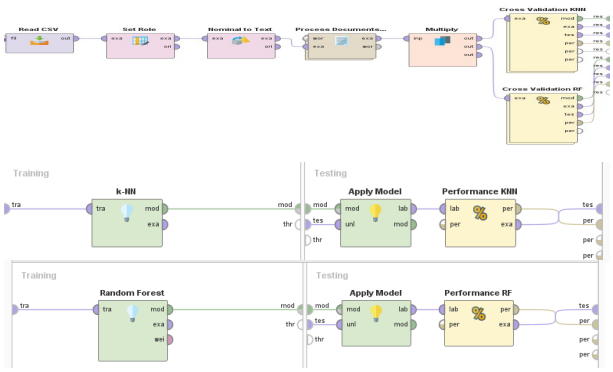



**Figure 8. Sentiment Analysis Model Without PSO**

The model with PSO optimization was seen in Figure 9.

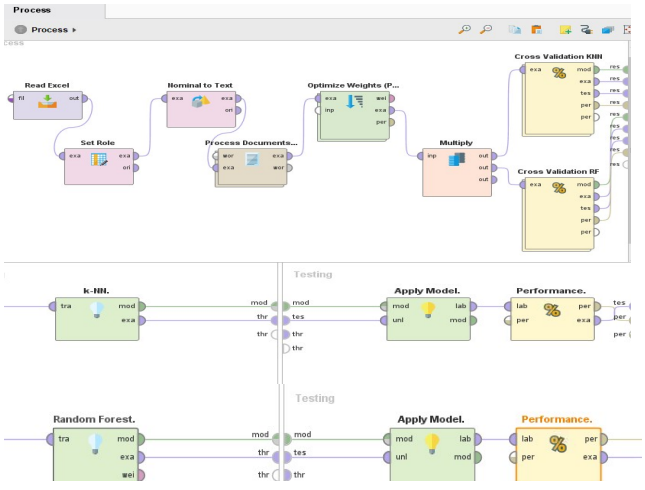



**Figure 9. KNN and RF models with PSO**

The model for visualizing of the results was as follows.


**Figure 10. Model of Data Visualization**

To provide the detailed insight into the model's performance in analyzing sentiment, other evaluation metrics were calculated, namely precision, recall and F1-score for each model.

Based on the model in Figure 10, 15 words that often appear to describe the public's view of the work copyright regulation are filtered as in Table 3.

Table 3. Data visualization of often words that appear

| Number | *Word* | *In docu-ments* | *total* | *In class (pos.)* | *In class (neg)* | *In class (neut.)* |
|---|---|---|---|---|---|---|
| 1 | kerja | 2642 | 3921 | 1370 | 2484 | 67 |
| 2 | cipta | 2612 | 3598 | 1254 | 2285 | 59 |
| 3 | perppu | 899 | 1896 | 287 | 1576 | 33 |
| 4 | dukung | 1118 | 1301 | 178 | 1100 | 23 |
| 5 | perpu | 1136 | 1251 | 779 | 443 | 29 |
| 6 | ciptaker | 974 | 1058 | 71 | 972 | 15 |
| 7 | ciptakerja | 800 | 856 | 496 | 346 | 14 |
| 8 | undang | 531 | 748 | 408 | 335 | 5 |
| 9 | perintah | 589 | 707 | 355 | 348 | 4 |
| 10 | nomor | 484 | 535 | 325 | 203 | 7 |
| 11 | buruh | 384 | 497 | 416 | 75 | 6 |
| 12 | atur | 447 | 492 | 280 | 210 | 2 |
| 13 | ganti | 387 | 399 | 231 | 166 | 2 |
| 14 | ekonomi | 361 | 382 | 56 | 326 | 0 |
| 15 | indonesia | 312 | 337 | 111 | 224 | 2 |

According to table 3 above, the most frequently words were appeared in the RapidMiner wordcloud results to be represented a visual representation of the text data and it showed how often the words appear in a particular document or body of text. The more often a word appears in a document, the larger in the word cloud. The word "Kerja" appears 3.921 times in all documents, while the word "Cipta" appears 3.598 times.

Furthermore, the words that frequently appear are described as follows.


**Figure 11. Result of Visualization with Wordcloud**

In Figure 11, it showed that the words that often appear in this sentiment analysis were the verbs, copyright, perpu, job creation, and government. The word copyright was tweeted the most because it was related to the object of this study.

3.5. Result of Evaluation Stage

The metric was used to evaluate the model's performance in testing tests, namely: accuracy in each algorithm. Model was created by using RapidMiner 9.10 software as shown in Figure 12.

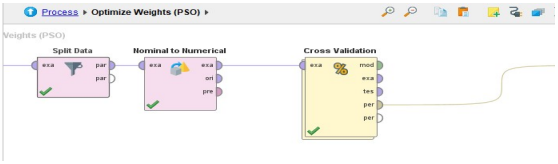



**Figure 12. Model of Performance Evaluation**

## 4. Discussion

4.1. Testing of *Algoritma K-Nearest Neighbors* (KNN) dan *Random Forest* (RF)

The results of model testing using testing and validation data were as shown in table 4.

Table 4. Test Results Without PSO

| Method | *PerformanceVektor (Performance)* |
|---|---|
| | *accuracy* |
| KNN | 80.40% |
| RF | 77.21% |

In table 4, it was known that the accuracy results of the KNN method were 80.40% in sentiment analysis and RF were 77.21%. This accuracy showed that the model was able to predict sentiment correctly. An accuracy of 80.40% showed that KNN was able to predict with a fairly good level of accuracy. The RF method achieved an accuracy of 77.21% in sentiment analysis. This value showed a good level of accuracy, but it was slightly lower than that achieved by KNN. The test resulted from KNN and RF were shown in table 5.

Table 5. Test Results of KNN

| **accuracy: 80.40% +/- 1.31% (micro average: 80.40%)** | | | | |
|---|---|---|---|---|
| | **true Negative** | **true Positive** | **true Neutral** | **class precision** |
| pred. Negatif | 1079 | 178 | 19 | 84.56% |
| pred. Positif | 377 | 1420 | 21 | 78.11% |
| pred. Netral | 17 | 1 | 16 | 47.06% |
| class *recall* | 73.25% | 88.81% | 28.57% | |

The RF algorithm test results are shown in Table 6.

Table 6. Test Results of RF

| **accuracy: 77.21% +/- 2.60% (micro average: 77.21%)** | | | | |
|---|---|---|---|---|
| | **true Negative** | **true Positive** | **true Neutral** | **class precision** |
| pred. Negatif | 1008 | 192 | 21 | 82.56% |
| pred. Positif | 465 | 1407 | 35 | 73.78% |
| pred. Netral | 0 | 0 | 0 | 0.00% |
| class *recall* | 68.43% | 87.99% | 0.00% | |

Based on the results in tables 5 and 6, the resulting confusion matrix accuracy of KNN was 80.40% and RF was 77.21%. it showed that most of the samples had been classified correctly. This accuracy reflected the percentage of correct predictions from the total amount of data available. In this case, the KNN and RF models without PSO achieve good accuracy from all predictions made by the model are correct. This level of accuracy indicated the ability of the KNN model to carry out sentiment analysis with a good level of accuracy.

Calculation of recall, precision and F1-Score values for each class was seen in Table 7.

Table 7. Recall, precision and f1-score results before PSO

| **Method** | **Class** | **Result** | | |
|---|---|---|---|---|
| | | ***Recall*** | ***Precision*** | ***F1-score*** |
| KNN | Positive | 90.30% | 32.01% | 47.23% |
| | Negative | 88.81% | 79.03% | 83.63% |
| | Neutral | 94.12% | 4.09% | 7.85% |
| RF | Positive | 97.59% | 75.16% | 85.09% |
| | Negative | 88.06% | 75.16% | 81.03% |
| | Neutral | 0% | 0% | 0% |

Based on table 7, it showed that RF had better performance than KNN in detecting Positive sentiment. RF had a higher recall and a better F1-Score for the Positive category. However, for Negative sentiment, both had almost comparable performance, although RF had a slight edge in F1-Score. So overall, both KNN and RF had good performance in detecting Positive and Negative sentiment, but both face difficulties in predicting Neutral sentiment.

Based on testing with the PSO optimized model, the accuracy result was in Table 8.

Table 8. Recapitulation of Test Results with PSO

| **Method** | ***PerformanceVector (Performance)*** |
|---|---|
| | ***accuracy*** |
| KNN | 85.23% |
| RF | 80.53% |

There was a significant improvement in both models after optimization with PSO as in table 8. The prediction success rate was higher with PSO optimization even though there was a difference in the magnitude of the improvement. It indicated that the ability of each model to classify data was different.

Based on the description above, it concluded that the PSO feature optimized the sentiment of Twitter social media users towards the Perpu on Job Creation using the KNN and RF classification algorithms. It was proven by a comparison of the test results of the two models. It showed a significant increase between the model test results before and after optimization with PSO, especially in the KNN algorithm.

The results of testing the KNN algorithm with PSO optimization was seen in Table 9.

Table 9. Accuracy Results of KNN with PSO

| accuracy: 85.23% +/- 1.53% (micro average: 85.23%) | | | | |
|---|---|---|---|---|
| | **true Negatif** | **true Positif** | **true Netral** | **class precision** |
| pred. Negatif | 1318 | 265 | 29 | 81.76% |
| pred. Positif | 154 | 1333 | 12 | 88.93% |
| pred. Netral | 1 | 1 | 15 | 88.24% |
| class *recall* | 89.48% | 83.36% | 26.79% | |

RF test results was seen in Table 10.

Table 10. Accuracy results of RF with PSO

| accuracy: 80.53% +/- 2.98% (micro average: 80.53%) | | | | |
|---|---|---|---|---|
| | **true Negative** | **true Positive** | **true Neutral** | **class precision** |
| pred. Negatif | 1148 | 228 | 26 | 81.88% |
| pred. Positif | 325 | 1371 | 30 | 79.43% |
| pred. Netral | 0 | 0 | 0 | 0.00% |
| class *recall* | 77.94% | 85.74% | 0.00% | |

The calculation of recall, precision and F1-Score values for each class was as follows.

Table 11. Recall, precision and f1-score results after PSO

| **Method** | **Class** | **Result** | | |
|---|---|---|---|---|
| | | ***Recall*** | ***Precision*** | ***F1-score*** |
| KNN | Positif | 99.10% | 89.62% | 94.19% |
| | Negatif | 83.45% | 89.62% | 86.44% |
| | Netral | 93,75% | 8.82% | 16.05% |
| RF | Positif | 97.86% | 80.87% | 88.65% |
| | Negatif | 85.75% | 80.87% | 83.23% |
| | Netral | 0% | 0% | 0% |

Based on the data above, it concluded that overall, KNN had good performance in detecting Positive and Negative sentiment, but had problems in predicting Neutral sentiment. Likewise, RF had good performance in detecting Positive and Negative sentiment, but it equally had difficulty predicting Neutral sentiment. Increasing accuracy in classifying Neutral sentiment was a potential area of improvement for both methods.

## 4.2. Comparison of Test Results

To see which algorithm contributed the most to accuracy, this study compared the accuracy results of the two models before and after optimization with PSO. The result of this comparison was as shown in table 12 below.

Table 12. Test Result

| **Method** | **Prediction Accuracy Without PSO** | **Prediction Accuracy With PSO** |
|---|---|---|
| KNN | 80.40% | 85.23% |
| RF | 77.21% | 80.53% |

Based on table 12 above, it was seen that the accuracy value of the KNN algorithm was higher than the RF algorithm. There was a difference in accuracy between before and after optimization. The use of PSO significantly increased prediction accuracy by around 4.83% on the KNN method and an increase of 3.32%. With the implementation of PSO, KNN prediction accuracy increased significantly from 80.40% to 85.23%. Similarly, RF also had an increase in prediction accuracy when PSO was applied. Accuracy increases from 77.21% to 80.53% with PSO.

Next, to see the level of agreement between observers, it was substantially from the Kappa value presented in table 13.

Table 13. Kappa Test Results

| **Method** | **Kappa Before PSO** | **Kappa After PSO** |
|---|---|---|
| KNN | 0.616 | 0.712 |
| RF | 0.548 | 0.615 |

Based on table 13, it was seen that the Kappa value in the KNN method has increased significantly after implementing PSO. It was seen that the Kappa value increased significantly from 0.616 to 0.712 after implementing PSO. A Kappa value that is closer to 1 indicates a higher level of agreement between model predictions and actual data. Therefore, the application of PSO positively affects the ability of the KNN model to provide more accurate predictions.

After implementing PSO, there was an increase of 0.096 (9.6%), namely from 0.616 to 0.712 in the good classification. This improvement shows that PSO is able to optimize KNN performance in classifying data so that the level of agreement between observers increases substantially.

In RF, it was seen that the Kappa value also had a significant increase from 0.548 to 0.615 after implementing PSO, namely from moderate to good classification. In this case, there was an increase of 0.067 (6.7%). Thus, it concluded that PSO succeeded in increasing the level of agreement between RF model predictions and reality.

Overall, these results showed that the application of PSO had provided significant improvements in prediction quality for both KNN and RF methods, as reflected in the increase in Kappa values.

## 4.3. Contextualization of Findings and Comparison with Previous Research

While this study demonstrates the performance improvement of KNN and RF models after optimization using Particle Swarm Optimization (PSO), it is important to acknowledge that direct comparisons with absolute accuracy from previous studies are often complex and not feasible on an apples-to-apples basis. This is due to several crucial factors that differentiate the research conditions:

1. Dataset Characteristics and Complexity: This study focuses on sentiment data from Twitter related to the Job Creation Law (Perpu Cipta Kerja), a highly specific, dynamic, and controversial topic in Indonesia. Twitter data inherently contains high levels of noise, including the use of informal

language, abbreviations, slang, typos, emoticons, and contextual ambiguity (e.g., sarcasm or irony), which are much more complex to analyze compared to general sentiment datasets often used in the literature (e.g., movie or product reviews), which tend to be cleaner or have undergone extensive curation. This complexity can fundamentally limit the accuracy that classification models can achieve in this domain.

2. 2. Text Feature Preprocessing and Representation: The performance of a classification model is highly dependent on the quality of the data preprocessing (e.g., text normalization, stop word handling, stemming) and the text feature representation method (e.g., TF-IDF, Bag-of-Words, or word embeddings). Differences in the implementation of these steps, or the use of more sophisticated feature representations (such as Transformer-based pretrained language models) by different studies, may explain the variation in reported accuracy. Our study used [Name your feature representation method, e.g., TF-IDF], which is effective but may have limitations in capturing all the semantic nuances of highly informal social media data.
3. Main Optimization Objectives and Research Contributions: The main objective of applying PSO in this study is to optimize the hyperparameters of the KNN and RF models, specifically on our dataset. The internally consistent and significant improvements in accuracy and Kappa values (KNN increased by 4.83% and RF by 3.32% accuracy, as well as Kappa values), clearly demonstrate the effectiveness of PSO in finding the best parameter configurations for these algorithms in this challenging data environment. This is a substantial methodological contribution, demonstrating that PSO is a valuable strategy for fine-tuning machine learning models on complex and noisy sentiment data.

To provide a broader context, this study compares the results obtained with other studies that used similar classification algorithms and/or optimization techniques in sentiment analysis.

Table 14. Comparison of Performance with Previous Research

| Research (References) | Dataset/Domain | Algorithm | Optimization Techniques | Accuracy (%) | F1-Score (%) |
|---|---|---|---|---|---|
| This research | Perpu Cipta Kerja (Twitter) | KNN+PSO | PSO | 85.23 | 65.56 |
| This research | Perpu Cipta Kerja (Twitter) | RF+PSO | PSO | 80.53 | 57.96 |
| Lavate & Srivastava (2023) [35] | IoT Network Traffic (CIC-IDS2017) | RF+PSO | PSO | ~99.9 | – |
| Ghale-Nou et al. (2020) [36] | Sentinel-2 Crop Mapping | RF+PSO | PSO | +0.22–1.06 (improve) | +0.54–2.97 (increase) |
| Rajendran et al. (2020) [37] | Remote-Sensing LULC | RF (dengan PSO & LSTM) | Human-Group PSO | +2.56 (improve) | – |
| Nisa et al. (2025) [38] | Wardah Instagram Comments (Beauty Products) | KNN | - | 78.23 | 78.0 |

From Table 14, it can be seen that the accuracy of the KNN+PSO (85.23%) and RF+PSO (80.53%) models in this study shows competitive performance, especially considering the complexity of the dataset used, namely Twitter data regarding the Perpu Cipta Kerja, which tends to be noisy and unstructured.

For example, compared to Nisa et al. (2025) who used KNN without optimization on Instagram comments of beauty products and obtained an accuracy of 78.23% and an F1 Score of 78.0%, the KNN+PSO model in this study was able to provide performance improvements in terms of both accuracy (+6.99%) and F1 Score (-12.44%), despite different data challenges.

Furthermore, in a study by Lavate & Srivastava (2023), which used RF+PSO on IoT network traffic (CIC IDS2017), the reported accuracy reached ~99.9%. While this appears significantly higher, it's important to note that IoT data has a very different structure and noise than social media data like Twitter.

Meanwhile, Ghale Nou et al. (2020) and Rajendran et al. (2020) showed an increase in accuracy and F1 Score performance after using PSO on Random Forest in the remote sensing domain, showing an accuracy increase of around +0.22–1.06% and +2.56%, respectively, which confirms that the use of optimization techniques such as PSO can improve model performance even in different domains.

Specifically, for neutral sentiment classification, both in this study and in various other studies, such as in the social media comment domain (Nisa et al., 2025), neutral sentiment remains a challenge. This is likely due to the often-unexplicit ambiguity of neutral sentiment, as well as its small data proportion compared to positive and negative sentiment. Therefore, improving the accuracy of neutral sentiment classification in the future may require specialized techniques such as data balancing, model fine-tuning, or the use of advanced embedding techniques..

### 4.4. Variations in Experimental Conditions and Contributions to Further Research

This research has successfully demonstrated the effectiveness of PSO optimization on the KNN and RF algorithms for sentiment analysis of the Job Creation Regulation. To provide a more concrete contribution to the algorithm and dataset, the experiment can be expanded by varying the following conditions for future research:

1. Exploration of Advanced Feature Representation Methods: In addition to the TF-IDF method already used, future research can explore the use of word embeddings (such as Word2Vec, GloVe, FastText) or Transformer-based pre-

trained language models (such as BERT, IndoBERT). These methods have the ability to capture richer semantic and contextual meaning from complex tweet texts, potentially resulting in significant improvements in model accuracy and robustness.

2. Analysis of the impact of different Pre-processing: Testing different pre-processing schemes (e.g., with/without stemming or lemmatization, different levels of slang word normalization, specific handling of emoticons and punctuation) and comparing their impact on model performance can provide critical insights into how the quality and characteristics of input data affect sentiment classification results.
3. In-depth investigation of 'neutral' sentiment classification: Given the persistent challenges experienced by both models in classifying 'Neutral' sentiment, future research could focus on techniques for addressing class imbalance (e.g., oversampling like SMOTE or undersampling) or exploring classification models specifically designed for ambiguous sentiment. A more in-depth error analysis of tweets misclassified as 'Neutral' could also help identify patterns and characteristics that make this class difficult to predict.
4. More comprehensive PSO hyperparameter optimization: Further exploration of the internal parameters of the PSO itself (e.g., number of particles, maximum number of iterations, inertia coefficient, cognitive and social coefficients) can be performed to ensure that the most optimal PSO configuration has been found within a wider search space, potentially resulting in further performance improvements.
5. External and Cross-Domain Validation: Testing the optimized model on sentiment datasets from other domains (e.g., other public policy issues, different product reviews) or at different time periods can prove the generalizability and robustness of the developed model.

## 5. Conclusion

Based on the test results, it was found that the Particle Swarm Optimization (PSO) feature successfully and significantly optimized the performance of sentiment analysis on Twitter social media user opinions regarding the Job Creation Perpu, particularly when applied to the K-Nearest Neighbors (KNN) and Random Forest (RF) classification algorithms. consistent increase in accuracy was observed in both algorithms after optimization using PSO. The KNN accuracy value before using PSO was 80.40% then increased to 85.23% after optimization with PSO. The RF accuracy value was 77.21% before PSO and increased to 80.53% post-PSO optimization. This indicates that the PSO-based KNN algorithm is a highly effective method and is recommended for sentiment analysis in this study, given its most substantial improvement.

This research provides two significant main contributions: 1) methodological contributions, and 2) practical and policy contributions. The methodological contribution of this research is that the integration of PSO with the KNN and RF classification algorithms shows that hyperparameter optimization can significantly improve the accuracy of sentiment analysis models, especially when applied to complex text data such as social media opinions. This finding emphasizes the importance of intelligent optimization strategies in developing more accurate and robust sentiment analysis systems. It offers a new approach to addressing the challenge of optimal parameter selection for sentiment classification, which can be applied to other domains beyond the public policy context.

The practical and policy contribution of this research is that the more accurate sentiment analysis provides clear and measurable insights into public perception of the Job Creation Regulation (Perpu Ciptra Kerja). By identifying the proportion of positive, negative, and neutral sentiment more precisely, the government and relevant stakeholders can better understand public reactions. This information is crucial as feedback for policy evaluation, more effective public communication, or even consideration in formulating future policies that better align with public aspirations. This enables more data-driven decision-making that is responsive to the dynamics of public opinion.

For further study development, it is recommended to explore the potential of other classification algorithms that can be used for sentiment analysis, such as Support Vector Machines (SVM), Naive Bayes, or Deep Learning methods (e.g., Recurrent Neural Networks or Transformer-based models) for performance comparison. In addition, the use of additional features in sentiment analysis, such as implementing more advanced Natural Language Processing (NLP) techniques to extract key features from tweet texts (e.g., word embeddings or more granular emotion analysis), should also be considered to enhance the depth and accuracy of the analysis.

## Acknowledgement

This study was supported by research grants organized by the Directorate of Research, Technology, and Community Service (DRTPM), Master Thesis Research Scheme (PTM) with number 187/LL2/AL.04/2023.

## References

[1] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, dan M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *J. Informatics Web Eng.*, vol. 3, no. 1, hal. 67–75, 2024, doi: 10.33093/jiwe.2024.3.1.5.

[2] Aldinata, A. M. Soesanto, V. C. Chandra, dan D. Suhartono, "Sentiments comparison on Twitter about LGBT," *Procedia Comput. Sci.*, vol. 216, hal. 765–773, 2023, doi: 10.1016/j.procs.2022.12.194.

[3] M. F. Fakhrezi, Adian Fatchur Rochim, dan Dinar Mutiara Kusomo Nugraheni, "Comparison of Sentiment Analysis

Methods Based on Accuracy Value Case Study: Twitter Mentions of Academic Article," *Resti*, vol. 7, no. 1 (2023), hal. 161–167, 2023, doi: 10.29207/resti.v7i1.4767.

[4] R. Othman, Y. Abdelsadek, K. Chelghoum, I. Kacem, dan R. Faiz, "Improving Sentiment Analysis in Twitter Using Sentiment Specific Word Embeddings," in *Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2019*, 2019, vol. 10, hal. 854–858, doi: 10.1109/IDAACS.2019.8924403.

[5] M. E. Purbaya, D. P. Rakhmadani, Maliana Puspa Arum, dan Luthfi Zian Nasifah, "Implementation of n-gram Methodology to Analyze Sentiment Reviews for Indonesian Chips Purchases in Shopee E-Marketplace," *Resti*, vol. 7, no. 3, hal. 609–617, 2023, doi: https://doi.org/10.29207/resti.v7i3.4726.

[6] Kamrozi, A. N. Hidayanto, K. Y. P.M., M. A. Virgananda, dan R. R. Suryono, "Sentiment Analysis of Cryptocurrency Trading Platform Service Quality on Playstore Data: A Case of Indodax," *Resti*, vol. 7, no. 3, hal. 445–456, 2023, doi: https://doi.org/10.29207/resti.v7i3.4769.

[7] Y. Astuti, Y. Ruldeviyani, F. Salbari, dan A. Prayogi, "Sentiment Analysis of Electricity Company Service Quality Using Naïve Bayes," *Resti*, vol. 7, no. 2 (2023), hal. 389–396, 2023.

[8] X. Li, J. Li, dan Y. Wu, "A global optimization approach to multi-polarity sentiment analysis," *PLoS One*, vol. 10, no. 4, hal. 1–18, 2015, doi: 10.1371/journal.pone.0124672.

[9] M. Ashrafzadeh, H. M. Taheri, M. Gharehgozlou, dan S. Hashemkhani Zolfani, "Clustering-based return prediction model for stock pre-selection in portfolio optimization using PSO-CNN+MVF," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 9, hal. 101737, 2023, doi: 10.1016/j.jksuci.2023.101737.

[10] Y. Yue, Y. Peng, dan D. Wang, "Deep Learning Short Text Sentiment Analysis Based on Improved Particle Swarm Optimization," *Electron.*, vol. 12, no. 19, 2023, doi: 10.3390/electronics12194119.

[11] G. Guo, H. Wang, D. Bell, Y. Bi, dan K. Greer, "Using kNN model for automatic text categorization," *Soft Comput.*, vol. 10, no. 5, hal. 423–430, 2006, doi: 10.1007/s00500-005-0503-y.

[12] S. Jiang, G. Pang, M. Wu, dan L. Kuang, "An improved K-nearest-neighbor algorithm for text categorization," *Expert Syst. Appl.*, vol. 39, no. 1, hal. 1503–1509, 2012, doi: 10.1016/j.eswa.2011.08.040.

[13] A. Adamu, M. Abdullahi, S. B. Junaidu, dan I. H. Hassan, "An hybrid particle swarm optimization with crow search algorithm for feature selection," *Mach. Learn. with Appl.*, vol. 6, no. June, hal. 100108, 2021, doi: 10.1016/j.mlwa.2021.100108.

[14] M. Rostami dan M. Oussalah, "A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest," *Informatics Med. Unlocked*, vol. 30, no. April, hal. 100941, 2022, doi: 10.1016/j.imu.2022.100941.

[15] M. Agrawal dan N. R. Moparthi, "An efficient multiple-word embedding-based cross-domain feature extraction and aspect sentiment classification," *Meas. Sensors*, vol. 28, no. January, hal. 100851, 2023, doi: 10.1016/j.measen.2023.100851.

[16] N. Jalal, A. Mehmood, G. S. Choi, dan I. Ashraf, "A novel improved random forest for text classification using feature ranking and optimal number of trees," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, hal. 2733–2742, 2022, doi: 10.1016/j.jksuci.2022.03.012.

[17] A. Shaamala, T. Yigitcanlar, A. Nili, dan D. Nyandega, "Machine learning applications for urban geospatial analysis: A review of urban and environmental studies," *Cities*, vol. 165, no. January 2024, hal. 106139, 2025, doi: 10.1016/j.cities.2025.106139.

[18] B. T. Pham *et al.*, "A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil," *Sustain.*, vol. 12, no. 6, hal. 1–16, 2020, doi: 10.3390/su12062218.

[19] S. Sengupta, S. Basak, dan R. A. Peters, "Particle Swarm Optimization: A Survey of Historical and Recent Developments with Hybridization Perspectives," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, hal. 157–191, 2019, doi: 10.3390/make1010010.

[20] L. Vanneschi dan S. Silva, "Particle Swarm Optimization," *Nat. Comput. Ser.*, hal. 105–111, 2023, doi: 10.1007/978-3-031-17922-8_4.

[21] T. Kusuma Wardana, Y. Arkeman, K. Priandana, dan F. Kurniawan, "Use of Plant Health Level Based on Random Forest Algorithm for Agricultural Drone Target Points," *Resti*, vol. 7, no. 3, hal. 637–645, 2023, doi: https://doi.org/10.29207/resti.v7i3.4959.

[22] T. Mustaqim, K. Umam, dan M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm," *J. Phys. Conf. Ser.*, vol. 1567, no. 3, hal. 8–15, 2020, doi: 10.1088/1742-6596/1567/3/032024.

[23] N. I. Saputri, Y. Sibaroni, dan S. S. Prasetiyowati, "Covid-19 Fake News Detection on Twitter Based on Author Credibility Using Information Gain and KNN Methods," *Resti*, vol. 7, no. 1 (2023), hal. 185–192, 2023.

[24] V. Dwi Antonio, S. Efendi, dan H. Mawengkang, "Sentiment analysis for covid-19 in Indonesia on Twitter with TF-IDF featured extraction and stochastic gradient descent," *Int. J. Nonlinear Anal. Appl*, vol. 13, no. 1, hal. 2008–6822, 2022, [Daring]. Tersedia pada: http://dx.doi.org/10.22075/ijnaa.2021.5735.

[25] G. Guslendra, S. Defit, dan R. Bastola, "K-Means and K-NN Methods For Determining Student Interest," *Int. J. Artif. Intell. Res.*, vol. 6, no. 1, 2022, doi: 10.29099/ijair.v6i1.222.

[26] A. Fadlil, Herman, dan D. Praseptian M, "K Nearest Neighbor Imputation Performance on Missing Value Data Graduate User Satisfaction," *Resti*, vol. 6, no. 4 (2022), hal. 570–576, 2022, doi: 10.29207/resti.v6i4.4173.

[27] Saparudin, R. Amalia, dan P. Rosyani, "Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO) Image Classification Using Random Forest Method and Sequential Minimal Optimazation (SMO)," *Justin*, vol. 09, no. 2 (2021), hal. 132–134, 2021, doi: 10.26418/justin.v9i2.44120.

[28] P. K. Sarangi, E. Singh, A. Kaushal, B. Sharma, M. Dutta, dan V. P. Dubey, "Analysing the Fluctuations in Commodity Prices and Forecasting the Future Directions Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 259, hal. 1827–1836, 2025, doi: 10.1016/j.procs.2025.04.138.

[29] J. Alzubi, A. Nayyar, dan A. Kumar, "Machine Learning from Theory to Algorithms: An Overview," *J. Phys. Conf. Ser.*, vol. 1142, no. 1, 2018, doi: 10.1088/1742-6596/1142/1/012012.

[30] E. S. Rahayu, A. Ma'arif, dan A. Çakan, "Particle Swarm Optimization (PSO) Tuning of PID Control on DC Motor," *Int. J. Robot. Control Syst.*, vol. 2, no. 2, hal. 435–447, 2022, doi: 10.31763/ijrcs.v2i2.476.

[31] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, hal. 115, 2020, doi: 10.33365/jti.v14i2.679.

[32] J. R. Dettori dan D. C. Norvell, "Kappa and Beyond: Is There Agreement?," *Glob. Spine J.*, vol. 10, no. 4, hal. 499–501, 2020, doi: 10.1177/2192568220911648.

[33] F. N. Hasan dan Mochamad Wahyudi, "Analisis Sentimen Artikel Berita Tokoh Sepak Bola Dunia Menggunakan Algoritma Support Vector Machine Dan Naive Bayes Berbasis Particle Swarm Optimization," *Bitkom Res.*, vol. 3, no. 4, hal. 1–3, 2018.

[34] Q. Xie, "Agree or Disagree ? A Demonstration of An Alternative Statistic to Cohen ' s Kappa for Measuring the Extent and Reliability of Agreement between Observers," *FCSM Res. Conf. Washingt. Conv. Cent.*, vol. 1, hal. 1–12, 2011, [Daring]. Tersedia pada: https://fcsm.sites.usa.gov/ files/ 2014/05/ J4_Xie_2013FCSM.pdf%0Ahttps://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/05/J4_Xie_2013FCSM.pdf.

[35] S. H. Lavate dan P. K. Srivastava, "A Hybrid Feature Selection Approach based on Random Forest and Particle Swarm Optimization for IoT Network Traffic Analysis," *Int. J. Electr. Electron. Res.*, vol. 11, no. 2, hal. 568–574, 2023, doi: 10.37391/IJEER.110244.

[36] E. Akbari, A. D. Boloorani, N. N. Samany, S. Hamzeh, S. Soufizadeh, dan S. Pignatti, "Crop mapping using random forest and particle swarm optimization based on multi-temporal sentinel-2," *Remote Sens.*, vol. 12, no. 9, hal. 1–21, 2020, doi: 10.3390/RS12091449.

[37] R. G. Babu, K. U. Maheswari, C. Zarro, B. D. Parameshachari, dan S. L. Ullo, "Land-Use and Land-Cover Classification Using a Human Group-Based Particle Swarm Optimization Algorithm with an LSTM Classifier on Hybrid Pre-Processing Remote-Sensing Images," *Remote Sens.*, vol. 12, no. 24, hal. 1–28, 2020, doi: 10.3390/rs12244135.

[38] A. A. Nisa, N. H. Safitri, L. Stianingsih, dan F. H. Saputri, "Analysis of Customer Sentiment towards Wardah Beauty Products on Instagram Using K-Nearest Neighbors (KNN) and Naïve Bayes Algorithms," *G-Tech J. Teknol. Terap.*, vol. 9, no. 2, hal. 919–928, 2025, doi: 10.70609/gtech.v9i2.6804.