DOI: 10.14421/ijid.2024.4500 IJID (International Journal on Informatics for Development), *e-ISSN: 2549-7448*
Vol. 13. No. 1, June 2024, Pp. 386-399

Analysis of Public Sentiment Towards POLRI's Performance using Naive Bayes and K-Nearest Neighbors

Yusuf Handika[1*]
Department of Informatics Engineering
Universitas Muhammadiyah Prof. Dr. Hamka
Jakarta, Indonesia
yusufhandika21@gmail.com

Isa Faqihuddin Hanif[2]
Department of Informatics Engineering
Universitas Muhammadiyah Prof. Dr. Hamka
Jakarta, Indonesia
isa@uhamka.ac.id

Firman Noor Hasan[3]
Department of Informatics Engineering
Universitas Muhammadiyah Prof. Dr. Hamka
Jakarta, Indonesia
firman.noorhasan@uhamka.ac.id

Article History
Received May 21st, 2024
Revised June 5th, 2024
Accepted June 8th, 2024
Published June, 2024

***Abstract*—Using Twitter as a platform for sharing information includes tracking public perceptions of the performance of the Indonesian National Police (POLRI). Public sentiment assists as a gauge for evaluating POLRI's operational capabilities and supports decision-making processes to enhance the organization's reputation. However, raw public opinion data often requires careful analysis for decision-making. Hence, conducting sentiment analysis of Twitter data is crucial. This analytical process involves extracting and classifying opinions into neutral, positive, and negative sentiments. This study employs two distinct sentiment analysis methods: the Naive Bayes algorithm and the K-Nearest Neighbors. Analysis of 1285 tweets reveals prevailing satisfaction with POLRI's performance, indicated by many positive sentiments. However, there is also a notable number of negative feelings. The assessment from confusion matrix results demonstrate that the Naive Bayes algorithm achieves 99.03% accuracy, while the K-Nearest Neighbors algorithm achieves 95.33% accuracy. By leveraging insights from public opinion data, POLRI can make more accurate and timely decisions, enabling it to better fulfill the community's needs and expectations. This strategic use of data enhances service quality and bolsters POLRI's favorable image among the public fosters more harmonious relationships and enhances public trust in law enforcement agencies.**

Keywords*—*algorithm; confusion matrix; decision making; sentiment analysis; Twitter

1 INTRODUCTION

The rapid development of digital technology has made the dissemination of information in society easy and swift [1]. The demand for efficiency and speed in data and information processing drives institutions to innovate to meet competitive public information needs [2]. Consequently, information in the form of data becomes a valuable asset for an institution and requires proper management to use it optimally [3]. Many people use digital platforms like Twitter to share information on various topics [4]. The performance of the Indonesian National Police (POLRI) is a crucial benchmark for evaluating public perception of the institution [5]. The police are considered a vital element [6] in maintaining security, upholding the law, and providing protection and services to the community [7]. Assessing POLRI's performance involves evaluating the quality and quantity of work carried out over a specific period [8]. Good performance can enhance POLRI's image in the eyes of the public, while subpar performance may tarnish that reputation [9]. As a part of public service, the effectiveness of POLRI is paramount as it directly impacts the community's [10] interests eliciting various responses from them [11].

On Twitter, public perceptions of the POLRI's image vary, ranging from recognition of achievements to criticism highlighting areas for improvement to meet public expectations. Public satisfaction with POLRI's performance can serve as an evaluation indicator to enhance the institution's image [5]. Therefore, reform in POLRI's performance is crucial [6] to improve its image and foster harmonious relations between the police and the community [7]. Public opinions can serve as an evaluation tool for POLRI to enhance performance quality and support decision-making efforts to improve the police institution's image. However, the need for more processing of public opinion data renders it unusable as a decision-making basis [8]. Hence, methods for examining opinions on Twitter, such as sentiment analysis, are essential [9]. Sentiment analysis effectively evaluates public opinions by extracting textual data on a specific topic [10] and classifying them as negative, neutral, or positive [11]. Several algorithms such as naive Bayes, a classification method that uses Bayes' theorem [12] with a statistical approach based on probability [13], and K-Nearest Neighbors (KNN) based on data proximity [14] can be used in examining public sentiment.

The Naive Bayes algorithm, based on Bayes' probability theorem, is known for its efficiency in text classification, especially on large datasets. It is a simple but effective process and has been widely used in various sentiment analysis applications [15]. However, the Naive Bayes algorithm lacks selection features that can affect accuracy because it assumes that they are independent of each other [16]. The KNN algorithm offers several advantages, including rapid training speed, ease of implementation, and effectiveness on large datasets, making it a practical choice for classification [17]. On the other hand, the K-Nearest Neighbors algorithm also has drawbacks, such as sensitivity to data scale and the use of Euclidean distance, which does not consider attribute relevance weights, potentially resulting in biased classifications [18].

In previous research [19], an analysis of 269 public sentiment data related to the POLRI was conducted using the Naive Bayes algorithm. The findings revealed a high level of accuracy, reaching 98.5%, with the ability to classify sentiments into three different classes. Another study [20] employing a similar algorithm, Naive Bayes, focused on the POLRI service application, achieving an accuracy rate of 70.3% from 1500 review data with two different class classifications. In another study [21], the KNN algorithm was used to examine 16.543 public sentiments on Twitter, achieving an accuracy rate of 98.5% and capable of classifying sentiments into three different classes. Another study [22] utilizing the same algorithm, KNN, achieved an accuracy rate of 83.3% from 1500 review data with two other class classifications.

This research examines public sentiment towards the performance of the POLRI on Twitter, focusing specifically on how this performance affects the institution's image on the platform by classifying opinions into negative, neutral, and positive classes using RapidMiner. Additionally, the researchers will compare the algorithm's accuracy and recall, precision, and f1-score for each sentiment from Naive Bayes and KNN using confusion matrices to examine sentiment. The results of this analysis are expected to serve as a valuable resource for decision-making to enhance performance and build a more positive and trustworthy image in society.

2 METHOD

In this research, Fig. 1 illustrates the picture of the data processing procedure implemented by the researcher. This picturing is the stage of research methodology that systematically underpins the conducted sentiment analysis.

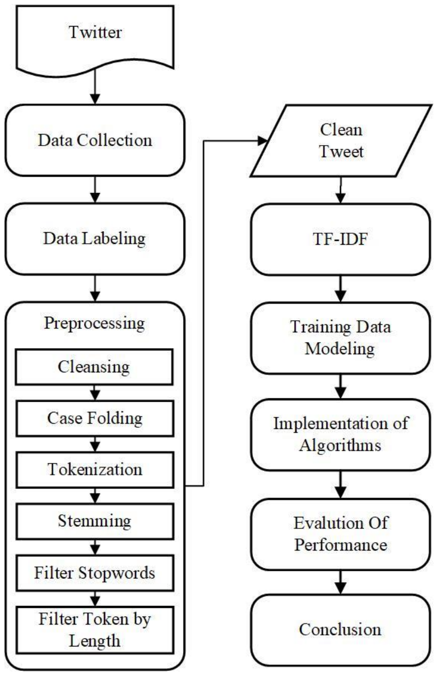

Figure 1. Research stages

The Twitter data analysis process began with collecting tweets and metadata from Twitter. This crucial step was essential to ensure the quality of the data for analysis. The data were then labelled for sentiment analysis, determining the public's reaction to a particular topic. After that, the tweets were cleaned through a preprocessing process. Data preprocessing involved cleansing, case folding, tokenizing, stemming, stop word filter and token by length filter to prepare the data for further analysis. The Frequency-Inverse Document Frequency (TF-IDF) method was used to determine keywords in the data. Training data modelling used machine learning algorithms to learn data patterns for implementing the Naive Bayes and KNN algorithms. Model performance evaluation used a confusion matrix to ensure the algorithm's accuracy and the recall and precision values for each sentiment. The conclusions from this analysis provided insight into public reactions and could be used to determine the current image of the police based on public sentiment regarding POLRI's performance.

2.1 Data Collection

This study used data from Twitter in Indonesian, obtained through a search using the keyword "kinerja POLRI." Data was collected using the crawling method via the Twitter API in the RapidMiner application and Fig. 2 shows an operator used in this step.

classification [23] by using Microsoft Excel. This labeling was only performed on a portion of the data designated as the training data. Out of the total data, 1028 data points were labeled, and 257 were left unlabeled, following an 80:20 split of the entire data set. The labeled training data were then used as a reference for the Naïve Bayes and KNN methods to classify the unlabeled test data. Next, the initial labelled training data was combined with the newly processed test data. Subsequently, a confusion matrix is used to evaluate the accuracy of the performed process. This labeling stage was crucial as it ensured the accuracy and validity of the sentiment analysis conducted [3]. Examples of tweets and their sentiment classification are presented in Table 1 to provide clearer expression variations and sentiment analysis used in this study.

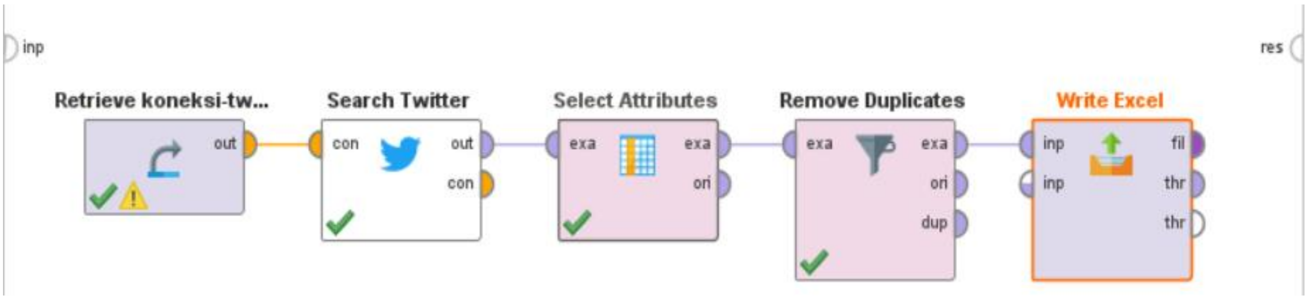

Figure 2. Operator in the data collection process

In the RapidMiner workspace shown in Fig 2, several operators were used in the data collection process: the retrieve connection operator which connects the application to Twitter, requiring an access token for authentication; the search Twitter operator used to input data search attributes; the select attributes operator used to select columns containing only public opinion texts; the remove duplicates operator used to eliminate sentiment data with identical values; and the write Excel operator which saved the sentiment data into an Excel file. From this crawling process, 1285 tweets were collected using the keyword "kinerja POLRI" from January 1, 2024, to May 16, 2024, and were saved into an Excel file. The importance of this data collection lay in its ability to support research by providing a solid foundation for in-depth analysis of public perceptions and helping to identify relevant trends and patterns in public opinion [12].

2.2 Data Labeling

In this stage, sentiment labels were assigned to the data to determine whether the data contained positive, negative, or neutral sentiments. This labeling stage was carried out manually by three researchers who classified the data based on majority opinion, ensuring the accuracy of the

Table 1. The Outcome of the Tweet Data Labeling Process

Tweet	Label
@mokhamadtoha2 @DivHumas_Polri @CCICPolri @ListyoSigitP Kita kecewa dengan kinerja POLRI dan penegak hukum di Indonesia	Negative
@Aryprasetyo85 @ListyoSigitP kinerja POLRI BOBROK 👎 👎 👎	Negative
Testimoni bapak Menkopolhukam RI atas kinerja Polri dalam Pengamanan Ops Ketupat Semeru 2023 https://t.co/KMCPGGSVwW #polisi_indonesia #ditlantaspoldajatim #MudikAmanBerkesan @NTMCLantasPolri @RTMCJatim @e100ss	Neutral
@janenuby terus maju untuk kinerja polri ini udah keren tetap Bersama Cegah Kejahatan #SinergiWujudkanKamtibmas	Positive
Selalu bangga ❤️ sama kinerja polri semangat selalu 👍 👍 #PolriIndonesiaHebatDensus88 https://t.co/zC8aB0pp4Z	Positive

2.3 Preprocessing Data

Preprocessing is the initial stage in preparing data for text mining, aiming to transform raw text into data ready for further processing [24]. This preprocessing process involves a series of steps, such as cleaning the text from irrelevant characters, tokenization to separate the text into individual words, removal of stop words, and text normalization [25]. The main goal of preprocessing is to gather significant and relevant words that will be used in sentiment analysis or other tasks in text mining [26]. By performing preprocessing correctly, we can enhance the data quality used in analysis, making the results more accurate and meaningful [27]. Therefore, preprocessing is a crucial and strategic step in supporting the success of further text analysis and ensuring that the data used is of good quality and relevance [28].

2.3.1 *Cleansing:* In RapidMiner, data cleansing involved the replace and filter operators which played a significant role in processing text data. This process explicitly removed irrelevant elements in tweet data, such as HTML tags, URLs, usernames, punctuation, and non-letter characters [29]. These steps kept the processed data focused on substantial text content, enabling more accurate and relevant analysis [30]. Thus, data cleansing became an essential initial step in text data processing and analysis using RapidMiner. The resulting text after the cleansing stage is shown in Table 2.

Table 2. The Outcome of Data Cleansing

Before	After	Label
@mokhamadtoha2 @DivHumas_POLRI @CCICPOLRI @ListyoSigitP Kita kecewa dengan kinerja POLRI dan penegak hukum di Indonesia	Kita kecewa dengan kinerja POLRI dan penegak hukum di Indonesia	Negative
@Aryprasetyo85 @ListyoSigitP kinerja POLRI BOBROK 👎 👎 👎	kinerja POLRI bobrok	Negative
Testimoni bapak Menkopolhukam RI atas kinerja POLRI dalam Pengamanan Ops Ketupat Semeru 2023 https://t.co/KMCPGGSV #polisi_indonesia #ditlantaspoldajatim #MudikAmanBerkesan @NTMCLantasPOLRI @RTMCJatim @e100ss	Testimoni bapak Menkopolhukam RI atas kinerja POLRI dalam pengamanan Ops Ketupat Semeru	Neutral
@janenuby terus maju untuk kinerja POLRI ini udah keren tetap Bersama Cegah Kejahatan #SinergiWujudkanKamtib	terus maju untuk kinerja POLRI ini udah keren tetap Bersama Cegah Kejahatan	Positive
Selalu bangga ❤️ sama kinerja POLRI semangat selalu 👍 👍 #POLRIIndonesiaHebat https://t.co/zC8aB0pp4Z	Selalu bangga sama kinerja POLRI semangat selalu	Positive

2.3.2 *Case Folding:* Using the Transform Case operator in RapidMiner to convert all words into lowercase aims to make the text format consistent. This process is crucial to ensure there is no differentiation between words with capital and lowercase letters, thus facilitating the processing of text analysis overall [31]. The resulting text after the case folding stage is shown in Table 3.

Table 3. The Outcome of the Tweet Data Case Folding

Before	After	Label
Kita kecewa dengan kinerja POLRI dan penegak hukum di Indonesia	kita kecewa dengan kinerja polri dan penegak hukum di indonesia	Negative
kinerja POLRI bobrok	kinerja polri bobrok	Negative
Testimoni bapak Menkopolhukam RI atas kinerja POLRI dalam pengamanan Ops Ketupat Semeru	testimoni bapak menkopolhukam ri atas kinerja polri dalam pengamanan ops ketupat semeru	Neutral
terus maju untuk kinerja POLRI ini udah keren tetap Bersama Cegah Kejahatan	terus maju untuk kinerja polri ini udah keren tetap bersama cegah kejahatan	Positive
Selalu bangga sama kinerja POLRI semangat selalu	selalu bangga sama kinerja polri semangat selalu	Positive

2.3.3 *Tokenization:* Tokenization splits text into words using the tokenize operator in text processing. The purpose of tokenization is to group words that appear, to let them be processed further in subsequent stages [32]. The resulting text after the tokenization stage is shown in Table 4.

Table 4. The Outcome of the Tweet Data Tokenization

Before	After	Label
kita kecewa dengan kinerja polri dan penegak hukum di indonesia	kita, kecewa, dengan, kinerja, polri, dan, penegak, hukum, di, indonesia	Negative
kinerja polri bobrok	kinerja, polri, bobrok	Negative
testimoni bapak menkopolhukam ri atas kinerja polri dalam pengamanan ops ketupat semeru	testimoni, bapak, menkopolhukam, ri, atas, kinerja, polri, dalam, pengamanan, ops, ketupat, semeru	Neutral
terus maju untuk kinerja polri ini udah keren tetap bersama cegah kejahatan	terus, maju, untuk, kinerja, polri, ini, udah, keren, tetap, bersama, cegah, kejahatan	Positive
selalu bangga sama kinerja polri semangat selalu	selalu, bangga, sama, kinerja, polri, semangat, selalu	Positive

2.3.4 *Stemming:* In general, sentences on Twitter often contain non-standard words that do not follow the standard spelling of the Indonesian language. Therefore, in this stage, these non-standard words were converted to adhere to the Enhanced Indonesian Spelling (EYD) and the use of stemming was significant as it could help identify the exact root words even in different forms, thus enhancing the efficiency and accuracy of text analysis overall

[33]. The resulting text after the stemming stage is shown in Table 5.

Table 5. The Outcome of the Tweet Data Stemming

Before	After	Label
kita, kecewa, dengan, kinerja, polri, dan, penegak, hukum, di, indonesia	kita, kecewa, dengan, kinerja, polri, dan, tegak, hukum, di, indonesia	Negative
kinerja, polri, bobrok testimoni, bapak, menkopolhukam, ri, atas, kinerja, polri, dalam, pengamanan, ops, ketupat, semeru	kinerja, polri, bobrok testimoni, bapak, menkopolhukam, ri, atas, kinerja, polri, dalam, aman, ops, ketupat, semeru	Negative Neutral
terus, maju, untuk, kinerja, polri, ini, udah, keren, tetap, bersama, cegah, kejahatan	terus, maju, untuk, kinerja, polri, ini, sudah, keren, tetap, sama, cegah, jahat	Positive
selalu, bangga, sama, kinerja, polri, semangat, selalu	selalu, bangga, sama, kinerja, polri, semangat, selalu	Positive

2.3.5 *Filter Stopwords:* In this step, a filtering process is performed to remove stopwords or common words that often appear and are usually ignored because they are considered insignificant [34]. Examples of words included in the stopwords category are pronouns, conjunctions, and others. This process uses the Filter Stopwords (dictionary) operator, referring to a stopwords dataset obtained from kaggle.com as a source. The importance of the Filter Stopwords in this process must be noted as they help ensure that only relevant and meaningful information is retained in the text analysis, producing more accurate and meaningful results [35]. The resulting text after the stopwords filtering stage is shown in Table 6.

Table 6. The Outcome of the Tweet Data Filter Stopwords

Before	After	Label
kita, kecewa, dengan, kinerja, polri, dan, penegak, hukum, di, indonesia	kecewa, kinerja, polri, tegak, hukum, indonesia	Negative
kinerja, polri, bobrok testimoni, bapak, menkopolhukam, ri, atas, kinerja, polri, dalam, pengamanan, ops, ketupat, semeru	kinerja, polri, bobrok testimoni, menkopolhukam, ri, kinerja, polri, aman, ops, ketupat, semeru	Negative Neutral
terus, maju, untuk, kinerja, polri, ini, udah, keren, tetap, bersama, cegah, kejahatan	terus, maju, kinerja, polri, sudah, keren, tetap, sama, cegah, jahat	Positive
selalu, bangga, sama, kinerja, polri, semangat, selalu	selalu, bangga, sama, kinerja, polri, semangat, selalu	Positive

2.3.6 *Filter Token By Length:* The next step is to filter words based on word length, where words with several letters outside the specified range will be removed. In this study, the minimum word length was set to four letters, and the maximum word length was set to 25 letters using the Filter Tokens (by Length) operator. The resulting text after the Filter Token By Length stage is shown in Table 7.

Table 7. The Outcome of the Tweet Data Filter Token by Length

Before	After	Label
kecewa, kinerja, polri, tegak, hukum, indonesia	kecewa, kinerja, polri, tegak, hukum, indonesia	Negative
kinerja, polri, bobrok testimoni, menkopolhukam, ri, kinerja, polri, aman, ops, ketupat, semeru	kinerja, polri, bobrok testimoni, menkopolhukam, kinerja, polri, aman, ketupat, semeru	Negative Neutral
terus, maju, kinerja, polri, sudah, keren, tetap, sama, cegah, jahat	terus, maju, kinerja, polri, sudah, keren, tetap, sama, cegah, jahat	Positive
selalu, bangga, sama, kinerja, polri, semangat, selalu	selalu, bangga, sama, kinerja, polri, semangat, selalu	Positive

3 RESULT AND DISCUSSION

3.1 TF-IDF Weighting

After the preprocessing stage, the next step is to assign word weights using the TF-IDF method. This method determines the data value based on the frequency of word occurrences in the documents [36]. The frequency of each word's appearance in the document is calculated and then weighted based on the occurrence against the total number of documents containing that word. This process uses the Process Documents from Data operator, which applies the TF-IDF method. A representation of the TF-IDF weighting process is shown in Fig. 3.

The researcher also pictured the data from the TF-IDF weighting results by creating a word cloud to display the 20 most frequently occurring words. In Fig. 4 the word "polisi" dominates with a frequency of 1430, followed by "Indonesia" with 1313, and "POLRI" with 1224 occurrences.

3.2 Data Splitting

The division of training and testing data in sentiment analysis is a crucial step in evaluating the model accurately. The standard ratio commonly used in data division is 80:20. In this ratio, 80% of the data were allocated for training the model. In comparison, the remaining 20% was reserved for assessing the model's performance. In the previous labeling stage, out of 1285 data points, 1028 labeled data were assigned as training data, while the remaining 257 unlabeled data were designated as testing data. These training data were then employed to train the model using Naïve Bayes and KNN methods. Subsequently, the labeled training data were combined with the prediction results from the testing data to determine the overall sentiment of the dataset.

Row No.	word	in documents	total	in class (po...	in class (ne...	in class (net...
1	polisi	1167	1430	413	285	732
2	indonesia	1155	1313	459	207	647
3	polri	820	1224	443	22	759
4	kinerja	1043	1076	379	205	492
5	tahun	289	470	360	14	96
6	lebaran	214	419	8	0	411
7	jakarta	408	409	167	0	242
8	wujud	380	380	175	0	205
9	tingkat	364	368	46	10	312
10	presisi	327	327	22	2	303
11	kapolri	302	304	19	7	278
12	aman	283	291	45	12	234
13	prabowo	287	290	13	0	277
14	jenderal	287	287	9	0	278
15	republik	273	275	209	13	53

Figure 3. Words after tf-idf weighting

Figure 4. Data visualization in word cloud form

Several vital stages involve various operators in the workspace in the data training modeling process using RapidMiner Studio. Fig. 5 demonstrates the use of operators for the Naive Bayes algorithm. The initial stage begins with the "Process Document from Data" operator in the data preprocessing. This operator is then connected to the Naive Bayes algorithm. Subsequently, the output from this algorithm is stored using two "Store" operators. The first "Store" operator, renamed "Store Model," saves the model generated by the Naive Bayes process. The second "Store" operator, "Store Training Data," saves the training data used in this process.

Fig. 6 shows that a similar process is applied to the KNN algorithm. The stages start with the "Process Document from Data" operator for data preprocessing, which is then connected to the KNN algorithm. The output from the KNN algorithm is also stored using two "Store" operators. The first operator, named "Store Model," is used to save the model generated by the KNN algorithm. The second operator, "Store Training Data," saves the processed training data.

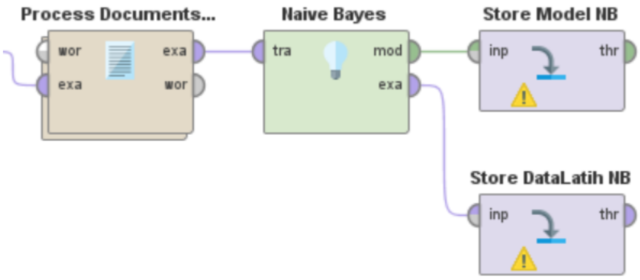

Figure 5. Data modeling operators in the Naive Bayes algorithm

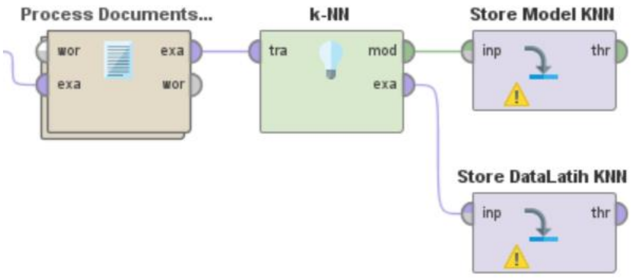

Figure 6. Data modeling operators in the Naive Bayes algorithm

3.3 *Algorithm Implementation*

The Naïve Bayes method was implemented as shown in Fig. 7. It commences with the "Read CSV" operator to access the prepared CSV file, followed by the connection to the "Filter Example" operator. This operator is configured with the "is missing" option to extract and process sentiment data that is still unlabeled or vacant in the training dataset. Subsequently, to facilitate data processing, the "Nominal to Text" operator is employed to enable the "Process Document from Data" operator to read and process data, with the parameter set to "TF-IDF" under the "vector creation" option. The "Union" operator also merges the test data with the previously constructed training data. Following this, reconnection to the "Filter Example" operator is established with the condition "is missing" to address the unlabeled sentiment texts for testing and processing. The "Replace Missing Value" operator addresses data processing gaps with the default condition set to "zero" to prevent system confusion with empty data values. Finally, the "Apply Model" operator is incorporated to store and display the resulting sentiments generated from the Naïve Bayes processing.

Next, the processing of the Naive Bayes algorithm which the results are given in Fig. 8. It displays the number of negative, neutral, and positive sentiments from the data. In Fig. 8, 1285 data have been examined using Naïve Bayes. The results show that positive sentiments form 43.6% or 560 tweets, while neutral sentiments are 38.1% or 489 tweets; on the other hand, negative sentiments make up 18.4% or 238 tweets.

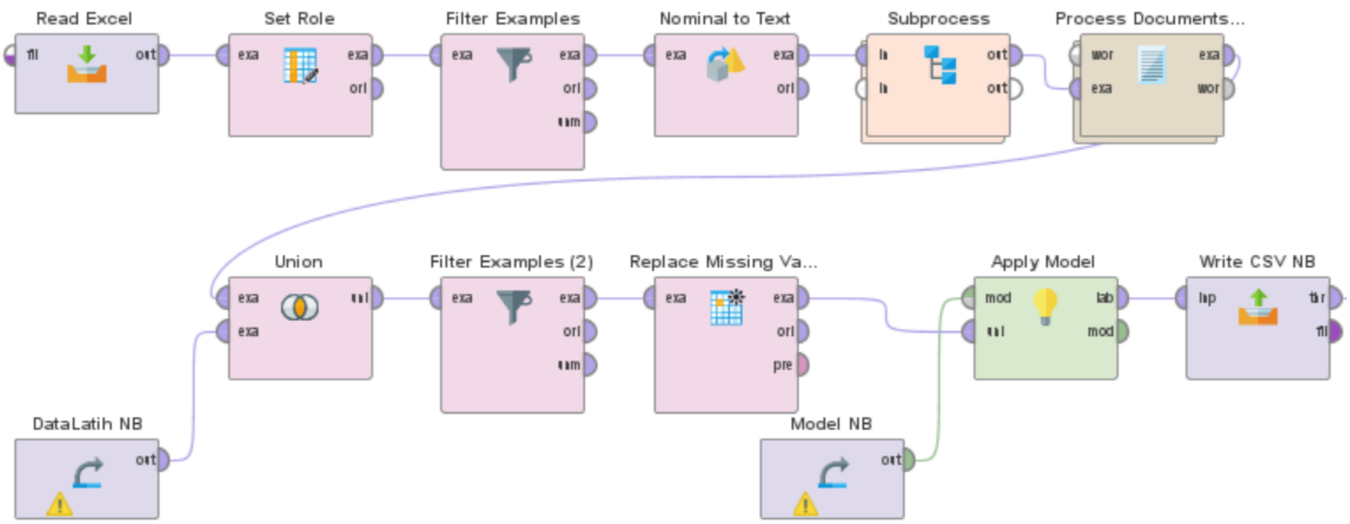

Figure 7. Implementing the Naïve Bayes operators for data testing

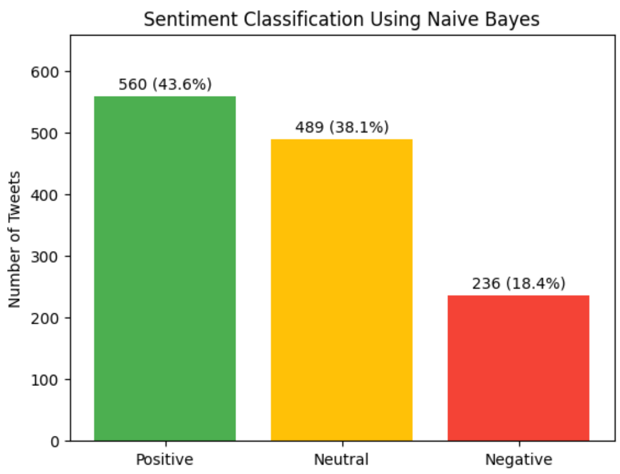

Figure 8. Percentage of sentiment results with Naive Bayes

Then, the KNN method shown in Fig. 9 is implemented. It is done using the "Read CSV" operator to read the prepared CSV file and then connecting it to the "Filter Example" operator. This operator is set to the "is missing" option to handle sentiments that are still empty or unlabeled in the training data, which will be taken and processed. Furthermore, to process the data, the "Nominal to Text" operator is required so that the "Process Document from Data" operator can read and process the data with the parameter set to "TF-IDF" in the "vector creation" option. The "Union" operator also merges the data to be tested with the created training data. Afterwards, a connection is made with the "Filter Example" operator with the condition "is missing" to handle empty sentiment texts that will be tested and processed. To fill in the gaps during data processing, the "Replace Missing Value" operator with the default condition "zero" is required so that the system is not confused by empty data values. Finally, the "Apply Model" operator is added to store and display the sentiment results generated from the KNN processing. Next, the processed dataset labeled using the KNN can be pictured in Fig. 10, to display the number of negative, neutral, and positive sentiments from the data.

In Fig. 10, 1285 data have been examined using KNN. The results show that positive sentiments are 41.8% or 537 tweets, while neutral sentiments are 38.8% or 498 tweets; on the other hand, negative sentiments make up 19.5% or 250 tweets.

Based on the results, it is evident that the majority of the public expresses positive sentiments, followed by neutral sentiments, and some negative sentiments. Analysis of positive sentiments indicates an appreciation for the performance of the POLRI in handling cases and providing good public service. This can serve as input for POLRI to strengthen positive aspects by maintaining or improving effective policies. Meanwhile, neutral sentiments present opportunities to enhance the image by identifying characteristics that require further attention. Analysis of negative sentiments highlights dissatisfaction and complaints regarding the lack of professionalism in handling cases and unfair treatment in public service by POLRI. This is a basis for evaluating and addressing complaints and issues through enhanced professionalism training. By using sentiment analysis, POLRI can make more accurate and strategic decisions to improve its image in the eyes of the public, ensuring effective responses to public needs and expectations.

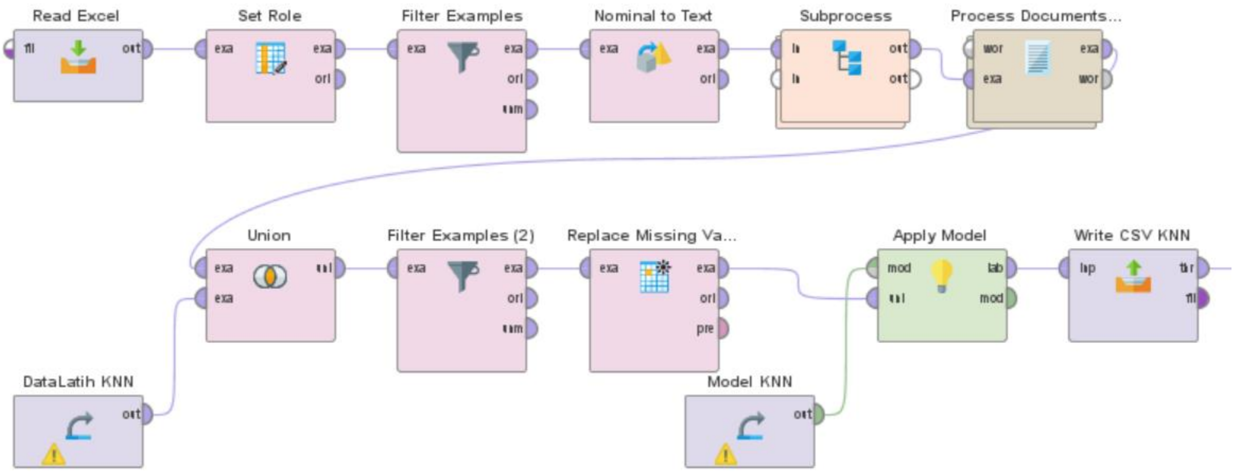

Figure 9. Implementing the KNN operators for data testing

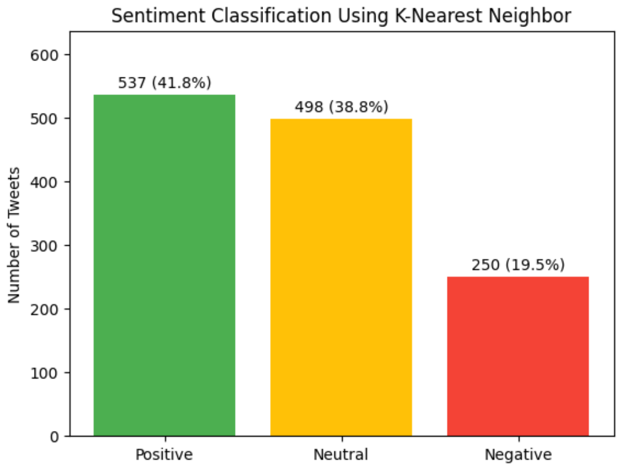

Figure 10. Percentage of sentiment results with KNN

3.4 *Evaluation Testing*

After conducting research in the form of sentiment analysis on the dataset using the Naïve Bayes and KNN methods, the next step is testing with the Confusion Matrix. This testing aims to calculate the accuracy rate based on evaluation and validation. The Confusion Matrix calculates the accuracy rate using three primary parameters: accuracy, precision, and recall. Accurate assessment of these metrics is essential to understand how well the model distinguishes between different sentiment classes, providing insights into its effectiveness and reliability. The values obtained from the confusion matrix can be used to evaluate the model's performance by calculating the algorithm's accuracy, as well as the precision, recall and f1-score for each sentiment class with Equations 1-4 provide the necessary formulas for these performance metrics [37].

$$Accuracy = \frac{(TP_{negative} + TP_{positive} + TP_{netral})}{(Total\ Cases)} \quad (1)$$

$$Precision = \frac{(TP)}{(TP+FP)} \quad (2)$$

$$Recall = \frac{(TP)}{(TP+FN)} \quad (3)$$

$$F1\ score = 2\ x \frac{(Precision\ x\ Recall)}{(Precision + Recall)} \quad (4)$$

Subsequently, the accuracy calculation for the Naïve Bayes algorithm is performed using the RapidMiner application. Fig. 11 shows the operators in calculating the Confusion Matrix with the Naïve Bayes algorithm in RapidMiner.

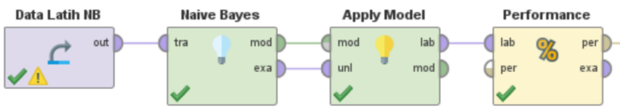

Figure 11. Operators for calculating confusion matrix on Naive Bayes

The previously created Naive Bayes data model will be reconnected with the "Naïve Bayes" and "Apply Model" operators. Then, it is concluded with the "Performance" operator to present the confusion matrix results as shown in Table 8.

Table 8. Confusion Matrix Result with Naive Bayes

	Prediction Negative	**Prediction Neutral**	**Prediction Positive**
True Negative	194	0	9

True Neutral	0	452	0
True Positive	0	1	372

From the results shown in Table 8 above, we can also see 194 cases considered True Negative. There is a null case in the negative sentiment category, considered neutral. There are 9 cases of negative sentiment in the harmful sentiment category, which is predicted as positive. In the neutral sentiment category, there are 452 cases of True Neutral. There is a null case in the neutral sentiment category, which is predicted as a negative sentiment. There is a null case in the neutral sentiment category, which is predicted as a positive sentiment. In the positive sentiment category, there are 372 cases as True Positive. There are 0 cases of positive sentiment in the positive sentiment category, which is predicted as negative. There is 1 case of positive sentiment in the positive sentiment category, which is forecast to be neutral.

The manual calculation was done to obtain the accuracy, recall, precision, and F1-score values for each sentiment with the Naive Bayes algorithm. It is as follows:

Accuracy with Naive Bayes:

$$TP_{negative} = 194$$

$$TP_{neutral} = 452$$

$$TP_{positive} = 372$$

$$Total\ Cases = 194 + 0 + 9 + 452 + 0 + 0 + 372 + 0 + 1$$

$$Total\ Cases = 1028$$

$$Accuracy = \frac{194+\ 452+372}{1028}$$

$$Accuracy = \ 0.99027\ or\ 99.03\ \%$$

Sentiment Negative with Naive Bayes:

$$TP_{negative} = 194$$

$$FP_{negative} = 0 + 0$$

$$FP_{negative} = 0$$

$$FN_{negative} = 0 + 9$$

$$FN_{negative} = 9$$

$$Precision_{negative} = \frac{(TP_{negative})}{(TP_{negative}+FP_{negative})}$$

$$Precision_{negative} = \frac{194}{194+0}$$

$$Precision_{negative} = \ 1\ or\ 100\%$$

$$Recall_{negative} = \frac{(TP_{negative})}{(TP_{negative}+FN_{negative})}$$

$$Recall_{negative} = \frac{194}{194+9}$$

$$Recall_{negative} = \ 0.95566\ or\ 95.57\%$$

$$F1\ score_{negative} = \ 2\ x \frac{(Precision_{negative}\ x\ Recall_{negative})}{(Precision_{negative} + Recall_{negative})}$$

$$F1\ score_{negative} = \ 2\ x \frac{(1\ x\ 0.95566)}{(1+\ 0.95566)}$$

$$F1\ score_{negative} = \ 2\ x \frac{0.95566}{1.95566}$$

$$F1\ score_{negative} = 0.97733\ or\ 97.73\%$$

Sentiment Neutral with Naive Bayes:

$$TP_{neutral} = 452$$

$$FP_{neutral} = 0 + 1$$

$$FP_{neutral} = 1$$

$$FN_{neutral} = 0 + 0$$

$$FN_{neutral} = 0$$

$$Precision_{neutral} = \frac{(TP_{neutral})}{(TP_{neutral}+FP_{neutral})}$$

$$Precision_{neutral} = \frac{452}{452+1}$$

$$Precision_{neutral} = \ 0.99779\ or\ 99.78\%$$

$$Recall_{neutral} = \frac{(TP_{neutral})}{(TP_{neutral}+FN_{neutral})}$$

$$Recall_{neutral} = \frac{452}{452+0}$$

$$Recall_{neutral} = \ 1\ or\ 100\%$$

$$F1\ score_{neutral} = \ 2\ x \frac{(Precision_{neutral}\ x\ Recall_{neutral})}{(Precision_{neutral} + Recall_{neutral})}$$

$$F1\ score_{neutral} = 2\ x \frac{(0.99779\ x\ 1)}{(0.99779 + 1)}$$

$$F1\ score_{neutral} = 2\ x \frac{0.99779}{1.99779}$$

$$F1\ score_{neutral} = 0.99889\ or\ 99.89\%$$

Sentiment Positive with Naive Bayes:

$$TP_{positive} = 372$$

$$FP_{positive} = 9 + 0$$

$$FP_{positive} = 9$$

$$FN_{positive} = 0 + 1$$

$$FN_{positive} = 1$$

$$Precision_{positive} = \frac{(TP_{positive})}{(TP_{positive} + FP_{positive})}$$

$$Precision_{positive} = \frac{372}{372+9}$$

$$Precision_{positive} = 0.97637\ or\ 97.64\%$$

$$Recall_{positive} = \frac{(TP_{positive})}{(TP_{positive} + FN_{positive})}$$

$$Recall_{positive} = \frac{372}{372+1}$$

$$Recall_{positive} = 0.99731\ or\ 99.73\%$$

$$F1\ score_{positive} = 2\ x \frac{(Precision_{positive}\ x\ Recall_{positive})}{(Precision_{positive} + Recall_{positive})}$$

$$F1\ score_{positive} = 2\ x \frac{(0.97637\ x\ 0.99731)}{(0.97637 + 0.99731)}$$

$$F1\ score_{positive} = 2\ x \frac{0.97376}{1.97369}$$

$$F1\ score_{positive} = 0.98673\ or\ 98.67\%$$

The confusion matrix results shown in Table 8 indicate an accuracy rate of 99.03% for applying the Naive Bayes algorithm. The recall values are 99.73% for positive sentiment, 100% for neutral sentiment, and 95.57% for negative sentiment. The precision values are 97.64% for positive sentiment, 99.78% for neutral sentiment, and 100% for negative sentiment. Meanwhile, the F1-score are 98.67% for positive sentiment, 99.89% for neutral sentiment, and 97.73% for negative sentiment.

Subsequently, the accuracy calculation for the KNN algorithm is performed using the RapidMiner application. Fig. 12 shows the operators in calculating the Confusion Matrix for the KNN algorithm in RapidMiner. The previously created KNN data model will be reconnected with the "KNN" and "Apply Model" operators and concluded with the "Performance" operator to present the confusion matrix results as shown in Table 9.

Table 9. Confusion Matrix Result with KNN

	Prediction Negative	**Prediction Neutral**	**Prediction Positive**
True Negative	185	1	17
True Neutral	0	450	2
True Positive	27	1	345

From the results in Table 9 above, we can also find 185 cases considered as True Negative. There is 1 case of negative sentiment in the negative sentiment category, considered neutral. There were 17 cases of negative sentiment in the harmful sentiment category, which were predicted as positive. In the neutral sentiment category, there were 450 cases of True Neutral. There is a null case in the neutral sentiment category, which is predicted as a negative sentiment. There are 2 cases of neutral sentiment in the neutral sentiment category, which are predicted as positive sentiment. In the positive sentiment category, there are 345 cases as True Positive. There are 27 cases of positive sentiment in the positive sentiment category, which are predicted as negative. There is 1 case of positive sentiment in the positive sentiment category, which is forecast to be neutral.

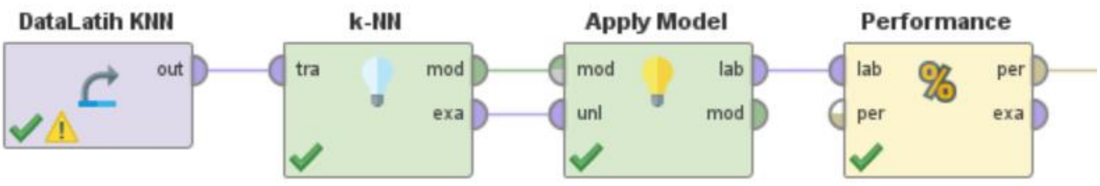

Figure 12. Operators for calculating confusion matrix on KNN

The manual calculation was also done to obtain the accuracy, recall, and precision values for each sentiment from the KNN. It is as follows:

Accuracy with KNN:

$$TP_{negative} = 185$$

$$TP_{neutral} = 450$$

$$TP_{positive} = 345$$

$$Total\ Cases = 185 + 1 + 17 + 450 + 0 + 0 + 345 + 27 + 1$$

$$Total\ Cases = 1028$$

$$Accuracy = \frac{185 + 450 + 345}{1028}$$

$$Accuracy = 0.95330\ or\ 95.33\ \%$$

Sentiment Negative with KNN:

$$TP_{negative} = 185$$

$$FP_{negative} = 0 + 27$$

$$FP_{negative} = 27$$

$$FN_{negative} = 1 + 17$$

$$FN_{negative} = 18$$

$$Precision_{negative} = \frac{(TP_{negative})}{(TP_{negative} + FP_{negative})}$$

$$Precision_{negative} = \frac{185}{185+27}$$

$$Precision_{negative} = 0.87264\ or\ 87.26\%$$

$$Recall_{negative} = \frac{(TP_{negative})}{(TP_{negative} + FN_{negative})}$$

$$Recall_{negative} = \frac{185}{185+18}$$

$$Recall_{negative} = 0.91133\ or\ 91.13\%$$

$$F1\ score_{negative} = 2\ x \frac{(Precision_{negative}\ x\ Recall_{negative})}{(Precision_{negative} + Recall_{negative})}$$

$$F1\ score_{negative} = 2\ x \frac{(0.87264\ \ x\ 0.91133)}{(0.87264 + 0.91133)}$$

$$F1\ score_{negative} = 2\ x \frac{0.79526}{1.78397}$$

$$F1\ score_{negative} = 0.89156\ or\ 89.16\%$$

Sentiment Neutral with KNN:

$$TP_{neutral} = 450 FP_{neutral} = 1 + 1$$

$$FP_{neutral} = 2$$

$$FN_{neutral} = 0 + 2$$

$$FN_{neutral} = 2$$

$$Precision_{neutral} = \frac{(TP_{neutral})}{(TP_{neutral} + FP_{neutral})}$$

$$Precision_{neutral} = \frac{450}{450+2}$$

$$Precision_{neutral} = 0.99557\ or\ 99.56\%$$

$$Recall_{neutral} = \frac{(TP_{neutral})}{(TP_{neutral} + FN_{neutral})}$$

$$Recall_{neutral} = \frac{450}{450+2}$$

$$Recall_{neutral} = 0.99557\ or\ 99.56\%$$

$$F1\ score_{neutral} = 2\ x \frac{(Precision_{neutral}\ x\ Recall_{neutral})}{(Precision_{neutral} + Recall_{neutral})}$$

$$F1\ score_{neutral} = 2\ x \frac{(0.99557\ x\ 0.99557)}{(0.99557 + 0.99557)}$$

$$F1\ score_{neutral} = 2\ x \frac{0.99117}{1.99115}$$

$$F1\ score_{neutral} = 0.99557\ or\ 99.56\%$$

Sentiment Positive with KNN:

$$TP_{positive} = 345$$

$$FP_{positive} = 17 + 2$$

$$FP_{positive} = 19$$

$$FN_{positive} = 27 + 1$$

$$FN_{positive} = 28$$

$$Precision_{positive} = \frac{(TP_{positive})}{(TP_{positive}+FP_{positive})}$$

$$Precision_{positive} = \frac{345}{345+19}$$

$$Precision_{positive} = 0.94780 \; or \; 94.78\%$$

$$Recall_{positive} = \frac{(TP_{positive})}{(TP_{positive}+FN_{positive})}$$

$$Recall_{positive} = \frac{345}{345+28}$$

$$Recall_{positive} = 0.92493 \; or \; 92.49\%$$

$$F1\ score_{positive} = 2\,x\,\frac{(Precision_{positive}\ x\ Recall_{positive})}{(Precision_{positive} + Recall_{positive})}$$

$$F1\ score_{positive} = 2\,x\,\frac{(0.94780\ \ x\ 0.92493)}{(0.94780 + 0.92493)}$$

$$F1\ score_{positive} = 2\,x\,\frac{0.87665}{1.87273}$$

$$F1\ score_{positive} = 0.93622 \; or \; 93.62\%$$

The confusion matrix results in Table 9 indicate an accuracy rate of 95.33% for applying the KNN algorithm. The recall values are 92.49% for positive sentiment, 99.56% for neutral sentiment, and 91.13% for negative sentiment. The precision values are 94.78% for positive sentiment, 99.56% for neutral sentiment, and 87.26% for negative sentiment. Meanwhile, the F1-score are 93.62% for positive, 99.56% for neutral, and 89.16%for negative sentiments.

4 CONCLUSION

From the data collection from Twitter, 1285 data points were obtained. Using the Naive Bayes algorithm, the number of positive sentiments obtained was 560 data points or 43.6%, neutral sentiments amounted to 489 data points or 38.1%, and negative sentiments amounted to 236 data points or 18.4%. In the implementation of the K-Nearest Neighbors (KNN) algorithm, the number of positive sentiments obtained was 537 data points or 41.8%, neutral sentiments amounted to 498 data points or 38.8%, and negative sentiments amounted to 250 data points or 19.5%. From these two algorithms, it can be concluded that many people are satisfied with the performance of the Indonesian National Police (POLRI), as indicated by the high number of positive sentiment points. However, some people are still dissatisfied, as seen from the negative sentiment points. This can be valuable input for evaluation to improve the performance of the police to enhance the institution's image.

The algorithm evaluation testing process using the confusion matrix yielded the following results: The application of the Naive Bayes algorithm attained an accuracy of 99.03%, with recall values of 99.73% for positive sentiment, 100% for neutral sentiment, and 95.57% for negative sentiment. The precision values obtained were 97.64% for positive sentiment, 99.78% for neutral sentiment, and 100% for negative sentiment. Meanwhile, the F1-score are 98.67% for positive sentiment, 99.89% for neutral sentiment, and 97.73% for negative sentiment. Using the KNN algorithm, an accuracy of 95.33% was achieved, with recall values of 92.49% for positive sentiment, 99.56% for neutral sentiment, and 91.13% for negative sentiment. The precision values obtained were 94.78% for positive sentiment, 99.56% for neutral sentiment, and 87.26% for negative sentiment. Meanwhile, the F1-score are 93.62% for positive sentiment, 99.56% for neutral sentiment, and 89.16%for negative sentiment. These results indicate that both algorithms have a relatively high level of accuracy and can be used in the sentiment analysis process. The results of this dataset processing may be considered for future decision-making and innovations by the POLRI, aiming to enhance harmonious relations with the Indonesian public.

AUTHOR'S CONTRIBUTION

Yusuf Handika passionately initiated this groundbreaking research and meticulously labeled the data, while Isa Faqihuddin Hanif diligently took charge of data collection and contributed to data labeling. The detailed and insightful data analysis was expertly conducted by Firman Noor Hasan, who additionally played a significant role in data labeling. Furthermore, Yusuf Handika also significantly contributed to designing the study and providing critical manuscript reviews. Isa Faqihuddin Hanif was crucial in overseeing the research process to ensure smooth execution and strict adherence to the established methodology.

COMPETING INTERESTS

Following the publication ethics of this journal, Yusuf Handika, Isa Faqihuddin Hanif, Firman Noor Hasan as the authors of this article declare that this article is free from conflict of interest (COI) and competing interest (CI).

ACKNOWLEDGMENT

The authors extend the deepest gratitude to everyone who has provided valuable input, moral and material support, constructive corrections, and assistance in various forms throughout the research process, enabling this study to be completed successfully.

REFERENCES

[1] I. R. Afandi, N. Pratiwi, A. A. Rizki, M. Irva, and M. F. Aulia,

"Perancangan Sistem Informasi Pelayanan Pembuatan Surat Online Di Desa Ciangsana Bebasis Website," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 571–577, 2022, doi: 10.36040/jati.v6i2.5318.

[2] L. A. Susanto, "Komparasi Model Support Vector Machine Dan K-Nearest Neighbor Pada Analisis Sentimen Aplikasi Polri Super App," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4152.

[3] P. Afikah, I. R. Afandi, and F. N. Hasan, "Implementasi Business Intelligence Untuk Menganalisis Data Kasus Virus Corona di Indonesia Menggunakan Platform Tableau," *Pseudocode*, vol. 9, no. 1, pp. 25–32, 2022, doi: 10.33369/pseudocode.9.1.25-32.

[4] K. Azkya and H. N. Y. Haloho, "Pengaruh Berita Penanganan Kasus oleh Kepolisian Terhadap Citra POLRI di Kalangan Remaja," *KALBISOCIO J. Bisnis dan Komun.*, vol. 10, no. 2, pp. 64–78, 2023, doi: https://doi.org/10.53008/kalbisocio.v10i2.2145.

[5] A. Saputra, D. Sutrasno, and W. Setiawan, "Analisa Analisis Tingkat Kepercayaan Masyarakat terhadap Kinerja Polri Tahun 2022," *J. Litbang Polri*, vol. 26, no. 2, pp. 78–89, 2023, doi: https://doi.org/10.46976/litbangpolri.v26i2.215.

[6] H. Y. Siregar, A. Akim, and Y. Djuyandi, "Analisis Kebijakan Pengawasan Program Dan Kegiatan Kepolisian Di Polres Sorong Selatan," *J. Pemikir. Dan Penelit. Bid. Adm. Sos. Hum. Dan Kebijak. Publik*, vol. 6, no. 1, p. 9, 2023, doi: https://doi.org/10.24198/responsive.v6i1.40029.

[7] E. D. Shantika and A. Handayani, "ANALISIS BUDAYA KERJA KEPOLISIAN DAERAH LAMPUNG DALAM PELAYANAN TERHADAP MASYARAKAT (Studi Pada Bidang Profesi dan Pengamanan Polda Lampung)," vol. 3, no. 6, 2024, doi: https://doi.org/10.6578/triwikrama.v3i6.2883.

[8] I. A. Ricky, I. F. Hanif, F. N. Hasan, E. S. Sinduningrum, Z. Halim, and N. Nunik, "Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression," *J. Linguist. Komputasional*, vol. 5, no. 2, p. 77, 2022, [Online]. Available: https://t.co/23c4krbjp

[9] S. H. W. Putra and D. Febriawan, "Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes pada Google Play Store," *Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, 2024, doi: https://doi.org/10.30865/klik.v4i4.1600.

[10] L. Lestari, M. D. R Wahyudi, and U. Kiftiyani, "Veil and Hijab: Twitter Sentiment Analysis Perspective," *IJID (International J. Informatics Dev.*, vol. 9, no. 1, p. 52, 2020, doi: 10.14421/ijid.2020.09108.

[11] I. M. Noor and M. Turan, "Sentiment Analysis on New Currency in Kenya using Twitter Dataset," *IJID (International J. Informatics Dev.*, vol. 8, no. 2, p. 81, 2020, doi: 10.14421/ijid.2019.08206.

[12] F. M. Sarimole and W. Septian, "Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 890–899, 2024, doi: https://doi.org/10.55338/saintek.v5i1.2789.

[13] I. R. Afandi, F. Noor Hasan, A. A. Rizki, N. Pratiwi, and Z. Halim, "Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes," *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 63–70, 2022, doi: https://doi.org/10.26418/jlk.v5i2.107.

[14] A. F. L. Ramadhina and E. Sofian, "Perbandingan Metode K-Nearest Neighbor dan Pohon Keputusan dalam Analisis Sentimen Data Ulasan Aplikasi Pinjaman Online Berizin OJK Di Google Play," *J. Sist. Komput. dan Kecerdasan Buatan*, vol. 7, no. 2, pp. 115–124, 2024, doi: https://doi.org/10.47970/siskom-kb.v7i1.543.

[15] T. A. Azzahra *et al.*, "Perbandingan Efektivitas Naïve Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube," *J. Media Inform. Budidarma*, vol. 8, no. 1, pp. 312–322, 2024, doi: 10.30865/mib.v8i1.7186.

[16] S. R. Azizah, R. Herteno, A. Farmadi, D. Kartini, and I. Budiman, "Kombinasi Seleksi Fitur Berbasis Filter dan Wrapper Menggunakan Naive Bayes pada Klasifikasi Penyakit Jantung," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 6, pp. 1361–1368, 2023, doi: 10.25126/jtiik.1067467.

[17] A. N. Iffah'da and Anita Desiani, "Implementasi Algoritma K-Nearest Neighbor (K-NN) dan Single Layer Perceptron (SLP) Dalam Prediksi Penyakit Sirosis Biliari Primer," *J. Ilm. Inform.*, vol. 7, no. 1, pp. 65–74, 2022, doi: 10.35316/jimi.v7i1.65-74.

[18] Muhammad Jauhar Vikri and R. Rohmah, "Penerapan Fungsi Exponential Pada Pembobotan Fungsi Jarak Euclidean Algoritma K-Nearest Neighbor," *Gener. J.*, vol. 6, no. 2, pp. 57–64, 2022, doi: 10.29407/gj.v6i2.18070.

[19] R. N. Fahmi, N. Nursyifa, and A. Primajaya, "Analisis Sentimen Pengguna Twitter Terhadap Kasus Penembakan Laskar Fpi Oleh Polri Dengan Metode Naive Bayes Classifier," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 61–66, 2021, doi: http://dx.doi.org/10.26798/jiko.v5i2.437.

[20] E. R. Kaburuan and N. R. Setiawan, "Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 105–116, 2023, doi: 10.32736/sisfokom.v12i1.1614.

[21] E. P. Harahap, H. D. Purnomo, A. Iriani, I. Sembiring, and T. Nurtino, "Trends in sentiment of Twitter users towards Indonesian tourism: analysis with the k-nearest neighbor method," *Comput. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 13–22, 2024, doi: 10.11591/csit.v5i1.pp13-22.

[22] T. D. Arista, Yusra, M. Fikry, and L. Oktavia, "Klasifikasi Sentimen Masyarakat di Twitter terhadap Kenaikan Harga BBM dengan Metode K-NN," *J. Komput. dan Inform.*, vol. 5, no. 1, pp. 140–150, 2023, doi: https://doi.org/10.59934/juki.v5i1.189.

[23] D. P. Ray, F. N. Hasan, and A. R. Dzikrillah, "Analisis Sentimen Terhadap KPU 2024 Berdasarkan Tweet Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 197–206, 2024, doi: 10.30865/klik.v4i4.1587.

[24] R. A. Saputra, D. P. Ray, and F. Irwiensyah, "Analisis Sentimen Aplikasi Tokocrypto Berdasarkan Ulasan Pada Google Play Store Menggunakan Metode Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 2028–2036, 2024, doi: 10.30865/klik.v4i4.1707.

[25] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes," *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 115–122, 2023, doi: https://doi.org/10.55338/saintek.v4i3.1368.

[26] M. Fudhail, F. Supeli, and Setiaji, "Klasifikasi Sentimen Positif Dan Negatif Pada Aplikasi Vidio Dengan Algoritma K-Nearest Neighbor," *Indones. J. Comput. Sci.*, vol. 2, no. 1, pp. 7–15, 2023, doi: https://doi.org/10.31294/ijcs.v2i1.1874.

[27] H. Z. Muflih, A. R. Abdillah, and F. N. Hasan, "Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1613–1621, 2023, doi: 10.30865/klik.v4i3.1303.

[28] J. F. Sianipar, Y. R. Ramadhan, and I. Jaelani, "Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 360–367, 2023, doi: 10.30865/klik.v4i1.1033.

[29] K. A. Lubis, M. T. A. Bangsa, and A. Yudertha, "ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP PINDAHNYA IBU KOTA INDONESIA DENGAN

MENGGUNAKAN KLASIFIKASI NAÏVE BAYES," *J. TEKNOINFO*, vol. 18, no. 1, pp. 226–238, 2024, doi: https://doi.org/10.33365/jti.v18i1.3708.

[30] E. D. Nurhazizah and I. Puspitasari, "OPINION MINING FUNGSI KPI (KEY PERFORMANCE INDIKATOR) DENGAN ALGORITMA NAÏVE BAYES CLASIFIER DAN SUPPORT VECTOR MACHINE (SVM)," *J. Cahaya Mandalika*, vol. 3, no. 2, pp. 290–302, 2023, doi: https://doi.org/10.36312/jcm.v3i2.1534.

[31] A. Saputra and F. N. Hasan, "ANALISIS SENTIMEN TERHADAP APLIKASI COFFEE MEETS BAGEL DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER," vol. 2, no. 2, pp. 465–474, 2023, doi: https://doi.org/10.54443/sibatik.v2i2.579.

[32] M. Samantri and Afiyati, "Perbandingan Algoritma Support Vector Machine dan Random Forest untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 1, pp. 1–9, 2024, doi: 10.35870/jtik.v8i1.1202.

[33] G. Radiena and A. Nugroho, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine," *J. Pendidik. Teknol. Inf.*, vol. 6, no. 1, pp. 1–10, 2023, doi: 10.37792/jukanti.v6i1.836.

[34] Fauzan Baehaqi and N. Cahyono, "Analisis Sentimen Terhadap Cyberbullying Pada Komentar Di Instagram Menggunakan Algoritma Naïve Bayes," *Indones. J. Comput. Sci.*, vol. 13, no. 1, pp. 1051–1063, 2024, doi: 10.33022/ijcs.v13i1.3301.

[35] D. Pratmanto and F. F. D. Imaniawan, "Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors," *Comput. Sci.*, vol. 3, no. 2, pp. 110–117, 2023, doi: 10.31294/coscience.v3i2.1917.

[36] J. E. B. Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 42–53, 2024, doi: 10.34010/jamika.v14i1.11946.

[37] P. W. Hardjita, Nurochman, and R. Hidayat, "Sentiment Analysis of Tweets on Prakerja Card using Convolutional Neural Network and Naive Bayes," *IJID (International J. Informatics Dev.*, vol. 10, no. 2, pp. 82–91, 2022, doi: 10.14421/ijid.2021.3007.