

# Modeling of Naïve Bayes and Decision Tree Algorithms to Analyze Sentiment Related to Jaklingko Public Transportation on Social Media X (Twitter)

Agustria Purnamasari<sup>1\*</sup>, Arief Agoestanto<sup>2</sup>

<sup>1,2</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, University Negeri Semarang  
Sekaran, Gunungpati, Semarang City, 50229, Indonesia

Corresponding author's e-mail: \* [agustriapurnamasari@students.unnes.ac.id](mailto:agustriapurnamasari@students.unnes.ac.id)

## ABSTRACT

**Keywords:**  
**Sentiment Analysis;**  
**Naïve Bayes;**  
**Decision Tree**

Technological developments have changed the way people access information and share opinions through the internet and social media such as X (Twitter). Public sentiment is now crucial in evaluating public services, especially in the era of advanced information technology. As a metropolitan city, public transportation in DKI Jakarta plays an important role in economic, business and government activities. The Jaklingko initiative launched by the DKI Jakarta Provincial Government aims to provide efficient and affordable public transportation. This research implements and compares Naive Bayes and Decision Tree classification methods to perform sentiment analysis of twitter users opinions regarding Jaklingko into positive and negative categories. Data was collected by crawling using `tweet_harvest` which obtained 6001 tweets about Jaklingko, then text preprocessing was carried out. Word weighting is done using the TF-IDF method to give value to each term in the document, and sentiment labeling is done using the Vader Lexicon library. The data is divided into training and testing data with a ratio of 80%:20% for the classification process. Evaluation of the method is done using confusion matrix. The results showed that the accuracy of Naive Bayes reached 84.9% and Decision Tree reached 84.2%. The wordcloud visualization depicts negative words including vehicle stoppage, bad driver attitude, and envy from people in other cities. Meanwhile, positive words included free system, useful programs, and user convenience. This research provides an in-depth understanding of public opinion towards Jaklingko, with potential implications for improving public transportation services in Jakarta.

## 1. INTRODUCTION

Advances in technology and information that are increasingly sophisticated certainly play an important role in human life both individuals and groups. Currently, the development of digital technology has enabled computing technology and storage media that allow humans to collect and prepare data from various available sources. The increasing amount of data collected causes an increase in the volume of data, but most users rarely utilize it and even the data can be deleted because it is considered useless and only fills the storage space. In fact, these data have the potential to provide valuable information in the future. From this problem, data mining is needed to study data to produce useful information. Data mining is the process of extracting and identifying useful information and knowledge that can be utilized for the future from very large data using mathematics, statistics, artificial intelligence (AI) and machine learning. In addition, data mining is also an iterative and interactive method that can be generalized to find new data patterns or models from massive data so that it will be more useful in the future [1].

Technological developments have changed the way information is presented and accessed by humans. One aspect that supports these developments in providing information is the internet. In using the internet, the opportunity to explore many things has expanded to include the latest information, data and statistics, knowledge, interaction on social media. Social media is considered a new public arena, with the belief that it facilitates individuals to express their aspirations more freely. Twitter is one of the most popular and in demand social media today. Twitter or now called X is considered a social media platform that is simple to use and very efficient in disseminating information quickly. In the current era, social media has become the main tool for people to interact, give opinions, and respond to various problems that are happening. Along with the rapid use of twitter social media, there are many topics discussed by its users. One of the topics that has recently been busy is the attributes ahead of the 2024 presidential and vice-presidential elections. With this phenomenon, the topic of transportation “Jaklingko” reappeared as a debate among twitter users. Jaklingko is an evolution of the OK-Otrip program which establishes an integration system as public transportation in Jakarta. This program is an effort to provide one-trip transportation at one price introduced by the Provincial Government of DKI Jakarta under the leadership of Anies Bawedan and Sandiaga Uno. On social media twitter, an account re-uploaded a conference video regarding the formation and determination of Jaklingko. The video was uploaded to twitter social media on January 14, 2023 and has been viewed by more than 5 million twitter users. So that the topic of Jaklingko has become a topic that is debated again by the community regarding the implementation of its benefits. Through issues circulating on twitter social media, people express their opinions about Jaklingko through tweets. The tweets uploaded can contain positive or negative responses. The process of analyzing these opinions is called sentiment analysis.

Sentiment analysis is a data processing technique used to analyze, process, and extract information from text data containing opinions about entities such as products, services, organizations, individuals, or specific topics. Ratings in sentiment analysis are usually expressed as positive or negative [2].

Previously, research was conducted by [2] which discussed sentiment analysis regarding Jaklingko public transportation using the SVM method which obtained an accuracy result of 71%. The author also said that when there are many developments in text preprocessing methods, it is hoped that their better, detailed and thorough application can increase the accuracy value in sentiment data classification.

Various studies on sentiment analysis have been carried out. Research conducted by [3] which examines the sentiment of Twitter users on the covid-19 vaccine obtained an accuracy of 81% for the Naïve Bayes algorithm and 79% for the Decision Tree algorithm. Research conducted by ([4] which examines the sentiment of user reviews of the care protect application obtained an accuracy of 83.7% for the Naïve Bayes algorithm and 82.8% for the Decision Tree algorithm, this shows the performance of the Naïve Bayes algorithm is superior. In research [5] which compares three methods, namely: Naïve Bayes, SVM, and LSTM obtained accuracy results that for the Naïve Bayes method has an accuracy of 84.65%, the SVM method has an accuracy of 84.89%, and for the LSTM method has an accuracy rate of 82.97%, this shows the Naïve Bayes and SVM algorithms both have higher performance in their calculations. Research was also conducted by [6] who examined the sentiment of Instagram user comments regarding the independent campus program using the Naïve Bayes Classifier and Decision Tree methods obtained for each of their respective accuracies, namely 83% for Naïve Bayes and 81% for Decision Tree.

Based on the above background, this research is titled “Modeling Naïve Bayes Algorithm and Decision Tree to Analyze Sentiments related to Jaklingko Public Transportation on Social Media X (Twitter)”. This research aims to compare the accuracy between the naïve bayes and decision tree methods.

## 2. RESEARCH METHODS

In general, the stages in this research are data collection, data analysis, and conclusions. The data used in this research is data obtained from the twitter crawling process that discusses jaklingko public transportation. This research data was taken using the keyword on the variable, namely “jaklingko”. The data used in this study were taken from January 10 to January 31, 2024 regarding tweets containing opinions about jaklingko. The twitter data that will be used consists of username, created at, and text. In retrieving and processing data in this study, it is assisted by using Google Collaboratory with the Python programming language.

### 2.1 Text Mining

Text mining is the process by which computers perform mining to find something new, previously unknown, or retrieve implied information indirectly, which is extracted automatically from various text data sources [7]. In general, text mining has two stages, which are as follows.

#### 1. Preprocessing Text

Text preprocessing is an initial process that aims to extract important information from text and use it to generate new datasets. Text preprocessing aims to transform unstructured text data into more structured ones. In general, the steps in text preprocessing are as follows:

- a. Case Folding  
Case folding is the process of uniformizing characters in a document. In this case every text document consistently uses small. Alphabetic characters from “A” to “Z” will be converted into “a” to “z”. So that case folding is needed to convert the entire text in the document into a standard form.
- b. Cleansing  
Cleansing is the process of cleaning text from back slice marks, RT, mentions, links, URLs, hashtags, numbers, punctuation marks, and single characters. This process aims to make the dataset clean from meaningless characters that affect sentiment.
- c. Tokenizing  
Tokenizing is the stage of separating sentences into tokens or pieces of words to make the sentence more meaningful and facilitate the next process.
- d. Spelling Normalization  
The process of spelling correction or expansion of abbreviated words into the correct format is called normalization. Word correction aims to reduce the number of dimensions of expanded words.
- e. Filtering  
At this stage, we take the key words from the generated tokens. Then there is a wordlist algorithm (saving important words) or stoplist (eliminating unimportant words) which will be used in filtering.
- f. Stemming  
The stemming process is the process of removing prefixes and suffixes (affixes) in each word to get the basic word of a document. This stemming stage is carried out with the aim of narrowing the number of words and obtaining basic words that have the same meaning as uniform writing.

## 2. TF-IDF Weighting

Term Weighting. TF-IDF is a frequently used term weighting method and serves as a tool to compare new weighting methods with existing ones. (TF) Term Frequency - (IDF) Inverse Document Frequency (TF-IDF) weighting is a calculation process to find how closely a term is attached to a document by determining the weight of each word. This method combines the frequency of words that appear in documents and the inverse of the frequency of documents that include that word.

According to [8] in getting the TF-IDF value, the  $IDF_w$  value is needed which can be obtained using the following equation.

$$IDF_w = \log\left(\frac{N}{df_w}\right)$$

$N$  : number of documents tested

$df_w$  : number of words in the document

After doing equation, the TF-IDF value will be obtained using the following equation.

TF-IDF : weight of a word in a document

$tf$  : the frequency of occurrence of the word in the document

IDF is the number of documents that contain the term. The log of IDF calculation is here to provide some adjustment for finer words. In this approach, each document is viewed as a vector consisting of 1 component corresponding to each term found in the dictionary, along with its respective weight value. Conversely, the weight for terms not found in the document is 0.

## 2.2 Naïve Bayes

According to [9] Naive Bayes is a classification technique using probability and statistical principles introduced by British scientist Thomas Bayes. Naive Bayes is a classification method based on the Bayes Theorem. This method uses previous data to make future predictions. One of the main characteristics of the naive bayes method is the strong assumption of the independence of each condition [10] The Bayes Theorem is formulated as follows:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Where:

$A$  : sample data with unknown class (label)

$B$  : hypothesis A data is data with a specific class (label)

$P(A|B)$  : probability of event A if B happens

$P(A)$  : prior probability of hypothesis A

$P(B|A)$  : probability that B determined by A is true

$P(B)$  : prior probability of evidence B

## 2.3 Decision Tree

A decision tree is a diagram that starts with one node and then each node has a branch for each option, with each branch generating a new branch. The decision tree model is a tree consisting of root nodes, internal nodes and terminal nodes. While root nodes and internal nodes are variables/features, terminal nodes are class labels. Decision Tree based on C4.5 algorithm is a commonly used classification technique to identify important associations in data. The C4.5 algorithm is a method that is used to perform predictive classification, where in the classification of a decision tree consists of a node that forms the root, the root node has no input. Other nodes that are not roots but have exactly one input are called test nodes, and other nodes are called leaves. The leaves here represent the most appropriate target value of one of the classes [11]. The steps in building a decision tree using the C4.5 algorithm are as follows.

1. The highest gain value of all the tributes is selected as the root

The following equation is used to calculate the highest gain value.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{S} \times \text{Entropy}(S_i)$$

Description:

$S$  : a set of cases

$A$  : attribute

$n$  : number of partitions  $S$

$|S_i|$  : number of cases found in the  $i$ -th partition

$|S|$  : number of cases found in  $S$

The entropy value can be found using the following equation.

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Description:

$n$  : number of partitions  $S$

$p_i$  : proportion of  $S_i$  to  $S$

2. Build branches for each value.
3. Split the case on the branch.
4. Repeating the steps until the same class is obtained from both branches and cases.

## 2.4 Confusion Matrix

According to [12], matrix classification is the most effective way to classify data. Visually, it is a two-way matrix with one axis showing the actual class distribution and the other axis showing the predicted class. Confusion matrix is usually used to help evaluate a model in classification. Evaluation using confusion matrix is done to determine the accuracy of a model.

**Table 1. Confusion Matrix**

Two Class Classification	Prediction result class		Total
	1	0	
Actual Class 1	True Positive (TP)	False Negative (FN)	$P$
Actual Class 0	False Positive (FP)	True Negative (TN)	$N$
Total	$P'$	$N'$	$P + N$

Description:

True Positive (TP) : correct prediction for the number of positive data

False Positive (FP) : wrong prediction for the number of positive data

False Negative (FN) : correct prediction for the amount of negative data

True Negative (TN) : wrong prediction for the amount of negative data

To find the accuracy level of a model, the equation

$$\text{accuracy} = \frac{TP + TN}{P + N}$$

Furthermore, the precision and recall processes are also needed in classification evaluation. Precision is the level of accuracy of the model's predictive value in predicting positive and correct classes. To calculate the precision value, the equation is used.

$$\text{precision} = \frac{TP}{TP + FP}$$

Meanwhile, recall is a measure of the completeness of all positive data. The recall value can be calculated with the equation.

$$\text{recall} = \frac{TP}{TP + FN} = \frac{TP}{P}$$

### 3. RESULT AND DISCUSSION

#### 3.1 Descriptive Analysis

Based on the results of crawling twitter with the keyword 'Jaklingko', a total of 6001 tweets were obtained with a vulnerable time from January 10 to January 31, 2024. Furthermore, the data is made into a graph so that information is easier to obtain. Figure 1 shows a graph of tweet data based on the time it was created.

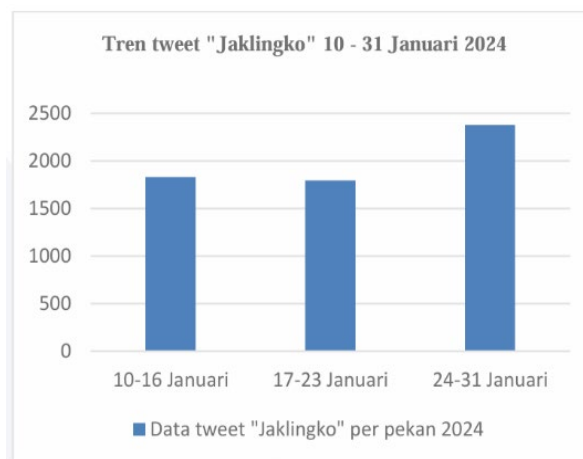


Figure 1. "Jaklingko" Tweet Data as of Week 2024

Based on this data, it is found that the average tweet about Jaklingko in a day approximately reaches 250 tweets so that it can be said that the topic of jaklingko is still widely discussed on twitter social media. jaklingko is still widely discussed on twitter social media. Tweets the most tweets on January 15 and January 27, 2024 with 615 tweets and 612 tweets respectively as many as 615 tweets and 612 tweets.

#### 3.2 Text Mining

##### 3.2.1 Preprocessing Text

To perform text mining, there are stages that must be carried out to process data sources, both structured and unstructured data. Text preprocessing in text mining is an initial process that aims to retrieve important information from text and use that information to generate new datasets. Various information or data is then processed to get an organized data. The preprocessing stage aims to harmonize words, remove characters such as numbers and punctuation marks, and discard words that do not need to be used so that the opinion data becomes more structured. The following are the stages of text preprocessing in this research.

**Table 2. Text Preprocessing Stages**

Preprocessing Text	Before	After
<i>Case folding</i>	@ui_fess Sebagai orang non jkt yg merantau ke jkt, gue sangat terbantu dengan berbagai program yang beliau lakuin selama jadi gubernur jkt, salah satunya jaklingko. Gue juga suka sama beliau yg memperjuangkan hak masyarakat buat dapet ruang terbuka yang nyaman dan gratis.	@ui_fess sebagai orang non jkt yg merantau ke jkt, gue sangat terbantu dengan berbagai program yang beliau lakuin selama jadi gubernur jkt, salah satunya jaklingko. gue jugasuka sama beliau yg memperjuangkan hak masyarakat buat dapet ruang terbuka yang nyaman dan gratis.
<i>Cleansing</i>	@ui_fess sebagai orang non jkt yg merantau ke jkt, gue sangat terbantu dengan berbagai program yang beliau lakuin selama jadi gubernur jkt, salah satunya jaklingko. gue juga suka sama beliau yg memperjuangkan hak masyarakat buat dapet ruang terbuka yang nyaman dan gratis.	sebagai orang non jkt yg merantau ke jkt gue sangat terbantu dengan berbagai program yang beliau lakuin selama jadi gubernur jkt salah satunya jaklingko gue juga suka sama beliau yg memperjuangkan hak masyarakat buat dapet ruang terbuka yang nyaman dan gratis
<i>Tokenizing</i>	sebagai orang non jkt yg merantau ke jkt gue sangat terbantu dengan berbagai program yang beliau lakuin selama jadi gubernur jkt salah satunya jaklingko gue juga suka sama beliau yg memperjuangkan hak masyarakat buat dapet ruang terbuka yang nyaman dan gratis	"sebagai", "orang", "non", "jkt", "yg", "merantau", "ke", "jkt", "gue", "sangat", "terbantu", "dengan", "berbagai", "program", "yang", "beliau", "lakuin", "selama", "jadi", "gubernur", "jkt", "salah", "satunya", "jaklingko", "gue", "juga", "suka", "sama", "beliau", "yg", "memperjuangkan", "hak", "masyarakat", "buat", "dapet", "ruang", "terbuka", "yang", "nyaman", "dan", "gratis"
<i>Spelling Normalization</i>	"sebagai", "orang", "non", "jkt", "yg", "merantau", "ke", "jkt", "gue", "sangat", "terbantu", "dengan", "berbagai", "program", "yang", "beliau", "lakuin", "selama", "jadi", "gubernur", "jkt", "salah", "satunya", "jaklingko", "gue", "juga", "suka", "sama", "beliau", "yg", "memperjuangkan", "hak", "masyarakat", "buat", "dapet", "ruang", "terbuka", "yang", "nyaman", "dan", "gratis"	"sebagai", "orang", "non", "jakarta", "yang", "merantau", "ke", "jakarta", "saya", "sangat", "terbantu", "dengan", "berbagai", "program", "yang", "beliau", "lakukan", "selama", "jadi", "gubernur", "jakarta", "salah", "satunya", "jaklingko", "saya", "juga", "suka", "sama", "beliau", "yang", "memperjuangkan", "hak", "masyarakat", "buat", "dapet", "ruang", "nyaman", "dan", "gratis"
<i>Filtering Stopword Removal</i>	"sebagai", "orang", "non", "jakarta", "yang", "merantau", "ke", "jakarta", "saya", "sangat", "terbantu", "dengan", "berbagai", "program", "yang", "beliau", "lakukan", "selama", "jadi", "gubernur", "jakarta", "salah", "satunya", "jaklingko", "saya", "juga", "suka", "sama", "beliau", "yang", "memperjuangkan", "hak", "masyarakat", "buat", "dapet", "ruang", "terbuka", "yang", "nyaman", "dan", "gratis"	"sebagai", "orang", "non", "jakarta", "yang", "merantau", "ke", "jakarta", "saya", "sangat", "terbantu", "dengan", "berbagai", "program", "yang", "beliau", "lakukan", "selama", "jadi", "gubernur", "jakarta", "salah", "satunya", "jaklingko", "saya", "juga", "suka", "sama", "beliau", "yang", "memperjuangkan", "hak", "masyarakat", "buat", "dapet", "ruang", "nyaman", "dan", "gratis"
<i>Stemming</i>	"orang", "non", "jakarta", "merantau", "jakarta", "terbantu", "program", "beliau", "lakukan", "gubernur", "jakarta", "salah", "satunya", "jaklingko", "suka", "beliau", "memperjuangkan", "hak", "masyarakat", "ruang", "terbuka", "nyaman", "gratis"	"rantau", "bantu", "suka", "juang", "hak", "buka", "nyaman", "gratis"

In the text preprocessing that has been done with the python program, the net tweet results are 4017 tweets from a total of 6001 tweets which are then ready to be translated as a basis for continuing the next process.

In this research, it is necessary to translate documents to continue the next process. The translation process is done automatically using google translate in the python program by utilizing the google translate

library. The research is translated from Indonesian to English because researchers will use the vader sentiment library which is designed in English with the aim of increasing the accuracy of the sentiment data labeling process [4].

**Table 3. Translate Result**

Before	After
“rantau”, “bantu”, “suka”, “juang”, “hak”, “buka”, “nyaman”, “gratis”	overseas help likes fight right open comfortably free

### 3.2.2 TF-IDF Weighting

Opinion data that has gone through the preprocessing stage which is still in text form will be converted into vector form with the Term Frequency Inverse Document Frequency technique. Word weighting using Jaklingko tweet data against queries, aims to change the text data of each opinion has a value.

**Table 4. Frequency of occurrence and weight of top 15 terms by Python**

No	Word	Frequency	Term	Weight
1	Free	668	Free	0.095022
2	Good	303	Good	0.040574
3	Stop	188	Angry	0.028899
4	Angry	184	Help	0.025755
5	Easy	171	Stop	0.024760
6	Help	151	Easy	0.024023
7	Like	140	Stupid	0.021161
8	Stupid	136	Thank	0.021081
9	Thank	131	Like	0.019348
10	Comfortable	121	Comfortable	0.018117
11	Helpful	109	Helpful	0.017902
12	Bad	109	Love	0.016995
13	Love	104	Sad	0.016354
14	Happy	98	Bad	0.016183
15	Sad	94	Happy	0.015448

## 3.3 Sentiment Class Labeling

Class labeling here utilizes the lexicon-based method library, where this method uses a dictionary as the basis of language or lexical automatically by the lexicon dictionary by calculating the sentiment score. How to determine the sentiment class is by calculating the score of the number of positive words minus the score of the number of negative words. Sentences that have a score  $\geq 0$  will be classified into the positive sentiment class, then sentences that have  $< 0$  are classified into the negative sentiment class. In this study, the results of the number of sentiments obtained from the labeling process using Vader Lexicon Sentiment through python programming are for positive sentiment obtained 2648 tweets and negative sentiment obtained 1369 tweets. Then divide the data into training and testing data by 80%:20%.

## 3.4 Classification

### 3.4.1 Naïve Bayes Classification

The first method used to classify opinion data through twitter social media user tweets regarding Jaklingko public transportation is naïve bayes. In this study, the entire process of the naïve bayes algorithm classification stage was carried out using python programming language. The package used to find the accuracy level of the naïve bayes algorithm is the naïve\_bayes package. After the programming coding is



run, it will produce an accuracy output from the model. Next is to evaluate the model to measure the accuracy of the model accuracy using confusion matrix, based on Table 1, the number of positive and negative labels has been obtained so that the process here will calculate the level of accuracy, precision, and recall. Based on the results of the confusion matrix calculation, it can be seen that the accuracy of the naïve bayes model reaches 84.9%. The precision value obtained to measure the accuracy of the model in the classification process reaches 83.9%. While the recall value obtained to measure the accuracy of all positive data reaches 96.5%.

### 3.4.2 Decision Tree

The second method used to classify opinion data through twitter social media user tweets regarding jaklingko public transportation is decision tree. In this study, the entire process of the classification stages of the decision tree algorithm was carried out using python programming language. The package used to find the accuracy level of the decision tree algorithm is package tree. After the programming coding is run, it will produce an accuracy output from the model. Next is to evaluate the model to measure the accuracy of the model accuracy using confusion matrix. Based on Table 1, the number of positive and negative labels is quite balanced so that in this process only calculates the level of accuracy, precision, and recall. Based on the results of the above calculations, it can be seen that the accuracy of the decision tree model reaches 84.2%. the precision value obtained to measure the accuracy of the model in the classification process reaches 91%. while the recall value obtained to measure the accuracy of the entire data reaches 85%.

### 3.4.3 Comparison of Naïve Bayes and Decision Tree Classification

This comparison is done by looking at the highest accuracy results of each classification model. The purpose of this comparison is to find which model is better in text mining classification. Comparison of the accuracy results of the Naïve Bayes and Decision Tree models can be seen in the following table.

**Table 5. Comparison of Naïve Bayes and Decision Tree Classification**

Model	Accuracy	Precision	Recall
<i>Naïve Bayes</i>	84,9%	83,9%	96%
<i>Decision Tree</i>	84,2%	91%	85%

The naïve bayes method performs classification by producing an accuracy rate of 84.9% with precision and recall of 83.9% and 96% respectively. While the decision tree method performs classification by producing an accuracy rate of 84.2% with precision and recall of 91% and 85% respectively. The naïve bayes model has a higher recall rate (96%) than the decision tree (85%), meaning that naïve bayes tends to be better at finding most of the true positive samples in the dataset. Meanwhile, the decision tree model has a higher precision rate (91%) compared to naïve bayes (83.9%), which shows that decision trees tend to be better at avoiding prediction errors, especially on false positives. The test results on tweet opinions about jaklingko on twitter social media show that this research has superior accuracy than research [3] which obtained accuracy for naïve bayes of 83% and decision tree of 79%. This is also in line with research conducted by [4] which obtained accuracy results for naïve bayes of 83.7% and decision tree of 82.8% for sentiment analysis research, where this research is superior to the naïve bayes method.

## 3.5 Wordcloud Visualization

Text data visualization shows the final results of sentiment analysis in the form of barplots and word clouds containing positive and negative words. The purpose of this visualization is to present information in the form of words that often appear and are expressed by twitter users in expressing opinions. In this research, wordcloud visualization will be divided into 2, namely negative and positive visualization.



transportation which shows that users feel helped in saving transportation expenses by the free jaklingko system, saying that jaklingko is a good and convenient transportation system so that users like the system. The use of naive bayes and decision tree methods in classifying positive and negative sentiments on twitter social media user opinions regarding Jaklingko public transportation resulted in an accuracy rate for the naive bayes method of 84.9% and the decision tree method of 84.2%.

## REFERENCES

- [1] F. A. Hermawati, *Data Mining*. Yogyakarta: CV. Andi, 2013.
- [2] F. Ladayya, D. Siregar, W. E. Pranoto, and H. D. Muchtar, "Analisis Sentimen pada Program Transportasi Publik JakLingko dengan Metode Support Vector Machine," *Jurnal Statistika dan Aplikasinya*, vol. 6, no. 2, pp. 381–392, 2022.
- [3] A. Zamsuri, E. Asril, M. Sadar, and Y. Turnandes, "Analisis Sentimen Opini Terhadap Vaksin Covid-19 Pada Media Sosial Twitter Menggunakan Naïve Bayes Dan Decision Tree," *ZONasi: Jurnal Sistem Informasi*, vol. 5, no. 1, 2023.
- [4] C. Ramdani, A. Rachman, and R. Setiawan, "Comparison of the Multinomial Naive Bayes Algorithm and Decision Tree with the Application of AdaBoost in Sentiment Analysis Reviews PeduliLindungi Application," *International Journal of Information System & Technology Akreditasi*, vol. 6, no. 158, pp. 419–430, 2022.
- [5] D. A. Kristiyanti and Sri Hardani, "Sentiment Analysis of Public Acceptance of Covid-19 Vaccines Types in Indonesia using Naïve Bayes, Support Vector Machine, and Long Short Term Memory (LSTM)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 722–732, Jun. 2023, doi: 10.29207/resti.v7i3.4737.
- [6] B. Wicaksono and N. Cahyono, "Analisis Sentimen Komentar Instagram Pada Program Kampus Merdeka Dengan Algoritma Naive Bayes Dan Decision Tree," 2024. [Online]. Available: <https://situs.com>.
- [7] R. Feldman and J. Sanger, *The text mining handbook: Advanced approaches in analyzing unstructured data*. 2007.
- [8] S. Rizkia, E. B. Setiawan, and D. Puspandari, "Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF," *eProceedings of Engineering*, vol. 6, no. 2, 2019.
- [9] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi" 2014.
- [10] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 1, no. 1, pp. 42–51, 2021.
- [11] F. Fatmawati and N. Narti, "Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 1, pp. 1–12, May 2022, doi: 10.35746/jtim.v4i1.196.
- [12] K. Ramasubramanian and A. Singh, *Machine Learning Using R: With Time Series and Industry-Based Use Cases in R*, Second Edition. Springer Science+Business Media, 2019. doi: 10.1007/978-1-4842-4215-5.

