



# Sentiment Analysis of Google Maps User Reviews on the Play Store Using Support Vector Machine and Latent Dirichlet Allocation Topic Modeling

Violita Aditya Zahrah<sup>1\*</sup>, Nurdin<sup>2</sup>, Risawandi<sup>1</sup>

<sup>1</sup>Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

<sup>2</sup>Department of Information Technology, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

\*Corresponding author E-mail: [violita.200170130@mhs.unimal.ac.id](mailto:violita.200170130@mhs.unimal.ac.id)

The manuscript was received on 17 March 2024, revised on 28 April 2024, and accepted on 28 September 2024, date of publication 16 October 2024

## Abstract

These days, traveling is made easier by utilizing easily accessible online directions such as Google Maps. Google Maps provides real-time routes by displaying and presenting the closest routes that users can take. However, lately, the routes provided by Google Maps services often get users lost by presenting routes such as forests, narrow roads, and even dead ends. Therefore, this study aims to determine the level of user satisfaction and sentiment into two categories, namely positive and negative, based on reviews on the Google Play Store platform using the Support Vector Machine (SVM) algorithm and topic modeling using Latent Dirichlet Allocation (LDA) to find out the collection of topics that are the main topics of conversation by users regarding Google Maps services. The results of this study show that the SVM algorithm is feasible to use in sentiment analysis classification with an accuracy value of 86%, precision of 93%, recall of 53%, and f1-score of 52%. In addition, topic modeling is applied to generate coherence values for each topic, which shows that the higher the coherence value, the more specific the topic is. The highest coherence value generated in this study was two topic models with a coherence value of 35.15%, but this study took five with a coherence value of 33.39%. The five topic models to be applied in this study are selected because they have a good enough coherence value to identify the main topics and hidden topics in Google Maps user reviews with the Latent Dirichlet Allocation model. The topic model shows five aspects users often discuss: Google Maps route accuracy, system and service errors, navigation application directions, lost time history, and convoluted route provision.

**Keywords:** Sentiment, Topic Modeling, Support Vector Machine, Latent Dirichlet Allocation, Google Maps.

## 1. Introduction

Currently, developments that have entered the technological era, like today, make it easier for someone to travel between cities or regions. This is evidenced by the ease with which users can access online maps as travel directions. Travel directions can be accessed online using an internet quota. Google Maps is one of the directions trusted and recognized by the general public [1]. The routes presented by Google Maps also consider road density and traffic conditions in real-time so that users can find out how long the trip will take. Not only does it provide the best route, but Google Maps users can also use this application to give criticism and suggestions. They can load images from a location visited. This allows other users to assess other visitors to a place regarding service, presentation, and appearance [2]. As an online navigator, users often consider Google Maps a flawed system due to its poor functional performance, which usually causes errors in route instructions. Therefore, this user review will be used as a sentiment or emotion analysis research sample to measure user satisfaction with the Google Maps application. Sentiment analysis



Techniques are often applied in several business sectors to determine the opinions and levels of public satisfaction with a product or service the company provides; this is useful for continuing to develop and assist companies in understanding customer needs [3].

This study applies the Support Vector Machine classification algorithm to classify sentiment. Support Vector Machine is one of the classification methods that is done by maximizing the hyperplane boundary. The hyperplane concept uses the maximum margin to produce better generalization at the classification stage.[4] In addition to classification, this study also conducted topic modeling using Latent Dirichlet Allocation (LDA), one of the topic modeling methods used to generate a list of topics by assigning a weight value to each document or data. Applying LDA to this study will identify the discussion topics often appearing in Google Maps user reviews. [5] The use of the SVM method has been compared several times with other classification methods, and this was also done by [6] with the research title "Comparison of Support Vector Machine and Naïve Bayes Classification Methods for Sentiment Analysis on Textual Reviews on the Google Play Store." The comparison carried out using reviews on the Google Play Store resulted in an SVM accuracy level reaching 81.46% and Naïve Bayes with a percentage of 75.41%, so it can be concluded that SVM has proven to be one of the best classification methods.

Other research was also conducted by [7] with the topic "Sentiment Analysis for Volcano Visitor Reviews using the Support Vector Machine Method" using 1745 data, resulting in an accuracy of 92.94%. This proves that using SVM in classification produces quite good values. Based on the many user comments and the increasing number of users complaining about route errors obtained from Google Maps online navigator directions, to determine the level of user satisfaction in using the navigator application with the most downloads, this study was conducted to assess the performance produced by the application so that users can find out the performance of Google Maps which continues to make updates and to be a means of assessment for companies in making improvements and updates to the system Sentiment analysis of the Google Maps application is essential to understand the user experience of the application service to improve the quality of reviews and feature development which is also done to improve the company's reputation and services. This is an advantage that can be obtained based on the results of the analysis of positive and negative user sentiment towards the application. With this research on the sentiment of Google Maps application users, it is hoped that this will be able to improve the Google Maps service into a better navigation application.

## 2. Literature Review

### 2.1. Previous Research

research conducted by Idris I, Mustafa Y, and Salihi I. (2023) entitled "Sentiment Analysis of the Use of the Shopee Application Using the Support Vector Machine (SVM) Algorithm" this study analyzes an online application object, namely Shopee. The Shopee review data used in this study were 3000, resulting in a Support Vector Machine classification accuracy rate of 98% [8].

Research conducted by Hilma R, Ula M, and Fachrurrazi S. (2023) entitled "Cyberbullying Sentiment Analysis on Twitter Social Media Using the Support Vector Machine and Naïve Bayes Methods" this study analyzed cyberbullying circulating on Twitter social media using 100 tweet data. This test uses two classification methods, namely Support Vector Machine (SVM) and Naïve Bayes, with an accuracy result of SVM of 72% and an accuracy of Naïve Bayes of 69%. This study proves that the level of classification accuracy using SVM is better than Naïve Bayes [9].

Research conducted by Chen M, Guo Z, Abbas K, and Huang W. (2022) entitled "Analysis of the Impact of Investor Sentiment on Stock Price Using the Latent Dirichlet Allocation Topic Model" this study applies a topic model to find out the primary information on the investor business market. The LDA used in this study produces information that the excess return rate is positively related to investor sentiment categorized by theme, the influence of various topics varies, company development has a long-term influence, the condition of the market index has a direct influence, and company dividends have a direct and long-term influence. The impact of investor sentiment on excess returns is asymmetric [10].

Research conducted by Damanik FJ and Setyohadi DB. (2021) entitled "Analysis of public sentiment about COVID-19 in Indonesia on Twitter Using Multinomial Naïve Bayes and Support Vector Machine", this study takes the topic of COVID-19, which has become a pandemic in various countries. The data used in this study are tweets sent by the Indonesian public regarding the COVID-19 virus via Twitter social media. The classification results obtained in this study on the opinions of the Indonesian people regarding COVID-19 tend to be more positive and neutral, with an average percentage of positive and neutral reviews of 40% and negative views of 20%. With the calculation accuracy using the SVM method, the results were 93% in the sentiment analysis classification and Multinomial Naïve Bayes 92% [11].

Research by Riyanto and Aziz A. (2021) entitled "Application of the Vector Machine Support Method in Twitter Social Media Sentiment Analysis regarding the Covid-19 Vaccine Issue in Indonesia" also uses overcoming COVID-19 in Indonesia as an object. In response to this outbreak, the Indonesian government has made many efforts, including vaccinating the community. Still, many rumors and perceptions regarding the advantages and disadvantages of vaccination are circulating on social media. Therefore, this study aims to determine public sentiment toward COVID-19 vaccination through opinions that can be seen on Twitter social media. This study uses the SVM method to classify public opinion sentiment with an accuracy value of 84.4% [12].

Suswandi and Erkamim (2023) conducted research entitled "Sentiment Analysis of Shopee app reviews using Random Forest and Support Vector Machine," using Shopee as an object by comparing classification methods. This study's classification of sentiment analysis uses the SVM and Random Forest methods to classify positive and negative reviews. The results of this study indicate that the accuracy level using Random Forest is 82.21% and SVM is 84.71% [13].

Research conducted by Muttaqin MN and Kharisudin I. (2021) entitled "Sentiment analysis of the Gojek application using Support Vector Machine and K Nearest Neighbor" using the Gojek application as an object. This study compares the KNN and SVM methods in classifying reviews with the calculation of the results obtained; the accuracy value using KNN is 82.14%, and the accuracy value using SVM is 87.98%. So, in this study, it can be concluded that the SVM method classifies reviews better in the case study of sentiment analysis on Gojek application reviews [14].

### 2.2. Data Mining

The definition of data mining is the process of extracting data from an initially unstructured database into more structured patterns or information, where this process can increase efficiency in data usage [15]. Data mining consists of several techniques, one of which is classification. This process aims to observe and find models based on the similarity of object characteristics by classifying data into several classes [16]. Data Mining dinilai sebuah pembelajaran yang sangat penting dalam mempermudah pengolahan data dengan aspek yang terjangkau dan efisien [17].

### 2.3. Text Mining

Text mining is a computerized process that automatically extracts information from different sources. The main goal of this process is to combine a lot of information collected from various sources [18]. Text mining is also a process of converting words in textual data into data that computer systems can understand and recognize by converting words into characters [19].

### 2.4. Sentiment Analysis

Sentiment analysis is a stage in identifying problems that arise to obtain an overall understanding of the system to be developed [20]. This analysis classifies or classifies the polarity of the text in a data, document, opinion, or sentence, which determines whether the data has positive or negative aspects. In addition, by analyzing sentiment, the results of the information in a text or sentence can also be seen through the problems or objects discussed that tend to be positive or negative opinions [21].

### 2.5. Google Maps

Google Maps is a web application developed by Google as an online map service. Users visit websites or application services for specific purposes or needs [22]. This application allows users to search for Locations, find directions, locate businesses or places to explore, and view satellite images of locations worldwide. Google Maps also makes traveling more accessible by providing real-time location information.

### 2.6. Text Preprocessing

The text preprocessing stage is carried out to improve the quality of the data used, which includes cleaning the data and restoring the words in the data to standard words according to the KBBI. Text preprocessing aims to reduce the dimensions of the data to be more efficient and precise, making it easier for computers to process data [23].

### 2.7. TF-IDF

Term Frequency - Inverse Document Frequency (TF-IDF) is used to balance words and to assess the relationship between words in a document. This method combines two ideas to calculate the weight of the word frequency in a particular document and to know the inverse frequency of documents containing the word. This TF-IDF weight calculation can be used to determine how relevant the words contained in the document are. Therefore, the weight of each document on the keyword can be calculated using the TF- IDF formula [3].

$$tf = 0,5 + 0,5 \times \frac{ft,d}{\max(ft,d)} \quad (1)$$

$$idf = \log \frac{N}{df} \quad (2)$$

$$W = tf \times idf \quad (3)$$

Description:

d: document

idf: words in the document

tf: number of words in the searched document

### 2.8. Support Vector Machine

Support Vector Machine (SVM) is an algorithm used to analyze data and introduce patterns for classification. Classification using the SVM concept will eventually find the best hyperplane to use as a class separator. SVM can also be used on datasets with a high dimensionality or complexity by using the trick part. The SVM concept will only be used in the classification process with the help of several selected data points [4].

In general, several kernel equations can be used in SVM calculations: linear kernels, polynomial kernels, radial essential function (RBF) kernels, and sigmoid kernels. In this study, the data used is text data and is classified as a real-world problem. Therefore, the classification process cannot be assumed to be linear, so the kernel used in this study is the RBF kernel because it is considered capable of classifying high-dimensional data.

$$f(x) = w \cdot x + b \tag{4}$$

atau

$$f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b$$

Description:

w: Hyperplane parameters sought

x: input data points support vector

b: bias value

a<sub>i</sub>: weight value of each i-th data point

y<sub>i</sub>: i-th data class

x<sub>i</sub>: i-th data

K(x,x<sub>i</sub>) : kernel function description:

w: Hyperplane parameters sought

x: input data points support vector

b: bias value

a<sub>i</sub>: weight value of each data point i

y<sub>i</sub>: data class i

x<sub>i</sub>: data i

K(x,x<sub>i</sub>) : kernel function

### 2.9. Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a topic modeling method because it can summarize, correlate, and process extensive data. LDA can also display a list of topics and compare their occurrence. The LDA model is based on a generative probabilistic model for a collection of writings called a corpus. In the corpus, each document is associated with a mixture of randomly selected hidden topics with characters assigned based on the distribution of words in the document [24]. The LDA model can be represented as follows.

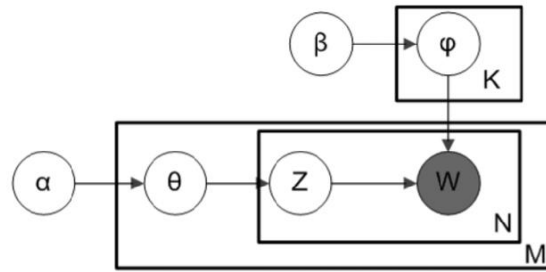
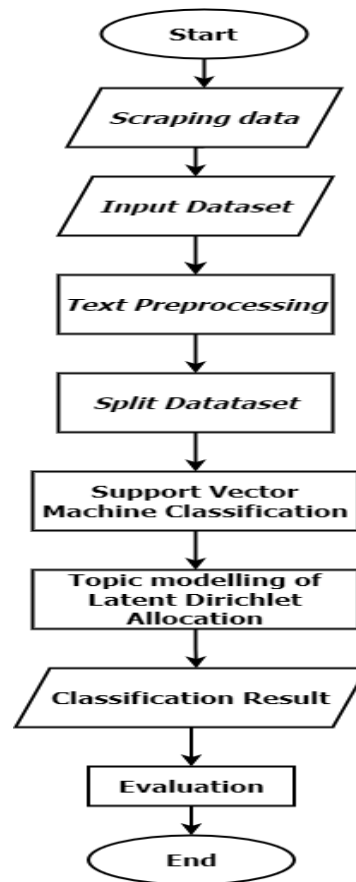


Fig 1. Representation of Latent Dirichlet Allocation Model

In LDA modeling, there are levels. The document set M consists of parameters  $\alpha$  and  $\beta$  determining the topic distribution at the corpus level. The value of  $\alpha$  in a document indicates that the mix of topics is increasing, and the value of  $\beta$  suggests the distribution of words within that topic. If n, the variable  $\theta$  indicates how the topics are distributed in the documents. A higher value of  $\theta$  indicates that more topics are present in the document, but if the value of  $\theta$  is smaller, it is more specific to a particular topic. The variable Z indicates the topic of a particular word in the document, and the variable W indicates the words associated with that topic.

### 3. Methods

The research method is an activity to obtain data and principles that will be used as guidelines in conducting research, and this includes collecting, writing, and studying data systematically. The stages carried out in this study are as follows:



**Fig 2.** Flowchart system

Following the system framework designed above, the system run will go through stages from start to end, and each stage will be explained as follows.

- a. Start  
The start is the initialization process of the system.
- b. Scraping data  
Scraping is the stage of collecting review data on the Google Play Store. This scraping stage uses the `google_play_scraper` library to retrieve review data.
- c. Input dataset  
The stage for importing labeled datasets into the system. After inputting the dataset, the system will display the data if successful.
- d. Text Preprocessing  
This stage is processing text into words that are easier for the system to understand. At this stage, the data input into the system will be simplified by cleaning the data or removing symbols, numbers, affixes, and words with no other meaning.
- e. Split Dataset  
The dataset will be divided into 80% training data and 20% testing data. Training data is data used to train the model on the system, and testing data is used to evaluate the performance results and accuracy of the model on the system.
- f. Sentiment analysis with Support Vector Machine  
User reviews will be classified to identify positive and negative sentiments at this stage using the Support Vector Machine method.
- g. Topic modeling with Latent Dirichlet Allocation  
At this stage, topic modeling will be done to find hidden information and top topics in a document using the Latent Dirichlet Allocation calculation.
- h. Evaluation  
At this stage, it will measure and calculate the accuracy, calculation, or specification of the algorithm used, namely the SVM method, according to the correct standard value.
- i. Results  
At the results stage, reviews classified as positive or negative classes will be displayed, and topic modeling of user reviews will be displayed using Word Cloud.
- j. Finished  
This stage is the last stage of the system.

## 4. Results and Discussion

### 4.1. Research Result

This study tested the Support Vector Machine algorithm implemented in the system to analyze and classify Google Maps user reviews. User reviews are classified into positive and negative sentiments. This sentiment analysis system uses Python as a programming language to build the system. This sentiment analysis produces Google Maps user reviews labeled positive or negative. It delivers an accuracy value of 86%, precision of 93, recall of 53%, and f1-score of 52% with the implementation of the Support Vector Machine algorithm as a classification algorithm in sentiment analysis.

### 4.2. Scraping Data

User review data on the Play Store platform through the scraping stage using the google\_play\_scraper library to collect data in CSV form. The review data taken is Google Maps user review data with the "most relevant" category, which amounts to 500 data points. After the scraping stage, here are 5 Google Maps user review data.

**Table 1.** Result of scraping data

No	Reviews
1	Please improve the maps again. I often get lost because the maps keep missing, and sometimes they run by themselves, even though I haven't walked yet. And my location is also not read accurately. The blue sign keeps moving.
2	Why does my Google Maps often have interference, or is the arrow pointer usually stuck/does not work? So, it is tough to find an address, and it is not helpful even though Google Maps is very accurate when first used (buy a smartphone or first use). Please. Checked, yes, because I am a driver by profession. In addition, if you can restore the way "Google Maps" works like before, where the "arrow or red dot" can move automatically like before.
3	The latest Google maps are terrible, so if we use Google Assistant to point to Jakarta, he doesn't go directly to the destination. Instead, he goes to the maps first, we have to press the MAPs, and then we press START. It is very complicated and impractical, whereas yesterday's maps were excellent. We have to say he is already heading where we are going; we don't need to go to the maps and press first. It isn't straightforward. especially while driving. Could you return it to the old maps?
4	Why is it that my saved marks are now lost? Now I can't save the mark anywhere. The green flag signs and the heart sign are all gone. It's getting increasingly difficult for me to use Google Maps; even reading the road becomes difficult because the display changes, not like before. Anyway, I am very disappointed with the changes in Google Maps.
5	Thank you for making the journey more leisurely; the update problem is getting worse: *) location history doesn't work even though you move locations *) the blue dot is always grey even though the permissions have been set *) It shoots everywhere so that it can't be read by the application system that requires a map apk, the total distance traveled doesn't appear as if the developer is aware of this, but why are they silent, what's the point of having a star rating, the goal is to refer to better changes but the reality is different.

### 4.3. Labelling Data

Furthermore, Google Maps user review data is labeled by calculating the text polarity value on each data based on the lexicon dictionary. The following review data has been labeled with positive sentiment and negative sentiment.

**Table 2.** Result of labeling data

No	Reviews	Sentiment
1	Please improve the maps again. I often get lost because the maps keep missing, and sometimes they run by themselves, even though I haven't walked yet. And my location is also not read accurately. The blue sign keeps moving.	Negative
2	Why does my Google Maps often have interference, or is the arrow pointer usually stuck/does not work if used? So, it isn't easy to find an address, and it is not helpful even though Google Maps is very accurate when first used (buy a smartphone or first use). Please. Checked, yes, because I am a driver by profession. In addition, if you can restore the way "Google Maps" works like before, where the "arrow or red dot" can move automatically like before.	Negative
3	The latest Google maps are terrible, so if we use Google Assistant to point to Jakarta, he doesn't go directly to the destination. Instead, he goes to the maps first. We have to press the MAPs, and then we press START. It is very complicated and impractical, whereas yesterday's maps were excellent. We have to say he is already heading where we are going. We don't need to go to the maps and press first, but it isn't straightforward. especially while driving. Could you return it to the old maps?	Negative
4	Why is it that my saved marks are now lost? Now I can't save the mark anywhere. The green flag signs and the heart sign are all gone. It's getting increasingly difficult for me to use Google Maps; even reading the road becomes difficult because the display changes, unlike before. Anyway, I am very disappointed with the changes in Google Maps.	Negative
5	Thank you for making the journey more accessible. The update problem is getting worse: *) location history doesn't work even though you move locations *) the blue dot is always grey even though the permissions have been set *) It shoots everywhere so that it can't be read by the application system that requires a map apk, the total distance traveled doesn't appear as if the developer is aware of this, but why are they silent, what's the point of having a star rating, the goal is to make a reference for better changes but the reality is different.	Negative

### 4.4. Preprocessing Data

After the review data is labeled based on the lexicon dictionary's word polarity value, the data goes through the text preprocessing stage. At this stage, the data will be processed to make it easier for computer systems to understand the model by cleaning data, normalizing data, removing symbols and numbers, converting words to lowercase, removing data with neutral sentiments, stopwords, stemming, and tokenizing. Here are some of the stages:

### 1. Cleaning text

The text cleaning process is part of data preprocessing, which cleans the data from symbols, numbers, lowercase, and punctuation. The following are the cleaning stages in the script.

- a. Import pandas library as pd to read data and import streamlet as the framework used
- b. Run the clean\_maps\_data function to process data cleaning

Here are the results after the data-cleaning process on the text.

**Table 3.** Result of cleaning data

No	Reviews	Sentiment
1	please fix the maps again. I often get lost because the maps are always left behind, and sometimes the maps move by themselves even though i haven't walked and my location is also not read accurately. The blue marks keep moving	Negative
2	why is my Google Maps often having problems or when I use the arrow sign, the directions often get stuck and don't work, making it very difficult to find an address, or it doesn't help even though i first used it when I just bought a smartphone or when i first used google maps it was perfect and accurate please check this because i work as an additional driver if possible please restore the way google maps works as before where the arrow or red dot can move automatically like before	Negative
3	the latest google maps is terrible so if we use Google Assistant to direct us to Jakarta it doesn't go directly to the destination instead, it goes to the map first, we have to press the map, then we press start it is very complicated it not practical even though the previous maps were excellent we just have to talk, its already heading to where were going we don't need to go to the map first and then press it first ah its very complicated especially while driving it returns the maps to how they used to be	Negative
4	why are the signs that I saved now gone i can't save signs anywhere the green flag sign, the heart sign all of them are gone its getting more challenging for me to use Google Maps even reading the road is challenging because the display has changed it not like it used to be in very disappointed with the changes to google maps	Negative
5	thank you for making it easier to travel. The problem of updating is that the location history does not work; even though the location is changed. The blue dot is always grey; even though the permissions have been set, it shoots everywhere so that the application system that requires the map application cannot read the total distance traveled does not appear it seems like the developer is aware of this but why is it silent what is the purpose of the star rating the goal is to provide a reference for better changes but the reality is different	Negative

### 2. Normalization

This stage is the stage of correcting words that are not following the good and correct Indonesian language order.

- a. Declaration of term words into standardized words stored in the norm function
- b. Run the following function to convert term words into standard words

The following are the review data results after going through the word normalization stage.

**Table 4.** Result of normalization

No	Reviews	Sentiment
1	please fix the maps again. I often get lost because the maps are always left behind, and sometimes the maps move by themselves even though I haven't walked; my location is also not read accurately; the blue marks keep moving	Negative
2	why is my Google Maps often having problems, or when I use the arrow sign, the directions often get stuck and don't work, making it very difficult to find an address, or it doesn't help even though I first used it when I just bought a smartphone or when I first used google maps it was perfect and accurate please check this because I work as an additional driver if possible please restore the way google maps works as before where the arrow or red dot can move automatically like before	Negative
3	the latest Google Maps is very bad, so if we use Google Assistant to direct us to Jakarta, it doesn't go directly to the destination; instead, it goes to the map first. We have to press the map, and then we press start; it is very complicated and not practical even though the previous maps were excellent; we have to talk; it is already heading to where we are going; we don't need to go to the map first and then press it first ah it is very complicated, especially while driving it returns the maps to how they used to be	Negative
4	why are the signs that I saved now gone? I can't save signs anywhere. The green flag sign the heart sign all of them are gone; it is getting more challenging for me to use Google Maps; even reading the road is challenging because the display has changed; it is not like it used to be im very disappointed with the changes to google maps	Negative
5	thank you for making it easier to travel. The problem with updating is that the location history does not work; even though the location is changed, the blue dot is always grey; even though the permissions have been set, it shoots everywhere so that the application system that requires the map application cannot read the total distance traveled does not appear it seems like the developer is aware of this but why is it silent what is the purpose of the star rating the goal is to provide a reference for better changes but the reality is different	Negative

### 3. Stopword removal

This stage is processed to remove words that do not have essential meanings. The following stages are carried out to apply stopword removal.

- a. Import library stopwords and word\_tokenize
- b. Declare the stopwords\_removal function and run the function

The following review data has gone through the stopwords removal stage.

**Table 5.** Result of stopword removal

No	Reviews	Sentiment
1	ask for the maps to be fixed, I often get lost, the maps are always left behind, and sometimes the maps run by themselves even though the road and location are read accurately, the blue signs keep moving	Negative
2	Google Maps often has problems when used. The arrow sign shows the road is frequently congested, and it is challenging to find an address; it helps even though when I first used it, I had just bought my first smartphone; the use of Google Maps is perfect and accurate; it was checked because I work as an additional driver if I return to how google maps used to work the arrow sign with the red dot would move automatically first.	Negative
3	the latest Google Maps is very bad, so if you use Google Assistant, it will direct you to Jakarta directly towards your destination, but you have to press the map first and then press start It is very complicated in practice even though yesterday's maps were excellent you have to talk and go you need to go to the map first and then press start oh its very complicated to drive you have to return the maps first	Negative
4	Now, the same sign is gone somewhere. Wherever the green flag and heart signs are gone, it is getting more challenging to use Google Maps, even to read the road so hard that the display changed before. I am disappointed with the changes to Google Maps	Negative
5	thank you for making it easier to travel. The update problem is getting more and more historical locations; the blue dot is always grey, the permission settings are shooting up where the application system reads the application requires a map application, and the total distance traveled the developer displays it is aware it is silent what is the purpose of the star rating its purpose is to refer to better changes the reality is different	Negative

#### 4. Stemming dan Tokenization

The tokenization stage is processed to cut the review sentence data into its constituent words. The stemming stage is processed to convert words in the review text into essential words. The following are the stages of the stemming and tokenizing script:

- a. Import StemmerFactory
- b. Run the stemmer and tokenize function script

The following review data has gone through the stemming and tokenizing stages.

**Table 6.** The result of stemming and tokenizing

No	Reviews	Sentiment
1	ask for a good map often get lost the map stays on and sometimes the map moves by itself even though the road and location are read accurately the blue signs keep moving	Negative
2	Google Maps often has problems using the arrow sign to show the way. It is usually congested, and the road is tough to find an address help even though when I first used it, I bought a smartphone for the first time using Google Maps. It was perfect and accurate; check the driver's profession. If you return to how Google Maps works, use the arrow sign the red dot and call it automatic movement.	Negative
3	the new Google Maps is terrible, so if you use Google Assistant to speak directions to Jakarta and go straight to the destination go to the destination, you have to press the maps first, and then press start. It isn't straightforward in practice, even though yesterday's Maps were excellent. Speak and go to the destination; you need to go to the maps first, then press start; oh, it is not very easy; you have to steer and go back to the maps first	Negative
4	Now the save sign is gone, and now the save sign is where the green flag sign and heart sign are; it's getting harder to use Google Maps even to read the road, and it's hard to show change. first, I was disappointed to change Google Maps	Negative
5	thank you! An easy way to update the problem is more history, location road change, location blue dot is always grey, permission setting is fast where to read the application system needs a map application total distance traveled show the developer aware silent why is there a star rating its purpose is for reference change it better its natural it is different	Negative



#### 4.5. Modelling Data

In this modeling process, data that has gone through the preprocessing stage will be trained with the Support Vector Machine classification model. The SVM model will train positive sentiment data and sentiment. In this modeling process, data that has gone through the preprocessing stage will be trained with the Support Vector Machine classification model. The SVM model will train positive and harmful sentiment data by generating classification accuracy values using the Support Vector Machine.

- Import library `train_test_split` to divide data into training data and test data, import library `TfidfVectorizer` to calculate word weighting, import `SVC` for data classification using the Support Vector Machine algorithm, import library `accuracy_score`, `recall_score`, `precision_score`, and `f1_score` to display accuracy, recall, precision, and f1 values.
- Training data and test data are divided based on reviews in the "cleaned\_tokenized\_stemmed" column, namely the data resulting from the stemming and tokenizing process.
- Run the following script function for the data modeling process.

The following are the results of the accuracy, precision, recall, and f1-score of training models using the Support Vector Machine algorithm classification

```
Classification Report      precision recall f1-score support
```

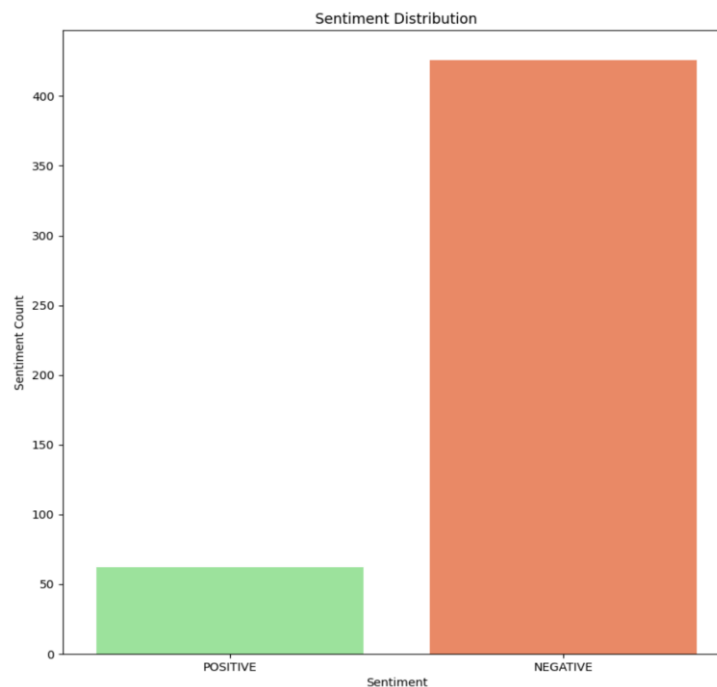
NEGATIVE	0.86	1.00	0.92	83
POSITIVE	1.00	0.07	0.12	15
accuracy		0.86		98

```
macro avg 0.93 0.53 0.52 98 weighted avg 0.88 0.86 0.80 98
```

**Fig 3.** Hyperparameter Tuning Support Vector Machine

#### 4.6. Visualization of Sentiment Classification Results

After that, review data is visualized by displaying a bar chart to determine the number of positive and negative sentiments. By doing visualization, users can compare the number of positive and negative sentiments



**Fig 4.** Sentiment classification results

#### 4.7. Positive Sentiment Visualization

Visualization is also done on each positive word in the review data by displaying a word cloud. The number of words influences the visualization size of each word in each user review of Google Maps. Therefore, the more a word is, the larger the visualization size of the word. In the positive word data, the largest size is "often" because many users use that word in their reviews.



Fig 5. Positive sentiment visualization

#### 4.8. Negative Sentiment Visualization

Visualization is also done on each negative word in the review data by displaying a word cloud. The number of words influences the visualization size of each word in each user review of Google Maps. Therefore, the more a word is, the larger the visualization size of the word. In the harmful word data, the largest size is "which" because many users use it in their reviews.



Fig 6. Negative Sentiment Visualization

#### 4.9. Latent Dirichlet Allocation Topic Modelling

The Latent Dirichlet Allocation topic model is built from the Google Maps user review dataset that has gone through the data preprocessing stage, which is built to facilitate the extraction of information on the reviews that have been used. The LDA model is evaluated by iteratively building the model to find the best coherence value for each topic model. Fig 7 shows the best coherence value **in Model Topic 2 at 35.15% and Model Topic 5 at 33.39%**. In the case of this model topic, no fixed and specific rule says that a particular coherence value can be said to be good or bad. Therefore, we used a sample of model topics with an average percentage of model topic 5 to extract information related to Google Maps user reviews.

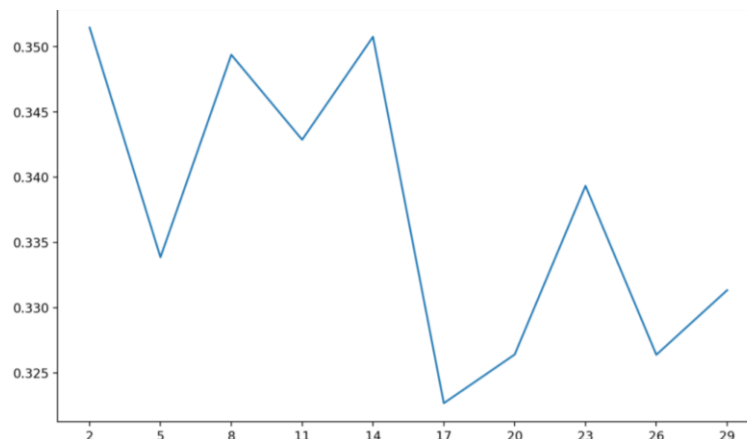


Fig 7. Coherence value of LDA

Based on the coherence value of the LDA model topic, it can be observed in Table 7 that the probability value/confidence level of the model in clustering has a good value in the range of 0.49 to 0.85.

**Table 7. Review Probability Value**

No	Reviews	Topic Number	Probability
1	thank you for your fast, friendly, and disciplined service	2	0.6657
2	sometimes, I get direction errors and often go off-route	1	0.7763
3	The GPS signal keeps getting lost. Strangely, a strong internet network GPS signal gets lost. Google	1	0.7311
4	good location accurate original accurate and precise route	4	0.7629
5	the timeline has always been there, so it is easy to check the updated date now that it is all gone. See reviews of many protests related to the new timeline	5	0.4958

Based on the results of modeling the topic of Google Maps user reviews in **Table 8**, many users are disappointed with the Google Maps service that provides route errors, the timeline of trips provided by Google Maps often has errors, and the service offers uphill routes. It does not match the type of vehicle, but not a few users also appreciate the Google Maps service that helps travel.

**Table 8. Topic Modelling Results**

No	Reviews	Topic Number	Aspect
1	beneficial application for searching for successful address maps continues	4	Helping with travel
2	there is no difference between the roads for large vehicles and small vehicles; they often target the direction of complex roads, use cars, ride motorbikes, tell them to go through a small alley	3	Providing routes that do not match vehicles
3	screw it up now so it's slow, and when it is close to auto-zoom, the old application is entirely rubbish	5	Slow service
4	the current map is less accurate and often does not reach the exact destination point	2	Route error
5	damn, the more you update, the more error you get when typing the location. Even a preview appears if you retype the address to find it, but it gives you an address outside the area, and city google maps doesn't know the location	1	Error travel timeline

In addition, topic visualization can also be seen on the plot by selecting 1 of 5 topics and relevance metric values that can be adjusted by the user with the provision that when the slider is set closer to 1, the system will focus more on words that are very specific to the topic. In this modeling, it can be understood that the visualization of each circle represents one topic by showing the number of each topic. The size of the circle shows the level of dominance of the topic in the corpus, so the larger the size of the circle, the more dominant the topic is. Each topic displays the 30 most relevant words, with a bar chart visualization to show the frequency of words on the topic. Bar charts in red show the frequency of words in the whole corpus, while those in blue show the frequency of words in the topic.

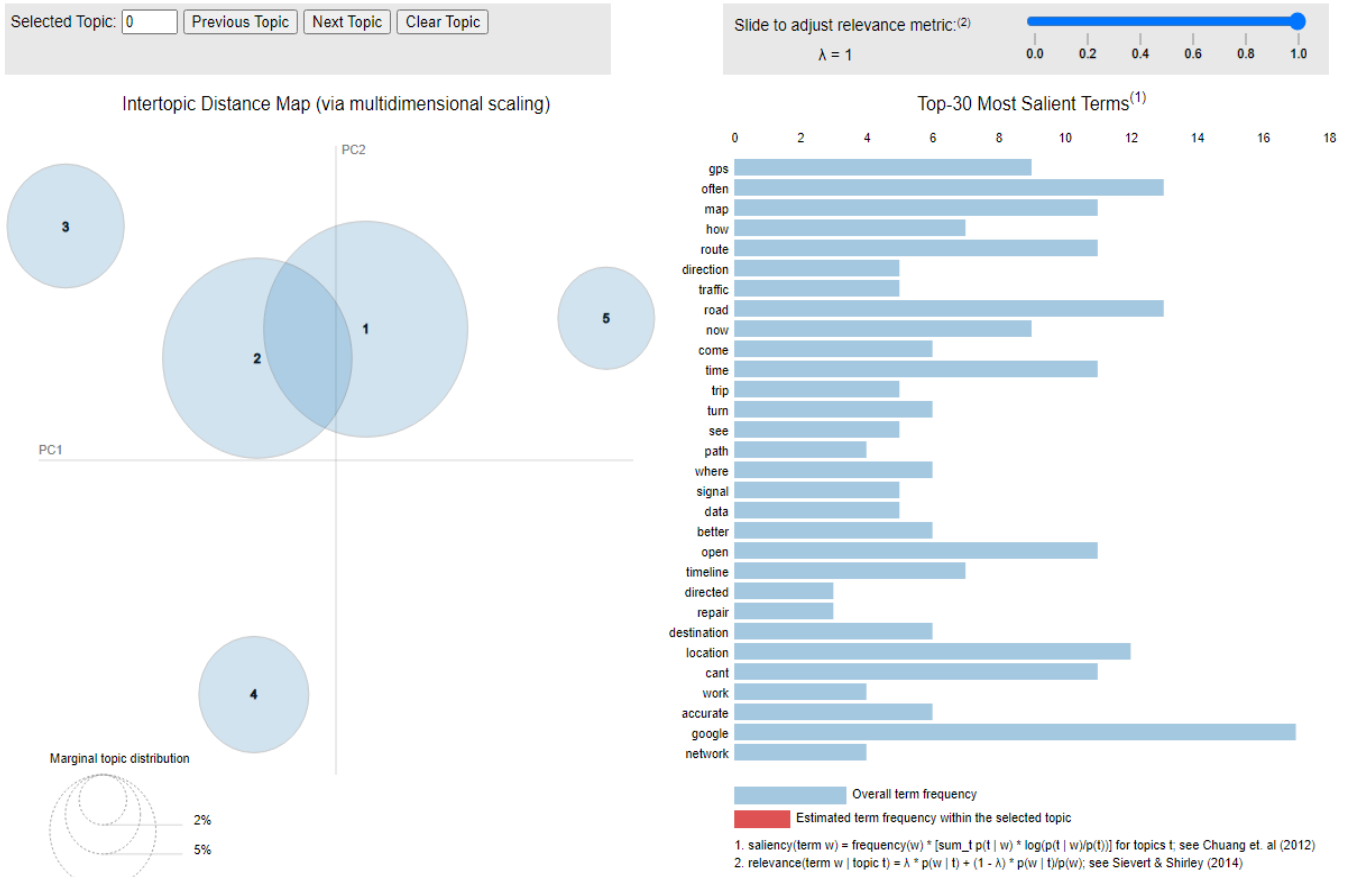


Fig 8. Visualization of Topic Modelling

4.10. Evaluation Classification

The confusion matrix is the method to evaluate the classifier's performance at each label level. The confusion matrix has diagonal lines that indicate correctly classified samples, and each row is summed with the total number of actual samples. In this study, the Support Vector Machine classification method produced an accuracy value of 86%, precision of 93%, recall of 53%, and f1-score of 52% in classifying Google Maps user review sentiments, as shown in Fig. 9 below.

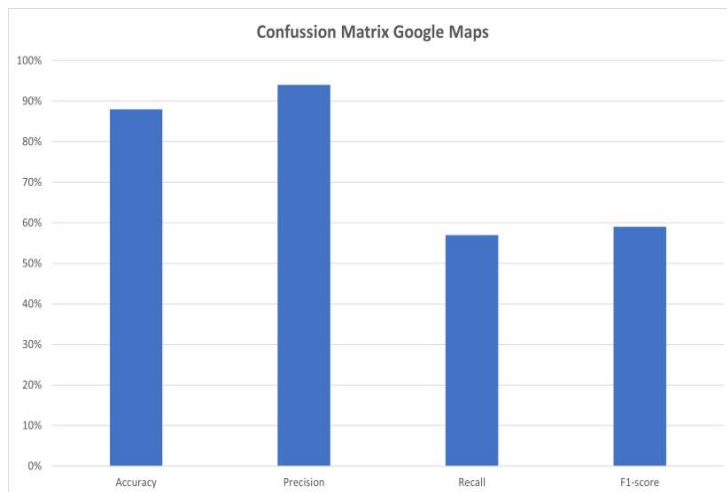


Fig 9. Evaluation Classification Support Vector Machine

## 5. Conclusion

Based on the results of sentiment analysis that has been carried out on Google Maps user reviews, the following conclusions are obtained:

1. The results show that using the Support Vector Machine algorithm to build a sentiment classification model is suitable, which can be proven by classifying Google Maps user reviews into positive and negative classes.
2. The results of sentiment analysis of Google Maps application user reviews using the Support Vector Machine algorithm produce an accuracy value of 86%, precision of 93%, recall of 53%, and f1-score of 52%.
3. This research also applies topic modeling using Latent Dirichlet Allocation, which displays terms on five main topics that discuss users using application services through 500 reviews taken on the Google Play Store platform.
4. Based on the evaluation results of the analysis obtained, the Support Vector Machine algorithm has a good accuracy value in performing sentiment classification.
5. Latent Dirichlet Allocation topic modeling can also produce main and hidden topics in Google Maps user reviews, namely the Google Maps service that provides route errors, the travel timeline supplied by Google Maps often has errors, the service offers uphill road routes and does not match the type of vehicle, but not a few users also appreciate the Google Maps service that helps travel.

## References

- [1] A. Rifa'i, H. Sujaini, and D. Prawira, "Sentiment Analysis Objek Wisata Kalimantan Barat Pada Google Maps Menggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 7, no. 3, p. 400, 2021, doi: 10.26418/jp.v7i3.48132.
- [2] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps," *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.
- [3] V. W. D. Thomas and F. Rumaisa, "Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, p. 1767, 2022, doi: 10.30865/mib.v6i3.4218.
- [4] R. Rosdiana, M. Ula, and H. A. K. Aidilof, "Implementasi Pemodelan Citra Model Svm (Support Vector Machine) Dalam Penentuan Pengklasifikasian Jenis Suara Kontes Burung," *Jurnal Informatika Kaputama (JIK)*, vol. 5, no. 2, pp. 317–324, 2021, doi: 10.59697/jik.v5i2.264.
- [5] B. A. Tondang, Muhammad Rizqan Fadhil, Muhammad Nugraha Perdana, Akhmad Fauzi, and Ugra Syahda Janitra, "Analisis pemodelan topik ulasan aplikasi BNI, BCA, dan BRI menggunakan latent dirichlet allocation," *INFOTECH : Jurnal Informatika & Teknologi*, vol. 4, no. 1, pp. 114–127, 2023, doi: 10.37373/infotech.v4i1.601.
- [6] L. B. Imawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naive Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [7] A. D. Cahyani, "Analisa Kinerja Metode Support Vector Machine untuk Analisa Sentimen Ulasan Pengguna Google Maps," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 3, pp. 604–613, 2023, doi: 10.47065/josyc.v4i3.3426.
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [9] M. Ula and S. Fachrurrazi, "Analisis Sentimen Cyberbullying pada Media Sosial Twitter menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *TECHSI - Jurnal Teknik Informatika*, vol. 14, no. 2, p. 91, 2023, doi: 10.29103/techsi.v14i2.12103.
- [10] M. Chen, Z. Guo, K. Abbass, and W. Huang, "Analysis of the impact of investor sentiment on stock price using the latent dirichlet allocation topic model," *Front Environ Sci*, vol. 10, pp. 82–87, 2022, doi: 10.3389/fenvs.2022.1068398.
- [11] [11]F. J. Damanik and D. B. Setyohadi, "Analysis of public sentiment about COVID-19 in Indonesia on Twitter using multinomial naive bayes and support vector machine," *IOP Conf Ser Earth Environ Sci*, vol. 704, no. 1, 2021, doi: 10.1088/1755-1315/704/1/012027.
- [12] Riyanto and A. Azis, "Application of the Vector Machine Support Method in Twitter Social Media Sentiment Analysis Regarding the Covid-19 Vaccine Issue in Indonesia," *Journal of Applied Data Sciences*, vol. 2, no. 3, pp. 102–108, 2021, doi: 10.47738/jads.v2i3.40.
- [13] M. Erkamim, "Sentiment Analysis of Shopee App Reviews Using Random Forest and Support Vector Machine," *Jurnal Prodi Teknik Informatika FIK Universitas Muslim Indonesia*, vol. 15, no. 3, pp. 427–435, 2023, doi: <https://doi.org/10.33096/ilkom.v15i3.1610.427-435>.
- [14] M. N. Muttaqin and I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor," *UNNES Journal of Mathematics*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [15] Nurdin and D. Astika, "Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Pada Supermarket Sejahtera Lhokseumawe," *Techsi*, vol. 6, no. 1, pp. 134–155, 2015, doi: 10.29103/TECHSI.V7I1.184.
- [16] H. Nuraliza, O. N. Pratiwi, and F. Hamami, "Analisis Sentimen IMDb Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance," *Jurnal Mirai Manajemen*, vol. 7, no. 1, pp. 1–17, 2022, doi: <https://doi.org/10.37531/mirai.v7i1.2222>.
- [17] M. Faisal and Z. Fitri, "Information and Communication Technology Competencies Clustering for students for Vocational High School Students Using K-Means Clustering Algorithm," *International Journal of Engineering, Science & InformationTechnology (IJESTY)*, pp. 111–120, 2022, doi: 10.52088/ijesty.v1i4.318.
- [18] M. Qamal et al., "ANALISIS SENTIMEN TOKO ONLINE MENGGUNAKAN dilakukan oleh Mehdi Mursalat Ismail dan Kemas Muslim Lhaksamana dengan judul ' Sen timen Analisis Pada Media Online Mengenai Pemilihan Presisen 2019 dengan Menggunakan Metode Naive Bayes ',," *Jurnal Teknologi Terapan and Sains 4.0*, vol. 2, no. 3, pp. 642–650, 2021, doi: DOI: <https://doi.org/10.29103/tts.v2i3.6771>.
- [19] J. Quillo-Espino, R. M. Romero-González, and A. Lara-Guevara, "Advantages of Using a Spell Checker in Text Mining Pre-Processes," *Journal of Computer and Communications*, vol. 06, no. 11, pp. 43–54, 2018, doi: 10.4236/jcc.2018.611004.

- [20] U. Umarrazi and N. Nurdin, "Peramalan J Umlah Keuntungan Mi E I Nstan Pada Sumber Rezeki Kota Lhokseumawe Menggunakan Metode Tri Ple Exponenti Al Smooti Ng," *Sisfo: Jurnal Ilmiah Sistem Informasi*, vol. 1, no. 2, pp. 185–218, 2017, doi: 10.29103/sisfo.v1i2.248.
- [21] R. Wahyudi and G. Kusumawardana, "Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *Jurnal Informatika*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [22] S. Shibul Muna, Nurdin and Taufiq "Tokopedia and Shopee Marketplace Performance Analysis Using Metrix Google Lighthouse," *International Journal of Engineering, Science & InformationTechnology (IJESTY)*, pp. 106–110, 2022, doi: 10.52088/ijesty.v1i4.312.
- [23] N. Nurdin, M. Suhendri, Y. Afrilia, and R. Rizal, "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naïve Bayes Classifier (NBC)," *Sistemasi*, vol. 10, no. 2, pp. 268–279, 2021, doi: 10.32520/stmsi.v10i2.1193.
- [24] K. B. Putra and R. P. Kusumawardani, "Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)," *Jurnal Teknik ITS*, vol. 6, no. 2, pp. 4–9, 2017, doi: 10.12962/j23373539.v6i2.23205.