

Research Article

Sentiment Analysis of the *Kabur Aja Dulu* Trend on X as a Basis for Designing a Public Sentiment Monitoring System Using Naïve Bayes and SVM

Sutisna^{1*}, Tri Wahyudi², Dwi Swasono Rachmad³, Fachrur Rozi⁴

¹Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia;
email: sutisna@stikomcki.ac.id

²Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia;
email: triwahyudi@stikomcki.ac.id

³Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia;
email: dwi.swasono@stikomcki.ac.id

⁴Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, Indonesia;
email: rozifachrur511@gmail.com

*Corresponding Author: sutisna@stikomcki.ac.id

Abstract: Social media X (Twitter) has become the main platform for the Indonesian public to express opinions, including on the trend of 'kabur aja dulu' (let's just run away for a bit). This research aims to classify the sentiments of the public using the Naïve Bayes and Support Vector Machine (SVM) methods, and to compare the accuracy of both in sentiment analysis. Data was collected via the Twitter API with the hashtag #kaburajadulu, resulting in 2,067 tweets, which, after the cleansing process and manual labeling, left 385 data points. The analysis process followed the CRISP-DM stages, which include business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Model evaluation was conducted using a confusion matrix with accuracy, precision, and recall metrics. The classification results show that 82% of tweets have a positive sentiment and 18% negative. The Naïve Bayes algorithm achieved an accuracy of 86.49%, slightly lower than SVM, which reached 88.05%. In conclusion, Support Vector Machine is more effective in sentiment classification on public opinion data. This research contributes to the digital mapping of public opinion and recommends the development of automatic labeling methods as well as the exploration of advanced algorithms in the future.

Keywords: Naive Bayes; Public Opinion Monitoring; Sentiment Analysis; Social Media X; SVM.

Received: 04 June 2025

Revised: 15 July 2025

Accepted: 12 August 2025

Published: 30 August 2025

Curr. Ver.: 30 August 2025



Copyright: © 2025 by the authors.

Submitted for possible open

access publication under the

terms and conditions of the

Creative Commons Attribution

(CC BY SA) license

([https://creativecommons.org/li](https://creativecommons.org/licenses/by-sa/4.0/)

[censes/by-sa/4.0/](https://creativecommons.org/licenses/by-sa/4.0/))

1. Introduction

The rapid development of information technology has brought significant changes in the way people communicate and express their opinions. Today, social media has become an inseparable part of daily life, serving as a platform for communication, information sharing, and opinion expression. The use of social media can generate both positive and negative impacts on individuals and society (Rahman et al., 2021). One consequence of this development is the increasing use of social media as the primary medium for people to express opinions, emotions, and responses to various events. Social media platform X (formerly known as Twitter) is among the most widely used platforms for openly, quickly, and real-time sharing of public opinions.

Between February 2025 and July 2025, a popular trend emerged on social media X under the phrase “kabur aja dulu” (“just run away first”). This trend was widely used by users to express attitudes or responses toward situations perceived as unfavorable, stressful, or confusing. The trend represents an interesting social phenomenon because it reflects how people respond to ongoing social, political, and cultural dynamics. However, the meaning

behind its usage varies considerably depending on the context and emotional tone of the users, making a deeper analysis necessary to understand the sentiments embedded within the trend.

Sentiment analysis is the process of identifying and classifying users' opinions or feelings toward a particular topic into positive, negative, or neutral sentiment categories. It can also be defined as a textual data analysis technique that employs Natural Language Processing (NLP) and Machine Learning (ML) to automatically evaluate the emotional tone expressed by the author (Detty et al., 2023; Asrumi et al., 2023). Through sentiment analysis, insights can be obtained regarding public perceptions of specific issues. Such information is highly valuable for policymakers, social institutions, and business organizations in formulating strategies and making data-driven decisions based on public opinion.

To perform sentiment analysis on large-scale and unstructured data such as social media posts, a systematic and technology-based approach is required. Naïve Bayes is widely recognized for its simplicity and efficiency in handling large datasets, while Support Vector Machine (SVM) offers superior classification accuracy, particularly when dealing with complex data patterns (Mulyoto, 2024; Mahendro & Abimanto, 2023). The combination of machine learning algorithms such as Naïve Bayes (NB) and Support Vector Machine (SVM) has become a popular approach due to their effectiveness in classifying informal text, slang expressions, and emoticons commonly found on social media platforms (Pratama & Maharani, 2025).

Based on the foregoing discussion, this study aims to conduct sentiment analysis of the "kabur aja dulu" trend on social media X and to develop a public sentiment monitoring system using Naïve Bayes and SVM algorithms. The proposed system is expected to serve as a tool for automatically and comprehensively understanding public opinion, while also providing a foundation for strategic decision-making across various sectors.

2. Literature Review

Sentiment Analysis

Sentiment analysis is a Natural Language Processing (NLP) and Machine Learning (ML) technique used to automatically evaluate text based on the emotions or opinions expressed by the author, typically categorized as positive, negative, or neutral sentiments (Detty et al., 2023). Sentiment analysis has become an important area of NLP due to the increasing interaction between humans and computers, information extraction, and the identification of opinions and emotions from existing textual data. It involves understanding words, phrases, and expressions that indicate positive, negative, or neutral attitudes within a text. Sentiment analysis is commonly used to extract valuable insights, opinions, perceptions, and emotions from both structured and unstructured textual data for decision-making and research purposes (Detty et al., 2023; Asrumi et al., 2023).

Twitter (X)

Twitter, currently known as X, is a popular social media and microblogging platform that allows users to share thoughts, opinions, and information in real time. Users can freely express their views regarding products, services, facilities, political issues, and various topics under public discussion. Due to its openness and extensive user engagement, Twitter has become one of the primary sources of data for sentiment analysis research (Rahman et al., 2021; Maulaya & Junadhi, 2022).

Application Programming Interface (API)

An Application Programming Interface (API) is a set of protocols, functions, commands, and rules that enable communication and interaction between different software applications. APIs serve as intermediaries that allow systems to exchange data and functionality efficiently without requiring direct access to each other's internal processes. Through APIs, software components can communicate and integrate seamlessly, facilitating interoperability among various applications and services (Bororing & Gunawan, 2024).

Google Colab

Google Colab, or Google Colaboratory, is a cloud-based platform that enables users to write and execute Python code in an interactive environment. Developed by Google, Colab provides access to powerful computational resources, including Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), without requiring additional software installation on local computers. Google Colab is widely used by researchers, developers, and students due to its accessibility and its support for machine learning, data science, and data analysis projects (Mulyoto, 2024).

Python

Python is an interpreted programming language that is easy to learn and can be executed across various platforms. It emphasizes code readability and simplicity while providing extensive capabilities for software development, data analysis, machine learning, and scientific computing. Python is supported by a large and comprehensive standard library, making it one of the most widely used programming languages today (Mulyoto, 2024).

Classification

Classification is the process of categorizing data into predefined classes based on the characteristics, attributes, and patterns associated with each class. This process begins by establishing a set of training data as a reference for classification criteria, which is then used to classify new data into the appropriate categories. The primary objective of classification is to identify patterns that distinguish one class from another, allowing unknown objects or instances to be assigned to the most suitable category. In sentiment analysis studies, classification is commonly used to categorize text into positive, negative, or neutral sentiments (Amna et al., 2023).

Naïve Bayes

Naïve Bayes is a classification algorithm based on Bayes' Theorem that assumes the features used for prediction are conditionally independent of one another. Although this assumption may not always hold true in real-world scenarios, Naïve Bayes often produces effective results, especially when dealing with relatively small datasets or complex feature relationships. Due to its simplicity, computational efficiency, and satisfactory performance, Naïve Bayes is widely used in sentiment analysis and text classification tasks (Mulyoto, 2024; Hakim & Sugiyono, 2024).

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm used for both classification and regression tasks. Compared to many other classification methods, SVM is supported by a strong mathematical foundation and can effectively handle both linear and non-linear classification problems. SVM works by identifying an optimal hyperplane that separates data points belonging to different classes. Through the use of kernel functions, SVM can transform data into higher-dimensional spaces, enabling the classification of data that are not linearly separable. These characteristics make SVM a popular and effective algorithm for sentiment analysis applications (Rahman et al., 2021; Pamungkas & Kharisudin, 2021; Mahendro & Abimanto, 2023).

Dataset

A dataset is a systematically organized collection of data used for analysis and research purposes. Datasets may consist of various types of data, including text, images, audio, or numerical values. They are commonly utilized in computer science, data science, statistics, and scientific research for training machine learning models, validating hypotheses, generating predictions, and extracting meaningful insights from data (Amna et al., 2023).

RapidMiner

RapidMiner is a powerful and versatile analytics platform used for data processing, analysis, and modeling. The platform provides an intuitive visual environment that enables users to perform various analytical tasks without requiring extensive programming skills. RapidMiner supports a wide range of machine learning algorithms and statistical analysis techniques, as well as tools for data preprocessing, feature extraction, data visualization, and model evaluation. Through RapidMiner, users can import data from multiple sources and conduct comprehensive analytical processes to generate valuable insights for decision-making (Amna et al., 2023).

Cross-Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM (Cross-Industry Standard Process for Data Mining) is a widely recognized methodology for conducting data mining projects. It was developed in 1996 by several industry organizations, including NCR, Daimler-Benz, and SPSS, to provide a standardized framework for solving business and research problems through data mining techniques. CRISP-DM consists of six phases that guide the entire data mining process, from understanding business objectives to deploying the final model (Zuhri et al., 2020).

Business Understanding

The Business Understanding phase focuses on defining business objectives and identifying the problems that need to be addressed. This stage requires a thorough understanding of business processes, regulations, and operational requirements to ensure that the data mining project aligns with organizational goals (Zuhri et al., 2020).

Data Understanding

Data Understanding involves collecting, exploring, and examining data to gain insights into its characteristics and quality. Activities in this phase include data collection, data description, data quality assessment, and preliminary data exploration (Zuhri et al., 2020).

Data Preparation

Data Preparation is the process of transforming raw data into a suitable format for modeling. This phase often consumes the largest portion of project resources and includes activities such as:

a. **Data Validation**

Data should be reviewed to ensure accuracy, consistency, and reliability throughout the analysis process.

b. **Outlier Handling**

Outliers may significantly influence model performance and should be carefully identified and managed before modeling.

c. **Handling Missing and Inconsistent Data**

Missing or inconsistent data should be treated appropriately according to the objectives of the analysis to maintain data quality and model reliability (Amna et al., 2023).

Modeling

Modeling involves applying appropriate analytical techniques and machine learning algorithms to develop predictive or descriptive models. Common objectives in data mining include estimation, prediction, classification, clustering, and association analysis. Depending on the research objectives, one or more modeling approaches may be combined to achieve more detailed results (Amna et al., 2023).

Machine learning algorithms are generally categorized into two groups: supervised learning and unsupervised learning. Supervised learning requires a target or dependent variable, whereas unsupervised learning does not.

Evaluation

The Evaluation phase involves interpreting and assessing the performance of the developed model. The objective is to determine whether the model meets the business and research goals established during the Business Understanding phase and whether it is suitable for practical implementation (Zuhri et al., 2020).

Deployment

Deployment is the final phase of the CRISP-DM methodology, where the developed model is implemented in real-world applications. Deployment planning should consider how the model's outputs will generate value, support decision-making, and integrate with existing operational systems (Zuhri et al., 2020).

3. Materials and Method

Research Data

This study employed a quantitative approach using textual data collected from the social media platform X (formerly Twitter). The dataset consisted of public tweets containing the hashtag #KaburAjaDulu, which reflects public opinion regarding social and governmental issues discussed in the online community.

Data were collected through the Twitter API integrated with RapidMiner. Initially, 5,000 tweets were retrieved, and after the preprocessing stage, 3,437 valid tweets remained for analysis. The dataset consisted of two attributes: Tweet and Sentiment. The Tweet attribute represents users' comments posted on X, while the Sentiment attribute represents the sentiment category assigned to each tweet, namely positive or negative.

The primary data source consisted of tweets obtained directly from X, whereas secondary data were collected through a literature review of journal articles, conference proceedings, books, and other relevant references related to sentiment analysis, machine learning, and social media analytics.

Research Framework

This study applied two supervised machine learning algorithms, namely Naïve Bayes (NB) and Support Vector Machine (SVM), to classify public sentiment regarding the #KaburAjaDulu trend. The research adopted the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which provides a systematic approach for conducting data mining projects.

The research process began with collecting tweets from X through the Twitter API, followed by data preprocessing, sentiment labeling, classification modeling, and model evaluation. The performance of the classification models was subsequently compared to determine the most effective algorithm for sentiment classification.

Data Collection

Data collection was conducted using the Twitter API available through the RapidMiner extension. Tweets written in Indonesian and containing the hashtag #KaburAjaDulu were retrieved from the platform. The collected tweets then underwent preprocessing before being used in the classification stage.

The data collection process involved retrieving tweets, cleaning the textual data, assigning sentiment labels, and preparing the dataset for further analysis. During preprocessing, irrelevant elements such as URLs, mentions, hashtags, punctuation marks, special symbols, and duplicate records were removed to improve data quality. Sentiment labels were assigned manually to classify each tweet into either positive or negative sentiment categories.

CRISP-DM Implementation

Business Understanding

The business understanding phase focused on identifying the research problem and defining the study objectives. The primary objective of this research was to analyze public sentiment toward the #KaburAjaDulu trend on social media. Public opinions expressed through tweets were selected as the main source of information for sentiment classification. At this stage, the research scope, objectives, and expected outcomes were established to guide subsequent phases of the study.

Data Understanding

The data understanding phase involved exploring and examining the collected dataset to understand its characteristics and quality. The dataset consisted of tweets obtained through the Twitter API using the hashtag #KaburAjaDulu. Each record contained textual content and a corresponding sentiment label. Exploratory analysis was conducted to identify data distribution, detect inconsistencies, and assess overall data quality before proceeding to the preprocessing stage.

Data Preparation

Data preparation was performed to transform raw textual data into a structured dataset suitable for machine learning algorithms. This stage included data cleansing, tokenization, stemming, noise removal, normalization, and sentiment labeling.

Data cleansing was carried out to remove URLs, hashtags, mentions, punctuation marks, special characters, and duplicate entries. Tokenization was applied to divide text into individual words or tokens. Stemming was then used to convert words into their root forms, while noise removal eliminated irrelevant symbols, numbers, and redundant terms. Text normalization was conducted to standardize informal words, abbreviations, and slang expressions into their formal forms. Finally, sentiment labels were assigned to categorize each tweet as either positive or negative. The output of this phase was a clean and structured dataset ready for classification modeling.

Modeling

The modeling phase employed Naïve Bayes and Support Vector Machine as classification algorithms. Both algorithms were implemented using RapidMiner to classify tweet sentiments based on the prepared dataset. The dataset was divided into training and testing subsets to build and validate the classification models. The purpose of this phase was to develop predictive models capable of identifying sentiment polarity from social media text data.

Evaluation

The evaluation phase was conducted to assess the performance of the classification models. Model performance was measured using a confusion matrix, which provides evaluation metrics such as accuracy, precision, recall, and F1-score. The evaluation results were analyzed to compare the effectiveness of Naïve Bayes and Support Vector Machine in classifying public sentiment related to the #KaburAjaDulu trend. The best-performing model was determined based on its classification accuracy and overall predictive performance.

4. Results and Discussion

Implementation and Testing

In this study, the algorithms used were Naive Bayes and Support Vector Machine (SVM). The data used were obtained from tweets collected from the social media platform Twitter (X) using Python programming through Google Colab. After the data were collected, they were processed using the CRISP-DM framework, which consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. To obtain classification accuracy values, RapidMiner was used to test the Naive Bayes and Support Vector Machine (SVM) algorithms. The Cross Industry Standard Process for Data Mining (CRISP-DM), which consists of six stages, was used to process the data.

Business Understanding

The dynamics of public sentiment regarding the “kabur aja dulu” trend may influence various aspects of social life, including industry, the economy, and other sectors. Although this trend is generally perceived negatively, many people also view the “kabur aja dulu” trend as something positive. Therefore, this study focuses specifically on the “kabur aja dulu” trend.

Data Understanding

At this stage, the researcher attempted to understand the data used from Twitter (X). The data were collected using Python programming and consisted of 2,067 tweets with two attributes, namely sentiment and tweet. The data were collected from February 10, 2025, to July 10, 2025.

Data Preparation

At the data preparation stage, several steps were carried out as follows:

a. Data Collection

Data were collected using Python programming with the assistance of Google Colab. The process included installing Python packages, obtaining a Twitter (X) token, crawling data based on text and date periods, and then saving the data. These steps are shown in Figure 4.1 below.



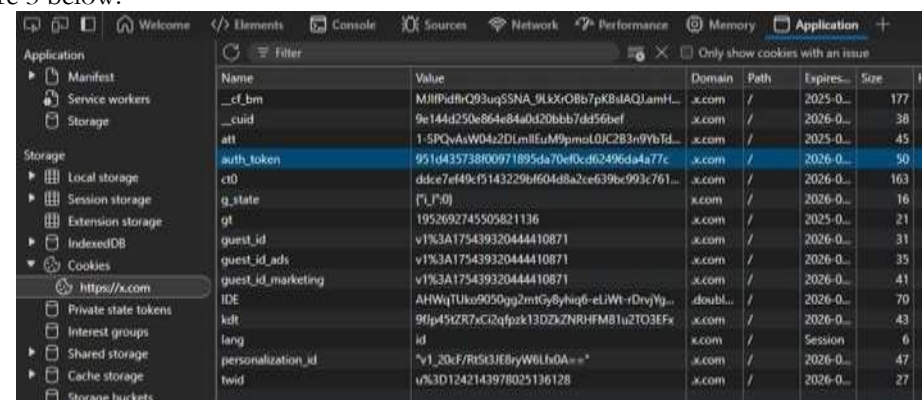
```

$ pip install pandas
$ pip install tweepy
$ pip install requests

```

Figure 1. Python Package Installation.

After the Python package was successfully installed, the next step was to obtain a token from the Twitter (X) account. After that, the token was entered as shown in Figure 2 and Figure 3 below.



Name	Value	Domain	Path	Expires...	Size
__cf_bm	MJHPidfrQ93uqSSNA_9LkxORB7pK8lAQJAmH...	.x.com	/	2025-0...	177
__cuid	9e144d250e864e84d0d20bb7dd56bef	.x.com	/	2026-0...	38
atl	1-5PQvAsW04z2DlmlEuM9pmeLOJC2B3rdYbTd...	.x.com	/	2025-0...	45
auth_token	951d435738f00971895da70ef0cd62496da4a77c	.x.com	/	2026-0...	50
ct0	ddce7ef49c15143229bf604d8a2ce639bc993c761...	.x.com	/	2026-0...	163
g_state	[{"_id":}]	.x.com	/	2026-0...	16
gl	1952692745505821136	.x.com	/	2025-0...	21
guest_id	v1%3A17543932044410871	.x.com	/	2026-0...	31
guest_id_ads	v1%3A17543932044410871	.x.com	/	2026-0...	35
guest_id_marketing	v1%3A17543932044410871	.x.com	/	2026-0...	41
IDE	AHWqTUKo9050jg2mtGy8yhiq6-eLjWt-iDvYjg...	.doubl...	/	2026-0...	70
kdt	9ffp45tZr7xCi2qfzk13DZ9z2NRHfMB1u2TO3EFx	.x.com	/	2026-0...	43
lang	id	.x.com	/	Session	6
personalization_id	~v1_20cF/Rt5i3jE8yW6Lb0A=*	.x.com	/	2026-0...	47
twid	v%3D1242143978025136128	.x.com	/	2026-0...	27

Figure 2. Obtaining the Token from X.



```

#@title Twitter Auth Token

twitter_auth_token = '951d435738f00971895da70ef0cd62496da4a77c' # change this auth token

```

Figure 3. Entering the Token.

After the token was entered and successfully authenticated, the keyword “kaburajadulu” was entered. The search limit was set to 2,000 text-based data entries, and the selected period was from February 10, 2025, to July 10, 2025, as shown in Figure 4. below.

```
# Crawl Data
filename = "kabar.csv"
search_keyword = "kaburajodulu since:2025-02-10 until:2025-07-10 lang:id"
limit = 2000

!pip -y tweet-harvest@2.6.1 -o "(filename)" -s "[search_keyword]" --tab 'LATEST' --l [limit] --token [twitter_auth_token]
```

Figure 4. Data Crawling Process.

After the data crawling process was completed, the prepared file was saved and exported as shown in Figure 5.

created_at	tweet_text	id	lang	source	reply_count	retweet_count	quote_count
2025-06-20 08:00:00	... #kaburajodulu ...	1758901234567890	id	android:google	0	0	0
2025-06-20 08:00:00	... #kaburajodulu ...	1758901234567891	id	android:google	0	0	0
2025-06-20 08:00:00	... #kaburajodulu ...	1758901234567892	id	android:google	0	0	0
2025-06-20 08:00:00	... #kaburajodulu ...	1758901234567893	id	android:google	0	0	0
2025-06-20 08:00:00	... #kaburajodulu ...	1758901234567894	id	android:google	0	0	0
2025-06-20 08:00:00	... #kaburajodulu ...	1758901234567895	id	android:google	0	0	0

Figure 5. Data Export Process.

After the export process, 2,067 data entries were obtained. The collected data were stored in CSV or Microsoft Excel format.

b. Data Cleansing

At this stage, the researcher cleaned the data using RapidMiner before assigning sentiment labels. This process was carried out to ensure that the latest CSV dataset did not contain unnecessary symbols or irrelevant data. As shown in Figure 6, this process removed various disturbances, such as URLs, account mentions, retweets, numerical digits, and special characters.



Figure 6. Cleansing Process Model.

After completing the cleansing process using RapidMiner, the initial number of tweet data in the CSV file was 2,067. After passing through the cleansing process, the tweet data were reduced to 385 entries.

c. Labeling

The next stage was classification or sentiment determination based on each tweet. During the sentiment determination process, the researcher manually assigned sentiment labels.

d. Tokenizing

This process was carried out to transform text from complex sentences into a sequence of tokens or simpler units. The purpose of tokenizing is to prepare text data so that it can be processed more effectively and accurately by sentiment analysis algorithms. The tokenizing process, the tokenizing results for Naive Bayes, and the tokenizing results for Support Vector Machine (SVM) can be seen in Figure 7, Figure 8, and Figure 9 below.



Figure 7. Data Tokenizing Model

Open in Turbo Prep Auto Model Interactive Analysis Filter (385 / 385 examples): all

Row No.	labeling	prediction(la...	confidence_...	confidence_...	aamin	abal	abis	abroad	absolutely	adlan
1	posittf	posittf	1	0	0	0	0	0	0	0
2	posittf	posittf	1	0	0	0	0	0	0	0
3	posittf	posittf	1	0	0	0	0	0	0	0
4	posittf	negattf	0	1	0	0	0	0	0	0
5	posittf	posittf	1	0	0	0	0	0	0	0
6	posittf	posittf	1	0	0	0	0	0	0	0
7	posittf	posittf	1	0	0	0	0	0	0	0
8	posittf	posittf	1	0	0	0	0	0	0	0
9	posittf	posittf	1	0	0	0	0	0	0	0
10	negattf	negattf	0	1	0	0	0	0	0	0
11	posittf	posittf	1	0	0	0	0	0	0	0
12	negattf	negattf	0	1	0	0	0	0	0	0
13	posittf	posittf	1	0	0	0	0	0	0	0
14	negattf	negattf	0	1	0	0	0	0	0	0
15	negattf	negattf	0	1	0	0	0	0	0	0
16	posittf	posittf	1	0	0	0	0	0	0	0
17	posittf	posittf	1	0	0	0	0	0	0	0
18	posittf	posittf	1	0	0	0	0	0	0	0
19	posittf	posittf	1	0	0	0	0	0	0	0
20	posittf	posittf	1	0	0	0	0	0	0	0
21	posittf	posittf	1	0	0	0	0	0	0	0
22	posittf	posittf	1	0	0	0	0	0	0	0

Figure 8. Naive Bayes Tokenizing Model.

Row No.	labeling	prediction(la...	confidence_...	confidence_...	aamin	abal	abis	abroad	absolutely	adlan
1	posittf	posittf	0.749	0.251	0	0	0	0	0	0
2	posittf	posittf	0.749	0.251	0	0	0	0	0	0
3	posittf	posittf	0.750	0.250	0	0	0	0	0	0
4	posittf	posittf	0.749	0.251	0	0	0	0	0	0
5	posittf	posittf	0.749	0.251	0	0	0	0	0	0
6	posittf	posittf	0.749	0.251	0	0	0	0	0	0
7	posittf	posittf	0.749	0.251	0	0	0	0	0	0
8	posittf	posittf	0.749	0.251	0	0	0	0	0	0
9	posittf	posittf	0.750	0.250	0	0	0	0	0	0
10	negattf	posittf	0.612	0.388	0	0	0	0	0	0
11	posittf	posittf	0.749	0.251	0	0	0	0	0	0
12	negattf	negattf	0.402	0.598	0	0	0	0	0	0
13	posittf	posittf	0.749	0.251	0	0	0	0	0	0
14	negattf	negattf	0.267	0.713	0	0	0	0	0	0
15	negattf	posittf	0.676	0.324	0	0	0	0	0	0
16	posittf	posittf	0.748	0.252	0	0	0	0	0	0
17	posittf	posittf	0.749	0.251	0	0	0	0	0	0
18	posittf	posittf	0.749	0.251	0	0	0	0	0	0
19	posittf	posittf	0.749	0.251	0	0	0	0	0	0
20	posittf	posittf	0.749	0.251	0	0	0	0	0	0
21	posittf	posittf	0.749	0.251	0	0	0	0	0	0
22	posittf	posittf	0.749	0.251	0	0	0	0	0	0

Figure 9. Support Vector Machine Tokenizing Model.

Modeling

At this stage, the researcher carried out preprocessing on the dataset to prepare clean and noise-free data. In addition, this process also calculated word weighting for the needs of the subsequent modeling process. After passing through the text preparation stage, the number of tweet data became 385 entries, which were clean data used for the next stage. At this stage, the performance of two classification algorithms was measured simultaneously, namely the Naive Bayes algorithm, as shown in Figure 10, and the Support Vector Machine (SVM) algorithm, as shown in Figure 11.

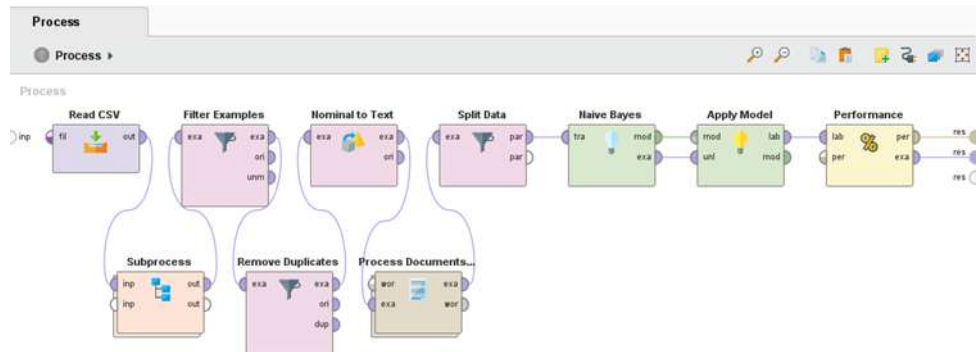


Figure 10. Naive Bayes Data Process Model

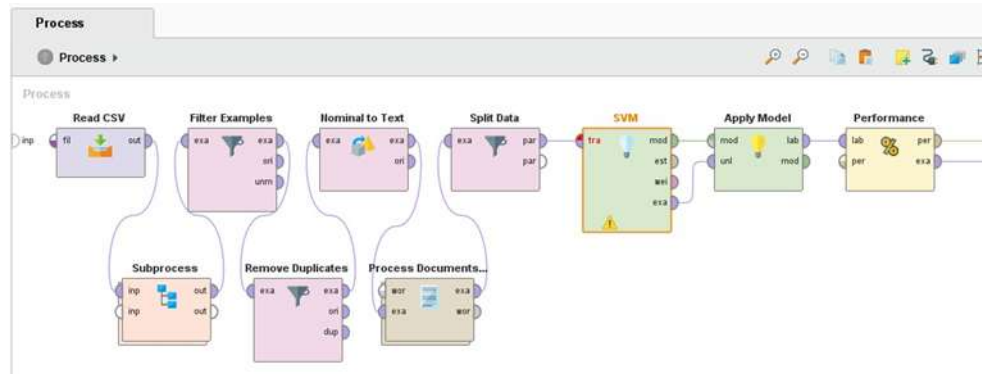


Figure 11. Support Vector Machine Data Process Model

Evaluation and Results

In this study, testing was carried out using the Naive Bayes and Support Vector Machine algorithms, with the following results:

a. Naive Bayes Algorithm

The total number of data collected was 385 entries. The results of the modeling stage using the Naive Bayes algorithm are as follows: data correctly predicted as positive had a precision percentage of 100%, data with negative sentiment had a precision percentage of 56.67%, data with positive sentiment had a recall or specificity percentage of 83.60%, and data with negative sentiment had a recall percentage of 56.67%. The accuracy value generated using the Naive Bayes model was 86.49%, as shown in Figure 12 below.

accuracy: 86.49%

	true positf	true negatf	class precision
pred positf	265	0	100.00%
pred negatf	52	88	56.67%
class recall	83.60%	100.00%	

Figure 12. Naive Bayes Accuracy Results.

The following results were described by the RapidMiner application, as shown in Figure 13.

PerformanceVector

```

PerformanceVector:
accuracy: 86.49%
ConfusionMatrix:
True: positif negatif
positif: 265 0
negatif: 52 88
precision: 86.47% (positive class: negatif)
ConfusionMatrix:
True: positif negatif
positif: 265 0
negatif: 52 88
recall: 100.00% (positive class: negatif)
ConfusionMatrix:
True: positif negatif
positif: 265 0
negatif: 52 88
AUC (optimistic): 1.000 (positive class: negatif)
AUC (optimistic): 0.700 (negative class: negatif)
AUC (pessimistic): 0.700 (positive class: negatif)
    
```

Figure 13. Description of Naive Bayes Accuracy.

From these results, the plot view for the Naive Bayes algorithm is shown in Figure 14. below.

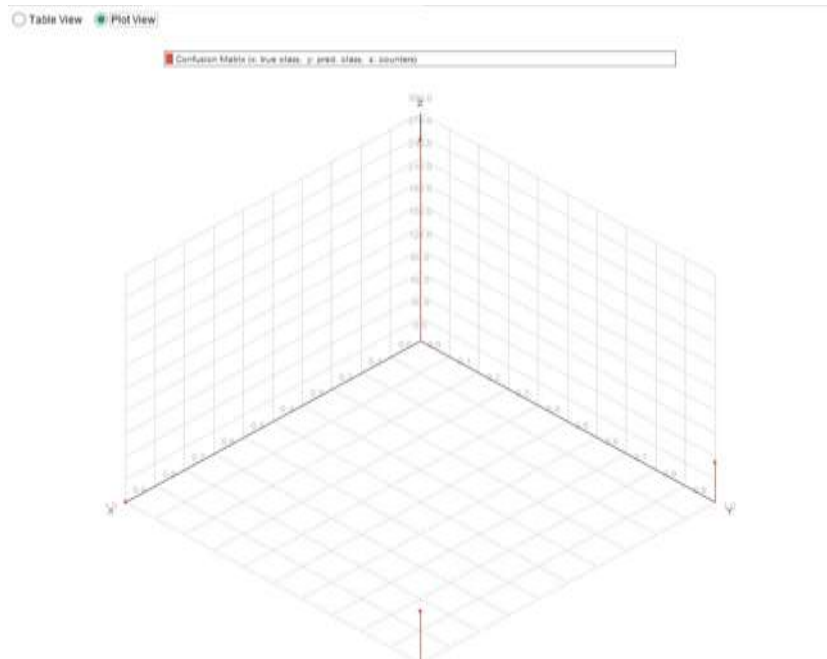


Figure 14. Naive Bayes Plot View.

The following results also display the AUC (Optimistic) graph for the Naive Bayes algorithm, as shown in Figure 15 below.



Figure 15. AUC (Optimistic) Graph of Naive Bayes.

Based on the results above, it can be concluded that the Naive Bayes algorithm is capable of adequately classifying sentiment using Twitter (X) tweet data related to the “kabur aja dulu” trend.

b. Support Vector Machine (SVM) Algorithm

The total dataset collected consisted of 385 data entries. The following are the results of the modeling stage using the Support Vector Machine algorithm. Based on the accuracy calculation results, the data correctly predicted as positive produced a precision percentage of 87.33%, while data with negative sentiment had a precision value of 100%. Positive sentiment had a recall or specificity value of 100%, while negative sentiment had a recall value of 32.36%. The accuracy value generated using the Support Vector Machine model was 88.05%, as shown in Figure 16 below.

accuracy: 88.05%

	true positif	true negatif	class precision
pred. positif	317	46	87.33%
pred. negatif	0	22	100.00%
class recall	100.00%	32.35%	

Figure 16. Support Vector Machine Accuracy Results.

The following results were described by the RapidMiner application, as shown in Figure 17.



Figure 17. Description of Support Vector Machine Accuracy.

From these results, the plot view for the Support Vector Machine algorithm is shown in Figure 18 below.

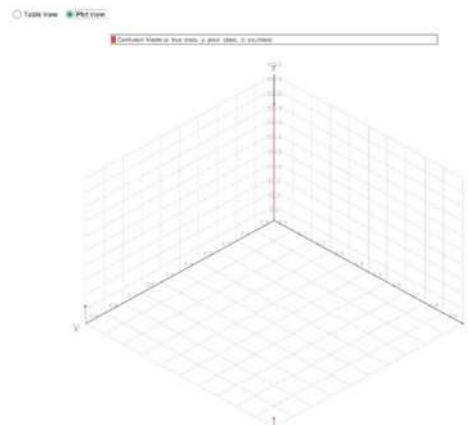


Figure 18. Support Vector Machine Plot View.

The following results also display the AUC (Optimistic) graph for the Support Vector Machine algorithm, as shown in Figure 19 below.



Figure 19. AUC (Optimistic) Graph of Support Vector Machine.

Based on the results above, it can be concluded that the Support Vector Machine algorithm is capable of adequately classifying sentiment using Twitter (X) tweet data related to the “kabur aja dulu” trend.

5. Conclusion and Suggestion

Conclusion

Based on the research and application of the Naïve Bayes and Support Vector Machine algorithms in sentiment analysis of the “Kabur aja dulu” Trend, it can be concluded that: a) According to the research conducted to classify the opinions of Indonesian society about the “Kabur aja dulu” trend, it was found that out of 385 data entries, 317, or 82%, showed positive sentiment and 68 data entries, or 18%, showed negative sentiment. b) The research results show that the majority of Indonesian society responded positively to the “Kabur aja dulu” trend. Based on the classification of the “Kabur aja dulu” trend dataset using the Naive Bayes algorithm model, an accuracy rate of 86.49% was obtained, whereas an accuracy rate of 88.05% was obtained using the Support Vector Machine algorithm. c) Therefore, it can be said that the Support Vector Machine algorithm is better at classifying sentiment data about the “Kabur aja dulu” trend than the Naive Bayes algorithm.

Suggestion

The suggestions provided based on observations and analysis during this research are: a) For future research, it is recommended to further develop this study and connect it with other algorithms to achieve even better accuracy. b) The results of this study are not yet complete and efficient, so it is expected that future research can develop this study using or adding other algorithms to determine objects more quickly and accurately. c) From this research, it is hoped that sentiment labeling can be determined automatically to accelerate the sentiment determination process.

References

- Al-Areef, M. H., & Saputra, K. (2023). Nomor 2. 22, 270–279. Retrieved from <https://ojs.trigunadharma.ac.id/index.php/jis/index>
- Amna, et al. (2023). *Data mining*. Padang: PT Global Eksekutif Teknologi.
- And, I., & Expert, D. (2022). Prediksi penyakit diabetes menggunakan algoritma support vector machine (SVM). Retrieved from <https://e-journal.unper.ac.id/index.php/informatics>
- Asrumi, Suharijadi, D. D. A., & Putri, W. D. (2023). *Analisis sentimen dan penggalian opini*. Jawa Tengah: Eureka Media Aksara.
- Azzahra, S. A., & Wibowo, A. (2020). Analisis sentimen multi-aspek berbasis konversi ikon emosi dengan algoritme Naïve Bayes untuk ulasan wisata kuliner pada web TripAdvisor. 7(4). <https://doi.org/10.25126/jtiik.202071907>
- Baihaqi, K. A., et al. (2023). A comparison support vector machine, logistic regression and naïve bayes for classification sentiment analysis user mobile app. *International Journal of Artificial Intelligence Research*, 7(1), 64. <https://doi.org/10.29099/ijair.v7i1.962>
- Bororing, G. M. G., & Gunawan, F. (2024). Sistem pemesanan dan pembayaran makanan berbasis web terintegrasi dengan application programming interface (API). *Jurnal Informatika dan Bisnis*, 13(1), 37–48. <https://doi.org/10.46806/jib.v13i1.1147>
- Hakim, Z. R., & Sugiyono, S. (2024). Analisa sentimen terhadap Kereta Cepat Jakarta–Bandung menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor. *Jurnal Sains dan Teknologi*, 5(3), 939–945. <https://doi.org/10.55338/saintek.v5i3.1423>
- Jelita, M. (2023). Text mining dengan topic modelling LDA dari pertanyaan gelar wicara literasi Perpustakaan Nasional RI. *Media Pustakawan*, 31(3), 253–265. <https://doi.org/10.37014/medpus.v31i3.5237>
- Kahi, F. R. B., Talakua, A., & Reynaldi, R. (2024). Analisis sentimen masyarakat di Twitter terhadap pemerintahan Anies Baswedan menggunakan metode Naive Bayes classifier. *Jurnal Minfo Polgan*, 13(1), 324–336. <https://doi.org/10.33395/jmp.v13i1.13636>
- Mahendro, I., & Abimanto, D. (2023). *Algoritma Support Vector Machine untuk menganalisa kepuasan mahasiswa terhadap pembelajaran e-learning*. Semarang: CV. Pustaka STIMART-AMNI.
- Maulaya, A. K., & Junadhi. (2022). Analisis sentimen menggunakan support vector machine masyarakat Indonesia di Twitter terkait Bjorka. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), 495–500. <https://doi.org/10.37859/coscitech.v3i3.4358>

- Mulyoto, S. P. M. Si. A. (2024). *Naive Bayes pada Google Colabs*. Jawa Tengah: Eureka Media Aksara.
- Nurtikasari, Y., Alam, S., & Hermanto, T. I. (2022). Analisis sentimen opini masyarakat terhadap film pada platform Twitter menggunakan algoritma Naive Bayes. *INSOLOGI: Jurnal Sains dan Teknologi*, 1(4), 411–423. <https://doi.org/10.55123/insologi.v1i4.770>
- Pamungkas, F. S., & Kharisudin, I. (2021). Analisis sentimen dengan SVM, 4, 628–634. Retrieved from <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Pratama, R. F. P., & Maharani, W. (2025). Comparative analysis of Naive Bayes and SVM for improved emotion classification on social media. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 11–20. <https://doi.org/10.29408/edumatic.v9i1.29087>
- Purnamasari, D., et al. (2023). *Pengantar metode analisis sentimen*. Depok: Gunadarma.
- Rahman, F. N., Lestari, S., & Cipta Karya Informatika Jakarta, S. T. I. (2024). Analisis sentimen masyarakat terhadap pemerintah di era kabinet Joko Widodo berdasarkan social media X menggunakan Naive Bayes dan K-Nearest Neighbor (KNN). *Journal of Information Technology and Computer Science (INTECOMS)*, 7(5).
- Rahman, O. H., Abdillah, G., & Komarudin, A. (2021). Klasifikasi ujaran kebencian pada media sosial Twitter menggunakan support vector machine. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 17–23. <https://doi.org/10.29207/resti.v5i1.2700>
- Safira, A., Masyarakat, A. S., & Hasan, F. N. (2023). Analisis sentimen masyarakat terhadap paylater menggunakan metode Naive Bayes classifier. *Jurnal Sistem Informasi*, 5(1).
- Safitri, T., Umidah, Y., & Maulana, I. (2023). Analisis sentimen pengguna Twitter terhadap BTS menggunakan algoritma support vector machine. Retrieved from <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Sitanggang, A., Umidah, Y., Adam, R. I., & Adam, R. I. (2024). Analisis sentimen masyarakat terhadap program makan siang gratis pada media sosial X menggunakan algoritma Naive Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3). <https://doi.org/10.23960/jitet.v12i3.4902>
- Studi, P., & Informatika, T. (2021). Analisa sentimen masyarakat Indonesia terhadap kinerja Prabowo Subianto sebagai anggota kabinet Indonesia Maju di platform Twitter menggunakan R dengan algoritma Naive Bayes. 2(1), 1–10.
- Taufiqurrahman, H., Anggraeny, F. T., & Al Haromainy, M. M. (2023). Perbandingan algoritma Naive Bayes dan K-Nearest Neighbor pada analisis sentimen ulasan aplikasi MyPertamina.
- Yuniar, M. N. (2023). Klasifikasi kualitas air bersih menggunakan metode Naive Bayes. *Jurnal Sains dan Teknologi*, 5(1), 243–246. <https://doi.org/10.55338/saintek.v5i1.1383>
- Zuhri, K., Adha, N., & Saputri, O. (2020). Analisis sentimen masyarakat terhadap Pilpres 2019 berdasarkan opini dari Twitter menggunakan metode Naive Bayes classifier. Retrieved from <https://journal-computing.org/index.php/journal-cisa/index>
- Zuhri, K., Adha, N., & Saputri, O. (2020). Analisis sentimen masyarakat terhadap Pilpres 2019 berdasarkan opini dari Twitter menggunakan metode Naive Bayes classifier. Retrieved from <https://journal-computing.org/index.php/journal-cisa/index>
- Zulqornain, J. A., & Adikara, P. P. (2021). Analisis sentimen tanggapan masyarakat aplikasi TikTtok menggunakan metode Naive Bayes dan Categorical Proportional Difference (CPD). Retrieved from <http://j-ptiik.ub.ac.id>