Published online on the page : https://ejournal.iainbukittinggi.ac.id/index.php/ijokid/

Knowbase : International Journal of Knowledge in Database

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

Implementation of the Support Vector Machine Method for Sentiment Analysis Using Twitter Data

Widya Wahyuni[1,*]

[1]*Universitas Putra Indonesia "YPTK" Padang, Padang, Indonesia*

Article Information

Article History:
Accepted by the Editor: December 26, 2022
Final Revision: December 26, 2022
Published Online: December 28, 2022

Key Word

Sentiment Analysis
Support Vector Machine
Twitter

Correspondence

E-mail: widyawahyuni040197@gmail.com*

A B S T R A C T

The development of feminism, which is centered on women all over the world who want to be free of male pressure, oppression, and inequality, has continued to the present day. Various public opinions about feminism are now being expressed on various social media platforms. There has been a long debate about feminism's critics and supporters in terms of equalizing women's intellectual and the role of women in making decisions. Not only that, but the desire to end acts of violence and injustice against women is a form of feminism that is often taken for granted, even in the legal realm. The purpose of this study was to examine public sentiment based on opinions shared on social media. Hashtags related to feminism from social media are the main data that will be used to analyze public opinion sentiments about feminism. In this study, 500 tweets were used, and the data was later separated into positive, negative, and neutral opinions before being analyzed using the Support Vector Machine (SVM) method. The results of this study obtained an accuracy of 72%, indicating that the use of SVM to perform sentiment analysis on Twitter data is quite good.

1. Introduction

The global development of the internet can be said to have progressed rapidly over time. This development has also had a significant impact on the dissemination of information that can be said to be important for the recipient, as evidenced by the rise of social media as a means of disseminating information [1]. According to 2019 data, there are approximately 2.77 billion people who use social media [2]. The use of social media is increasing dramatically. This is demonstrated by the ubiquitous presence of social media, followed by the development of mobile devices and applications, as well as people's lives that have been mediated by technology [3]. Users of social media networks can share content such as text, videos, photos, and graphic [4]. Sharing opinions is one of the many cases mentioned by users [5]. Opinions and sentiments expressed frequently on social media are very much in line with daily life, and thus must be analyzed to monitor public opinion and aid in automatic decision-making [6].

The advent of web 2.0 and an increase in active social media users has resulted in people becoming accustomed to sharing their opinions and thoughts in a forum, blog, microblog, or social media platforms such as Twitter, Instagram, and Facebook, among others [7]. In the digital age, social media platforms generate massive amounts of data in a variety of forms. The rate at which data is generated results in a flood of big data that has limitless potential and will continue to grow. Twitter is a social media platform that is

widely used around the world [8]. Twitter has a feature that allows users to share their thoughts and opinions in real time. Furthermore, several Twitter features, such as mentions, replies, and hashtags, allow users to interact with other users [9]. A Twitter post is a brief message. This brief message is known as a 'tweet, and it is limited to 280 characters. All accounts are set to public, allowing anyone to read the tweet [10]. Every day, approximately 500 million tweets are sent, generating massive amounts of data from unstructured text [11].

Feminism is one of the many opinions spread on social media, particularly Twitter. Feminism is not a new concept in addressing women's issues [12]. The term feminism itself has been used since the 1880s [13] and first appeared in France due to hereditary traditions that despised women [14]. Simply put, feminism is the belief that men and women have equal rights or the advocacy of rights so that women have the same rights as men [13]. Furthermore, feminism is regarded as successful in changing institutional and legal practices concerning gender discrimination, employment, and women's participation in public life [15]. According to a feminist perspective, vulnerability may occur in both men and women, implying that gender is not only a social construct but also relational and functional [16]. It's just that feminism, as a concept or belief, is sometimes associated with negativity and controversy [13]. This is because feminism means different things to different people. For example, from a liberal feminist standpoint, women and men have equal rights. Meanwhile, man-hating feminists believe that feminists despise men and seek to exploit them. According to cultural feminism, it is defined as respecting traditional gender roles while also respecting the contributions of both men and women. Based on this, it is clear that one's own understanding of feminism influences whether an opinion has a positive or negative value sentiment [8].

Based on the occurrence and the large amount of data spread on social media, particularly Twitter, manually analyzing public sentiment is insufficient. As a result, a sentiment analysis is required to conduct a sentiment analysis of feminist opinions. Sentiment analysis is used to detect patterns in existing sentiments, and these patterns are then classified to determine whether the opinion contains negative, positive, or negative sentiments. Because of Twitter's rapid growth in popularity among the general public, sentiment analysis is a source of concern. Furthermore, sentiment analysis can be used for business, social, educational, or recreational purposes [17].

To make decisions based on consideration, sentiment analysis is considered appropriate. Data can be used to interpret one's aspects of a particular topic and information that can be used to make predictions in product sales, stock markets, or topic selection [7]. Text mining is used to process data in the form of text. Text mining is one of the Data Mining fields. Text mining seeks to discover something previously unknown but capable of being used in knowledge from unstructured or semi-structured text data. Text mining can also be used to solve problems involving large amounts of data, multiple dimensions, mutable data, and "noise" data. Text mining, as opposed to data mining, is concerned with unstructured or semi-structured data [18]. Knowledge graphs are frequently used to connect extracted information to form new facts or hypotheses [19]. In order to provide solutions, text mining employs and develops a variety of techniques, including data mining, linguistics, natural language processing information retrieval, statistics and mathematics, and visualization. The goal of text mining is to extract useful information from a collection of documents [20].

Machine learning is a popular technique in sentiment analysi [21]. The system is trained using a set of training data, and after training, it will attempt to make an automatic classification or make predictions. The system's efficiency will be evaluated based on the correct predicted value from the test data set [22]. Several methods that are often used are Naïve Bayes, K-NN, Random Forest and Support Vector Machine [21]. The Support Vector Machine method was used in this study (SVM). The most popular machine learning method for sentiment analysis is SVM. Labels can be used to train algorithms, which can then display word separation in classes [23]. SVM is also a type classifier, as defined formally by separating the hyperplane. These steps are required to enter the labeled training data and generate the best hyperplane [1].

Furthermore, the data to be analyzed must be preprocessed, as data obtained from various sources, particularly social media, is typically unstructured. Because the raw format of this data can contain many errors, including spelling and grammatical errors, the text must be cleaned and processed before analysis. The aim of the preprocessing step is not only to improve the words but also to reduce the dimensional data input because many words are unnecessary and must be removed before they affect the text's polarity [24].

Meishita and Iqbal conducted sentiment analysis research in 2022, using the SVM, Nave Bayes, and Logistic methods to examine the sentiments of Tokopedia application users. Using 3125 reviews submitted by Tokopedia users. The web scraping method is used in this study to collect or extract semi-structured data from websites. This study's reviews are divided into two categories: positive and negative sentiments. Positive sentiment is associated with the ease of use of the Tokopedia application, while negative sentiment is associated with application system errors and frequent logging or bugs [25].

To assess sentiment toward MyIndiHome, the SVM and Nave Bayes methods are also used. Sulton Nur Hakim conducted this study in 2021, using data collected from the Google Play application between November 1 and December 15, 2020, with a total of 2,539 participants. According to the study's findings, Indihome's error rate remains relatively high. The SVM method has a higher average total accuracy of 86.54% than the Nave Bayes method [26].

The Nave Bayes and SVM methods were used to analyze public opinion on Covid-19 vaccination. Frizka, Hanif, and Ema conducted this study in 2021 to gauge public opinion on Covid-19, during which time many people expressed their views on social media. This study makes use of 1000 tweets from the year 2020. This study's findings will be used to compare Nave Bayes and SVM. SVM outperforms the Nave Bayes algorithm [27].

This study was carried out because the development of feminism is quite significant in Indonesia, and feminism is not a concept that originates in Indonesia. Because the majority of Indonesia is patriarchal, the mindset of feminism has at least changed the mindset of some women. Because there are pros and cons to feminism, sentiment analysis is used to determine the level of sentiment from society toward feminism. Sentiment from society can be divided into negative or positive sentiments. The findings of this study will demonstrate the accuracy of sentiment analysis performed with the Support Vector Machine algorithm. Table 1 contains several studies that are relevant to this article.

Table 1. State Of The Art

No	Author	Data	Method	Result
1	Herlinawati., Yuliani., Faizah., Gata., & Samudi. (2020)	1.007 record zoom review on playstore	Naïve Bayes and Support Vector Machine	The results of this study show that the Support Vector Machine (SVM) algorithm is 6.85% superior to the Nave Bayes (NB) algorithm in analyzing positive and negative labels.
2	Arsi., & Waluyo. (2021)	1.320 tweets	Support Vector Machine (SVM)	The test results from a study of 1236 tweets (404 positive and 832 negative) using SVM obtained an accuracy of 96.68%, precision of 95.82%, recall of 94.04% and AUC of 0.979.
3	Ulfah., & Anam. (2020)	Comments from 33 news	Support Vector Machine (SVM)	The results of this study can be used as a reference for news portals to apply a filtering system with proven accuracy of 53.88%, recall 49.69%, precision 48.77%, classification error 46.12% and measurement error 49.23%.

4	Sharmin Akter Milu, Md. Ismail Siddiqi Emon, Sabiha Sunjida Ahmed, M. J. Alam, Sheikh Shahparan Mahtab, Jamal Ahmed Bhuiyan, Md. Fahad Mojumder & Mahedy Hasan (2020)	The data set or sample was extracted from a local online store using pursehub.	Support Machine Vector, Naïve Bayes, Logistic Regression	The research results obtained accuracy scores where 85.31%, 88.05%, 88.11%, 81.82% for Naive Bayes, logistic regression, SVM, random forest. Precision scores were obtained 85.56%, 88.54%, 87.59%, 79.14% for Naive Bayes, logistic regression, SVM, random forest, respectively. Recall scores were obtained 84.95%, 88.72%, 88.80%, 85.30% for Naive Bayes, logistic regression, SVM, random forest, respectively.
5	Hassonah, Al-Sayyed, Rodan, Ala'M, Aljarah, & Faris (2020)	6900 tweets from the social network Twitter	Support Vector Machine and feature selection techniques, namely the Relief algorithm and Multi-Verse Optimizer	The findings of this study show that the researcher's method outperforms the Technique and other classifiers, with a yield and reduced number of features of 96.85% of the original feature set.
6	Kiran Kumar, Tejaswi., Vasanthi., Srihitha, Phanindra Kumar (2022)	Vietnamese sentiment dataset with 5000 secret sentences	Naïve Bayes, Maximum Entropy and Support Vector Machine	The results of this study prove that the Maximum Entropy algorithm is better than Naïve Bayes and Support Vector Machine with the best score of 91.36%, with high accuracy.

Based on the data in Table 1, it is possible to conclude that the use of the Support Vector Machine method in conducting sentiment analysis in a study with multiple datasets has a high level of accuracy. As a result, the State of Art has a say in deciding which method to use in this research.

2. Method

This research was conducted in a structured and progressive manner, with critical methods used at each stage to direct it to the correct process. A study is carried out using objective data and information that will be used as a reference point in future research. It is hoped that by utilizing these data, this research will produce high-quality results. The dataset used in this study came from the Twitter platform and included feminism-related queries. The collected data yielded 500 records, which will be divided into training and test data in an 80:20 ratio. Figure 1 displays the research method.

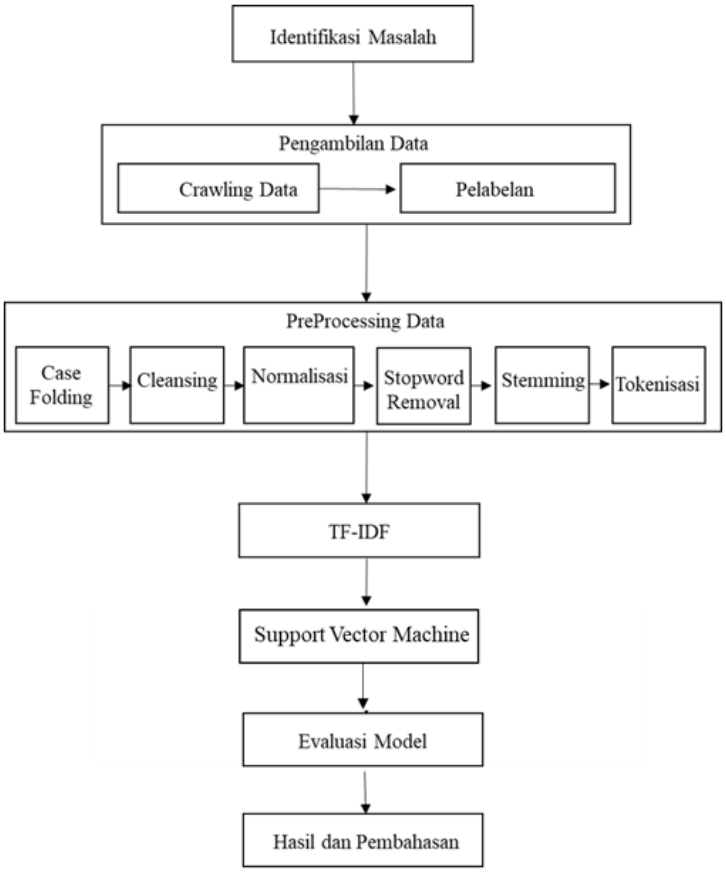

Figure 1. Research Methods

3. Results and Discussion

3.1. Identification of Problems

Problem identification is carried out in an attempt to explain the problem so that it can later be measured. The identification results reveal that the problem is analyzing sentiments about feminism in general and about gender equality using the Support Vector Machine algorithm.

3.2. Data Collection

The data that will be processed later comes from Twitter. This information is derived from the RapidMiner software's crawling results. At this point, the RapidMiner Connection is linked to Search Twitter so that the AccessToken obtained from Twitter can later crawl or collect data in the form of tweets discussing feminism issues. In this study, queries related to feminism use RapidMiner. The results of crawling data that have been previously obtained will be exported to Excel with the .csv extension, because the .csv extension is easier to carry out further processing at the data processing stage. Prior to data processing, the results of data collection will be cleaned up, which includes manually removing duplicate tweets and labeling. The crawling data process is depicted in Figure 2.

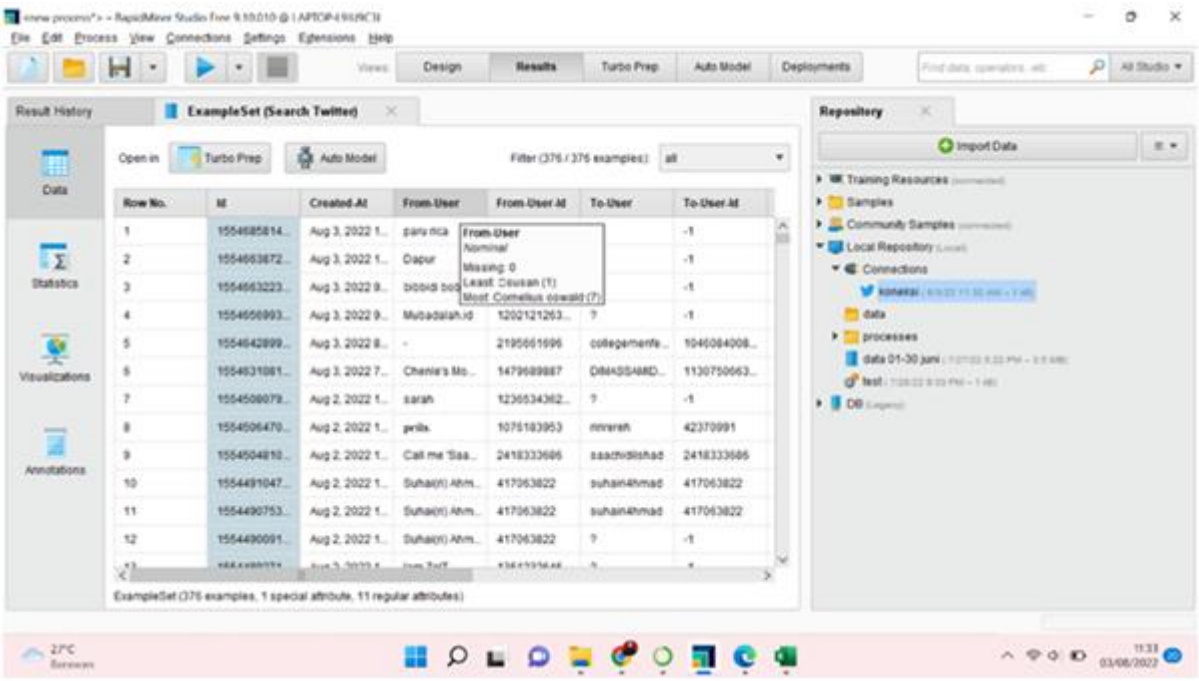

Figure 2. Process of Crawling Data with RapidMiner

3.3. Case Folding

The purpose of this process is to convert all of the letters in the dataset to lowercase or lowercase only. Table 2 displays the results of case folding.

Table 2. Result of Case Folding

Before	After
@literarybase Aku bilang gitu karena beliau tetep dukung novelis perempuan yang feminis abis dan tetap bilang kalau ceritanya keren. Ada di buku apa? Breast and Eggs karya Kawakami Mieko.	@literarybase aku bilang gitu karena beliau tetep dukung novelis perempuan yang feminis abis dan tetap bilang kalau ceritanya keren. ada di buku apa? breast and eggs karya kawakami mieko.
@neveurlandx @infotwitwor_ Cwe feminis pinter tp egonya bikin jijik	@neveurlandx @infotwitwor_ cwe feminis pinter tp egonya bikin jijik

3.4. Cleansing

The available dataset must first be cleansed. The goal is to determine whether or not the dataset contains any missing values or damaged data. If the dataset contains missing values, the data must be cleaned or go through the cleansing stage first. Cleaning was not performed in this study because the dataset's results did not contain any missing values. The data description process shown in Figures 3 and 4 shows the results of cleansing.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 599 entries, 0 to 598
Data columns (total 5 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Tanggal  599 non-null    object
 1   User     599 non-null    object
 2   Tweet    599 non-null    object
 3   Id       599 non-null    float64
 4   Label    599 non-null    object
dtypes: float64(1), object(4)
memory usage: 23.5+ KB
```

Figure 3. Data Description Process

	Tanggal	User	Tweet	Id	Label
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
594	False	False	False	False	False
595	False	False	False	False	False
596	False	False	False	False	False
597	False	False	False	False	False
598	False	False	False	False	False

599 rows × 5 columns

Figure 4. Missing Value Detection Process

3.5. Normalizationi

Normalization is the process of removing a variety of punctuation marks and numbers, symbols, URLs, and usernames from data so that it can be more structured for further processing. Table 3 shows the results of the normalization process on the training and test data.

Table 3. Data Normalization Process

Before	After
@literarybase aku bilang gitu karena beliau tetep dukung novelis perempuan yang feminis abis dan tetap bilang kalau ceritanya keren. ada di buku apa? breast and eggs karya kawakami mieko.	aku bilang gitu karena beliau tetep dukung novelis perempuan yang feminis abis dan tetap bilang kalau ceritanya keren ada di buku apa breast and eggs karya kawakami mieko
@neveurlandx @infotwitwor_ cwe feminis pinter tp egonya bikin jijik	cwe feminis pinter tp egonya bikin jijik

3.6. Stopword Removal

At this stage, non-essential phrases are removed based on the phrases contained in the stopword. If a word matches the word contained in the stopword, the word will be removed from the document because it is deemed unimportant. Table 4 shows the stopword removal process on training and test data.

Table 4. Data Normalization Process

Before	After
aku bilang gitu karena beliau tetep dukung novelis perempuan yang feminis abis dan tetap bilang kalau ceritanya keren ada di buku apa breast and eggs karya kawakami mieko	bilang gitu beliau tetep dukung novelis perempuan feminis abis bilang ceritanya keren. buku apa breast and eggs karya kawakami mieko
cwe feminis pinter tp egonya bikin jijik	cwe feminis pinter tp egonya bikin jijik

3.7. Stemming

Any affixes in the word will be removed during the stemming process. Stemming is done using a rules-based approach. Table 5 displays the stemming process results.

Table 5. Stemming Process

Before	After
bilang gitu beliau tetep dukung novelis perempuan feminis abis bilang ceritanya keren. buku apa breast and eggs karya kawakami mieko	bilang gitu beliau tetep dukung nove perempuan femini abi bilang cerita keren buku apa breast and egg karya kawakami mieko
cwe feminis pinter tp egonya bikin jijik	cwe femini pinter tp ego bikin jijik

3.8. Tokenization

At this point, the process of splitting text documents into sentences based on training and test data will be carried out word by word. Table 6 depicts the tokenization process.

Table 6. Stemming Process

Before	After
cwe femini pinter tp ego bikin jijik	'cwe', 'femini', 'pinter', 'tp', 'ego', 'bikin', 'jijik'
cwe femini pinter tp ego bikin jijik	'cwe', 'femini', 'pinter', 'tp', 'ego', 'bikin', 'jijik'

3.9. TF-IDF

The next preprocessing step is to weight each word (term) in each document, also known as TF-IDF, by calculating tf, Wtft, d, df, and idf.

3.9.1. Calculation of tf and df values

Table 7 shows the manual calculation process from TF-IDF on TF and DF.

Table 7. Calculation of tf and df values

Term	TF								DF
	Training Data					Testing Data			
	D1	D2	D3	D4	D5	D1	D2	D3	
Sejalan	0	0	0	0	0	1	0	0	1
Perhatian	0	0	0	0	0	1	0	0	1
Pengembangan	0	0	0	0	0	1	0	0	1
Perempuan	0	0	1	0	1	1	0	1	4
Babi	0	0	0	0	0	0	1	0	1

3.9.2. Calculation of wtft, d and idf values

The next step is to calculate the value of wtft,d based on the values of tf and df in Table 7, then compute the value of idf in each term. Table 8 will show the result of the calculation along with an example of manual calculation as follows:

$$Wtf_{t,d} = \begin{cases} 1+log_{10} tf_{t,d}, if\ tf_{t,d}>0 \\ 0, \quad otherwise \end{cases} \text{.....(1)}$$

Table 8. Calculation of Wtft, d Values on Training and Test Data

Term	Wtft,d								IDF
	Training Data					Testing Data			
	D1	D2	D3	D4	D5	U1	U2	U3	
Sejalan	0	0	0	0	0	1	0	0	0,9030
Perhatian	0	0	0	0	0	1	0	0	0,9030
Pengembangan	0	0	0	0	0	1	0	0	0,9030
Perempuan	0	0	1	0	1	1	0	1	0,310
Babi	0	0	0	0	0	0	1	0	0,9030
Cewek	1	0	0	1	0	0	0	1.3	0,424
Women	0	0	0	0	0	1	0	0	0,9030

3.9.3. Calculation of TF-IDF

The following stage is to complete the word weighting on the training and test data by calculating the TF-IDF value. The calculation results will be entered into Table 9.

Table 9. Calculation of Wtft, d Values on Training and Test Data

Term	Wtft, d								IDF
	Training Data					Testing Data			
	D1	D2	D3	D4	D5	U1	U2	U3	
Sejalan	0	0	0	0	0	0,9030	0	0	0,9030
Perhatian	0	0	0	0	0	0,9030	0	0	0,9030
Pengembangan	0	0	0	0	0	0,9030	0	0	0,9030
Perempuan	0	0	0,310	0	0,310	0,310	0	0,310	0,310
Babi	0	0	0	0	0	0	0,9030	0	0,9030
Cewek	0,424	0	0	0,424	0	0	0	0,551	0,424
Women	0	0	0	0	0	0,9030	0	0	0,9030

3.10. Support Vector Machine

At this point, the Support Vector Machine (SVM) process will begin with the results of the previously obtained TF-IDF. Using SVM sequential training, SVM looks for boundaries between data classes (hyperplanes). Figure 5 depicts the stages of SVM sequential training.

1. Input data hasil pembobotan TF-IDF
2. Menghitung kernel *polynomial*
3. Menghitung *matrix hessian*
4. Menghitung *sequential training*
5. Menghitung nilai bias
6. Menghitung nilai data *testing*
7. Hasil *support vector machine*

Figure 5. Sequential Training Stages

3.10.1. Calculating kernel polynomials

This process employs a polynomial of degree up to d, with a simple polynomial kernel to implement. The above formula employs a polynomial kernel with degree = 2, namely: K(xi,xj) = (xi.xj + c) d. The polynomial at x 1 calculation denotes the computation of the first data document. Table 10 displays the results of the polynomial kernel calculations.

Table 10. Kernel Calculation Results

	D1	D2	D3	D4	D5
D1	34,78457812	1	1,39187141	1,848381641	1
D2	1	45,43005892	1	1,065683365	1,065683365
D3	1,39187141	1	8,44887001	1,39187141	1,20143521
D4	1,848381641	1	1,39187141	14,4839538	1
D5	1	1	1,20143521	1	7,436081779

3.10.2. Calculating matrix hessian

The first stage is to initialize I = 0 and perform hessian matrix calculations using formula 2, and the results of the hessian matrix calculations are shown in Table 11.

$$D_{i,j} = y_i, y_j \ (K(x_i, x_j) + \lambda^2) \quad \text{.....(2)}$$

Table 11. Results of Hessian Matrix Calculations on Sample Data

	D1	D2	D3	D4	D5
D1	35,03457812	-1,25	1,64187141	2,098381641	-1,25
D2	-1,25	45,68005892	1,25	-1,315683365	1,315683365
D3	-1,64187141	1,25	8,69887001	-1,64187141	1,45143521
D4	2,098381641	-1,25	-1,64187141	14,7339538	-1,25
D5	-1,25	1,25	1,45143521	-1,25	7,686081779

3.10.3. Calculating sequential training

At this stage, the sequential training process aims to process training data from the support vector machine. The sequential training procedure is divided into several stages, which include initializing the variables, calculating the epsilon delta alpha and alpha values, and calculating the bias values until the data testing values are obtained.

Initialization of variables, namely αi, γ, C, and ε. At this stage several variables are initialized for processing. These variables include:

c = 1

λ = 0,5

γ = 0,001

C = 1

ε = 0,0001

Maximum iteration (imax) = 3

Calculate the value of Ei, δαi, and α. The first stage is to calculate the error rate value. The calculation of the search for the error rate value is presented in the following search:

$$Ei = \sum_{j=1}^{i} a_j D_{ij} \quad \text{.....(3)}$$

When the value of the error rate in the first iteration is obtained, the value of the first iteration error rate in the first data will be used to calculate the delta alpha value (δαi). The first iteration alpha delta value on the first data is calculated as follows:

$$\delta\alpha i = \min\left(\max[\,\gamma(1-i), \alpha_i], C - \alpha_i\right. \quad \text{.....(4)}$$

The alpha delta value is used to calculate a new alpha value that will be used to calculate the error rate value for the next iteration. The formula for calculating the new alpha is as follows:

$$\alpha i = \alpha i + \delta\alpha \quad \text{.....(5)}$$

The above calculation also holds true for subsequent iterations on different data points. The calculation result or the value search result Ei, δαi, and α then presented in Table 12, 13, 14.

Table 12. Calculation Results of Ei on Sample Data

	D1	D2	D3	D4	D5
E1	0	0	0	0	0
E2	0,032991088	0,045680059	0,011400305	0,012624781	0,0079532
E3	0,064893765	0,08927345	0,022670643	0,025090176	0,015843147

Table 13. Calculation Results of δαi on Sample Data

	D1	D2	D3	D4	D5
*δα*1	0,001	0,001	0,001	0,001	0,001
*δα*2	0,000967009	0,00095432	0,0009886	0,000987375	0,000992047
*δα*3	0,000935106	0,000910727	0,000977329	0,00097491	0,000984157

Table 14. αi Calculation Results on Sample Data

	D1	D2	D3	D4	D5
$\delta\alpha 1$	0,001	0,001	0,001	0,001	0,001
$\delta\alpha 2$	0,000967009	0,00095432	0,0009886	0,000987375	0,000992047
$\delta\alpha 3$	0,000935106	0,000910727	0,000977329	0,00097491	0,000984157

Calculating the value of the bias. The bias value is calculated by first determining x+ and x-. This is accomplished by inspecting the kernel data with the highest category's new alpha value (iteration 3). The highest value of x+ in this calculation is found in the fifth data, and the highest value of x- is found in the sixth data. Table 15 shows the procedure for calculating the bias value.

$$b = -\frac{1}{2}\left(\sum_{i=0}^{N} \alpha_i\, y_i (K(x_i, x^-) + \sum_{i=0}^{N} \alpha_i\, y_i (K(x_i, x^+)\right) \quad \text{.....(6)}$$

Table 15. Support Vector (SV) Calculation Results on Sample Data

	D5(+)	D6(-)	a.y(k+)	ay(k-)
D1	1	1,848381641	-0,002908247	-0,00537555
D2	1	1	0,002866311	0,00286631
D3	1,20143521	1,39187141	0,003564773	0,00412982
D4	1	14,4839538	-0,002963448	-0,04292244
D5	7,436081779	1	0,022132751	0,0029764
	Total		0,022692141	-0,03832546

The sentiment results will be seen after the value of the testing data is obtained. Based on the search results from sign(h(x)), it can be concluded that test data 1 is positive, and thus test data 1 and test data 2 have a positive sentiment class. Because test data 3 has a negative value, it is classified as negative sentiment. Table 16 is the Support Vector (SV) calculation results on sample data.

Table 16. Table of Support Vector (SV) Calculation Results on Sample Data

$\alpha_i y_{i(K(x_i x_i)}$	U1	U2	U3
D1	-0,002866311	-0,002866311	-0,01203431
D2	0,002866311	0,002866311	0,002866311
D3	0,003443687	0,002866311	0,003443687
D4	-0,002866311	-0,002866311	-0,004362033
D5	0,003443687	0,003443687	0,003443687
Total	0,004021063	0,003443687	-0,00664266
Total+bias	0,011837724	0,003443687	-0,00664266

3.11. Evaluation

This stage is carried out to see the level of accuracy using the support vector machine and naïve Bayes methods. Before calculating the level of accuracy, a confusion matrix table is first made from the results of the methods that have been applied before. Table 17 shows the predictions of the confusion matrix support vector machine. In Table 18 there are SVM Performance Evaluation.

Table 17. Results of the Confusion Matrix Support Vector Machine

		Predicted	
		Positive	**Negative**
Actual	**Positive**	2	0
	Negative	1	0

Table 18. Results of SVM Performance Evaluation

	Precision	Recall	Fi-score
Positive	50%	100%	66%
Negative	100%	50%	66%
Average	75%	75%	66%
Accuracy		72%	

According to Table 18, the accuracy of the support vector machine classification is 72%, indicating that the level of the support vector machine method in conducting sentiment analysis on the topics discussed is categorized as quite good.

4. Conclusion

A sentiment analysis process is required to forecast a phenomenon or business needs. In this study, the results of sentiment analysis on Twitter regarding public opinion on feminism will provide an overview to the community of how much the development of feminism is currently sticking out. Furthermore, people with an interest in data mining will discover that sentiment analysis can be used to research social phenomena in addition to business purposes. The performance evaluation results for the algorithm used show that SVM has an accuracy of 80%, which is considered good. As a result, the sentiment analysis on Twitter using the SVM method with the topic of feminism on the 500 data points collected was deemed satisfactory.

References

[1] P. Sharma and A. K. Sharma, "Experimental investigation of automated system for twitter sentimen analysis to predict the public emotions using machine learning algorithms," Mater. Today Proc., no. xxxx, 2020.
[2] G. R. A. & F. L. Gustavo Nogueira de Sousa, "Social Network Advertising Classification Based on Content Categories," in Lecture Notes in Business Information Processing, vol. 373 LNBIP, 2019, pp. 169–180.
[3] E. Marttila, A. Koivula, and P. Räsänen, "Does excessive media sosial use decrease subjective well-being? A longitudinal analysis of the relationship between problematic use, loneliness and life satisfaction," Telemat. Informatics, vol. 59, no. August 2020, pp. 1–11, 2021.
[4] T. R. Soomro and M. Hussain, "Media sosial-Related Cybercrimes and Techniques for Their Prevention," Appl. Comput. Syst., vol. 24, no. 1, pp. 9–17, 2019.
[5] Y. Gorodnichenko, T. Pham, and O. Talavera, "Media sosial, sentimen and public opinions: Evidence from #Brexit and #USElection," Eur. Econ. Rev., vol. 136, p. 103772, 2021.
[6] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "A framework for big data analytics in commercial social networks: A case study on sentimen analysis and fake review detection for marketing decision-making," Ind. Mark. Manag., no. July, pp. 1–15, 2019.
[7] A. R. Pathak, "Adaptive model for sentimen analysis of media sosial data using deep learning," ICICCT 2019 – Syst. Reliab. Qual. Control. Safety, Maint. Manag., pp. 416–423, 2020.
[8] S. M. Ogletree, P. Diaz, and V. Padilla, "What is Feminism? College Students' Definitions and Correlates," Curr. Psychol., vol. 38, no. 6, pp. 1576–1589, 2019.
[9] V. Gupta and R. Hewett, "Real-time tweet analytics using hybrid hashtags on twitter big data streams," Inf., vol. 11, no. 7, 2020.
[10] D. Antonakaki, P. Fragopoulou, and S. Ioannidis, "A survey of Twitter research: Data model, graph structure, sentimen analysis and attacks," Expert Syst. Appl., vol. 164, no. July 2020, p. 114006, 2021.
[11] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentimen analysis," Procedia Comput. Sci., vol. 152, pp. 341–348, 2019.
[12] H. Hidayati, "Kesetaraan Gender dalam Pelestarian Lingkungan Perspektif Al-Qur'an," TAFAKKUR J. Ilmu Al-Qur'an dan …, pp. 186–199, 2021.
[13] K. McLaughlin and S. N. Aikman, "That is What a Feminist Looks Like: Identification and Exploration of the Factors Underlying the Concept of Feminism and Predicting the Endorsement of Traditional Gender Roles," Gender Issues, vol. 37, no. 2, pp. 91–124, 2020.
[14] M. Tamboukou, "Women Workers' Education," p. 961, 2020.
[15] K. Alexeyeff, "Cinderella of the south seas? Virtuous victims, empowerment and other fables of development feminism," Womens. Stud. Int. Forum, vol. 80, no. April, p. 102368, 2020.
[16] P. Yadav, N. Saville, A. Arjyal, S. Baral, P. Kostkova, and M. Fordham, "A feminist vision for transformative change to disaster risk reduction policies and practices," Int. J. Disaster Risk Reduct., vol. 54, no. February 2020, p. 102026, 2021.
[17] A. Kumar and G. Garg, "Systematic literature review on context-based sentimen analysis in social multimedia," Multimed. Tools Appl., vol. 79, no. 21–22, pp. 15349–15380, 2020.
[18] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentimen analysis of media sosial Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," Procedia Comput. Sci., vol. 161, pp. 765–772, 2019.
[19] S. Zheng, S. Dharssi, M. Wu, J. Li, and Z. Lu, "Text mining for drug discovery," Methods Mol. Biol., vol. 1939, pp. 231–252, 2019.
[20] D. Ariyanti and K. Iswardani, "Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes," J. IKRA-ITH Inform., vol. 4, no. 3, pp. 125–132, 2020.
[21] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," J. Teknol. Inf. dan Ilmu Komput., vol. 8, no. 1, p. 147, 2021.
[22] R. Singh and R. Singh, "Applications of sentimen analysis and machine learning techniques in disease outbreak prediction – A review," Mater. Today Proc., no. xxxx, 2021.
[23] T. Cura, "Use of support vector machines with a parallel local search algorithm for data classification and feature selection," Expert Syst. Appl., vol. 145, p. 113133, 2020.
[24] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentimen analysis: Approaches, challenges and trends," Knowledge-Based Syst., vol. 226, p. 107134, 2021.

[25] M. I. Putri and I. Kharisudin, "Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace Tokopedia," Prism. Pros. Semin. Nas. Mat., vol. 5, pp. 759–766, 2022.

[26] M. D. Prasetio, R. Y. Xavier, H. Rachmat, W. Wiyono, and D. S. E. Atmaja, "Sentimen analysis on myindihome user reviews using support vector machine and naïve bayes classifier method," Int. J. Ind. Optim., vol. 2, no. 2, p. 141, 2021.

[27] F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," J. Komtika (Komputasi dan Inform., vol. 5, no. 1, pp. 19–25, 2021.