Email: admin@antispublisher.com

e-ISSN: 3032-1077 JAIDE, Vol. 2, No. 5, May 2025 Page 359-371 © 2025 JAIDE:

Journal of Artificial Intelligence and Digital Economy

# Design of Hoax Filtering Plugin in Twitter Application Using SVM Method

## Agung Mulya Rasyiidi<sup>1</sup>, Cindy Taurusta<sup>2</sup>

1,2 Muhammadiyah University of Sidoarjo, Indinesia



DOI: https://doi.org/10.61796/jaide.v2i5.1528

#### **Sections Info**

#### Article history:

Submitted: February 25, 2025 Final Revised: March 11, 2025 Accepted: April 20, 2025 Published: May 30, 2025

Keywords: SVM TF-IDF Twitter

#### ABSTRACT

Objective: This study aims to analyze and classify public sentiment toward prominent individuals on social media, focusing on distinguishing between acceptable opinion and hate speech. Method: A dataset comprising 700 tweets about public figures was collected and processed using the Support Vector Machine (SVM) classification algorithm, with Term Frequency-Inverse Document Frequency (TF-IDF) employed for feature weighting. The approach was chosen to evaluate the effectiveness of SVM in handling textual sentiment data. Results: The implementation of the SVM model demonstrated reliable performance in categorizing tweets according to sentiment polarity and identifying instances of hate speech. The TF-IDF weighting method enhanced the accuracy of feature representation within the model. Novelty: This study contributes to computational linguistics and social media analytics by providing an empirical evaluation of SVM's capability in sentiment classification related to public figures, offering potential applications for automated monitoring of online discourse and hate speech detection.

#### INTRODUCTION

Freedom of opinion, also known as freedom of speech, is often misunderstood [1]. The right to speak freely in discourse, stated in the 1945 Constitution of the Republic of Indonesia (UUD RI 1945), is the opportunity to offer points of view and reflections either orally, recorded in print, or verbally [2]. However, hate speech can be considered hate speech if it has caused hostility, attacked, or spread against individuals or groups, either directly or indirectly [3].

Misclassification due to this ambiguity is feared to limit freedom of speech and opinion [4]. On the other hand, allowing people to express themselves freely will trigger hatred, discrimination, and violence against certain groups [5]. Hate speech can be spread on social media [6]. Twitter is one of the most influential social media platforms in this century [7]. Currently, there are 106 million Twitter user accounts, and this number continues to increase [8]. 55 million tweets are sent every day [9].

Indonesia is in eighth place as the country with the most tweets [10]. A study on Twitter social media sentiment analysis tested the level of customer satisfaction of mobile telecommunications service providers [11]. The accuracy of the SVM method used to assess tweets written in Indonesian was 73.43% in several previous studies [12], [15].

The goal of this website is to use Support Vector Machines to sort tweets on Twitter into those that insult public figures and those that do not. Therefore, this final project will test the development of an opinion investigation application on Twitter, which uses Support Vector Machine (SVM) techniques to characterize tweets containing insulting discourse and hoax news.

#### **RESEARCH METHOD**

## A. Research Stages

The first step in this research was collecting data from Twitter. The next step was to collect data on public comments about fake news. The obtained data then underwent preprocessing. The data was manually labeled after going through the preprocessing process. Afterward, a Support Vector Machine (SVM) algorithm was used to classify the data. Positive, negative, and neutral were the three categories used to categorize the obtained data.

## B. SVM Modeling Scheme

The data classification process using the Multiclass SVM method can be seen in Figure 1.

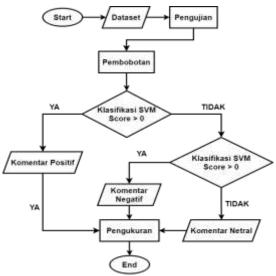


Figure 1. SVM Classification Process

## C. Support Vector Machine

The goal of the Support Vector Machine (SVM) method is to select the most effective function for separating two categories of objects. The answer for the two-class setting can use the following conditions:

$$\begin{aligned} & \min_{wij} \frac{1}{2} \left( w^{ij} \right)^T w^{ij} + c \sum_r t_r^{ij} \\ & \text{subject to:} \\ & w^{ij} x_r + b^{ij} \geq t_r^{ij}, jika \ y_r = i \\ & w^{ij} x_r + b^{ij} \leq t_r^{ij}, jika \ y_r \neq i \\ & t_r^{ij} \geq 0, \end{aligned}$$

SVM was originally used to divide data into two classes. Support Vector Machine can be developed to support multiclass classification. The separator can be a line in two dimensions, a plane in three dimensions, or a hyperplane in more than three dimensions. SVM was originally used to divide data into two classes. SVM can be developed to support multiclass classification. The two-class case requires different strategies for multiclass SVM. There are several Multiclass SVM techniques, one of which is the One-Against-One Multiclass SVM strategy. By creating a number of binary SVM models that will later be compared with each other in the One-Again-One method. There must be a number of k (k-1) / 2 binary SVM models built to classify data into k classes.

# D. Kernel Engineering

Information mining or AI procedures are widely developed with the assumption of linearity, so the resulting algorithms are more suitable for linear cases. Frequent cases are usually nonlinear. To overcome this indirect nature, a portioning strategy can be used. A higher-dimensional feature space F is mapped to the input data space x using a kernel method. In SVM, the kernel function typically used is:

Linear kernel: x<sup>T</sup>x

Kernel polynomial:  $(x^Tx_i + 1)^p$ 

RBF Kernel:  $\left(-\frac{\pi}{2}\|\mathbf{x}-\mathbf{x}^{\dagger}\|_{2}\right)$ 

## E. Confusion Matrix

The characterization assessment in this study uses a Confusion Matrix strategy. The prediction matrix will be compared with the original classes, which contain real information and predicted classification values, in this method. Once the framework effectively characterizes the tweets, the action is expected to determine how precise or accurate the groupings have been created by the framework. The situation for the leading-order test using the Confusion Matrix technique should be seen in Figure 2.

		Actual Values	F
		1 (Positive)	0(Negative)
Predicted Value	1 (Positive)	TP (True Positive)	FP (False Positive)
	0 (Negative)	FN (False Negative)	TN (True Negative)

**Figure 2.** *Confusion Matrix* 

## **RESULTS AND DISCUSSION**

#### A. Data collection

The data in this study is secondary. Indonesian-language text taken from Twitter using the Python programming language is included in the dataset. Access keys and tokens are required as proof of authentication by Twitter developers to use the Twitter API. All data collected is public opinion regarding government programs. The information collection was conducted to obtain 1,400 pieces of

information.

## B. Data Preprocessing

Preprocessing is the stage where information obtained by crawling over 1,400 pieces of data is first processed. Cleansing, case folding, tokenizing, filtering, and stemming are the methods used for data preprocessing.

# C. Cleansing

Cleansing is a phase that aims to remove characters or images from the join URL (http://situs.com), username or references to it (@username), hashtags (#), retweets, and emojis. Table 1 displays the data results after the Cleansing process.

**Table 1.** Data Results After Cleansing Process

No	Raw Data	
		Cleansing Results
1	RT @Kompascom: Kaesang:	Kaesang, the father of
	My father could deceive the	simplicity, deceives the
	Indonesian people with his	Indonesian people
	simplicity. Why can't I?	
2	RT @adi098: The Corruption	KPK prohibits Brimob
	Eradication Committee	personnel from bringing
	(KPK) is prohibited from	weapons into the DPR
	bringing armed Brimob	building
	personnel into the DPR	
	building.	
3	RT @agege: Lawyer Hotman	Hotman Paris Hutapea's
	Paris Hutapea Becomes	event as Habib Rizieq
	Legal Counsel for Habib	Shihab's legal
	Rizieq Shihab	representative
4	RT @slavko: They Said They	He said he wanted to
	Wanted to Send the UI	send the chairman of
	Student Executive Board	BEM UI Asmat, but
	Chair to Asmat, But Foreign	foreign journalists
	Journalists Who Wanted to	wanted to cover it, just
	Cover the Event Were	throw them out.
	Expelled	
5	RT @Metrotv: Ministry of	Ministry of Home Affairs
	Home Affairs Revokes	revokes regional
	Regional Regulation on	regulation on alcoholic
	Alcoholic Beverages	beverages

## D. Folding Case

The process of converting all uppercase letters in a document to their standard form (lowercase) is called case folding. Other characters are treated as boundaries or separators. Table 2 displays the results of the case folding procedure.

Table 2. Data Results After Case Folding Process

Nic	Classics Possilis Felding Process			
No	Cleaning Results	Folding Case Results		
1	Kaesang: My father could	Kaesang, the father of		
	deceive the Indonesian	simplicity, deceives the		
	people with his simplicity.	Indonesian people		
	Why can't I?			
2	The Corruption Eradication	KPK prohibits Brimob		
	Committee (KPK) is	personnel from bringing		
	prohibited from bringing	weapons into the DPR		
	armed Brimob personnel	building		
	into the DPR building.			
3	Lawyer Hotman Paris	Hotman Paris Hutapea's		
	Hutapea Becomes Legal	event as Habib Rizieq		
	Counsel for Habib Rizieq	Shihab's legal		
	Shihab	representative		
4	They said they wanted to	He said he wanted to		
	send the UI Student	send the chairman of		
	Executive Board (BEM UI)	BEM UI Asmat, but		
	chairman to Asmat, but	foreign journalists		
	foreign journalists who	wanted to cover it, just		
	wanted to cover the event	throw them out.		
	were expelled.			
5	Ministry of Home Affairs	Ministry of Home Affairs		
	Revokes Regional	revokes regional		
	Regulation on Alcoholic	regulation on alcoholic		
	Beverages	beverages		

# E. Tokenizing

The next step after case folding is tokenization. To produce sentences or words that can stand alone, the data is processed without punctuation. Table 3 displays the data results after the tokenization procedure.

Table 3. Data Results After Tokenizing Process

No	Cleaning Results	Folding Case Results
1	Kaesang: My father could	Kaesang, the father of
	deceive the Indonesian	simplicity, deceives the
	people with his simplicity.	Indonesian people
	Why can't I?	
2	The Corruption Eradication	KPK prohibits Brimob
	Committee (KPK) is	personnel from bringing
	prohibited from bringing	

No	Cleaning Results	Folding Case Results
	armed Brimob personnel	weapons into the DPR
	into the DPR building.	building
3	Lawyer Hotman Paris	Hotman Paris Hutapea's
	Hutapea Becomes Legal	event as Habib Rizieq
	Counsel for Habib Rizieq	Shihab's legal
	Shihab	representative
4	They said they wanted to	He said he wanted to
	send the UI Student	send the chairman of
	Executive Board (BEM UI)	BEM UI Asmat, but
	chairman to Asmat, but	foreign journalists
	foreign journalists who	wanted to cover it, just
	wanted to cover the event	throw them out.
	were expelled.	
5	Ministry of Home Affairs	Ministry of Home Affairs
	Revokes Regional	revokes regional
	Regulation on Alcoholic	regulation on alcoholic
	Beverages	beverages

# F. Filtering

Using a stoplist algorithm (which removes less important words) or a wordlist algorithm (which stores important words), filtering is a stage that aims to eliminate common words that appear frequently in large numbers and have no significant meaning. The data results after going through the filtering process can be seen in Table 4.

**Table 4.** Data Results After Filtering Process

No	Tokenizing Results	Filtering Results
1	Kaesang: My father could	Kaesang, the father of
	deceive the Indonesian	simplicity, deceives the
	people with his simplicity.	Indonesian people
	Why can't I?	
2	The Corruption Eradication	KPK prohibits Brimob
	Committee (KPK) is	personnel from bringing
	prohibited from bringing	weapons into the DPR
	armed Brimob personnel	building
	into the DPR building.	
3	Lawyer Hotman Paris	Hotman Paris Hutapea's
	Hutapea Becomes Legal	event as Habib Rizieq
	Counsel for Habib Rizieq	Shihab's legal
	Shihab	representative

No	<b>Tokenizing Results</b>	Filtering Results
4	They said they wanted to	He said he wanted to
	send the UI Student	send the chairman of
	Executive Board (BEM UI)	BEM UI Asmat, but
	chairman to Asmat, but	foreign journalists
	foreign journalists who	wanted to cover it, just
	wanted to cover the event	throw them out.
	were expelled.	
5	Ministry of Home Affairs	Ministry of Home Affairs
	Revokes Regional	revokes regional
	Regulation on Alcoholic	regulation on alcoholic
	Beverages	beverages

# G. Stemming

Stemming is the process by which each word is transformed from a compound word to its root word. The Nazief and Adriani algorithm from the Sastrawi library is used for stemming.

Table 5. Data Results After Stemming Process

No	Filtering Results	Stemming Results
1	Kaesang: My father could	Kaesang, the father of
	deceive the Indonesian	simplicity, deceives the
	people with his simplicity.	Indonesian people
	Why can't I?	
2	The Corruption Eradication	KPK prohibits Brimob
	Committee (KPK) is	personnel from bringing
	prohibited from bringing	weapons into the DPR
	armed Brimob personnel	building
	into the DPR building.	
3	Lawyer Hotman Paris	Hotman Paris Hutapea's
	Hutapea Becomes Legal	event as Habib Rizieq
	Counsel for Habib Rizieq	Shihab's legal
	Shihab	representative
4	They said they wanted to	He said he wanted to
	send the UI Student	send the chairman of
	Executive Board (BEM UI)	BEM UI Asmat, but
	chairman to Asmat, but	foreign journalists
	foreign journalists who	wanted to cover it, just
	wanted to cover the event	throw them out.
	were expelled.	

No	Filtering Results	Stemming Results
5	Ministry of Home Affairs	Ministry of Home Affairs
	Revokes Regional	revokes regional
	Regulation on Alcoholic	regulation on alcoholic
	Beverages	beverages

## H. Manual Labeling

Researchers used manual data set labeling involving linguistics to determine tweet labels before beginning the data classification process. The three types of data labels used in this study were positive, negative, and neutral. Table 6 shows some of the results of the manual labeling.

**Table 6.** Example of Manual Labeling Results

No	No Tweet		
1	Kaesang: My father could		
	deceive the Indonesian	Positive	
	people with his simplicity.		
	Why can't I?		
2	The Corruption Eradication		
	Committee (KPK) is		
	prohibited from bringing	Neutral	
	armed Brimob personnel		
	into the DPR building.		
3	Lawyer Hotman Paris		
	Hutapea Becomes Legal	Positive	
	Counsel for Habib Rizieq	rositive	
	Shihab		
4	They said they wanted to		
	send the UI Student		
	Executive Board (BEM UI)		
	chairman to Asmat, but	Positive	
	foreign journalists who		
	wanted to cover the event		
	were expelled.		
5	Ministry of Home Affairs		
	Revokes Regional	Neutral	
	Regulation on Alcoholic	reunai	
	Beverages		

## I. Data Testing Results

The results obtained from testing SVM with linear kernel and RBF can be seen in table 7.

<b>Table 7.</b> Comparison of L	Linear Kernel	and RBF
---------------------------------	---------------	---------

Kernel	Accuracy	Precision	Recall	F1-Score
Linear	54%	0.266	0.50	0.347
RBF	79%	0.782	0.783	0.782

Based on Table 7, the linear kernel produced an accuracy of 54%, while the RBF kernel produced an accuracy of 79%. From both tests, it can be seen that the RBF kernel is superior to the linear kernel in terms of accuracy.

# J. Implementation

# 1. Dashboard Page

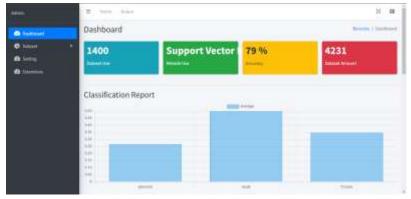


Figure 3. Admin Dashboard Page

Figure 3 shows a dashboard display where you can see the dataset used, the method used, the accuracy of the Support Vector Machine method, and the total amount of incoming data, as well as providing a classification diagram display of the SVM method.

## 2. All Dataset Page



Figure 4. All Data Page

In Figure 4 is an image that displays all the datasets used, where on this page the admin can manage data such as editing and deleting.

## 3. Dataset Usage Page

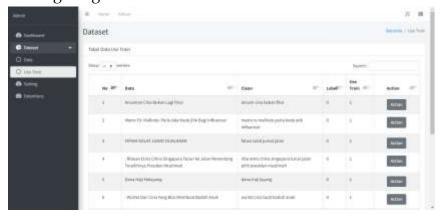


Figure 5. Dataset Usage Page

Figure 5 shows the use of datasets, where on this page the admin can manage data such as editing and deleting.

# 4. Settings Page

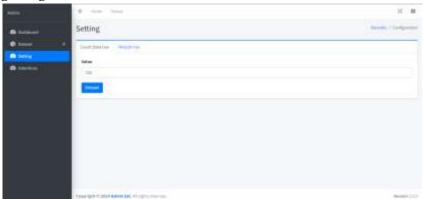
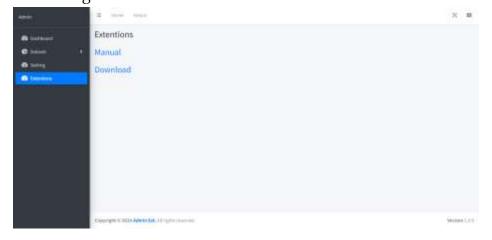


Figure 6. Settings Page

Figure 6 shows an image showing the settings page where the admin can manage data and set the method to be used.

## 5. Extension Page



**Figure 7.** Extension Page

Figure 7 shows an extension page that is used to download and install an extension for the browser for data testing.

6. Page When the Process of Performing Data Testing



# Predicting...

https://agung.givenjeremia.com/api/checking? keyword=Dilansir%20dari%20liputan6.com%2C%20klai m%20kabar%20Presiden%20Jokowi%20mengundurkan %20diri%20dari%20jabatan%20sebagai%20Presiden

**Figure 8.** Page When the Data Testing Process is in progress

Figure 8 shows a page where the program processes test data.

7. Data Test Results Page



Dilansir dari liputan6.com, klaim kabar Presiden Jokowi mengundurkan diri dari jabatan sebagai Presiden

Correct 54%



Figure 9. Data Test Results Page

Figure 9 shows a page that appears after the data testing process, namely the display of the data test results.

#### **CONCLUSION**

**Fundamental Finding:** Based on the analysis of 1,400 tweets obtained from Twitter, this study concludes that the Support Vector Machine (SVM) method can effectively classify hate speech content. Using the RBF kernel and a configuration

matrix with a maximum iteration of 1,400 and a threshold of 0.3, the SVM model achieved an accuracy rate of 79%, demonstrating its strong potential for text classification in social media contexts. **Implication**: These results imply that SVM-based models can serve as a reliable foundation for developing automated systems to detect and mitigate hate speech on digital platforms, thereby supporting healthier online discourse and policy enforcement. **Limitation**: Nonetheless, the study is limited by the relatively small dataset size and the potential biases inherent in tweet sampling, which may restrict the model's generalizability to broader or multilingual datasets. **Future Research**: Future studies should explore larger and more diverse datasets, experiment with hybrid deep learning–SVM models, and assess model performance across different languages and cultural contexts to improve robustness and applicability.

#### **REFERENCES**

- [1] I. Riadi, R. Umar, and FD Aini, "Comparative Analysis of Traffic Anomaly Detection Using Naive Bayes Methods and Support Vector Machine (SVM)," *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 17–24, 2019, doi: 10.33096/ilkom.v11i1.361.17-24.
- [2] MI Fikri, TS Sabrila, and Y. Azhar, "Comparison of Naïve Bayes and Support Vector Machine Methods in Twitter Sentiment Analysis," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020.
- [3] I. Aida Sapitri and M. Fikry, "Classification of WhatsApp Application Review Sentiments on Google Play Store Using Support Vector Machine," *J. TEKINKOM*, vol. 6, no. 1, pp. 1–7, 2023, doi: 10.37600/tekinkom.v6i1.773.
- [4] A. Muhammadin and IA Sobari, "Sentiment Analysis on Kredivo Application Reviews Using SVM and NBC Algorithms," *J. Software Engineering Reputation*, vol. 2, no. 2, pp. 85–91, 2021, doi: 10.31294/reputation.v2i2.785.
- [5] D. Alita, Y. Fernando, and H. Sulistiani, "Implementation of the Multiclass SVM Algorithm on Indonesian-Language Public Opinion on Twitter," *J. Tekno Kompak*, vol. 14, no. 2, p. 86, 2020, doi: 10.33365/jtk.v14i2.792.
- [6] SD Asri, D. Ramayanti, AD Putra, and YT Utami, "Vehicle Wheel Detection Using Circle Hough Transform (Cht) and Support Vector Machine (Svm)," *J. Teknoinfo*, vol. 16, no. 2, p. 427, 2022, doi: 10.33365/jti.v16i2.1952.
- [7] CF Hasri and D. Alita, "Application of the Naã Method Ve Bayes Classifier And Support Vector Machine In Sentiment Analysis Of The Impact Of Corona Virus On Twitter," *J. Inform. and Software Engineering*, vol. 3, no. 2, pp. 145–160, 2022, doi: 10.33365/jatika.v3i2.2026.
- [8] R. Tineges, A. Triayudi, and ID Sholihati, "Sentiment Analysis of Indihome Services Based on Twitter Using the Support Vector Machine (SVM) Classification Method," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [9] E. Suryati, Styawati, and AA Aldino, "Online Transportation Sentiment Analysis Using Word2vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm," *J. Teknol. Dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: https://doi.org/10.33365/jtsi.v4i1.2445.
- [10] AS Rahayu, A. Fauzi, and R. Rahmat, "Comparison of Naive Bayes Algorithm and Support Vector Machine (SVM) in Spotify Sentiment Analysis," *J. Sist. Comput. and*

- Inform., vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.
- [11] D. Darwis, ES Pratiwi, and AFO Pasaribu, "Application of SVM Algorithm for Sentiment Analysis on Twitter Data of the Corruption Eradication Commission of the Republic of Indonesia," *Edutic Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [12] RA Rizal, IS Girsang, and SA Prasetiyo, "Face Classification Using Support Vector Machine (SVM)," *REMIK* (Research and E-Journal of Computer Management and Information), vol. 3, no. 2, p. 1, 2019, doi: 10.33395/remik.v3i2.10080.
- [13] N. Fitriyah, B. Warsito, and DAI Maruddani, "Gojek Sentiment Analysis on Twitter Social Media Using Support Vector Machine (SVM) Classification," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [14] MM Maarif and N. Setiyawati, "Sentiment Analysis of LinkedIn App Reviews on Google Play Store Using Support Vector Machine," *Progresif J. Ilm. Comput.*, vol. 20, no. 1, p. 454, 2024, doi: 10.35889/progresif.v20i1.1614.
- [15] TR Biantong, MT Furqon, and AA Soebroto, "Implementation of the Support Vector Machine Method for Implementation of the Support Vector Machine Method for," *J. Pengemb. Teknol. Inf. dan Ilmu Comput.*, vol. 3, no. June, pp. 185–192, 2019.

#### Agung Mulya Rasyiidi

Muhammadiyah University of Sidoarjo, Indonesia

\* Cindy Taurusta (Corresponding Author)

Muhammadiyah University of Sidoarjo, Indonesia

Email: cindytaurusta@umsida.ac.id