Bulletin of Social Informatics Theory and Application
Vol. 8, No. 2, December 2024, pp. 236-246


# Sentiment analysis of Indonesian government policy in the era of social commerce: public perception and reaction

Sugiarti [a,1,*], Primandani Arsi [a,2], Pungkas Subarkah [a,3], Dhanar Intan Surya Saputra [a,4], Jay V [b,5]

[a] Amikom University Purwokerto, Watumas, Purwanegara, North Purwokerto District, Banyumas Regency, Central Java, Indonesia
[b] University of East London, London E16 2RD, United Kingdom
[1] giarty.kireina@gmail.com; [2] ukhti.prima@amikompurwokerto.ac.id; [3] subarkah@amikompurwokerto.ac.id; [4] dhanarsaputra@amikompurwokerto.ac.id;
[5] jay285@gmail.com
* corresponding author

ARTICLE INFO

**Article history**
Received February 4, 2024
Revised May 4, 2024
Accepted December 6, 2024

**Keywords**
Sentiment analysis
Public perspective
Social commerce

ABSTRACT

This research aims to investigate public sentiment towards Indonesian government policies in the context of social commerce, with a focus on Minister of Trade Regulation Number 31 of 2023. We use data obtained from 1013 Twitter posts to analyze responses to these policies. Various classification algorithms, including Logistic Regression, Support Vector Machine (SVM), Random Forest, Naive Bayes, and K-Nearest Neighbors (KNN), are applied to evaluate people's sentiments. Our main findings show that although the SVM algorithm achieved the highest accuracy rate of 87%, overall public sentiment was negative, with only 20.2% positive sentiment and 79.8% negative sentiment. The implication of this research is that government policies do not receive a positive response from the public. However, research limitations include the limited amount of data and the data source which only comes from Twitter. Future research could broaden the scope by using larger data sets and considering data sources beyond Twitter to gain a more holistic picture of public sentiment. This will help in formulating more effective and responsive policies in the future.



## 1. Introduction

The swift progression of technology and evolving consumer behaviors have profoundly impacted the dynamics of social commerce [1]. Utilizing social media, digital platforms, and online reviews has allowed businesses to engage with consumers more effectively [2]. However, this transformation also introduces new challenges related to data privacy, transaction security, and consumer protection [3].

Digital transformation has altered the way society conducts transactions, especially in the context of business and trade [4]. In the broader context of e-commerce, social commerce has become a crucial pillar in economic and commercial activities in Indonesia. To regulate and monitor this rapid development, the Indonesian government has established policies through Minister of Trade Regulation No. 31 of 2023 concerning licensing, advertising, guidance, and supervision of businesses in trade through electronic systems [5]. This government regulation addresses various aspects of online trading that have the potential to shape and influence the trading ecosystem through social commerce in Indonesia [6].

The significance of government policies in regulating the use of e-commerce in the trade industry cannot be overlooked because it directly impacts businesses, consumers, and society as a whole [7]. Government policies can affect the ability of economic players to innovate, compete, and grow. These policies are also considered necessary to maintain security in online transactions and protect consumer data. Therefore, it is crucial to understand the perceptions and reactions of the public and economic

https:doi.org/10.31763/businta.v8i2.710

players to these regulations-whether government policies align with public expectations or potentially hinder businesses in developing their ventures to reach a wider audience.

Based on the aforementioned issue, an idea emerges to conduct research based on the public's response to government decisions [8]. In a previous similar study regarding sentiment analysis based on online social media indicated an accuracy rate of around 75% was achieved using the Random Forest algorithm. The study also suggested exploring other algorithms such as Logistic Regression, SVM, or KNN to compare whether better results could be achieved [9]. Building upon the findings of the previous research, this research is conducted utilizing several algorithms including Logistic Regression, Support Vector Machine (SVM), Random Forest, Naive Bayes, and K-Nearest Neighbor (KNN) to determine which algorithm yields a higher accuracy level. This research utilizes public comments data extracted from Twitter, comprising initially 1120 data points [10]. However, after the data scraping process, 1013 datasets were obtained. These datasets were subsequently analyzed to determine sentiment responses concerning government regulations related to social commerce.

This research analyzes how the public and economic actors to government policy, with a focus on perceptions of the Minister of Trade Regulation. Hopefully this research can make a significant contribution to the understanding and development of social commerce in Indonesia, which will have an impact on the economy and society as a whole.

## 2. Method

This research employs data gathered from public responses on the Twitter social media platform. The data collection took place on November 4-5, 2023, and total of 1013 dataset were obtained through the process of crawling data from Twitter. Subsequently, the acquired data underwent analysis using various classification algorithm methods. Fig. 1 are the stages of analysis conducted.

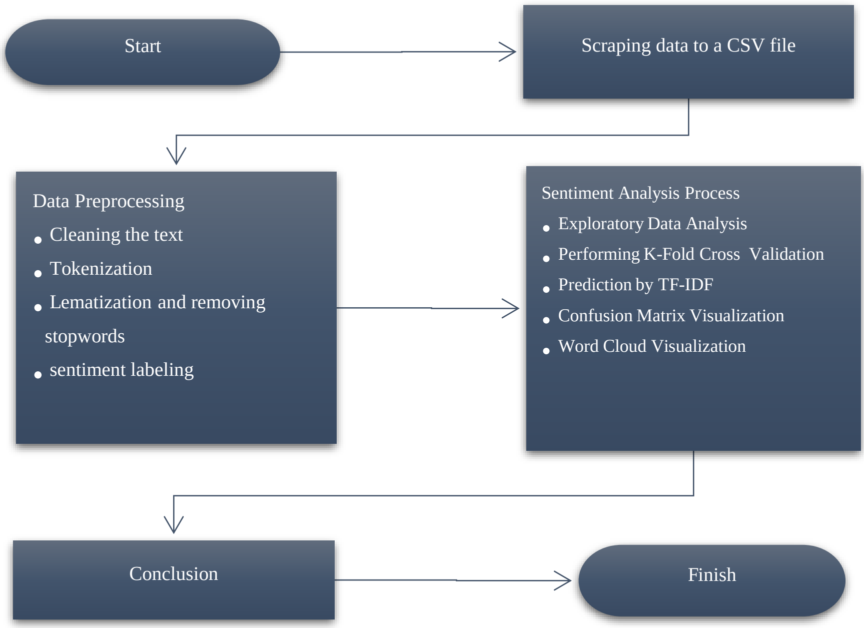



**Fig. 1.** Formation of datasets

## 3. Results and Discuccion

### 3.1. Web Scraping

Web scraping, also known as data scraping, is the practice of obtaining information from websites. Scrapy is an open-source library that provides a general web crawling framework. It navigates online sites and extracts data by examining the HTML tag structure. TweetScraper is designed to scrape tweets with Scrapy, allowing users to enter search terms into the Twitter website. There are two ways for collecting tweets: 1) Using Twitter's APIs, and 2) crawling tweet webpages. The Twitter standard

API is the official mechanism for collecting tweets, allowing authorized users to effortlessly gather all the data types handled by Twitter show as Fig. 2. [11].

| | created_at | id_str | full_text | quote_count | reply_count | retweet_count | favorite_count | lang | user_id_str | conversation_id_str | username | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 401 | Thu Sep 28 06:10:43 +0000 2023 | 1,71E+18 | Jd apa yg dilakukan pemerintah saat ini udh te... | 0 | 0 | 0 | 0 | in | 851458460 | 1,71E+18 | aalooee_veeraa | https://twitter.com/aal |
| 171 | Sun Oct 08 11:25:42 +0000 2023 | 1,71E+18 | @tanyakanrl Pemerintah ngapus e-commerce sama ... | 0 | 0 | 0 | 0 | in | 1,37E+18 | 1,71E+18 | lemontwea | https://twitter.com/lemc |
| 43 | Wed Oct 18 06:30:32 +0000 2023 | 1,71E+18 | udah deh tuh ngubek-ngubek dari kondisi layana... | 0 | 1 | 0 | 0 | in | 1,46E+18 | 1,71E+18 | LXDCCCIII | https://twitter.com/LXD |
| 373 | Thu Sep 28 09:01:00 +0000 2023 | 1,71E+18 | #TiktokShop telah membuat para pelaku umkm mem... | 0 | 0 | 0 | 0 | in | 1,55E+18 | 1,71E+18 | vanilaclatte | https://twitter.com/v |
| 826 | Thu Mar 16 16:37:08 +0000 2023 | 1,64E+18 | Pemerintah Bersama E-commerce Sepakat Berantas... | 0 | 0 | 0 | 0 | in | 9,06E+17 | 1,64E+18 | Jayaneg14 | https://twitter.com/Jaya |

**Fig. 2.** Sample Data Tweets

## 3.2. Data Preparation

Data preparation involves transforming data or tweets into a "bag-of-words" representation to reduce data dimensions while preserving text meaning [12]. Because the text is highly dimensioned unstructured data, it must be cleaned and processed first before analysis. Preprocessing the data entails a variety of procedures, depending on the type of analysis. Before proceeding to the analysis stage, the text preparation procedure begins by cleaning the textual data. Normally, text preparation entails recognizing and deleting non-textual elements from the data, such as hash tags and hyperlinks [13].

Data preprocessing consists of tokenization, case folding, cleaning, removing stopwords, stemming, and filtering [14]. Cleaning involves eliminating noise such as punctuation, hashtags, usernames, URLs, and emoticons. The process of breaking up phrases into smaller units called tokens is known as tokenization. The technique of transforming every letter into the same form is known as case folding, such as lowercase. Removing stopwords is used to eliminate unimportant words. Stemming involves deriving the base word by eliminating affixes. Meanwhile, filtering is process of selecting tweets with non-standard Indonesian words and then replacing them with synonyms (standard words) or adding those non-standard words to optimize the calculation of the frequency of occurrences of words with similar meanings as show in Fig. 3. [15].

| | full_text | label | cleaned_text | nilai | full_text_len | punct | tokens | lemmatized_review |
|---|---|---|---|---|---|---|---|---|
| 0 | Era Digital tak bisa dihentikan tapi dimanfaat... | POSITIVE | era digital tak bisa dihentikan tapi dimanfaat... | 1 | 241 | 2.9 | [era, digital, tak, bisa, dihentikan, tapi, di... | era digital dihentikan dimanfaatkan bijak peme... |
| 1 | Jika benar Facebook, Youtube, dan Tiktok seriu... | NEGATIVE | jika benar facebook youtube dan tiktok seriu... | 0 | 235 | 4.7 | [jika, benar, facebook, youtube, dan, tiktok, ... | facebook youtube tiktok serius garap commerce ... |
| 2 | @jonijoniyesmama kata indah yg selalu dilontar... | POSITIVE | jonijoniyesmama kata indah yg selalu dilontar... | 1 | 263 | 3.0 | [jonijoniyesmama, kata, indah, yg, selalu, dil... | jonijoniyesmama indah dilontarkan anjing suka ... |
| 3 | Kombinasi e-commerce dan media sosial menunjuk... | NEGATIVE | kombinasi e commerce dan media sosial menunjuk... | 0 | 154 | 5.8 | [kombinasi, e, commerce, dan, media, sosial, m... | kombinasi commerce medium sosial potensi pemer... |
| 4 | Tok! Pemerintah Rilis 4 Barang Impor Positive ... | POSITIVE | tok pemerintah rilis barang impor positive ... | 1 | 79 | 8.9 | [tok, pemerintah, rilis, barang, impor, positi... | pemerintah rilis barang impor positive list vi... |

**Fig. 3.** Sample Tweets after Reprocessing Data

### 3.3. Sentiment Analysis Process

Sentiment analysis of text essentially goes through several stages before obtaining the desired output results. The input text is transformed into multiple vectors of a feature extraction, which are then further processed to create a training dataset. Generally, the sentiment analysis process consists of several stages including dataset collection, manual labeling of data, dataset cleaning (preprocessing), feature extraction, classification, and evaluation of output results [16].

Analysis sentiment is an aspect of the data analysis field that aims to examine and extract information from texts, including opinions, evaluations, attitudes, emotions, judgments, and sentiments of individuals towards a product, person, organization, or issue [17]. In this study, the sentiment labeling process for responses is conducted by counting the quantity of both positive and negative terms used in each response. The public's response is classified as positive if the number of positive words is greater, but is classified as negative if the number of negative words is greater. Pie chart based on labeling of tweet reviews as show in Fig. 4.

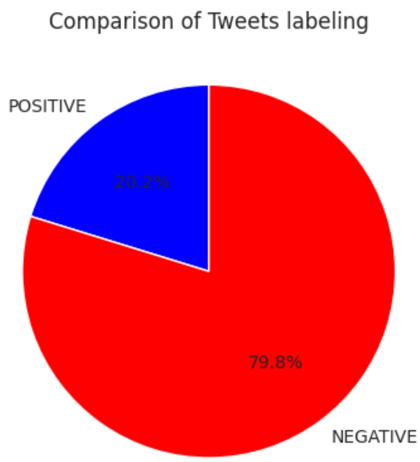



**Fig. 4.** Pie chart based on labeling of tweet reviews

#### 3.3.1. Performing K-Fold Cross Validation

One technique for evaluating a model's correctness that is based on a corpus dataset and forecasting how well it will perform in real-world situations is cross-validation. The ratio of the k segments that make up the data is the same or almost the same. K times through the process is iterated for testing and training. The model is trained on k-1 segments during each iteration, and the remaining segment serves as the test or validation data. This procedure allows for the averaging of results across multiple iterations [18].

In this research, several algorithm models are employed to obtain the most accurate results from the comparison of these models.

- Naïve Bayes

  Due to its high effectiveness, ease of use, quick processing, and basic implementation with a very simple structure, Naive Bayes is a prominent algorithm employed for data mining. An algorithm for classification called Naïve Bayes may categorize a given variable using statistical techniques and probability [11]. Depending on its probability model, the Naïve Bayes classifier can be trained for highly effective supervised learning. Notably, Naïve Bayes doesn't need a lot of training data.

  The formula used for this algorithm is as follows [19]:

$$P(H|X)\frac{(P(X|H)\times P(H))}{P(X)} \quad (1)$$

  Explanation:

  $X$ : Information of uncertain category

$H$ : The idea that data X is a member of a given category

$P\ (H|X)$ : X is the probability of hypothesis H

$P\ (H)$ : Primary probability of hypothesis H

$P\ (X|H)$ : Probability of X in light of assumption H's condition

- Logistic Regression

One classification approach in machine learning that is used to forecast the likelihoods of categorical dependent variables is called Logistic Regression. This approach, which uses binary or probabilistic variables to examine the relationship between several variables, is a generic type of linear regression. A logistic function is employed in the logistic regression classification method to represent the likelihoods of different categories within the given dataset. It produces probability predictions and subsequently separates data into positive, negative, and neutral groups depending on a predefined limit. Because it may produce helpful probability predictions for assessing the sentiment of texts, logistic regression is another frequently used classification technique in sentiment analysis [20].

The principal purpose of logistic regression is in categorization [21]. The formula which represents the categorization border, is found using logistic regression. The optimization method is employed by the training classifier to identify the optimal regression coefficient in the formula. An unrestricted set of inputs is used in classification based on logistic regression, and the output is produced by a function that classifies the input data. For example, the function yields either 0 or 1 in the two classifications represents two categories when classifying, to facilitate executing. The array of the aforementioned function's parameter is from positive infinity to negative infinity based on the verified demands and the analysis above. Ranges for the dependent variable are 0 and 1. Several functions fulfill the aforementioned requirements.

- Random Forest

Recursive binary split is the approach that Random Forest (RF) algorithm employs to get to the final element in a tree structure according on regression and classification structures [22], [23]. This algorithm has benefits, including as the ability to generate relatively few errors, strong classification performance, the ability to manage massive volumes of training data well, and a useful technique for approximating missing data. With subsets chosen at random via bootstrap from the training sample and the input variables at each node, the Random Forest generates numerous of distinct structures [24].

A classifier called Random Forest is consisting of a collection of tree-structured classifiers denoted as $\{h(x,k), k = 1, \ldots,\}$, where each tree casts a vote for the most preferred class at input x, and the $\{k\}$ represent independent and identically distributed random vectors. The class with the most votes wins when numerous trees are generated, each one casts a vote for a specific class [25].

- Support Vector Machine (SVM)

The algorithm of this method is used in machine learning for regression and classification problems. Prediction (predicting continuous values) and binary classification (dividing data into two classes) can both be accomplished with SVM. One of SVM's advantages is capability to perform well with complicated data. Preprocessed training data and TF-IDF computations are used in the model development procedure during implementation. Depending on the kind of kernel being utilized, different parameters will be employed during the model creation process [26].

- K-Nearest Neighbor

The K-Nearest Neighbor technique, often abbreviated as KNN, is frequently used for text and data classification. Classifying objects according to their properties and training samples is the aim of this approach. It is among the techniques that makes use of supervised learning algorithms. Supervised learning seeks to identify new patterns in data by connecting preexisting patterns, whereas unsupervised learning works with input that lacks any preexisting models [27]. This is the main distinction between supervised and unsupervised learning

#### 3.3.2. Prediction by TF-IDF

To determine the weight of every term, the TF-IDF algorithm is applied first. [28] TF-IDF quantifies the relevance of a word within a single document by considering its frequency relative to a collection of documents is indicated by a numerical statistic. The weighting technique known as TF–IDF in Eq. (1) is widely used and increases in value in direct proportion to the incidence of a term within the document.

$$TF - IDF(t, d, D) = tf(t, d). \tag{2}$$

$tf\ (t, d)$ represents Term rate - The relative rate of phrase t in the text d.

In Eq. (1), the $tf\ (t, d)$ is computed as follows,

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \tag{3}$$

Where is the term count in a document, $ft, d$.

The amount of information a word contains is indicated by its document frequency inverse (t, D) in Equation (1).

The calculation of $idf\ (t, D)$ is,

$$idf(t, D) = log \frac{N}{|\{d \in D : t \in d\}|} \tag{4}$$

Where, $\{d \in D: t \in d\}|$ is the number of documents that contain the term t, and N is the total number of documents in a corpus ($N = |D|$).

The sentiment of a given text can be predicted using classification techniques. A model with predefined labels is developed using the given dataset. A sentiment classification's performance can be assessed in 4 (four) common ways [29].

- **Accuracy** is calculated as the ratio of the sum of true positives (TP) and true negatives (TN) to the total number of cases, which includes true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN): (TP+TN)/(TP+TN+FP+FN)
- **Precision** is determined by dividing the number of true positives (TP) by the total number of positive predictions, which includes true positives (TP) and false positives (FP): TP/(TP+FP)
- **Recall** is given by the ratio of true positives (TP) to the sum of true positives (TP) and false negatives (FN): TP/(TP+FN)
- **F1 Score** is computed as the harmonic mean of precision and recall, calculated as: 2×(Precision×Recall)/(Precision+Recall)

After conducting training and testing data using the aforementioned algorithms, the obtained results for accuracy, precision, recall, and F-1 are as follows

#### 3.3.3. Confusion matrix (TF-IDF) visualization

The performance evaluation section of a sentiment classification model provides important insights into how well the model performs in classifying sentiment. The parameters are used for evaluation is F1 score, recall, accuracy, and precision [30].

Accuracy is an overall measure of how often a model provides correct predictions.If a model achieves an accuracy of 0.87, it means that 87% of its predictions were accurate. Precision measures how many positive predictions are correct out of all the predictions classified as positive by the model. High precision suggests that the model rarely gives false positive predictions. Example: if positive precision is 0.83, that means 83% of all of the model's predictions classified as positive were accurate.

The number of positive classifications that the model properly identified is measured by recall, also known as recall. A high recall indicated that most of the samples that are genuinely positive can be identified by the model. In the event that the positive recall, for instance, is 0.41, only 41% of all true positive samples are recognized by the model.

The precision and recall weighted average is called the F1-Score. To achieve a compromise between recall and precision, the F1 score proves useful. Positive precision and positive recall, for instance, are balanced if the positive F-1 is 0.55.

Visualization is a helpul instrument for understanding model achievement in more detail. It shows how well the model classifies samples into the correct classes and how often it makes errors. Fig. 5 to Fig. 9 show the confusion matrix for each classification algorithm used in this research.

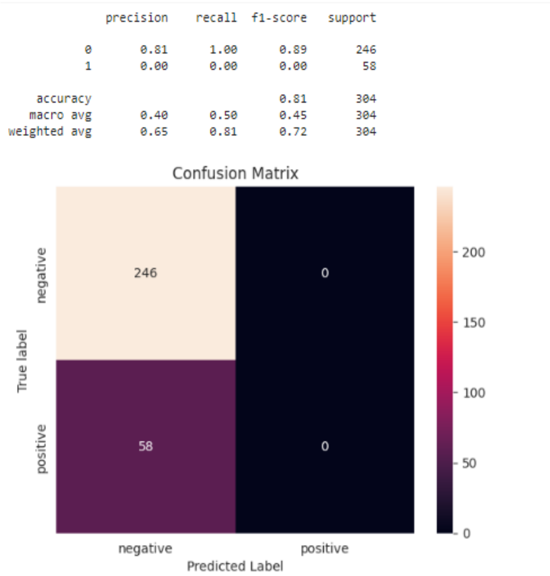



**Fig. 5.** Confusion Matrix using Naive Bayes Algorithm

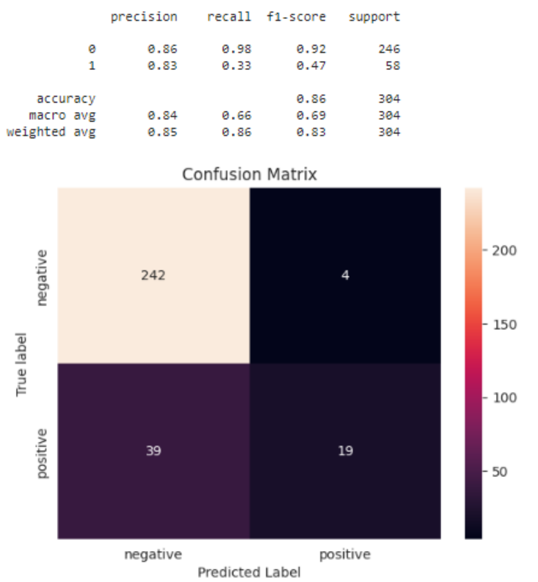



**Fig. 6.** Confusion Matrix using Random Forest Algorithm

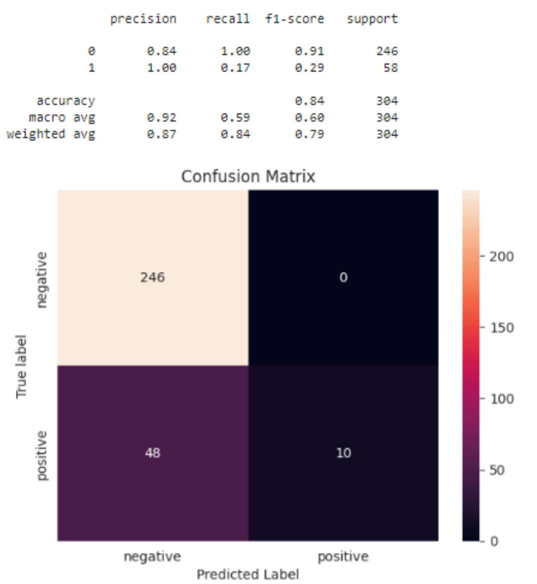



**Fig. 7.** Confusion Matrix using Logistic Regression Algorithm

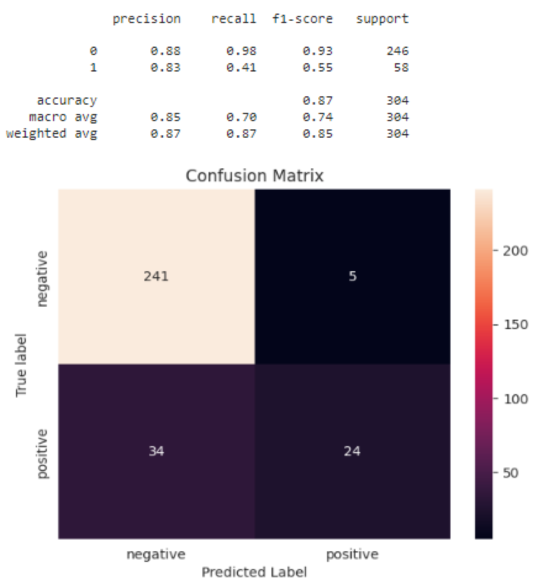
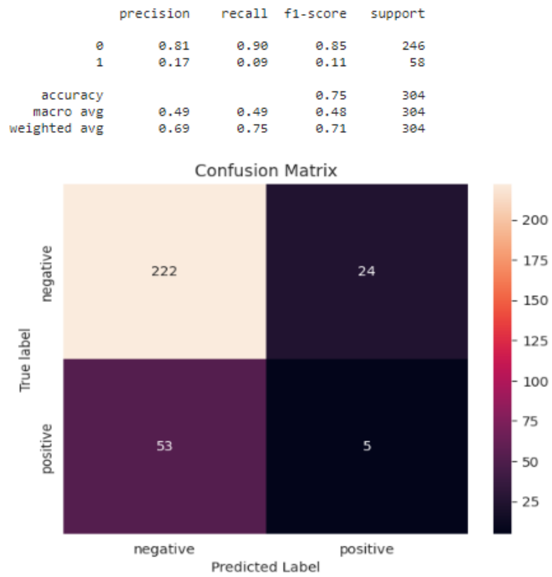



**Fig. 8.** Confusion Matrix using SVM Algorithm



| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.81 | 0.90 | 0.85 | 246 |
| 1 | 0.17 | 0.09 | 0.11 | 58 |
| accuracy | | | 0.75 | 304 |
| macro avg | 0.49 | 0.49 | 0.48 | 304 |
| weighted avg | 0.69 | 0.75 | 0.71 | 304 |



**Fig. 9.** Confusion Matrix using KNN Algorithm

Based on the evaluation results in Table 1. of each algorithm, they are given in detail. For example, for the SVM algorithm, an accuracy of 0.87 indicates that the overall model provides correct predictions in 87% of cases. A positive precision of 0.83 indicates that of all the model's positive predictions were correct in 83% of cases. A positive recall of 0.41 indicates that the model only recognized 41% of all true positive samples. Higher values suggest overall superior performance; a positive F-1 of 0.55 illustrates a trade-off between good recall and precise positive predictions.

**Table.1** Evaluation Results

| Algoritma | Score | | | | | | |
|---|---|---|---|---|---|---|---|
| | Accuracy | Precision | | Recall | | F-1 | |
| | | Positive | Negative | Positive | Negative | Positive | Negative |
| Naïve Bayes | 0.81 | 0 | 0.81 | 0 | 1.00 | 0 | 0.89 |
| Logistic Regression | 0.84 | 1.00 | 0.84 | 0.17 | 1.00 | 0.29 | 0.91 |
| Random Forest | 0.86 | 0.83 | 0.86 | 0.33 | 0.98 | 0.47 | 0.92 |
| SVM | 0.87 | 0.83 | 0.88 | 0.41 | 0.98 | 0.55 | 0.93 |
| KNN | 0.75 | 0.17 | 0.81 | 0.09 | 0.90 | 0.11 | 0.85 |

According to the outcomes of this evaluation, it can be determined that SVM represents the most effective algorithm for classifying public sentiment towards government policies in the context of social commerce in Indonesia. Comparing of Algorithm Accuracy show in Fig. 10.

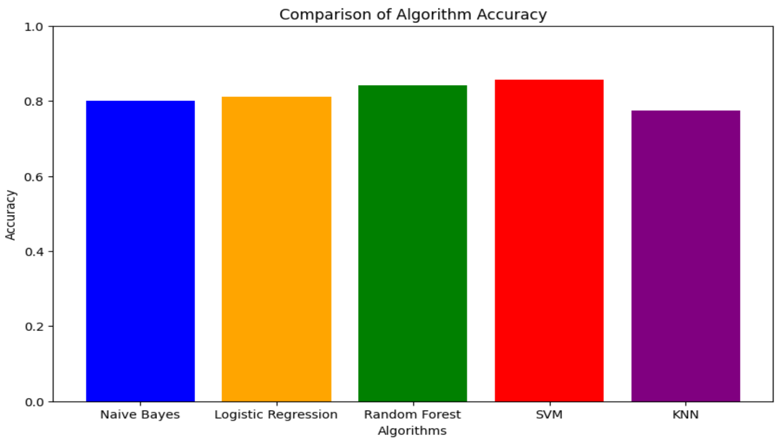



**Fig. 10.** Comparing of Algorithm Accuracy

### 3.3.4. Word Cloud Visualization

Word Cloud is a visual illustration made up of various words, considering each word's size reflecting its occurrence or significance. The more prominent a word is, the more often it appears in the text. Word cloud of positive responses as show in Fig. 11 and Word cloud of negative responses as show in Fig. 12.


Fig. 11. Word cloud of positive responses


Fig. 12. Word cloud of negative responses

## 4. Conclusion

Analysis of public sentiment towards Indonesian government's policies in the era of social commerce, using sentiment data obtained from 1013 Twitter posts, the results indicate that classification using Support Vector Machine (SVM) achieves a higher accuracy rate of 87% compared to other algorithms. In this study, it was found that public sentiment regarding the Indonesian government's policies in the era of social commerce, specifically through Minister of Trade Regulation No. 31 of 2023 concerning licensing, advertising, guidance, and supervision of businesses in trade through electronic systems, garnered a positive sentiment of 20.2% and a negative sentiment of 79.8%. Therefore, it can be concluded that the policies implemented by the government did not receive a positive response from the public. Using a more comprehensive dataset is recommended for future research and consider sources of data beyond Twitter alone to achieve higher accuracy results.

## References

[1] N. Zhao and H. Li, "How can social commerce be boosted? The impact of consumer behaviors on the information dissemination mechanism in a social commerce network," *Electron. Commer. Res.*, vol. 20, no. 4, pp. 833–856, 2019, doi: 10.1007/s10660-018-09326-3.

[2] Y. K. Dwivedi *et al.*, "Setting the future of digital and social media marketing research: Perspectives and research propositions," *Int. J. Inf. Manage.*, vol. 59, p. 102168, Aug. 2021, doi: 10.1016/j.ijinfomgt.2020.102168.

[3] Adedoyin Tolulope Oyewole, Bisola Beatrice Oguejiofor, Nkechi Emmanuella Eneh, Chidiogo Uzoamaka Akpuokwe, and Seun Solomon Bakare, "Data Privacy Laws and Their Impact on Financial Technology Companies: a Review," *Comput. Sci. IT Res. J.*, vol. 5, no. 3, pp. 628–650, 2024, doi: 10.51594/csitrj.v5i3.911.

[4] M. Mattila, M. Yrjölä, and P. Hautamäki, "Digital transformation of business-to-business sales: what needs to be unlearned?," *J. Pers. Sell. Sales Manag.*, vol. 41, no. 2, pp. 113–129, 2021, doi: 10.1080/08853134.2021.1916396.

[5] Minister Of Trade Regulation Of The Republic Of Indonesia Number 31 Of 2023 Concerning Business Licenses, Advertising, Guidance, And Supervision Of Business Players In Trade Through Electronic Systems, no. 3. Indonesia, pp. 31-41, 2023. [Online]. Available at: https://jdih.kemendag.go.id/peraturan/peraturan-menteri-perdagangan-nomor-31-tahun-2023-tentang-perizinan-berusaha-periklanan-pembinaan-dan-pengawasan-pelaku-usaha-dalam-perdagangan-melalui-sistem-elektronik.

[6] H. N. Utami, E. Alamanos, and S. Kuznesof, "'A social justice logic': how digital commerce enables value co-creation at the bottom of the pyramid," *J. Mark. Manag.*, vol. 37, no. 9–10, pp. 816–855, 2021, doi: 10.1080/0267257X.2021.1908399.

[7] E. Terryn and E. Van Gool, "The Role of European Consumer Regulation in Shaping the Environmental Impact of e-Commerce," *SSRN Electron. J.*, no. November, 2020, doi: 10.2139/ssrn.3732911.

[8] E. Sørensen and J. Torfing, "Radical and disruptive answers to downstream problems in collaborative governance?," *Public Manag. Rev.*, vol. 23, no. 11, pp. 1590–1611, 2021, doi: 10.1080/14719037.2021.1879914.

[9] Bahrawi, "Sentiment Analysis Using Random Forest Algorithm-," *J. Inf. Technol. ITS Util.*, vol. 2, no. 2, pp. 29–33, 2019, doi: 10.30818/jitu.2.2.2695.

[10] E. Michela, J. M. Rosenberg, R. Kimmons, O. Sultana, M. A. Burchfield, and T. Thomas, "'We Are Trying to Communicate the Best We Can': Understanding Districts' Communication on Twitter During the COVID-19 Pandemic," *AERA Open*, vol. 8, no. 1, 2022, doi: 10.1177/23328584221078542.

[11] J. You, J. Lee, and H.-Y. Kwon, "A Complete and Fast Scraping Method for Collecting Tweets," in *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jan. 2021, pp. 24–27, doi: 10.1109/BigComp51126.2021.00014.

[12] Y. S. Mehanna and M. Mahmuddin, "The Effect of Pre-processing Techniques on the Accuracy of Sentiment Analysis Using Bag-of-Concepts Text Representation," *SN Comput. Sci.*, vol. 2, no. 4, p. 237, Jul. 2021, doi: 10.1007/s42979-021-00453-7.

[13] A. A. Onyekachi, "Sentiment Analysis for E-Commerce Product Reviews," *Int. J. Eng. Mod. Technol.*, vol. 9, no. 1, pp. 18–32, 2023. [Online]. Available at: https://iiardjournals.org/abstract.php?j=IJEMT&pn=Sentiment%20Analysis%20for%20E-Commerce.

[14] N. Garg and K. Sharma, "Text pre-processing of multilingual for sentiment analysis based on social network data," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 1, pp. 776–784, 2022, doi: 10.11591/ijece.v12i1.pp776-784.

[15] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Comparison of Random Forest, Naïve Bayes, and Support Vector Machine Algorithms in Analyzing Twitter Sentiment Regarding Public Opinion on the Removal of Honorary Employees," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 563–572, 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.

[16] E. Mona Cindo, Dian Palupi Rini, "Literature Review: Classification Methods in Sentiment Analysis," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2019, pp. 66–70. [Online]. Available at: https://prosiding.seminar-id.com/index.php/sainteks/article/view/124.

[17] X. Zou, Y. Hu, Z. Tian, and K. Shen, "Logistic Regression Model Optimization and Case Analysis," in *International Conference on Computer Science and Network Technology (ICCSNT) Logistic*, 2019, pp. 135–139, doi: 10.1109/ICCSNT47585.2019.8962457.

[18] B. Agarwal and N. Mittal, "Machine Learning Approach for Sentiment Analysis," Springer, Cham, 2016, pp. 21–45, doi: 10.1007/978-3-319-25343-5_3.

[19] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Comparison of Random Forest, Naïve Bayes, and Support Vector Machine Algorithms in Twitter Sentiment Analysis Regarding Public Opinion on the Elimination of Honorary Workers," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 563–572, Nov. 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.

[20] F. Sodik and I. Kharisudin, "Sentiment Analysis with SVM, Naïve Bayes, and KNN for Studying Indonesian Public Responses to the Covid-19 Pandemic on Twitter Social Media," *Prisma*, vol. 4, pp. 628–634, 2021.[Online]. Available at: https://journal.unnes.ac.id/ .

[21] V. Oktaviani, B. Warsito, H. Yasin, R. Santoso, and Suparti, "Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method," *J. Phys. Conf. Ser.*, vol. 1943, no. 1, p. 012147, Jul. 2021, doi: 10.1088/1742-6596/1943/1/012147.

[22] E. Edwin, "Government Policy Assessment Application Related To Online Operations Of Ojek During The Pandemic (Covid-19) Using The Naïve Bayes Classifier Algorithm," *J. Algoritm. Log. dan Komputasi*, vol. 3, no. 2, Mar. 2021, doi: 10.30813/j-alu.v3i2.2651.

[23] A. C. (aristin) khotimah and E. (Ema) Utami, "Comparison Naïve Bayes Classifier, K-nearest Neighbor and Support Vector Machine in the Classification of Individual on Twitter Account," *J. Tek. Inform.*, vol. 3, no. 3, pp. 673–680, Jun. 2022. [Online]. Available at: https://www.neliti.com/publications/497600/comparison-na%C3%AFve-bayes-classifier-k-nearest-neighbor-and-support-vector-machine.

[24] E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review," *J. Data Anal. Inf. Process.*, vol. 08, no. 04, pp. 341–357, Sep. 2020, doi: 10.4236/jdaip.2020.84020.

[25] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *J. Inf. Syst. Informatics*, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.

[26] N. I. P. Munggaran and E. B. Setiawan, "DISC Personality Prediction with K-Nearest Neighbors Algorithm (KNN) Using TF-IDF and TF-Chi Square Weighting," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 9446–9457, 2019, [Online]. Available at: https://repository.telkomuniversity.ac.id/pustaka/files/152887/jurnal_eproc/prediksi-kepribadian-disc-dengan-k-nearest-neighbors-algorithm-knn-menggunakan-pembobotan-tf-idf-dan-tf-chi-square.pdf.

[27] F. Smarra, G. D. Di Girolamo, V. De Iuliis, A. Jain, R. Mangharam, and A. D'Innocenzo, "Data-driven switching modeling for MPC using Regression Trees and Random Forests," *Nonlinear Anal. Hybrid Syst.*, vol. 36, p. 100882, May 2020, doi: 10.1016/j.nahs.2020.100882.

[28] F. Y. Pamuji and V. P. Ramadhan, "Comparison of Random Forest and Decision Tree Algorithms for Predicting Immunotheraphy Success," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 1, pp. 46–50, Jul. 2021, doi: 10.26905/jtmi.v7i1.5982.

[29] P. Fremmuzar and A. Baita, "SVM Kernel Test in Sentiment Analysis of Telkomsel Services on Twitter Social Media," *Komputika J. Sist. Komput.*, vol. 12, no. 2, pp. 57–66, Sep. 2023, doi: 10.34010/komputika.v12i2.9460.

[30] R. Bertolini, S. J. Finch, and R. H. Nehm, "Enhancing data pipelines for forecasting student performance: integrating feature selection with cross-validation," *Int. J. Educ. Technol. High. Educ.*, vol. 18, no. 1, p. 44, Dec. 2021, doi: 10.1186/s41239-021-00279-6.