

Sentiment Analysis of Citayam Fashion Week Phenomenon Using Support Vector Machine

Muhammad Rosyadi¹, Erlin²

Department of Informatics Engineering, Institut Bisnis dan Teknologi Pelita Indonesia^{1,2}
m.rosyadi@student.pelitaindonesia.ac.id¹, erlin@lecturer.pelitaindonesia.ac.id²

Article Info

Article history:

Received Feb 2, 2023
Revised Mar 3, 2023
Accepted Jun 14, 2023

Keyword:

Sentiment analysis
Support Vector Machine
Phenomena
Citayam Fashion Week

ABSTRACT

Citayam Fashion Week is a phenomenon that displays a model doing a fashion show using distinctive and unique clothing when crossing a zebra cross on a catwalk. This phenomenon has received extraordinary attention and discussion from various circles and led to numerous pros and cons among the public and observers of society in Indonesia. Therefore, it is of great importance to conduct a study on sentiment analysis of this phenomenon to determine society's sentiment tendency to provide government references and help decision-makers improve their policies. Sentiment analysis was performed using the Support Vector Machine based on the polynomial kernel. The results show that the accuracy, recall, precision, and F1-Score value of 95.61%, 95.66%, 96%, and 95.55%, respectively. This study proved that the Support Vector Machine classifier with the polynomial kernel provides higher algorithm performance on text classification. Therefore, the government can use the result of this study to evaluate the existence of the Citayam Fashion Week which may be followed by other phenomena.

© This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Muhammad Rosyadi
Department of Informatics Engineering
Institut Bisnis dan Teknologi Pelita Indonesia
Email: m.rosyadi@student.pelitaindonesia.ac.id

1. INTRODUCTION

One of the phenomena that have become a trending topic and hot topic of conversation in 2022 is Citayam Fashion Week. Initially, this phenomenon was just a gathering of young people from Citayam, Bojonggede, Bogor, and Depok who hung out and looked for entertainment at a well-known place in Jakarta that expressed various kinds of unique and distinctive fashion models. The teenagers are demonstrating like a model doing a Fashion Show [1]. The Citayam fashion week action is demonstrated in the zebra crossing area as the catwalk. The teenagers are waddling in distinctive and unique clothing while crossing the street [2].

This phenomenon gave rise to various kinds of public opinion on social media, one of which is Twitter. Many parties support Citayam Fashion Week. The public can retrieve various kinds of information and create a picture of sentiment through tweets. Therefore, this research on public opinion was carried out to see the text of public sentiment so that it could be used as a consideration for the government to give permission or vice versa for this phenomenon.

Sentiment analysis is a science that analyzes opinions, people's expressions in the form of emojis, and people's sentiments about something. Sentiment will determine the category of each opinion or public opinion in the form of the text so that it is in the form of positive or negative sentiment [3]. Information from sentiment analysis text can be used for consideration in making a decision and evaluating a matter [4]. Research using sentiment analysis has continued to increase in popularity because it can provide advantages in various aspects. One aspect that can apply sentiment analysis is satisfaction with a product, service, or phenomenon [5].

There are several algorithms used in sentiment analysis research, such as Support Vector Machine, Naive Bayes and K-nearest neighbors, Logistic Regression, and other algorithms. The research was carried out by Giovani et al (2020) concerning Sentiment Analysis for the Ruang Guru Application on Twitter Using a Classification Algorithm that uses the Support Vector Machine, Naive Bayes, and K-Nearest Neighbors algorithms based on the Particle Swarm Optimization (PSO) feature which evaluates the success of the teacher's room application. It was obtained that the accuracy of the Support Vector Machine algorithm was 78.55%, Naive Bayes was 67.32%, and K-nearest neighbors were 77.21%, respectively [6].

In 2021, a research project was also conducted by Himawan and Eliyani on the comparison of the accuracy of Tweet Sentiment Analysis for the Provincial Government of DKI Jakarta during the Pandemic Period using the Support Vector Machine, Naive Bayes, and Random Forest classifier. The result showed that the accuracy of Random Forest and Naive Bayes classifiers behind the accuracy of the Support Vector Machine classifier obtained values of 75.81%, 75.22%, and 77.58%, respectively [7]. Research conducted by Kelvin et al (2022) on the comparative analysis of Corona Virus Disease-2019 (Covid19) sentiment on Twitter Using the Logistic Regression Method and Support Vector Machine (SVM) to see people's responses to the coronavirus obtained the results of Support Vector Machine accuracy of 91.15% and the Logistic Regression method of 87.68% [8]. Based on the results of previous research on the Support Vector Machine algorithm, which had good accuracy, the Support Vector Machine (SVM) algorithm is popular as it is commonly used by most researchers. This is further strengthened by research that the SVM algorithm can be used in sentiment analysis in grouping public opinion. According to Erlin et al (2021) concerning Sentiment Analysis for AMD Ryzen Processors using the Support Vector Machine Method, which discusses public opinion trends towards AMD Ryzen Processors, the results of the SVM method were very good, with an accuracy of 96.76% In 2023 research conducted by Idris et al on Sentiment Analysis of the Use of the Shopee Application Using the Support Vector Machine (SVM) Algorithm, which evaluated comment data from Shopee users from the reviews provided with the Support Vector Machine algorithm, obtained an accuracy of 98% [10]. According to Erlin et al (2022) in the research on Sentiment Analysis for the Elimination of the National Examination Using the Support Vector Machine Method, which discusses public opinion regarding the abolition of the national exam, the accuracy results obtained with the SVM method were 96.97% [11]. These three studies proved that the Support Vector Machine is the best algorithm for sentiment analysis.

Based on the literature review of the Support Vector Machine (SVM) method used by previous research to classify public opinion and sentiment in the form of text originating from tweets on Twitter into positive, negative, and neutral categories, to the best of the author's knowledge, there is no research discussing on the Citayam Fashion Week phenomenon, even though this phenomenon may generate other similar phenomena that need to be studied for its existence. Therefore, this study showed the ability of the SVM method to classify opinions or texts.

2. RESEARCH METHOD

The research method is described in steps that have been arranged in a structured manner so that data collection, processing, and testing produce good information. The steps for text classification about the Citayam Fashion Week phenomenon are shown in Figure 1. The first step is collecting data through social media Twitter using the scraping technique. Next, labeling, preprocessing, and classification were carried out using the Support Vector Machine method, and the final step was evaluating the classification model using the confusion matrix. The research used several libraries of Python such as Pandas, Scrape, and Matplotlib. The scrape library was used to retrieve data without access keys, hence, the step to create a Twitter Development account is not

required. The pandas' library was used to process data, from data cleaning and manipulation to data analysis. Next, the Matplotlib library was used for data visualization.

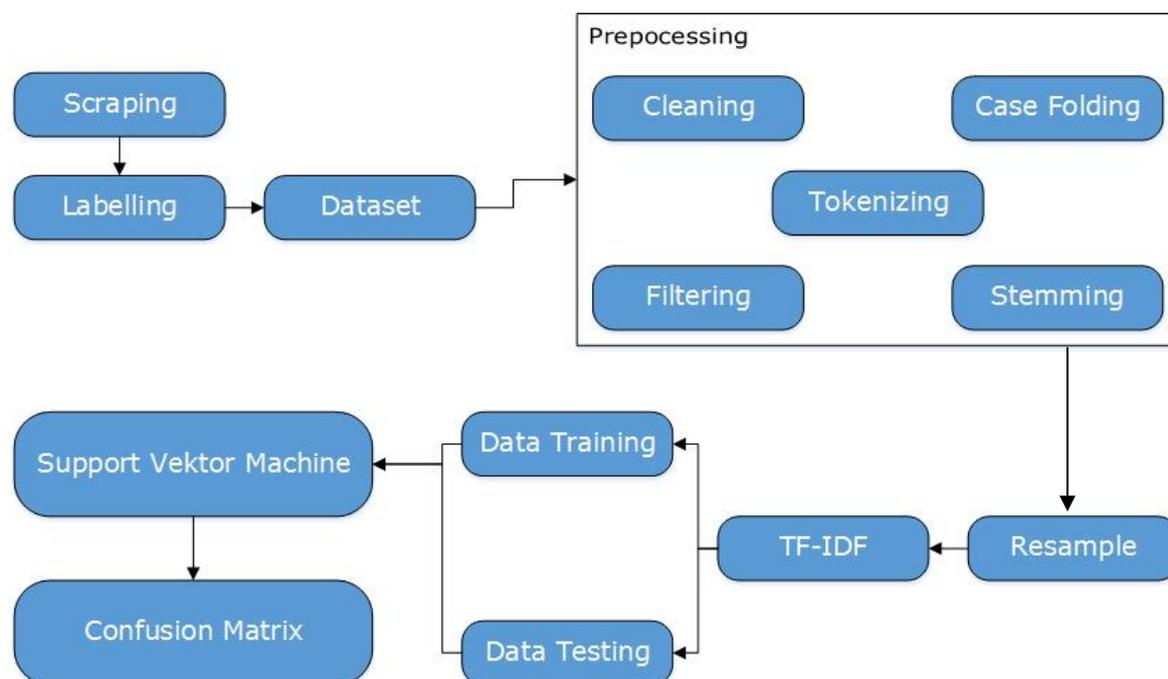


Figure 1. Research Framework

2.1. Data Collection

The data used in this study were taken from Twitter in the form of tweets from Twitter users who used Indonesian related to the Citayam Fashion Week phenomenon using the library contained in the Python programming language. In this phase, Scrape and Pandas libraries were used in the data collection process.

The amount of data used was 1219 data. The data then proceed to the labeling process. The labeling process was done manually. In this case, the grouping of data was conducted by labeling the data into three sentiment classes, namely positive, negative, and neutral sentiments. The labeling process was carried out to train classification modeling in machine learning based on the sentiment classes that have been made before. Data that have been labeled further became a dataset.

2.2. Preprocessing

Doing sentiment requires the Preprocessing stage. The available dataset still contained many non-standard words, so preprocessing was carried out to obtain maximum accuracy and increase system performance [12]. The stages in pre-processing were cleaning, case folding, tokenizing, filtering, and stemming.

2.3. Resampling Technique

After the pre-processing process was carried out, it was continued with the resampling technique. This technique can solve unbalanced data problems so that the data becomes balanced. An imbalance in the dataset makes classification modeling difficult to predict because the majority class will dominate the classification model, so the minority class is neglected [13]. In this study, balancing the data was carried out using the Upsampling technique.

2.4. TF-IDF Weighting

TF-IDF weighting is a process of carrying out a data transformation from textual data into numeric data to carry out a weighting on features or each word. TF IDF is a measure in the form of statistics used to evaluate a word's importance in a document. TF is the frequency of words appearing

in each document, describing how important the word is in each document. DF is the number of times the document contains the word and describes how common the word is. IDF is an inverse of the DF process [14]. The TF-IDF value can be determined using equation (1).

$$W_{if} = tf_{if} + \left\{ \log \frac{D}{df_i} \right\} \quad (1)$$

2.5. Classification of Support Vector Machines

Support Vector Machine is a technique that aims to make predictions in regression or classification. SVM includes a supervised learning class. In the implementation of SVM, there needs to be training and testing. Researchers widely applied SVM in solving everyday life cases, such as its application in gene expression analysis, financial prediction, and weather to the medical field. The application of a Support Vector Machine in classification is to find the best hyperplane as a separator for two data classes. The simple idea of using SVM is to make the most of the margin, which is the separation between data classes. SVM works with datasets and uses trick kernels. SVM will select several data points that will contribute to producing a model used in the classification process [15]. SVM has seven kernels which can be seen in Table 1.

Table 1. Kernel Support Vector Machine (Sianturi et al., 2019)

No	Types of Kernel	Function
1	Linear	$K(x,y) = x.y$
2	Polynomial of degree d	$K(x,y) = (x.y)^d$
3	Polynomial of degree up to d	$K(x,y) = (x.y+c)^d$
4	Gaussian RBF	$K(x,y) = \exp\left(\frac{-\ x-y\ ^2}{2\sigma^2}\right)$
5	Sigmoid (Tangen Hiperbolik)	$K(x,y) = \tanh(\sigma(x.y)+c)$
6	Invers Multi Kuadratik	$K(x,y) = \frac{1}{\sqrt{\ x-y\ ^2 + c^2}}$
7	Additive	$K(x,y) = \sum_{i=1}^n k_i(x_i, y_i)$

Based on Table 1, the use of a linear kernel when the classified data are separated by a hyperplane and the use of a non-linear kernel when the data can only be separated using a curved line or plane contained in a high-dimensional space. This study uses the Polynomial kernel.

3. RESULTS AND DISCUSSION

The discussion of the sentiment analysis results of the text classification of the Citayam Fashion Week phenomenon, which begins began with the scraping method and ends with the evaluation of the model, will be presented in this chapter.

3.1. Scraping Data

Web scraping is a way to extract data and information from a website, which are then stored in a specific format. In this research, the tweets on social media Twitter were scrapped for data analysis. The tweets were taken from July 10, 2022, to September 29, 2022, because during that time the phenomenon was booming and became a lot of talk by the public. Tweet data were taken using the Indonesian language by the keywords "#citayamfashionweek" and "Citayam Fashion Week." Data collection was carried out using import libraries that Python provided. The tweet data taken were in the form of the date/time, username, and contents of the tweet. In this study, the total amount of data is 1219 which was expected to represent the results of public opinion. The highest number of tweets in the dataset was on 20 July 2022. Figure 2 is a data display that can be seen based on the highest number of tweets by date.

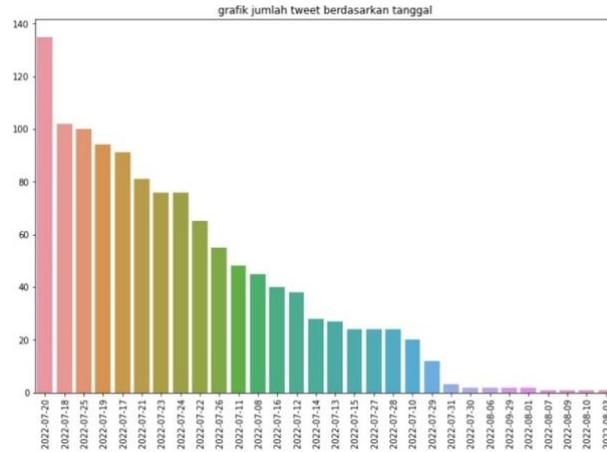


Figure 2. Data display that based on the highest number of tweets by date

After the data was collected, each tweet was labeled. In this case, the tweets were made into three labels, Positive, Negative, and Neutral. The result of labeling can be displayed using the pie chart as shown in Figure 3, which is the value of 69.9%, 4%, and 26.1% for positive, neutral, and negative, respectively. Furthermore, Table 2 is a snippet of tweets that have been labeled into three classifications.

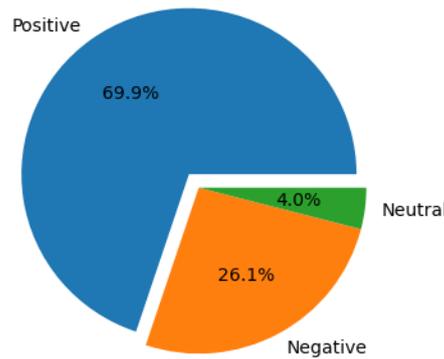


Figure 3. Distribution of Positive, Negative, and Neutral Tweets

Table 2. Labelling of Data Collection

No	Date	Username	Tweet	Label
1	2022-08-26 03:24:55+00:00	Wartech_24	Seru banget pagi ini.. tidak mau kalah dengan #citayamfashionweek di Jakarta.. Mantan model senior @arzeti_bilbina on the spot membimbing kaum millennial #lubuklinggau untuk melenggok di catwalk.. @cakimiNOW	Positive
2	2022-08-07 16:29:31+00:00	BuleKeriting	industri fashion bisa jadi peluang ekonomi #citayamfashionweek	Positive
3	2022-08-07 14:03:45+00:00	Momon_alfatih	Citayam Fashion Week kalah kelas dengan Pagelaran Fashion jalanan ini #citayamfashionweek	Negative

3.2. Preprocessing Data

The data that have been collected need to be preprocessed because the sentences contained in the tweet did not fully use standard words and good Indonesian language. Preprocessing itself was

done using the library of Python. The preprocessing has five stages, Cleaning, Case Folding, Tokenizing, Filtering, and Stemming.

At the cleaning stage, a process was carried out to remove words that have no meaning and meaning. The cleaning process was carried out so that the resulting sentiment analysis had very high accuracy. According to Shafira [16], the cleaning stages remove mentions and Hashtags, URLs, RTs, numbers, punctuation marks, emojis, and Whitespace characters. The results of the cleaning can be seen in Table 3.

Table 3. Cleaning Process

Tweet	Text_Clean
Seru banget pagi ini.. tidak mau kalah dengan #citayamfashionweek di Jakarta.. Mantan model senior @arzeti_bilbina on the spot membimbing kaum millennial #lubuklinggau untuk melenggok di catwalk.. @cakimiNOW industri fashion bisa jadi peluang ekonomi #citayamfashionweek Citayam Fashion Week kalah kelas dengan Pagelaran Fashion jalanan ini #citayamfashionweek @Wandystjk Selamat untuk para penggagas #citayamfashionweek yang telah menginspirasi daerah lain	seru banget pagi ini tidak mau kalah dengan di jakarta mantan model senior bilbina on the spot membimbing kaum millennial untuk melenggok di catwalk industri fashion bisa jadi peluang ekonomi citayam fashion week kalah kelas dengan pagelaran fashion jalanan ini selamat untuk para penggagas yang telah menginspirasi daerah lain
Gue iseng2 ikutan nimbrung soal Citayam Fashion Week. Baru dimuat kemaren Di Jakarta Post. Enjoy ya gaes #CitayamFashionWeek #budaya #TikTok #influencer â€” https://t.co/rDIHeRZWcc	gue iseng ikutan nimbrung soal citayam fashion week baru dimuat kemaren di jakarta post enjoy ya gaes
https://t.co/nQS6dxgm0e	

Case folding is a preprocessing process that equates to the characters in the tweet. Case folding will change all the letters in the tweet to lowercase [17]. The next stage was tokenizing, separating each word in a sentence [18]. The results of word separation from Tokenizing were processed for further text analysis. Then, it was proceed with the filtering stage. Filtering is a step taken to select important words from token data [19]. Another term for common words or words that have no meaning was the stopword. Lastly, Stemming was conducted to change words into their primary forms by removing affixes. Affixes were removed before and after words [20]. The stemming stage was carried out using the help of a library in the Python3 programming language called Sastrawi. The results of the case folding, tokenizing, filtering, and stemming processes can be seen in Table 4

Table 4. Preprocessing Process

text_clean	text_preprocessed
seru banget pagi ini tidak mau kalah dengan di jakarta mantan model senior bilbina on the spot membimbing kaum millennial untuk melenggok di catwalk	['seru', 'banget', 'pagi', 'kalah', 'jakarta', 'mantan', 'model', 'senior', 'bilbina', 'on', 'the', 'spot', 'bimbing', 'kaum', 'millennial', 'lenggok', 'catwalk']
industri fashion bisa jadi peluang ekonomi citayam fashion week kalah kelas dengan pagelaran fashion jalanan ini	['industri', 'fashion', 'peluang', 'ekonomi'] ['citayam', 'fashion', 'week', 'kalah', 'kelas', 'pagelaran', 'fashion', 'jalan']
selamat untuk para penggagas yang telah menginspirasi daerah lain	['selamat', 'gagas', 'inspirasi', 'daerah']
gue iseng ikutan nimbrung soal citayam fashion week baru dimuat kemaren di jakarta post enjoy ya gaes	['gue', 'iseng', 'ikut', 'nimbrung', 'citayam', 'fashion', 'week', 'muat', 'kemaren', 'jakarta', 'post', 'enjoy', 'ya', 'gaes']

3.3. Resampling Technique

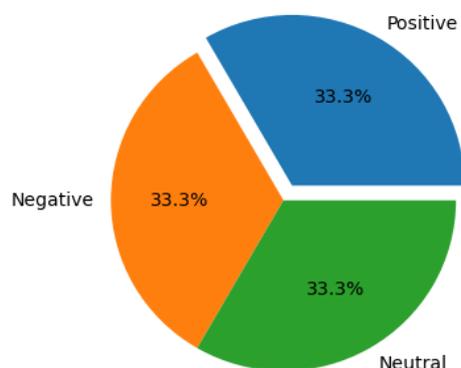


Figure 4. Percentage of Resample Data

The resample technique created a resample by manipulating data to balance the amount of data. The resample technique used was upsampling. At first, the data used were 1219 and divided into 851 positive tweets, 320 negative tweets, and 49 neutral tweets. After performing the resampling technique, the data became balanced with 835 positive tweets, 835 negative tweets, and 835 neutral tweets, so the total tweet data amounts to 2505. The following Figure 4 shows the Percentage of labels after Resample.

3.4. TF-IDF Weighting

Feature extraction is a method used to retrieve features that describe a characteristic of an object. In this study, the feature extraction used was Tf-Idf. Tf-Idf is a method that calculates word weight. Tf-Idf weighting was chosen because this method is efficient and accurate. In simple terms, the Tf-Idf method determines how often a word appears in a document.

3.5. Classification with Support Vector Machine

The processed data were further divided into 2, namely training data and testing data. The training data were used to model the classification, while the testing data were used to measure the performance of the classification. The amount of training data greatly affects the level of accuracy. In this study, three training and testing data comparisons were carried out. The classification used SVM with a kernel polynomial function that maps non-linear data to obtain a new learning model dataset for each trial. The important things that need to be done in the kernel are the selection of parameters from the SVM classification engine and poly kernel functions, namely parameters C and d (degree). The value of C in this study is 0.01, while the degree is 20.

The experimental results from the Support Vector Machine classification process for three comparisons of training data and testing data can be seen in Table 5. The distribution of training data and testing data with the best accuracy value is 90:10 with an accuracy value of 95,61%.

No	Training Data: Testing Data	Accuracy
1	90:10	95.61%
2	80:20	92.01%
3	70:30	90.82%

Based on Table 5. The results of the Support Vector Machine classifier showed that the comparison of training data and testing data was 90:10 is the best-split data, which has a higher level of model accuracy compared to comparisons of training data and other testing data with an accuracy of 95.61%.

3.6. Evaluation of Models

A matrix obtains the results of the classification with a size of 3x3 as a representative of the actual class and its predictions. In this study, the accuracy of the data is to be calculated as a comparison of training data and testing data 90:10 because it had higher data accuracy than comparisons of training data and other testing data. In this case, the data accuracy is calculated using the Confusion matrix.

Table 6. Confusion Matrix

Actual Data	Prediction data		
	Negative	Neutral	Positive
Negative	TNg	NgN	FN
Neutral	NNg	TN	NP
Positive	FP	PN	TP

Table 6 is explained that the Confusion matrix is the predicted result using the SVM classification engine, which measures the performance of each class by calculating precision, recall, and F1-score, description:

TNg	:	Negative predicted negative word class
NNg	:	Neutral predicted negative word class
FN	:	Negative word class predicted positive word
NgN	:	Neutral word class predictable negative word
TN	:	Neutral predictable word class
NP	:	Neutral word class predicted as positive word
FP	:	Positive predicted negative word class
PN	:	Neutral predicted positive word class
TP	:	Correct predicted word class has a positive value.

Precision calculates the prediction class accuracy according to the actual class for accurate results. The precision value can be determined using equation (2).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

Calculation of the precision of each word class can be determined using equations (3),(4), and (5).

$$Positive = \frac{TP}{TP + NP + FN} \quad (3)$$

$$Neutral = \frac{TN}{TN + PN + NgN} \quad (4)$$

$$Negative = \frac{TNg}{TNg + FP + NNg} \quad (5)$$

The recall is used to measure the sensitivity of the measurement to the dataset or the system's predictive ability according to the truth level. The recall value can be determined using equations (6),(7), and (8).

$$Positive = \frac{TP}{TP + PN + FP} \quad (6)$$

$$Neutral = \frac{TN}{TN+NP+NNg} \quad (7)$$

$$Negative = \frac{TNg}{TNg+FN+NgN} \quad (8)$$

	precision	recall	f1-score	support
Negative	1.00	0.89	0.94	84
Neutral	0.98	1.00	0.99	85
Positive	0.90	0.98	0.94	82
accuracy			0.96	251
macro avg	0.96	0.96	0.96	251
weighted avg	0.96	0.96	0.96	251

[[75 0 9]
[0 85 0]
[0 2 80]]

Figure 5. Performance Measure of Support Vector Machine

Negative Class

$$Precision\ Negative = \frac{75}{75 + 0 + 0} = \frac{75}{75} = 1.0$$

$$Recall\ Negative = \frac{75}{75 + 9 + 0} = \frac{75}{84} = 0.89$$

$$F1 - Score\ Negatif = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 1.0 \times 0.89}{1.0 + 0.89} = \frac{1.78}{1.89} = 0.94$$

Neutral Class

$$Precision\ Neutral = \frac{85}{85 + 2 + 0} = \frac{85}{87} = 0.98$$

$$Recall\ Neutral = \frac{85}{85 + 0 + 0} = \frac{85}{85} = 1.0$$

$$F1 - Score\ Neutral = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 0.98 \times 1.0}{0.98 + 1.0} = \frac{1.96}{1.98} = 0.99$$

Positive Class

$$Precision\ Positive = \frac{80}{80 + 0 + 9} = \frac{80}{89} = 0.90$$

$$Recall\ Positive = \frac{80}{80 + 2 + 0} = \frac{80}{82} = 0.98$$

$$F1 - Score\ Positive = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 0.90 \times 0.98}{0.90 + 0.98} = \frac{1.96}{1.88} = 0.94$$

Accuracy



(c)

Figure 6. Word Cloud (a) Positive (b) Negative (c) Neutral

4. CONCLUSION

Based on the classification results of a tweet using the Support Vector Machine classifier on the Citayam Fashion Week phenomenon, it is proved that the Support Vector Machine success to classify tweets with high accuracy. The sentiment analysis also shows that more people left positive sentiment tweets than negative and neutral sentiment, with 69,9%, 26,1%, and 4%, respectively. In other words, more people agree about the existence of Citayam Fashion Week than those who disagree and are neutral. Further positive, negative, and neutral sentiment analysis was conducted using a word cloud. The presentation of the word cloud can be used as an illustration of understanding the intent of the reviews and comments written by the general public. The results of this sentiment analysis can be used by the government and related parties to decide on policies to be taken about this phenomenon. This research can be used as a reference for further research and try to use other classification algorithm methods to be able to compare the model test results to get the best classification algorithm.

REFERENCES

- [1] F. Nazila, "Asal Usul Citayam Fashion Week yang Viral, Ide Inisiatif dari Jeje Slebew dan Bonge," *SuaraMerdeka.com*, 2022. <https://www.suaramerdeka.com/nasional/pr-043980114/asal-usul-citayam-fashion-week-yang-viral-ide-inisiatif-dari-jeje-slebew-dan-bonge?page=2> (accessed Aug. 18, 2022).
- [2] A. N. Dzulfaroh, "Citayam Fashion Week: Awalnya Tempat Nongkrong Rakyat Jelata, Kini 'Diperebutkan' Orang Kaya," <https://www.kompas.com/tren/read/2022/07/25/083718865/citayam-fashion-week-awalnya-tempat-nongkrong-rakyat-jelata-kini?page=all>, 2022. <https://www.kompas.com/tren/read/2022/07/25/083718865/citayam-fashion-week-awalnya-tempat-nongkrong-rakyat-jelata-kini?page=all> (accessed Sep. 12, 2022).
- [3] R. Siringoringo and J. Jamaludin, "Text Mining dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online," *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 1, pp. 41–48, 2019, doi: 10.34012/jutikomp.v2i1.456.
- [4] M. A. Maulana, A. Setyanto, and M. P. Kurniawan, "Analisis Sentimen Media Sosial Universitas Amikom," *Semin. Nas. Teknol. Inf. dan Multimed. 2018 Univ. AMIKOM Yogyakarta, 10 Februari 2018*, pp. 7–12, 2018.
- [5] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.

- [6] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [7] R. D. Himawan and E. Eliyani, "Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 1, p. 58, 2021, doi: 10.26418/jp.v7i1.41728.
- [8] K. Kelvin, J. Banjarnahor, E. I. -, and M. NK Nababan, "Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 47–52, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365.
- [9] Erlin, J. Sianturi, A. Hajjah, and Agustin, "Analisis Sentimen Prosesor AMD Ryzen menggunakan Metode Support Vector Machine," *SATIN-Sains dan Teknol. Inf.*, vol. 7, no. 2, pp. 129–141, 2021, doi: 10.33372/stn.v7i2.804.
- [10] I. surya kumala Idris, Y. A. Mustafa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, pp. 32–35, 2023.
- [11] Erlin, I. Suliani, H. Asnal, L. Suryati, and R. Efendi, "Sentiment Analysis for Abolition of National Exams in Indonesia using Support Vector Machine," *Eng. Lett.*, vol. 30, no. 4, pp. 1342–1352, 2022.
- [12] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [13] W. Nugraha and R. Sabaruddin, "Teknik Resampling untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, dan SVM Resampling Technique for Handling Class Imbalance in the Classification of Diabetes using C4.5, Random Forest, and SVM," *Techno.COM*, vol. 20, no. 3, pp. 352–361, 2021, [Online]. Available: <https://www.kaggle.com/uciml/pima-indians-diabetes-database>.
- [14] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Journal of Intelligent Systems and Computation 43," *J. Intell. Syst. Comput.*, pp. 43–49, 2019, [Online]. Available: <https://t.co/9WloaWpfD5>.
- [15] F. A. Sianturi, P. M. Hasugian, and A. Simangunsong, *Data Mining: Teori dan Aplikasi Weka*. IOCS Publisher, 2019.
- [16] S. N. Aprisadianti, "Analisis Sentimen Twitter terhadap Content Creator Sisca Kohl Menggunakan Regular Expression," no. 13519040, 2021.
- [17] D. Alita and A. R. Isnain, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *J. Komputasi*, vol. 8, no. 2, pp. 50–58, 2020, doi: 10.23960/komputasi.v8i2.2615.
- [18] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [19] I. Susianti, S. S. Ningsih, M. Al Haris, and T. W. Utami, "Analisis Sentimen Pada Twitter Terkait New Normal Dengan Metode Naïve Bayes Classifier," *Pros. Semin. Edusainstech FMIPA UNIMUS*, pp. 354–363, 2020, [Online]. Available: <https://prosiding.unimus.ac.id/index.php/edusaintek/article/view/576/578>.
- [20] I. Z. Simanjuntak, "Analisa Kombinasi Algoritma Stemming Dan Algoritma Soundex Dalam Pencarian Kata Bahasa Indonesia," *Inf. dan Teknol. Ilm.*, vol. 10, no. 1, pp. 24–30, 2022, [Online]. Available: <http://ejurnal.stmik-budidarma.ac.id/index.php/inti/article/view/5040>.
- [21] M. Galih Pradana, "Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining," *J. Ilm. Inform.*, vol. 8, no. 1, pp. 38–43, 2020.