Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 2 (2022) 269 - 274 ISSN Media Electronic: 2580-0760

Sentiment Analysis of Beauty Product E-Commerce Using Support Vector Machine Method

Muhammad Rio Pratama[1], Faza Abdillah Gunawan S.[2], Rafdi Reyhan Zhafari[3], Rendy[4], Helena Nurramdhani Irmanda[5]
[1,2,3,4,5]Department of Information System, Faculty of Computer Science,
University of Pembangunan Nasional Veteran Jakarta
[1]muhammadrp@upnvj.ac.id, [2]faza@upnvj.ac.id, [3]rafdirz@upnvj.ac.id, [4]rendy@upnvj.ac.id, [5]helenairmanda@upnvj.ac.id

Abstract

One of the important aspects of e-commerce is product quality, such as the quality of the trade and the application itself. Customers who buy goods will provide an assessment in the form of a review. If an item is dominated by negative reviews, other customers will be reluctant to buy at that store, so customers look for other stores and this affects the store's revenue. Therefore, the purpose of this study is to classify e-commerce beauty product reviews using the Support Vector Machine to create a model to categorize beauty product reviews and analyze accuracy. The research phase begins by collecting 50,000 datasets consisting of 35,000 training data and 15,000 test data. After the data is collected, the data labeling stage is carried out which is labeled positive and negative. Then the preprocessing step is carried out so that the data is ready to be processed in the feature extraction step. The feature extraction step aims to explore potential information that represents words. Furthermore, the resulting data is evaluated to obtain an accuracy value and determine whether the model made is feasible to use. The results showed that the Support Vector Machine can classify beauty product reviews well with an accuracy of 80.06%.

Keywords: classification, support vector machine, text mining, data mining.

1. Introduction

Along with the rapid development of technology in all parts of the world, it has had an impact, one of which is the emergence of many online or online buying and selling sites are popularly known as electronic commerce. Electronic commerce or better known as e-commerce is the activities of buying and selling goods, services, transmitting funds or data using electronic devices connected to the internet network. [1]. E-commerce is regulated in Law Number 7 of 2014 concerning Trade which prioritizes national interests and can be used to protect the domestic market and domestically made products, as a regulation of domestic trade, and provide protection to consumers. [2].

Today, e-commerce is preferred by consumers because they don't need to come directly to a physical store, and every year the number of e-commerce users continues to increase. In 2021 it will decrease. According to data from Bank Indonesia, Indonesia experienced a decline in online shopping transactions or e-commerce, which fell to Rp58,2 trillion in the third quarter of 2021, which previously reached Rp75,4 trillion in the second quarter of 2021 and Rp51,6 trillion in the second quarter of 2021. I-2021 [3]. Based on the decrease in the number of transactions in e-commerce, of course, there is crime or fraud in cyberspace that is increasing. According to previous research, there are 3 (three) forms of fraud, namely discount prices on Online Shopping Day (Harbolnas), goods that do not match orders or product catalogs, and sellers pretending to sell goods [4].

The number of e-commerce used by the people of Indonesia, more crimes in the network occur, so an analysis of the online market or e-commerce is needed based on consumer assessments of the products sold. Overcome this problem, consumer reviews of products on e-commerce can help other consumers be more careful in conducting transactions on the network. The e-commerce object in this research is Amazon. Reporting from an article written by the media in the coil.com network, Amazon is the largest e-commerce site in the world [5].

Based on licensing data issued by BPOM, there is an increase in licensing of beauty or cosmetic products, which is around 75.500 notifications during 2020, an

Accepted: 15-02-2022 | Received in revised: 23-04-2022 | Published: 29-04-2022

increase from 73.000 notifications in 2019 [6]. It shows that the interest and potential of the beauty product business continues to grow even though the pandemic has hit Indonesia since early March 2020 [7]. Based on this background, beauty products are then determined as the subject of this research because the development of beauty products is currently faster. It can be seen in everyday, beauty or cosmetic products are often used to support women to be more attractive.

Information contained in this beauty product review is valuable and can be used as a policy-making tool that is process through text mining. Data processing using text mining will be more effective when management and processing use software assistance [8]. Then sentiment analysis is carried out which consists of grouping the polarity of the text in a sentence or document to find out whether the opinion is positive or negative. By applying this method, expected to get the best final result, and it can be seen which direction the public sentiment has been processed [9]. Sentiment analysis is a computational study that collects opinions from individuals expressed in the form of the text [10]. Based on this background, this research aims to classify Amazon beauty product reviews and measure the accuracy of sentiment analysis.

In management and processing, several methods can be used and implemented such as Support Vector Machine, Decision Tree, Naïve Bayes, and K-Nearest Neighbors. From the choice of methods, Support Vector Machine is one of the methods that produce the best accuracy value. Accuracy result is obtain from previous research on text classification including news topic classification using SVM. The results of this research conclude that SVM is a classification method that provides the most accurate prediction result and is higher when compared to other methods, which is 92,24% [11]. Research on the classification of wood species using In addition, SVM has also been used to classify a person's facial characteristics which result in an accuracy value of 90% for the true detection level [13]. In the movie review dataset, researcher used 40.000 reviews to train the classifier and 10.000 reviews to test the model performance. Researcher found out that accuracy of the classification increases as researcher increase the feature size. Linear Support Vector Machine (LSVM) achieved an accuracy of 89.91% [14]. Besides, dataset was built from tweets discussing several issues and events in Saudi Arabia with a total of 4.200 tweets were annotated. The proposed Support Vector Machine model achieved an accuracy of 89.83% by applying several techniques including light stemming, feature extraction (Ngrams, emoji features and the tweet topic feature), model parameter optimization, and feature set reduction. The SVM classifier achieved an accuracy of 89.83% [15]. On the other hand, there is a low accuracy value in the implementation of this Support Vector Machine. The THUCNews dataset is generated according to the historical data filtering of the Sina News RSS subscription channel from 2005 to 2011. It contains 74 million news documents or 2.19 GB of data. 19.630 features were received. 1.998 samples were selected for training and 509 were used for testing. Sentiment analysis achieved an accuracy of 41.90%, so the researcher chose the multi-class method to improve the result [16]. Based on these studies, the application of this research will implement the support vector machine method for sentiment analysis, review of beauty products in electronic commerce.

2. Research Methods

The research method used in this research to conduct sentiment analysis of beauty product reviews on electronic commerce is using the support vector machine method, the steps of the research carried out can be seen in Figure 1.

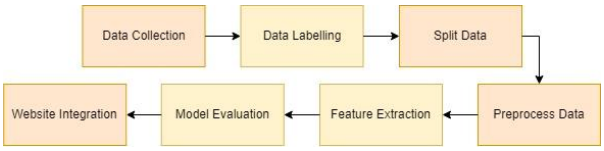

Figure 1. The Research Steps

Based on Figure 1, there are 8 (eight) steps to be carried out in this research, including data collection, data labeling, split data, preprocess data, feature extraction, model evaluation, and website integration. The discussion of each step is as follows:

2.1 Data Collection

The data used as input in this system is a review dataset obtained from beauty product review data on Amazon, the link to access the dataset is as follows http://jmcauley.ucsd.edu/data/amazon/links.html. The dataset that will be collected and processed is 50,000 records review.

```
#   Column
---  ------
0   id
1   reviewerID
2   reviewerName
3   reviewText
4   overall
5   summary
6   unixReviewTime
```

Figure 2. Data Collection Feature

Based on Figure 2, of all the existing features, namely “id”, “reviewerID”', “reviewerName”, “reviewText”', “overall”, “summary”, and “unixReviewTime”, the researcher only took the features “reviewText” and “overalls” in this research.

2.2 Data Labeling

DOI: https://doi.org/10.29207/resti.v6i2.3876

After the preprocessing step, the next step is labeling the data. In this step, the data must have a label based on the rating value. The rating in question is that if the user gives a rating of 4 or 5, it will show the positive class, and a rating of 3 and below will show the negative class, shown in Figure 3.

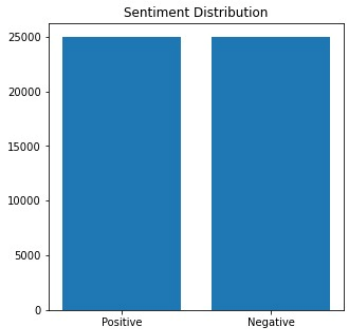

Figure 3. Sentiment Distribution

Figure 3 has been labeled with a positive class and a negative class. The positive class used in this research was 25.000 records, and the negative class used was 25.000 records.

2.3 Split Data

The next step is the split or separation of the data used to see the proportion of the review data. In this research, 70% of the training data and 30% of the testing data were used from the total review data of 50.000 records.

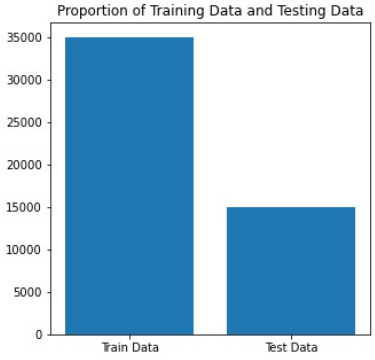

Figure 4. Proportion of Training Data and Testing Data

Figure 4 shows the proportion of training data and test data. 35.000 training data were obtained and 15.000 testing data were obtained. The determination of training data and testing data is carried out randomly so that the proportion between product review categories is balanced.

2.4 Preprocess Data

After the data is split, the preprocessing step is carried out so that the data is ready to be processed at the next step. The preprocess step consists of data preparation, tokenization, data cleansing, case folding, and stop words removal. The discussion of each steps is as follows:

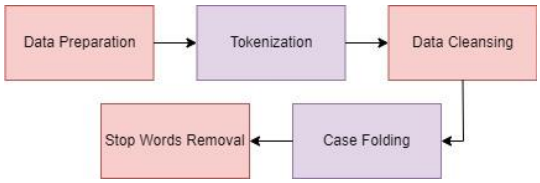

Figure 5. Preprocess Data Steps

Based on Figure 5, the first preprocessing step is data preparation which aims to prepare data to be processed at a later step. The second step is tokenization which serves to break sentences into tokens or single words. The third step is data cleansing which aims to turn dirty data into quality data so it can produce accurate information. The fourth step is case folding which changes all words in the review text into lowercase letters, and other characters can be removed. Then the last step is stop words removal, which is removing a collection of words that are considered meaningless.

2.5 Feature Extraction

After the data is labeled, the next step is feature extraction. In this step, the text is extracted to be numeric because the computer does not process data other than numeric data. Feature extraction is used to explore potential information and represent words in the feature vector. This feature vector will use as input for the classification method in the next step [14]. Techniques that can be applied in this feature extraction are Term Frequency and Inverse Document Frequency which function to count the occurrences of each word in the document and throughout the document. TF value compared to IDF value [15].

The final result of this TF-IDF value is an output matrix containing unique words from the dataset and values from the calculation of TF-IDF features. The calculation of word weight can be seen in Equation 1 and Equation 2[16].

$$W_{ij} = tf_{ij}\, x\, idf_i \quad (1)$$

$$W_{ij} = tf_{ij}\, x\, log(D/df_j) \quad (2)$$

Based on Equation 1 and Equation 2, it can conclude that w_{ij} is the word from all documents, while tf_{ij} is the frequency of occurrence of words in a document idf_j is the inverse of the number of documents containing words.

2.6 Model Evaluation

The classification results obtained are then evaluated to obtain an accuracy value which is then analyzed and shows the classification model made is feasible or not. Describe the performance of the classification model, the 2x2 confusion matrix table was created which was applied to this research, which is shown in Table 1.

DOI: https://doi.org/10.29207/resti.v6i2.3876

Table 1. Confusion Matrix 2x2

		Prediction Results	
		P	N
Accuracy	P	TP	FN
Results	N	FP	TN

Based on Table 1, a 2x2 confusion matrix consists of 2 parts, namely the prediction results and the actual results. Prediction results are results obtained from the SVM classification which consists of 2 (two) classes, namely P (positive) and N (negative). Cases are into four grades, namely: TP, FP, FN, and TN. TP is a correct positive prediction, FP is a false positive prediction, FN is a false negative prediction, and TN is a correct negative prediction [17].

The accuracy value is calculated ffter creating a 2x2 confusion matrix to assess the performance of the SVM to classify, while the formula is determined in Equation 3 [17].

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad (3)$$

The accuracy value in Equation 3 is the value obtained from the quotient of all the correct test data with the overall testing data. The TP value is the true positive predictive value and the TN value is the true negative predictive value. While the P value and N value are the records used, namely testing data.

In addition to accuracy, to evaluate the success of the prediction model, it is also possible to calculate precision, recall, and specificity for each review class shown in Equation 4, Equation 5, and Equation 6.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{P} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (6)$$

Precision in Equation 4, is used to calculate the ratio of the number of data that is predicted to be positive compared to the overall data that is predicted to be positive. Recall in Equation 5 is used to get a comparison of the number of correct positive predictions compared to the total number of positive classes. Specificity in Equation 6 is used to get a comparison of the number of correct negative predictions compared to the total number of negative classes.

2.7 Website Integration

The last step is website integration as our contribution in this research, namely by integrating the system that has been created to process data with a user interface to make it more accessible to users. This is an update through a website intermediary to see the results of sentiment analysis based on reviews of e-commerce beauty products that have not been found in previous research. In this step, we use flask and pickle as a medium to connect the program and the designed website.

3. Results and Discussions

The results and discussions of the sentiment analysis review of beauty products using SVM are explained based on the block diagram that has been designed previously in chapter 2 of the research method section.

3.1 Preprocess

The preprocessing consists of six steps, namely data preparation, tokenization, data cleansing, case folding, and stop words removal using Python programming language with NLTK library.

The data preparation step is the collection of e-commerce beauty product review data from Amazon by users, taking one review in Table 2.

Table 2. Input and Output of Data Tokenization

Input	Output
Very oily and creamy. Not at all what expected.…	'very', 'oily', 'and', 'creamy', 'not', 'at', 'all', 'expected'

Based on the Table 2, the data tokenization step or split into tokens/words from the data review.

The data cleansing step or deletion becomes clean data that can be used for the next step. The case folding step or changing the text of the product review content to lowercase and eliminating characters other than letters, the results are in Table 3.

Table 3. Output of Data Cleansing

Output
very oily and creamy not at all expected

The stop words removal step or remove words that are considered meaningless using NLTK. Some stop words in NLTK include: ['in', 'on', 'at', 'and', 'but', …]. If it is implemented, an example of applying NLTK in a sentence, then the words "and" and "at" are lost, so the results are in Table 4.

Table 4. Output of Stop Words Removal

Output
'oily, 'creamy', 'expected'

3.2 Feature Extraction

Feature extraction in this research uses TF-IDF weighting with formulas such as in Equation 1 and Equation 2. TF-IDF will assess how important a word is in a document. To perform calculations using TF-IDF, you can use the library in the Python Sklearn, namely *TfidfVectorizer()*.

DOI: https://doi.org/10.29207/resti.v6i2.3876

The number of features that will be used is selected using the maximum top feature sorted by term frequency of all review data. In this study, the number of features that are not too large is needed but allows the classification model to function properly. Therefore, the maximum used feature is 50.000 records.

3.3 Model Evaluation

The classification process carried out using the support vector machine method begins with training data of 35.000 records. In this research, the classification process uses the library support vector machine and sklearn python. Therefore, this classification includes a binary classification that describes 2 (two) categories, namely the positive class and the negative class of product reviews in e-commerce.

The model obtained from the training process will be stored and used for the testing process. The testing process is carried out on 30% of the total data used or as many as 15.000 records. In the testing process, the class predicted by the model will be obtained. This class will be compared to the actual class. To find out whether the model made is successful or not, an evaluation stage is needed. The evaluation method that can be done is by calculating the accuracy value. As described in Table 1, the test results data are listed in the form of a confusion matrix in Table 5.

Table 5. Confusion Matrix Classification Results

		Prediction Results	
		P	N
Accuracy	P	6.061,5	1443
Results	N	1.548	5.947,5

Based on the classification results in Table 5, there are 4 (four) types of possible cases that will occur, including:

TP (True Positive) is the data for the positive class that is predicted with a true value of 6.061,5 or 40,41%. TN (True Negative) is data for the negative class that is predicted to be correct with the true value with a total of 5.947,5 or 39,65%. FP (False Positive) is data for negative class but is predicted as positive class data with a total of 1.443 or 9,62%. FN (False Negative) is data for positive class but predicted as negative class data with a total of 1.548 or 10,32%.

Based on the values of TP, FP, FN, and TN for each positive class and negative class, the values of precision, recall, and specificity are calculated in the calculations in Equation 4, Equation 5, and Equation 6 are shown in Table 6.

Table 6. Value of Precision, Recall, and Specificity Each Class

Class	Precision	Recall	Specificity
Positive	80,05%	81%	80,77%
Negative	80,04%	79%	80,47%

Table 6, it can be seen that the precision value explains that the percentage of predictions for the correct review class from the overall class predicted by the positive class is 80,05% and the percentage of predictions for the negative class is 80,04%. The highest precision value was obtained by the positive class.

The recall value explains that the percentage of correct predictions of the review class from the overall class that is predicted to be positive is 81% and the percentage of negative class predictions is 79%. The highest recall value was obtained by the positive class.

The specificity value explains that the percentage of predictions for the correct review class from the overall class that is predicted to be positive is 80,77% and the percentage of negative class predictions is 80,47%. The highest specificity value was obtained by the positive class.

In addition to precision, recall, and specificity based on the calculations in Equation 3, in the positive class and negative class, the overall accuracy value for the product review classification model using the support vector machine is calculated, namely the sum of TP + TN respectively 6.061,5 and 5.947,5 so the resulting 12.009 divided by the total number of test data of 15.000. Thus, the accuracy value of the calculation can be generated at 80,06% and it can be concluded that the product review classification model using the support vector machine method can work well.

3.4 Website Integration

The website integration process begins with entering sentences and will classify them into positive or negative classes. In this research, flask and pickle are used as media to connect programs and websites that are created. The results of website integration can be seen in Figure 6.

Figure 6. Website Integration

Based on Figure 6, the input given is = “Very oily and creamy. Not at all what expected…” and resulted in a classification of sentiment analysis, namely = “Negative”.

4. Conclusion

Based on the results of the positive and negative classes classification research with a dataset of 50.000 records consisting of 35.000 training data and 15.000 testing

DOI: https://doi.org/10.29207/resti.v6i2.3876

data, it can be concluded that the support vector machine can proper classify the review class with an accuracy value of 80,06%. On the other hand, there are values of precision, recall, and specificity in the positive class which is higher than the negative class by 80,05%, 81%, and 80,77%. However, the negative class has a fairly high value of precision, recall, and specificity and is considered good and acceptable. The highest values of precision, recall, and specificity is influenced by the amount of data as much as 15.000. Suggestions for further research are to add more data for both positive and negative class categories. Thus, it is expected that the values of accuracy, precision, recall, and specificity will increase.

Acknowledgment

The researchers thank PT Huawei Tech Investment, the Ministry of Education, Culture, Research and Technology, and the Faculty of Computer Science, University of Pembangunan Nasional Veteran Jakarta who has assisted in the process and funded this research.

Reference

[1] R. W. Primadineska, "Pengaruh Penggunaan Sistem Pembayaran Digital terhadap Perilaku Beralih di Era Pandemi COVID-19," *Telaah Bisnis*, vol. 21, no. 2, p. 89, 2021, doi: 10.35917/tb.v21i2.215.

[2] Undang-Undang Republik Indonesia, "Undang-Undang Republik Indonesia No. 7 Tahun 2014 Tentang Perdagangan," *LN.2014/No. 45, TLN No. 5512, LL SETNEG 56 HLM*, pp. 1–56, 2014, [Online]. Available: https://peraturan.bpk.go.id/Home/Details/38584/uu-no-7-tahun-2014.

[3] Putri, Cantika Adinda. 2021, November 18. *Kuartal III-2021, Transaksi E-Commerce Turun Jauh ke Rp 58 T* [Online]. Available: https://www.cnbcindonesia.com/tech/20211118173007-37-292654/kuartal-iii-2021-transaksi-e-commerce-turun-jauh-ke-rp-58-t.

[4] S. N. Fauzi and L. Primasari, "Tindak Pidana Penipuan dalam Transaksi di Situs Jual Beli Online (E-Commerce)," *Recidive*, vol. 7, no. 3, pp. 250–261, 2018.

[5] Izzudin. 2019, Februari 18. *Fakta Menarik Amazon.com, Situs E-commerce Terbesar Di dunia* [Online]. Available: https://kumparan.com/temali/fakta-menarik-amazon-com-situs-e-commerce-terbesar-dunia-1550208014712120226/full.

[6] Autoridad Nacional del Servicio Civil, "**済無**No Title No Title No Title," *Angew. Chemie Int. Ed. 6(11), 951–952.*, pp. 2013–2015, 2021.

[7] Nabila mecadinisa, "Melihat Perkembangan Industri Kosmetik di Indonesia Pasca Covid-19," *Fimela.Com*. 2021, [Online]. Available: https://www.fimela.com/beauty/read/4578615/melihat-perkembangan-industri-kosmetik-di-indonesia-pasca-covid-19.

[8] M. G. Pradana, "Penggunaan fitur wordcloud dan document term matrix dalam text mining," *J. Ilm. Inform.*, vol. 8, no. 1, pp. 38–43, 2020.

[9] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komputer.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[10] C. S. K. Aditya, M. Hani'ah, A. A. Fitrawan, A. Z. Arifin, and D. Purwitasari, "Deteksi Bot Spammer pada Twitter Berbasis Sentiment Analysis dan Time Interval Entropy," *J. Buana Inform.*, vol. 7, no. 3, pp. 179–186, 2016, doi: 10.24002/jbi.v7i3.656.

[11] L. G. Irham, A. Adiwijaya, and U. N. Wisesty, "Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine," *J. Media Inform. Budidarma*, vol. 3, no. 4, p. 284, 2019, doi: 10.30865/mib.v3i4.1410.

[12] N. Neneng, N. U. Putri, and E. R. Susanto, "Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern," *Cybernetics*, vol. 4, no. 02, pp. 93–100, 2021, doi: 10.29406/cbn.v4i02.2324.

[13] A. Setiyono and H. F. Pardede, "Klasifikasi Sms Spam Menggunakan Support Vector Machine," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 275–280, 2019, doi: 10.33480/pilar.v15i2.693.

[14] B. Das and S. Chakraborty, "An Improved Text Sentiment Classiciation Model Using TF-IDF and Next Word Negation"", 2018, arXiv preprint arXiv:1806.06407.

[15] S. N. Alyami And S. O. Olantunji, "Application Of Support Vector Machine For Arabic Sentiment Classification Using Twitter-Based Dataset," *J. Information & Knowledge Management*, vol. 19, no. 01, 2020, 2040018.

[16] G. Wang and S. Y. Shin, "An Improved Text Classification Method For Sentiment Classification," *J. Information and Communication Convergence Engineering*, vol. 17, no. 1, pp. 41-48, 2019.

[17] H. N. Irmanda and Ria Astriratma, "Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 915–922, 2020, doi: 10.29207/resti.v4i5.2313.

[18] P. M. Prihatini, "Implementasi Ekstraksi Fitur Pada Pengolahan Dokumen Berbahasa Indonesia," *J. Matrix*, vol. 6, no. 3, pp. 174–178, 2016.

[19] M. Z. Naf'an, A. Burhanuddin, and A. Riyani, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," *J. Linguist. Komputasional*, vol. 2, no. 1, pp. 23–27, 2019, doi: 10.26418/jlk.v2i1.17.

[20] A. Suresh, "What is a confusion matrix?," *Medium*. 2020, [Online]. Available: https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5.

DOI: https://doi.org/10.29207/resti.v6i2.3876