

Sentiment Analysis of Brand Ambassador Influence on Product Buyer Interest Using KNN and SVM

by Jsslynn Jsslynn

Submission date: 17-Apr-2024 10:30PM (UTC+0700)

Submission ID: 2180401220

File name: IJAIDM_Paper.docx (1.54M)

Word count: 4161

Character count: 24178

Sentiment Analysis of Brand Ambassador Influence on Product Buyer Interest Using KNN and SVM

¹Natasya Kurnia Putri, ^{2*}Anik Vega Vitianingsih, ³Slamet Kacung, ⁴Anastasia Lidya Maukar, ⁵Verdi Yasin

^{1,2,3,5}Informatics Department, Universitas Dr. Soetomo, Surabaya, Indonesia

⁴Industrial Engineering Department, President University, Bekasi, Indonesia

⁵Informatics Department, Sekolah Tinggi Manajemen Informatika dan komputer Jayakarta, Indonesia

Email: ¹natasyaaptr23@gmail.com, ²vega@unitomo.ac.id*, ³slamet@unitomo.ac.id, ⁴almaukar@president.ac.id,

⁵verdiyasin29@gmail.com

Article Info

Article history:

Received May 12th, 2019

Revised Jun 20th, 2020

Accepted Jul 26th, 2020

Keyword:

Sentiment Analysis

Brand Ambassador Influence

Product Buyer Interest

K-Nearest Neighbor

Support Vector Machine

ABSTRACT

In the dynamic marketing field, companies usually use strategies involving celebrities or influencers to promote their products or brands. The currently popular strategy is using Korean boy bands as brand ambassadors. This collaboration certainly gets a lot of opinion responses through tweets on X app social media. This research aims to analyze sentiment to determine how the product buyer's interest responds to brand suitability, brand image management, and the influence of issues that arise in this collaboration. The research stages consist of data collection, pre-processing, labeling, weighting, and classification with K-Nearest Neighbor and Support Vector Machine and performance evaluation using a confusion matrix. The dataset used was 696 tweets taken using web scrapping techniques. This research uses the Lexicon-based method to divide the dataset into positive, negative, and neutral classes. The SVM method shows superior test results by achieving an accuracy rate of 83.34% compared to the KNN method, which produces an accuracy value of 71.2% in its calculations.

Copyright © 2022 Puzzle Research Data Technology

Corresponding Author:

Natasya Kurnia Putri,

Informatics Department, Universitas Dr. Soetomo

Email: natasyaaptr23@gmail.com

DOI: <http://dx.doi.org/10.24014/ijaidm.v2i2.xxxx> (10 pt)

1. INTRODUCTION

In the current marketing era, competition between business actors is increasing, thus encouraging companies to explore innovative and successful marketing strategies. Marketing success depends on product excellence and how brands build and maintain relationships with consumers [1]. Marketing strategies utilizing brand ambassadors are becoming increasingly popular to create a strong brand identity and influence consumer perceptions of a product [2]. One example of a Brand Ambassador with global appeal is NCT, a K-pop music group with a wide and diverse fan base [3]. Lemonilo, known for its emphasis on health and quality, utilizes the influence of NCT Brand Ambassadors to expand its reach and increase its appeal [4]. The success of this strategy depends not only on NCT's popularity but also on consumers' perceptions and sentiments towards the association between the Brand Ambassador and the promoted product.

Although the role of Brand Ambassadors has received significant attention in marketing literature, there is still a need for further research to understand their impact and purchase intention. The X app is a social media platform where users can share their opinions on various topics [5]. The X app has become a widely accepted digital communication instrument, gaining significant popularity among internet users [5]. The X app is considered a vast source of information for sentiment analysis and decision-making [6]. Sentiment analysis can be the right approach to understanding how the public responds and consumer perceptions of Brand Ambassadors. The research to be conducted is sentiment analysis. Computerization is required to address the challenge of manually analyzing large amounts of data from the X app. Analysis can be done more effectively and efficiently using Python [7].

Previous research has looked at sentiment analysis concerning cyberbullying. In an investigation on cyberbullying on Twitter using SentiStrength, [8] found that 45.3% of cases involved bullying, 36.2% involved non-bullying, and 18.5% involved neutral behavior. Using Support Vector Machine Algorithm Evaluation, Naive Bayes, Decision Trees, and Gradient Boosting, researchers [9] investigated Twitter ChatGPT. Of the algorithms, the Support Vector Machine algorithm with a 90:10 data division had the highest accuracy value with an 80% achievement. Researchers [10] investigated Motorku using the Naive Bayes Classifier technique in the X app. The study's Naive Bayes technique produced findings with 76% accuracy, 76% precision, and 97% recall. There has already been a study on brand ambassadors concerning beauty reviews. Using the Naive Bayes approach, researchers [11] conducted sentiment analysis in the study and achieved an accuracy rate of approximately 82.65%. Additionally, the accuracy values produced by using the Support Vector Machine approach were approximately 83.60%. Scholars concentrate on sentiment analysis about brand ambassadors, particularly within the framework of Tokopedia e-commerce. Without normalizing the data, using Bayesian Network classification in a study by [12] produced accuracy results of roughly 66.6667%, precision of 68.1%, and recall of 66.7%. As a result of data normalization, accuracy rose to 76.5556%, recall was 76.6%, and precision was 77.4%.

This research aims to conduct sentiment analysis of buyer interest responses to products using the K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) methodologies. Combining these two methods can provide a thorough understanding of consumer sentiment toward products, which is influenced by the presence of Brand Ambassadors. The dataset for this research is obtained from the X app social media through a web scraping technique. The next stage will involve data pre-processing, which includes cleaning, case folding, tokenizing, normalizing, filtering, and stemming. After the pre-processing process is complete, a labeling process is performed to ascertain the sentiment class, specifically categorizing it as positive, negative, or neutral, using a Lexicon-based dictionary. The next stage involves TF-IDF weighting, after which the system will process the data using KNN and SVM methods. The last stage involves model assessment, where the evaluation uses a Confusion Matrix to generate metrics such as accuracy, precision, and recall.

The results of this study can significantly contribute to the positive or negative identification of products that the image of Brand Ambassadors may influence. In addition, this research is useful for providing an in-depth understanding of consumer responses and perceptions of the influence of Brand Ambassadors on products. Through sentiment analysis, readers can better understand the positive, negative, or neutral feelings associated with the interaction between Brand Ambassadors and shoppers.

2. RESEARCH METHOD

In order to determine which method has the best accuracy, this study compares and contrasts the K-Nearest Neighbours and Support Vector algorithms. Reviews talk about how brand ambassadors affect consumers' curiosity. Figure 1 illustrates the several stages involved in doing research.

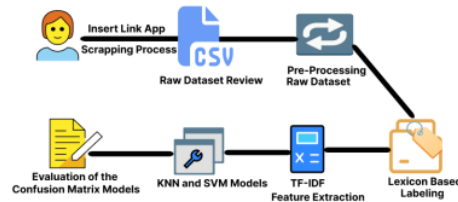


Figure 1. Research Process Design

Data collection for the study starts with online scraping methods. After that, it moves on to the pre-processing phase, which consists of the following six procedures: cleaning, tokenizing, normalization, folding cases, filtering, and stemming. After that, the data is labeled to differentiate between neutral, negative, and positive groups. The TF-IDF approach is used to extract features following the labeling procedure. Subsequently, the K-Nearest Neighbour and Support Vector Machine are used for classification. Evaluation is performed using the Confusion Matrix to assess accuracy, precision, recall, and F1 score after receiving the data from KNN and SVM. Following precise data collection, an analysis is performed to draw findings.

2.1. Data Collection

Data collection and analysis are carried out at this stage using *web scraping* techniques such as Python and the Twitterscraper repository. Datasets were collected from the X social media platform using the keyword "lemonilo" and hashtags such as #NCTDream, #Lemonilo, and #LemoniloXNCTDream. The total data collected reached 696 tweets, which were then saved in *.csv file format.

15

2.2. Pre-processing Data

The pre-processing data takes place before sentiment analysis. After crawling, raw datasets containing a great deal of noise and still unstructured text are the result [13]. It must, therefore, be transformed into data that is simpler to comprehend. Preparing the data and eliminating unnecessary material is the first step in obtaining more precise details. The pre-processing phase involves several procedures, as Figure 2 illustrates.

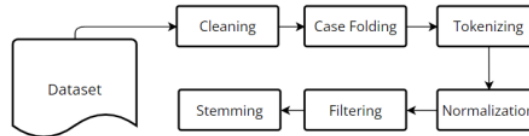


Figure 2. Pre-processing Flow

- *Cleaning*: The process of removing punctuation and removing foreign characters such as numbers, usernames (@), URLs (https://), and hashtags (#) [14].
- *Case Folding*: Use case folding to turn all capital characters into lowercase letters. [15].
- *Tokenizing*: The process involves segmenting the words in each sentence and creating meaningful individual words.
- *Normalization*: Replacing inappropriate words with more suitable or superior alternatives [16].
- *Filtering*: The process of removing irrelevant or unnecessary text in the context of sentiment analysis [17].
- *Stemming*: The process of changing words in the text with affixes into base or root words [18].

2.3. Data Labeling

The second level of pre-processing involves classifying the tweet data into three categories: positive, neutral, and negative sentiment. Calculating score values with the Lexicon dictionary allows for labeling the data processing automation. According to the findings of the score computation, a sentence will be classified as a neutral class if the score is equal to zero, as a negative class if the score is less than zero, and as a positive class if the score is more significant than zero [19]. Equation (1) can therefore be used to obtain the sentiment score.

$$Skor = (\sum \text{kata positif} - \sum \text{kata negatif}) \quad (1)$$

2.4. Word Weighting

The feature extraction approach known as TF-IDF assigns a weight to each extracted feature on a word inside a phrase [20]. Weights are assigned to each word's significance and generality within a phrase to assess each word's significance and generality. The frequency with which these terms appear in the document affects the TF-IDF value [20]. Consequently, the weight of the association between the words and the document increases as the frequency of words in a document increases and appears in more papers. On the other hand, the weight of the association between a word and the document likewise diminishes when the term appears in the document less frequently [20]. Equation 2 calculates the TF-IDF weight from a mathematical perspective. Where

$$TF.IDF_{std}(t) = tf_d^t \times \log \frac{N}{df_t} \quad (2)$$

2.5. Model Analysis

After the weighting stage, the classification model is carried out using the KNN and SVM algorithms, which are carried out separately and aim to distinguish which model produces the most optimal accuracy value. 1). *KNN*: K-Nearest Neighbours is a machine learning approach that uses datasets to classify or perform regression. [21]. The KNN process involves classifying data by categorizing objects according to the closest neighboring data. The KNN method works by finding the K nearest neighbors of the data to be classified or regressed and then determining the class or regression value based on most of the class or regression value of the k nearest neighbors [22]. Applying the KNN method in text classification can provide more optimal results by incorporating the *Cosine Similarity* formula. This formula gives weight to each word in the analyzed text document, resulting in a more optimal value [23]. The KNN algorithm in Figure 3 has several advantages, such as being easy to implement and understand, not requiring certain assumptions about data distribution, and being suitable for data with many features.

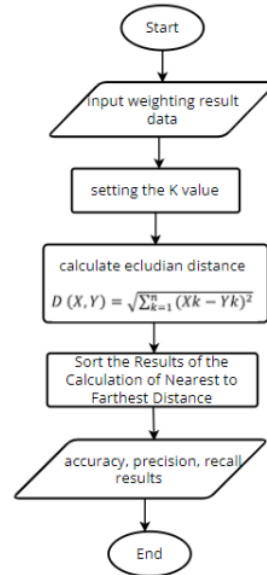


Figure 3. KNN Flow

The basic formula of K-Nearest Neighbor can be seen in Equation (3). Where, D is the distance between two points x and y , the X variable denotes test data, while Y represents sample data and n represents the data dimension.

$$D(X, Y) = \sqrt{\sum_{k=1}^n (xk - yk)^2} \quad (3)$$

2). SVM

Figure 4 shows the SVM flow through the process of finding the best hyperplane to maximise the distance between two data classes.

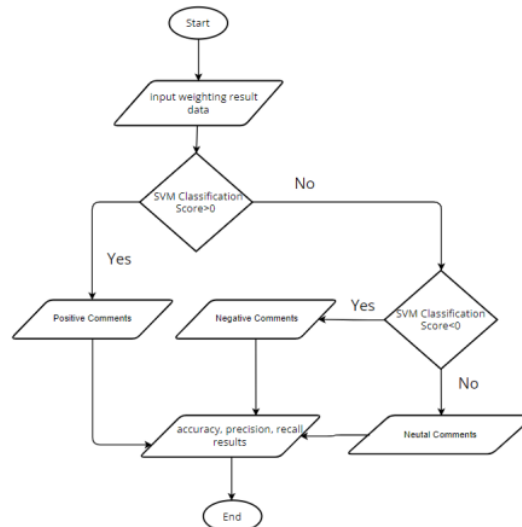


Figure 4. SVM Flow

This *hyperplane* acts as a decision boundary that separates the data into two distinct classes [24]. In addition, SVM also utilizes kernels to implicitly transform the data into a higher feature space, allowing the

model to handle nonlinear data. Using these kernels will enable SVMs to capture complex relationships between data features, which in turn enhances the capability of SVMs in handling nonlinear classification problems [25]. Equation (4) provides the basic formula for the Support Vector Machine. Where x is the accessible data, b is the switch value, and w is the weight vector.

$$f(x) = w \cdot x + b \quad (4)$$

2.5. Evaluation

The last step involves evaluating the model's performance to test the categorization outcomes and gauge the accuracy of the system. The Confusion Matrix is used in the evaluation phase after successfully acquiring the KNN and SVM classification models. One way to assess the classification process's success or failure was to utilize a confusion matrix [26]. The criteria for assessing categorization performance are recall, precision, and accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

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

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

Where, true positive (TP) indicates a situation where the model successfully predicts the data as positive, and it is positive; false positive (FP) occurs when the model predicts the data as positive, but it is negative; true negative (TN) involves a condition where the model successfully predicts the data as negative, and it is negative; while False Negative (FN) occurs when the model incorrectly predicts the data as negative, even though it is positive.

3. RESULTS AND ANALYSIS

This research successfully collected 695 tweets utilizing the keyword "lemonilo", and several hashtags such as #NCTDream, #Lemonilo, and #LemoniloXNCTDream relating to the appointment of NCT Dream as Lemonilo brand ambassador on the X apps account through the application of crawling techniques. Table 1. is the result of the data collection process carried out by crawling Twitter.

Table 1. Comment Data Sample

Comment
<i>Mie nya enak, modelnya cakep yang satu jodoh saya tuh, bintang lima * * * * *</i>
<i>Asli enakan indomie! sedaap sebenarnya mungkin krna belum terbiasa sama lemonilo.tapi nnti bakal terbiasa kok 😊</i>
<i>Ini pada ngomong mau beli sekardus cuma buat incar PC doang 😊</i>
<i>Enakkk. Aku bukan penyuka mie instanst, jarang makan karna lemak banget. Tp lemonilo itu enak, lebih sehat juga</i>

3.1. Pre-processing Data

Once the data collection stage is completed, processing cannot be done immediately due to considerable noise in the data set. As a result, pre-processing plays a crucial role in data cleaning, which later tries to remove unneeded attributes. The outcome of the data-cleaning procedure is shown in Table 2. As the table demonstrates, the data has undergone many processing steps, including cleaning, case folding, tokenization, normalisation, filtering, and stemming.

Table 2. Pre-processing Results

Dataset	
<i>Mie nya enak, modelnya cakep yang satu jodoh saya tuh, bintang lima * * * * *</i>	
Preprocessing	
Cleansing	<i>Mie nya enak modelnya cakep yang satu jodoh saya tuh, bintang lima</i>
Case Folding	<i>mie nya enak modelnya cakep yg satu jodoh saya tuh bintang lima</i>
Tokenization	<i>['mie', 'nya', 'enak', 'modelnya', 'cakep', 'yg', 'satu', 'jodoh', 'saya', 'tuh', 'bintang', 'lima']</i>
Normalization	<i>['mie', 'nya', 'enak', 'modelnya', 'cakep', 'yang', 'satu', 'jodoh', 'saya', 'tuh', 'bintang', 'lima']</i>
Filtering	<i>'mie', 'enak', 'modelnya', 'cakep', 'satu', 'jodoh', 'saya', 'bintang', 'lima'</i>
Steaming	<i>'mie', 'enak', 'model', 'cakep', 'satu', 'jodoh', 'saya', 'bintang', 'lima'</i>

Title of manuscript is short and clear, implies research results... (First Author)

The next stage is data classification, done by applying the KNN algorithm model. Figure 8 shows the classification results generated by the K-Nearest Neighbor algorithm and visualized in a bar chart, showing that the number of tweet frequencies expressing negative sentiment is 3, positive sentiment is 60, and neutral sentiment is 75.

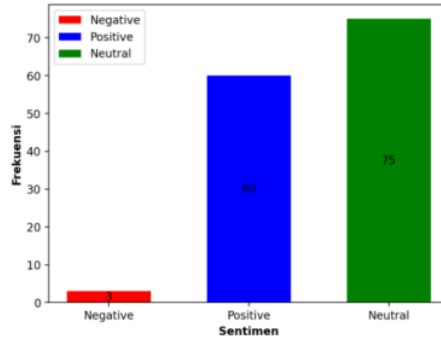


Figure 8. KNN Classification Result

3.4. SVM Classification

In the next step, this study uses the KNN algorithm model and SVM data classification. Figure 9 displays the categorization results obtained from the Support Vector Machine algorithm as a bar chart, indicating that there are, on average, 69 tweets with positive sentiment and 69 with negative sentiment.

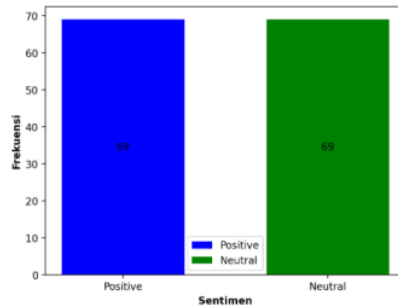


Figure 9. SVM Classification Result

3.5. Model Evaluation

The classification procedure evaluation is the final stage, where the effectiveness of the applied model is examined. This evaluation ensures the model's reliability by measuring accuracy, precision, and retrieval using the Confusion Matrix.

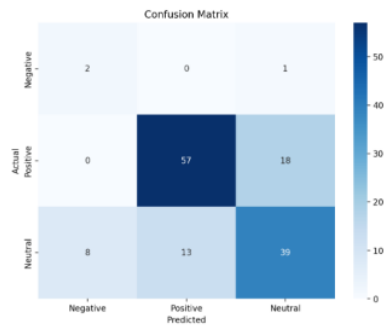


Figure 10. The Confusion Matrix Evaluation Results Using the KNN

The KNN approach performs well, with accuracy reaching 71%, precision reaching 56%, recall reaching 69%, and f1-score reaching 78%, according to the test data in Figure 11.

In this study, a comparison of the results of the K-Nearest Neighbor and Support Vector Machine algorithms was carried out using a streamlet. The accuracy results of each algorithm are 71% for KNN and 83% for SVM. Based on these findings, it can be concluded that SVM has a superior level of accuracy when compared to KNN. Analysis of buyer responses shows that the percentage of tweets expressing positive sentiment is more dominant than those expressing negative sentiment. Specifically, tweets regarding how NCT Dream's brand ambassador influenced the purchase intention of Lemonilo products showed a positive response. The word "delicious" is also frequently used with pleasant connotations.

Based on this study's results, several suggestions emerged as potential areas of focus that future researchers could explore. While SVM outperformed KNN in terms of accuracy, it is advised that the approach be expanded to include more machine learning techniques or even attempt a hybrid approach to increase the model's dependability. In order to improve the model's accuracy in sentiment classification, it is also advised that future researchers consider applying oversampling or undersampling strategies.

REFERENCES

- [1] P. K. P. D. B. E. T. K. PELANGGAN, "Pengaruh Kualitas Pelayanan dan Promosi," *Dahmiri*, vol. 2, no. 2, pp. 183–192, 2018.
- [2] N. Nurhasanah, Z. P. Febriyani, and P. SK, "Pengaruh brand ambassador dan social media ads terhadap purchase intention melalui brand image produk skincare ms glow," *J. Ris. Ekon. dan Bisnis*, vol. 16, no. 2, p. 156, 2023, doi: 10.26623/jreb.v16i2.6466.
- [3] Dessy Angelina, U. Hayati, and G. Dwilestari, "Penerapan Metode Support Vector Machine Pada Sentimen Analisis Pengguna Twitter Terhadap Konser K-Pop," *Kopertip J. Ilm. Manaj. Inform. dan Komput.*, vol. 7, no. 1, pp. 14–23, 2023, doi: 10.32485/kopertip.v7i1.251.
- [4] D. A. Yusuf, A. L. Tumbel, and D. Woran, "Pengaruh Kualitas Produk Dan Brand Ambassador Kpop Nct Dream Terhadap Keputusan Pembelian Pada Produk Mie Lemonilo Di Manado," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 10, no. 3, p. 965, 2022, doi: 10.35794/emba.v10i3.43526.
- [5] W. Yulita, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naive Bayes Classifier," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 2, p. 1, 2021, doi: 10.33365/jdmsi.v2i2.1344.
- [6] Y. Cahyono and S. Saprudin, "Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier dengan Seleksi Feature Chi Squared Statistic," *J. Inform. Univ. Pamulang*, vol. 4, no. 3, p. 87, 2019, doi: 10.32493/informatika.v4i3.3186.
- [7] R. I. Pristiyanti, M. A. Fauzi, and L. Muflikhah, "Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 3, pp. 1179–1186, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] U. Khaira, R. Johanda, P. E. P. Utomo, and T. Suratno, "Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength," *Indones. J. Artif. Intell. Data Min.*, vol. 3, no. 1, p. 21, 2020, doi: 10.24014/ijaidm.v3i1.9145.
- [9] S. Rabbani, D. Safitri, F. Try Puspa Siregar, R. Rahmadden, and L. Efrizoni, "Evaluation of Support Vector Machine, Naive Bayes, Decision Tree, and Gradient Boosting Algorithms for Sentiment Analysis on ChatGPT Twitter Dataset," *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 1, p. 11, 2023, doi: 10.24014/ijaidm.v7i1.24662.
- [10] A. Mustolih, P. Arsi, and P. Subarkah, "Sentiment Analysis Motorku X Using Applications Naive Bayes Classifier Method," *Indones. J. Artif. Intell. Data Min.*, vol. 6, no. 2, p. 231, 2023, doi: 10.24014/ijaidm.v6i2.24864.
- [11] R. Slamet, W. Gata, A. Novtariany, K. Hilyati, F. Ainun Jariyah, and U. Nusa Mandiri, "Twitter Sentiment Analysis of South Korea Artists As Brand Ambassadors of Local Beauty Products," *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, 2022.
- [12] Regina, T. H. Saragih, and D. Kartini, "Analisis Sentimen Brand Ambassador Bts Terhadap Tokopedia Menggunakan Klasifikasi Bayesian Network Dengan Ekstraksi Fitur Tf-Idf," *J. Inform. Polinema*, vol. 9, no. 4, pp. 383–390, 2023, doi: 10.33795/jip.v9i4.1333.
- [13] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.
- [14] T. Pano and R. Kashef, "A complete vader-based sentiment analysis of bitcoin (BTC) tweets during the ERA of COVID-19," *Big Data Cogn. Comput.*, vol. 4, no. 4, pp. 1–17, 2020, doi: 10.3390/bdcc4040033.
- [15] T. Mustaqim, K. Umam, and M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm," *J. Phys. Conf. Ser.*, vol. 1567, no. 3, 2020, doi: 10.1088/1742-6596/1567/3/032024.
- [16] E. Cambria, Q. Liu, S. Decherchi, F. Xing, and K. Kwok, "SenticNet 7: A Commonsense-based Neurosymbolic AI Framework for Explainable Sentiment Analysis," *2022 Lang. Resour. Eval. Conf. Lr. 2022*, no. June, pp. 3829–3839, 2022.
- [17] S. Smetanin, "The Applications of Sentiment Analysis for Russian Language Texts: Current Challenges and Future Perspectives," *IEEE Access*, vol. 8, pp. 110693–110719, 2020, doi: 10.1109/ACCESS.2020.3002215.
- [18] A. S. Rizki, A. Tjahyanto, and R. Trialih, "Comparison of stemming algorithms on Indonesian text processing," *Telkonnika (Telecommunication Comput. Electron. Control.*, vol. 17, no. 1, pp. 95–102, 2019, doi: 10.12928/TELKOMNIKA.v17i1.10183.
- [19] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.

Title of manuscript is short and clear, implies research results... (First Author)

- [20] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.
- [21] R. Hidayat, *Cara Praktis Membangun Website Gratis*. Jakarta: PT. Elex Media Komputindo, 2010.
- [22] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [23] K. Pramayasa, I. M. D. Maysanjaya, and I. G. A. A. D. Indradewi, "Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE," *SINTECH (Science Inf. Technol. J.)*, vol. 6, no. 2, pp. 89–98, 2023, doi: 10.31598/sintechjournal.v6i2.1372.
- [24] Ferdi and Vina Ayumi, "Analisa Sentimen Mengenai Kenaikan Harga Bbm Menggunakan Metode Naïve Bayes Dan Support Vector Machine," *JSAI (Journal Sci. Appl. Informatics)*, vol. 6, no. 1, pp. 1–10, 2023, doi: 10.36085/jsai.v6i1.4628.
- [25] A. Setiawan, Zunan., Fajar, Muhammad., Priyatno, Arif Mudi., Putri, Anggi Yhurinda Perdana., Aryuni, Mediana., Yulianti, Siti., Junaidi, Satrio., Wijaya, *Buku Ajar DATA MINING*.
- [26] N. M. Farhan and B. Setiaji, "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023.

BIBLIOGRAPHY OF AUTHORS (10 PT)

	Natasya Kurnia Putri is an active final-semester student in the Informatics Department at Universitas Dr. Soetomo, Surabaya, Indonesia. The Author is interested in data science.
	Anik Vega Vitianingsih, S.Kom., MT, is a lecturer at Universitas Dr. Soetomo, Surabaya, Indonesia's Informatics Department. Geographic information systems study subjects that fall under machine learning-based spatial analysis are attractive to the Author.
	Slamet Kacung, S.Kom., M.Kom., is an informatics Department at Universitas Dr. Soetomo. The Author is interested in research issues linked to computer vision.
	Anastasia Lidya Maukar is a lecturer in information systems development and industrial management at President University in Indonesia. She has the degrees of S.T., M.Sc., and M.MT. Her academic background comprises two master's degrees: one in information systems development from the University of Hertfordshire and another in industrial management from the Institut Teknologi Sepuluh Nopember Surabaya (ITS). Research topics include information systems, databases, statistics, and production systems.
	Verdi Yasin is an Assistant Professor of Informatics Engineering at Jayakarta College of Informatics and Computer Management. His research focuses on Software Engineering, Business Intelligence, and Artificial Intelligence.

Sentiment Analysis of Brand Ambassador Influence on Product Buyer Interest Using KNN and SVM

ORIGINALITY REPORT

13%

SIMILARITY INDEX

9%

INTERNET SOURCES

5%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Institut Pertanian Bogor Student Paper	4%
2	Submitted to President University Student Paper	2%
3	ejournal.uin-suska.ac.id Internet Source	1%
4	Submitted to Universitas Jember Student Paper	1%
5	Submitted to UIN Sultan Syarif Kasim Riau Student Paper	<1%
6	"Inventive Communication and Computational Technologies", Springer Science and Business Media LLC, 2023 Publication	<1%
7	Zhang-Fa Yan, Yu-Lin Shen, Wei-Jun Liu, Jie-Min Long, Qingyang Wei. "An E-Commerce Coupon Target Population Positioning Model Based on Random Forest and eXtreme Gradient Boosting", 2018 11th International	<1%

Congress on Image and Signal Processing,
BioMedical Engineering and Informatics
(CISP-BMEI), 2018

Publication

8

web.fulcrumapp.com

Internet Source

<1 %

9

Dian Sa'Adillah Maylawati, Anjany Risqiati, Cepy Slamet, Muhammad Ali Ramdhani, Nur Lukman, Popon Dauni, Nunik Destria Arianti. "Chatbot for Virtual Travel Assistant with Random Forest and Rapid Automatic Keyword Extraction", 2021 IEEE 7th International Conference on Computing, Engineering and Design (ICCED), 2021

Publication

<1 %

10

ebin.pub

Internet Source

<1 %

11

www.techscience.com

Internet Source

<1 %

12

Dede Kurniadi, Shopi Nurhidayanti, Indri Tri Julianto, Teguh Wahyono, Yosep Septiana, Hetty Rohayani. "Classification of Television Programs Based on Public Opinion in Social Media Using Random Forest and Decision Tree", 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), 2023

Publication

<1 %

13 Dinar Ajeng Kristiyanti, Samuel Ady Sanjaya, Vinsencius Christio Tjokro, Jason Suhali. "Dealing imbalance dataset problem in sentiment analysis of recession in Indonesia", IAES International Journal of Artificial Intelligence (IJ-AI), 2024
Publication

14 M Khairul Anam, Muhammad Ihza Mahendra, Wirta Agustin, Rahmaddeni Rahmaddeni, Nurjayadi Nurjayadi. "Framework for Analyzing Netizen Opinions on BPJS Using Sentiment Analysis and Social Network Analysis (SNA)", INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi, 2022
Publication

15 conference.upnvj.ac.id
Internet Source

16 forum.upbatam.ac.id
Internet Source

17 jutif.if.unsoed.ac.id
Internet Source

18 www.arpnjournals.org
Internet Source

19 www.grafiati.com
Internet Source

20 Dian Rezky Wulandari, Muhammad Ary Murti, Hilman Fauzi T.S.P. "Sentiment Analysis Based on Text About President Candidate 2024 in Indonesia Using Artificial Intelligence with Parameter Optimization Algorithm", 2023 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS), 2023
Publication <1 %

21 Sasa Ani Arnomo, Yulia Yulia. "Clustering the potential bandwidth upgrade of FTTH broadband subscribers", ILKOM Jurnal Ilmiah, 2021
Publication <1 %

22 jurnal.ubl.ac.id
Internet Source <1 %

23 "Futuristic Trends in Network and Communication Technologies", Springer Science and Business Media LLC, 2021
Publication <1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On