ANALYSIS OF SKINCARE REVIEW SENTIMENT ON THE TOKOPEDIA MARKETPLACE USING THE K-NEAREST NEIGHBOR ALGORITHM

Andi Tyas Saputra [1]; Ruhul Amin [2];

Informatics [1]
Nusa Mandiri University 2
https://www.nusamandiri.ac.id/
12250094@nusamandiri.ac.id [1], ruhul.ran@nusamandiri.ac.id [2]



Works are distributed under the Creative Commons Attribution-NonCommercial 4.0 International License.

Abstract— *The research problem stems from the massive volume of unstructured skincare product reviews on Tokopedia containing informal language, which makes it difficult for consumers and sellers to extract public sentiment efficiently. The research objective is to classify skincare review sentiment (Skintific) on Tokopedia into positive, neutral, and negative categories, evaluate model performance, and deploy an interactive web-based prediction application. The proposed method utilizes the K-Nearest Neighbor (KNN) algorithm combined with Term Frequency-Inverse Document Frequency (TF-IDF) feature weighting and Euclidean Distance (K=5). The research steps encompass scraping 2,000 raw reviews (yielding 1,911 clean data points), text preprocessing (case folding, cleaning, tokenization, stopword removal, and stemming), an 80:20 data split (1,528 training data and 383 testing data), and interface implementation using the Streamlit framework. The obtained results demonstrate that the KNN model achieved a global accuracy of 95.56% (366 of 383 test samples correctly classified). The performance for the majority class (positive) was optimal, reaching 96% precision, 100% recall, and a 98% F1-score. Although classification on minority classes (neutral and negative) was limited due to extreme data imbalance, the Streamlit web application successfully facilitates interactive review sentiment prediction and visualization for single-text inputs and bulk file testing.*

Keywords: *Sentiment Analysis, K-Nearest Neighbor, Skincare, Tokopedia, TF-IDF, Streamlit..*

Digest – *The problem in this study is the large volume of skincare product reviews on Tokopedia which is unstructured and uses non-standard language, so that buyers and sellers have difficulty in extracting consumer sentiment information efficiently. The purpose of this study is to classify the sentiment of skincare product reviews (Skintific) on Tokopedia into positive, neutral, and negative categories, measure the performance of classifiers, and build a web-based prediction application. The proposed method utilizes the K-Nearest Neighbor (KNN) algorithm with weighting of the Term Frequency-Inverse Document Frequency (TF-IDF) feature and Euclidean Distance calculation (K=5). The research stages included scraping 2,000 review data (1,911 net reviews were obtained), pre-processing of text (case folding, cleaning, tokenizing, stopword removal, and stemming), data sharing in an 80:20 ratio (1,528 training data and 383 test data), and implementation of an interactive interface using the Streamlit framework. The results obtained showed that the KNN model achieved an overall accuracy of 95.56% (366 out of 383 predicted test data were precise). Performance in the majority (positive) class was optimal with 96% accuracy, 100% recall, and 98% F1-score. Despite the limitation of classification in the minority class (neutral and negative) due to data imbalance, the Streamlit web application successfully facilitated the visualization and prediction of review sentiment in an interactive manner for single-text and large-scale file testing.*

Keywords: *Sentiment Analysis, K-Nearest Neighbor, Skincare, Tokopedia, TF-IDF, Streamlit.*

INTRODUCTION

Rapid advances in the field of information and communication technology have shifted people's habits in transacting, especially in the beauty product sector. The presence of online marketplaces such as Tokopedia has now become the main channel for consumers to obtain beauty products because it presents better practicality and flexibility of access (Prameswari and Wahyono 2025).

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 18, No. 2 September 2021 | DOI https://doi.org/10.33480/techno.v18i2.XXXX

In these conditions, consumer opinions and reviews function as a very crucial information reference. In addition to providing an overview of the quality of a product, these reviews also reflect the level of satisfaction and real user experience, which is indirectly an important consideration for potential buyers in making transaction decisions (Scott, S. and S. S. 2024). Given that the beauty sector relies heavily on subjective perceptions and the diversity of product types, sentiment analysis is present as a crucial instrument for mapping consumer perspectives. This approach helps businesses identify product strengths and weaknesses based on customer perspectives, which in turn becomes a foothold in innovating products and formulating more targeted marketing tactics. To implement it, the K-Nearest Neighbor (KNN) algorithm can be relied on because of its straightforward nature but still tough in classifying text data (Scott, S. and S. S. 2024).

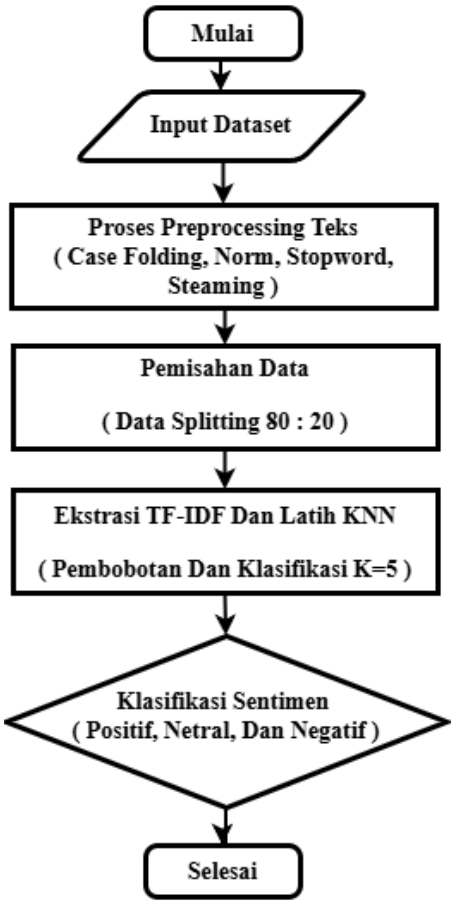
As one of the top e-commerce platforms in the country, Tokopedia presents a variety of beauty products from various brands and sellers. However, the surge in the volume of consumer reviews that continues to flow every day poses its own obstacles in processing and categorizing sentiment trends from thousands of text reviews efficiently (Irwannia and Lubis 2025).

In order to measure the level of public acceptance of beauty products, the evaluation of consumer opinion needs to be carried out through the mapping of reviews into positive and negative polarities. Sentiment analysis itself is a text processing technique aimed at examining and exploring the meaning of the text that contains the user's views, emotions, or expressions. The main focus of this approach is to recognize and classify the textual data into certain sentiment classes (Irwannia and Lubis 2025). The main advantage of the K-Nearest Neighbor algorithm lies in its mechanism that determines data categories based on the level of similarity or distance between samples in the feature space.

In sentiment analysis, reviews that contain words or phrases with certain similarities will be collected into one same category, so that the process of recognizing general patterns of consumer sentiment becomes easier and more structured. Various previous studies have shown that the K-Nearest Neighbor method is able to provide a fairly good level of accuracy in the task of classifying sentiment(Tanggu Mara, Sediyono, and Purnomo 2021).

MATERIALS AND METHODS

The methodology of this study is designed in a structured manner through several main stages which include data collection, automatic sentiment labeling, text preprocessing, data splitting, TF-IDF-based feature extraction, KNN model training (K=5), and evaluation of classification performance metrics. The overall system workflow architecture is shown in Figure 1.



Gambar 1. Flowchart

1. Data Collection and Sentiment Labeling Data reviews of skincare products Skintific are collected from Tokopedia e-commerce through web scraping techniques. After removing incomplete lines (missing values), a total of 1,911 net reviews were obtained. The initial sentiment labeling is carried out automatically based on the customer's star rating value, with the criteria: Ratings 1 and 2 are categorized as Negative Sentiment (Label 0), Rating 3 as Neutral Sentiment (Label 1), and Ratings 4 and 5 as Positive Sentiment (Label 2).

2. Text PreprocessingUnstructured textual review data must go through a series of text cleanups to remove noise and transform it

into a standard form(Hanafiah et al. 2023). The stages of preprocessing include:

a. Case Folding works to convert entire letters to lowercase (lowercase) and remove non-alphabetic characters (numbers and punctuation) using Regular Expression.

b. Word normalization is used to convert slang, abbreviations, and typos into Indonesian standard words (e.g.: yg = yang, bgt = banget, ori = original) based on the normalization dictionary(in Asih 2024).

c. Stopword Removal functions to remove common words that do not contain specific sentiment information using the Literary library plus product custom stopwords.

d. Tokenization serves to cut review sentences into single word sets (tokens).

e. Stemming Stemmer is used to convert affixed words into root words (e.g.: moisten = damp, absorb = absorb) by implementing a caching-based Stemmer Literature library for computational efficiency(Anggraeni, Cholissodin, and Marji 2021).

Tahap Preprcessing	Text Before processing	Text Before processing
Case Folding And Cleaning	Moisturizing, smooth and lightweight! Fits 100% on the skin.	Gentle and light moisturizing is perfect on the skin
Normalization and Stopwords	Lightweight, smooth, easy to absorb, sgt suitable on my skin	Lightweight, smooth, easy to absorb, very suitable for skin
Stemming Sastrawi	The product is very moisturizing and brightens the face	Very Moisturized Products Brighten Face

Table 1. Example of Text Pre-processing Stages

3. Data SplittingThe net review dataset was divided into two separate parts using the Stratified Train-Test Split technique with a ratio of 80% training set and 20% test data(Imani, Wijoyo, and Amalia 2023). A stratified scheme was chosen to keep the proportion of class label distribution (Positive, Neutral, Negative) in the training data and test data balanced according to the distribution of the original dataset. Out of a total of 1,911 net reviews, 1,528 samples were obtained as training data and 383 samples as test data.

4. TF-IDF (Term Frequency – Inverse Document Frequency) Feature Extraction A representation of clean text is transformed into a weighted numerical matrix using the TF-IDF method. Term Frequency (TF) measures the frequency of occurrence of the word t in document d, while Inverse Document Frequency (IDF) gives higher weight to unique words that rarely appear throughout the corpus of document D (Manning et al., 2008). The TF-IDF mathematical equation is formulated as follows:

$$TF\text{-}IDF(t, d, D) = TF(t, d) \times \log(|D| / |\{d \in d : t \in d\}|) \quad (1)$$

5. K-Nearest Neighbor (KNN) Classification Algorithm The KNN algorithm is a non-parametric classification method that works based on the degree of distance proximity between data. The TF-IDF vector of the test data was compared with all vectors of the training data using the Euclidean Distance metric (Equation 2). The sentiment class was determined based on the majority voting of the nearest neighbor K(Fieryando and Kristianto 2024). In this study, the value of K=5 of the nearest neighbor was configured.

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 18, No. 2 September 2021 | DOI https://doi.org/10.33480/techno.v18i2.XXXX

$$d(x, y) = \sqrt{[\sum (x_i - y_i)^2]} \quad (2)$$

6. Performance Evaluation Metrics The classification of the model is evaluated using standard Machine Learning metrics calculated based on the confusion matrix(Enhancing K-nearest neighbor algorithm.pdf n.d.) (Confusion Matrix) 3x3, including Accuracy, Precision, Recall, and F1-Score (Equations 3 to 6):

$$Accuracy = (TP + TN + TNe) / Total\ Data\ Test \quad (3)$$

$$Precision = TP/(TP+FP) \quad (4)$$

$$Recall = TP / (TP + FN) \quad (5)$$

$$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (6)$$

RESULTS AND DISCUSSION

1. Sentiment Dataset Distribution

Of the total 1,911 net review data of Skintific products on Tokopedia, the distribution of the sentiment class shows the level of data imbalance which is very dominant in the Positive class. Details of the distribution of the data sample are presented in Table 2 and the distribution graph in Figure 2.

Table 2. Distribution of Sentiment Classes

Category Late hour	Class Labels	Jumlah Sample	Percentage (%)
Sentimen Negatif	0	1.817	95.08 %
Sentimen Netral	1	68	3.56 %
Positive Sentiment	2	26	1.36 %

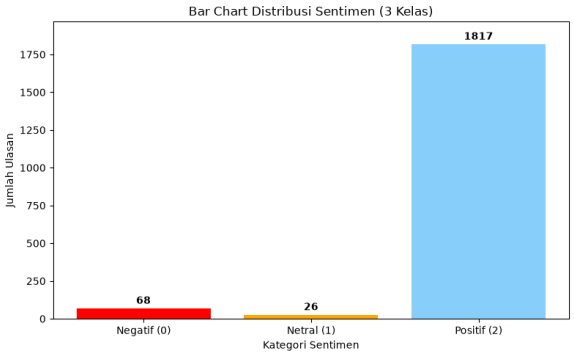


Figure 2. Sentiment Distribution Graph

2. TF-IDF Feature Extraction Results

The TF-IDF feature extraction process on 1,528 training data resulted in a dimensional feature matrix (1528, 1757), which showed that there were 1,757 unique vocabulary words. Specific words that rarely appeared but had high sentiment differentiating power (such as 'streaked', 'itch', 'nyesel') gained a greater IDF weight (above 6.0), while common words that appeared frequently gained a smaller IDF weight(Septiani and Isabela 2022).

3. Performance Evaluation of KNN Classification (K=5)

The KNN model test of 383 test data (20% of the total dataset) resulted in an overall accuracy value of 95.56% (366 correctly predicted test data). Details of the classification metrics for each sentiment class are presented in Table 3.

Table 3. Results of Evaluation of KNN Classification Metrics

Sentiment Class	Precision	Recall	F1-Score	Amount of Data
Sentimen Negatif (0)	1.00	0.14	0.25	14
Sentimen Netral (1)	0.00	0.00	0.00	5
Positive Sentiment (2)	0.96	1.00	0.98	364
Total Accuracy	-	-	0.96 (95.56 %)	383
Macro Average	0.65	0.38	0.41	383
Weighted Average	0.94	0.96	0.94	383

The detailed distribution of the actual prediction results versus the model's guesses is presented in the confusion matrix (Confusion Matrix 3x3) in Table 4 and Figure 3.

Table 4. Matrix Confusion Matrix 3X3 model KNN (K=5)

Actual/Predictive	Negative Precedents (0)	Neutral Predictions (1)	Positive Predictions (2)
Current Negative (0)	2 (true negative)	0	12 (false positive)
Neutral Actuals (1)	0	0 (true neutral)	5 (false positive
Positive Actuals (2)	0	0	364 (true positive)

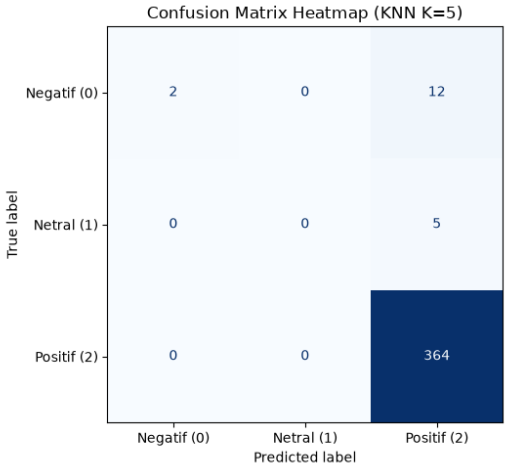


Figure 3. Heatmap Confusion Matrix KNN Classification (K=5)

4. Critical Analysis and Discussion of Data Imbalance

Based on empirical analysis of the test results, the accuracy paradox phenomenon was found. Although the global accuracy reached a high value of 95.56%, the Macro Average F1-Score only recorded 41% and the Macro Average Recall was 38%. This condition is due to the absolute dominance of the Positive class (1,817 reviews out of 1,911 total data).

Systematically, the majority voting mechanism in the KNN algorithm (K=5) calculates the Euclidean distances between TF-IDF vectors. Since the training data is dominated by the Positive class, when the sample of the test data of the minority class (Neutral and Negative) is plotted in the vector space, the 5 closest neighbors around it almost certainly consist of samples labeled Positive. As a result:

a. In the Negative class (14 test data): Only 2 reviews were predicted to be correct (14% recall), while the other 12 reviews were mispredicted as Positive grades.

b. In the Neutral class (5 test data): None of the reviews were successfully predicted to be correct (True Neutral = 0, Recall 0%), where all 5 neutral reviews were incorrectly classified as Positive class.

These findings confirm that the high global accuracy value of the e-commerce consumer opinion dataset does not necessarily reflect the reliability of the model across sentiment classes. Therefore, handling data imbalances using oversampling methods (such as SMOTE) is crucial to improve the model's generalization capabilities in minority classes.



Figure 3. Wordcloud Dominant Keywords of Positive Sentiment

5. Streamlit Web Interface Implementation

A trained (knn_model.sav) KNN binary model is integrated into a web-based interactive dashboard application interface using the Streamlit framework (Nurhidayat and Dewi 2023). The dashboard provides five main menus, including: (1) Dashboard & Metrics Overview, (2) Interactive TF-IDF Feature Weighting, (3) Single Text Prediction, (4) Excel File Batch Prediction, and (5) Dataset Exploration. This interface makes it easy for users and sellers to conduct sentiment testing in a hands-on and practical way.

CONCLUSION

Based on the results of the experiment and discussion of sentiment analysis of Skintific skincare product reviews on Tokopedia using the K-Nearest Neighbor (KNN) algorithm and TF-IDF weighting, several conclusions can be drawn as follows:

1. The implementation of the KNN algorithm (K=5) with TF-IDF feature extraction and an 80:20 data separation scheme successfully classified the review sentiment with a very high global accuracy of 95.56% (366 out of 383 predicted test data were correct).

2. The performance of the model was very dominant and optimal in the majority class (Positive Sentiment), with a Precision value of 0.96, Recall of 1.00, and F1-Score of 0.98 (364 positive reviews predicted to be correct).

3. The extreme data imbalance in real-world reviews limits the model's predictability in a

minority class, where the Negative class gets an F1-score of 0.25 and the Neutral class gets an F1-score of 0.00, resulting in a Macro Average F1-score of 41%.

4. The use of TF-IDF has proven to be effective in removing stopwords and giving higher weight to typical sentiment words, and the integration of the Streamlit dashboard makes it easy to test sentiment interactively.

Suggestions for further research include: (1) Applying synthetic data balancing techniques such as SMOTE (Synthetic Minority Over-sampling Technique) before KNN model training; (2) Conducting tuning experiments with a combination of K values (e.g. K=1, 3, 7, 9) and distance metric variations (Manhattan or Cosine Similarity); (3) Expanding the sources of review data collection from other e-commerce platforms such as Shopee and Lazada; and (4) Conducting comparative analysis with other classification algorithms such as Support Vector Machine (SVM) and Naive Bayes.

REFERENCE

F. D. Prameswari and T. Wahyono, "A Model of Sentiment Analysis of Online Shopping Application Users Using the Bidirectional Encoder Representations from Tran MethodAnggraeni, A, I Cholissodin, and M Marji. 2021. "Sentiment Analysis of Indihome Product Services Using Information Gain and K-Nearest Neighbor Method." *... Information Technology and Science ...* 5(8): 3616–24. https://jptiik.multi.web.id/index.php/j-ptiik/article/view/9642.

in Asih, Endah Mustika. 2024. "Analysis of Shopee as the Most Popular E-Commerce in Indonesia." *Journal of Antarctic Business Economics* 2: 2024–73. www.mckinsey.com,.

"Enhancing K-nearest Neighbor Algorithm.Pdf."

Fieryando, Fieryando, and Budhi Kristianto. 2024. "Sentiment Analysis of TikTok Shop with K-Nearest Neighbor, Decision Tree, and Naive Bayes." *Journal of Buana Informatics* 15(01): 21–29. doi:10.24002/jbi.v15i1.8205.

Hanafiah, Anggi, Arbi Haza Nasution, Yudhi Arta, Rizky Wandri, Hafiza Oktasia Nasution, Jerika Mardafora, and Universitas Islam Riau. 2023. "Sentiment Analysis of Customer Reviews of Shopee Products Based on Wordcloud Using Naïve Bayes Classifier Algorithm." *Journal of Information Technology and Computer Science (INTECOMS)* 6(1): 230–36.

Imani, Rahman Kukuh, Satrio Hadi Wijoyo, and Faizatul Amalia. 2023. "Application of the K-Nearest Neighbor Algorithm for the Classification of Students' Graduates' Ability to Compete for Jobs (Case Study: SMK 'SORE' Tulungagung)." 1(1).

Irwannia, Fira, and Andre Hasudungan Lubis. 2025. "Product Sentiment Analysis of Shopee Customer Reviews Using the K-Nearest Neighbors (KNN) Algorithm." *Journal of Informatic and Computer Science Engineering* 3(October): 1172–80. http://journal.mahesacenter.org/index.php/incodingmahesainstitut@gmail.com1.

Nurhidayat, Robi, and Kania Evita Dewi. 2023. "Application of K-Nearest Neighbor Algorithm and N-Gram Feature Extraction in Aspect-Based Sentiment Analysis." *Computing: Journal of Computer Science and Informatics* 12(1): 91–100. https://ojs.unikom.ac.id/index.php/komputa/article/view/9458.

Prameswari, Fla Desya, and Teguh Wahyono. 2025. "A Model of Sentiment Analysis of Online Shopping App Users Using the Bidirectional Encoder Representations from Transformers (BERT) Method." *Journal of Algorithms* 22(2): 1716–26. doi:10.33364/algoritma/v.22-2.3035.

Putri, Nitami Lestari, Budi Warsito, and Bayu Surarso. 2024. "The Effect of Sentiment Classification on Feature Engineering-Based Amazon Product Reviews and *K-Nearest Negihbor*." *Journal of Information Technology and Computer Science* 11(1): 65–74. doi:10.25126/jtiik.20241117376.

Septiani, Dwi, and Ica Isabela. 2022. "Term Frequency Inverse Document Frequency (Tf-IDF) Analysis in Information Recollection in Text Documents." *Synthesis* 1: 81–88.

Tanggu Mara, Andry, Eko Sediyono, and Hindriyanto Purnomo. 2021. "Application of K-Nearest Neighbors Algorithm to Sentiment Analysis of Network Learning Methods (DARING) at Wira Wacana Christian University Sumba." *Jointer - Journal of*

Informatics Engineering 2(01): 24–31. doi:10.53682/jointer.v2i01.30.

sformers (BERT)," *J. Algorithm.*, vol. 22, no. 2, pp. 1716–1726, 2025, doi: 10.33364/algorithm/v.22-2.3035.

Kuncoro, T. (1996). *Development of Internship Training Curriculum at STM Nasional Malang Department of Building, Building Study Program: A Study Based on the Needs of the Construction Services Business World*. Thesis not published. Malang: PPS IKIP MALANG.

Frieyadie, F. (2014). The Use of the RAD Model for the Development of an Online Bus Ticket Sales Information System. Journal of Pilar Nusa Mandiri, 10(2), 204–208. Retrieved from http://ejournal.nusamandiri.ac.id/ejurnal/index.php/pilar/article/view/359

Yossa, S., & . Z. (2013). Analysis of the Influence of Employee Ability, Task Division, and Motivation on Employee Performance at PT. Pelabuhan Indonesia Ii (Persero) Palembang Branch. Journal of Management and Business Sriwijaya, 11(4), 263–286. Retrieved from http://www.ejournal.unsri.ac.id/index.php/jmbs/article/view/3198