Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i3.817 Published By DRPM-UBD

Sentiment Analysis of Skincare Products Using the Naive Bayes Method

Karina Nurfebia[1], Sriani[2]

[1,2]Computer Science, North Sumatera State Islamic University, Medan, Indonesia
Email: [1]karinanurfebia0502@gmail.com, [2]sriani@uinsu.ac.id

Abstract

The number of reviews about skincare products can be used as an evaluation of product quality and satisfaction from consumers who have used it as well as considerations for other consumers to try the product. With the number of reviews, it is important to classify reviews into positive, negative, and neutral classes so that the level of product quality from each classification class can be known. The number of reviews causes the review classification process to be unable to be carried out automatically, so sentiment analysis is carried out. To determine the classification of positive sentiment, negative sentiment, and neutral sentiment on the skincare product, the Naive Bayes algorithm method is used. Naive Bayes was chosen because it is easy to implement and has a probability value to classify data. To determine the percentage of results from the specified classification, the Confusion Matrix will be used. The results of the classification process using the Naive Bayes method produce data into 3 types, namely 65 positive classes, 87 neutral classes, and 24 negative classes with an accuracy value of 73%, precision 77%, recall 61%, and f1-score 63%.

Keywords: Product Reviews, Sentiment Analysis, Naïve Bayes, Confusion Matrix

1. INTRODUCTION

Beauty products, particularly skincare items, are designed to maintain skin health and enhance one's appearance [1]. With the market flooded with a wide variety of skincare brands, not all products meet the desired quality standards or cater effectively to consumer needs [2]. To avoid purchasing unsuitable products, many consumers turn to online reviews, which serve as a vital source of information. Reviews in this context are written evaluations that express users' opinions and feelings about a product [3, 4]. These reviews not only help potential buyers make informed decisions but also generate greater public interest in the products [5]. For example, platforms like FemaleDaily.com provide numerous reviews of different skincare products, reflecting the diverse experiences of consumers [6].

For manufacturers, customer reviews are invaluable for improving product quality and enhancing customer satisfaction, which is considered a key performance indicator [7]. However, the overwhelming volume of online reviews makes it

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

difficult to manually analyze and classify them effectively. This has led to the use of sentiment analysis techniques, which assess the underlying sentiments such as emotionality, negativity, polarity, and subjectivity in the text [8]. Sentiment analysis helps categorize user feedback into positive, neutral, or negative sentiments, enabling companies to understand consumer perspectives more clearly [9]. Typically, this process involves dividing reviews into various sentiment classes to capture the full spectrum of consumer opinion [10].

Many existing studies on sentiment analysis of beauty product reviews, such as those for "Hadalabo Gokujyun Ultimate Moisturizing Milk," have relied primarily on binary classification methods that identify reviews as either positive or negative. One such study employed the Naïve Bayes algorithm, known for its effectiveness in text classification due to its simplicity, speed, and high accuracy, achieving an accuracy rate of 80.45% when applied to a dataset of 220 reviews [11]. In this study, 110 negative reviews were correctly classified, while 85 positive reviews were accurately identified, demonstrating the algorithm's effectiveness in binary sentiment classification. However, these methods did not consider neutral sentiments, which could provide more nuanced insights into consumer feedback.

Given this limitation, a more comprehensive approach to sentiment analysis is needed to include neutral reviews, which often reflect mixed or moderate opinions not captured by binary classification. To address this, the present research focuses on analyzing reviews of "Hadalabo Gokujyun Ultimate Moisturizing Milk" from the FemaleDaily website, expanding the sentiment classification to three categories: positive, negative, and neutral. By doing so, the study aims to capture a broader range of consumer sentiment, providing a more complete understanding of customer experiences and preferences.

Using the Naïve Bayes method on a larger dataset of 876 comments, this research aims to improve the sentiment classification process and provide deeper insights into consumer opinions. The inclusion of a neutral category allows for a finer granularity in sentiment analysis, which could lead to more accurate and meaningful interpretations of customer feedback. This enhanced sentiment analysis can help manufacturers better understand their customers, refine their products, and develop more targeted marketing strategies, ultimately contributing to improved product quality, customer satisfaction, and market competitiveness.

2. METHODS

The research process outlines a clear framework for the necessary actions to be taken. In this study, the system design is illustrated using a system flow diagram, which provides a precise depiction of each step involved, as shown in Figure 1.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

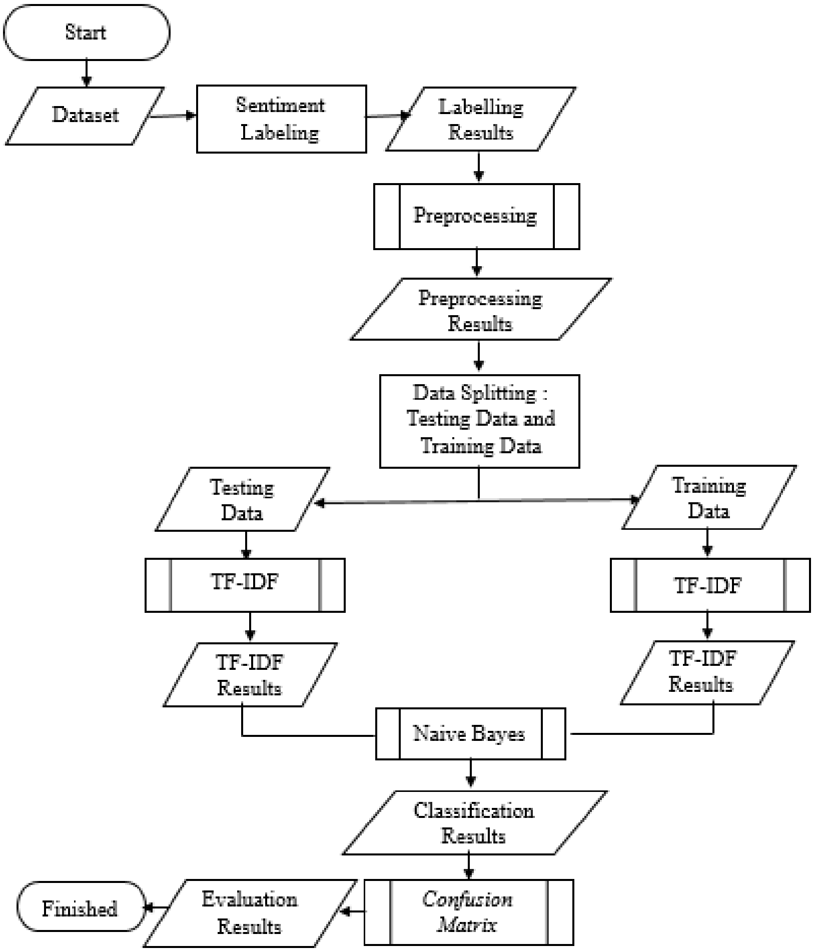

Figure 1. Research Flow

2.1 Dataset

The data collection stage in this study, the author collected data manually, namely collecting each review into a dataset. The variables used in this study are reviews on the femaledaily.com site regarding the Hadalabo Gokujyun Ultimate Moisturizing Milk skincare product.

2.2 Sentiment Labelling

The sentiment labeling stage, the author gives each review a label using the lexicon-based method with positive, neutral, and negative sentiment class criteria.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.3 Preprocessing

Data preprocessing is a technique used to enhance the algorithm's performance by transforming the unprocess edorraw data into an appropriate form for efficient handling of data by the machine learning algorithm [12]. In this stage the reviews are selected so that they are more structured and the information obtained is clearer. There are several stages in the processing process, namely:

a) Cleaning : In this step, the data will be cleaned or unnecessary characters will be removed, such as punctuation, numbers, URLs, usernames [13].
b) Case Folding : this stage will change the text to lowercase, remove punctuation, and remove numbers in the review.
c) Normalization: this stage the abbreviated words are changed into words that comply with the KBBI [14].
d) Tokenizing: this stage the text will be broken down into words [15].
e) Stopword: this stage filters words that do not have significant meaning in a document.
f) Stemming: at this stage, affixing words in the text will be removed.

2.4 Data Splitting

At the data splitting stage, the dataset will be divided into two data, namely training data and test data with a ratio of 80:20. The 80:20 ratio is used in addition to being a common practice in sentiment analysis and machine learning modeling, the 80:20 ratio is used because with 80% of the data for training, the model has enough examples to learn patterns and characteristics from the data. And the remaining 20% of the data is used to test the performance of the model, the data provides a more objective and realistic evaluation of the model's ability to predict new data.

2.5 TF-IDF

At this stage each word will be given a weight, the weight calculation will be calculated from the level of frequency of the word's appearance in each document. The greater the number of occurrences in the document, the greater the weight, and words that do not appear will be worth 0. TF-IDF features are used for determining the feature weights and it defines the vector[16]. The term frequency (TF) is the number of occurrences of a specific word in a given document. In contrast, the inverse document frequency (IDF) measures how common or uncommon the word is across all documents in a collection use Equation 1 to 3 [17].

$$TF\text{-}IDF\ (d,t) = TF(d,t) * IDF(t) \quad (1)$$

$$TF(d,t) = \frac{\text{number of words } t \text{ in the document } d}{\text{total words in the document } d} \quad (2)$$

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

$$IDF(t)=log\frac{D}{Df(t)} \quad (3)$$

Description: *IDF(t)* is Inverse Document Frequency value of word "t". *D* is total number of documents in the document collection to be analyzed. *Df(t)* is number of documents containing word *"t"*. *Tf* is number of term values that appear in a document.

2.6 Naive Bayes

At this stage, the training data that has been trained the model will be loaded with test data where the data has gone through the previous process so it can be grouped into certain categories based on the words contained in it. At this classification stage, the naive Bayes algorithm method will be used bayes theorem to compute conditional probabilities use formula as shown in Equation 4 [18]. The naive Bayes algorithm has been proven to be able to obtain quite satisfactory results when used for text classification. This classification technique can work quickly with high accuracy on large amounts of data.

$$P(H/X)= \frac{P(X\backslash H)}{P(X)} . P(H) \quad (4)$$

Description: X : Unknown class data. H : Data hypothesis is a specific class. $P(H|X)$: Probability of hypothesis H based on condition X. $P(H)$: Probability of hypothesis H. $P(X|H)$: Probability of X based on condition in hypothesis H. $P(X)$: Probability of X

2.7 Confusion Matrix

The accuracy, precision, and recall testing of previous results will be carried out using formula as shown in Equation 5 to 7. The results of this evaluation are carried out to check the correctness of the classification results that have been carried out. Accuracy, precision, and recall are commonly used metrics for evaluating the performance of classification models [19].

$$accurary= \frac{TP+TNt+TN}{TP+TNt+TN+FP+FNt+TN} \quad (5)$$

$$recall= \frac{TP}{TP+FNt+FN} \quad (6)$$

$$precision= \frac{TP}{TP+FP} \quad (7)$$

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3 RESULTS AND DISCUSSION

3.1 Dataset

Several reviews of the Hadalabo Gokujyun Ultimate Moisturizing Milk product on the Femaledaily.com website were manually collected into a data set, with a total of 877 reviews. Some of dataset as shown in Table 1.

Tabel 1. Dataset Containing Reviews

No.	Text
1.	Really recommend for those who want to focus on hydrating. The thing is, this product contains hyaluronic acid and alpha arbutin to brighten the skin, also no alcohol, no perfume. Especially if you use toner and moisturizer, your face automatically looks like a baby's butt. really plumpyyttðŸ'–
2.	I use this because my face tends to dry to normal, after using this and the lotion my face becomes healthy and supple. Even though after a while my face becomes a bit oily, it's still normal in my opinion, because I'm looking for skincare that moisturizes and makes my face glow. I will continue to buy this before I get a moisturizer with better results than this.
...	...
877	The product is delicious, it looks thick and milky but when you apply it it doesn't stick to your face... it's even better if you store the product in the refrigerator first, then use it later. But because I have oily skin, I only need to use a little, because if I use a little too much, it can cause flooding...

3.2 Sentiment Labelling

The data set containing reviews with a total of 877 reviews will be labeled using the lexicon-based method. The dataset containing reviews will be divided into several sentiment classes, namely neutral, positive, and negative sentiment as shown in Table 2 and Figure 2.

Tabel 2 Amount of Sentiment Class Data

Sentiment Class	Amount
Positive	313
Negative	133
Netral	431
Total	877

Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The reviews have gone through the labeling stage where the reviews have been divided into several sentiment classes with a total of 431 neutral classes, 313 positive, and 133 negative.

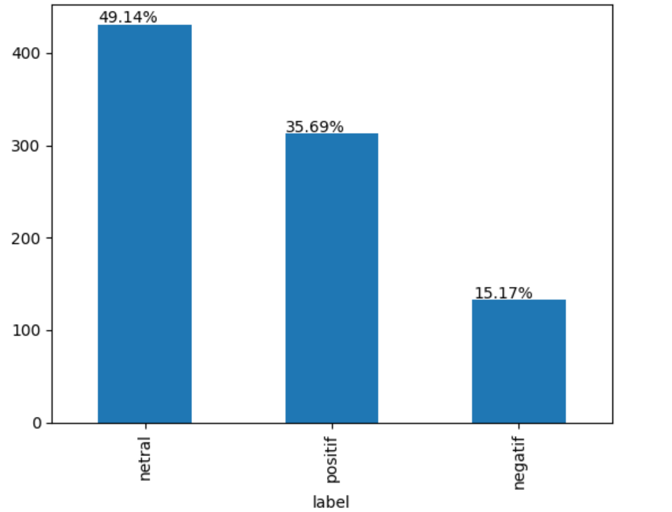

Figure 2. Sentiment Class Results on Skincare Product Reviews

3.3 Preprocessing Data

Sentiment data collection on public opinion on one of the skincare products, namely hadalabo gokujyun ultimate moisturizing milk on the femaledaily.com website. The process of collecting public opinion data on one of the skincare products on the femaledaily.com website manually is to compile the opinions into a csv file. The number of reviews used as a data set is 877.

The text preprocessing for sentiment analysis follows several key steps, as shown in Table 3. First, the Cleaning step involves removing unnecessary characters such as punctuation, numbers, URLs, usernames, and hashtags (e.g., "!”~&?!><#%{}([0-9]+:;‟)). Next, in Case Folding, all uppercase letters in the text are converted to lowercase to maintain consistency. The Normalization step corrects and expands abbreviated words to their full form, ensuring that the text is standardized and easily processed. Following this, the Tokenizing process breaks down sentences into individual words to identify meaningful terms. During Stopword Removal, common words that do not add significant value to the analysis are eliminated. Finally, Stemming removes affixes, such as prefixes and suffixes, from words, reducing them to their base form.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Tabel 3. Preprocessing Results Process

Cleaning	For me this hadalabo suits my skin very well. my skin type is normal to dry so it gets cleaner using this the texture is almost the same as other creamy cleansing milks for me the effect is not only clean but it also brightens it
Case Folding	for me this hadalabo suits my skin very well. my skin type is normal to dry so it gets cleaner using this the texture is almost the same as other creamy cleansing milks for me the effect is not only clean but it also brightens it
Normaliz	for me this hadalabo suits my skin very well. my skin type is normal to dry so it gets cleaner using this the texture is almost the same as other creamy cleansing milks for me the effect is not only clean but it also brightens it
Tokenizing	'make', 'me', 'hadalabo', 'fit', 'suitable', 'my skin', 'type', 'my skin', 'normal', 'tend', 'dry', 'so', 'clean', 'wear', 'this', 'texture', 'almost', 'same', 'with', 'milk', 'cleanser', 'other', 'creamy', 'on', 'me', 'anyway', 'effect', 'besides', 'clean', 'can', 'make', 'so', 'brighten', 'yes'
Stopword Removal	'hadalabo', 'suitable', 'suitable', 'my skin', 'type', 'my skin', 'normal', 'tends', 'dry', 'clean', 'use', 'texture', 'milk', 'cleansing', 'creamy', 'effect', 'clean', 'brighten'
Stemming	'hadalabo', 'suitable', 'suitable', 'skin', 'type', 'skin', 'normal', 'tend', 'dry', 'clean', 'wear', 'texture', 'milky', 'clean', 'creamy', 'effect', 'clean', 'bright'

In Table 3 it can be seen that a review goes through a preprocessing stage where the preprocessing stage has several stages, namely cleaning, case folding, normalizing, tokenizing, stopword removal, and stemming. So that the data that initially contains a sentence is processed into vocabulary that has meaning.

3.4 TF-IDF Weighting

The next stage after passing the sentiment class labeling stage is the TF-IDF (Term Frequency-Inverse Document Frequency) weighting stage, which at this stage uses the technique of calculating each weighting word (term) in the document data. The results of TF-IDF as shown in Table 4.

Tabel 4. Inverse Document Frequency Calculation Process for Training Data

Vocabulary	TF (Neutral)	TF (Positive)	TF (Negative)	DF	N/DF	IDF
hadalabo	1	0	0	1	3	0.477121
suitable	2	0	1	2	1.5	0.176091

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Vocabulary	TF (Neutral)	TF (Positive)	TF (Negative)	DF	N/DF	IDF
skin	2	2	0	2	1.5	0.176091
type	1	0	0	1	3	0.477121
normal	1	0	0	1	3	0.477121
……						
appear	0	0	1	1	3	0.477121
so	0	0	1	1	3	0.477121
good	0	0	1	1	3	0.477121
indeed	0	0	1	1	3	0.477121
product	0	0	1	1	3	0.477121

In the Table 4 each vocabulary will be calculated its weight, where each vocabulary will be calculated the number of term values that appear in a document (TF). Then the number of documents containing the vocabulary is calculated (DF). Next, the total number of documents in the collection of documents to be analyzed (N/DF) will be calculated. And the last stage will be calculated the inverse value of the frequency document from the vocabulary(IDF).

3.5 Naive Bayes Classification

In the test data classification process by multiplying all the probability values. Higher values constitute a new class of data. The test data classification process use formula as shown in Equation 8.

$$P\left(\frac{\frac{positive}{neutral}}{negative}\right)=P\left(\frac{\frac{positive}{neutral}}{negative}\right)\times \pi p(w\,|positive/neutral/negative) \qquad (8)$$

The classification stage where the test data will go through a classification process by multiplying all probability values. After being multiplied by all probability values, the higher value is the new data class. The classification model performance as shown in Figure 3.

```
MultinomialNB Accuracy: 0.7272727272727273
MultinomialNB Precision: 0.7680460782155697
MultinomialNB Recall: 0.6065281461833186
MultinomialNB f1_score: 0.6275585393810381
confusion_matrix:
 [[ 6 10  8]
 [ 0 79  8]
 [ 1 21 43]]

              precision    recall  f1-score   support

     negatif       0.86      0.25      0.39        24
      netral       0.72      0.91      0.80        87
     positif       0.73      0.66      0.69        65

    accuracy                           0.73       176
   macro avg       0.77      0.61      0.63       176
weighted avg       0.74      0.73      0.71       176
```

Figure 3. Naive Bayes Classification Model

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Test data containing 20% of the total data has been classified using the naive bayes method where the test data has a new class from the previous class. The classification results using the naive bayes method on the test data produced 6.73% for positive sentiment, 12.54% neutral sentiment, and 0.8% negative sentiment. In this classification result, the number of neutral sentiments is greater than the positive and negative sentiment classes.

3.6 Confusion Matrix

The classification accuracy is evaluated using a confusion matrix, which calculates the performance of the model based on the following metrics: True Positive (TP), True Neutral (TNt), True Negative (TN), False Positive (FP), False Neutral (FNt), and False Negative (FN). Figure 4 is performance evaluation using Confusion Matrix. The accuracy values will be calculated as follows.

$$6 + 79 + 43 = 128 \;/\; 6 + 10 + 8 \; + 0 + 79 + 8 + 1 + 21 + 43 = 176$$
$$128 \;/\; 176 = 0{,}73 * 100 = 73\%$$

The accuracy obtained in the classification of skincare product reviews analyzed is: Score Accuracy = 0.73 * 100 = 73%.

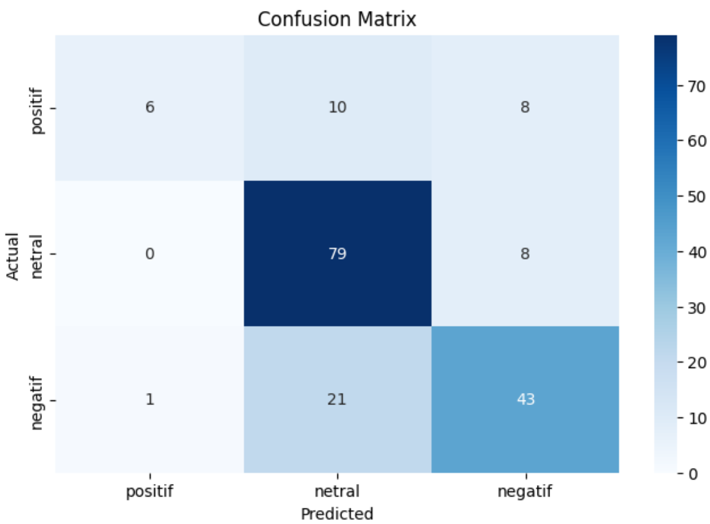

Figure 4. Confusion Matrix

3.7 Discussion

The dataset of skincare product reviews was categorized into three sentiment classes: positive, negative, and neutral. To achieve accurate sentiment classification, the analysis involved key steps, such as text preprocessing and

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

weighting using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The preprocessing steps—cleaning, normalizing, tokenizing, removing stopwords, and stemming ensured that the text data was appropriately prepared for analysis. The TF-IDF method was used to assign weights to each word in the dataset, highlighting the significance of terms that distinguish different sentiments. This approach allowed for a more nuanced understanding of the data by emphasizing terms that are more relevant to specific sentiment categories.

The Naive Bayes algorithm was then applied to classify the sentiment of the reviews, yielding an accuracy of 73% based on a sample of 176 reviews. The analysis revealed that the neutral sentiment was the most dominant among the reviews, suggesting that many users had moderate or mixed opinions about the skincare product. This finding underscores the value of including a neutral category in sentiment analysis, as it captures a range of sentiments that may be overlooked in a binary classification model. The predominance of neutral sentiment also reflects a more balanced perspective from the reviewers, indicating that their experiences with the product were varied and not strongly polarized.

While the Naive Bayes method proved effective in classifying the reviews, the study identified several limitations. The dataset was restricted to a single product, "Hadalabo Gokujyun Ultimate Moisturizing Milk," which limits the generalizability of the findings to other skincare products. Additionally, the achieved accuracy, while moderate, suggests that there is room for improvement. Future studies could enhance sentiment classification accuracy by experimenting with alternative machine learning algorithms, which may better capture the complexity of human language and sentiment. Moreover, expanding the language dictionary to include a broader range of terms and expressions could improve the detection and classification of subtle nuances in sentiment.

The findings highlight the importance of using comprehensive and refined analytical techniques in sentiment classification to gain deeper insights into consumer opinions. By broadening the scope of the analysis to include a more diverse dataset and employing more advanced methods, future research could further improve the effectiveness of sentiment analysis in understanding consumer behavior and preferences in the skincare market.

4 CONCLUSION

This study aimed to enhance sentiment analysis of skincare product reviews by categorizing them into three sentiment classes: positive, negative, and neutral, using a dataset focused on "Hadalabo Gokujyun Ultimate Moisturizing Milk." Through a comprehensive preprocessing approach and the application of the TF-IDF weighting method, the Naive Bayes algorithm was employed to classify the

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

reviews, achieving an accuracy of 73%. The results indicated that neutral sentiments were predominant, suggesting a wide range of user opinions that were not strongly polarized. While the Naive Bayes classifier demonstrated effectiveness, the study faced limitations, including the use of a dataset restricted to a single product and moderate accuracy levels. Future research should consider employing alternative algorithms and expanding the language dictionary to capture more nuanced sentiments, thereby improving the accuracy and applicability of sentiment analysis across a broader range of skincare products.

REFERENCES

[1] M. Hamka, N. Alfatari, and D. Ratna Sari, "Analisis Sentimen Produk Kecantikan Jenis Serum Menggunakan Algoritma Naïve Bayes Classifier," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 64, 2022, doi: 10.30865/json.v4i1.4740.

[2] T. Astuti and Y. Astuti, "Analisis Sentimen Review Produk Skincare Dengan Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO)," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 1806, 2022, doi: 10.30865/mib.v6i4.4119.

[3] D. N. Sari, D. N. Sari, F. Adelia, F. Rosdiana, B. B. Butar, and M. Hariyanto, "Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier," *JIKA (Jurnal Inform.*, vol. 4, no. 3, p. 109, 2020, doi: 10.31000/jika.v4i3.3086.

[4] S. Zeybek, E. Koç, and A. Seçer, "MS-TR : A Morphologically enriched sentiment Treebank and recursive deep models for compositional semantics in Turkish MS-TR : A Morphologically enriched sentiment Treebank and recursive deep models for compositional semantics in Turkish," *Cogent Eng.*, vol. 8, no. 1, 2021, doi: 10.1080/23311916.2021.1893621.

[5] R. Nurhidayat and K. E. Dewi, "Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek," *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 91–100, 2023, doi: 10.34010/komputa.v12i1.9458.

[6] S. K. Wardani and Y. A. Sari, "Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5582–5590, 2021.

[7] R. K. Botchway, A. B. Jibril, Z. K. Oplatková, and M. Chovancová, "Deductions from a Sub-Saharan African Bank's Tweets: A sentiment analysis approach," *Cogent Econ. Financ.*, vol. 8, no. 1, 2020, doi: 10.1080/23322039.2020.1776006.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[8] W. Van Atteveldt, M. A. C. G. Van Der Velden, and M. Boukes, "The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms," *Commun. Methods Meas.*, vol. 15, no. 2, pp. 121–140, 2021, doi: 10.1080/19312458.2020.1869198.

[9] H. Fang, G. Jiang, and D. Li, "Sentiment analysis based on Chinese BERT and fused deep neural networks for sentence-level Chinese e-commerce product reviews," *Syst. Sci. Control Eng.*, vol. 10, no. 1, pp. 802–810, 2022, doi: 10.1080/21642583.2022.2123060.

[10] K. W. Trisna and H. J. Jie, "Deep Learning Approach for Aspect-Based Sentiment Classification: A Comparative Review," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2021.2014186.

[11] N. S. Fathullah, Y. A. Sari, and P. P. Adikara, "Analisis Sentimen Terhadap Rating dan Ulasan Film dengan menggunakan Metode Klasifikasi Naïve Bayes dengan Fitur Lexicon-Based," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 2, pp. 590–593, 2020.

[12] K. Ramasamy, A. Sundaramurthy, and D. Velusamy, "Assessment and classification of grid stability with cost-sensitive stacked ensemble classifier," *Automatika*, vol. 64, no. 4, pp. 783–797, 2023, doi: 10.1080/00051144.2023.2218164.

[13] M. Furqan, M. Ikhsan, and R. Aini, "Algoritma Support Vector Machine Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Pandemi Virus Corona Di Media Sosial," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 4, no. 4, pp. 908–915, 2023.

[14] M. Furqan, S. Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.

[15] A. F. Setyaningsih, D. Septiyani, and S. R. Widiasari, "Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia," *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 224–235, 2023, doi: 10.37012/jtik.v9i1.1409.

[16] K. S. Kumar *et al.*, "Sentiment Analysis of Short Texts Using SVMs and VSMs-Based Multiclass Semantic Classification Sentiment Analysis of Short Texts Using SVMs and," *Appl. Artif. Intell.*, vol. 38, no. 1, 2024, doi: 10.1080/08839514.2024.2321555.

[17] F. Westin, "Time Period Categorization in Fiction : A Comparative Analysis of Machine Learning Techniques Time Period Categorization in Fiction : A Comparative," *Cat. Classif. Q.*, vol. 62, no. 2, pp. 124–153, 2024, doi: 10.1080/01639374.2024.2315548.

[18] O. C. Edo *et al.*, "Fintech adoption dynamics in a pandemic: An experience from some financial institutions in Nigeria during COVID-19 using machine learning approach," *Cogent Bus. Manag.*, vol. 10, no. 2, 2023, doi: 10.1080/23311975.2023.2242985.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[19] V. Ganganwar and R. Rajalakshmi, "Employing synthetic data for addressing the class imbalance in aspect-based sentiment classification," *J. Inf. Telecommun.*, vol. 8, no. 2, pp. 167–188, 2024, doi: 10.1080/24751839.2023.2270824.