



**ANALISIS SENTIMEN PADA REVIEW PRODUK
DI MARKETPLACE SHOPEE
MENGUNAKAN NAÏVE BAYES CLASSIFIER**

TUGAS AKHIR

Ika Agustina
41518110066

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**



**ANALISIS SENTIMEN PADA REVIEW PRODUK
DI MARKETPLACE SHOPEE
MENGUNAKAN NAÏVE BAYES CLASSIFIER**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Ika Agustina
41518110066

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518110066

Nama : Ika Agustina

Judul Tugas Akhir : Analisis Sentimen pada *Review* Produk di *Marketplace*
Shopee Menggunakan Naïve Bayes Classifier

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 25 Agustus 2022



SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Ika Agustina
NIM : 41518110066
Judul Tugas Akhir : Analisis Sentimen pada *Review* Produk di *Marketplace* Shopee Menggunakan Naïve Bayes Classifier

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 25 Agustus 2022

UNIVERSITAS
MERCU BUANA



Ika Agustina

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Ika Agustina
 NIM : 41518110066
 Judul Tugas Akhir : Analisis Sentimen pada *Review* Produk di *Marketplace* Shopee Menggunakan Naive Bayes Classifier

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	✓
		Jurnal Nasional Terakreditasi	✓		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: Lontar Komputer ; Jurnal Ilmiah Teknologi Informasi			
	ISSN	: ISSN Print 2088-1541 ISSN Online 2541-5832			
	Link Jurnal	: https://ojs.unud.ac.id/index.php/lontar/			
	Link File Jurnal Jika Sudah di Publish	:			

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan.

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
Dosen Pembimbing TA



Ir. Emil Robert Kaburuan, Ph.D., IPM.

Jakarta, 25 Agustus 2022



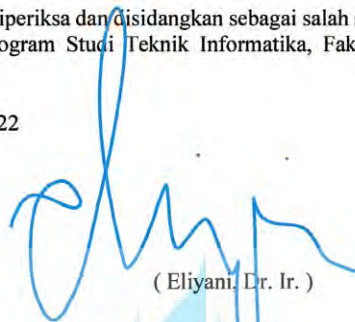
Ika Agustina

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110066
Nama : Ika Agustina
Judul Tugas Akhir : Analisis Sentimen pada *Review* Produk di
Marketplace Shopee Menggunakan Naïve Bayes
Classifier

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 16 Agustus 2022



(Eliyani, Dr. Ir.)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110066
Nama : Ika Agustina
Judul Tugas Akhir : Analisis Sentimen pada *Review* Produk di
Marketplace Shopee Menggunakan Naïve Bayes
Classifier

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 12 Agustus 2022



(Afiyati, S.Si, MT)


UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110066
Nama : Ika Agustina
Judul Tugas Akhir : Analisis Sentimen pada *Review* Produk di
Marketplace Shopee Menggunakan Naïve Bayes
Classifier

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 24 Agustus 2022



(Achmad Kodar, Drs. MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41518110066
Nama : Ika Agustina
Judul Tugas Akhir : Analisis Sentimen pada *Review* Produk di *Marketplace* Shopee Menggunakan Naïve Bayes Classifier

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022

Menyetujui,



(Ir. Emil Robert Kaburuan, Ph.D., IPM.)
Dosen Pembimbing

Mengetahui,



(Wawan Gumawan, S.Kom, MT)

Koord. Tugas Akhir Teknik Informatika



(Ir. Emil Robert Kaburuan, Ph.D., IPM.)

Ka. Prodi Teknik Informatika

UNIVERSITAS
MERCU BUANA

KATA PENGANTAR

Puji syukur kita panjatkan kehadiran Allah SWT. Tuhan Yang Maha Esa, Maha Pengasih dan Maha Penyayang yang telah melimpahkan segala rahmat dan hidayah-Nya kepada penulis sehingga penulis dapat menyelesaikan laporan tugas akhir dengan judul “Analisis Sentimen pada *Review* Produk di *Marketplace* Shopee Menggunakan Naïve Bayes Classifier” sesuai dengan rencana karena banyaknya dukungan dari berbagai pihak.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, tugas akhir ini tidak akan dapat diselesaikan tepat waktu. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Emil Robert Kaburuan, S.T., M.A., Ph.D selaku Ka. Prodi Teknik Informatika Fakultas Ilmu Komputer Universitas Mercubuana, Jakarta, sekaligus menjadi Dosen Pembimbing yang telah membimbing saya dalam pengerjaan tugas akhir ini.
2. Bapak Wawan Gunawan, S.Kom, MT, selaku Koordinator Tugas Akhir Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Mercubuana, Jakarta.
3. Orang tua, keluarga, sahabat dan teman-teman yang telah mendukung, membantu dan memberikan semangat kepada penulis dalam menyelesaikan skripsi ini.
4. Semua pihak yang telah membantu kelancaran dalam penyusunan skripsi, yang tidak dapat penulis sebutkan satu persatu.

Akhir kata, penulis berharap penyusunan tugas akhir ini dapat bermanfaat bagi semua pihak. Penulis menyadari bahwa dalam penyusunan laporan tugas akhir ini masih terdapat banyak kekurangan, oleh karena itu penulis sangat mengharapkan saran dan kritik yang membangun.

Jakarta, 28 Juli 2022

Ika Agustina

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI	v
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xii
NASKAH JURNAL	1
KERTAS KERJA.....	12
BAB 1. LITERATUR REVIEW.....	13
BAB 2. ANALISIS DAN PERANCANGAN.....	19
BAB 3. SOURCE CODE.....	24
BAB 4. DATASET.....	32
BAB 5. TAHAPAN EKSPERIMEN.....	34
BAB 6. HASIL SEMUA EKSPERIMEN.....	41
DAFTAR PUSTAKA	49
LAMPIRAN DOKUMEN HAKI.....	51
LAMPIRAN KORESPONDENSI	53

NASKAH JURNAL

Sentiment Analysis on Product Reviews from Shopee Marketplace using the Naïve Bayes Classifier

Ika Agustina^{a1}, Emil Robert Kaburuan^{a2}

^aFaculty of Information Technology, Mercubuana University
Jakarta, Indonesia

¹ikaag.08@gmail.com

²emil.kaburuan@mercubuana.ac.id

Abstract

Online shopping has become a popular shopping method ever since the number of internet users increased. Online shopping activities have become very easy and flexible because it can be completed anywhere and anytime. The products provided are also complete. The products sold often do not always match the actual conditions because the product can only be seen through pictures. Users who have purchased a product can share their opinions on it using the review feature. The products purchased thousands or millions of times, however, have many reviews. To take an overview of the product, it is important to go through every positive and negative review, so it takes a lot of time and effort. Reviews of products from the Shopee marketplace will be classified into positive or negative sentiments towards women's home wear clothing or house dress in this study. The research starts with data crawling, text preprocessing, training data, testing, evaluation model and then concludes with a general description based on the most frequently discussed topics in the reviews for each sentiment class. Classification is done using the Naïve Bayes Classifier algorithm. The accuracy obtained is 90,03%. The total dataset is 2907 data.

Keywords: online shopping, sentiment analysis, naïve bayes classifier, product reviews, shopee

1. Introduction

Shopping is part of everyday life [1]. Shopping activities that were previously done offline by visiting shops or markets now can be done online by using gadgets only. Online shopping provides consumers with more information and opportunities to compare products and prices, with a better product selection, and more convenience and ease in finding the desired product online [2]. There are already many marketplaces available, one of them is Shopee which has provided various needs such as food, clothing, accessories, electronic devices, and even household equipment. Online shopping has many advantages, but there are also disadvantages. The products sold often do not always match the actual conditions, like the shape, color, and size, because it can only be seen based on the picture, it is not like the original condition as shown in the picture.

Reviews of a product are critical in deciding product purchases because they can provide an overview of product quality based on other consumers' experiences [3]. The decisions we make are influenced by the opinions of others in some cases [3]. Looking at the reviews given by other consumers to get an overview of a product is essential for purchasing products online [4]. Reviews of a product can increase interest in buying and using the product. Users can provide reviews about products that have been purchased with the review feature from consumers that the marketplace has provided. Sellers can use these reviews as material for evaluation, and potential buyers also get an overview of the

products they are interested in based on the experiences of other consumers. The reviews also can help sellers and buyers know each product's quality.

Researchers have researched sentiment analysis for product review previously. Researchers have researched sentiment analysis for product review previously. Many research has been developed using Naïve Bayes Classifier for films [3], applications [5], [6], restaurant [7], and delivery services [8], with each research having fairly high accuracy. Sentiment analysis also has been used for application reviews using SVM [9] and KNN [10], and there is also using long short-term memory for comments written on social media [11]. The research compares the naive Bayes classifier with the lexicon based holistic, and it can conclude that the naive Bayes classifier method has a better precision value and accuracy level than the lexicon-based holistic method [11].

Text mining classification will discussed in this study from a review on a product of home wear clothes for women, commonly called a house dress, from one of the shops in Shopee. Reviews will be classified using the Naïve Bayes Classifier algorithm into positive and negative sentiments. Based on previous research, the Naïve Bayes Classifier method has a pretty good performance and has been widely used in research in the field of text mining, and has a high level of accuracy. Therefore, this study uses the Naïve Bayes Classifier. After sentiment classification, it will analyze the general description of a product based on the reviews given by users, including the product's advantages and disadvantages from each sentiment.

2. Reseach Methods

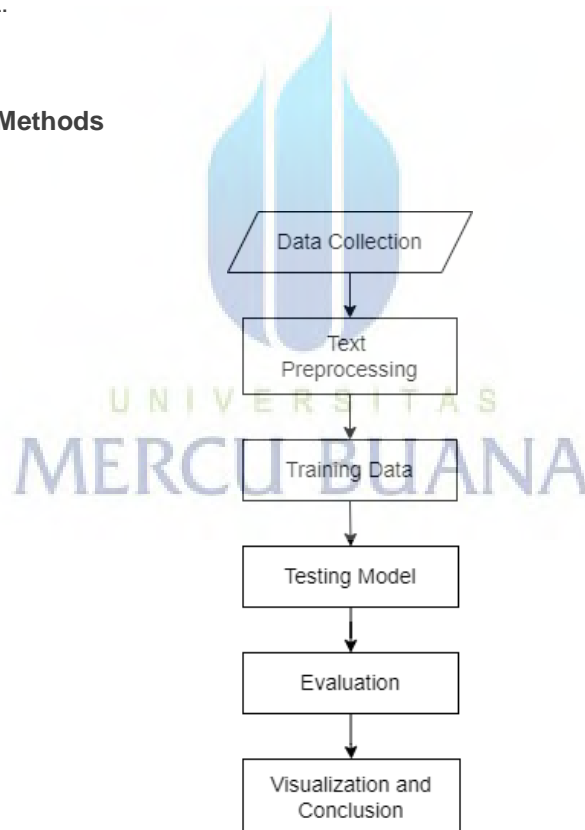


Figure 1. System Main Flowchart

Figure 1 shows the steps carried out in this study. The process starting from data collection, text preprocessing, training data, model testing, evaluation and visualization as well as conclusions.

2.1. Data Collection

Data was collected using a web crawling method from reviews written by users purchased one of the home wear clothes products for women or house dress sold on the Shopee marketplace. The data collected from the review and rating columns on the product review page. Crawling is done using the API provided by Shopee and the Python programming language use Google Colaboratory.

2.2. Text Preprocessing

Text preprocessing is carried out to clean data and change what was originally unstructured data to be more structured. The stages of text processing are divided into several, namely: case folding, text cleaning, word normalization, stemming, translating datasets into English language, and stopword removing. The result of text preprocessing is data that is ready to be processed for the data training and sentiment classification process.

2.3. Training Data

Training data is used to build a suitable model for the classification. The dataset is divided into train data and test data. The amount of train data is 70% of the total dataset. The algorithm used is Naïve Bayes Classifier.

2.4. Testing Model

The results of the model that have been trained are then tested on the test data to see the model's accuracy. The test uses the Naïve Bayes Classifier algorithm to determine the sentiment class of each review. The amount of test data is 30% from dataset. Naïve Bayes Classifier is an algorithm that predicts the probability of each sentiment class and then chooses which class has the greatest probability. The Naïve Bayes Classifier algorithm has a pretty good performance and has been widely used in research in the field of text mining, and has a high level of accuracy [12]. Comparison calculations between the terms in the testing data and each existing class can be done with Equation (1) [13].

$$P(a_j|v_j) = \frac{nc+mp}{n+m} \quad (1)$$

information:

n = the number of training examples for which $v = v_j$

nc = number of examples for which $v = v_j$ and $a = a_j$

p = a priori estimate for $P(a_j|v_j)$

m = the equivalent sample size

Equation (2) is used to calculate the classification of the test data to find which class has the greater probability after calculating the comparison between the terms in the testing data. [13].

$$V_{nb} = \underset{v \in V}{\operatorname{argmax}} P(v) \prod P(a_i|v_j) \quad (2)$$

2.5. Evaluation and Calculating Accuracy

The classification results are then calculated for accuracy by comparing the classification results using Naïve Bayes with the manual labeling sentiment using confusion matrix. A confusion matrix is a tool used to evaluate the classification model to estimate whether objects are right or wrong [14]. Results from the confusion matrix will also be used to calculate accuracy, recall, and precision.

2.6. Visualization and Product Overview

The first visualization uses a bar chart to display the result of each class of sentiment classification results. Positive and negative sentiment class will be displayed as a word cloud to find out what words or topics are most often discussed in product reviews. A general overview of the product is concluded based on user reviews from the words with the highest frequency of occurrence showing in word cloud.

3. Result and Discussion

3.1. Data Collection

The rating and review will be taken into the datasets from the review page. A rating is a standard symbol representing the overall consumer satisfaction with the seller's or marketer's product or service (usually denoted using 1 to 5 stars), where more stars or higher scores reflect better satisfaction with the product or service [15]. A total of 2907 reviews were collected. The next step is manual labeling which will be used as an actual prediction to compare the predicted results from the program. Reviews are divided into two classes of sentiment, positive and negative reviews. Manual sentiment labeling resulted in 2314 or 79.63% in positive sentiment and 593 or 20.98% in negative sentiment. Data training will use 70% dataset, which is 2034, and 30% for testing data, which is 873.

Table 1. Sample Dataset

Label	Review
Positive	Bahan bagus, pengiriman cepat,
Negative	Salah kirim min, tolong dicek lagi 😞

3.2. Preprocessing

- Case folding: converts all letters to lowercase. All the letters in the text reviews are changed to lowercase as shown in Table 2.

Before	After
Bahan bagus, pengiriman cepat,	bahan bagus, pengiriman cepat,
Salah kirim min, tolong dicek lagi 😞	salah kirim min, tolong dicek lagi 😞

Table 2. Case Folding

- Text cleaning: removing punctuation, emoji, and numbers. Some emojis and punctuation marks were removed from the text as shown in Table 3.

Before	After
bahan bagus, pengiriman cepat,	bahan bagus pengiriman cepat
salah kirim min, tolong dicek lagi 😞	salah kirim min tolong dicek lagi

Table 3. Text cleaning

- Text normalization: removing repetitive characters in a word and converting slang words into standard words.

Table 4. Text normalization

Before	After
bahan bagus pengiriman cepat	bahan bagus pengiriman cepat
salah kirim min tolong dicek lagi	salah kirim min tolong dicek lagi

- d. Stemming: changing words their root forms. The word "dicek" is changed to the root word "cek" as shown in Table 5.

Before	After
bahan bagus pengiriman cepat	bahan bagus kirim cepat
salah kirim min tolong dicek lagi	salah kirim min tolong cek lagi

Table 5. Stemming

- e. Translate to English

Before	After
bahan bagus kirim cepat	good material fast delivery
salah kirim min tolong cek lagi	sent wrong one please check again

Table 6. Translate to English

- f. Stopword removal: removes meaningless or irrelevant words. Examples of removed words is "again".

Before	After
good material fast delivery	good material fast delivery
sent wrong one please check again	sent wrong one please check

Table 7. Stopword removal

3.3. Classification Results using Naïve Bayes Classifier

Classification is carried out on all datasets using a previously trained model. The sample results of the classification are as follows.

Table 8. Sample Classification Data Test

No	Review	Actual Prediction	Nave Bayes Classification
1	good material fast delivery	Positive	Positive
2	sent wrong one please check	Negative	Negative

The following is an example of the calculation of the classification results shown in Table 8 to calculate the probability of each class in the review. Probability of the sentiment class in the training data, and the calculation of the comparison between the terms and testing data in each existing class using Equation (1) need to be calculate first. To obtain the document

probability for each class, multiply the class probability with the word probability. The next step is to decide which probability is the largest, then that is the sentiment class.

Sentiment class probability from the training data can be calculated with the equation:

$$\begin{aligned} P(\text{Positive}) &= \frac{\text{amount positive classes}}{\text{all training data}} \\ P(\text{Positive}) &= 1619 / 2034 \\ P(\text{Positive}) &= 0,795968 \end{aligned}$$

$$\begin{aligned} P(\text{Negative}) &= \frac{\text{amount negative classes}}{\text{all training data}} \\ P(\text{Negative}) &= 415 / 2034 \\ P(\text{Negative}) &= 0,203539 \end{aligned}$$

Frequency of occurrence of words for each class in the sample test data shown in Table 9.

Table 9. Frequency of Occurrence of Words in Training Data

Word	Positive Class	Negative Class	Total
good	866	83	949
material	636	107	743
fast	273	13	286
delivery	240	20	260
sent	40	86	126
wrong	27	39	66
one	96	116	212
please	20	43	63
check	9	14	23

Next is calculate the probability of a word for positive or negative class using equation (1).

Test data 1:

$$P(\text{positive}|\text{good}) = \frac{866 + 4 \cdot 0,795968}{949 + 4} = 0,912050$$

$$P(\text{positive}|\text{material}) = \frac{636 + 4 \cdot 0,795968}{66 + 4} = 0,855668$$

$$P(\text{positive}|\text{fast}) = \frac{273 + 4 \cdot 0,795968}{212 + 4} = 0,952358$$

$$P(\text{positive}|\text{delivery}) = \frac{240 + 4 \cdot 0,795968}{63 + 4} = 0,921151$$

$$P(\text{negative}|\text{good}) = \frac{83 + 4 \cdot 0,203539}{949 + 4} = 0,087948$$

$$P(\text{negative}|\text{material}) = \frac{107 + 4 \cdot 0,203539}{743 + 4} = 0,144330$$

$$P(\text{negative}|\text{fast}) = \frac{13 + 4 \cdot 0,203539}{286 + 4} = 0,047635$$

$$P(\text{negative}|\text{delivery}) = \frac{20 + 5 \cdot 0,203539}{260 + 4} = 0,078842$$

Test data 2:

$$P(\text{positive}|\text{sent}) = \frac{40 + 5 \cdot 0,795968}{126 + 5} = 0,335724$$

$$P(\text{positive}|\text{wrong}) = \frac{27 + 5 \cdot 0,795968}{66 + 5} = 0,436336$$

$$P(\text{positive}|\text{one}) = \frac{96 + 5 \cdot 0,795968}{212 + 5} = 0,460737$$

$$\begin{aligned}
P(\text{positive}|\text{please}) &= \frac{20+5 \cdot 0,795968}{63+5} = 0,352645 \\
P(\text{positive}|\text{check}) &= \frac{9+5 \cdot 0,795968}{23+5} = 0,463566 \\
P(\text{negative}|\text{sent}) &= \frac{86+5 \cdot 0,203539}{126+5} = 0,664257 \\
P(\text{negative}|\text{wrong}) &= \frac{39+5 \cdot 0,203539}{66+5} = 0,563630 \\
P(\text{negative}|\text{one}) &= \frac{116+5 \cdot 0,203539}{212+5} = 0,539252 \\
P(\text{negative}|\text{please}) &= \frac{43+5 \cdot 0,203539}{63+5} = 0,647319 \\
P(\text{negative}|\text{check}) &= \frac{14+5 \cdot 0,203539}{23+5} = 0,536346
\end{aligned}$$

The next step is to find the maximum value from the multiplication of the probability value and the P-value for each class using equation (2), as follows:

Test data 1:

$$\begin{aligned}
V(\text{Positive}) &= 0,795968 * 0,912050 * 0,855668 * 0,952358 * 0,921151 = 0.54494241 \\
V(\text{Negative}) &= 0,203539 * 0,087948 * 0,144330 * 0,047635 * 0,078842 = 0.000009 \\
V_{nb} &= \text{argmax}(V(\text{Positive}) | V(\text{Negative})) \\
V_{nb} &= \text{argmax}(0.54494241 | 0.000009) \\
V_{nb} &= 0.54494241
\end{aligned}$$

Test data 2:

$$\begin{aligned}
V(\text{Positive}) &= 0,795968 * 0,335724 * 0,436336 * 0,460737 * 0,352645 * 0,463566 \\
&= 0,008782 \\
V(\text{Negative}) &= 0,203539 * 0,664257 * 0,563630 * 0,539252 * 0,647319 * 0,536346 \\
&= 0,014267 \\
V_{nb} &= \text{argmax}(V(\text{Positive}) | V(\text{Negative})) \\
V_{nb} &= \text{argmax}(0,008782 | 0,014267) \\
V_{nb} &= 0,014267
\end{aligned}$$

Calculation with Equation (2) showing results the 1st test data obtained a maximum value of 0.54494241 in positive class probability, so the sentiment class value is Positive. The maximum value obtained in the second test data is 0.54494241 in negative class probability, so the sentiment class value is Negative.

3.4. Evaluation and Calculation Accuracy

Validation of the classification results using a confusion matrix. Confusion matrix of the manual classification results on the classification results from the model that has been built showing in the Table 10.

Table 10. Confusion Matrix

	Predict Class
--	---------------

		Predicted “+”	Predicted “-”
Actual Class	Actual “+”	670	25
	Actual “-”	62	116

In the confusion matrix, it can be concluded as follows. The number of positive sentiment classes that were correctly predicted was 364. The number of wrongly predicted positive sentiment classes is 19. The number of correctly predicted negative sentiment classes is 325. The number of wrongly predicted negative sentiment classes is 58. The cause of the error prediction is probably due to the imbalance of the dataset between positive and negative sentiment, which causes the tendency of the model to predict sentiment as positive.

Accuracy, precision and recall show in the calculation below.

$$\begin{aligned}
 \text{Accuracy} &= (TP+TN) / (TP+FP+FN+TN) \\
 &= (670+116) / (670+25+62+116) \\
 &= 0.9003 \\
 &= 0.9003 * 100 \% \\
 &= 90.03 \%
 \end{aligned}$$

$$\begin{aligned}
 \text{Precision} &= TP / (TP + FP) \\
 &= 670 / (670 + 25) \\
 &= 0.964 \\
 &= 96.4\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= TP / (TP + FN) \\
 &= 670 / (670 + 62) \\
 &= 0.9153 \\
 &= 91.53\%
 \end{aligned}$$



3.5. Visualization and Product Overview

The entire 2907 dataset is classified after the model is successfully built and evaluated, to conclude an overview of the negligee product and visualized using a bar chart to indicate the number of each sentiment class. Then display words that appear most often in each sentiment class using a word cloud to conclude the general picture of the negligee product.

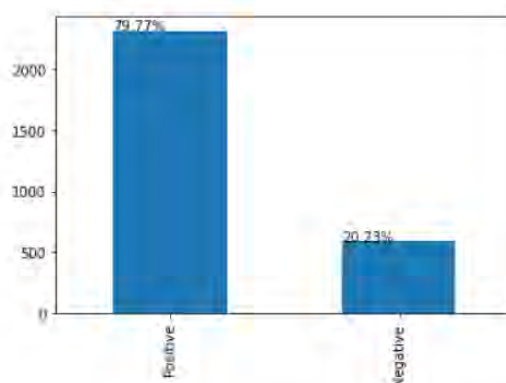


Figure 2. Bar Chart Classification Result

The classification obtained 79.77% or 2319 positive sentiments and 20.23% or 588 negative sentiments as shown in the Figure 2.



Figure 3. Word Cloud Positive Sentiment

Figure 3 showing the words that appear most often in the positive sentiment class, which are: "good", "thank", "material", "cool", "color", "negligee", "price", "fast delivery", "good material", "according price", "cool material", "thank seller", "pretty good", and others. The conclusion of positive sentiment is, the negligee product sold in one of the shops at Shopee is quite liked by buyers. With a fairly cheap price, it turns out that the quality is good, the material is cold when used. The delivery is fast too. The color or motif of the negligee is also following what was ordered.

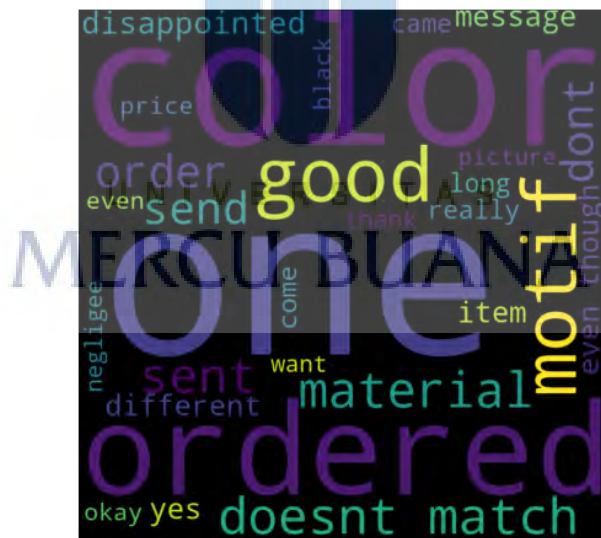


Figure 4. Word Cloud Negative Sentiment

Figure 4 showing the words that appear most often in the negative sentiment class, which are: "color", "one", "ordered", "motif", "doesn't match", "different", "disappointed", "came", "different", "disappointed", "came", "message", "price", "black", "order", "good", "long", "picture", "send", "even", "negligee", "sent", "want", "different", "material", "motif", "dont", "item", "okay", "yes", "doesn't", "match". The conclusion for the negative sentiment is, that the buyer is disappointed because there was a mistake in the order. The patterns and colors ordered do not match what was sent. This could be because the seller is not careful in processing orders, or it could be that the variation chosen by the buyer is empty, but the seller does not confirm and replaces it randomly according to the existing stock, so the buyer feels disappointed.

4. Conclusion

This research has sentiment analysis can be used to find out the general picture of the product based on reviews from customers who have made a purchase. The product discussed in this final project is home clothing for women or what is commonly called a negligee from one of the shops in Shopee. The Naïve Bayes Classifier algorithm can classify reviews on negligee products into positive and negative sentiments with a fairly high accuracy of 90.03%, precision 96.4%, and recall 91.53%. For all the reviews that have been classified, it can be seen which words appear most often using the word cloud in each sentiment class to conclude an overview of the product based on reviews from customers.

From a total of 2907 data obtained as much as 79.77% or 2319 positive sentiments and 20.23% or 588 negative sentiments, so it can be concluded that the buyer's opinion about the negligee product at the store is quite good. On positive reviews, customers like their negligee products because the price is low and the quality of the material is good and cold when worn. Delivery was fast, and the variety ordered was following what was sent. In negative reviews, customers are disappointed because there was an error in their order, the motif or color ordered did not match what was sent.

In addition to the Naïve Bayes classifier, several method have been conducted for sentiment analysis. Research [16] was conducted using the KNN method for sentiment analysis of Shopee application reviews and adding the Jaro Winkler Distance algorithm for word improvement. The test resulted in an accuracy of 0.876, precision of 0.810, a recall of 0.942, and an f-measure of 0.882. Research [17] conducted a sentiment analysis using review data on Google Play in order to compare the accuracy between the Support Vector Machine method and the Decision Tree. Through classification, the accuracy results are 90.20% for the Support Vector Machine method while 89.80% for the Decision Tree method.

Future works will be done by using an algorithm other than the Naïve Bayes Classifier to get the highest accuracy. Then it can be better if implemented as a system or application that automatically performs from data crawling to visualization so that the system can be more useful for various parties.

References

- [1] Sukhwinder and V. Kaur, "Comparative Study on Online," *International Journal of Creative Research Thoughts*, vol. 6, no. 1, pp. 1460–1470, 2018.
- [2] N. Vasic, M. Kilibarda, and T. Kaurin, "The Influence of Online Shopping Determinants on Customer Satisfaction in the Serbian Market," *International Journal of Creative Research Thoughts*, vol. 14, no. 2, pp. 0–0, 2019, doi: 10.4067/s0718-18762019000200107.
- [3] S. R. Reddy. V., D. V. L. N. Somayajulu, and A. R. Dani, "Classification of Movie Reviews Using Complemented Naive Bayesian Classifier," *International Journal of Intelligent Computing Research*, vol. 2, no. 3, pp. 148–153, 2011, doi: 10.20533/ijicr.2042.4655.2011.0019.
- [4] R. A. Rangsang and H. Millayani, "The Effect Of Online Consumer Review On Customer Purchase Decision Process In The E-commerce Site Bibli. com," *e-Proceeding of Management*, vol. 8, no. 6, pp. 8501–8513, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/management/article/view/17071>.
- [5] M. Rezki, D. N. Kholifah, M. Faisal, P. Priyono, and R. Suryadithia, "Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naïve Bayes," *Jurnal Infortech*, vol. 2, no. 2, pp. 264–270, 2020, doi: 10.31294/infortech.v2i2.9286.
- [6] A. K. Janah, E. D. Wahyuni2, and A. A. Arifiyanti, "Klasifikasi Emosi Ulasan Aplikasi Traveloka," *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, vol. 1, no. 3, pp. 716–722, 2020.

- [7] D. A. Muthia, "Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, vol. 2, no. 2, pp. 39–45, 2017.
- [8] A. Febriyanti, "Analisis Sentimen Persepsi Pengguna Jne Menggunakan Algoritma Naive Bayes Classifier," no. 16522259, 2018.
- [9] S. Ailiyya, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Tokopedia Menggunakan Support Vector Machine," vol. 3, no. 2017, pp. 54–67, 2020, [Online]. Available: <http://repositorio.unan.edu.ni/2986/1/5624.pdf>.
- [10] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [11] A. Paputungan, Casey; Jacobus, "Sentiment Analysis of Social Media Users Using Long-Short Term Memory Method," *Jurnal Teknik Elektro dan Komputer vol.10*, vol. 10, no. 2, pp. 99–106, 2021.
- [12] C. Fadlan, S. Ningsih, and A. P. Windarto, "Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra," *Jurnal Teknik Informatika Musirawas (JUTIM)*, vol. 3, no. 1, p. 1, 2018, doi: 10.32767/jutim.v3i1.286.
- [13] A. T. Hardianti, A. R. Manga, and H. Darwis, "Penerapan metode Naive Bayes pada klasifikasi judul jurnal," *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 2, p. 97, 2018, [Online]. Available: <http://e-journals.unmul.ac.id/index.php/SAKTI/article/view/1838/pdf>.
- [14] S. Juniarsih, E. F. Ripanti, and E. E. Pratama, "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 3, p. 239, 2020, doi: 10.26418/justin.v8i3.39118.
- [15] L. Dennis, F. Ramdhana, T. C. E. Faustine, and R. B. Hendijani, "Influence of Online Reviews and Ratings on the Purchase Intentions of Gen Y Consumers: the Case of Tokopedia," *International Journal of Management (IJM)*, vol. 11, no. 6, pp. 26–40, 2020, doi: 10.34218/IJM.11.6.2020.003.
- [16] L. Shanty Wato Wele Kean, "Analisis Sentimen Review Shopee Berbahasa Indonesia Menggunakan Improved K-Nearest Neighbor dan Jaro Winkler Distance," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 7, pp. 2548–964, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [17] K. A. Rokhman, B. Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," *Journal of Information System Management (JOISM)*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.

KERTAS KERJA

Ringkasan

Pada penelitian ini dilakukan analisa sentimen terhadap ulasan salah satu produk daster yang ada di Shopee menjadi sentimen positif dan negatif. Dataset berjumlah 2907 ulasan yang kemudian diberikan label secara manual berdasarkan sentimennya sebagai data latih. Jumlah data latih adalah 70% dari dataset yaitu sebanyak 2034 ulasan dan sebanyak 30% data atau 873 ulasan digunakan sebagai data uji. Evaluasi model menggunakan confusion matrix untuk menghitung akurasi, presisi dan recall. Akurasi yang didapatkan adalah 90,03% dengan presisi sebesar 96,4% dan recal 91,53%. Setelah model berhasil dibuat dan dievaluasi selanjutnya dilakukan klasifikasi terhadap seluruh dataset untuk visualisasi dan menyimpulkan gambaran umum dari produk berdasarkan masing-masing sentimen.

Dari total 2907 data didapatkan sebanyak 79,77% atau sebanyak 2319 sentimen positif dan 20,23% yaitu sebanyak 588 sentimen negatif, sehingga dapat disimpulkan pendapat pembeli mengenai produk daster pada toko tersebut cukup baik. Pada ulasan positif, pelanggan suka dengan produk dasternya karena harganya murah dan kualitas bahannya bagus dan dingin ketika dipakai. Pengirimannya cepat dan variasi yang dipesan pun sesuai dengan yang dikirimkan. Pada ulasan negatif, pelanggan kecewa dikarenakan ada kesalahan dalam pesannya yaitu motif atau warna yang dipesan tidak sesuai dengan yang dikirimkan.

