

Analisis Sentimen Ulasan Toko *Online* Halona *Beauty Care* Untuk Peningkatan Layanan Menggunakan Algoritma *Naïve Bayes*

Mutia Mustika Rani¹, Feri Candra²

^{1,2}Teknik Informatika, Fakultas Teknik, Universitas Riau, Pekanbaru

^{1,2}Kampus Bina Widyia, Kota Pekanbaru, 28293, Indonesia

Email: mutiamustikarani1807@gmail.com¹, feri@eng.unri.ac.id²

Abstract – The rapid development of technology has had a tremendous impact on human life, including in the field of e-commerce. E-commerce provides users with the opportunity to transact goods and services without distance and time limitations, as long as users have internet access. However, due to the rise of online fraud, it is not uncommon for customers to hesitate to buy products online. This proves that customer trust has a significant effect on the intention to buy products in e-commerce. Reviews from customers will be very influential to prospective customers as a benchmark reference for finding information. This study aims to classify and analyze customer reviews of Halona Beauty Care online stores in Shopee using the Naive Bayes Classifier (NBC) algorithm. The results of the analysis using NBC showed an accuracy of 93%. The results of positive sentiment can be used as a reference for maintaining service quality, while the results of negative sentiment are used as evaluation material for service improvement.

Keywords – *Sentiment Analysis, Naïve Bayes, Review*

Intisari - Pesatnya perkembangan teknologi telah memberikan dampak yang luar biasa bagi kehidupan manusia, termasuk dalam bidang *e-commerce*. *E-commerce* memberikan kesempatan kepada pengguna untuk melakukan transaksi barang dan jasa tanpa batasan jarak dan waktu, selama pengguna memiliki akses internet. Namun, karena maraknya penipuan *online*, tidak jarang pelanggan ragu untuk membeli produk secara *online*. Hal ini membuktikan bahwa kepercayaan pelanggan berpengaruh signifikan terhadap niat membeli produk di *e-commerce*. *Review* dari *customer* akan sangat berpengaruh kepada calon pelanggan sebagai acuan tolok ukur untuk menemukan informasi. Penelitian ini bertujuan untuk mengklasifikasikan dan menganalisis *review customer* terhadap toko *online Halona Beauty Care* yang ada di shopee menggunakan algoritma *Naive Bayes Classifier (NBC)*. Hasil analisis dengan menggunakan NBC menunjukkan hasil akurasi sebesar 93%. Hasil dari sentimen *positive* dapat dijadikan sebagai acuan untuk menjaga kualitas layanan, sedangkan hasil dari sentimen *negative* digunakan sebagai bahan evaluasi untuk peningkatan layanan.

Kata Kunci – Analisis Sentimen, *Naïve Bayes*, Ulasan

I. PENDAHULUAN

Perkembangan teknologi internet telah memberikan dampak yang sangat luar biasa bagi kehidupan manusia saat ini, salah satu contohnya terdapat pada bidang perdagangan yaitu kegiatan membeli dan menjual barang dan jasa secara *online*. Akan tetapi, karena maraknya penipuan *online*, tidak jarang pelanggan ragu untuk membeli produk secara *online*, misalnya produk yang tercantum di halaman toko tidak sesuai dengan produk aslinya. Hal ini membuktikan bahwa kepercayaan pelanggan berpengaruh signifikan terhadap niat membeli di *e-commerce* [1].

Kurangnya rasa kepercayaan pelanggan terhadap *seller* menghadirkan tantangan bagi para pengusaha untuk dapat mengevaluasi layanan. Penjual perlu mengidentifikasi masalah yang

muncul berdasarkan *review* yang diberikan *customer* [2]. Ulasan dari *customer* dapat dijadikan sebagai tolak ukur untuk mendapatkan informasi mengenai suatu komentar terkait layanan pada suatu toko. *Review* dari *customer* biasanya berisi saran maupun keluhan. Tentunya ulasan ini akan sangat berpengaruh kepada calon pelanggan.

Untuk membaca dan mengklasifikasikan setiap ulasan yang diberikan *customer* bukanlah hal yang mudah apabila diproses secara manual. Maka, dibutuhkan suatu metode yang dapat secara cepat dan otomatis dalam mengklasifikasikan ulasan tersebut baik itu *positive* maupun *negative*. Oleh karena itu diperlukan suatu metode khusus untuk mengumpulkan data dalam jumlah yang besar. Metode *web scraping* adalah metode pengumpulan data yang paling efektif[3]. *Web scraping* merupakan teknik untuk mengambil suatu informasi dari internet, umumnya informasi tersebut diambil dari halaman *web*, lalu mengekstrak informasi yang diperlukan dari halaman tersebut untuk tujuan tertentu [4][5].

Kemudian pada proses klasifikasinya menggunakan algoritma *Naïve Bayes*. *Naïve Bayes Classifier* merupakan salah satu algoritma yang dapat mengklasifikasikan dengan cepat. *Naïve Bayes Classifier* juga merupakan salah satu algoritma yang paling efektif dan efisien, bahkan ketika digunakan untuk menganalisis data dalam jumlah besar[6].

Berdasarkan permasalahan yang telah dijelaskan, maka penelitian ini akan menerapkan algoritma *Naïve Bayes Classifier* untuk menganalisa sentimen pada ulasan toko *online* Halona *Beauty Care* yang ada di *e-commerce* Shopee. Penulis memilih *platform e-commerce* Shopee dikarenakan platform ini merupakan salah satu *platform e-commerce* yang paling banyak dikunjungi di Indonesia menurut *iprice.co.id*. Analisis sentimen merupakan salah satu penelitian yang paling sering digunakan dalam *Natural Language Processing* (NLP) sejak awal tahun 2000 [7]. Analisis sentimen adalah proses mengekstrak informasi yang didapat dari berbagai sumber data seperti internet dan beragam *platform* media sosial yang berbentuk pandangan seseorang terhadap suatu isu. Analisis sentimen juga dapat berupa opini kepuasan pelayanan dan kebijakan. Dengan bantuan analisis sentimen, informasi yang tadinya tidak terstruktur diubah menjadi informasi yang lebih terstruktur [8].

II. SIGNIFIKANSI STUDI

A. Studi Literatur

Analisis sentimen yaitu proses mengekstrak suatu informasi yang didapat dari berbagai sumber data seperti internet dan berbagai platform media sosial yang berbentuk pandangan seseorang terhadap suatu isu. Tujuan dari analisis sentimen adalah untuk mengetahui opini dari seorang penulis dengan memperhatikan suatu topik tertentu [8]. Analisis sentimen merupakan salah satu teknik yang digunakan dalam *text mining*. *Text mining* adalah proses untuk memperoleh informasi berkualitas tinggi dari suatu teks dengan cara mempelajari pola statistik [9].

Text mining erat kaitannya dengan *Natural Language Processing* (NLP). NLP mempelajari tentang bagaimana mengembangkan teknologi yang dapat memungkinkan komputer untuk mampu dan memahami makna Bahasa manusia [10]. NLP sering digunakan dalam *text mining* untuk mengidentifikasi kata-kata kunci, topik, serta melakukan analisis sentimen dan klasifikasi teks. NLP disini digunakan pada tahap *preprocessing* yang merupakan langkah penting pada suatu analisis sentimen. Setelah tahap *preprocessing*, perlu dilakukan ekstraksi fitur untuk menghitung frekuensi kemunculan suatu kata menggunakan TF-IDF (Term Frequency-Inverse Document Frequency).

Algoritma *Naïve Bayes Classifier* merupakan salah satu algoritma *text mining* yang digunakan untuk pengklasifikasian teks. Algoritma ini merupakan metode klasifikasi terbaik yang dikemukakan oleh Thomas Bayes dengan memprediksi peluang di masa depan

berdasarkan data di masa lalu [11]. Contoh perhitungan *Naive Bayes Classifier* dapat dilihat sebagai berikut:

TABEL I
CONTOH DATA PERHITUNGAN NAIVE BAYES

No.	Ulasan	Sentimen
1.	Pelayanan bagus	Positif
2.	Admin slow respon	Negative
3.	Packing bagus rapi	Positif
4.	Dikasih bonus	Positif
5.	Produk bocor	Negative

Persamaan probabilitas prior:

$$P(H) = \frac{N_j}{N} \tag{1}$$

Keterangan:

N_j = Jumlah data pada suatu kelas

N = Jumlah total data

Menghitung peluang kemunculan kategori positif dan negative (probabilitas kelas):

$$P(Positif) = \frac{3}{5} = 0,6$$

$$P(Negative) = \frac{2}{5} = 0,4$$

Frekuensi kemunculan *term*:

TABEL II
FREKUENSI KEMUNCULAN TERM

No.	Term	Positif	Negatif
1.	Pelayanan	1	
2.	Bagus	2	
3.	Admin		1
4.	Slow		1
5.	Respon		1
6.	Packing	1	
7.	Rapi	1	
8.	Dikasih	1	
9.	Bonus	1	
10.	Produk		1
11.	bocor		1

Menghitung probabilitas kemunculan *term* kata:

$$P(term|kelas) = \frac{\text{total kemunculan term}+1}{\text{total kelas}} \tag{2}$$

TABEL III
PROBABILITAS KEMUNCULAN TERM

No.	Term	Positif	Negatif
1.	Pelayanan	0,666	0,5
2.	Bagus	1	0,5
3.	Admin	0,333	1

No.	Term	Positif	Negatif
4.	Slow	0,333	1
5.	Respon	0,333	1
6.	Packing	0,666	0,5
7.	Rapi	0,666	0,5
8.	Dikasih	0,666	0,5
9.	Bonus	0,666	0,5
10.	Produk	0,333	1
11.	bocor	0,333	1

Peluang kemunculan kategori positif dan negatif:

1. Pelayanan bagus

$$P(\text{Positif}|\text{contoh}) = P(\text{Positif}) \times P(\text{"Pelayanan"}|\text{positif}) \times P(\text{"bagus"}|\text{positif})$$

$$P(\text{Positif}|\text{contoh}) = 0,6 \times 0,666 \times 1 = 0,3996$$

$$P(\text{Negatif}|\text{contoh}) = P(\text{Negatif}) \times P(\text{"Pelayanan"}|\text{negatif}) \times P(\text{"bagus"}|\text{negatif})$$

$$P(\text{Negatif}|\text{contoh}) = 0,4 \times 0,5 \times 0,5 = 0,1$$

Berdasarkan perhitungan diatas, terlihat bahwa hasil dari peluang kemunculan bernilai positif adalah 0,3996 dan peluang kemunculan bernilai negatif adalah 0,1. Dapat diambil kesimpulan bahwa $0,3996 > 0,1$. Artinya contoh data 1 merupakan sentimen **positif**.

2. Admin slow respon

$$P(\text{Positif}|\text{contoh}) = 0,6 \times 0,333 \times 0,333 \times 0,333 = 0,0221556$$

$$P(\text{Negatif}|\text{contoh}) = 0,4 \times 1 \times 1 \times 1 = 0,4$$

Berdasarkan perhitungan diatas, terlihat bahwa hasil dari peluang kemunculan bernilai positif adalah 0,0221556 dan peluang kemunculan bernilai negatif adalah 0,4. Dapat diambil kesimpulan bahwa $0,0221556 < 0,4$. Artinya contoh data 2 merupakan sentimen **negatif**.

3. Packing bagus rapi

$$P(\text{Positif}|\text{contoh}) = 0,6 \times 0,666 \times 1 \times 0,666 = 0,2661336$$

$$P(\text{Negatif}|\text{contoh}) = 0,4 \times 0,5 \times 0,5 \times 0,5 = 0,05$$

Berdasarkan perhitungan diatas, terlihat bahwa hasil dari peluang kemunculan bernilai positif adalah 0,2661336 dan peluang kemunculan bernilai negatif adalah 0,05. Dapat diambil kesimpulan bahwa $0,2661336 > 0,05$. Artinya contoh data 3 merupakan sentimen **positif**.

4. Dikasih bonus

$$P(\text{Positif}|\text{contoh}) = 0,6 \times 0,666 \times 0,666 = 0,2661336$$

$$P(\text{Negatif}|\text{contoh}) = 0,4 \times 0,5 \times 0,5 = 0,1$$

Berdasarkan perhitungan diatas, terlihat bahwa hasil dari peluang kemunculan bernilai positif adalah 0,2661336 dan peluang kemunculan bernilai negatif adalah 0,1. Dapat diambil kesimpulan bahwa $0,2661336 > 0,1$. Artinya contoh data 4 merupakan sentimen **positif**.

5. Produk bocor

$$P(\text{Positif}|\text{contoh}) = 0,6 \times 0,333 \times 0,333 = 0,0665334$$

$$P(\text{Negatif}|\text{contoh}) = 0,4 \times 1 \times 1 = 0,4$$

Berdasarkan perhitungan diatas, terlihat bahwa hasil dari peluang kemunculan bernilai positif adalah 0,0665334 dan peluang kemunculan bernilai negatif adalah 0,4. Dapat diambil kesimpulan bahwa $0,0665334 < 0,4$. Artinya contoh data 5 merupakan sentimen **negatif**.

Untuk mengukur performa dari algoritma *Naïve Bayes Classifier* dapat diuji menggunakan *confusion matrix*. Dimana data aktual akan dibandingkan dengan data hasil klasifikasi dari model untuk melihat seberapa akurat model bekerja.[12] Tabel IV merupakan tabel dari *confusion matrix*.

TABEL IV
CONFUSION MATRIX

Kelas	Prediksi	
	+	-
+	TP	TN
-	FP	FN

Keterangan:

- TP = *True positive*, yaitu jumlah data dengan kategori *positive* yang diklasifikasi *positive*.
- TN = *True negative*, yaitu jumlah data kategori *negative* yang diklasifikasikan *negative*.
- FP = *False positive*, yaitu jumlah data dengan kategori *positive* diklasifikasikan *negative*
- FN = *False negative*, yaitu jumlah data dengan kategori *negative* diklasifikasikan *positive*.

Confusion matrix dapat digunakan untuk menghitung berbagai *performance metrics* dari kinerja yang dibuat. Terdapat tiga *performance metrics* yang sering digunakan dalam mengukur kinerja model dari sistem yaitu akurasi, presisi dan *recall*. Rumus untuk mendapatkan nilai akurasi, presisi dan *recall* dapat dilihat sebagai berikut:

1. Akurasi

Akurasi adalah kedekatan hasil yang diukur dengan nilai sebenarnya. Pada saat melakukan suatu prediksi, diharapkan dapat memprediksi semua data dengan tepat, tetapi kinerja sistem tidak dipungkiri tidak bisa 100% tepat dan benar sehingga dalam memprediksi harus diukur tingkat akurasinya. Untuk dapat tingkat akurasi dari hasil prediksi maka digunakan persamaan (3) sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

2. Presisi

Presisi yaitu menggambarkan tingkat akurasi data yang diminta menggunakan prediksi yang diberikan oleh model. Presisi adalah rasio prediksi positif yang benar untuk semua prediksi positif. Dari semua kelas positif yang diprediksi dengan benar, berapa banyak data yang merupakan kelas positif sebenarnya. Nilai presisi dapat dihitung dengan persamaan (4) sebagai berikut:

$$Presisi = \frac{TP}{TP+FP} \tag{4}$$

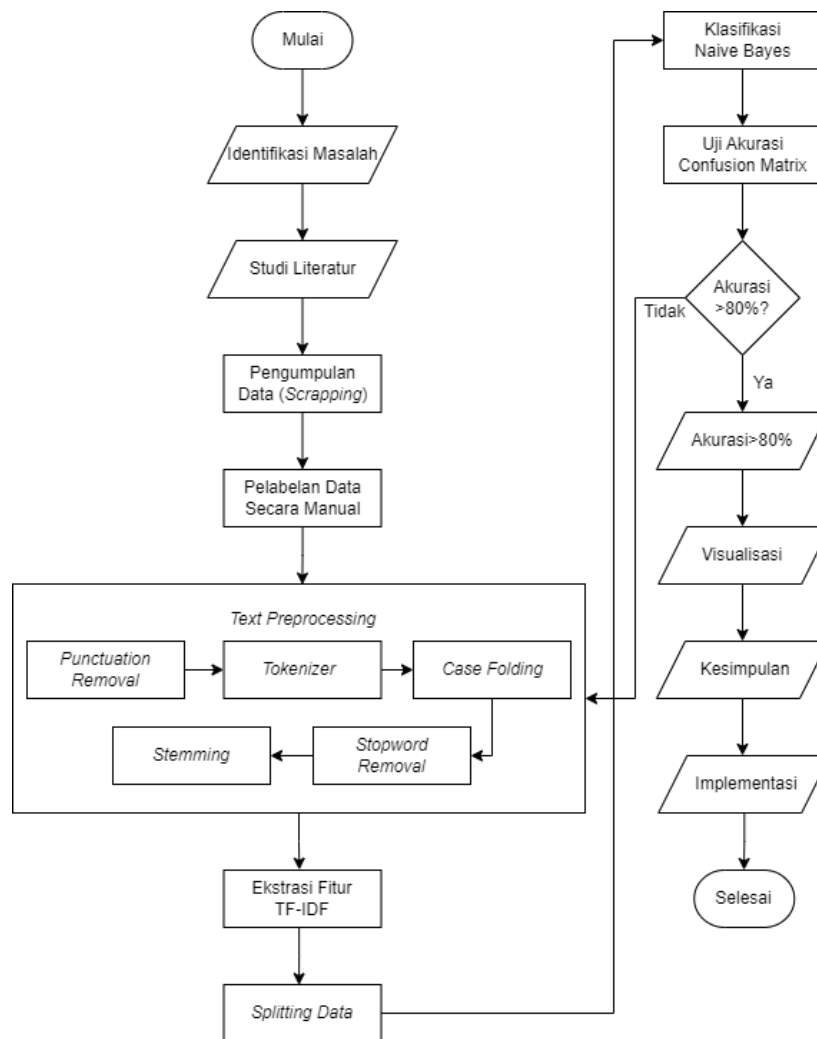
3. Recall

Recall menggambarkan keberhasilan model dalam mengambil informasi. *Recall* adalah persentase prediksi positif yang benar dibandingkan dengan semua data yang dipanggil kembali dengan benar dan dapat diperoleh dengan persamaan (5).

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

B. Tahapan Penelitian

Dalam penelitian ini dilakukan analisis sentimen pada salah satu toko *online* yang terdapat di *e-commerce shopee* yaitu *Halona Beauty Care*. Proses analisis akan menghasilkan klasifikasi teks ulasan berdasarkan kelas yang tercantum pada *dataset*. Untuk melakukan *preprocessing* data, penulis menggunakan Bahasa pemrograman *Python*. Sedangkan untuk pengklasifikasian *dataset* penulis menggunakan algoritma *Naive Bayes Classifier*. Gambar 1 menunjukkan tahapan-tahapan pada penelitian ini:



Gambar 1. Tahapan penelitian

Pada Gambar 1 dapat dilihat bahwa alur pada penelitian ini dimulai dari identifikasi masalah, kemudian dilanjutkan dengan studi literatur untuk mencari teori-teori terkait yang akan dijadikan sebagai acuan. Selanjutnya yaitu mengumpulkan data dengan menerapkan teknik *web scraping* menggunakan Bahasa pemrograman *Python*. Data yang sudah dikumpulkan tersebut dilanjutkan ke tahap *text preprocessing* untuk membersihkan *noise data*. Kemudian dilakukan ekstraksi fitur menggunakan *TF-IDF*. Setelah itu dilakukan *splitting data* dengan perbandingan pembagian 80:20. Tahap selanjutnya yaitu pemodelan menggunakan algoritma *Naive Bayes Classifier* yang kemudian akan diuji performanya menggunakan *confusion matrix*. Jika akurasi tidak mencapai 80% maka periksa kembali tahap *text preprocessing*, karena salah satu faktor yang mempengaruhi akurasi adalah banyaknya data. Faktor lainnya juga terdapat pada *splitting data*, karena banyaknya *data training* juga mempengaruhi hasil dari akurasi. Jika akurasi sudah

mencapai 80% atau bahkan lebih maka dilanjutkan ke tahap visualisai dengan menggunakan *wordcloud*. Kemudian kesimpulan yang sudah didapat dari hasil penelitian dapat diimplementasikan terhadap toko *online* Halona *Beauty Care* agar dijadikan sebagai pembelajaran untuk kedepannya.

III. HASIL DAN PEMBAHASAN

A. Scraping Data

Data didapatkan melalui penerapan teknik *web scraping* dengan menggunakan Bahasa pemrograman *Python*. Teknik ini mengekstrak data dari halaman web HTML dan mengimpornya ke dalam file *Comma Separated Values (CSV)*. Data yang digunakan pada penelitian ini adalah data *review* yang terdapat pada toko *online* Halona *Beauty care* yang ada di *Shopee*. Jumlah data yang terkumpul yaitu sebanyak 2223 data yang merupakan ulasan yang terdapat pada toko Halona. Data yang diambil merupakan data ulasan dari Juli 2020 hingga 15 Mei 2023. Gambar 2 berikut ini memperlihatkan hasil dari *scraping data*.

	A	B	C	D
1	Review			
2	Daya serap:Baikk			
3	Pengirimannyaaa cepatt, pengemasannya juga cepatt, packaging aman pake bubble wrap, sellernyaaa baik banget deh!			
4	Daya serap:bagus			
5	pesanan lengkap ga ada yg kurang,malah dapet free masker,pengiriman cepet,bakal langganan ditoko ini,sellernya juga			
6	Daya serap:			
7	Daya serap:bagus			
8	Produk baik, pengemasan baik, pengiriman baik dan terimakasih seller.			
9	Barang sudah sampai, pengiriman cepat. Barang tidak cacat			
10	Daya serap:ademm betul			
11	Alhamdulillah barang sampe dengan selamat dan bagus bagus banget barangnya dapet gratisan hair mask juga sumpah s			
12	Barangnya original, packaging aman banget pake kardus, pelayanan ramah ðŸ™Š»à™Š,			
13	Mantep pol, suka, selalu belanja disini			
14	Terimakasih suka sekali. Pengiriman cepat. Admin nya juga ramah bangetà™Š, à™Š, semoga cocok biar bisa langganan disi			
15	Bagus banget dikemasnya cepat			
16	Untuk packing allhamdulillah aman, tetapi ada 1kemasan yg bocorr dri pabrik sepertinyaðŸ™Š„			
17	Packaging aman...			
18	MasyaAllah tabarakallah, belanja di halona beauty pokoknya best banget, variasi produknya juga best banget			
19	Daya serap:B			
20	Daya serap:Alhamdulillah			
21	Pengiriman cepat, barang original. Semoga cocok sama maskernya			
22	Daya serap:seperti bagus			
23	Aku kira mengirimannya bakalan lama taunya cepet bangetðŸ™Š-ðŸ™Š-ðŸ™Š-			
24	Pengirimannya cepet banget, pengemasan juga rapii, resellernya ramah thxðŸ™Š*			
25	Daya serap:belum dicoba			
26	Seller nya baik bangetttt dapet klip 3 dan packing nya rapiiii, sukakkk, semoga lancar rezeki nya kakkk ðŸ™ŠœðŸ™ŠœðŸ™Šœ			
27	Barangnya bagussss pengemasan cepat penjual ramah banget à™Š,			
28	Barang sesuai pesanann gadaa yg kurang,pengirimannya cepatt banget,ðapatt bonusnya lagi makasii yaa kak hhe,ga			

Gambar 2. Data Hasil Scraping

B. Pelabelan Data

Pelabelan ini dilakukan oleh seorang ahli Bahasa. Tahap ini dilakukan secara manual selama lebih kurang satu bulan. Dari total 2223 ulasan, terdapat 2115 ulasan dengan sentimen *positive* dan 107 ulasan dengan sentimen *negative*. Tabel I merupakan hasil dari data yang sudah diberikan label.

TABEL V
CONFUSION MATRIX

No.	Ulasan	Label
1.	Senang belanja disini, pengiriman cepat dan aman terus tiap order selalu dapat free gift makasih ya kak	<i>positive</i>
2.	pelayanan slow respon dicat balas nya lama	<i>negative</i>
3.	setiap beli selalu dikasi bonus masker. sellernya baik bangetttttt	<i>positive</i>
4.	Udah pesan lama tapi baru kasih nilai hahaha. Pokoknya toko ini amanah malah dikasih gift juga kita. Terimakasih kk	<i>positive</i>
5.	sedih banget isinya udah keluar2 karna kemasan bocor	<i>negative</i>

C. *Preprocessing*

1. *Case folding*

Tahap ini hampir selalu disertakan Ketika melakukan text preprocessing, karena *case folding* bertujuan untuk merubah huruf besar menjadi *lowercase* agar tidak terdeteksi memiliki perbedaan arti. Kata “Kirim” dan “kirim” diartikan sebagai dua kata yang berbeda oleh program. Seperti yang ditunjukkan oleh Tabel VI.

TABEL VI
CASE FOLDING

Sebelum	Sesudah
Senang belanja disini, pengiriman cepat dan aman terus tiap order selalu dapat free gift makasih ya kak pelayanan slow respon dichat balas nya lama setiap beli selalu dikasi bonus masker. sellernya baik bangetttttt	senang belanja disini, pengiriman cepat dan aman terus tiap order selalu dapat free gift makasih ya kak pelayanan slow respon dichat balas nya lama setiap beli selalu dikasi bonus masker. sellernya baik bangetttttt
Udah pesan lama tapi baru kasih nilai hahaha. Pokoknya toko ini amanah malah dikasih gift juga kita. Terimakasih kk	udah pesan lama tapi baru kasih nilai hahaha. pokoknya toko ini amanah malah dikasih gift juga kita. terimakasih kk
sedih banget isinya udah keluar2 karna kemasan bocor	sedih banget isinya udah keluar2 karna kemasan bocor

2. *Punctuation Removal*

Punctuation removal adalah tahapan untuk pembersihan data dari hal-hal yang tidak diperlukan seperti angka, karakter, tanda baca, *emoticon*, *single character*, *multiple whitespace*, duplikat huruf dan lainnya. Tabel VII menunjukkan hasil dari *punctuation removal*.

TABEL VII
PUNCTUATION REMOVAL

Sebelum	Sesudah
senang belanja disini, pengiriman cepat dan aman terus tiap order selalu dapat free gift makasih ya kak pelayanan slow respon dichat balas nya lama setiap beli selalu dikasi bonus masker. sellernya baik bangetttttt	senang belanja disini pengiriman cepat dan aman terus tiap order selalu dapat free gift makasih ya kak pelayanan slow respon dichat balas nya lama setiap beli selalu dikasi bonus masker sellernya baik banget
udah pesan lama tapi baru kasih nilai hahaha. pokoknya toko ini amanah malah dikasih gift juga kita. terimakasih kk	udah pesan lama tapi baru kasih nilai hahaha pokoknya toko ini amanah malah dikasih gift juga kita terimakasih kk
sedih banget isinya udah keluar2 karna kemasan bocor	sedih banget isinya udah keluar karna kemasan bocor

3. *Tokenizer*

Tokenizer merupakan proses pemisahan teks dalam suatu dokumen menjadi potongan kata independen sehingga tidak saling mempengaruhi.[13] Proses ini juga bertujuan untuk memudahkan dalam menghitung frekuensi kemunculan kata. Tabel VIII menunjukkan hasil dari *tokenizer*.

TABEL VIII
TOKENIZER

Sebelum	Sesudah
senang belanja disini pengiriman cepat dan aman terus tiap order selalu dapat free gift makasih ya kak	senang, belanja, disini, pengiriman, cepat, dan, aman, terus, tiap, order, selalu, dapat, free, gift, makasih, ya, kak
pelayanan slow respon dichat balas nya lama	pelayanan, slow, respon, dichat, balas, nya, lama

setiap beli selalu dikasi bonus masker sellernya baik bangetttttt	setiap, beli, selalu, dikasi, bonus, masker, sellernya, baik, banget
udah pesan lama tapi baru kasih nilai hahaha pokoknya toko ini amanah malah dikasih gift juga kita terimakasih kk	udah, pesan, lama, tapi, baru, kasih, nilai, hahaha, pokoknya, toko, ini, amanah, malah, dikasih, gift, juga, kita, terimakasih, kk
sedih banget isinya udah keluar karna kemasan bocor	sedih, banget, isinya, udah, keluar, karna, kemasan, bocor

4. *Stemming*

Stemming merupakan proses mengubah kata menjadi kata dasar dengan menghilangkan imbuhan pada kata dalam dokumen. Tahap ini juga bertujuan untuk memperkecil jumlah indeks yang berbeda dari suatu data sehingga sebuah kata yang memiliki *suffix* maupun *prefix* akan kembali ke bentuk dasarnya seperti yang ditunjukkan oleh Tabel IX.

TABEL IX
STEMMING

Sebelum	Sesudah
senang, belanja, disini, pengiriman, cepat, dan, aman, terus, tiap, order, selalu, dapat, free, gift, makasih, ya, kak	senang, belanja, sini, kirim, cepat, dan, aman, terus, tiap, order, selalu, dapat, free, gift, makasih, ya, kak
pelayanan, slow, respon, dichat, balas, nya, lama	layan, slow, respon, chat, balas, nya, lama
setiap, beli, selalu, dikasi, bonus, masker, sellernya, baik, banget	setiap, beli, selalu, kasi, bonus, masker, seller, baik, banget
udah, pesan, lama, tapi, baru, kasih, nilai, hahaha, pokoknya, toko, ini, amanah, malah, dikasih, gift, juga, kita, terimakasih, kk	udah, pesan, lama, tapi, baru, kasih, nilai, hahaha, pokok, toko, ini, amanah, malah, kasih, gift, juga, kita, terimakasih, kk
sedih, banget, isinya, udah, keluar, karna, kemasan, bocor	sedih, banget, isi, udah, keluar, karna, kemas, bocor

5. *Stopword Removal*

Tujuan dari proses ini yaitu untuk menghapus kata-kata yang tidak berguna, seperti kata-kata yang sering muncul tetapi tidak berpengaruh terhadap ekstrasi sentimen nantinya. Contohnya seperti kata “dan”, “yang”, “terus”, “juga” dan lainnya. Tabel X menunjukkan hasil dari *stopword removal*.

TABEL X
STOPWORD REMOVAL

Sebelum	Sesudah
senang, belanja, sini, kirim, cepat, dan, aman, terus, tiap, order, selalu, dapat, free, gift, makasih, ya, kak	senang, belanja, sini, kirim, cepat, aman, order, dapat, free, gift
layan, slow, respon, chat, balas, nya, lama	layan, slow, respon, chat, balas, lama
setiap, beli, selalu, kasi, bonus, masker, seller, baik, banget	setiap, beli, bonus, masker, seller, baik, banget
udah, pesan, lama, tapi, baru, kasih, nilai, hahaha, pokok, toko, ini, amanah, malah, kasih, gift, juga, kita, terimakasih, kk	udah, pesan, lama, baru, kasih, nilai, toko, amanah, malah, kasih, gift
sedih banget isinya udah keluar karna kemasan bocor	sedih banget isinya udah keluar karna kemasan bocor

D. *Ekstraksi Fitur*

Data review yang telah melewati tahap *text preprocessing* masih berupa teks. Saat menganalisis data, klasifikasi harus dalam bentuk numerik atau angka. Maka dari itu, data dilakukan konversi terlebih dahulu ke bentuk numerik menggunakan pembobotan kata TF-IDF.

Jika sudah dilakukan ekstraksi fitur maka tahap selanjutnya yaitu *splitting data*. Gambar 3 menunjukkan hasil dari pembobotan kata menggunakan TF-IDF.

```
TfidfTransformer()
(0, 1799)    0.6173562194023555
(0, 1402)    0.35776038243550623
(0, 1159)    0.3482861756272978
(0, 976)     0.24949458056779805
(0, 360)     0.23250355591164756
(0, 282)     0.1987055910401008
(0, 202)     0.3675749409640975
(0, 41)      0.28049037169237073
(1, 861)     0.4119111373827993
```

Gambar 3. Hasil Pembobotan Kata

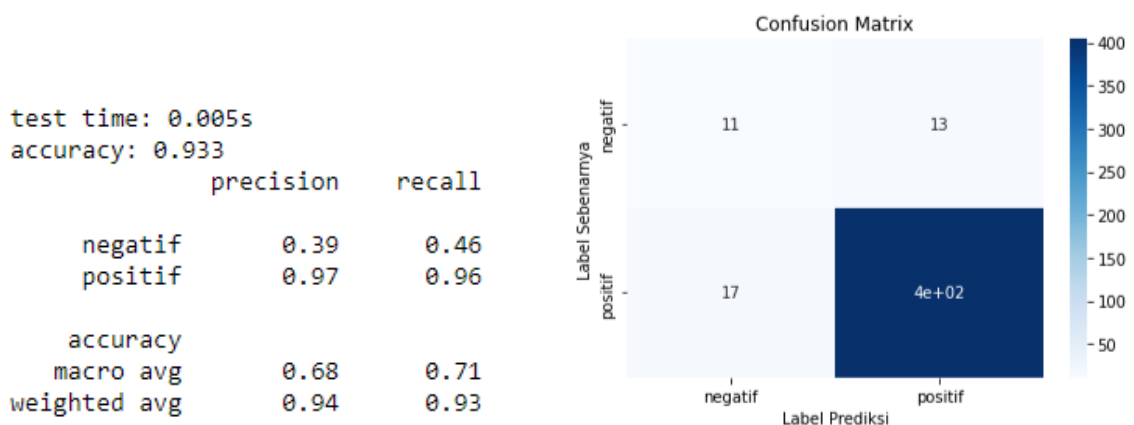
Dilihat dari Gambar 3. Angka yang terletak pada sebelah kiri merepresentasikan indeks dari dokumen, angka 0 merupakan kalimat pertama dari dokumen. Angka 1 merupakan kalimat kedua dari dokumen. Angka yang berada di tengah merepresentasikan indeks dari feature name, 1799 artinya kalimat pada indeks ke-0 mengandung token indeks ke-1799. 861 artinya kalimat pada indeks ke-2 mengandung token indeks ke-861. Sedangkan angka di sebelah kanan merepresentasikan bobot dari TF-IDF. 0,6173562194023555 merupakan bobot dari token ke-1799. Dan 0,4119111373827993 merupakan bobot dari token ke-861.

E. Splitting Data

Splitting data bertujuan untuk membagi dua bagian yaitu *data training* dan *data testing* dengan perbandingan 80:20. 80% sebagai *data training* dan 20% sebagai *data testing*. Total data yang digunakan sebagai *data training* yaitu sebanyak 1778 dan 445 sebagai *data testing*.

F. Klasifikasi Naïve Bayes

Setelah melakukan *text preprocessing* dan ekstraksi fitur selanjutnya dilakukan tahap klasifikasi dengan menggunakan algoritma *Naïve Bayes Classifier*. Proses ini bertujuan untuk mengkategorikan data ke dalam sentimen *positive* dan sentimen *negative*. Untuk menguji kinerja algoritma *Naive Bayes*, dilakukan pengujian menggunakan *confusion matrix*. *Confusion matrix* adalah salah satu *tools* penting dari metode evaluasi yang digunakan pada *machine learning* [14]. Gambar 4 menggambarkan hasil dari pengujian penggunaan metode *confusion matrix*:



Gambar 4. Hasil *Confusion Matrix*

Dari Gambar 4 dapat dilihat bahwa ulasan yang diprediksi benar berjumlah 404 ulasan, ulasan *positive* yang diprediksi salah berjumlah 13 ulasan, ulasan *negative* yang diprediksi benar berjumlah 17 ulasan dan ulasan *negative* yang diprediksi salah berjumlah 11 ulasan. Dari penjabaran tersebut diperoleh hasil akurasi sebesar 93%, nilai presisi 68 dan recall 71%.

G. Visualisasi

Visualisasi bertujuan untuk mengekstrak informasi berupa topik yang paling banyak dibicarakan oleh pengunjung toko *online Halona beauty care*, sehingga dari banyaknya ulasan yang ada dapat diambil informasi yang dianggap penting. Pada penelitian ini, hasil analisis klasifikasi divisualisasikan menggunakan *wordcloud*. *Wordcloud* adalah representasi data yang menunjukkan kumpulan kata-kata penting yang sering muncul dalam data tersebut. Semakin besar kata tersebut muncul di *wordcloud* maka semakin besar pula frekuensi kemunculan kata tersebut dalam data [15]. Gambar 5 berikut ini merupakan *wordcloud* dari hasil klasifikasi keseluruhan ulasan.



Gambar 5. *Wordcloud* Keseluruhan

Wordcloud ulasan *positive* dapat dilihat pada Gambar 6 berikut ini:



Gambar 6. *Wordcloud* Ulasan *Positive*

Wordcloud ulasan *positive* dapat dilihat pada Gambar 7 berikut ini:



Gambar 7. *Wordcloud* Ulasan *Negative*

Berdasarkan Gambar hasil keseluruhan *wordcloud*, *wordcloud* ulasan *positive* dan *wordcloud* ulasan *negative* diperoleh informasi bahwa kata yang sering muncul pada keseluruhan *wordcloud* dan *wordcloud* sentimen *positive* hampir serupa. Terdapat kata “kirim”, “cepat”, “bonus”, “bagus”, dan “aman” untuk 5 kata teratas, artinya keseluruhan ulasan pada toko *online* halona *beauty care* didominasi oleh ulasan-ulasan *positive*. Sedangkan pada *wordcloud* sentimen *negative* kata-kata yang sering muncul yaitu “kirim”, “bocor”, “gak”, “bubblwrap” dan “pecah”, artinya *customer* halona *beauty care* memberikan ulasan bahwa pengiriman lama, produk yang dikirimkan bocor dan pecah, dan *packing* tidak menggunakan *bubblewrap*. Dari ulasan-ulasan tersebut diharapkan kedepannya agar toko *online* halona *beauty care* dapat memperbaiki dan meningkatkan pelayanannya.

IV. KESIMPULAN

Dari penelitian yang dilakukan, dapat disimpulkan bahwa penerapan metode *Naive Bayes Classifier* pada analisis sentimen toko *online* Halona *beauty care* menghasilkan akurasi sebesar 93%. Terdapat 2115 ulasan dengan sentimen *positive* dan 107 ulasan dengan sentimen *negative*. Berdasarkan hasil analisis sentimen dan visualisasi *negative* dapat disimpulkan bahwa beberapa faktor yang dapat diperbaiki dari toko Halona *beauty care* adalah kekecewaan pelanggan terhadap pengiriman, pengemasan dan pelayanan. Sedangkan untuk ulasan-ulasan *positive* yang diberikan *customer* dapat dipertahankan dan ditingkatkan lagi. Kekurangan pada penelitian ini yaitu tidak menggunakan *spell correction* yang berguna untuk memperbaiki kesalahan dalam penulisan kata.

REFERENSI

- [1] M. R. Picaulty, “Pengaruh Kepercayaan Pelanggan Terhadap Niat Pembelian Gadget Di Shopee Indonesia,” *J. Manaj. Maranatha*, vol. 18, no. 1, pp. 31–40, 2018, doi: 10.28932/jmm.v18i1.1094.
- [2] L. O. Sihombing, Hannie, and B. A. Dermawan, “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.
- [3] Z. Ayu N. G., “Implementasi Naive Bayes Classifier dan Asosiasi Untuk Analisis

- Sentimen Data Ulasan Aplikasi E-commerce Shopee Pada Situs Google Play,” pp. 1–26, 2018.
- [4] F. R. Wibowo, D. S. Rusdianto, and A. Arwan, “Pengembangan Sistem Pengumpulan Promo E-Commerce Berbasis Website Dengan Menerapkan Teknik Web Scraping Dalam Proses Pengambilan Data Promo,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2887–2893, 2019.
- [5] D. D. A. Yani, H. S. Pratiwi, and H. Muhandi, “Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace,” *J. Sist. dan Teknol. Inf.*, vol. 7, no. 4, p. 257, 2019, doi: 10.26418/justin.v7i4.30930.
- [6] M. Ahmad, S. Aftab, S. S. Muhammad, and S. Ahmad, “Machine Learning Techniques for Sentiment Analysis: A Review,” *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 3, pp. 27–32, 2017.
- [7] F. Alberto P., E. Fersini, E. Messina, and B. Liu, *Sentiment Analysis in Social Networks*. Todd Green, 2017.
- [8] T. N. Mahardika and H. Februariyanti, “Analisis Sentimen Ulasan Ekpedisi J&T Expres Menggunakan Algoritma Naive Bayes,” *Anal. Sentimen Ulas. Ekpedisi J&T Expres Menggunakan Algoritm. Naive Bayes*, vol. 5, no. 1, pp. 20–29, 2022.
- [9] A. Deolika, K. Kusriani, and E. T. Luthfi, “Analisis Pembobotan Kata Pada Klasifikasi Text Mining,” *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [10] A. Imron, “Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier,” *Tek. Inform.*, pp. 10–13, 2019.
- [11] C. C. Anggarwal, *Data Classification: Algorithms and Applications*. 2015.
- [12] E. Putra, “Klasifikasi Sentimen Masyarakat Terhadap Transgender Berdasarkan Komentar di Instragram Menggunakan Metode Naive Bayes Classifier,” 2019.
- [13] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, “Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing,” *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [14] A. F. Hidayatullah and A. SN, “Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter,” *Semin. Nas. Inform.*, 2015.
- [15] A. V. Febrianti, “Analisis Sentimen Data Ulasan Pengunjung Objek Wisata Lawang Sewu Kota Semarang Pada Situs Tripadvisor,” 2020.

UCAPAN TERIMA KASIH

Terima kasih untuk semua pihak-pihak terlibat yang sudah membantu serta mendukung penulis dalam pelaksanaan penelitian ini.