

Evaluasi *Naïve Bayes* dan *K-Nearest Neighbor* dalam Klasifikasi Sentimen Ulasan Produk Skincare MSGLOW di Tokopedia

I Kadek Arya Sugianta^{1*}, Ni Kadek Winda Patrianingsih²

^{1,2} Program Studi Informatika Fakultas Bisnis, Sosial, Teknologi, Dan Humaniora Universitas Bali Internasional
Jln. Seroja Gang Jeruk No.9A-Denpasar Timur

¹aryasugianta@iikmpbali.ac.id

²windapatrianingsih@iikmpbali.ac.id

Intisari— Industri skincare di Indonesia mengalami pertumbuhan pesat, terutama dalam segmen perawatan wajah pria. Platform e-commerce seperti Tokopedia menjadi sarana utama bagi konsumen untuk memberikan ulasan produk, sehingga analisis sentimen terhadap ulasan pelanggan menjadi penting dalam memahami persepsi pasar. Penelitian ini mengevaluasi kinerja algoritma *Naïve Bayes* dan *K-Nearest Neighbor* (KNN) dalam klasifikasi sentimen ulasan produk skincare di e-commerce. Data dikumpulkan melalui web scraping dari ulasan pelanggan terhadap produk MSGLOW Men Energizer Facial Wash di Tokopedia, menghasilkan 475 data ulasan. Proses pengolahan teks dilakukan dengan metode text mining, termasuk case folding, tokenizing, stop words removal, dan stemming. Analisis dilakukan menggunakan perangkat lunak RapidMiner, dengan klasifikasi sentimen menjadi positif, negatif, dan netral. Evaluasi model dilakukan menggunakan confusion matrix untuk mengukur akurasi. Hasil penelitian menunjukkan bahwa *Naïve Bayes* memiliki akurasi sebesar 92,16%, lebih unggul dibandingkan KNN yang hanya mencapai 68,82%. Temuan ini menegaskan bahwa pemilihan algoritma yang tepat berperan penting dalam meningkatkan akurasi analisis sentimen di e-commerce. Selain itu, penelitian ini memberikan wawasan bagi industri skincare dalam memahami opini pelanggan dan mengoptimalkan strategi pemasaran berbasis data.

Kata kunci— analisis sentimen, e-commerce, klasifikasi sentimen, *Naïve Bayes*, *K-Nearest Neighbor*

Abstract— The skincare industry in Indonesia has experienced rapid growth, particularly in the men's facial care segment. E-commerce platforms such as Tokopedia serve as a primary medium for consumers to share product reviews, making sentiment analysis crucial for understanding market perceptions. This study evaluates the performance of *Naïve Bayes* and *K-Nearest Neighbor* (KNN) algorithms in sentiment classification of skincare product reviews in e-commerce. Data was collected through web scraping, obtaining 475 customer reviews of MSGLOW Men Energizer Facial Wash from Tokopedia. Text processing was conducted using text mining techniques, including case folding, tokenizing, stop words removal, and stemming. The analysis was performed using RapidMiner, categorizing sentiments into positive, negative, and neutral. Model evaluation was conducted using the confusion matrix to measure accuracy. The results indicate that *Naïve Bayes* achieved an accuracy of 92.16%, outperforming KNN, which only reached 68.82%. These findings highlight the importance of selecting the appropriate algorithm to enhance the accuracy of sentiment analysis in e-commerce. Furthermore, this study provides valuable insights for the skincare industry in understanding customer opinions and optimizing data-driven marketing strategies.

Keywords— sentiment analysis, e-commerce, sentiment classification, *Naïve Bayes*, *K-Nearest Neighbor*

I. PENDAHULUAN

Perkembangan teknologi digital yang pesat telah mendorong transformasi di berbagai sektor, termasuk perdagangan elektronik (*e-commerce*). Di Indonesia, digitalisasi terus mengalami peningkatan, sebagaimana tercermin dalam laporan Kementerian Komunikasi dan Informatika (KOMINFO) yang mencatat tingkat digitalisasi platform mencapai 60% pada tahun 2022, meningkat dari 55% pada tahun 2021[1]. Seiring dengan pertumbuhan ini, jumlah pengguna internet di Indonesia juga meningkat secara signifikan. Berdasarkan data Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), jumlah pengguna internet di Indonesia mencapai 215,63 juta orang

pada tahun 2022–2023, meningkat sebesar 2,67% dibandingkan periode sebelumnya [2]. Peningkatan jumlah pengguna internet ini mencerminkan perkembangan literasi digital yang semakin baik di Indonesia. Kondisi tersebut menunjukkan bahwa semakin banyak individu yang memanfaatkan teknologi digital, salah satunya melalui penggunaan platform *e-commerce* [3].

Pasar *e-commerce* di Indonesia memiliki potensi yang besar, sebagaimana terlihat dari pertumbuhan jumlah platform *e-commerce* setiap tahunnya. Pada tahun 2022, nilai ekonomi *e-commerce* di Indonesia tercatat mencapai US\$59 miliar [4]. Salah satu sektor yang mengalami pertumbuhan pesat dalam *e-commerce* adalah industri skincare, terutama produk perawatan

wajah pria. Faktor-faktor seperti meningkatnya kesadaran akan perawatan diri, pengaruh budaya Korea (*K-wave*), serta pemulihan ekonomi pascapandemi menjadi pendorong utama pertumbuhan pasar ini. Pada tahun 2023, nilai industri kecantikan di Indonesia diperkirakan mencapai USD 5,184 miliar atau sekitar Rp 80,9 triliun. Produk facial wash menjadi salah satu produk perawatan kulit yang paling diminati, dengan 54% responden mengaku menggunakannya sebagai bagian dari rutinitas skincare mereka [5].

Produk perawatan wajah, khususnya *facial wash*, merupakan salah satu produk *skincare* dengan tingkat permintaan tertinggi. Berdasarkan survei, sebanyak 54% responden mengaku menggunakan *facial wash* sebagai produk *skincare* yang paling sering digunakan [6]. Lebih lanjut, penggunaan *facial wash* juga dipengaruhi oleh faktor sosial ekonomi. Pada kelompok sosial ekonomi tinggi, sebanyak 73% responden pria rutin menggunakan *facial wash*, sementara pada kelompok sosial ekonomi menengah dan rendah, angka tersebut masing-masing mencapai 68% dan 49% [7].

MSGLOW merupakan salah satu merek lokal produk *skincare* yang memproduksi dan memasarkan *face wash*. Merek ini didirikan pada tahun 2013 dan terus berkembang hingga meraih penghargaan dari Museum Rekor Dunia Indonesia (*MURI*) pada Agustus 2021 sebagai merek dengan penjualan tertinggi serta jaringan distribusi terluas. Pada tahun 2022, MSGLOW juga dianugerahi penghargaan *WOW Brand* dalam kategori *skincare*. Penjualan MSGLOW melalui *e-commerce* mencapai angka yang signifikan, yaitu 51% atau setara dengan Rp38,1 miliar dalam periode 1–18 Februari 2021, dengan omzet bulanan sebesar Rp600 miliar atau Rp7,2 triliun per tahun[8]. Produk ini menjadi salah satu yang paling laris di Tokopedia, dengan angka penjualan yang tinggi di toko resminya. Namun, meskipun produk ini memiliki banyak peminat, terdapat berbagai ulasan pelanggan dengan sentimen yang beragam. Analisis terhadap ulasan pelanggan menjadi penting untuk memahami persepsi konsumen serta membantu perusahaan dalam meningkatkan kualitas produk dan strategi pemasaran.

Analisis sentimen berbasis *text mining* telah menjadi metode yang umum digunakan untuk mengolah opini pelanggan dalam skala besar. Berbagai algoritma pembelajaran mesin digunakan untuk klasifikasi sentimen, termasuk *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Kedua algoritma ini memiliki karakteristik yang berbeda: *Naïve Bayes* dikenal dengan pendekatan probabilistiknya yang sederhana namun efektif dalam klasifikasi teks, sedangkan KNN bekerja berdasarkan kedekatan antar data dalam ruang vektor. Oleh karena itu, perbandingan performa kedua algoritma ini menjadi penting dalam menentukan metode yang lebih akurat untuk analisis sentimen dalam konteks *e-commerce*.

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma *Naïve Bayes* dan KNN dalam klasifikasi sentimen ulasan produk skincare di *e-commerce*, dengan studi kasus pada produk MSGLOW Men Energizer Facial Wash di Tokopedia.

Data ulasan dikumpulkan menggunakan teknik web scraping, kemudian diproses menggunakan *text mining*, termasuk *case folding*, *tokenizing*, *stop words removal*, dan *stemming* sebelum diklasifikasikan menjadi sentimen positif, negatif, dan netral. Evaluasi model dilakukan menggunakan *confusion matrix* untuk mengukur tingkat akurasi masing-masing algoritma.

Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pemilihan algoritma terbaik untuk analisis sentimen dalam *e-commerce*, serta memberikan wawasan yang lebih dalam bagi industri skincare mengenai preferensi dan pengalaman pelanggan. Dengan pemilihan metode klasifikasi yang lebih akurat, perusahaan dapat mengoptimalkan strategi pemasaran dan meningkatkan kualitas layanan berdasarkan analisis sentimen pelanggan.

II. METODOLOGI PENELITIAN

Sumber Data

Jenis penelitian yang digunakan dalam penelitian ini adalah *deskriptif kuantitatif*. Penelitian *deskriptif kuantitatif* bertujuan untuk mendeskripsikan, meneliti, dan menjelaskan suatu fenomena yang telah dipelajari, serta menarik kesimpulan berdasarkan hasil yang diperoleh[9].

Sumber data dalam penelitian ini adalah data sekunder. Data sekunder merupakan data yang diperoleh bukan secara langsung dari sumber pertama, melainkan melalui pihak kedua. Data ini telah tersedia sebelumnya tanpa memerlukan proses pengumpulan langsung, seperti wawancara, dan dapat berasal dari berbagai sumber, seperti situs web, dokumen penting, dan lainnya[10]. Dalam penelitian ini, data sekunder diambil dari toko MSGLOW yang terdapat di platform *e-commerce* Tokopedia. Pengambilan data dilakukan dengan menggunakan teknik *web scraping*.

Tahapan proses *web scraping* dilakukan dengan menggunakan perangkat lunak *Axiom*. Proses ini dimulai dengan mengunduh dan instalasi ekstensi *Axiom* pada browser Chrome. Setelah instalasi, peneliti mengunjungi situs web yang akan di-*scrape*, yaitu toko MSGLOW pada platform Tokopedia. Selanjutnya, peneliti masuk ke bagian pengaturan, membuka *Developer Tools*, dan memilih opsi *Web Scraper*, yang akan membuka perangkat lunak *Axiom*.

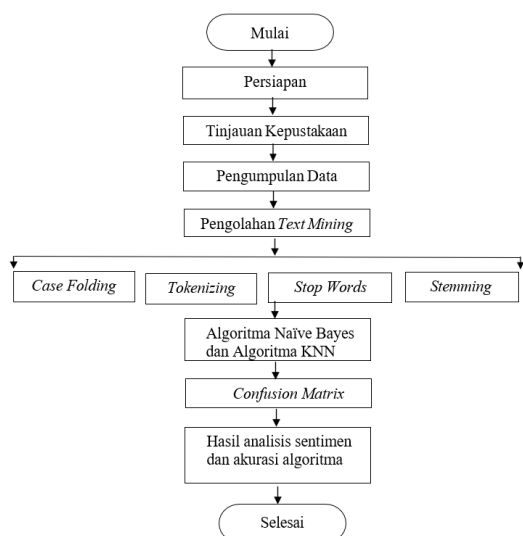
Pada tahap ini, peneliti membuat *sitemap* untuk menentukan alur proses *scraping* dan elemen-elemen yang akan diambil. Proses *scraping* dilakukan dengan memilih elemen-elemen tertentu, seperti teks atau tautan, sesuai kebutuhan penelitian. Dalam penelitian ini, elemen yang diambil terdiri dari teks ulasan dan tautan terkait. Setelah elemen-elemen dipilih, data di-*scrape* dan dapat langsung diunduh atau disimpan dalam format *Google Spreadsheet*.

Penelitian ini mengumpulkan sebanyak 475 ulasan produk skincare MSGLOW Men Energizer Facial Wash dari bulan Oktober 2022 hingga 14 Februari 2024. Proses *scraping* juga

disertai dengan pengaturan *delay* untuk menghindari kelebihan beban pada server situs web serta risiko pemblokiran akibat penggunaan *web scraper*.

Hasil *web scraping* dapat diunduh dalam format data, seperti CSV atau XLSX, yang kemudian digunakan sebagai input untuk perangkat lunak *RapidMiner*. Data tersebut akan diolah melalui tahapan *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) guna mendukung analisis lebih lanjut.

Rancangan Penelitian



Penelitian ini mengevaluasi performa algoritma *Naïve Bayes* dan *K-Nearest Neighbor (KNN)* dalam klasifikasi sentimen ulasan produk skincare MSGLOW yang dijual di platform e-commerce Tokopedia. Tahapan penelitian mencakup persiapan penelitian, penyusunan kajian pustaka, pengumpulan data, serta pengolahan data menggunakan teknik *text mining*.

Proses *text mining* terdiri dari empat tahapan utama: *case folding*, *tokenizing*, *stop word removal*, dan *stemming*. Selanjutnya, dilakukan klasifikasi sentimen menggunakan algoritma *Naïve Bayes* dan *KNN*. Evaluasi performa model klasifikasi dilakukan dengan menggunakan metrik berbasis *confusion matrix*. Rancangan penelitian ini dapat dilihat pada Gambar 1

Gambar 1 Rancangan Penelitian

Pengolahan Text Mining

Text mining merupakan sebuah proses penggalian data, mengeksplorasi serta menganalisis data teks yang kemudian akan dilakukan praproses teks. *Text mining* sendiri merupakan gabungan dan kumpulan bahasa alami yang tidak terstruktur[11]. Praproses teks ini meliputi *case folding*,

tokenizing, *stopwords*, dan *stemming* dengan penjelasan dibawah berikut:

a. Case Folding

Perubahan bentuk kata menjadi bentuk dasar atau huruf kecil (*lower case*) atau huruf besar semua (*UPPERCASE*) hal ini dilakukan agar teks menjadi seragam dan mempermudah dalam pemrosesan, sehingga meminimalisir terjadi nya ambiguitas huruf kapital. Hasil dari *case folding*, dengan merubah menjadi huruf kecil semua, "cocok dipakai utk kulit wajah saya, jadi ngga kering lagi dan lebih cerah. fungsi 3in1 benar² valuable!"

b. Tokenizing

Pada tahap ini dilakukan pemecahan simbol-simbol atau karakter non alphabet, contohnya seperti simbol, spasi, angka serta memisahkan elemen penting agar dapat mempermudah analisis data teks menggunakan tokenize. Berikut hasil proses *tokenizing* yang dapat dilihat pada Tabel 1.

TABEL 1
HASIL TOKENIZING

Komentar	Proses	Hasil
"Cocok dipakai utk kulit wajah saya, jadi ngga kering lagi dan lebih cerah. fungsi 3in1 benar ² valuable!"	Tokenizing	"cocok", "dipakai", "utk", "kulit", "wajah", "saya", "jadi", "ngga", "kering", "lagi", "dan", "lebih", "cerah", "fungsi", "3in1", "benar", "2", "valuable", "!"

c. Stop words

Tahap *stop words* dilakukan untuk menghilangkan kata-kata yang tidak berpengaruh atau tidak berguna dalam hasil akurasi analisis sentimen, sehingga lebih mengutamakan kata-kata yang penting. Berikut hasil dari *stop words* yang dapat dilihat pada Tabel 2.

TABEL 2 HASIL STOP WORDS

Komentar	Proses	Hasil
"Cocok dipakai utk kulit wajah saya, jadi ngga kering lagi dan lebih cerah. fungsi 3in1 benar ² valuable!"	Stop words	"cocok", "dipakai", "kulit", "wajah", "kering", "cerah", "fungsi", "valuable"

d. Stemming

Stemming merupakan tahapan terakhir, dan pada tahapan ini berguna untuk menghilangkan sufiks dan prefiks atau imbuhan dalam kata dan dalam kalimat-kalimat. Berikut hasil dari stemming yang dapat dilihat pada Tabel 3.

TABEL 3 HASIL STEMMING

Komentar	Proses	Hasil
"cocok dipakai utk kulit wajah saya, jadi ngga kering lagi dan lebih cerah. fungsi 3in1 benar ² valuable!"	Stemming	"cocok", "pakai", "kulit", "wajah", "kering", "cerah", "fungsi", "valuable"

Naïve Bayes

Algoritma Naïve Bayes merupakan algoritma klasifikasi yang bekerja dengan menghitung probabilitas statistik. Dalam hal menganalisis sentimen, algoritma Naïve Bayes akan menghitung probabilitas berdasarkan kemunculan kata-kata dalam data[12]. Persamaan algoritma Naïve Bayes dapat dilihat pada Persamaan 1.

$$P(H | X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Keterangan:

- X : Data berupa kelas yang belum diketahui
- H : Hipotesa data X merupakan suatu kelas yang spesifik
- $P(H|X)$: Probabilitas hipotesis data H berdasarkan kondisi
- $X P(H)$: Probabilitas hipotesis H (*prior probability*)
- $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis
- $H P(X)$: Probabilitas X

K-Nearest Neighbor

Algoritma *k-nearest neighbor* atau KNN merupakan algoritma yang bekerja dengan baik dalam klasifikasi dan regresi. Algoritma KNN bekerja dengan mencari tetangga terdekat (*nearest neighbor*) dari data baru, dan memberikan label pada data tersebut. Dari label tetangga (*neighbor*), nantinya akan digunakan untuk memprediksi label data baru yang memiliki jarak terdekat[13]. Rumus dari algoritma KNN dapat dilihat pada Persamaan 2.

$$d(x, y) = \sqrt{(\sum_{i=1}^n (x_i - y_i)^2)} \quad (2)$$

Keterangan:

- d : Jarak
- x : data uji
- y : sampel data
- i : variabel data
- n : dimensi data

Confusion Matrix

Penelitian ini menggunakan metrik evaluasi berbasis *confusion matrix*. *Confusion matrix* merupakan metode yang sesuai untuk



mengevaluasi performa model klasifikasi, dengan mengelompokkan hasil evaluasi ke dalam empat kategori, yaitu *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, dan *False Negative (FN)*[14]. Penelitian ini bertujuan untuk menganalisis sentimen pelanggan terhadap produk skincare MSGLOW Men Energizer Facial Wash, serta menentukan apakah produk tersebut dapat dikategorikan sebagai produk yang berkualitas berdasarkan data penjualan yang mencapai lebih dari 8 ribu unit.

III. HASIL DAN PEMBAHASAN

Pengumpulan Data

Pada penelitian Komparasi Algoritma Naïve Bayes Dan Algoritma *K-Nearest Neighbor* Dalam Menganalisis Sentimen Produk Skincare Msglow Men Energizer Facial Wash Pada E-Commerce Tokopedia Dengan Software Rapidminer, pengumpulan data dilakukan dengan *web scraping* ulasan-ulasan produk MSGLOW facial wash energizer men yang terdapat di toko MSGLOW di Tokopedia. Produk tersebut merupakan produk yang paling laris di toko MSGLOW di Tokopedia, dengan berbagai macam review, baik review yang positif, negatif, dan juga tergolong netral.

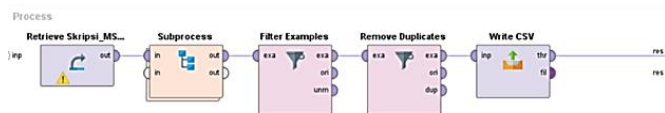
Data yang telah terkumpul kemudian di proses dan di masukkan kedalam Excel dengan format ".csv". Untuk memvalidasi aktual penelitian, digunakan dua algoritma, yaitu naïve bayes dan *K-Nearest Neighbour* yang juga akan dibandingkan, untuk mengetahui dari kedua algoritma tersebut, manakah yang lebih akurat, sehingga dapat lebih baik dalam mengvalidasi aktual data penelitian. Pada proses validasi, terbukti bahwa kedua algoritma memiliki tingkat keakuratan yang berbeda.

Data Cleaning

Pada penelitian produk MSGLOW facial wash enegerizer men memiliki beberapa tahapan proses yang dilakukan

menggunakan *software* Rapidminer, mulai dari *data cleaning* yang didalamnya mencakupi *text processing*, dilanjutkan dengan ke tahap pengujian, dan akhirnya mendapatkan hasil. Bagian pertama yang harus dilakukan ialah dengan pembersihan data atau *data cleaning*, pembersihan data yang diambil dengan *web scraping* ini di masukkan ke dalam file excel.

Data yang dibersihkan merupakan data yang masih memiliki simbol-simbol unik, kemudian emoji yang kerap sekali digunakan dalam berkomentar. semua itu digunakan dengan menggunakan operator *filter example* dan juga operator *replace*. Kemudian data kosong juga dihilangkan dengan menggunakan operator *filter example*, serta data yang duplikat dihilangkan dengan menggunakan operator *remove duplicates*, setelah data bersih, data disimpan ke dalam format CSV dengan menggunakan operator *write csv*. Proses *data cleaning* dapat dilihat pada Gambar 2.

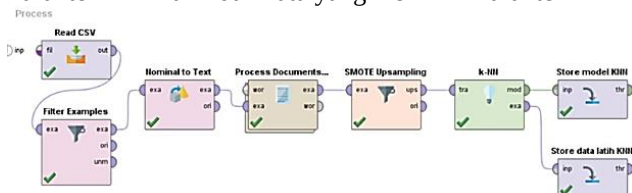


Gambar 2 Proses Data Cleaning

Text Processing

Pada bagian *text processing* digunakan operator *transform cases*, yang digunakan untuk menyetarakan kalimat antara huruf kapital atau huruf, pada penelitian ini digunakan *low case*/huruf kecil. Tokenize digunakan untuk memisahkan kata per kata, agar dapat mempermudah dalam analisis, setelah itu digunakan operator *filter stop words* untuk menghilangkan kata-kata yang tidak berpengaruh atau singkatan-singkatan yang tidak penting yang ada dalam data, dikarenakan Rapidminer belum memiliki dalam versi Bahasa Indonesia, sehingga peneliti menggunakan operator *filter stop words(dictionary)* yang memungkinkan untuk menggunakan file diluar Rapidminer yang menjadi acuan dalam penggunaan operator tersebut, peneliti menemukan *filter stop words* Bahasa Indonesia di Kaggle.

Dari hasil *stop words* yang telah dijalankan, masih terdapat beberapa kata yang belum terhapus, yang merupakan sebuah singkatan seperti “yg”, “bgt”, “iy”, dan seterusnya. Digunakan operator *filter tokens by length* untuk menghapus kata dengan karakter minimal 4 dan kata yang memiliki karakter



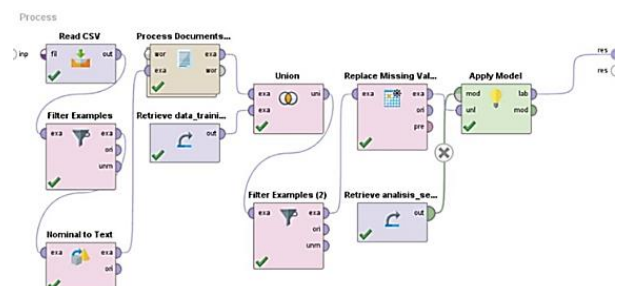
25. Setelah dijalankan, dilanjutkan dengan menggunakan operator *stemming(dictionary)*, yang harus dibuat secara

manual, dikarenakan Rapidminer belum memiliki versi Bahasa Indonesia. Proses *text processing* dapat dilihat pada Gambar 3.

Gambar 3 Proses Text Processing

Model dan Data Latih

Proses untuk menyimpan model serta data latih dari algoritma yang digunakan, yaitu algoritma KNN serta algoritma *Naive Bayes*. Data yang sudah dibagi 70-30 sebelumnya secara manual, diberikan label dengan sentimen positif, negatif, dan juga netral. Data kemudian di saring menggunakan operator *filter example*, penyaringan ini dilakukan agar dapat melatih data dan juga menyimpan model yang sudah dilatih dengan 70% data yang telah diberikan sentimen, yang nantinya



berguna untuk proses selanjutnya, yaitu proses pengujian. Selanjutnya setelah dilakukan *filter example*, dilakukan proses *text processing* dengan menggunakan operator *process documents from data*, digabungkan dengan operator *SMOTE upsampling* yang berguna untuk menyeimbangkan data sentimen, kemudian dijalankan dengan menambahkan algoritma KNN dan juga algoritma *Naive Bayes*. Algoritma KNN serta algoritma *Naive Bayes* di tambahkan ke dalam proses untuk di latih dengan data yang telah di bersihkan dan dibagi sebagai data latih. Proses penyimpanan model dan data latih dapat dilihat pada Gambar 4.

Gambar 4 Proses Penyimpanan Model dan Data Latih

Penggabungan Data

Pada tahap awal, data disaring untuk mengambil 70% data yang telah diberikan sentimen atau label. Proses ini dapat dilihat pada Gambar 5. Selanjutnya, data disaring menggunakan operator *Filter Example* untuk mengambil 30% data yang belum diberikan sentimen, yang akan digunakan sebagai data uji. Data kemudian melewati tahapan *text processing* dengan menggunakan operator *Process Documents from Data*. Untuk menggabungkan kedua jenis data, yaitu data uji dan data latih, digunakan operator *Union*. Data gabungan ini sebenarnya akan menjadi data uji yang telah disaring untuk mengambil data yang belum memiliki label. Selanjutnya, data kosong atau data yang berisi tanda tanya (?) diubah menjadi 0 (nol) menggunakan operator *Replace Missing Values*. Akhirnya, model yang sebelumnya telah disimpan digabungkan dan

diterapkan menggunakan operator *Apply Model*. Proses penggabungan data dapat dilihat pada Gambar 5.

Gambar 5 Proses Penggabungan Data

set (139 examples, 7 special attributes, 450 regular attributes)		Filter (139 / 139 examples): all			
prediction(Sentimen)	text ↓	confidence(positif)	confidence(neutr)	confidence(negatif)	alhamdulillah
positif	wangi kaya bedak wardah cewek original	1	0	0	0
positif	wangi encer kirim cepat	1	0	0	0
positif	udah pake glow iya langganan glow kemarin cari t...	1	0	0	0
positif	udah langganan beli	1	0	0	0
negatif	udah beli	0	0	1	0
positif	toko langganan packing super aman barang origi...	1	0	0	0
positif	terima kasih seller	1	0	0	0
positif	terima kasih quick response tersedia produk kata	1	0	0	0
positif	terima kasih diterimapacking rapi kirim cepatgoo...	1	0	0	0
netral	terima kasih	0	1	0	0

Hasil Prediksi *Naïve Bayes*

Berbeda dengan algoritma *KNN*, algoritma *Naïve Bayes* mampu memprediksi kalimat yang sama dengan sentimen yang lebih akurat, yaitu memberikan sentimen positif daripada sentimen netral. Salah satu contoh hasil prediksi menggunakan algoritma *Naïve Bayes* dapat dilihat pada Gambar 6.

Gambar 6 Hasil Prediksi Sentimen *Naïve Bayes*

accuracy: 68.82% +/- 5.44% (micro average: 68.82%)					
	true positif	true netral	true negatif	true negatif	class precision
pred. positif	100	10	0	0	90.91%
pred. netral	144	21	7	0	12.21%
pred. negatif	20	1	283	1	92.79%
pred. negatif	0	0	0	0	0.00%
class recall	37.83%	65.62%	97.59%	0.00%	

Pada hasil analisis sentimen produk MSGLOW *facial wash energizer men*, didapatkan tingkat akurasi sebesar 92.16%. Hasil ini sudah tergolong baik, dengan menggunakan algoritma *Naïve Bayes*. Akurasi prediksi algoritma *Naive Bayes* dapat dilihat pada Gambar 7.

Gambar 7 Akurasi Prediksi Algoritma *Naïve Bayes*

Hail Prediksi *K-Nearest Neighbour* (KNN)

Berdasarkan hasil prediksi menggunakan algoritma *K-Nearest Neighbour* (*KNN*) menunjukkan, bahwa kata “wangi kaya bedak wardah cewek original” diprediksi netral, yang sebenarnya merupakan sentimen yang termasuk positif. Hasil prediksi *KNN* dapat dilihat pada Gambar 8.

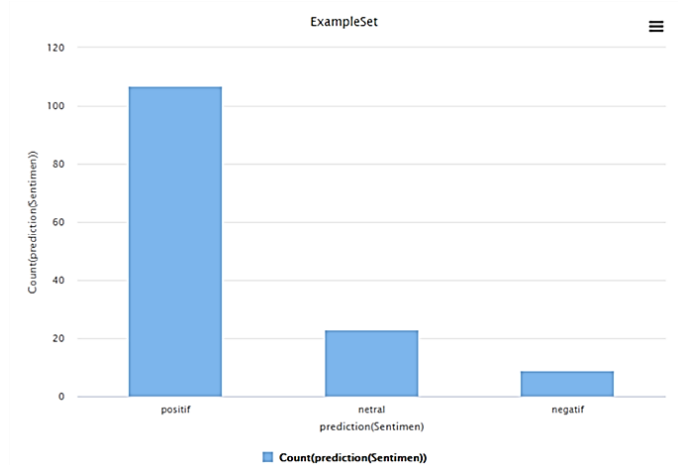
Gambar 8 Hasil Prediksi Sentimen *KNN*

Hasil analisis menggunakan algoritma KNN menunjukkan tingkat akurasi sebesar 68,82%. Tingkat akurasi ini berbeda secara signifikan dibandingkan dengan *algoritma Naïve Bayes*,

accuracy: 92.16% +/- 1.69% (micro average: 92.16%)					
	true positif	true netral	true negatif	true negatif	class precision
pred. positif	233	10	0	0	95.68%
pred. netral	24	18	0	0	42.86%
pred. negatif	7	4	290	1	96.03%
pred. negatif	0	0	0	0	0.00%
class recall	88.26%	56.25%	100.00%	0.00%	

di mana algoritma *Naïve Bayes* memberikan akurasi yang lebih

baik. Oleh karena itu, akurasi *KNN* dapat dianggap kurang



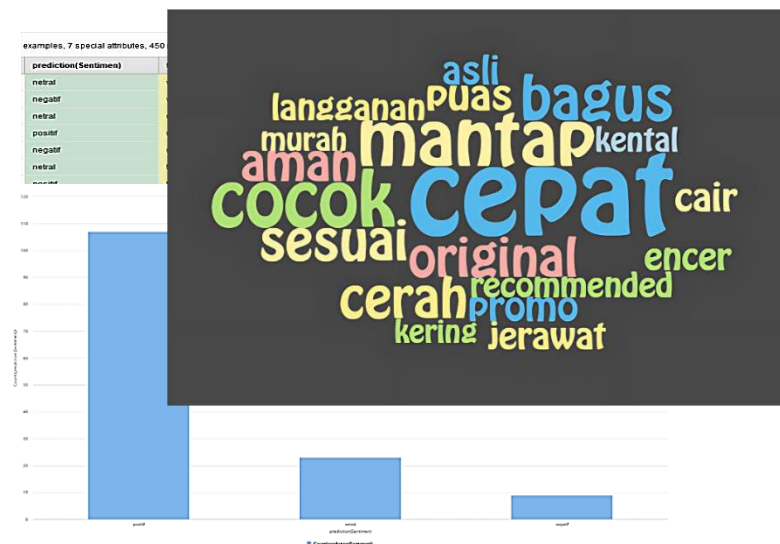
optimal dalam konteks ini. Akurasi prediksi algoritma *KNN* dapat dilihat pada Gambar 9.

Gambar 9 Akurasi Prediksi Algoritma *KNN*

Analisis Algoritma Naïve Bayes dengan K-Nearest Neighbour

Algoritma *Naïve Bayes* berhasil memprediksi sentimen, dan dari hasil analisis sentimen didapatkan 107 berupa sentimen positif, 23 sentimen netral, dan 9 merupakan sentimen negatif, dengan tingkat akurasi sebesar 92.16% yang dapat dilihat pada Gambar Grafik 10 di bawah ini.

Gambar 10 Grafik Hasil Prediksi Algoritma *Naïve Bayes*



Hasil prediksi sentimen dengan menggunakan algoritma KNN, didapatkan hasil dengan 105 berupa sentimen positif, 23 sentimen netral dan 11 sentimen negatif. Tingkat keakuratan

nya sebesar 68.82% yang dapat dilihat pada Gambar Grafik 11 di bawah ini.

Gambar 11 Grafik Hasil Prediksi Algoritma *K-Nearest Neighbour*

Word Cloud

Berdasarkan hasil yang diperoleh, kedua algoritma, *Naïve Bayes* dan *K-Nearest Neighbor (KNN)*, meskipun memiliki tingkat akurasi yang berbeda, tetap dapat menghasilkan output yang sejalan dengan hasil *word cloud* dalam analisis sentimen ulasan produk skincare MSGLOW di Tokopedia. Sebanyak 20 kata dengan frekuensi kemunculan tertinggi dalam keseluruhan dokumen dipilih untuk dianalisis.

Word cloud yang digunakan dalam penelitian ini tidak merepresentasikan frekuensi kata mentah, melainkan telah dihitung berdasarkan *Term Frequency-Inverse Document Frequency (TF-IDF)* sebagai bagian dari proses *text processing*. Dengan demikian, *word cloud* yang dihasilkan menampilkan kata-kata penting dengan nilai frekuensi tinggi dalam data ulasan yang dianalisis. Semakin besar ukuran kata dalam *word cloud*, semakin tinggi frekuensi kemunculannya dalam ulasan yang diteliti. Gambar 12 merupakan hasil dari *word cloud*.

Gambar 12 *Word Cloud*

IV. KESIMPULAN

Berdasarkan hasil analisis sentimen, produk *MSGLOW Facial Wash Energizer Men* tergolong memiliki kualitas yang baik, dengan proses pengiriman yang cepat serta sesuai dengan deskripsi produk. Sebagian besar pengguna menyatakan bahwa produk ini cocok digunakan dan memberikan efek mencerahkan pada wajah. Namun, terdapat inkonsistensi dalam tekstur produk, di mana beberapa pengguna melaporkan konsistensi yang lebih encer atau cair, sementara yang lain mendapatkan tekstur yang lebih kental. Hal ini menunjukkan adanya variasi dalam produksi yang dapat memengaruhi pengalaman pengguna.

Hasil analisis sentimen menggunakan perangkat lunak *RapidMiner* menunjukkan bahwa algoritma *Naïve Bayes* memiliki akurasi lebih tinggi dalam klasifikasi sentimen ulasan produk *MSGLOW Facial Wash Energizer Men* di Tokopedia. Dari 463 ulasan yang telah melalui tahap pra-pemrosesan, algoritma *Naïve Bayes* mencapai akurasi 92,16%, sedangkan algoritma *K-Nearest Neighbors (KNN)* hanya mencapai akurasi 68,82%. Perbedaan ini menunjukkan bahwa *Naïve Bayes* memiliki performa yang lebih unggul dibandingkan *KNN* dalam menganalisis sentimen ulasan produk ini.

Untuk penelitian selanjutnya, disarankan beberapa pengembangan sebagai berikut:

1. Menggunakan dataset yang lebih besar dan beragam guna meningkatkan generalisasi model dalam klasifikasi sentimen ulasan.

2. Mengeksplorasi algoritma lain seperti *Support Vector Machine (SVM)* atau *Deep Learning* untuk meningkatkan akurasi prediksi.
3. Mengintegrasikan metode *aspect-based sentiment analysis* guna memahami aspek spesifik produk yang paling memengaruhi kepuasan pelanggan.
4. Melakukan analisis lebih lanjut terhadap faktor-faktor yang menyebabkan variasi tekstur produk, sehingga dapat memberikan rekomendasi perbaikan kepada pihak produsen.

Dengan pengembangan lebih lanjut, diharapkan hasil analisis sentimen ini dapat memberikan wawasan yang lebih akurat bagi produsen dalam meningkatkan kualitas produk serta pengalaman pelanggan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Bali Internasional atas dukungan yang diberikan dalam proses penelitian ini. Penghargaan juga disampaikan kepada Tokopedia sebagai penyedia data ulasan produk yang digunakan dalam penelitian ini.

Selain itu, penulis berterima kasih kepada rekan-rekan dosen di Program Studi Informatika serta pihak-pihak yang telah memberikan masukan berharga dalam penyusunan artikel ini. Dukungan, bimbingan, serta diskusi yang konstruktif dari berbagai pihak sangat membantu dalam penyelesaian penelitian ini.

REFERENSI

- [1] Yunita., "Digitalisasi Adalah Kunci," *Kominfo*, 2023. <https://www.komdigi.go.id/berita/sorotan-media/detail/digitalisasi-adalah-kunci>
- [2] Apjii, "Survei Apjii Pengguna Internet Di Indonesia Tembus 215 Juta Orang," 2023. <https://apjii.or.id/berita/d/survei-apjii-pengguna-internet-di-indonesia-tembus-215-juta-orang>
- [3] A. T. Natania And R. Dwijayanti, "Pemanfaatan Platform Digital Sebagai Sarana Pemasaran Bagi Ukm," *J. Pendidik. Tata Niaga*, Vol. 12, No. 1, 2024.
- [4] H. A. A Wirapraja, N T Hariyanti, "Kajian Literatur Pengaruh Digital Influencer Marketing Terhadap Perkembangan Strategi Bisnis," *Konstelasi Konvergensi Teknol. Dan Sist. Inf.*, Vol. 3, 2023, [Online]. Available: <https://ojs.uaajy.ac.id/index.php/konstelasi/article/view/7098/3055>
- [5] W. Rafdinal And P. N. Bandung, *Perkembangan Hallyu Di Indonesia : Pemasaran Budaya Korea Selatan Melalui K-Drama Dan K-Pop*, No. February. 2024.
- [6] K. Yuliantari, "Brand Loyalty Perawatan Wajah (Skin Care) Wardah Gentle Face Wash," Vol. 8, No. 2, 2020.
- [7] P. A. Kusuma, A. Natalina, A. N. Kakiay, U. Gunadarma, U. Gunadarma, And U. Gunadarma, "Faktor-Faktor Yang Mempengaruhi Keputusan Pembelian Produk Facial Wash Garnier Men," Vol. 3, No. 1, Pp. 63–70, 2024.
- [8] T. Ruwanti, B. Nirwana, And H. Muslimin, "Analisis Strategi Komunikasi Pemasaran Ms Glow Kepada Konsumen Melalui Instagram Dalam Meningkatkan Brand Awareness," No. 2, Pp. 1–14, 2024.
- [9] A. Syarifudin, "Pengembangan Media Pembelajaran Interaktif

- Berbasis Android Pada Pembelajaran Tematik Kelas Iv Sd/Mi,” Vol. 2507, No. February, Pp. 1–9, 2020.
- [10] A. Dia *Et Al.*, “Efektivitas Metode Edutainment Ular Tangga Terhadap Peningkatan Aktivitas Belajar Siswa Pada Pembelajaran Ppkn Di Upt Sd Negeri 74 Gresik,” *Cendikia Pendidik.*, Vol. 2, No. 2, 2023.
- [11] N. Hardi *Et Al.*, “Analisis Sentimen Physical Distancing Pada Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier,” Vol. 10, Pp. 131–138, 2021.
- [12] M. Farid, S. Wibowo, N. F. Puspitasari, And B. Satya, “Penerapan Data Mining Dan Algoritma Naïve Bayes Untuk Metode Klasifikasi Abstraksi Pendahuluan Tinjauan Pustaka Metode Penelitian,” Vol. 3, No. 2, 2022.
- [13] R. Sakti And A. Daulay, “Analisis Kritis Dan Pengembangan Algoritma K-Nearest Neighbor (Knn): Sebuah Tinjauan Literatur,” Vol. 4, No. 2, Pp. 131–141, 2024.
- [14] A. Dewi, A. Safira, And Y. Yamasari, “Penerapan Algoritma Naïve Bayes (Nb) Untuk Klasifikasi Penyakit Jantung,” *Jinacs*, Vol. 05, Pp. 447–455, 2024.