

**ANALISIS SENTIMEN REVIEW PRODUK KECANTIKAN DI SHOPEE
MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR***

LAPORAN TUGAS AKHIR

Laporan ini disusun guna memenuhi salah satu syarat untuk menyelesaikan
Program Studi Teknik Informatika S-1 pada Fakultas Teknologi Industri
Universitas Islam Sultan Agung



Disusun Oleh :

**IRENA DIVA WIKY PUTRI
32602100059**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG
2025**

**SENTIMENT ANALYSIS OF BEAUTY PRODUCT REVIEWS ON
SHOPEE USING THE K-NEAREST NEIGHBOR ALGORITHM**

FINAL PROJECT

This report was prepared to fulfill one of the requirements for completing the Undergraduate Informatics Engineering Study Program at the Faculty of Industrial Technology, Sultan Agung Islamic University.



Arranged by:

IRENA DIVA WIKY PUTRI

32602100059

**MAJORING OF INFORMATICS ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
SULTAN AGUNG ISLAMIC UNIVERSITY**

2025

**LEMBAR PENGESAHAN
TUGAS AKHIR**

**ANALISIS SENTIMEN REVIEW PRODUK KECANTIKAN DI SHOPEE
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR**

**IRENA DIVA WIKY PUTRI
NIM 32602100059**

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 27 Februari 2025

TIM PENGUJI UJIAN SARJANA :

Moch Taufik, ST., MIT

NIK. 210604034

(Ketua Penguji)

10 - 3 - 2025

Mustafa, ST, MM, M. Kom

NIK. 210610040

(Anggota Penguji)

11 - 3 - 2025

Badie'ah ST., M. Kom

NIK. 210615044

(Pembimbing)

11 - 3 - 2025

Semarang, 13 Maret 2025

Mengetahui,

Kaprodi Teknik Informatika
Universitas Islam Sultan Agung



Moch. Taufik, S.T., M.IT

NIK. 210604034

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Irena Diva Wiky Putri

NIM : 32602100059

Judul Tugas Akhir : Analisis Sentimen Review Produk Kecantikan Di Shopee
Menggunakan Algoritma *K-Nearest Neighbor*

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 24 Februari 2025

Yang Menyatakan



Irena Diva Wiky Putri

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Irena Diva Wiky Putri
NIM : 32602100059
Program Studi : Teknik Informatika
Fakultas : Teknologi Industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : Analisis Senetimen Review Produk Kecantikan di Shopee Menggunakan Algoritma K-Nearest Neighbor

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 24 Februari 2025

Yang menyatakan,




Irena Diva Wiky Putri

KATA PENGANTAR

Segala puji dan syukur saya panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan Tugas Akhir ini dengan judul “Analisis Sentimen Review Produk Kecantikan Di Shopee Menggunakan Algoritma *K-Nearest Neighbor*”. Tugas Akhir ini disusun sebagai salah satu syarat kelulusan dalam menempuh studi serta untuk memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika, Fakultas teknologi Industri, Universitas Islam Sultan Agung Semarang.

Dalam penyusunan Tugas Akhir ini, saya mendapatkan banyak bantuan, baik dalam aspek materi maupun teknis dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, saya ingin mengucapkan terima kasih kepada :

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Hj. Novi Marlyana, S.T., M.T., IPU., ASEAN Eng
3. Dosen Pembimbing Ibu Badie'ah, ST, M.Kom yang telah meluangkan waktu, membimbing, dan memberi ilmu.
4. Orang tua penulis Bapak Sukiyo dan Ibu Widy Astutik yang telah mengizinkan untuk menyelesaikan laporan ini serta memberikan dukungan lahir batin dan doa agar tugas akhir ini dapat berjalan dengan lancar.
5. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan laporan ini masih terdapat banyak kekurangan, untuk itu penulis mengharap kritik dan saran dari pembaca untuk sempurnanya laporan ini. Semoga dengan ditulisnya laporan ini dapat menjadi sumber ilmu bagi setiap pembaca.

Semarang, 24 Februari 2025

Irena Diva Wiky Putri

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN JUDUL	ii
LEMBAR PENGESAHAN TUGAS AKHIR	iii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR	iv
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
ABSTRAK	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	5
1.3 Pembatasan Masalah.....	6
1.4 Tujuan Tugas Akhir.....	6
1.5 Manfaat	6
1.6 Sistematika penulisan.....	7
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	7
2.1 Tinjauan Pustaka	8
2.2 Dasar Teori	8
2.2.1 Analisis Sentimen.....	8
2.2.2 <i>Review</i>	9
2.2.3 <i>Marketplace Shopee</i>	10
2.2.4 <i>Text Preprocessing</i>	11
2.2.5 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	12
2.2.6 <i>K-Nearest Neighbor (KNN)</i>	13
2.2.7 <i>Confusion Matrix</i>	15
BAB III METODE PENELITIAN	16
3.1 Metode Penelitian.....	16
3.1.1 Studi Literature.....	17

3.1.2	Pengumpulan Data	17
3.1.3	<i>Labelling</i> Data.....	17
3.1.4	<i>Text Preprocessing</i>	18
3.1.5	Splitting Data	18
3.1.6	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	19
3.1.7	<i>K-Nearest Neighbor</i>	19
3.1.8	Evaluasi (<i>Confusion Matrix</i>).....	19
3.1.9	Visualisasi	20
3.2	Analisis Kebutuhan	20
3.3	Analisis Sistem.....	22
BAB IV HASIL DAN PEMBAHASAN		24
4.1	Hasil Penelitian	24
4.1.1	<i>Scraping</i> Data.....	24
4.1.2	<i>Labelling</i> Data.....	25
4.1.3	<i>Text Preprocessing</i>	26
4.1.4	<i>Splitting</i> Data.....	28
4.1.5	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	28
4.1.6	<i>K-Nearest Neighbor</i>	29
4.1.7	Evaluasi Model (<i>Confusion Matrix</i>).....	29
4.1.8	Visualisasi	31
4.2	Hasil Implementasi Antarmuka.....	31
4.2.1	Halaman Awal Sistem Prediksi Sentimen Ulasan	31
4.2.2	Halaman hasil Sistem Prediksi Sentimen Ulasan	32
BAB V KESIMPULAN DAN SARAN		30
5.1	Kesimpulan	30
5.2	Saran.....	30
DAFTAR PUSTAKA		
LAMPIRAN		

DAFTAR GAMBAR

Gambar 3. 1 Kerangka Penelitian	16
Gambar 3. 2 Flowchart Analisa Sistem.....	23
Gambar 4. 1 Hasil Pembobotan TF-IDF	29
Gambar 4. 2 Confusion Matrix	30
Gambar 4. 3 Hasil Distribusi Sentimen Berdasarkan Label	31
Gambar 4. 4 Halaman Awal Sistem	32
Gambar 4. 5 Halaman Hasil Prediksi Sentimen Positif	32
Gambar 4. 6 Halaman Hasil Prediksi Sentimen Negatif.....	33



DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	15
Tabel 3. 1 Contoh Hasil Labelling	17
Tabel 3. 2 Contoh Hasil Sebelum dan Sesudah Preprocessing	18
Tabel 4. 1 Hasil dari Scraping	25
Tabel 4. 2 Hasil Labelling	25
Tabel 4. 3 Hasil Text Preprocessing	26
Tabel 4. 4 Classification Report	29



ABSTRAK

Perkembangan teknologi memungkinkan pengguna untuk memberikan *review* terhadap produk yang dibeli di *platform e-commerce* seperti Shopee. Salah satu metode untuk menilai kualitas suatu produk adalah dengan membaca ulasan dari konsumen berdasarkan pengalaman mereka. Saat ini, terdapat berbagai macam produk kecantikan di pasaran, baik dalam hal variasi jenis maupun merek. Akan tetapi, tidak semua produk kecantikan memiliki mutu yang tinggi dan sejalan dengan keperluan pelanggan.. Konsumen yang menulis opini, ulasan, dan pengalaman secara online kini sedang meningkat. Penelitian ini bertujuan dalam mengelompokkan *review* produk kecantikan ke dalam kategori positif atau negatif. Klasifikasi algoritma K-Nearest Neighbor dipilih sebab mengusung konsep yang sederhana serta mudah dipahami dan diterapkan. Hasil penelitian yang dilakukan menghasilkan nilai akurasi sebesar 82%, nilai *precision* sebesar 83%, nilai *recall* sebesar 82% dan nilai *f1-score* sebesar 82%. Hal ini mengindikasikan bahwa model memiliki kinerja yang unggul dalam mengelompokkan ulasan. Hasil analisis ini dapat dimanfaatkan sebagai dasar evaluasi untuk memperbaiki mutu produk.

Kata kunci: Shopee, Review Produk Kecantikan, Analisis Sentiment, K-Nearest Neighbor

ABSTRACT

Technological advancements enable users to provide reviews of products purchased on e-commerce platforms such as Shopee. One method for assessing product quality is by reading consumer reviews based on their experiences. Currently, there are various beauty products available on the market, differing in both type and brand. However, not all beauty products have good quality and meet consumer needs. The number of consumers sharing their opinions, reviews, and experiences online is increasing. This study aims to classify beauty product reviews into positive or negative categories. The K-Nearest Neighbor classification method was chosen because of its simple concept, which is easy to understand and implement. The research results show an accuracy of 82%, a precision of 83%, a recall of 82%, and an F1-score of 82%. This indicates that the model performs well in classifying reviews. The analysis results can be used as an evaluation tool to improve product quality.

Keywords: Shopee, Beauty Product Reviews, Sentiment Analysis, K-Nearest Neighbor

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi dan komunikasi telah mengubah cara pelanggan berbelanja, khususnya dalam industri kecantikan. *Marketplace* online seperti Shopee telah menjadi *platform* utama bagi konsumen untuk membeli produk kecantikan, menawarkan kemudahan dan aksesibilitas yang lebih besar. Dalam konteks ini, ulasan produk dari pengguna menjadi salah satu referensi informasi yang sangat bernilai. Ulasan ini tidak hanya memberikan gambaran tentang kualitas produk tetapi juga mencerminkan pengalaman dan kepuasan konsumen, yang dapat mempengaruhi keputusan pembelian calon pembeli lainnya.

Dalam industri kecantikan, di mana produk sering kali memiliki variasi yang luas dan preferensi yang subjektif, analisis sentimen berperan sebagai alat utama dalam memahami pandangan konsumen. Dengan menggunakan analisis sentimen, penjual dapat mengidentifikasi aspek-aspek produk yang disukai atau tidak disukai oleh konsumen. Hal ini memungkinkan mereka untuk melakukan perbaikan produk dan strategi pemasaran yang lebih baik. Salah satu pendekatan yang dapat diterapkan dalam analisis sentimen yaitu algoritma *K-Nearest Neighbor* (KNN), yang dikenal karena kesederhanaan dan efektivitasnya dalam klasifikasi data (Ridwan dkk., 2024).

Shopee sebagai salah satu *marketplace e-commerce* terkemuka di Indonesia, menawarkan beragam produk kecantikan dari beragam brand dan berbagai toko. Dengan meningkatnya jumlah ulasan yang dihasilkan setiap harinya, tantangan muncul dalam menganalisis dan mengklasifikasikan sentimen dari ribuan ulasan tersebut. Untuk mengetahui apakah produk kecantikan itu disukai atau tidak, kita perlu menganalisis opini dari ulasan pelanggan dengan cara mengelompokkannya menjadi ulasan yang positif dan negatif. Analisis sentimen merupakan proses menganalisis teks yang berasal dari opini, perasaan, dan emosi yang diungkapkan dalam suatu tulisan. Proses ini bertujuan untuk mengklasifikasikan informasi dalam bentuk teks ke dalam kategori positif atau negatif.

Salah satu keunggulan dari metode *K-Nearest Neighbor* adalah kemampuannya untuk memberikan hasil klasifikasi berdasarkan kedekatan data dengan titik data lainnya dalam ruang fitur (Juni dkk., 2024). Dalam analisis sentimen, ulasan dengan kata-kata atau frasa serupa akan dikelompokkan bersama, sehingga memudahkan identifikasi pola-pola umum dalam sentimen konsumen. Penelitian sebelumnya menunjukkan bahwa *K-Nearest Neighbor* dapat mencapai tingkat ketepatan yang memadai dalam pengklasifikasian sentimen, meskipun hasilnya dapat bervariasi tergantung pada kualitas data dan teknik pra-pemrosesan yang digunakan (Amalia dkk., 2023.)

Studi ini menerapkan algoritma *K-Nearest Neighbor* untuk mengkaji opini publik dari ulasan produk kecantikan yang ada di *platform* Shopee. Diharapkan hasil dari penelitian ini tidak hanya memberikan pemahaman lebih lanjut mengenai persepsi konsumen terhadap produk kecantikan tetapi juga memberikan rekomendasi bagi penjual dalam meningkatkan kualitas produk layanan mereka berdasarkan ulasan dari pengguna.

1.2 Perumusan Masalah

Merujuk pada latar belakang permasalahan, oleh karena itu dapat diidentifikasi permasalahan yang akan dikaji adalah sebagai berikut:

1. Sentimen konsumen terhadap produk kecantikan di *platform* Shopee dapat diklasifikasikan menggunakan algoritma *K-Nearest Neighbor*.
2. Sejauh mana algoritma *K-Nearest Neighbor* (KNN) mampu memberikan hasil yang akurat dalam mengidentifikasi sentimen dari ulasan produk kecantikan di Shopee.
3. Metode apa yang dapat diterapkan untuk mengidentifikasi dan mengelompokkan opini pelanggan menjadi sentimen positif atau negatif dari ulasan produk kecantikan di Shopee

1.3 Pembatasan Masalah

Pembatasan ruang lingkup penelitian dari struktur penulisan tugas akhir ini adalah sebagai berikut:

1. Penelitian ini hanya menganalisis ulasan produk sunscreen, moisturizer dan toner yang diambil dari platform *e-commerce* Shopee.
2. Analisis sentimen hanya diklasifikasikan menjadi ulasan positif dan ulasan negatif.

1.4 Tujuan Tugas Akhir

Adapun tujuan dari penelitian tugas akhir ini dari penulisan penelitian tugas akhir ini memiliki tujuan sebagai berikut:

1. Mengimplementasikan algoritma *K-Nearest Neighbor* untuk analisis sentimen pada *review* produk.
2. Mengevaluasi performa algoritma *K-Nearest Neighbor* dalam analisis sentimen produk kecantikan.

1.5 Manfaat

Manfaat dari penelitian tugas akhir ini adalah sebagai berikut:

1. Memahami kualitas produk kecantikan berdasarkan ulasan pengguna lain sebelum melakukan pembelian.
2. Hasil analisis sentimen dapat membantu memahami opini dan meningkatkan kualitas produk kecantikan.

1.6 Sistematika penulisan

Struktur penulisan yang akan diterapkan oleh penulis dalam penyusunan laporan tugas akhir ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab 1 membahas urgensi studi yang dibahas, mencakup perumusan latar belakang, perumusan permasalahan, serta batasan ruang lingkup penelitian, sasaran serta manfaat yang diharapkan, serta sistematika penulisan.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab 2 berisi landasan teori yang diterapkan dalam penelitian, termasuk referensi dari penelitian terdahulu yang mendukung perancangan sistem. Selain itu, bab ini membantu penulis dalam memahami teori yang berkaitan dengan *K-Nearest Neighbor* (KNN) sepanjang berlangsungnya penelitian.

BAB III : METODE PENELITIAN

Bab 3 memaparkan proses dan tahapan penelitian, mulai dari pengumpulan dataset hingga tahap pemodelan topik data yang tersedia.

BAB IV : HASIL DAN ANALISIS PENELITIAN

Bab 4 menjelaskan hasil penelitian, termasuk output akhir sistem, pengelompokan data uji, serta tingkat ketepatan sistem.

BAB V : KESIMPULAN DAN SARAN

Bab 5 menguraikan kesimpulan dari hasil penelitian, merangkum seluruh proses yang dilakukan dari tahap awal hingga selesai.

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Algoritma *K-Nearest Neighbor* bekerja dengan cara mengelompokkan data baru berdasarkan kemiripannya dengan data yang sudah ada. Untuk menentukan kelompok data baru, algoritma ini melihat data-data terdekat. Nilai 'k' yang optimal sangat bergantung pada karakteristik data yang digunakan. Semakin besar nilai 'k', semakin kecil kemungkinan adanya kesalahan klasifikasi akibat data yang berisik, namun hal ini juga dapat menyebabkan batas antar kelompok menjadi kurang jelas (Fritama dkk., 2023). Dalam hal ini, hasil penelitian menunjukkan bahwa algoritma *K-Nearest Neighbor* memberikan akurasi tertinggi, sehingga algoritma ini yang digunakan oleh peneliti

Menurut (Khasanah dkk., 2024), *Marketplace* adalah bentuk lanjutan dari *e-commerce*, yang berfungsi sebagai *platform* daring untuk kegiatan bisnis dan transaksi antara penjual dan pembeli. Di *marketplace*, pembeli dapat dengan mudah menemukan produk yang mereka butuhkan. Shopee, sebagai salah satu *marketplace* terkemuka di Indonesia, mencatatkan jumlah pengunjung bulanan tertinggi pada kuartal pertama tahun 2020, yaitu mencapai 71 juta pengunjung. Produk kecantikan merupakan salah satu barang yang dijual di Shopee. Kualitas produk tersebut dapat dinilai melalui ulasan dari pembeli sebelumnya. Ulasan produk menjadi sumber informasi penting mengenai kualitas suatu produk dan memiliki pengaruh besar terhadap keputusan konsumen. Shopee menjadi sumber utama data ulasan yang digunakan oleh peneliti dalam penelitian ini.

Dalam salah satu penelitian (Kristianto, 2024.), Analisis sentimen adalah proses yang digunakan untuk mengolah opini yang diberikan oleh pengguna aplikasi. Selain itu, analisis sentimen juga berperan dalam mengekstrak, mengolah data tekstual, dan menangkap informasi dari data tersebut. Pada penelitian ini pembagian data latih dan data uji yaitu 80:20 mendapatkan hasil nilai akurasi sebesar 80%. Penelitian ini berkaitan dengan pembagian data peneliti menggunakan algoritma *K-Nearest Neighbor*.

Penelitian yang dilakukan oleh (Kristiyanti, 2015), *review* produk merupakan Suatu pandangan yang dapat diandalkan. *Review* produk kecantikan dapat membantu konsumen dalam mengetahui kualitas produk kecantikan tersebut. *Review* produk kecantikan akan dikelompokkan menjadi sentimen positif dan negatif. Hal ini bertujuan agar konsumen dapat dengan cepat dan akurat mengetahui pendapat konsumen lain tentang produk tersebut. Penelitian ini akan menggunakan klasifikasi sentimen positif dan negatif.

Penelitian lain yang dilakukan oleh (Putri dkk., 2024), Dalam algoritma *K-Nearest Neighbor*, suatu data baru dikategorikan berdasarkan kelas mayoritas dari 'tetangga' terdekatnya. Data tersebut dimasukkan ke dalam kelas yang paling banyak muncul di antara 'K' tetangga terdekatnya. Algoritma ini juga mampu menentukan batas keputusan untuk kasus dengan lebih dari dua kelas. Nilai 'K', yang merupakan jumlah tetangga, adalah parameter yang dapat disesuaikan untuk mengatur kompleksitas algoritma. Dalam penelitian ini, nilai 'K' diuji dari 1 hingga 13, dan akurasi terbaik sebesar 83% diperoleh pada 'K' sama dengan 10. Penelitian ini berfokus pada pengaruh pemilihan nilai 'K' terhadap hasil klasifikasi.

Analisis sentimen merupakan teknik dalam *text mining* yang digunakan untuk mengklasifikasikan opini pengguna menjadi sentimen positif atau negatif. Dalam penelitian ini, digunakan metode *Improved K-Nearest Neighbor (Improved K-NN)* yang merupakan pengembangan dari K-NN untuk meningkatkan akurasi klasifikasi teks, serta *Jaro Winkler Distance* yang berfungsi memperbaiki kata tidak baku akibat kesalahan ketik atau penggunaan bahasa gaul. Sebelum proses klasifikasi, data diproses menggunakan TF-IDF (*Term Frequency - Inverse Document Frequency*) untuk memberikan bobot pada setiap kata sehingga hanya kata-kata yang paling relevan yang digunakan. Model ini kemudian dievaluasi dengan *K-Fold Cross Validation*, dan hasil pengujian menunjukkan bahwa pada nilai $k=10$, akurasi tertinggi yang diperoleh adalah 87,6%, dengan precision 0,810, recall 0,942, dan f-measure 0,882 (Kean dkk., 2019). Dalam penelitian ini, menunjukkan bahwa algoritma *K-Nearest Neighbor* memberikan akurasi tertinggi.

Menurut (Pradhana dkk., 2024), penelitian mengenai analisis sentimen dalam *e-commerce* telah berkembang pesat, terutama dalam memahami opini konsumen

terhadap produk tertentu. Studi sebelumnya menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki akurasi tinggi dalam klasifikasi sentimen ulasan produk, dengan beberapa penelitian mencapai lebih dari 90% akurasi. Dalam konteks ulasan produk skincare, pendekatan *text mining* seperti TF-IDF dan Word2Vec digunakan untuk meningkatkan pemahaman terhadap sentimen pengguna. Selain itu, teknik *balancing* data seperti SMOTE juga telah diterapkan untuk mengatasi ketidakseimbangan kelas, meningkatkan performa model klasifikasi

Dalam salah satu penelitian (Mara dkk., 2021), analisis sentimen telah banyak diterapkan dalam berbagai bidang, termasuk pendidikan dan *e-commerce*, dengan tujuan memahami opini pengguna terhadap suatu layanan atau produk. Studi sebelumnya menunjukkan bahwa algoritma *K-Nearest Neighbors* (KNN) dapat memberikan hasil klasifikasi yang baik dalam analisis sentimen, seperti yang diterapkan pada analisis sentimen metode pembelajaran daring di Universitas Kristen Wira Wacana Sumba, di mana KNN berhasil mencapai akurasi 87% dan AUC 0.916.

Penelitian yang dilakukan oleh (Saifurridho & Hayati, 2024), analisis sentimen telah banyak diterapkan dalam berbagai studi untuk memahami opini pengguna terhadap suatu layanan atau produk, termasuk dalam *platform e-commerce* seperti Shopee. Algoritma *K-Nearest Neighbor* (KNN) menjadi salah satu metode yang sering digunakan dalam klasifikasi sentimen karena kemampuannya dalam menangani data dengan jarak kedekatan yang lebih presisi. Penelitian sebelumnya menunjukkan bahwa KNN mampu mencapai akurasi hingga 70% dalam klasifikasi sentimen ulasan aplikasi Shopee di *Google Playstore*, dengan pembagian data latih dan uji 70:30

Penelitian lain yang dilakukan oleh (Prastika dkk., 2021), dalam penelitian mengenai analisis sentimen dalam ulasan produk kecantikan menunjukkan bahwa metode *K-Nearest Neighbor* (KNN) dapat digunakan untuk mengklasifikasikan sentimen berbasis aspek dengan tingkat akurasi yang bervariasi tergantung pada parameter yang digunakan. Studi sebelumnya mengungkapkan bahwa KNN dapat memberikan hasil yang lebih baik jika dikombinasikan dengan teknik seleksi fitur

seperti *Information Gain* (IG), yang berfungsi untuk menghilangkan fitur yang tidak relevan sehingga meningkatkan akurasi klasifikasi. Dalam penelitian yang dilakukan terhadap ulasan produk kecantikan di *platform Female Daily*, implementasi KNN dengan IG sebagai metode seleksi fitur menghasilkan akurasi sebesar 74,21% dengan nilai k optimal sebesar 23. Selain itu, penelitian ini menyoroti pentingnya preprocessing data seperti stemming dan normalisasi dalam meningkatkan performa model klasifikasi. Hasil studi ini menunjukkan bahwa kombinasi KNN dan IG dapat menjadi pendekatan yang efektif dalam analisis sentimen berbasis aspek, meskipun tantangan seperti distribusi data yang tidak seimbang masih perlu diatasi untuk meningkatkan akurasi klasifikasi secara keseluruhan

2.2 Dasar Teori

2.2.1 Analisis Sentimen

Analisis sentimen pada ulasan merupakan proses evaluasi produk di internet untuk mengidentifikasi perasaan atau opini terhadap produk secara keseluruhan. Proses ini berfungsi sebagai tugas klasifikasi yang mengelompokkan orientasi teks ke dalam kategori positif atau negatif. Tujuan utama dari analisis sentimen adalah untuk mengetahui dan menilai opini atau sikap seorang penulis terhadap suatu hal yang spesifik. Perilaku mengindikasikan opini, alasan atau penilaian, kondisi kecenderungan (Atmodjo dkk., 2024).

Analisis sentimen adalah jenis studi komputasi yang didasarkan pada pengolahan bahasa alami dan komputasi linguistik yang melihat pendapat, penilaian, evaluasi, sikap, emosi dan sentimen seseorang terhadap sesuatu seperti barang atau jasa, individu, organisasi, acara, topik dan elemen lainnya. Biasanya, data pengujian berupa ulasan produk *online* yang menunjukkan perasaan emosional, seperti sedih, senang atau marah, untuk menentukan apakah mendapatkan penilaian positif atau negatif (Muktafin dkk., 2020)

Analisis sentimen, atau yang juga dikenal sebagai penggalian opini, adalah cabang ilmu data mining yang digunakan untuk menganalisis, memproses, dan mengekstrak informasi dari teks. Informasi ini dapat berkaitan dengan berbagai hal,

seperti layanan, produk, individu, organisasi, peristiwa, atau topik tertentu. Tujuannya adalah untuk mendapatkan wawasan dari kumpulan data yang ada. Analisis sentimen merupakan bagian dari pemrosesan bahasa alami (NLP) yang bertujuan untuk mengidentifikasi subjektivitas dalam teks, serta mengekstrak dan mengklasifikasikan opini (Rani & Candra, 2023).

Dalam penelitian ini, analisis sentimen dilakukan menggunakan pembelajaran mesin karena kemampuannya untuk memprediksi polaritas sentimen, yaitu positif atau negatif, berdasarkan data pelatihan dan pengujian. Proses analisis sentimen sebagaimana diilustrasikan sebagai teks tidak teratur, mencakup teks *review*. (Rivaldi dkk., 2024)

2.2.2 Review

Review yang umumnya digunakan untuk menilai karya seni atau sastra, bertujuan untuk mengevaluasi suatu produk dan memberikan informasi tentang kelebihan serta kekurangannya. Dalam konteks produk kecantikan, *review* sangat dinantikan karena banyak konsumen, terutama perempuan, yang ingin memastikan kualitas produk sebelum membeli. *Review* dari sumber terpercaya menjadi faktor penting dalam memengaruhi keputusan pembelian pelanggan (Hanafiah dkk., 2023).

Review adalah sebuah ulasan atau evaluasi yang diberikan seseorang terhadap suatu produk, layanan, karya seni, atau pengalaman tertentu berdasarkan pendapat dan pengalamannya. *Review* bertujuan untuk memberikan informasi yang lebih jelas kepada orang lain mengenai kelebihan dan kekurangan dari objek yang ditinjau. Dalam dunia digital, *review* banyak ditemukan di berbagai platform seperti e-commerce, media sosial, blog, dan situs ulasan, dimana pengguna dapat membagikan pengalaman mereka mengenai suatu produk atau layanan.

Review dapat bersifat subjektif maupun objektif. *Review* subjektif didasarkan pada pengalaman pribadi dan preferensi masing-masing individu, sehingga dapat berbeda antara satu dengan yang lain. Sedangkan, *review* objektif lebih mengutamakan fakta, data, atau standar tertentu yang digunakan sebagai acuan dalam menilai suatu produk atau layanan.

Review dapat ditulis dalam berbagai format, mulai dari teks singkat hingga ulasan yang lebih mendalam dan terstruktur. Beberapa *review* menggunakan sistem rating atau skor untuk memberikan gambaran secara kuantitatif mengenai kepuasan pengguna. Selain itu, ada juga *review* yang menyertakan foto, video, atau perbandingan dengan produk serupa untuk memberikan gambaran yang lebih lengkap.

Secara umum, *review* memiliki peran penting dalam membangun kepercayaan antara konsumen dan produsen. Ulasan yang baik dan jujur dapat meningkatkan reputasi sebuah produk atau layanan, sementara *review* yang buruk dapat menjadi masukan bagi produsen untuk melakukan perbaikan. Oleh karena itu, banyak perusahaan dan pelaku bisnis yang memperhatikan *review* pelanggan sebagai bagian dari strategi pemasaran dan peningkatan kualitas layanan.

2.2.3 Marketplace Shopee

Marketplace Shopee adalah *platform e-commerce* yang memungkinkan pengguna untuk membeli dan menjual berbagai produk secara online dengan mudah, menawarkan berbagai fitur seperti pembayaran digital, layanan pengiriman, serta promo dan diskon menarik. *Marketplace* Shopee merupakan *platform e-commerce* yang memungkinkan pengguna untuk membeli dan menjual berbagai produk secara online. *Marketplace* adalah suatu sarana yang digunakan untuk memasarkan layanan melalui internet. Dalam konteks ini, “*place*” mengacu pada tempat atau platform yang memungkinkan pemasaran produk atau layanan melalui internet. *Marketplace* Shopee sebagai salah satu *marketplace* terbesar di Asia Tenggara, Shopee menawarkan fitur-fitur yang memudahkan transaksi online, seperti pembayaran, layanan pengiriman cepat, dan fitur interaktif untuk berinteraksi dengan penjual dan pembeli.

Desain pada Shopee memiliki desain UI yang ramah pengguna dan UX yang responsif, memudahkan pengguna dalam mencari, memilih dan membeli produk yang dibutuhkan serta mengomentari sebuah produk. Shopee menawarkan berbagai metode pembayaran, termasuk transfer bank, *Cash On Delivery* (COD), dan *e-wallet*. Yang memberikan fleksibilitas kepada pengguna. Dengan kerjasama berbagai jasa logistik, Shopee memastikan pengiriman yang cepat dan tepat waktu.

Shopee sering mengadakan diskon besar, gratis ongkir, *cashback* untuk menarik lebih banyak pengguna (Arrasyid dkk., 2024).

2.2.4 Text Preprocessing

Preprocessing adalah tahapan penting dalam klasifikasi teks, yaitu proses membersihkan dan mempersiapkan data teks agar siap dianalisis. Data teks daring sering kali mengandung elemen-elemen yang tidak relevan atau 'noise'. Tujuan preprocessing adalah untuk mengubah data mentah menjadi format yang lebih terstruktur dan relevan (Kamal & Ratnasari, 2023).

Preprocessing merupakan proses mengolah data mentah agar menjadi lebih terstruktur dan siap untuk dianalisis (Ariqoh dkk., 2023). Tujuan dari *preprocessing* adalah meningkatkan kualitas data, memastikan keakuratan hasil analisis, serta mengatasi berbagai permasalahan dalam data, seperti noise, inkonsistensi, dan data yang tidak lengkap. Tahapan *preprocessing* sebagai berikut:

- a. *Cleaning* adalah proses menghilangkan unsur-unsur yang tidak relevan dalam teks, seperti tanda baca, untuk mengurangi 'noise' atau gangguan, spasi berlebih, kata-kata yang tidak baku yang tidak diperlukan dalam analisis. *Cleansing* bertujuan untuk mengurangi kesalahan dalam analisis teks dan meningkatkan kualitas *output* akhir. Dalam teks *preprocessing*, *cleansing* sering kali dilakukan pada tahap awal untuk memastikan data teks yang akan diolah lebih lanjut dan sudah bersih dan terstruktur dengan baik. Pada proses analisis ini seperti klasifikasi teks, analisis sentimen, atau pemodelan topik untuk berjalan lebih efektif dan akurat.
- b. *Case Folding* adalah salah satu tahap penting dalam *text preprocessing*, di mana semua huruf diubah menjadi huruf kecil. Tujuan dari tahap ini adalah untuk menghindari sensitivitas terhadap perbedaan huruf besar dan kecil dalam pengenalan teks, sehingga kata yang sama tetap dianggap identik meskipun memiliki perbedaan dalam penulisan huruf besar atau kecil. Tujuan utama dari *case folding* adalah untuk menyamakan penggunaan huruf kapital dalam data teks, sehingga menghindari perbedaan makna atau interpretasi yang disebabkan oleh penggunaan huruf besar atau kecil. Dalam proses *case folding*, karakter-karakter yang bukan huruf alphabet, seperti tanda baca dan angka, seringkali

dihapus atau diabaikan karena dianggap sebagai delimiter yang tidak memiliki pengaruh signifikan terhadap analisis teks. Selain itu, spasi diawal dan akhir kalimat juga dihilangkan untuk membersihkan teks. *Case folding* sangat penting dalam teks *preprocessing* karena membantu meningkatkan konsistensi data dan memudahkan proses analisis teks berikutnya, seperti klasifikasi teks atau analisis sentimen (Dinda dkk., 2023).

2.2.5 *Term Frequency-Inverse Document Frequency (TF-IDF)*

Term Weighting-Inverse Document Frequency (TF-IDF) adalah proses mengubah teks menjadi representasi numerik yang mencerminkan pentingnya kata-kata dalam dokumen relatif terhadap dokumen. Nilai atau bobot sebuah kata akan meningkat seiring dengan bertambahnya jumlah kemunculan kata tersebut. Pada teknik pemrosesan bahasa alami ini *Natural Language Preprocessing (NLP)* dan pembelajaran mesin, seperti klasifikasi teks, penambangan teks, dan pencarian informasi. TF memiliki tugas untuk mengukur frekuensi kemunculan suatu kata dalam sebuah kalimat tertentu. IDF memiliki tugas untuk mengukur seberapa penting kata tersebut dalam sebuah dokumen (Prakoso dkk., 2022)

Dalam dokumen sering muncul kata-kata, seperti stopewords yang meliputi kata-kata umum seperti “dan”, “atau”, dan “juga”, mendapatkan bobot yang rendah dalam perhitungan TF-IDF. Hal ini disebabkan karena kata-kata tersebut tidak memberikan informasi yang signifikan untuk membedakan satu dokumen dari yang lain. Kata-kata yang spesifik dan jarang muncul di dokumen lain mendapatkan bobot yang lebih tinggi. Perhitungan TF-IDF dalam menyoroiti kata-kata kunci yang penting dan memastikan bahwa fitur-fitur yang paling informatif dan relevan diberikan perhatian yang lebih besar dalam model pembelajaran mesin. Hal ini dapat membantu dan meningkatkan akurasi dan model dalam klasifikasi teks.(Khasanah dkk., 2024)

Term Weighting adalah proses pemberian nilai pada setiap kata dalam dokumen untuk menunjukkan seberapa penting kata tersebut. Nilai ini disebut bobot, dan setiap kata dianggap sebagai fitur unik. TF (*Term Frequency*) adalah metode yang digunakan untuk menghitung frekuensi kemunculan setiap kata dalam

data, yang kemudian digunakan untuk menentukan bobot kata dalam kalimat (Harnelia dkk., 2024). Berikut adalah rumus proses TF :

$$tf = 0.5 + 0.5 \frac{tf}{\max(tf)} \dots \dots \dots (1)$$

Keterangan :

tf : Banyaknya kata yang muncul pada sebuah kalimat

$\max(tf)$: Panjang kata dari sebuah kalimat

Pada proses IDF menghitung jumlah kalimat yang berisikan term yang terdapat pada sebuah dataset. Berikut adalah rumus dari IDF:

$$idf = \ln \frac{N}{df} + 1 \dots \dots \dots (2)$$

Keterangan :

\ln : Logaritma natural

N : Jumlah semua kalimat

df : Jumlah kata pada kalimat

2.2.6 *K-Nearest Neighbor* (KNN)

K-Nearest Neighbor (KNN) merupakan sebuah teknik pembelajaran mesin yang produktif namun mudah, efisiensi guna mengklasifikasikan dan regresi. *K-Nearest Neighbor* sangat umum digunakan dalam memprediksi dan mengklasifikasikan. *K-Nearest Neighbor* bekerja dalam mengelompokkan data ke dalam *cluster* atau subset koheren dalam mengklasifikasikan data yang diinput dapat dilihat kemiripannya berdasarkan data yang sebelum di training. Tugas tersebut dimasukan ke kelas yang paling banyak tetangga terdekatnya. Hal ini menjadikan *K-Nearest Neighbor* (KNN) sebagai metode yang digunakan untuk mengklasifikasikan data berdasarkan contoh pelatihan terdekat atau tetangga dalam suatu wilayah tertentu (Putri dkk., 2024).

K-Nearest Neighbor adalah teknik klasifikasi yang dapat menentukan seberapa jauh data set histori dari data yang dimasukkan. Dengan beberapa beban untuk membandingkan atribut yang ada, teknik *K-Nearest Neighbor* dapat digunakan dalam pencarian kasus, menentukan instansi seberapa dekat dengan kasus sebelumnya. *K-Nearest Neighbor* bertujuan untuk menentukan pemisahan antara dua set data. Seberapa dekat pengukuran atau sebanding data set pengujian dan

pelatihan satu sama lain yang dapat ditentukan menggunakan nilai jarak yang dihitung. Nilai K pada *Nearest Neighbor* diartikan sebagai K-data terdekat dari data uji. (Heltroyce dkk., 2024)

K-Nearest Neighbor (K-NN) mengklasifikasikan suatu instance berdasarkan jarak dengan kategori mayoritas di sekitarnya. Kelas yang paling sering muncul akan menjadi hasil klasifikasi. Metode ini bertujuan untuk mengelompokkan objek berdasarkan atribut dan sampel pelatihan yang tersedia. *K-Nearest Neighbor* (K-NN) merupakan teknik yang sering digunakan untuk klasifikasi data. Metode ini mengelompokkan suatu objek berdasarkan data pelatihan dengan mencari tetangga terdekat yang memiliki jarak paling kecil atau nilai selisih yang minimal dengan objek tersebut.

K-Nearest Neighbor (KNN) adalah algoritma klasifikasi yang menentukan kelas suatu objek berdasarkan kedekatannya dengan data training terdekat, dengan menggunakan metode perhitungan jarak seperti *Euclidean Distance*. Langkah awal penerapan *K-Nearest Neighbor* yaitu menentukan nilai K, mewakili jumlah tetangga yang akan dipertimbangkan dalam proses klasifikasi. Setelah nilai K ditentukan, langkah berikutnya adalah mengukur jarak antara objek baru (data *testing*) dengan tetangga terdekatnya dalam data pembelajaran (data *training*). Jarak dapat dihitung menggunakan berbagai metrik, seperti *euclidean distance*.

Berikut adalah rumus KNN dengan menggunakan *Euclidean Distance*:

$$\text{Distance} = \sqrt{\sum_{j=1}^n (x_{\text{train}}^j - x_{\text{test}})^2} \dots \dots \dots (3)$$

Keterangan:

x_{train}^j : data *training* ke-j

x_{test} : data *testing*

j : baris ke-j dari tabel

n : banyaknya data *training*

Pemilihan nilai K penting dapat mempengaruhi untuk menekankan hasil klasifikasi. Nilai K terkecil dapat menghasilkan model menjadi sensitif terhadap noise, sedangkan nilai K yang besar dapat membuat hasil model yang terlalu umum.

Pemilihan K yang tepat merupakan salah satu aspek kritis dalam penerapan *K-Nearest Neighbor*. Setelah jarak diukur, objek baru akan diklasifikasikan berdasarkan mayoritas kelas dari tetangga terdekat. Melalui implementasi *K-Nearest Neighbor*, maka hasil klasifikasi dapat dicapai dengan cukup baik tanpa harus membuat asumsi yang kompleks mengenai distribusi data. Meskipun sederhana, *K-Nearest Neighbor* menjadi pilihan yang efektif terutama dalam kasus-kasus dimana hubungan antara fitur dan kelas tidak dapat dijelaskan dengan fungsi matematis yang didefinisikan dengan baik (Harnelia dkk., 2024).

2.2.7 Confusion Matrix

Confusion matrix adalah alat visual untuk mengevaluasi kinerja model klasifikasi, yang menampilkan perbandingan antara hasil prediksi dan hasil aktual dalam bentuk tabel. Evaluasi merupakan langkah-langkah yang digunakan untuk mengukur kinerja dan efektivitas model supaya berjalan sesuai dengan diagram atau alur yang telah dibuat. Proses ini merupakan langkah krusial pembangunan dalam model pembelajaran mesin, termasuk *K-Nearest Neighbor* (KNN). Salah satu cara menghitung metrik untuk menghitung kinerja dan efektivitas sistem adalah *confusion matrix*. Untuk memastikan kinerja dari model yang dikembangkan, perlu dilakukan pengujian model dengan menggunakan data uji. Pada proses ini menggunakan *confusion matrix*. Tabel *confusion matrix* menyajikan empat kemungkinan hasil yang berbeda, yang membandingkan antara prediksi model dengan nilai aktual (Arrasyid dkk., 2024).

Tabel 2. 1 *Confusion Matrix*

	Positif	Negatif
Positif	<i>True Positif (TP)</i>	<i>False Positif (FP)</i>
Negatif	<i>False Negatif (FN)</i>	<i>True Negatif (TN)</i>

Confusion matrix menggunakan empat istilah untuk menggambarkan hasil klasifikasi: *True Positif (TP)*, *True Negatif (TN)*, *False Positif (FP)*, dan *False Negatif (FN)*. Dengan menggunakan tabel *confusion matrix*, dapat menghitung nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

- a. *Accuracy* adalah metrik evaluasi yang digunakan untuk mengukur seberapa dekat hasil prediksi model dengan nilai aktual, dengan menghitung persentase

jumlah data yang diklasifikasikan dengan benar terhadap total data. Salah satu metrik evaluasi yang paling umum dan sederhana yang digunakan untuk mengukur kinerja model klasifikasi dalam pembelajaran mesin. Akurasi menunjukkan seberapa sering prediksi yang benar dari model. Akurasi dihitung dan diinterpretasikan, sehingga sering digunakan sebagai metrik dasar dalam mengevaluasi kinerja model. Prediksi digunakan untuk memberikan gambaran umum mengenai kinerja model tersebut. Berikut adalah rumus untuk menghitung *accuracy* :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \dots \dots \dots (4)$$

- b. *Precision* adalah metrik evaluasi yang sangat penting dalam pembelajaran mesin, khususnya dalam klasifikasi. *Precision* adalah metode yang digunakan untuk menguji seberapa relevan informasi yang diperoleh sistem dengan membandingkan jumlah informasi relevan terhadap seluruh informasi yang diambil, baik yang relevan maupun tidak. *Precision* mengukur tingkat keakuratan model dalam memprediksi kelas positif dengan membandingkan jumlah prediksi positif yang benar terhadap total prediksi positif. Dalam konteks ini, *precision* sangat berguna dalam memberikan penekanan kualitas prediksi positif, memastikan bahwa apa yang diprediksi positif benar-benar positif. Berikut adalah rumus untuk menghitung *precision*:

$$Precision = \frac{TP}{TP+FN} \dots \dots \dots (5)$$

- c. *Recall* atau dikenal sebagai *True Positive Rate* (TPR) atau *Sensitivity* merupakan metode mengukur seberapa banyak informasi relevan yang berhasil diambil oleh sistem dibandingkan dengan seluruh informasi relevan yang ada dalam koleksi data. *Recall* adalah metrik evaluasi dalam *machine learning* yang dapat mengukur seberapa baik model dalam mengenali semua *instance* positif yang ada dalam dataset. *Recall* sangat penting dapat menangkap semua *instance* positif lebih diutamakan daripada menghindari kesalahan positif palsu. *Recall* dapat menekankan kemampuan model dalam mendeteksi semua *instance* positif. Berikut adalah rumus untuk menghitung *recall* :

$$Recall = \frac{TP}{TP+FN} \dots \dots \dots (6)$$

d. *F1-Score* adalah metrik evaluasi yang sering digunakan dalam analisis klasifikasi untuk menilai kinerja model, terutama ketika data tidak seimbang. *F1-Score* adalah rata-rata harmonis dari *precision* dan *recall* yang digunakan untuk memberikan gambaran seberapa baik suatu model dalam melakukan klasifikasi, terutama ketika terdapat ketidakseimbangan antara kelas positif dan negatif. *F1-Score* memiliki nilai antara 0 hingga 1. Nilai 1 menunjukkan *precision* dan *recall* memiliki nilai yang bagus, sedangkan nilai 0 menunjukkan model tidak dapat mendeteksi kelas positif dengan baik. *F1-Score* merupakan keseimbangan antara *precision* dan *recall*, sehingga ketika distribusi kelas tidak seimbang dan memastikan model tidak hanya mengutamakan salah satu aspek. *F1-Score* digunakan ketika jumlah kelas positif sangat kecil dibandingkan kelas negatif, *F1-Score* memiliki kinerja model yang baik dibandingkan hanya mengandalkan *accuracy*. Berikut adalah rumus untuk menghitung *F1-Score*:

$$F1\ score = \frac{2 \times (precision \times recall)}{precision + recall} \dots\dots\dots (7)$$

Keterangan :

TP : *True Positif* adalah jumlah model yang benar-benar memprediksi kelas positif dan kelas sebenarnya juga positif. Model ini berhasil mengidentifikasi contoh kelas yang benar-benar positif.

TN : *True Negatif* adalah jumlah model yang benar-benar memprediksi kelas negatif dan kelas sebenarnya juga negatif. Model ini berhasil mengidentifikasi contoh kelas yang benar-benar negatif.

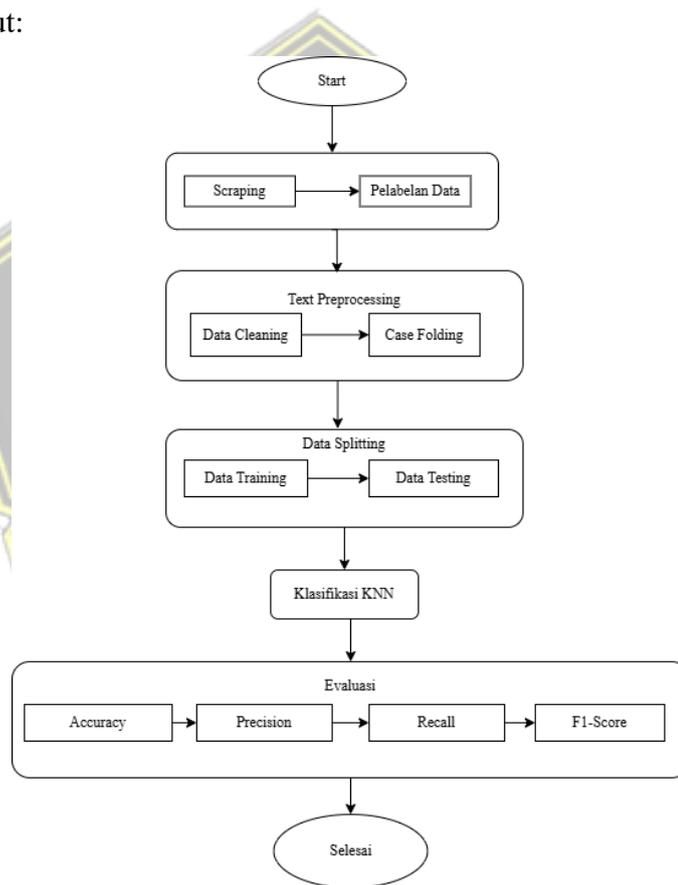
FP : *False Positif* adalah jumlah model yang memprediksi kelas positif padahal kelas sebenarnya adalah negatif. Model ini salah memprediksi contoh kelas positif ketika sebenarnya negatif.

FN : *False Negatif* adalah jumlah model yang memprediksi kelas negatif padahal kelas sebenarnya adalah positif. Model ini salah memprediksi contoh kelas negatif ketika sebenarnya positif.

BAB III METODE PENELITIAN

3.1 Metode Penelitian

Penelitian ini menggunakan metode atau algoritma yang digunakan adalah *K-Nearest Neighbor*. Dalam penelitian ini, metode tersebut digunakan untuk mengkalsifikasikan analisis sentimen berdasarkan *abstract* dan judul yang dimasukkan. Langkah-langkah yang harus dilakukan dalam penelitian ini adalah sebagai berikut:



Gambar 3. 1 Kerangka Penelitian

3.1.1 Studi Literature

Penelitian ini akan meninjau berbagai sumber, yaitu *e-book*, makalah, jurnal, tesis, skripsi, dan situs web, untuk memperdalam pemahaman tentang teori analisis sentimen dan algoritma *K-Nearest Neighbor*.

3.1.2 Pengumpulan Data

Pada tahap ini sebagai persiapan awal dan pengembangan model adalah dengan mengumpulkan data. Data yang digunakan adalah menggunakan data set ulasan yang ada di Shopee. Dengan cara *scraping* menggunakan *Google Collab*. Data yang sudah diambil disimpan dalam bentuk *excel*.

3.1.3 Labelling Data

Proses *labelling* dilakukan secara manual dengan bantuan *Microsoft Excel*, di mana setiap ulasan dikategorikan dan diberi label tambahan yang menunjukkan apakah ulasan tersebut mengandung sentimen positif atau negatif. Pada tabel 3.1 merupakan contoh *labelling* data.

Tabel 3. 1 Contoh Hasil *Labelling*

No.	Ulasan	Sentimen
1.	Sunscreen pas dipake gak lengket dan mudah menyerap keliatan cerah dan lembab, gak bikin white cast. Cocok di aku yang kulit berminyak, gak bikin kusam malah bikin glowing apalagi kalo kena panas jadi suka banget sama sunscreen ini.	Positif
2.	Buat toner yg bisa sekalian exfoliasi kulit, pemakaian malem hari ini sih udah enak banget	Positif
3.	Sayang banget ga cocok buat kulitku, malah jadi beruntusan dan muncul jerawat padahal sebelumnya udah sembuh 🙄 Pas setelah dipake juga jd kaya lengket gitu berat di muka jadinya	Negatif
4.	Sunscreen di saya ga ada pengaruh apa-apa yang ada buat muka saya jadi item berminyak muncul jerawat gradakan dan gatal, baunya aneh, jangan dibeli barang palsu!	Negatif

3.1.4 Text Preprocessing

Berikut penjelasan setiap tahap *preprocessing* adalah sebagai berikut :

- a. Langkah pertama dalam *preprocessing* adalah melakukan data *cleaning* untuk menghilangkan karakter selain huruf dan angka, serta menghilangkan spasi berlebih.
- b. Kemudian dilakukan Case Folding adalah proses normalisasi teks dengan mengubah semua huruf dalam suatu dokumen menjadi huruf kecil untuk memastikan konsistensi dalam pemrosesan teks untuk menghindari sensitivitas terhadap perbedaan huruf besar dan kecil dalam pengenalan teks. Tahap ini memastikan bahwa kata-kata yang sama, meskipun ditulis dengan huruf besar atau kecil, tetap dianggap identik.

Tabel 3. 2 Contoh Hasil Sebelum dan Sesudah Preprocessing

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
Sunscreen pas dipake gak lengket dan mudah menyerap keliatan cerah dan lembab, gak bikin white cast. Cocok di aku yang kulit berminyak, gak bikin kusam malah bikin glowing apalagi kalo kena panas jadi suka banget sama sunscreen ini.	sunscreen pas dipake gak lengket dan mudah menyerap keliatan cerah dan lembab gak bikin white cast cocok di aku yang kulit berminyak gak bikin kusam malah bikin glowing apalagi kalo kena panas jadi suka banget sama sunscreen ini

3.1.5 Splitting Data

Pembagian data dengan rasio 80:20 merupakan teknik yang umum digunakan dalam machine learning, di mana 80% data digunakan untuk pelatihan (training set) dan 20% sisanya digunakan untuk pengujian (testing set). Tujuan dari metode ini adalah memastikan model memiliki cukup data untuk belajar pola yang terdapat dalam dataset sekaligus memiliki data uji yang cukup untuk mengevaluasi kinerjanya. Dengan lebih banyak data latih, model dapat menangkap pola dengan lebih baik, namun tetap harus diuji dengan data yang belum pernah dilihat sebelumnya untuk menghindari overfitting. Jika model menunjukkan performa yang tinggi pada data latih tetapi buruk pada data uji, maka ada kemungkinan model

terlalu menyesuaikan diri dengan data pelatihan dan kurang mampu melakukan generalisasi.

3.1.6 Term Frequency-Inverse Document Frequency (TF-IDF)

Pada proses TF-IDF, merupakan teknik penting dalam pemrosesan teks dan analisis informasi. TF-IDF digunakan untuk mengevaluasi seberapa penting sebuah kata dalam suatu dokumen dengan membandingkan frekuensi kemunculannya dalam dokumen tersebut terhadap keseluruhan kumpulan dokumen. Term Frequency (TF) mengukur seberapa sering suatu kata muncul dalam sebuah dokumen, dengan memberikan bobot lebih tinggi pada kata-kata yang sering muncul. Sementara itu, Inverse Document Frequency (IDF) mengukur tingkat kepentingan sebuah kata di seluruh dokumen dalam korpus, dengan memberikan bobot lebih tinggi pada kata yang jarang muncul di banyak dokumen. Kata-kata umum, misal “dan”, “ini”, “yang”, akan mendapatkan bobot yang rendah, karena muncul di banyak dokumen dan tidak memberikan banyak informasi spesifik.

3.1.7 K-Nearest Neighbor

Pada tahap ini melakukan klasifikasi pada data set yang telah melalui tahap *preprocessing*. Pada algoritma *K-Nearest Neighbor* terdapat perhitungan jarak antara dua titik, yaitu titik berupa data *training* dan titik berupa data *testing* yang dapat didefinisikan dengan rumus *Euclidean Distance* sebagai persamaan. Dalam klasifikasi, data akan dikategorikan ke dalam teks yang paling banyak muncul di antara k tetangga terdekat. Sedangkan dalam regresi, nilai prediksi akan dihitung berdasarkan rata-rata atau metode lain dari nilai-nilai tetangga tersebut.

3.1.8 Evaluasi (Confusion Matrix)

Pada tahap ini, Confusion Matrix adalah metode evaluasi dalam pembelajaran mesin yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan prediksi model terhadap nilai sebenarnya. Confusion Matrix ini berbentuk tabel yang membandingkan prediksi model dengan hasil aktual, sehingga menunjukkan jumlah prediksi yang benar dan salah.

Table 3. 1 *Confusion Matrix*

	Positif	Negatif
Positif	<i>True Positif (TP)</i>	<i>False Positif (FP)</i>
Negatif	<i>False Negatif (FN)</i>	<i>True Negatif (TN)</i>

Confusion matrix menggunakan empat istilah untuk menggambarkan hasil klasifikasi: *True Positif (TP)*, yaitu jumlah data positif yang benar diprediksi; *True Negatif (TN)*, yaitu jumlah data negatif yang benar diprediksi; *False Positif (FP)*, yaitu jumlah data negatif yang salah diprediksi sebagai positif; dan *False Negatif (FN)*, yaitu jumlah data positif yang salah diprediksi sebagai negatif. Dari tabel *confusion matrix* ini, kita dapat menghitung metrik seperti *akurasi*, *presisi*, *recall*, dan *f1-score*

3.1.9 Visualisasi

Visualisasi adalah proses penyajian data dalam bentuk grafis atau gambar untuk memudahkan pemahaman, analisis, dan interpretasi informasi. Dengan menggunakan berbagai teknik visual seperti diagram batang, grafik garis, dan diagram lingkaran, visualisasi memungkinkan data yang kompleks dapat disederhanakan dan lebih mudah dipahami. Dalam konteks analisis data, visualisasi membantu dalam mengidentifikasi pola, tren, dan hubungan yang mungkin sulit dikenali hanya dengan melihat angka atau tabel.

3.2 Analisis Kebutuhan

Penelitian ini menganalisis *software* apa saja yang akan dibutuhkan sehingga mulai dari proses *input* hingga hasil akhir dapat berjalan sesuai dengan yang diharapkan. Berikut adalah daftar perangkat lunak yang digunakan:

1. *Python*

Python merupakan bahasa pemrograman tingkat tinggi yang fleksibel dengan fokus pada keterbacaan dan kesederhanaan. Dikembangkan oleh Guido van Rossum dan pertama kali dirilis pada tahun 1991, *Python* telah berkembang menjadi salah satu bahasa pemrograman yang paling populer dan banyak digunakan di seluruh dunia. Dalam penelitian ini, *Python* dipilih karena memiliki library yang lengkap serta bersifat open source.

2. *Google Colaboratory*

Google Colaboratory adalah sebuah aplikasi dari platform *Google* yang digunakan untuk mengembangkan dan menjalankan kode dalam lingkungan komputasi interaktif dalam penelitian ini, *google colab* digunakan untuk menulis dan menjalankan kode program, serta membuat penyimpanan catatan atau dokumentasi.

3. *Library Pandas*

Pandas adalah *library Python* yang digunakan untuk memproses data, mulai dari pembersihan, manipulasi, hingga analisis data, dengan menyediakan struktur data yang efisien seperti *DataFrame* dan *Series*. Dalam penelitian ini *library Pandas* digunakan untuk membuat laporan data pada saat testing dalam *google colab* dan juga digunakan dalam pengolahan data.

4. *Library Numpy*

NumPy (*Numerical Python*) adalah *library Python* yang digunakan untuk bekerja dengan array multidimensi serta menyediakan berbagai fungsi untuk komputasi numerik, termasuk operasi aljabar linear, transformasi Fourier, dan manipulasi matriks.

5. *Library Scikit-Learn*

Library scikit-learn adalah *library* yang digunakan untuk pembelajaran mesin (*machine learning*) yang ditulis dalam bahasa pemrograman *Python*.

6. *Library NLTK (Natural Language Tool Kit)*

Natural Language Tool Kit adalah *library* atau *toolkit* populer dalam bahasa pemrograman *Python* yang digunakan untuk pemrosesan *Natural Language Processing* (NLP) dalam penelitian ini NLTK digunakan dalam proses *preprocessing* atau pengolahan teks diantaranya *case folding*, *stopwords*, *tokenizatin*, *stemming* dan *normalized text*.

7. *Library Collections*

Library Collections adalah *library* dalam bahasa pemrograman *python* yang menyediakan alternatif struktur data yang lebih efisien dan fungsional dari struktur data *built-in* seperti *list*, *dict*, dan *set*. *Library* ini dirancang untuk memenuhi

kebutuhan pemrograman yang lebih spesifik dan kompleks dengan menyediakan berbagai tipe koleksi yang dapat digunakan untuk menyimpan dan mengelola data.

8. *Visual Studio Code*

Visual Studio Code dipilih sebagai text editor pada pengembangan aplikasi dalam penelitian ini, *Visual Studio Code* dipilih dikarenakan mendukung banyak Bahasa Pemrograman dan framework, multi-platform, performa yang sangat cepat, mempunyai banyak extensions yang dapat mempermudah proses pengembangan.

9. *MiniConda*

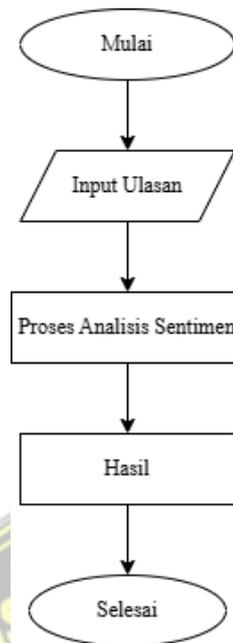
MiniConda adalah sebuah *package manager* dan *environment manager* yang digunakan untuk mengelola paket serta lingkungan dalam bahasa pemrograman *Python*.

10. *Framework Streamlit*

Streamlit adalah framework open-source berbasis *Python* yang dirancang untuk mempermudah pembuatan aplikasi web interaktif, khususnya dalam bidang sains data dan pembelajaran mesin. Pada penelitian ini, *framework streamlit* digunakan sebagai pengembangan aplikasi *website* dikarenakan core dari sistem yang akan dikembangkan sendiri berbahasa *Python*.

3.3 Analisis Sistem

Pada tahapan analisis sistem, dilakukan analisa guna membuat alur kerja sistem yang akan dibuat dengan memakai *flowchart*.



Gambar 3. 2 *Flowchart* Analisa Sistem

Langkah pertama adalah pengguna atau user membuka aplikasi. Kemudian terlihat halaman sistem prediksi sentimen ulasan dengan sebuah *text area*. User bisa menginput *query* pada *text area* dengan mengetikkan ulasan/*review* produk dari Shopee. Setelah itu, user bisa mengklik tombol “Prediksi” guna memproses *query* tersebut. Selanjutnya, *query* akan diproses dengan mencocokkan ulasan/*review* produk yang diinputkan dengan masing-masing topik yang telah dilatih menggunakan model *K-Nearest Neighbor* yang ada dalam sistem. Kemudian, menghasilkan apakah ulasan/*review* produk tersebut memiliki sentiment negatif atau positif.

BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pada bagian ini, peneliti membahas mengenai yang didapatkan setelah melalui tahap metode penelitian sebagai berikut:

4.1.1 *Scraping* Data

Pada tahap *Scraping* data menggunakan *Google Collab* yaitu dengan memasukkan URL produk kecantikan di Shopee. Kemudian mengekstrak data ulasan yang berisi data ulasan (teks ulasan, peringkat, dll.). Selanjutnya, data ulasan yang diekstrak disimpan ke dalam bentuk *excel*. Scraping yang telah dilakukan memperoleh data keseluruhan sebanyak 2.000 data ulasan. Pada tabel 4.1 merupakan hasil dari *scraping*.



Tabel 4. 1 Hasil dari *Scraping*

No.	username,rating,comment,product_variation,timestamp
1.	a*****s,5,"Kemasan:unik Po nya lama Barang datang langsung aplikasikan ke wajah dan emang bagus bngt. Gk berminyak,gak lengket dan cepet meresap. Emang yah benerÂ² skintific kalo bikin formula juara" ,,1725449689
2.	a*****y,5,"Tekstur:cair tapi ga cair Daya Serap:langsung terserap dengan baik Kondisi Barang:aman, rapi Pengiriman cepet, barangnya aman, packgingnya rapi, sunscreenya langsung meresap sempurna sih, the best ini. Harga juga oke" ,,1730356489
...
2000	376vgp4z_r,5,"Tekstur:kental dan mudah digunakan Daya Serap:sangat bgus dan efisien Kondisi Barang:diterima dalam keadaan baik. Barangnya sudah diterima dengan baik.. Pemakaiannya mudah, dan ringan diwajah.. Terimakasih bg kurir ðŸ™ ðŸ™ Terimakasih seller ðŸ™ ðŸ™" ,,1730965653

4.1.2 Labelling Data

Hasil dari proses scraping kemudian diberi label secara manual dengan mengidentifikasi setiap ulasan sebagai ulasan positif atau negatif. Proses ini bertujuan untuk memastikan data memiliki kategori yang sesuai sebelum digunakan dalam analisis lebih lanjut. Pada tabel 4.2 hasil dari *labelling* data.

Tabel 4. 2 Hasil *Labelling*

No.	Ulasan	Sentimen
1.	Sunscreen pas dipake gak lengket dan mudah menyerap keliatan cerah dan lembab, gak bikin white cast. Cocok di aku yang kulit berminyak, gak bikin kusam malah bikin glowing apalagi kalo kena panas jadi suka banget sama sunscreen ini.	Positif

2.	Buat toner yg bisa sekalian exfoliasi kulit, pemakaian malem hari ini sih udah enak banget	Positif
3.	Ini exfo toner paling calm dan gaada efek aneh2 waktu aku pake. Harga terjangkau dan packing aman. Selalu suka order disini, nggak pernah ngecewain ♥	Positif
4.	Tekstur nya agak cair, ga white case, ringan, cuma agak lama buat nyerep, but overall suka karna harga yg terjangkau	Positif
5.	Moist nya ini di aku kurang cocok karna bikin beruntusan dan lengket dimuka aku. Tapi kayaknya untuk makeup ini bagus tahan lama. Tapi maaf di aku bikin beruntusan skrg aku udah berentiin takut beruntusannya tambah banyak.	Negatif
6.	Sunscreen di saya ga ada pengaruh apa-apa yang ada buat muka saya jadi item berminyak muncul jerawat gradakan dan gatal, baunya aneh, jangan dibeli barang palsu!	Negatif
...	...	
2000.	Belum terlalu utk mengangkat sel kulit mati, cuma utk menghaluskan tekstur kulit lumayan ada perubahan. Nnti klo bnr2 ad perubahan sya riview lgi	Negatif

4.1.3 Text Preprocessing

Text preprocessing pada data ulasan dilakukan melalui beberapa tahap, salah satunya adalah mengurangi noise dengan menghapus tanda baca, angka, simbol, atau emotikon. Proses ini dilakukan menggunakan library RegEx (Regular Expression) pada data ulasan yang diperoleh dari marketplace Shopee mengubah semua teks ulasan menjadi huruf kecil agar teks ulasan menjadi yang seragam dikarenakan hasil data *scraping* terdapat huruf kapital dan huruf kecil.

Tabel 4. 3 Hasil *Text Preprocessing*

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
Sunscreen pas dipake gak lengket dan mudah menyerap keliatan cerah dan lembab, gak bikin white cast. Cocok di	sunscreen pas dipake gak lengket dan mudah menyerap keliatan cerah dan lembab gak bikin white cast cocok di

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
aku yang kulit berminyak, gak bikin kusam malah bikin glowing apalagi kalo kena panas jadi suka banget sama sunscreen ini.	aku yang kulit berminyak gak bikin kusam malah bikin glowing apalagi kalo kena panas jadi suka banget sama sunscreen ini
Packing bagus & aman, produk ori, pengiriman bagus . Baru pertamakali coba semoga cocok untuk kulit kusam . Kalo cocok nnti bisa jadi langganan . Makasih .	packing bagus aman produk ori pengiriman bagus baru pertamakali coba semoga cocok untuk kulit kusam kalo cocok nnti bisa jadi langganan makasih
Aku pesen sunscreen dikirimnya malah cleanser 😞	aku pesen sunscreen dikirimnya malah cleanser
Produk original, bagus di kulit wajah saya, bakal beli lagi 👍👍👍👍	produk original bagus di kulit wajah saya bakal beli lagi
Buat toner yg bisa sekaligus exfoliasi kulit, pemakaian malem hari ini sih udah enak banget	buat toner yg bisa sekaligus exfoliasi kulit pemakaian malem hari ini sih udah enak banget
Ini exfo toner paling calm dan gaada efek aneh2 waktu aku pake. Harga terjangkau dan packing aman. Selalu suka order disini, nggak pernah ngecewain ❤️	ini exfo toner paling calm dan gaada efek aneh waktu aku pake harga terjangkau dan packing aman selalu suka order disini nggak pernah ngecewain
Tekstur nya agak cair, ga white case, ringan, cuma agak lama buat nyerep, but overall suka karna harga yg terjangkau	tekstur nya agak cair, ga white case ringan cuma agak lama buat nyerep but overall suka karna harga yg terjangkau
Moist nya ini di aku kurang cocok karna bikin beruntusan dan lengket dimuka	moist nya ini di aku kurang cocok karna bikin beruntusan dan lengket dimuka

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
aku. Tapi kayaknya untuk makeup ini bagus tahan lama. Tapi maaf di aku bikin beruntusan skrg aku udah berentiin takut beruntusannya tambah banyak.	aku tapi kayaknya untuk makeup ini bagus tahan lama tapi maaf di aku bikin beruntusan skrg aku udah berentiin takut beruntusannya tambah banyak
Sunscreen di saya ga ada pengaruh apa-apa yang ada buat muka saya jadi item berminyak muncul jerawat gradakan dan gatal, baunya aneh, jangan dibeli barang palsu!	sunscreen di saya ga ada pengaruh apa-apa yang ada buat muka saya jadi item berminyak muncul jerawat gradakan dan gatal baunya aneh jangan dibeli barang palsu
Sayang banget ga cocok buat kulitku, malah jadi beruntusan dan muncul jerawat padahal sebelumnya udah sembuh 🙄 Pas setelah dipake juga jd kaya lengket gitu berat di muka jadinya	sayang banget ga cocok buat kulitku malah jadi beruntusan dan muncul jerawat padahal sebelumnya udah sembuh pas setelah dipake juga jd kaya lengket gitu berat di muka jadinya
Belum terlalu utk mengangkat sel kulit mati, cuma utk menghaluskan tekstur kulit lumayan ada perubahan. Nnti klo bnr2 ad perubahan sya riview lgi	belum terlalu utk mengangkat sel kulit mati cuma utk menghaluskan tekstur kulit lumayan ada perubahan nnti klo bnr ad perubahan sya riview lgi

4.1.4 *Splitting Data*

Pada tahap *splitting data*, data dibagi menjadi dua bagian: data training dan data testing. Data training digunakan oleh model machine learning untuk mempelajari pola dalam klasifikasi K-Nearest Neighbor (KNN), sedangkan data testing digunakan untuk menguji model dengan melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya. Prosentase pembagian data 80:20 dari total jumlah 2000 data ulasan sehingga mendapatkan 1600 data *training* dan 400 data *testing*.

4.1.5 *Term Frequency-Inverse Document Frequency (TF-IDF)*

Pada tahap ini TF-IDF, merupakan teknik yang sangat penting dalam pemrosesan teks dan analisis informasi. Teknik ini mengevaluasi seberapa penting

sebuah kata dalam dokumen relatif terhadap kumpulan dokumen yang lebih besar, dengan mempertimbangkan frekuensi keunculan kata dalam dokumen serta seberapa jarang kata tersebut muncul.

(0, 2312)	0.2054783669758333
(0, 197)	0.3366604595627254
(0, 91)	0.16477268697391273
(0, 2505)	0.15423205264931328
(0, 2271)	0.21604125936022375
(0, 1570)	0.12773329358828012
(0, 2421)	0.40965696259577167
(0, 633)	0.20160357574146778
(0, 2013)	0.17294923468478343
(0, 642)	0.2751659842487689
(0, 1658)	0.1673574090451821
(0, 1677)	0.29583813290938366
(0, 1444)	0.23892871806618962
(0, 2165)	0.34476236036924224
(0, 1729)	0.2557859829781398
(0, 1858)	0.24396191505980058
(1, 2505)	0.15598097040408
(1, 401)	0.20630431169194766
(1, 473)	0.3053632821954032
(1, 2350)	0.2643556300556648
(1, 740)	0.39306053451286804
(1, 2054)	0.1980875569511021
(1, 2057)	0.3416762803760328
(1, 1364)	0.2352524207692086
(1, 1378)	0.2154619786209967

Gambar 4. 1 Hasil Pembobotan TF-IDF

4.1.6 K-Nearest Neighbor

Pada proses pelatihan model K-Nearest Neighbor (KNN), yang merupakan metode pembelajaran mesin yang populer untuk klasifikasi, kode program berikut ini digunakan untuk mengimplementasikan berbagai langkah kritis dalam pelatihan model. Proses ini melibatkan pengolahan input, pelatihan model menggunakan dataset yang telah disiapkan, serta evaluasi kinerja model berdasarkan metrik yang relevan.

4.1.7 Evaluasi Model (*Confusion Matrix*)

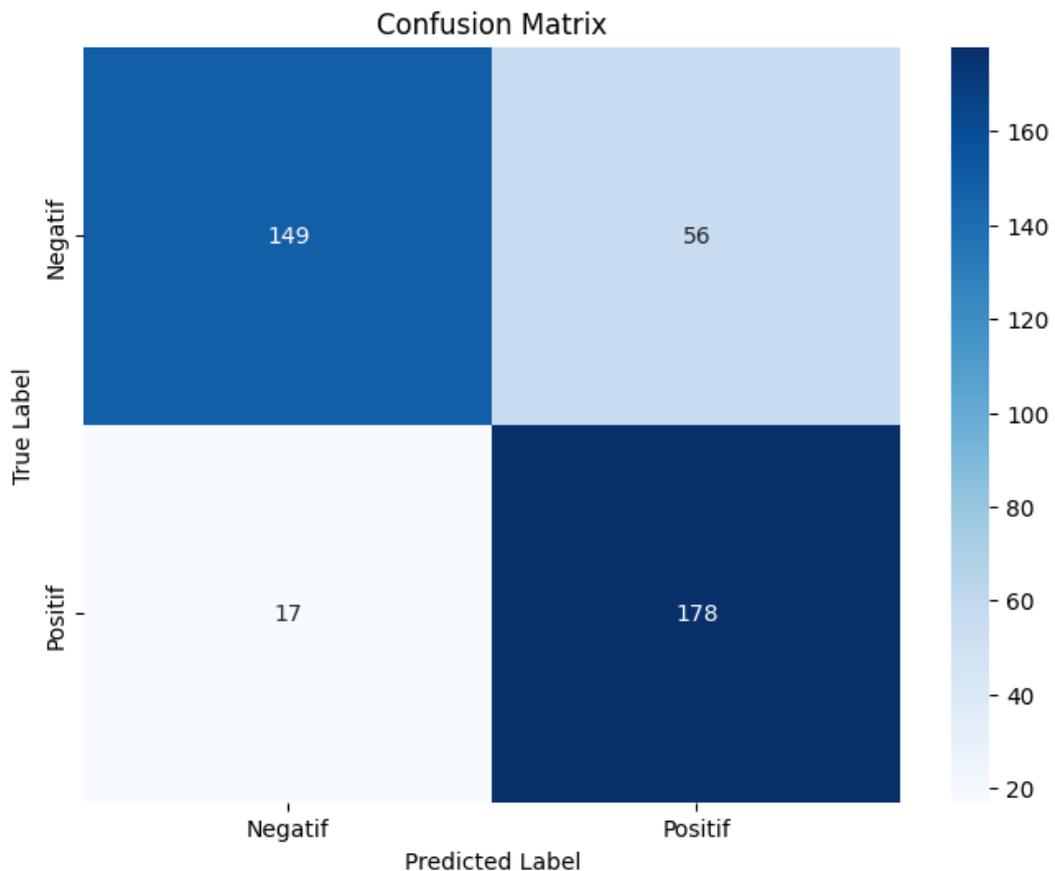
Tabel 4. 4 *Classification Report*

	Precision	Recall	F1-Score	Support
Negatif	0,90	0,73	0,80	205
Positif	0,76	0,91	0,83	195
Accuracy			0,82	400
Macro Avg	0,83	0,82	0,82	400
Weighted Avg	0,83	0,82	0,82	400

Pada gambar 3 dari hasil *Classification Report* mendapatkan hasil :

- Pada *accuracy* menunjukkan nilai proporsi dari semua prediksi yang benar dengan data asli didapatkan nilai 0.82, model memprediksi dengan benar 82% dari ulasan.

- b. Macro average recall menunjukkan persentase kinerja model dalam mengidentifikasi setiap ulasan sesuai dengan kelasnya, dengan nilai sebesar 82%.
- c. Macro average F1-Score menunjukkan persentase kinerja model secara keseluruhan, dengan nilai sebesar 82%.



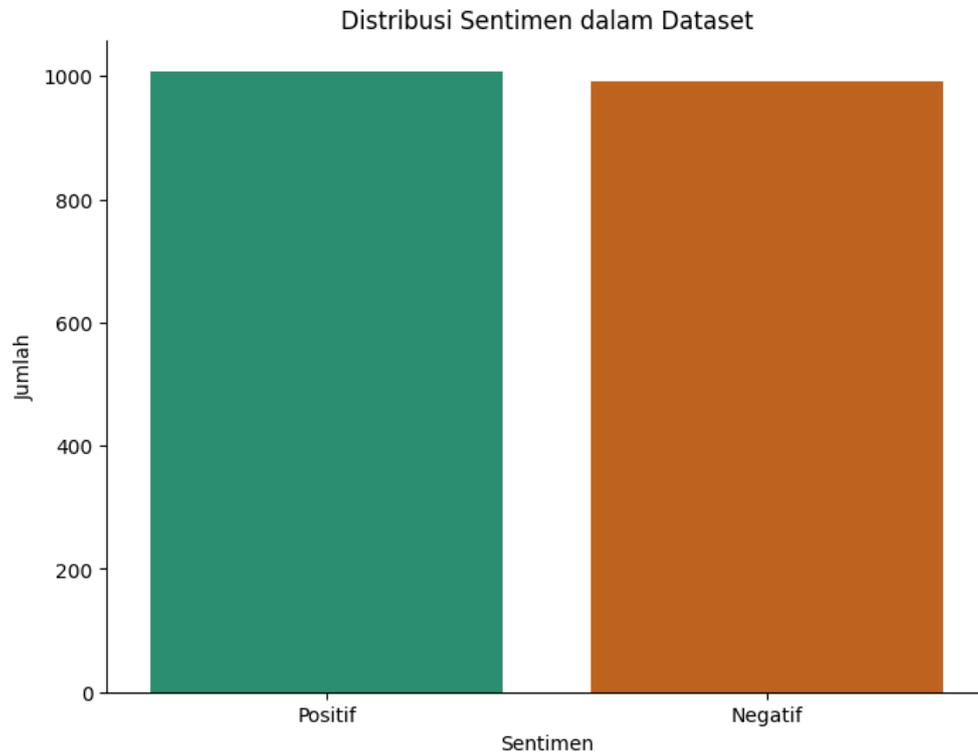
Gambar 4. 2 *Confusion Matrix*

Pada gambar 4 merupakan hasil dari *Confusion Matrix* yang didapatkan:

- a. Nilai True Negatif menunjukkan jumlah ulasan yang berhasil diprediksi sebagai negatif dengan benar oleh model, yaitu sebanyak 149 ulasan.
- b. Nilai *False* Positif atau jumlah ulasan positif yang diprediksi secara salah sebagai kelas negatif sebanyak 56 ulasan.
- c. Nilai *False* Negatif atau jumlah ulasan negatif yang diprediksi secara salah sebagai kelas positif sebanyak 17 ulasan.
- d. Nilai True Positif menunjukkan jumlah ulasan yang secara akurat diprediksi sebagai positif oleh model, yaitu sebanyak 178 ulasan.

4.1.8 Visualisasi

Pada tahap ini, visualisasi dilakukan untuk menampilkan distribusi sentimen dalam dataset.



Gambar 4. 3 Hasil Distribusi Sentimen Berdasarkan Label

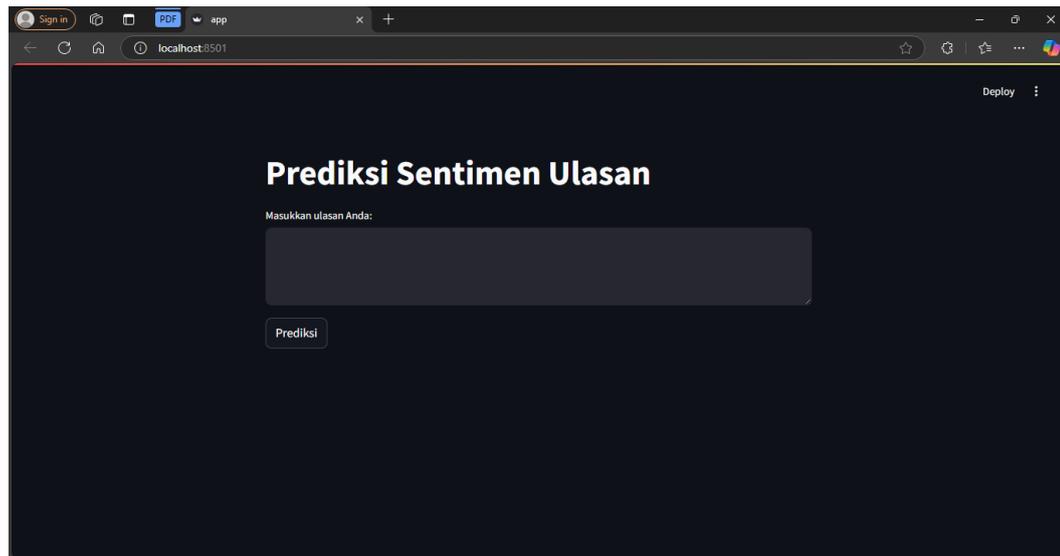
Pada gambar 4.5 menampilkan hasil distribusi ulasan/review pada masing-masing sentimen, dimana jumlah sentimen positif sebanyak 1008 ulasan dan sentimen negatif sebanyak 992 ulasan.

4.2 Hasil Implementasi Antarmuka

Setelah dilakukan tahap peodelan selesai, tahap berikutnya yaitu tahap implementasi sistem pada *platform website*, dalam hal ini memakai *streamlit*.

4.2.1 Halaman Awal Sistem Prediksi Sentimen Ulasan

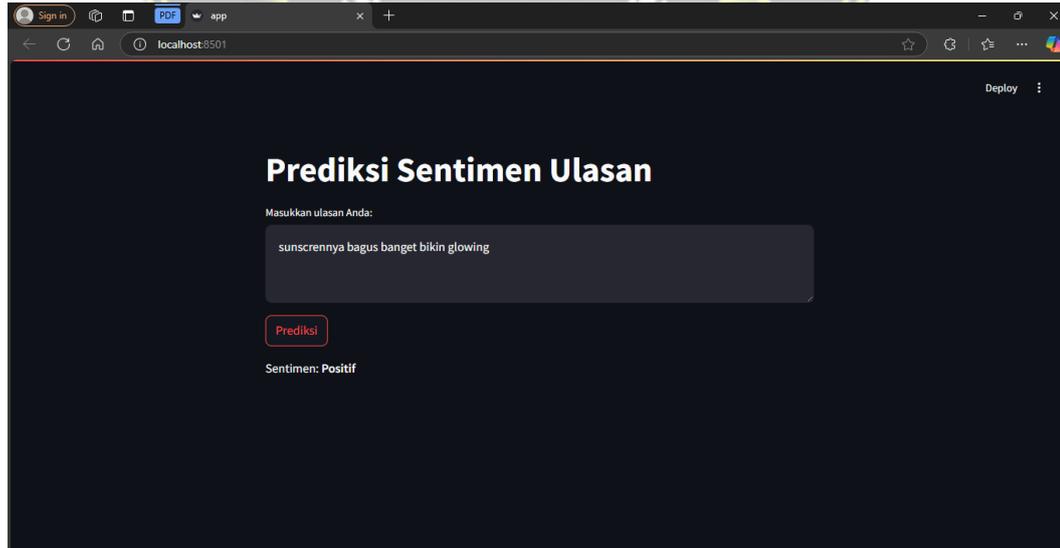
Dalam halaman depan sistem Prediksi sentimen ulasan. Pada bagian bawah *tittle* terdapat *input text area* untuk menginputkan teks berupa ulasan/*review* produk dan button bertuliskan “Prediksi” guna memproses input.



Gambar 4. 4 Halaman Awal Sistem

Pada gambar 4.6 tersebut, adalah tampilan awal yang akan dilihat oleh pengguna yang mengakses. Pada tampilan tersebut, pengguna bisa langsung memberikan *input* teks dalam *text area* yang ada.

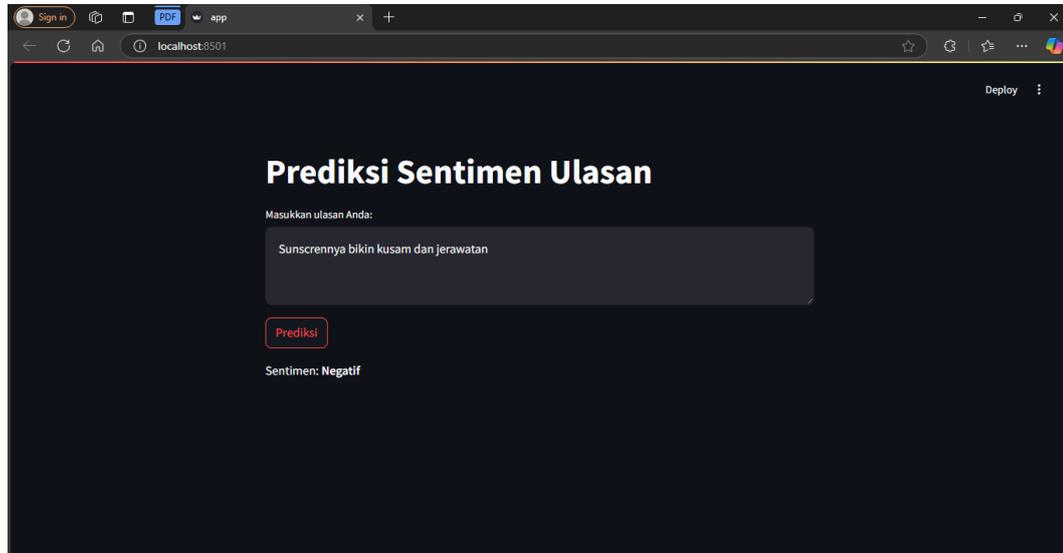
4.2.2 Halaman hasil Sistem Prediksi Sentimen Ulasan



Gambar 4. 5 Halaman Hasil Prediksi Sentimen Positif

Pada gambar 4.7 tersebut, adalah tampilan hasil yang akan dilihat oleh pengguna yang mengakses. Pada tampilan tersebut memasukkan ulasan “sunscreennya bagus banget bikin glowing”. Setelah pengguna mengklik *button* “Prediksi”, sistem akan memproses ulasan dan menampilkan hasil sentimen. pada

tampilan ini, sistem menentukan bahwa ulasan/*review* tersebut memiliki sentimen positif, yang ditampilkan dengan teks “Sentimen: Positif”



Gambar 4. 6 Halaman Hasil Prediksi Sentimen Negatif

Pada gambar 4.8 tersebut, adalah tampilan hasil yang akan dilihat oleh pengguna yang mengakses. Pada tampilan tersebut memasukkan ulasan “Sunscreennya bikin kusam dan jerawatan”. Setelah pengguna mengklik *button* “Prediksi”, sistem akan memproses ulasan dan menampilkan hasil sentimen. pada tampilan ini, sistem menentukan bahwa ulasan/*review* tersebut memiliki sentimen negatif, yang ditampilkan dengan teks “Sentimen: Negatif”

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini mengkaji sentimen dari ulasan produk yang ditemukan di *marketplace* Shopee. Data ulasan dikumpulkan menggunakan teknik *scraping* dan dianalisis menggunakan metode *K-Nearest Neighbor*. Berikut adalah hasil dari penelitian ini :

- a. Model *K-Nearest Neighbor* yang dirancang untuk mengidentifikasi sentimen positif dan negatif dari ulasan, menunjukkan hasil yang cukup baik. Evaluasi menggunakan *confusion matrix* menghasilkan akurasi 0,82, *precision* 83%, *recall* 82%, dan *f1-score* 82%. Model ini berhasil mengklasifikasikan sentimen ulasan dengan baik.
- b. Hasil penelitian menunjukkan bahwa algoritma *K-Nearest Neighbor* memiliki performa yang cukup baik dalam melakukan analisis sentimen.
- c. Pada penelitian ini, memberikan wawasan mengetahui kualitas produk kecantikan berdasarkan ulasan pengguna lain sebelum melakukan pembelian. Hasil analisis sentimen dapat membantu memahami opini dan meningkatkan kualitas produk kecantikan.

5.2 Saran

Penelitian selanjutnya diharapkan fokus pada penanganan data yang tidak sesuai pada klasifikasi sentimen. Dalam pengolahan data yang tidak sesuai, akurasi sistem klasifikasi diharapkan dapat lebih baik dan disarankan untuk mencoba membandingkan metode dan model *machine learning* lainnya.

DAFTAR PUSTAKA

- Amalia, D., Hafid Totohendarto, M., Alam, S., & sitasi, C. (2023). Analisis Sentimen Produk Populer Moisturizer Pada Female Daily Menggunakan Metode Naive Bayes. *Informatics for Educators and Professionals. Informatics for Educators And Professionals : Journal of Informatics*, 8(2), 108–121.
- Ariqoh, S., Agus Sunandar,) M, Muhyidin, Y., Informatika, T., Tinggi, S., Wastukencana, T., & Artikel, R. (2023). *Analisis Sentimen Pada Produk Cushion di Website Female Daily Menggunakan Metode Support Vector Machine (SVM)*. 2(3), 137–142. <https://doi.org/10.55123>
- Arrasyid, R., Putera, D., & Yusuf, A. (2024). *Analisis Sentimen Review Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing*. 18(2).
- Atmodjo, D., Firizqi, J., & Amalia, Z. (2024). Analisis Sentimen Produk Skincare Somethinc Niacinamide di Female Daily dengan Naïve Bayes Classifier. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(2), 946. <https://doi.org/10.30865/mib.v8i2.7571>
- Dinda, M., Azzahra, M., Hafid, T., & Alam, S. (2023). Analisis Sentimen Ulasan Produk Serum Wajah Pada Beauty Brand Somethinc Menggunakan Metode Naive Bayes Classifier. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Nomor 3). <https://reviews.femaledaily.com/products/treatment/s>
- Hanafiah, A., Nasution, A., Arta, Y., Wandri, R., Nasution, H., & Mardafora, J. (2023). Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud dengan Algoritma Naive Bayes Classifier. *Journal of Information Technology and Computer Science (INTECOMS)*, 6(1).
- Harnelia, & Saputra, R. (2024). Analisis Sentimen Review Skincare Skintific dengan Algoritma Support Vector Machine (SVM). *Jurnal Informatika dan Teknik Elektro Terapan*, 12(2). <https://doi.org/10.23960/jitet.v12i2.4095>
- Heltroyce, C., Feoh, G., & Made Dwi Ardiada, I. (2024). *Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Algoritma K-Nearest Neighbor* (Vol. 3, Nomor 1). <https://jurnal.undhirabali.ac.id/index.php/jakasakti/index>

- Juni, D., Andrasthea, G., & Februariyanti, H. (2024). Analisis Sentimen Ulasan Herborist Sistem Pengambilan Keputusan Menggunakan Klasifikasi Neighbor dan TF-ID. *JURNAL SWABUMI*, 12(2), 176–181.
- Kamal, W. W., & Ratnasari, C. I. (2023). *Analisis Sentimen Ulasan Produk: Kajian Pustaka*.
- Kean, L., Indriati, & Marji. (2019). *Analisis Sentimen Review Shopee Berbahasa Indonesia Menggunakan Improved K-Nearest Neighbor dan Jaro Winkler Distance* (Vol. 3, Nomor 7). <http://j-ptiik.ub.ac.id>
- Khasanah, F., Hatta, M., & Jaringan Sistem Informasi Robotik, J. (2024). *Analisis Sentimen Pada Ulasan Pembelian Produk dengan Metode Support Vector Machine* (Vol. 8, Nomor 2). <http://ojsamik.amikmitragama.ac.id>
- Kristianto, B. (2024). *Analisis Sentimen Terhadap TikTok Shop Dengan K-Nearest Neighbor, Decision Tree, dan Naive Bayes*.
- kristiyanti, dinar ajeng. (2015). Analisis Sentimen Review Produk Kosmetik Menggunakan Algoritma Support Vector Machine dan Particle Swarm Optimiation Sebagai Metode Seleksi Fitur. Dalam *Seminar Nasional Inovasi dan Tren*. www.amazon.com
- Mara, A. A. P. T., Sedyono, E., & Purnomo, H. (2021). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba. Dalam *JOINTER-JOURNAL OF INFORMATICS ENGINEERING* (Vol. 01, Nomor 02).
- Muktafin, E., Kusriani, & Luthfi, E. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Jurnal Eksplora Informatika*, 10(1), 32–42. <https://doi.org/10.30864/eksplora.v10i1.390>
- Pradhana, F., Musthafa, A., & Fitria, I. (2024). *Analisis Sentimen Ulasan Produk Daviena di Shopee Menggunakan Algoritma Support Vector Machine*.
- Prakoso, Q., Muliawati, A., & Isnainiyah, I. (2022). Analisis Sentimen terhadap Produk Skin Game di Forum Review Female Daily Menggunakan Metode Multinomial Naïve Bayes dan TF-IDF. *JURNAL INFORMATIK Edisi ke, 18*, 2022.
- Prastika, E., Faraby, S., & Dwifabri, M. (2021). *Analisis Sentimen pada Ulasan Produk Kecantikan Menggunakan K-Nearest Neighbor dan Information Gain* (Vol. 8, Nomor 5).

- Putri, N. L., Warsito, B., & Surarso, B. (2024a). Pengaruh Klasifikasi Sentimen Pada Ulasan Produk Amazon Berbasis Rekamaya Fitur dan *K-Nearest Neighbor*. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 65–74. <https://doi.org/10.25126/jtiik.20241117376>
- Rani, M., & Candra, F. (2023). *Analisis Sentimen Ulasan Toko Online Halona Beauty Care Untuk Peningkatan Layanan Menggunakan Algoritma Naïve Bayes*. 8(2), 2023.
- Raymond Ramadhan, Y., & Andayani Komara, M. (2023). Analisis Sentimen Review Produk Acne Spot Treatment di Female Daily Menggunakan Algoritma K-Nearest Neighbor. *Media Online*, 4(1), 134–143. <https://doi.org/10.30865/klik.v4i1.1070>
- Ridwan, A. N., Hartanti, D., & Sumarlinda, S. (t.t.). Pemodelan Analisis Sentimen pada Ulasan Produk Toko Online By_andwe dengan Algoritma Lexicon. *Jurnalnya Orang Pintar Komputer*, 13(4). <https://doi.org/10.30591/smartcomp.v13i4.6874>
- Rivaldi, R. C., Wismarini, T. D., Lomba, J. T., & Semarang, J. (2024). *Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP) (Studi Kasus Zalika Store 88 Shopee)*. 17(1), 120–128. <https://doi.org/10.51903/elkom.v17i1.1680>
- Saifurridho, M., & Hayati, U. (2024). Jurnal Informatika Terpadu ANALISIS ALGORITMA K-NEAREST NEIGHBOR TERHADAP SENTIMEN PENGGUNA APLIKASI SHOPEE. *Jurnal Informatika Terpadu*, 10(1), 21–26. <https://journal.nurulfikri.ac.id/index.php/JIT>
- Wilantapoera, A., Astuti, W., & Dwifebri, M. (2038). *Analisis Sentimen Kategori Aspek Pada Ulasan Produk Menggunakan Metode KNN Dengan Seleksi Fitur Mutual Information*.