

Perbandingan Analisis Sentimen Pengguna Aplikasi Shopee dan Lazada pada Situs Google Play Store Menggunakan Algoritma K-Nearest Neighbor dan Naive Bayes

Trisnaeni Fardaningsih¹, Anisa Lutfiyani^{2*}

^{1,2*} Teknik Informatika, Universitas Ma'arif Nahdlatul Ulama, Kebumen, Jawa Tengah, Indonesia

Email: ¹danibhawell7@gmail.com, ^{2*}an.lutfiyani@gmail.com

Abstract

This study aims to compare the performance of the Naive Bayes and K-Nearest Neighbor (KNN) algorithms in classifying user sentiment towards two popular e-commerce applications in Indonesia, namely Shopee and Lazada. The data used are 40,000 user reviews collected equally from the Google Play Store, 20,000 reviews for each application. The pre-processing process is carried out through the stages of cleaning, case folding, tokenization, stopword removal, normalization, and stemming. Furthermore, weighting is carried out using the TF-IDF method, then classified using Naive Bayes and KNN with a cosine similarity approach. The evaluation results show that in the Shopee application, Naive Bayes produces an accuracy of 84.72% and an F1-score of 81.56%, while KNN produces an accuracy of 84.17% and an F1-score of 82.31%. On the Lazada application, Naive Bayes achieved an accuracy of 82.53% and an F1-score of 79.89%, while KNN obtained an accuracy of 75.04% and an F1-score of 73.20%. Thus, Naive Bayes proved to be superior in terms of classification accuracy and stability, especially on Lazada data. Meanwhile, KNN showed superiority in the balance of precision and recall on Shopee data. This study contributes to the selection of appropriate algorithms for text-based sentiment analysis in the context of e-commerce in Indonesia.

Keywords: Sentiment Analysis, Naive Bayes, K-Nearest Neighbor, Shopee, Lazada.

Abstract

Penelitian ini bertujuan untuk membandingkan performa algoritma Naive Bayes dan K-Nearest Neighbor (KNN) dalam mengklasifikasikan sentimen pengguna terhadap dua aplikasi e-commerce populer di Indonesia, yaitu Shopee dan Lazada. Data yang digunakan berupa 40.000 ulasan pengguna yang dikumpulkan secara seimbang dari Google Play Store, masing-masing 20.000 ulasan untuk setiap aplikasi. Proses pra-pemrosesan dilakukan melalui tahapan cleaning, case folding, tokenisasi, stopword removal, normalisasi, dan stemming. Selanjutnya, pembobotan dilakukan dengan metode TF-IDF, lalu diklasifikasikan menggunakan Naive Bayes dan KNN dengan pendekatan cosine similarity. Hasil evaluasi menunjukkan bahwa pada aplikasi Shopee, Naive Bayes menghasilkan akurasi 84,72% dan F1-score 81,56%, sedangkan KNN menghasilkan akurasi 84,17% dan F1-score 82,31%. Pada aplikasi Lazada, Naive Bayes mencapai akurasi 82,53% dan F1-score 79,89%, sementara KNN memperoleh akurasi 75,04% dan F1-score 73,20%. Dengan demikian, Naive Bayes terbukti lebih unggul dalam hal akurasi dan stabilitas klasifikasi, terutama pada data Lazada. Sementara itu, KNN menampilkan keunggulan dalam keseimbangan precision dan recall pada data Shopee. Penelitian ini memberikan kontribusi dalam pemilihan algoritma yang sesuai untuk analisis sentimen berbasis teks dalam konteks e-commerce di Indonesia.

Kata Kunci: Analisis Sentimen, Naive Bayes, K-Nearest Neighbor, Shopee, Lazada.

1. PENDAHULUAN

Dalam beberapa tahun terakhir, perkembangan sektor perdagangan elektronik atau *e-commerce* mengalami lonjakan yang sangat signifikan, terutama di kawasan Asia Tenggara. Transformasi digital yang cepat mendorong masyarakat untuk semakin bergantung pada layanan daring dalam memenuhi kebutuhan sehari-hari. Aplikasi *mobile* menjadi media utama bagi konsumen dalam melakukan transaksi, baik untuk pencarian produk, perbandingan harga, hingga proses pembayaran dan pengiriman. Dua platform yang mendominasi pasar *e-commerce* di wilayah ini adalah *Shopee* dan *Lazada*. Keduanya menawarkan beragam kemudahan dan fitur yang menarik, seperti promo harian, sistem pembayaran digital, hingga dukungan logistik yang terintegrasi.

Berdasarkan data pada Mei 2025, aplikasi *Shopee* telah diinstal lebih dari 100 juta kali, sedangkan *Lazada* melebihi 500 juta unduhan di *Google Play Store*. Keduanya memiliki peringkat pengguna tinggi, rata-rata 4,7 dari 5 bintang, dan masing-masing memperoleh jutaan ulasan. Volume ulasan yang sangat besar ini berpotensi menjadi sumber informasi berharga untuk menilai kualitas layanan, pengalaman pengguna, serta persepsi publik terhadap performa aplikasi. Namun, ulasan dalam jumlah besar tidak dapat dianalisis secara manual karena akan memakan waktu lama dan rawan bias subjektif. Oleh karena itu, pendekatan otomatis berbasis analisis sentimen menjadi sangat relevan dan diperlukan. Analisis sentimen merupakan salah satu teknik dalam *Natural Language Processing (NLP)* yang bertujuan mengklasifikasikan opini pengguna dalam bentuk teks ke dalam kategori tertentu, seperti sentimen positif, negatif, atau netral. Dengan teknik ini, opini yang tersebar secara tidak terstruktur di media daring dapat diolah menjadi informasi yang bermanfaat secara sistematis. Untuk mendukung proses klasifikasi sentimen, dibutuhkan algoritma pembelajaran mesin (*machine learning*) yang mampu bekerja dengan data teks dalam jumlah besar dan memiliki performa klasifikasi yang andal.

Dalam penelitian ini, dua algoritma yang digunakan adalah *Naïve Bayes* dan *K-Nearest Neighbor (KNN)*. Keduanya merupakan metode pembelajaran mesin yang banyak digunakan dalam studi klasifikasi teks. *Naïve Bayes* adalah pendekatan berbasis probabilitas yang menghitung peluang suatu dokumen masuk ke dalam kelas tertentu berdasarkan distribusi kata dalam teks. Sementara *KNN* adalah metode *non-parametrik* yang menentukan kelas suatu data berdasarkan kemiripan terhadap data lain yang paling dekat. Meskipun sama-sama banyak digunakan, performa kedua algoritma ini dapat sangat bergantung pada karakteristik data dan konteks penggunaannya. Beberapa penelitian terdahulu telah menggunakan kedua algoritma ini untuk klasifikasi *sentimen*, namun sebagian besar hanya berfokus pada satu jenis aplikasi atau satu algoritma. Penelitian (Putra & Dwiki, 2021) menggunakan *KNN* untuk aplikasi investasi tanpa melakukan perbandingan algoritma. (Safira & Hasan, 2023) menggunakan *Naïve Bayes* pada fitur *Paylater*, tetapi dengan cakupan objek terbatas. Penelitian oleh (Winoto et al., 2024) membandingkan *KNN* dan *Naïve Bayes*, namun dilakukan pada aplikasi edukasi Duolingo, bukan *e-commerce*. (Limbong et al., 2022) dan (Muslim et al., 2024) juga melakukan studi serupa, tetapi terbatas pada satu aplikasi atau jumlah data yang tidak seimbang.

Dari studi tersebut, dapat disimpulkan bahwa masih terdapat kesenjangan penelitian (*research gap*), yaitu belum adanya studi yang secara langsung membandingkan performa *Naïve Bayes* dan *KNN* dalam konteks dua aplikasi *e-commerce* besar sekaligus, dengan data yang besar, seimbang, dan diambil dari sumber yang sama. Penelitian ini untuk mengisi gap tersebut dengan membandingkan kedua algoritma pada 40.000 data ulasan pengguna yang dikumpulkan secara seimbang dari aplikasi *Shopee* dan *Lazada* di

Google Play Store, masing-masing sebanyak 20.000 ulasan. Kebaruan (*novelty*) dari penelitian ini terletak pada pendekatannya yang menyeluruh dan terstandarisasi, mulai dari pengumpulan data secara otomatis menggunakan teknik *web scraping*, *pre-processing* teks (*cleaning*, *normalisasi*, *tokenisasi*, dll.), pelabelan sentimen berdasarkan skor, hingga klasifikasi dan evaluasi performa dengan metrik *akurasi*, *precision*, *recall*, dan *F1-score*. Tidak banyak studi sebelumnya yang melakukan perbandingan langsung dengan pendekatan eksperimen yang seragam terhadap dua aplikasi sekaligus dan dengan jumlah data sebesar ini. Secara ilmiah, penelitian ini memberikan kontribusi pada pengembangan metode klasifikasi teks dalam bidang pemrosesan bahasa alami dan *machine learning*. Dari sisi praktis, hasilnya dapat dimanfaatkan oleh pengembang aplikasi untuk memahami pola sentimen pengguna secara objektif dan berbasis data, serta oleh pelaku industri *e-commerce* dalam menyusun strategi peningkatan kualitas layanan, fitur aplikasi, maupun penanganan keluhan pelanggan secara lebih tepat sasaran.

Dengan pendekatan eksperimental yang sistematis serta penggunaan dataset besar dan representatif dari dua aplikasi *e-commerce* utama, penelitian ini diharapkan dapat menjadi rujukan dalam pemilihan algoritma klasifikasi sentimen yang efektif. Temuan ini juga memberikan kontribusi dalam pengembangan pendekatan analitik berbasis data yang dapat digunakan untuk mengevaluasi persepsi pengguna secara objektif dan mendukung peningkatan kualitas layanan dalam industri *e-commerce* di Indonesia dan Asia Tenggara.

1.1 *Analysis Sentimen*

Analisis sentimen merupakan metode otomatis yang menggunakan teknik *Natural Language Processing (NLP)* dan algoritma pembelajaran mesin untuk mengklasifikasi opini dalam teks, seperti komentar di media sosial, ke dalam kategori positif, negatif, atau netral (Sari & Wibowo, 2019). Teknik ini menganalisis kata dan frasa dalam teks untuk menentukan *sentimen* secara keseluruhan. Algoritma yang sering digunakan meliputi *K-Nearest Neighbor (KNN)* dan *Naive Bayes*, yang masing-masing memiliki keunggulan tersendiri dalam menangani data teks secara efisien. *Analisis sentimen* membantu memahami kepuasan pelanggan, mengidentifikasi tren pasar, serta mendukung pengambilan keputusan berbasis data.

1.2 *K-Nearest Neighbor (KNN)*

K-Nearest Neighbor (KNN) adalah salah satu algoritma paling sederhana dan sering digunakan dalam pengelompokan data dan maupun teks (Kusuma & Cahyono, 2023). *K-Nearest Neighbor (KNN)* bekerja dengan mencari data yang paling mirip berdasarkan jarak terdekat, lalu mengklasifikasikan objek baru ke dalam kelas mayoritas dari data tetangganya. Dalam konteks teks, kemiripan dihitung menggunakan metode cosine similarity, yang mengukur kesamaan arah antar vektor dokumen. Proses *KNN* meliputi: menentukan nilai *k* (jumlah tetangga), menghitung bobot kata (misalnya dengan *TF-IDF*), mengukur kemiripan antar dokumen menggunakan rumus cosine similarity:

$$\cos(\theta_{ij}) = \frac{\sum_k (d_{ik}d_{jk})}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}}$$

Hasilnya diurutkan dari kemiripan tertinggi, lalu dipilih *k* data terdekat untuk menentukan kelas dokumen yang diklasifikasikan.

1.3 Naive Bayes

Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang sederhana namun efisien. Algoritma ini menghitung *probabilitas* suatu data termasuk ke dalam kelas tertentu berdasarkan nilai-nilai atribut yang ada pada data. Menurut (Paradisona & Devega, 2024) *Naive Bayes*, adalah algoritma klasifikasi berbasis probabilitas yang digunakan untuk memprediksi kelas suatu data, seperti sentimen positif, negatif, atau netral, berdasarkan kata-kata dalam teks. Algoritma ini didasarkan pada Teorema Bayes dengan rumus:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Rumus ini digunakan untuk menghitung kemungkinan suatu data (X), seperti ulasan teks, termasuk ke dalam kelas tertentu (C), misalnya sentimen positif. Di dalamnya, $P(C|X)$ adalah kemungkinan data X termasuk ke dalam kelas C, $P(X|C)$ menunjukkan seberapa besar kemungkinan data X muncul jika diketahui termasuk kelas C, $P(C)$ adalah kemungkinan awal dari kelas C, dan $P(X)$ adalah kemungkinan keseluruhan data X. Meskipun *Naive Bayes* mengasumsikan bahwa setiap kata dalam teks tidak saling memengaruhi, algoritma ini tetap efektif karena sederhana, cepat, dan mampu mengolah data dalam jumlah besar. Prosesnya dimulai dari pelatihan model, perhitungan kemungkinan, dan pemilihan kelas dengan nilai kemungkinan tertinggi sebagai hasil akhir klasifikasi.

1.4 Web Scraping

Web scraping adalah proses otomatis untuk mengumpulkan data dari situs *web* atau media sosial, kemudian disimpan dalam file teks atau database untuk dianalisis lebih lanjut (Carina et al., 2024). Teknik ini bermanfaat dalam analisis sentimen, misalnya untuk mengumpulkan ulasan pengguna dari aplikasi *Shopee* dan *Lazada* di *Google Play Store*. Proses ini meliputi mengakses halaman *web*, mengambil konten, memilah data relevan seperti isi ulasan (*review content*), *score/rating*, tanggal ulasan, nama pengguna/id, dan menyimpannya untuk analisis. *Web scraping* menyediakan data mentah yang berguna untuk mengidentifikasi pola *sentimen* dan memahami respons pengguna terhadap produk atau layanan tertentu.

1.5 E-commerce

E-commerce adalah metode jual beli produk dan layanan secara daring melalui internet dan teknologi digital (Wahyuni et al., 2019). Transaksi berlangsung tanpa interaksi langsung antara penjual dan pembeli, dan mencakup pertukaran barang, jasa, maupun informasi. Menurut Philip Kotler, *e-commerce* merupakan proses bisnis elektronik antar perusahaan, konsumen, dan masyarakat. Unsur penting dalam *e-commerce* meliputi platform digital (*website/aplikasi*), pembayaran elektronik (kartu kredit, e-wallet, kripto), pengiriman/logistik, keamanan transaksi, serta pemasaran digital. Keuntungannya mencakup kemudahan akses, ragam produk, dan harga yang kompetitif.

1.6 Shopee

Menurut (Sea, 2025) *Shopee* yang diperkenalkan pada tahun 2015 merupakan platform *e-commerce* terkemuka di kawasan Asia Tenggara dan Taiwan yang memberikan pengalaman berbelanja online yang cepat, aman, dan nyaman melalui dukungan yang kuat dalam hal pembayaran dan logistik. Pengguna utama *Shopee* merupakan generasi muda yang terbiasa menggunakan gadget untuk berbagai aktivitas,

termasuk berbelanja, dan mereka juga menyediakan proses pembayaran yang terpecaya serta manajemen logistik yang terintegrasi bagi para penjual. Oleh karena itu, *Shopee* hadir dalam bentuk aplikasi *mobile* untuk mempermudah dan mempercepat kegiatan berbelanja.

1.7 Lazada

Lazada adalah salah satu *platform e-commerce* terbesar di Asia Tenggara, menyajikan beragam produk, dari gadget, fashion, alat rumah tangga, hingga barang kebutuhan sehari-hari. Konsumen dapat berbelanja secara online dengan mudah dan terjamin keamanannya. Tersedia baik dalam aplikasi *mobile* maupun situs *web*nya. Sebagaimana dinyatakan oleh (Lazada Group, 2025) didirikan oleh *Rocket Internet* pada tahun 2012 dan kini berada dibawah kepemilikan Alibaba Group. *Platform* ini beroperasi di berbagai negara, termasuk Indonesia, Malaysia, Filipina, Singapura, Thailand, dan Vietnam. menyediakan beraneka metode pembayaran, termasuk opsi pembayaran tunai saat pengiriman (*Cash on Delivery*) yang sangat diminati di Indonesia.

1.8 Google Play Store

Google Play Store adalah platform distribusi digital milik *Google* yang menyediakan layanan untuk mencari, mengunduh, dan memperbarui aplikasi Android, serta mengakses konten digital seperti musik, film, buku, dan game (Firdaus et al., 2024). Diluncurkan pada 22 Oktober 2008 sebagai Android Market dan direbranding menjadi *Play Store* pada Maret 2012, *platform* ini telah berkembang menjadi salah satu toko aplikasi terbesar di dunia. *Play store*, memainkan peran penting dalam ekosistem Android dengan memberikan akses yang mudah ke berbagai aplikasi dan konten digital bagi pengguna di seluruh dunia .

1.9 Ulasan

Ulasan adalah tanggapan atau evaluasi yang diberikan seseorang terhadap produk, layanan, atau pengalaman tertentu setelah menggunakan aplikasi tersebut. Ulasan biasanya dipublikasikan secara online melalui platform *e-commerce*, aplikasi, atau media sosial, dan berisi deskripsi pengalaman, kelebihan dan kekurangan, serta penilaian berupa bintang atau skor. Bagi konsumen, ulasan membantu dalam mengambil keputusan, sedangkan bagi perusahaan, ulasan menjadi sumber masukan untuk meningkatkan kualitas. Di era digital, ulasan berperan penting dalam membangun kepercayaan dan reputasi suatu produk atau layanan.

1.10 Python

Python dikembangkan oleh *Guido van Rossum* pada pertengahan tahun 1990 an di *centrum wiskunde and informatica (CWI)*, Belanda. *Python* adalah bahasa pemrograman berbasis interpreter yang dapat dijalankan di berbagai *platform* seperti *Windows, Linux*, dan lainnya (Nazar, 2024). *Python* memiliki beberapa kelebihan dan kekurangan yakni Kelebihan *Python* meliputi sintaks yang relatif sederhana dibandingkan bahasa pemrograman lain, sehingga mudah dipelajari. *Python* juga bersifat multifungsi dan efisien dalam penulisan kode dibandingkan bahasa seperti *C, C++*, atau *Java*. Selain itu, *Python* dapat berkolaborasi dengan berbagai bahasa pemrograman lain, seperti *C/C++*, yang membuatnya fleksibel dalam pengembangan. Namun, *Python* juga memiliki sejumlah kelemahan. Beberapa tugas tidak dapat ditangani secara optimal oleh *Python*, terutama dalam pengembangan aplikasi mobile seperti *Android* dan *iOS*, karena tingkat portabilitasnya rendah. Sebagai bahasa berbasis interpreter, *Python* kurang efisien dalam mengeksekusi program dasar dan memiliki keterbatasan dalam mengakses informasi

sistem secara langsung. Selain itu, *Python* tidak ideal digunakan sebagai bahasa pemrograman dasar untuk sistem yang membutuhkan kontrol tingkat rendah. Meskipun demikian, *Python* tetap populer karena menawarkan produktivitas dan fleksibilitas yang tinggi.

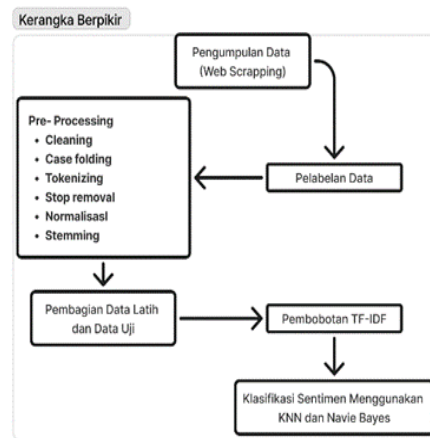
1.11 Google Collab

Google Collab merupakan *platform* berbasis *web* yang digunakan untuk menulis dan menjalankan kode *Python* langsung melalui browser, tanpa perlu instalasi perangkat lunak tambahan. Dikembangkan oleh *Google*, *platform* ini dirancang untuk mendukung pembelajaran mesin, analisis data, serta eksperimen ilmiah. Salah satu keunggulannya adalah akses gratis ke sumber daya komputasi yang kuat seperti *GPU* (*GraphicsProcessing Unit*) dan *TPU* (Rofiq & Sitio, 2021).

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif karena bertujuan untuk mengevaluasi dan membandingkan performa algoritma klasifikasi secara objektif berdasarkan data numerik. Proses dimulai dari pengambilan data ulasan melalui *web scraping*, dilanjutkan dengan pelabelan data, *pre-processing*, pembobotan *TF-IDF*, pelatihan model, dan evaluasi. Dua algoritma yang digunakan adalah *Naive Bayes* dan *K-Nearest Neighbor* (*KNN*), dengan kinerja dianalisis menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*.



Gambar 1 Kerangka Berpikir

2.1.1 Pengumpulan data (Web Scrapping)

Data dikumpulkan secara otomatis menggunakan teknik *web scraping* dengan library *google_play_scraper* versi 1.2.7 di *Python* (Zaitsev, 2020), tanpa perlu parsing *HTML* manual. Sebanyak 20.000 ulasan berbahasa Indonesia dari masing-masing aplikasi *Shopee* dan *Lazada* di *Google Play Store* dikumpulkan melalui *Google Colab*. Data meliputi ulasan, skor, nama pengguna, dan tanggal publikasi, lalu disimpan dalam format *CSV* untuk analisis lebih lanjut.

2.1.2 Pelabelan data

Pelabelan data dilakukan secara otomatis menggunakan metode *rule-based* dengan pustaka *pandas* di *Python*. Label sentimen dibagi menjadi tiga kategori berdasarkan skor ulasan: positif (rating 4–5), negatif (rating 1–2), dan netral (rating 3). Aturan pelabelan

diterapkan menggunakan fungsi if-else, yang membaca nilai skor dan menetapkan label sesuai kriteria tersebut. Pendekatan ini terbukti efisien dan konsisten dalam klasifikasi, serta banyak digunakan dalam penelitian sebelumnya (Firdaus et al., 2024), (Safira & Hasan, 2023), serta (Winoto et al., 2024).

2.1.3 Pre-processing

Sebelum data dapat digunakan untuk proses klasifikasi, teks ulasan harus melalui tahapan *pre-processing* agar sesuai untuk dianalisis. Tahapan ini terdiri dari:

1. *Cleaning*, *cleaning* adalah tahapan dalam proses pembersihan data yang mencakup pembersihan tanda baca, menghapus karakter atau symbol yang tidak diperlukan.

Table 1 Contoh Hasil Cleaning

Content	Text Cleaning
Alhamdulillah selama saya belanja di <i>shopee</i> , belum ada masalah, sukses selalu buat perusahaan <i>SHOPEE</i> 👍🙏	Alhamdulillah selama saya belanja di <i>shopee</i> belum ada masalah sukses selalu buat perusahaan <i>SHOPEE</i>

2. *Case folding* merupakan proses mengubah huruf kapital menjadi huruf kecil.

Table 2 Contoh Hasil Case Folding

Text Cleaning	Text Casefolding
Alhamdulillah selama saya belanja di <i>shopee</i> belum ada masalah sukses selalu buat perusahaan <i>SHOPEE</i>	alhamdulillah selama saya belanja di <i>shopee</i> belum ada masalah sukses selalu buat perusahaan <i>shopee</i>

3. *Stopword remove* yaitu penghapusan kata umum merujuk pada istilah yang dianggap gangguan dan perlu dihilangkan.

Table 3 contoh Hasil Stopword Remove

Text Casefolding	Text stopwords
alhamdulillah selama saya belanja di <i>shopee</i> belum ada masalah sukses selalu buat perusahaan <i>shopee</i>	alhamdulillah belanja <i>shopee</i> belum masalah sukses selalu perusahaan <i>shopee</i>

4. *Tokenizing* adalah langkah untuk membagi kalimat menjadi kata-kata dalam *dataset*.

Table 4 Contoh Hasil Tokenizing

Text stopwords	Text tokens
alhamdulillah belanja <i>shopee</i> belum masalah sukses selalu perusahaan <i>shopee</i>	alhamdulillah, belanja, <i>shopee</i> belum, masalah sukses, selalu, perusahaan, <i>shopee</i>

5. *Normalisasi* merupakan proses mengubah kata tidak baku menjadi baku sesuai dengan kamus KBBI.

Table 5 Contoh Hasil Normalisasi

Text tokens	Text normalisasi
alhamdulillah, belanja, <i>shopee</i> belum, masalah sukses, selalu, perusahaan, <i>shopee</i>	alhamdulillah, belanja, <i>shopee</i> , belum, masalah, sukses, selalu, perusahaan, <i>shopee</i>

6. *Stemming* adalah proses menghapus kata awal dan akhir kata untuk mendapatkan bentuk kata dasar.

Table 6 Contoh Hasil Stemming

Text normalisasi	Text steamindo
alhamdulillah, belanja, <i>shopee</i> , belum, masalah, sukses, selalu, perusahaan, <i>shopee</i>	alhamdulillah belanja <i>shopee</i> belum masalah sukses selalu usaha <i>shopee</i>

2.1.4 Pembagian data *training* dan data *testing*,

Setelah data dibersihkan, data dibagi menjadi data *training* (latih) dan data *testing* (uji) dengan rasio 80:20, menggunakan teknik *stratified sampling*. Metode ini menjaga proporsi label sentiment (positif, negative, netral) tetap seimbang di kedua subset data, sehingga hasil evaluasi model menjadi lebih *representative*.

2.1.5 Pembobotan *TF-IDF*,

Proses pembobotan dilakukan menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*, yang mengubah teks menjadi representasi numerik berbentuk vector berdasarkan Tingkat kepentingan kata dalam dokumen. Pembobotan ini penting agar algoritma dapat mengenali pola dalam data teks secara efektif. Pemilihan *TF-IDF*, dibanding metode lain seperti *count vectorizer*, lebih cocok dalam konteks pendekatan kualitatif karena mempertahankan pentingnya konteks kata dalam struktur makna ulasan.

2.1.6 Klasifikasi *sentiment*,

Klasifikasi dilakukan menggunakan dua algoritma, yaitu *K-Nearest Neighbor (KNN)* dan *Naive Bayes*. Eksperimen dibagi menjadi dua tahap: pertama untuk ulasan *Shopee* dan kedua untuk *Lazada*, masing-masing dijalankan pada notebook terpisah di platform *Google Colab*. Proses meliputi pelatihan (*training*) dan pengujian (*testing*) model, dengan evaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Untuk *Naive Bayes* digunakan *MultinomialNB* dari *sklearn.naive_bayes*, yang sesuai untuk data hasil representasi *TF-IDF*.

```
import pandas as pd
from sklearn.naive_bayes import MultinomialNB
#naive bayes
nb = MultinomialNB()
nb.fit(tfidf_train, y_train)
```

KNN diimplementasikan menggunakan *KNeighborsClassifier* dari *sklearn.neighbors*. Proses tuning dilakukan pada rentang $k = 1$ hingga 30, dan hasilnya menunjukkan bahwa nilai k terbaik adalah 3 untuk *Shopee* dan 5 untuk *Lazada*, seperti ditunjukkan pada Gambar 5 dan 6.

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier (n_neighbors=best_k)
knn.fit(tfidf_train, y_train)
```

Algoritma *KNN* menggunakan *cosine similarity* karena lebih fokus pada arah vektor dibanding panjangnya, sehingga cocok untuk teks pendek seperti ulasan pengguna. Metode ini memungkinkan pengenalan makna meskipun panjang ulasan berbeda. Selain itu, *KNN* juga diimplementasikan secara kustom menggunakan *cosine_distances* dari *sklearn.metrics.pairwise*. Tujuan klasifikasi ini tidak hanya untuk mendapatkan akurasi terbaik, tetapi juga memahami pola sentimen pengguna pada aplikasi *Shopee* dan *Lazada*.

3. HASIL DAN PEMBAHASAN

3.1 Pengambilan Data

Data dikumpulkan dengan metode *web scraping* menggunakan pustaka *google_play_scraper* versi 1.2.7 di Python (Zaitsev, 2020) tanpa perlu melakukan parsing *HTML* secara manual. Sebanyak 40.000 ulasan berhasil diperoleh, masing-masing 20.000 dari aplikasi *Shopee* dan *Lazada*, dan disimpan dalam format *CSV* yang berisi kolom *content* (isi ulasan) dan *score* (rating).


```
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'com.shopee.id',
    lang='id',
    country='id',
    sort=Sort.NEWEST,
    count=20000,
    filter_score_with=None
)
```

Gambar 2 Source Code Scraping Shopee Dan Lazada

Gambar 2 menampilkan cuplikan kode *scraping*, sedangkan Tabel 7 dan Tabel 8 menunjukkan contoh data mentah dari *Shopee* dan *Lazada*. Contohnya, ulasan “lemot buat bukak susah” dari *Shopee* dengan skor 1 akan dilabeli sebagai sentimen negatif.

Table 7 Hasil Scraping Data *Shopee*

content	score
lemot buat bukak susah	1
kenapa tiap masuk aplikasinya selalu tiba² masuk ke live&video 🤔	2
very good	5
shoppe terpercaya	5
Alhamdulillah selama saya belanja di <i>shopee</i> , belum ada masalah, sukses slalu buat perusahaan <i>SHOPEE</i> 👍🙏	5

Table 8 Hasil Scraping Data *Lazada*

content	score
sangat ok	5
tambah susah daftarnya	1
Ok mantap	5
aplikasi paling berat dudunia. hp samsung s22 ...	3
mudah di gunakan di saat lagi sibuk	5

3.2 Pre-Processing.

Setelah data terkumpul, dilakukan tahap *pre-processing* untuk menyiapkan data sebelum pelatihan dan pengujian model. Proses ini meliputi *cleaning* (menghapus tanda baca dan karakter tidak penting), *case folding* (mengubah huruf menjadi kecil), *tokenizing* (memecah teks menjadi kata), *stopword removal* (menghapus kata umum seperti “yang”, “dan”), *normalisasi* (mengubah kata tidak baku menjadi baku, seperti “gk” menjadi “tidak”), dan *stemming* (mengubah kata ke bentuk dasar, seperti “berbelanja” menjadi “belanja”). Hasil *pre-processing* ditampilkan pada Tabel 9–10 (*Shopee*) dan Tabel 11–12 (*Lazada*).

Table 9 Hasil *Preprocessing* Pada *Shopee*

Content	Score	Label	Text Casefolding
lemot buat bukak susah	1	Negatif	lemot buat bukak susah
kenapa tiap masuk aplikasinya selalu tiba² masuk ke live&video 🤔	2	Negatif	kenapa tiap masuk aplikasinya selalu tiba masuk ke livevideo
....			
....			
Alhamdulillah selama saya belanja di <i>shopee</i> , belum ada masalah, sukses slalu buat perusahaan <i>SHOPEE</i> 👍🙏	5	Positif	alhamdulillah selama saya belanja di <i>shopee</i> belum ada masalah sukses slalu buat perusahaan <i>shopee</i>

Table 10 Hasil *Preprocessing* Pada *Shopee*

Text stopwords	Text tokens	Text normalisasi	Text steamindo
lemot bukak susah	lemot,bukak,susah	lemot,bukak,susah	lot bukak susah
masuk aplikasinya masuk livevideo	masuk,aplikasinya,masu k,livevideo	masuk,aplikasinya,masu k,live vidio	masuk aplikasi masuk livevideo
....			
....			
alhamdulillah belanja shopeebelum masalahsukses slalu perusahaan shopee	alhamdulillah,belanja,sh opeebelum,masalahsuk es,slalu,perusahaan,sho pee	alhamdulillah,belanja,sh opee,belum,masalah,suk ses,slalu,perusahaan,sho pee	alhamdulillah belanja shopeebelum masalahsukses slalu usaha shopee

Table 11 Hasil *Preprocessing* Pada *Lazada*

Content	score	Label	Text casefolding
sangat ok	5	Positif	sangat ok
tambah susah daftarnya	1	Negatif	tambah susah daftarnya
....			
....			
aplikasi paling berat dudunia. hp samsung s22 ultra aja jadi kek hape bodoh klo buka	3	Netral	aplikasi paling berat dudunia hp samsung s ultra aja jadi kek hape bodoh klo buka

Table 12 Hasil *Preprocessing* Pada *Lazada*

Text stopwords	Text tokens	Text normalisasi	Text steamindo
ok	ok	ok	ok
susah daftarnya	susah,daftarnya	susah,daftarnya	susah daftar
....			
....			
aplikasi berat dudunia hp samsung s ultra aja kek hape bodoh klo buka	aplikasi,berat,dudunia, hp,samsung,s,ultra,aja, kek,hape,bodoh,klo,bu ka,	aplikasi,berat,didunia,h p,samsung,s,ultra,aja,k ek,hape,bodoh,klo,buk a,	aplikasi berat dudunia hp samsung s ultra aja kek hape bodoh klo buka

3.3 Persiapan Dataset

Setelah tahap *preprocessing*, data dilabeli secara otomatis menggunakan script *Python* berbasis aturan (*rule-based*), dengan fungsi *if-else* untuk mengkategorikan skor ulasan. Skor 4–5 dianggap sebagai sentimen positif, 1–2 sebagai negatif, dan skor 3 sebagai netral. Pendekatan ini memberikan keunggulan dari sisi kecepatan dan konsistensi, tanpa campur tangan manual. Contoh hasil pelabelan ditampilkan pada Tabel 13 dan Tabel 14.

Table 13 Hasil Pelabelan Pada *Shopee*

content	score	Label
lemot buat bukak susah	1	Negatif
kenapa tiap masuk aplikasinya selalu tiba ² masuk ke live&video 🥲	2	Negatif
very good	5	Positif
Bagus	5	Positif
shoppe terpercaya	5	Positif

Table 14 Hasil Pelabelan Pada *Lazada*

content	score	Label
sangat ok	5	Positif
tambah susah daftarnya	1	Negatif
Ok mantap	5	Positif
aplikasi paling berat dudunia. hp samsung s22 ultra aja jadi kek hape bodoh klo buka	3	Netral
mudah di gunakan di saat lagih sibuk	5	Positif

Proses pembobotan menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* untuk mengubah teks menjadi vektor numerik berdasarkan pentingnya kata dalam dokumen. Data dibagi 80% untuk pelatihan dan 20% untuk pengujian. Pada Shopee, digunakan 14.505 ulasan (11.604 data latih dan 2.901 data uji) dengan 10.310 fitur unik. Sementara pada Lazada, terdapat 13.424 ulasan (10.739 data latih dan 2.685 data uji) dengan 10.184 fitur unik. Hasil pembobotan ditampilkan pada Gambar 3 dan 4.

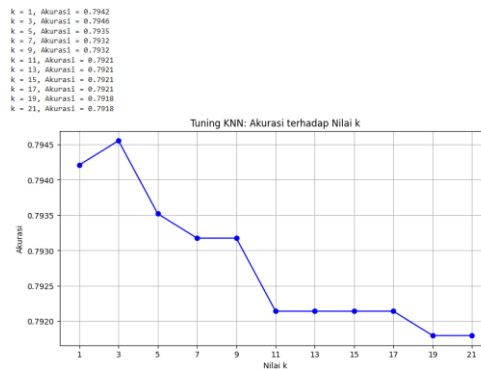
data latih (TF-IDF): (11604, 10310)
data uji (TF-IDF): (2901, 10310)

Gambar 3 Hasil Pembobotan TF-IDF Di Shopee

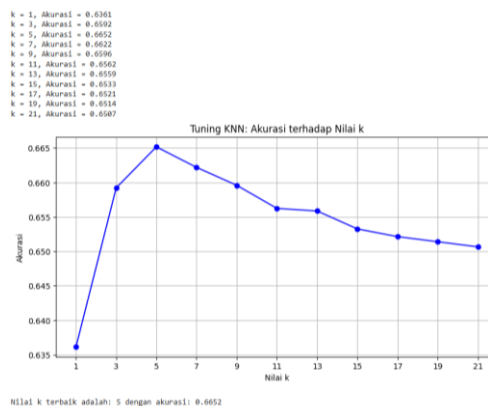
data latih (TF-IDF): (10739, 10184)
data uji (TF-IDF): (2685, 10184)

Gambar 4 Hasil Pembobotan TF-IDF Di Lazada

Dalam proses tuning parameter k pada algoritma *K-Nearest Neighbor (KNN)*, diperoleh hasil berbeda untuk masing-masing dataset. Pada dataset *Shopee*, akurasi terbaik sebesar 79,46% dicapai saat nilai $k = 3$, sebagaimana ditunjukkan pada Gambar 5. Hal ini menunjukkan bahwa mempertimbangkan tiga tetangga terdekat memberikan klasifikasi yang paling optimal. Sementara itu, pada dataset *Lazada*, nilai k optimal adalah 5 dengan akurasi tertinggi sebesar 66,52%, seperti terlihat pada Gambar 6. Nilai k yang lebih besar pada *Lazada* dapat disebabkan oleh kompleksitas ekspresi dalam ulasan pengguna yang membutuhkan konteks dari lebih banyak tetangga agar pola sentimen dapat dikenali secara efektif.



Gambar 5 Hasil Tuning Terhadap Nilai K Pada Shopee



Gambar 6 Hasil Tuning Terhadap Nilai K Pada Lazada

Penelitian ini bertujuan membandingkan performa dua algoritma pembelajaran mesin, yaitu *Naive Bayes* dan *K-Nearest Neighbor (KNN)*, dalam klasifikasi sentimen pengguna aplikasi *Shopee* dan *Lazada*. Berdasarkan hasil akhir evaluasi (Tabel 15) Dalam penelitian ini menggunakan pendekatan klasifikasi *Naive Bayes* dan *K-Nearest Neighbor (KNN)* diterapkan dengan perubahan *cosine similarity*.

Table 15 Hasil Evaluasi Model *Naive Bayes* Dan

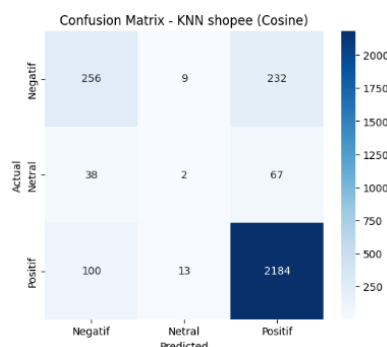
Model	Aplikasi	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	Shopee	0.8472	0.8084	0.8472	0.8156
	Lazada	0.8253	0.7751	0.8253	0.7989
K-Nearest Neighbor	Shopee	0.8417	0.8108	0.8417	0.8231
	Lazada	0.7504	0.7216	0.7504	0.7320

Berdasarkan Tabel 15, performa kedua algoritma sangat dipengaruhi oleh karakteristik data dari masing-masing aplikasi. Pada data *Shopee*, algoritma *Naïve Bayes* menghasilkan akurasi sebesar 0,8472 dan *F1-score* 0,8156, sedangkan *K-Nearest Neighbor (KNN)* mencatat akurasi 0,8417 dengan *F1-score* lebih tinggi, yaitu 0,8231. Ini menunjukkan bahwa *KNN* mampu memberikan klasifikasi yang lebih seimbang antara *precision* (0,8108) dan *recall* (0,8417). Sementara itu, pada data *Lazada*, *Naïve Bayes* menunjukkan performa yang lebih unggul secara keseluruhan dengan akurasi 0,8253 dan *F1-score* 0,7989, dibandingkan *KNN* yang hanya memperoleh akurasi 0,7504 dan *F1-score* 0,7320. Hal ini mengindikasikan bahwa *Naïve Bayes* lebih efektif dalam menangani ulasan yang panjang dan kompleks, seperti pada aplikasi *Lazada*, sedangkan *KNN* lebih optimal untuk data dengan struktur kalimat yang sederhana dan langsung seperti pada *Shopee*.

Ulasan pengguna pada aplikasi *Lazada* umumnya lebih panjang dan mengandung kombinasi antara pujian serta keluhan dalam satu kalimat. Karakteristik ini lebih sesuai untuk algoritma *Naïve Bayes* yang mampu menghitung probabilitas kata secara independen. Sebaliknya, algoritma *KNN* yang mengandalkan kemiripan literal melalui *cosine similarity* mengalami kesulitan dalam mengklasifikasikan konteks ulasan semacam ini. Sementara itu, ulasan pada aplikasi *Shopee* cenderung lebih singkat, langsung, dan memiliki struktur yang sederhana, sehingga lebih tepat diklasifikasikan menggunakan *KNN* yang menunjukkan keseimbangan *precision* dan *recall*, khususnya dalam mengenali ulasan negatif yang jumlahnya relatif sedikit.

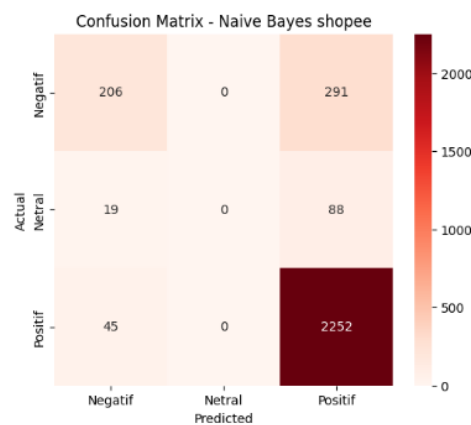
3.4 Evaluasi

Evaluasi model dilakukan dengan mengukur akurasi, precision, recall, dan *F1-score* yang didasarkan pada confusion matrik. Metrik ini memberikan gambaran performa klasifikasi untuk masing-masing kelas sentimen (positif, negatif, netral) dapat dilihat seperti dibawah :



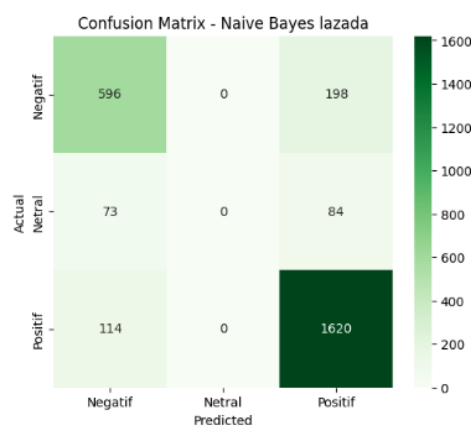
Gambar 7 Hasil Confusion Matrik Pada Shopee

Pada Gambar 7, Model *KNN* pada *Shopee* menunjukkan performa tinggi terutama dalam mengklasifikasikan kelas positif. Hal ini terlihat dari nilai *true positive (TP)* sebanyak 2.184 data berhasil dikenali dengan benar. Namun, model ini masih mengalami *false positive (FP)* yang cukup tinggi untuk kelas positif, yakni 232 data negatif yang salah diklasifikasikan sebagai positif. Distribusi klasifikasi untuk kelas netral (38 negatif, 2 netral, 67 positif) menunjukkan bahwa model belum optimal membedakan sentimen netral, meskipun dibandingkan dengan *Naive Bayes*, masih ada pengenalan kelas netral walau kecil. Oleh karena itu, klaim bahwa kelas netral “seimbang” perlu diluruskan: lebih tepat jika dikatakan masih dikenali meskipun jumlahnya minim, bukan seimbang secara proporsional.



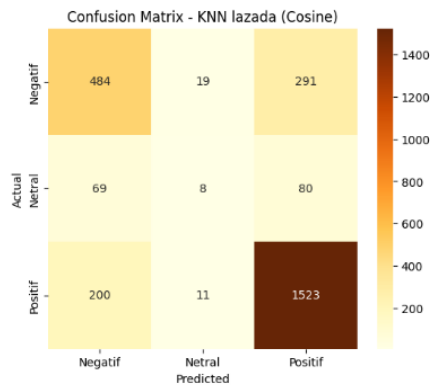
Gambar 8 Hasil *Confusion Matrik Naive Bayes Pada Shopee*

Pada Gambar 8, Model *Naive Bayes* menghasilkan *TP* sebesar 2.252 untuk kelas positif, lebih tinggi dari *KNN*. Namun, model ini gagal mengenali seluruh data pada kelas netral (19 data netral diprediksi sebagai negatif, 88 sebagai positif) dengan 0 true netral, menunjukkan kelemahan *Naive Bayes* dalam mengenali sentimen sedang/netral. Selain itu, 291 data negatif juga salah diklasifikasikan sebagai positif, yang menandakan tingginya *false positive*.



Gambar 9 Hasil *Confusion Matrik Naive Bayes Pada Lazada*

Pada Gambar 9, *Naive Bayes* pada *Lazada* memiliki performa moderat. *TP* kelas positif mencapai 1.620, tetapi kelemahan utama kembali terlihat pada kegagalan mengklasifikasikan kelas netral, karena 73 data netral diklasifikasikan sebagai negatif dan 84 sebagai positif, tanpa satupun dikenali secara akurat. Untuk kelas negatif, model cukup baik dengan 596 *true negative*, namun tetap menghasilkan 198 *false positive*.



Gambar 10 Hasil *Confusion Matrik* Di *Lazada*

Gambar 10, Kinerja *KNN* untuk *Lazada* lebih seimbang dibanding *Naive Bayes*. Model berhasil mengenali 1.523 data positif secara benar dan 484 data negatif. Meskipun begitu, *KNN* juga menghasilkan *false positive* (291 negatif → positif) dan menunjukkan kelemahan serupa dengan *Shopee* dalam mengenali sentimen netral (69 negatif, 8 netral, 80 positif). Dari hasil yang ditampilkan:

1. *Shopee - KNN* menunjukkan *F1-score* yang seimbang karena memiliki *precision* dan *recall* yang relatif merata, terutama untuk kelas positif, meskipun ada *trade-off* di kelas negatif.
2. *Shopee - Naive Bayes* unggul dalam *recall* untuk kelas positif namun memiliki *precision* yang rendah karena banyaknya *false positive*.
3. *Lazada - KNN* lebih baik dibanding *Naive Bayes* dalam mengenali sentimen netral dan negatif, namun masih belum ideal.
4. *Lazada - Naive Bayes* secara umum menunjukkan ketidakseimbangan karena *precision* dan *recall* yang timpang, serta pengenalan netral yang gagal total.

Kontribusi dari *confusion matrik* menunjukkan bahwa *KNN* lebih konsisten dalam mengenali sentimen netral dan negatif dibanding *Naive Bayes*. Akan tetapi, keduanya masih memiliki kelemahan dalam mendeteksi sentimen netral, terutama karena distribusi data yang mungkin tidak seimbang dan kompleksitas pola teks yang ambigu.

Temuan ini konsisten dengan beberapa penelitian sebelumnya yang dilakukan (Winoto et al., 2024) juga menunjukkan bahwa *Naive Bayes* unggul dibanding *KNN* dalam menganalisis aplikasi dengan struktur kalimat kompleks (Duolingo). (Limbong et al., 2022) menemukan bahwa *KNN* lebih unggul pada data *Shopee*, yang juga diamini dalam penelitian ini dalam konteks keseimbangan *precision* dan *recall*. (Safira & Hasan, 2023) yang hanya menggunakan *Naive Bayes* untuk analisis fitur Paylater menunjukkan performa tinggi, tetapi dengan keterbatasan pada cakupan fitur. Penelitian ini memperluas temuan mereka dengan menggunakan dataset yang besar, seimbang, dan berasal dari dua aplikasi berbeda, sehingga menyajikan perbandingan head-to-head yang lebih valid dan generalis.

4. KESIMPULAN

Berdasarkan hasil analisis dan evaluasi terhadap 40.000 data ulasan pengguna dari aplikasi *Shopee* dan *Lazada*, penelitian ini menyimpulkan bahwa algoritma *Naive Bayes* dan *K-Nearest Neighbor (KNN)* memiliki kelebihan masing-masing dalam klasifikasi sentimen. *Naive Bayes* menunjukkan performa yang lebih stabil dan akurat pada kedua aplikasi, terutama dalam menangani distribusi data yang tidak seimbang, dengan hasil terbaik pada aplikasi *Lazada* (akurasi 82,53% dan *F1-score* 79,89%). Sementara itu, *KNN*

menghasilkan kinerja yang sangat kompetitif pada aplikasi *Shopee* (akurasi 84,17% dan *F1-score* 82,31%), serta menunjukkan keunggulan dalam keseimbangan *precision* dan *recall* yang membantu klasifikasi sentimen secara proporsional. Hasil *confusion matrik* juga memperlihatkan bahwa *Naive Bayes* lebih efektif dalam mengklasifikasikan sentimen negatif dan positif secara konsisten, meskipun kurang dalam mengenali sentimen netral pada *Lazada*. Sebaliknya, *KNN* masih dapat mengenali sebagian kelas netral pada *Shopee* meskipun performa keseluruhannya lebih rendah pada *Lazada*. Penggunaan *cosine similarity* terbukti relevan dalam meningkatkan hasil klasifikasi *KNN* pada data berbasis teks. Dengan demikian, penelitian ini menjawab permasalahan klasifikasi sentimen terhadap ulasan pengguna *e-commerce* dengan menunjukkan bahwa pemilihan algoritma yang tepat harus mempertimbangkan jenis data, distribusi kelas, serta konteks aplikasi yang dianalisis. Temuan ini diharapkan menjadi rujukan untuk pengembangan sistem analisis opini publik berbasis pembelajaran mesin di Indonesia, khususnya dalam ranah *e-commerce*.

REFERENCES

- Carina, F. M., Salma, A., Permana, D., & Martha, Z. (2024). Sentiment Analysis of X Application Users on the Conflict Between Israel and Palestine Using Support Vector Machine Algorithm Fadhillah. *UNP JOURNAL OF STATISTICS AND DATA SCIENCE*, 2, 204–212.
- Firdaus, M. R., Rahaningsih, N., & Dana, R. D. (2024). Analisis Sentimen Aplikasi Shopee di Goole Play Store Menggunakan Klasifikasi Algoritma Naïve Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 6(1).
- Kusuma, I. H., & Cahyono, N. (2023). Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 8(3), 302–307.
- Lazada Group. (2025). *Lazada*.
https://group.lazada.com/en/about/?spm=a2o4j.homepage.footer_top.6.121653e0zm02Ed
- Limbong, J. J. A., Sembiring, I., & Hartomo, K. D. (2022). Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes dan K-Nearest Neighbor. *JTIK*, 9(2), 347–356. <https://doi.org/10.25126/jtiik.202294960>
- Muslim, S. N. S., Nurdiansyah, F., & Rahman, A. Y. (2024). Perbandingan Algoritma Naive Bayes dan KNN Dalam Analisis Sentimen Ulasan Pengguna Aplikasi Capcut. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3S1). <https://doi.org/10.23960/jitet.v12i3S1.5156>
- Nazar, R. (2024). Implementasi Pemrograman Python Menggunakan Google Colab. *Jurnal Informatika Dan Komputer (JIK)*, 15(1), 50–56.
- Paradisona, R., & Devega, M. (2024). Analisis Sentimen Terhadap Perbandingan Kepuasan Penggunaan Aplikasi Marketplace Menggunakan Metode Naïve Bayes Classifier Di Playstore. *Prosiding-Seminar Nasional Teknologi Informasi & Ilmu Komputer (SEMATER)*, 3(1), 54–66.
- Putra, A., & Dwiki, A. (2021). Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(2), 636–646.
- Rofiq, N., & Sitio, S. L. M. (2021). *Pengenalan Dasar Analisis Data Dengan Python di Google Collab*. 6.
- Safira, A., & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *ZONAsi: Jurnal Sistem Informasi*, 5(1), 59–70. <https://doi.org/10.31849/zn.v5i1.12856>
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko OnlineE Jd.IdMenggunakn Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686.

- Sea, C. F. P. (2025). *Shopee Careers*. Shopee. <https://careers.shopee.co.id/about>
- Wahyuni, N., Mutaqin, A. I. S., & Gunawan, A. (2019). Pengenalan Dan Pemanfaatan Marketplace E-Commerce Untuk Pelaku UKM Wilayah Cilegon. *Jurnal Pengabdian Dinamika*, 1(6), 31–39. <https://jurnal.untirta.ac.id/index.php/Dinamika/article/viewFile/8758/5837>
- Winoto, D., Aditia, V. D., Sorisa, C., Priskila, R., & Pranatawijaya, V. H. (2024). Analisis Sentimen Pada Ulasan Pengguna Terhadap Aplikasi Pembelajaran Bahasa Duologi: Menggunakan Algoritma Naive Bayes dan K-Nearest Neighbor. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 3).
- Zaitsev, S. (2020). *google-play-scraper*. <https://pypi.org/project/google-play-scraper/>