

Analisis Sentimen Ulasan Di *Playstore* Terhadap Aplikasi *Marketplace* Menggunakan *Naive Bayes* dan *Logistic Regression*

Aji Lukito¹, Fajar Nugraha², Noor Latifah³

^{1,2,3)} Program Studi Sistem Informasi, Universitas Muria Kudus

Jl. Lingkar UtaraUMK,Gondangmanis,Bae,Kudus–59327JawaTengah-Indonesia

Email: 202153050@std.umk.ac.id

ABSTRAK

Analisis Sentimen Ulasan di *Playstore* terhadap Aplikasi *Marketplace* (Shopee, Tokopedia, dan Lazada) Menggunakan *Naive bayes* dan *Logistic regression* bertujuan untuk mengklasifikasikan sentimen pengguna menjadi positif atau negatif berdasarkan ulasan terhadap tiga aplikasi *Marketplace* populer di Indonesia. Permasalahan utama dalam penelitian ini adalah bagaimana mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi *Marketplace* Indonesia di *Google Playstore* secara akurat menggunakan algoritma *Naive bayes* dan *Logistic regression*, serta membandingkan performa kedua algoritma dalam melakukan klasifikasi tersebut. Data dikumpulkan dari ulasan *Playstore*, kemudian diproses melalui tahapan *tokenization*, *case folding*, *stopword removal*, dan *stemming*. Selanjutnya, dilakukan klasifikasi menggunakan kedua algoritma, lalu hasilnya dibandingkan. Evaluasi dilakukan dengan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score* untuk menilai efektivitas masing-masing algoritma terhadap data yang tidak seimbang. Hasil penelitian ini diharapkan dapat menunjukkan algoritma yang lebih optimal untuk analisis sentimen, serta mendukung pengembangan strategi layanan *e-commerce* berbasis data secara lebih akurat dan efisien.

Kata kunci: analisis sentimen, *naive bayes*, *logistic regression*, *machine learning*, *marketplace*.

ABSTRACT

Sentiment Analysis of Reviews on the Play Store for Marketplace Applications (Shopee, Tokopedia, and Lazada) Using Naive bayes and Logistic regression aims to classify user sentiment as positive or negative based on reviews of three popular marketplace applications in Indonesia. The main problem in this study is how to accurately classify user sentiment toward Indonesian Marketplace apps on Google Playstore using the Naive bayes and Logistic regression algorithms, as well as comparing the performance of the two algorithms in performing this classification. Data was collected from Playstore reviews, then processed through the stages of tokenization, case folding, stopword removal, and stemming. Next, classification was performed using both algorithms, and the results were compared. Evaluation was conducted using the Accuracy, Precision, Recall, and F1-score metrics to assess the effectiveness of each algorithm on imbalanced data. The results of this study are expected to identify the more optimal algorithm for sentiment analysis and support the development of data-driven e-commerce service strategies in a more accurate and efficient manner.

Keywords: *sentimen analysis*, *naive bayes*, *logistic regression*, *machine learning*, *marketplace*

Pendahuluan

Shopee, Tokopedia, dan Lazada merupakan tiga aplikasi *Marketplace* terkenal di Indonesia yang memberikan layanan belanja online secara mobile. Ketiganya hadir untuk memenuhi kebutuhan masyarakat terhadap kemudahan bertransaksi secara digital, mulai dari pembelian produk kebutuhan sehari-hari hingga produk elektronik dan gaya hidup. Seiring meningkatnya penggunaan *smartphone* dan kemajuan teknologi digital, aplikasi *Marketplace* ini menjadi bagian penting dari gaya hidup konsumen modern di Indonesia. Melalui aplikasi *mobile* yang dirancang secara intuitif, pengguna dapat dengan mudah melakukan pencarian produk, bertransaksi, hingga memberikan penilaian dan ulasan terhadap layanan yang diterima.

Namun, dalam proses penggunaan aplikasi tersebut, penting bagi penyedia layanan untuk secara aktif memperhatikan umpan balik (*Feedback*) dari pengguna guna meningkatkan kualitas layanan dan kepuasan pelanggan. *Feedback* ini dapat berupa puji, saran, maupun keluhan, yang semuanya tercermin dalam bentuk ulasan pengguna di *platform* seperti *Google Playstore*. Ulasan-ulasan ini memegang peranan penting dalam proses evaluasi aplikasi karena berisi persepsi nyata pengguna terhadap aspek teknis aplikasi, kemudahan penggunaan, kecepatan layanan pengiriman, kualitas barang, hingga sistem promosi dan pembayaran (Tamami dkk., 2025).

Data ulasan pengguna yang tersedia dalam bentuk teks tidak terstruktur ini sebenarnya menyimpan informasi yang sangat berharga, namun sulit dianalisis secara manual dalam jumlah besar (Tjuet dkk., 2025). Oleh karena itu, pemanfaatan teknologi seperti *Machine learning* menjadi sangat penting untuk membantu mengolah dan menganalisis data ulasan secara sistematis. Salah satu pendekatan yang efektif dalam hal ini adalah analisis sentimen, yaitu proses klasifikasi opini pengguna menjadi kategori positif dan negatif, berdasarkan isi teks ulasan mereka.

Dalam penelitian analisis sentimen aplikasi media sosial, (Friska Aditia Indriyani dkk., 2023) meneliti opini pengguna terhadap aplikasi TikTok menggunakan algoritma *Naive Bayes* dan *SVM*. Mereka memperoleh hasil bahwa *SVM* memberikan *Accuracy* lebih tinggi (84%) dibandingkan *Naive Bayes* (79%). Hal ini mendukung temuan bahwa meskipun *Naive Bayes* lebih sederhana dan cepat, *SVM* mampu memberikan performa lebih baik pada *dataset* dengan fitur yang kompleks

(Lidinillah dkk., 2023) juga melakukan studi serupa terhadap ulasan di *platform* Steam menggunakan data dari *Twitte*. Mereka menggunakan *Logistic Regression* dan *SVM* sebagai algoritma pembanding. Dari hasil yang diperoleh, *SVM* mencatat performa terbaik dengan *Accuracy* sebesar 81%, *Precision* 85%, dan *Recall* 77%, yang menunjukkan kemampuan *SVM* dalam menangani data opini dari media sosial yang cenderung tidak terstruktur dan bervariasi

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk menerapkan dan membandingkan kinerja algoritma *Naive bayes* dan *Logistic regression* dalam mengklasifikasikan sentimen pengguna terhadap aplikasi Shopee, Tokopedia, dan Lazada berdasarkan ulasan di *Google Playstore*. *Naive bayes* merupakan metode statistik sederhana namun efisien dan cepat dalam

mengklasifikasikan teks. Sementara itu, *Logistic regression* memiliki kemampuan yang baik dalam menangani hubungan antar fitur dan menghasilkan prediksi yang stabil pada data ulasan yang beragam.

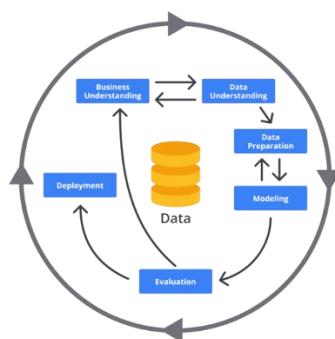
Dengan pendekatan ini, diharapkan hasil analisis dapat memberikan gambaran yang objektif dan akurat mengenai persepsi pengguna terhadap masing-masing *platform*, yang selanjutnya dapat dimanfaatkan oleh pengembang aplikasi untuk melakukan perbaikan, meningkatkan kepuasan pengguna, dan menyusun strategi layanan digital yang lebih efektif dan berkelanjutan

Metode Penelitian

Penelitian ini menggunakan pendekatan analisis sentimen untuk mengklasifikasikan opini pengguna terhadap tiga aplikasi *Marketplace* terpopuler di Indonesia, yaitu Shopee, Tokopedia, dan Lazada. Data yang digunakan diambil dari ulasan pengguna di *Google Playstore*. Penelitian ini dilakukan dengan metode *Machine Learning* yaitu *Naive Bayes* dan *Logistic Regression* yang digunakan untuk mengklasifikasikan sentimen menjadi positif atau negatif.

Kerangka Kerja

Penelitian ini mengadopsi kerangka kerja *CRISP-DM* (*Cross Industry Standard Process for Data Mining*). *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) merupakan kerangka kerja standar yang digunakan secara luas dalam proses Data Mining (Yudiana dkk., 2023) yang terdiri dari enam tahapan utama, dapat dilihat pada gambar 2



Gambar 1 Kerangka kerja CRISP-DM

Berdasarkan gambar 1 diatas *CRISP-DM* terdiri dari enam tahapan yaitu

1) *Bussines Understanding*

Tahap ini adalah memahami kebutuhan pelanggan secara mendalam. Kegiatan yang dilakukan pada tahap ini adalah menentukan tujuan bisnis, menilai situasi ketersediaan sumber daya, menentukan tujuan pengumpulan data, dan menghasilkan rencana proyek (Singgalen, 2023).

2) *Data Understanding*

Tahap pemahaman data, yaitu mengidentifikasi, mengumpulkan, dan menganalisis kumpulan data yang dapat membantu untuk mencapai tujuan proyek.

Kegiatan pada tahap ini adalah mengumpulkan data awal, menjelaskan data, menjelajahi data, dan memverifikasi kualitas data.

3) *Data Preparation*

Fase ini sering disebut “*data mining*”, yaitu menyiapkan kumpulan data akhir untuk pemodelan. Kegiatan pada fase ini adalah memperbaiki kualitas data agar sesuai dengan proses modeling yang akan dilakukan berikutnya.

4) *Modeling*

Membuat dan menilai berbagai model berdasarkan beberapa teknik pemodelan yang berbeda. Pada tahap ini, ada empat tugas, yaitu memilih teknik pemodelan, menghasilkan desain pengujian, membangun model, dan yang terakhir menilai model. Dalam penelitian akan menggunakan Naive Bayes dan Logistic Regression

5) *Evaluation*

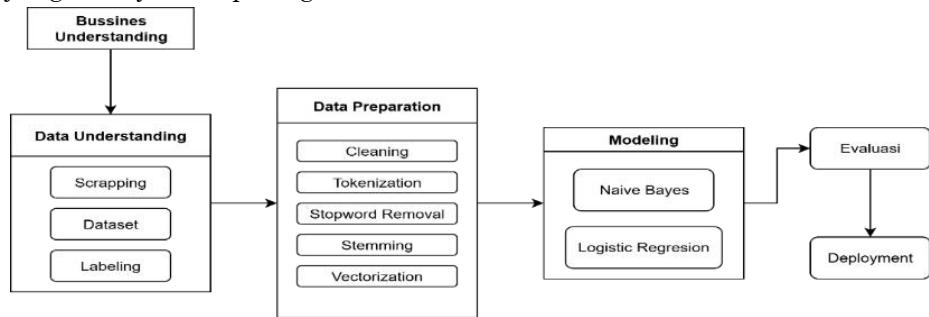
Evaluasi dilakukan terhadap hasil klasifikasi dari dua algoritma dengan mempertimbangkan beberapa metrik, yaitu ketepatan klasifikasi (*Accuracy*), keseimbangan antara positif dan negatif (*Precision* dan *Recall*), serta efektivitas menyeluruh yang diukur menggunakan *F1-score*. Selain itu, analisis juga mencakup identifikasi platform yang memiliki dominasi sentimen positif dan negatif terbanyak di antara Shopee, Tokopedia, dan Lazada.

6) *Deployment*

Tahap terakhir dan paling penting dari proses CRISP-DM. Tahap *deployment* dilakukan dengan mendokumentasikan hasil evaluasi model ke dalam bentuk laporan visual dan analitik yang dapat dimanfaatkan oleh pengembang aplikasi dan tim bisnis. Visualisasi seperti *confusion matrix*, pie chart, dan *wordcloud* disusun dalam laporan untuk memberikan gambaran mengenai persepsi pengguna terhadap aplikasi *Marketplace*. Dokumentasi ini mencakup seluruh hasil evaluasi performa model, baik dari algoritma *Naive bayes* maupun *Logistic regression*, yang disusun secara komprehensif agar dapat digunakan sebagai dasar pengambilan keputusan oleh pengembang aplikasi, analis data, maupun tim bisnis dalam konteks pengembangan dan peningkatan layanan aplikasi *marketplace*.

Metode

Dibawah ini merupakan alur metodologi berdasarkan kerangka kerja CRISP-DM yang ditunjukkan pada gambar 3 berikut.



Gambar 2 Alur penelitian

Berdasarkan gambar 2 diatas, metodologi ini akan mencakup tahapan - tahapan penting mulai dari pengumpulan data, pemrosesan, hingga evaluasi hasil. Berikut merupakan pembahasan lebih rinci tentang metodologi penelitian tersebut, yaitu :

a. Analisis Kebutuhan (Business Understanding)

Shopee, Tokopedia, dan Lazada adalah aplikasi marketplace populer di Indonesia yang bersaing ketat dalam menarik dan mempertahankan pengguna. Namun, penilaian mana yang terbaik sering kali subjektif dan tidak cukup hanya berdasarkan rating. Oleh karena itu, analisis sentimen dari ulasan pengguna di Google Playstore menjadi penting untuk memahami persepsi dan kepuasan pengguna secara lebih objektif dengan tujuan untuk membantu memahami opini pengguna terhadap masing-masing aplikasi dengan cara mengklasifikasikan ulasan menjadi sentimen positif dan negatif menggunakan algoritma *Naive Bayes* dan *Logistic Regression*.

Dalam hal ini, penulis menggunakan dua teknik pengumpulan data seperti berikut:

- 1) Data Teks Ulasan : Berupa teks atau komentar dari pengguna aplikasi Shopee, Tokopedia, dan Lazada di *Google Playstore*
- 2) Data Sentimen : Sentimen diklasifikasikan menjadi dua kelas utama: positif dan negatif, berdasarkan isi ulasan (label dapat dikembangkan dari *rating* atau anotasi manual)

Dua algoritma klasifikasi teks yang digunakan dalam penelitian ini adalah sebagai berikut :

- 1) *Naive Bayes* : Metode Naïve Bayes merupakan proses pengklasifikasian atau pengujian suatu teks baru yang belum pernah diketahui kategorinya berdasarkan basis pengetahuan yang dimiliki (Nugroho & , Yulison Herry Chrisnanto, 2016)

$$P(Y|X) = \frac{P(X|Y)*P(Y)}{P(X)} \quad (1)$$

- 2) *Logistic Regression* : Metode yang memiliki keunggulan dalam menangani relasi linier antar fitur dan memberikan prediksi yang relatif lebih stabil untuk jenis data teks yang beragam, metode ini dapat dimanfaatkan untuk mengklasifikasikan suatu teks ke dalam sentimen positif atau negatif, dengan melihat pola dan karakteristik kata-kata yang terkandung dalam teks tersebut. (Toyibah dkk., 2024)

$$\ln\left(\frac{p}{1-p}\right) = B_0 + B_1 X \quad (2)$$

b. Pengumpulan Data (*Gathering Dataset*)

Data dalam penelitian ini diperoleh melalui metode web *scrapping* menggunakan bahasa pemrograman *Python* dengan bantuan *library* seperti *google-play-scraping* untuk mengakses ulasan dari *Google Playstore*. Data yang dikumpulkan mencakup beberapa elemen penting, yaitu nama aplikasi (Shopee, Tokopedia, atau Lazada), isi ulasan pengguna, *rating* dalam skala 1 hingga 5 bintang, tanggal ulasan, serta *username* pengguna (jika tersedia). Setelah proses pengumpulan selesai, data akan difilter untuk menyisakan hanya ulasan yang ditulis dalam Bahasa Indonesia agar lebih relevan dengan analisis.

c. *Labeling*

Pada tahap ini, data yang telah diperoleh akan diberi label sesuai dengan skor ulasan dalam rentang 1 sampai 5 menggunakan *Python*, dan dibagi ke dalam dua kategori sentimen, yaitu :

- Negatif untuk skor ulasan 1 dan 2
- Positif untuk skor ulasan 3,4 dan 5

Pelabelan ini bertujuan untuk membuat suatu target variabel *sentiment*, yang nantinya akan digunakan pada model klasifikasi

d. *Preprocessing* Data

Untuk mempersiapkan data ulasan agar dapat diproses secara optimal oleh algoritma klasifikasi, dilakukan tahap *preprocessing* teks yang bertujuan untuk

membersihkan, merapikan, dan mengubah data teks menjadi bentuk yang dapat dikenali oleh mesin. Tahapan ini terdiri dari beberapa proses berikut:

- 1) *Cleaning* : pembersihan teks dari elemen-elemen yang tidak relevan atau mengganggu analisis seperti angka, tanda baca, emoji, URL, dan karakter khusus seperti simbol.
- 2) *Tokenization* : memecah teks utuh menjadi satuan kata atau token. Sebagai contoh, kalimat "Pengirimannya sangat cepat" akan dipecah menjadi ["pengirimannya", "sangat", "cepat"].
- 3) *Stopword Removal* : penghapusan kata-kata umum yang tidak memiliki makna penting dalam konteks analisis sentimen, seperti "dan", "di", "yang", dan sebagainya.
- 4) *Stemming* : proses mengubah kata ke bentuk dasarnya. Misalnya, kata "berbelanja", "belanjaan", dan "membelanjakan" akan dikembalikan ke bentuk dasar "belanja".
- 5) *Vectorization* : mengubah data teks menjadi representasi numerik yang dapat digunakan oleh algoritma klasifikasi.

e. Klasifikasi dan Analisis Data

Setelah melalui tahap *preprocessing*, data ulasan dibagi menjadi dua bagian, yaitu data *training* dan data *testing* dengan proporsi umum sebesar 80% untuk *training* dan 20% untuk *testing*. Data *training* digunakan untuk melatih model agar dapat mempelajari pola dari data, sedangkan data *testing* digunakan untuk mengukur seberapa baik model tersebut dapat menggeneralisasi prediksi terhadap data baru yang belum pernah dilihat sebelumnya. Pada tahap klasifikasi, dua algoritma yang digunakan adalah *Naive Bayes* dan *Logistic Regression*. Hasil klasifikasi dari masing-masing algoritma dibandingkan berdasarkan empat metrik evaluasi, yaitu:

- *Accuracy* : Mengukur seberapa banyak prediksi model yang benar dibandingkan seluruh data. berikut adalah rumus perhitungannya

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

- *Precision* : Mengukur seberapa tepat model dalam memprediksi data positif.berikut adalah rumus perhitungannya

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

- *Recall* : Mengukur seberapa baik model menemukan seluruh data positif.berikut adalah rumus perhitungannya

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

- *F1-Score* : Merupakan rata-rata harmonis dari *Precision* dan *Recall*.berikut adalah rumus perhitungannya

$$F1 Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Dengan menggunakan keempat metrik ini, performa masing-masing algoritma dapat dievaluasi secara menyeluruh dan objektif.

Hasil dan Pembahasan

Pengumpulan Dataset

Dataset yang digunakan dalam diperoleh melalui proses *crawling* atau pengambilan data secara otomatis dari ulasan pengguna yang terdapat di *Google Playstore* pada tiga aplikasi *marketplace*, yaitu Shopee, Tokopedia, dan Lazada. Pada tabel 1 ditunjukkan total data yang terkumpul dari masing-masing aplikasi.

Tabel 1. Jumlah data dari masing-masing aplikasi

Aplikasi	Jumlah ulasan
Shopee	3000
Tokopedia	3000
Lazada	3000

Jumlah 3.000 ulasan per aplikasi dipilih untuk menjaga keseimbangan jumlah data antar kelas dan antar aplikasi, sekaligus memastikan bahwa volume data cukup besar untuk melatih model *machine learning* dengan baik, namun tetap terkendali dengan efisien dalam proses *preprocessing* dan analisis.

Hasil *Data Preparation*

Proses pengolahan data mentah hasil *scraping* diubah menjadi kumpulan data yang tersusun rapi dan terstruktur, sehingga siap digunakan pada tahap pemodelan.

Proses *Cleaning*

Pada proses ini, dilakukan penghapusan atribut atau kolom data yang dianggap tidak relevan dengan tujuan analisis, baik karena tidak memiliki keterkaitan langsung dengan variabel penelitian maupun karena tidak memberikan kontribusi informasi yang berarti.

Hasil dari proses *cleaning* pada ketiga aplikasi *marketplace* yang ditunjukkan pada gambar 3 menunjukkan bagaimana teks ulasan yang semula berisi berbagai karakter, tanda baca serta variasi penulisan kata, telah berhasil dibersihkan menjadi bentuk yang lebih terstruktur dan mudah dibaca.

	content	score	cleaned_text	label
0	semuanya sih baik baik saja, cuman kalau bisa ...	5	semuanya sih baik baik saja cuman kalau bisa ...	positif
1	Aplikasi ga jelas, pesanan gagal kirim karena ...	1	aplikasi ga jelas pesanan gagal kirim karena ...	negatif
2	Akhir2 ini shopee saya bermasalah, tiap buka s...	1	akhir ini shopee saya bermasalah tiap buka s...	negatif
3	kecewa banget,pengguna lama,tapi tidak bisa me...	1	kecewa banget pengguna lama tapi tidak bisa me...	negatif
4	Dari aplikasi shopee saya sangat suka ... cuma...	5	dari aplikasi shopee saya sangat suka cuma...	positif

Gambar 3. Hasil *cleaning* aplikasi shopee

Gambar 3 memperlihatkan perbandingan antara teks asli dan hasil *cleaning*, sehingga dapat terlihat secara jelas perbedaan signifikan sebelum dan sesudah pemrosesan.

Proses *Tokenization*

Proses ini dapat diibaratkan seperti memotong kalimat panjang menjadi potongan-potongan kecil sehingga setiap kata dapat dianalisis secara terpisah. Gambar 4 menunjukkan hasil *tokenize* dari 3 aplikasi yang memperlihatkan

bagaimana sebuah teks utuh dipecah menjadi potongan-potongan kata yang siap diproses lebih lanjut dalam tahap analisis sentimen

	cleaned_text	score	tokens	label
0	semuanya sih baik baik saja cuman kalau bisa ...	5	[semuanya, sih, baik, baik, saja, cuman, kalau...	positif
1	aplikasi ga jelas pesanan gagal kirim karena ...	1	[aplikasi, ga, jelas, pesanan, gagal, kirim, k...	negatif
2	akhir ini shopee saya bermasalah tiap buka s...	1	[akhir, ini, shopee, saya, bermasalah, tiap, b...	negatif
3	kecewa banget pengguna lama tapi tidak bisa me...	1	[kecewa, banget, pengguna, lama, tapi, tidak, ...	negatif
4	dari aplikasi shopee saya sangat suka cuma...	5	[dari, aplikasi, shopee, saya, sangat, suka, c...	positif

Gambar 4. Hasil proses *tokenization*

Proses ini membantu menghilangkan ambiguitas dalam pemrosesan teks karena setiap kata diperlakukan sebagai satu entitas mandiri.

Proses *Stopword removal*

Pada tahap ini dilakukan proses penghapusan kata-kata yang frekuensinya sangat tinggi dalam teks, tetapi tidak memberikan kontribusi berarti terhadap penentuan sentimen suatu kalimat. Kata-kata seperti "yang", "di", "dan", "ke", atau imbuhan seperti "nya" termasuk ke dalam kategori ini proses *stopword removal* dilakukan dengan memanfaatkan dictionary yang berisi kumpulan (*corpus*) kata-kata stopword dalam bahasa Indonesia. Daftar ini digunakan sebagai acuan untuk mengenali dan menghapus kata-kata yang dianggap tidak relevan

Hasil penerapan *stopword removal* pada dataset ketiga aplikasi yang ditunjukkan Gambar 5 memperlihatkan bagaimana kata-kata umum dihapus, menyisakan kata-kata inti yang merepresentasikan opini pengguna secara lebih jelas

	cleaned_text	score	stopword	label
0	semuanya sih baik baik saja cuman kalau bisa ...	5	semuanya sih baik baik cuman kalau mohon selek...	positif
1	aplikasi ga jelas pesanan gagal kirim karena ...	1	aplikasi ga jelas pesanan gagal kirim lama ban...	negatif
2	akhir ini shopee saya bermasalah tiap buka s...	1	akhir shopee bermasalah tiap buka shoppe langs...	negatif
3	kecewa banget pengguna lama tapi tidak bisa me...	1	kecewa banget pengguna lama mencoba fitur payl...	negatif
4	dari aplikasi shopee saya sangat suka cuma...	5	aplikasi shopee sangat suka cuma dri jasa peng...	positif

Gambar 5. Hasil *stopword removal* aplikasi shopee

Proses *Stemming*

Pada tahap ini dilakukan proses mengubah kata yang sudah mengalami perubahan bentuk menjadi bentuk dasarnya. Gambar 16 menunjukkan proses *stemming* dilakukan menggunakan *library Sastrawi*, yang merupakan salah satu *stemmer* bahasa Indonesia yang sudah teruji keakuratannya. Sastrawi bekerja dengan mengidentifikasi pola imbuhan dalam bahasa Indonesia, kemudian menghapus imbuhan tersebut untuk mendapatkan kata dasar yang benar.

Hasil *stemming* pada dataset penelitian ini ditunjukkan pada gambar 6, di mana kata-kata yang memiliki variasi bentuk dikonversi menjadi bentuk dasar yang konsisten

	cleaned_text	score	stemmed	label
0	semuanya sih baik baik saja cuman kalau bisa ...	5	semua sih baik baik cuman kalau mohon seleksi ...	positif
1	aplikasi ga jelas pesanan gagal kirim karena ...	1	aplikasi ga jelas pesan gagal kirim lama bange...	negatif
2	akhir ini shopee saya bermasalah tiap buka s...	1	akhir shopee masalah tiap buka shoppe langsung...	negatif
3	kecewa banget pengguna lama tapi tidak bisa me...	1	kecewa banget guna lama coba fitur paylatter p...	negatif
4	dari aplikasi shopee saya sangat suka cuma...	5	aplikasi shopee sangat suka cuma dri jasa kiri...	positif

Gambar 6. Hasil *stemming* aplikasi shopee

Data *Splitting* dan *Vectorization*

Pembagian set data ke dalam data latih dan data uji menggunakan perbandingan 80:20 yang ditunjukkan pada gambar 20, yaitu 80% dari jumlah data menjadi data latih dan 20% sebagai data uji. Pada tabel 2 ditunjukkan Hasil data *splitting* dari 3 aplikasi, yaitu shopee, tokopedia, lazada

Tabel 2 Hasil data *splitting*

	Jumlah data latih	Jumlah Data Uji
Shopee	2,400	600
Tokopedia	2,400	600
Lazada	2,400	600

Setelah melalui tahapan *cleaning*, *tokenization*, *stopword removal*, dan *stemming*, dan data *splitting* data teks ulasan yang semula berupa kumpulan kata bebas perlu diubah menjadi bentuk yang dapat diolah oleh algoritma *machine learning*. Proses ini dikenal sebagai *vectorization*, yaitu mengubah teks menjadi representasi numerik. Pada penelitian ini, teknik yang digunakan adalah *Term Frequency–Inverse Document Frequency* (TF-IDF).

TF-IDF akan memberikan bobot tinggi pada kata yang jarang muncul di seluruh dokumen tetapi sering muncul di dokumen tertentu. Parameter yang digunakan dalam penelitian ini:

- $\text{max_df} = 0.5$ → mengabaikan kata yang muncul di lebih dari 50% dokumen.
- $\text{min_df} = 2$ → mengabaikan kata yang muncul di kurang dari 2 dokumen.
- $\text{ngram_range} = (1,1)$ → hanya unigram digunakan.

Untuk mendapatkan gambaran kata yang paling informatif, dilakukan perhitungan rata-rata bobot TF-IDF dari seluruh dokumen pada data latih, kemudian diurutkan dari yang tertinggi.

Pada tahap pemodelan, digunakan dua algoritma pembelajaran mesin untuk melakukan klasifikasi sentimen ulasan pengguna, yaitu *Naive Bayes* dan *Logistic regression*

Naive Bayes

Pada tabel 3 ditunjukkan hasil pengujian model *Naive Bayes* pada data uji dengan menggunakan dataset shopee, yang dilihat dari nilai *precision*, *recall*, dan *f1-score* untuk masing-masing kelas, serta akurasi keseluruhan yang diperoleh, sehingga memberikan gambaran yang jelas mengenai tingkat efektivitas model dalam melakukan analisis sentimen.

Tabel 3 Hasil pemodelan Naive Bayes

	precision	recall	F1-score	support
negatif	0.64	0.88	0.74	256
positif	0.88	0.64	0.74	344
accuracy			0.74	600

Nilai Precision mengukur proporsi prediksi benar dari semua prediksi positif, Recall mengukur kemampuan model dalam menemukan semua data positif, sedangkan *F1-score* adalah harmonisasi antara *precision* dan *recall*.

Berdasarkan hasil pengujian model *Naive bayes* yang ditunjukkan pada tabel, diperoleh nilai akurasi keseluruhan sebesar 74%, yang berarti dari total 600 data uji, sekitar 74% di antaranya berhasil diklasifikasikan dengan benar oleh model

Logistic Regression

Hasil pemodelan *logistic regression* ditunjukkan pada Tabel 4 menggunakan dataset yang sama yaitu shopee.

Tabel 4. Hasil pemodelan *logistic regression*

	<i>precision</i>	<i>recall</i>	<i>F1-score</i>	<i>support</i>
negatif	0.67	0.75	0.71	256
positif	0.79	0.72	0.76	344
<i>accuracy</i>			0.73	600

Berdasarkan hasil yang ditunjukkan pada tabel, model *Logistic regression* mencapai tingkat akurasi sebesar 0.73, yang berarti 73% dari seluruh prediksi yang dibuat model sesuai dengan label sebenarnya di data uji. Angka ini menunjukkan bahwa model cukup handal dalam membedakan ulasan positif dan negatif meskipun masih ada ruang untuk perbaikan.

Evaluasi atas implementasi model *machine learning* pada algoritma *naive bayes* dan *logistic regression* menghasilkan algoritma yang cukup lumayan baik dengan nilai *accuracy*, *precision*, *recall*, dan *f1-score* optimal dalam klasifikasi sentimen shopee, tokopedia, lazada. Tabel 5 menampilkan hasil evaluasi implementasi algoritma *machine learning* terhadap set data shopee.

Tabel 5 Perbandingan hasil pemodelan

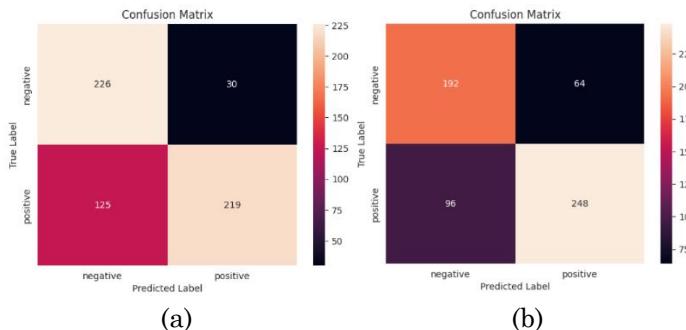
	<i>accuracy</i>	<i>precision</i>	<i>recall</i>	<i>F1-score</i>
<i>Naive bayes</i>	0.74	0.88	0.64	0.74
<i>Logistic regression</i>	0.73	0.79	0.72	0.76

Berdasarkan hasil evaluasi, algoritma *Naive bayes* memperoleh nilai akurasi sebesar 0,74, yang berarti 74% prediksi sentimen pada data uji sesuai dengan label sebenarnya. Nilai *Precision* sebesar 0,88 menunjukkan bahwa dari seluruh ulasan yang diprediksi positif oleh model, 88% benar-benar positif. Sementara itu, nilai *Recall* sebesar 0,64 mengindikasikan bahwa model hanya mampu menemukan 64% dari seluruh ulasan positif yang ada di dataset. Nilai *F1-score* sebesar 0,74 menjadi ukuran keseimbangan antara *precision* dan *recall*, menunjukkan performa yang cukup baik namun cenderung lebih kuat dalam ketepatan prediksi dibanding kelengkapan deteksi.

Sedangkan pada algoritma *Logistic regression*, akurasi yang diperoleh adalah 0,73, sedikit lebih rendah dari *Naive bayes*. Nilai *Precision* sebesar 0,79 mengindikasikan bahwa 79% prediksi positif sesuai dengan label aslinya. *Recall* sebesar 0,72 menunjukkan kemampuan model dalam menemukan ulasan positif secara lebih seimbang dibandingkan *Naive bayes*. *F1-score* sebesar 0,76 menandakan

bahwa *Logistic regression* memiliki keseimbangan yang sedikit lebih baik antara ketepatan prediksi dan kelengkapan deteksi dibanding *Naive bayes*.

Gambar 7a memperlihatkan confusion matrix dari data uji pada dataset Shopee dengan algoritma Naive Bayes. Hasilnya menunjukkan bahwa terdapat 30 data dengan sentimen positif yang salah diprediksi sebagai sentimen negatif (false negative), serta 125 data dengan sentimen negatif yang keliru diklasifikasikan sebagai sentimen positif (false positive)



Gambar 7. Confusion matrix Naive Bayes (a), dan Confusion matrix *logistic regression* (b)

Pada Gambar 7b ditampilkan confusion matrix dari data uji pada dataset Shopee dengan algoritma logistic regression. Dari hasil tersebut, terdapat 64 data dengan sentimen positif yang keliru diprediksi sebagai sentimen negatif (false negative), serta 96 data dengan sentimen negatif yang salah diklasifikasikan sebagai sentimen positif (false positive)

Deployment

Pada tahap *deployment*, hasil analisis sentimen disajikan dalam bentuk visualisasi yang memungkinkan perbandingan antar-tiga aplikasi *marketplace* (Shopee, Tokopedia, dan Lazada) secara lebih jelas dan informatif. Dua jenis visualisasi utama yang digunakan adalah *wordcloud* dan grafik kata terbobot (TF-IDF).

Wordcloud perbandingan aplikasi

Wordcloud digunakan untuk menampilkan kata-kata yang paling sering muncul dalam ulasan positif dan negatif pada masing-masing aplikasi. Pada gambar 8 dan 9 Setiap aplikasi memiliki *wordcloud* terpisah untuk sentimen positif dan negatif, sehingga memudahkan identifikasi kata kunci yang dominan.



Gambar 8. Wordcloud Sentimen Positif



Gambar 9. Wordcloud Sentimen Negatif

Grafik Kata Terbobot

Visualisasi ini menunjukkan kata-kata dengan bobot TF-IDF tertinggi pada masing-masing aplikasi yang ditunjukkan pada Gambar 10. TF-IDF digunakan untuk mengidentifikasi kata yang paling informatif dan relevan dalam membedakan ulasan.



Gambar 10. Distribusi kelompok kata yang sering muncul pada ke-tiga aplikasi

Simpulan

Berdasarkan hasil penelitian, kedua algoritma memiliki performa yang cukup sebanding, di mana Naive Bayes lebih unggul pada precision (88%) sehingga cocok digunakan jika fokus pada ketepatan identifikasi ulasan positif, sedangkan Logistic Regression lebih baik pada recall (72%) dan F1-score (76%), sehingga sesuai untuk meminimalkan ulasan positif yang terlewat sekaligus menjaga keseimbangan performa. Hasil perbandingan ulasan menunjukkan bahwa Shopee dan Lazada unggul pada kualitas produk dengan kata “barang” dominan dalam sentimen positif, sedangkan Tokopedia lebih menonjol pada pengalaman bertransaksi dengan kata “toko” dan “belanja”. Namun, keluhan pengiriman banyak muncul di Shopee dan Lazada, sementara Lazada juga sering dikritik terkait kualitas produk dan proses penjualan. Secara keseluruhan, Tokopedia dapat dianggap sebagai platform terbaik karena memberikan pengalaman berbelanja yang lebih memuaskan dan minim keluhan dibandingkan dua platform lainnya.

Daftar Pustaka

- Friska Aditia Indriyani, Ahmad Fauzi, & Sutan Faisal. (2023). Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine. *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, 10(2), 176–184. <https://doi.org/10.37373/tekno.v10i2.419>
- Lidinillah, E. R., Rohana, T., & Juwita, A. R. (2023). Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine. *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, 10(2), 154–164. <https://doi.org/10.37373/tekno.v10i2.4400>
- Nugroho, D. G., & Yulison Herry Chrisnanto, A. W. (2016). Analisis Sentimen pada Jasa Ojek Online. *Prosiding Sains* ..., 156–161. https://publikasiilmiah.unwahas.ac.id/PROSIDING_SNST_FT/article/view/1526 https://publikasiilmiah.unwahas.ac.id/index.php/PROSIDING_SNST_FT/article/download/1526/1609
- Singgalen, Y. A. (2023). Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM.

- Building of Informatics, Technology and Science (BITS)*, 4(4).
<https://doi.org/10.47065/bits.v4i4.3199>
- Tamami, G., Triyanto, W. A., & Muzid, S. (2025). *Sentiment Analysis Mobile JKN Reviews Using SMOTE Based LSTM*. 19(1), 13–24.
- Tjuet, R., Fitri, Z., & Siegar, S. C. (2025). *Analisis Sentimen Komentar Pada Saluran Youtube Beauty Vlogger Berbahasa Indonesia Menggunakan Metode Support Vector Machine*. 5(2), 164–175. <https://doi.org/10.35957/algoritme.v5i2.9692>
- Toyibah, Z. B., Putri, Y. N., Puandini, P., Widodo, Z. M., & Ni'mah, A. T. (2024). Perbandingan Kinerja Algoritma Multinomial Naïve Bayes dan Logistic Regression pada Analisis Sentimen Movie Ratings IMDB. *Jurnal Ilmiah Edutic : Pendidikan dan Informatika*, 10(2), 181–189.
<https://doi.org/10.21107/edutic.v10i2.28150>
- Yudiana, Y., Yulia Agustina, A., & Nur Khofifah, dan. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 01–20. <http://ejournal.lp2m.uinjambi.ac.id/ojp/index.php/ijieb>