



Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes

Anggista Oktavia Praneswara¹, Nuri Cahyono^{2*}

anggistaop@students.amikom.ac.id, nuricahyono@amikom.ac.id

Universitas Amikom Yogyakarta

Informasi Artikel

Diterima : 23 Okt 2023

Direview : 4 Nov 2023

Disetujui : 1 Des 2023

Kata Kunci

analisis sentimen, naïve bayes, tiktok shop, google playstore

Abstrak

Dalam era digital yang berkembang pesat, pandangan pengguna terhadap aplikasi mobile menjadi faktor kunci dalam kesuksesan suatu aplikasi. Pemahaman terhadap sentimen pengguna dapat membantu pengembang dan manajemen aplikasi untuk meningkatkan kualitas layanan dan kepuasan pengguna. Salah satu media sosial yang mengalami revolusi adalah TikTok, merupakan platform berbagi video pendek dengan menghadirkan inovasi e-commerce melalui TikTok Shop Seller Center. Akan tetapi terdapat masalah yakni pemerintah telah mengambil langkah tegas dengan menghentikan operasi platform ini. Oleh karena itu, dilakukan analisis sentimen untuk mengetahui apakah respon dari ulasan pengguna aplikasi TikTok Shop Seller Center cenderung positif atau negatif berbasis algoritma Naive Bayes. Metodologi penelitian melibatkan tahap scrapping data, cleaning data, preprocessing (case folding, stopword removing, tokenization, stemming), labelling, TF-IDF, pengujian data menggunakan confusion matrix dan visualisasi menggunakan wordcloud. Hasil penelitian mengenai analisis sentimen ulasan aplikasi TikTok Shop Seller Center pada Google Playstore berjumlah 5000 data, disimpulkan bahwa ulasan pengguna tergolong negatif dengan presentase 87,6% nilai akurasi, 85,8% nilai precision, 95,4% nilai recall dan 90,3% nilai F1-Score.

Keywords

sentiment analysis, naïve bayes, tiktok shop, google playstore

Abstrak

In the rapidly developing digital era, users' views on mobile applications are a key factor in the success of an application. Understanding user sentiment can help application developers and management to improve service quality and user satisfaction. One of the social media that is experiencing a revolution is TikTok, a short video sharing platform that presents e-commerce innovations through the TikTok Shop Seller Center. However, there is a problem, namely that the government has taken firm steps by stopping the operation of this platform. Therefore, sentiment analysis was carried out to find out whether responses from user reviews of the TikTok Shop Seller Center application tended to be positive or negative based on the Naive Bayes algorithm. The research methodology involves data scrapping, data cleaning, preprocessing (case folding, stopword removing, tokenization, stemming), labeling, TF-IDF, data testing using confusion matrix and visualization using wordcloud. The results of research regarding sentiment analysis of reviews of the TikTok Shop Seller Center application on Google Playstore totaling 5000 data, it was concluded that user reviews were classified as negative with a percentage of 87.6% accuracy value, 85.8% precision value, 95.4% recall value and 90.3% % F1-Score value.

A. Pendahuluan

Di era sekarang ini, peradaban manusia bergerak sangat cepat, sejak adanya revolusi industri 4.0, manusia dihadapkan pada tuntutan untuk menjadi lebih adaptif dan responsif terhadap perkembangan teknologi[1]. Aplikasi seluler telah menjadi salah satu sarana utama untuk berinteraksi dengan berbagai layanan dan produk[2]. Saat ini, platform media sosial sedang sangat populer digunakan sebagai sarana komunikasi, sumber informasi, alat pendidikan, dan bahkan sebagai wadah pemasaran[3]. Salah satu media sosial yang telah mengalami revolusi adalah TikTok yang merupakan sebuah platform berbagi video pendek yang telah menjadi sangat populer di seluruh dunia. Dengan munculnya tren belanja online yang semakin meningkat, TikTok telah menghadirkan inovasi terbaru dalam dunia e-commerce dengan meluncurkan TikTok Shop Seller Center. Bukan hanya menjadi platform hiburan, tetapi juga tempat yang potensial bagi para penjual untuk mempromosikan dan menjual produk mereka.

TikTok Shop Seller Center adalah sebuah fitur layanan e-commerce yang terdapat di dalam aplikasi TikTok, memungkinkan pengguna untuk melakukan kegiatan jual-beli secara online di dalam aplikasi tersebut. Platform ini menyediakan berbagai alat dan fitur yang memudahkan penjual untuk mempromosikan produk, berinteraksi dengan pelanggan, dan mengelola operasi bisnis secara lebih efisien. TikTok Shop Seller Center memungkinkan penjual untuk membuat toko online yang menarik dengan fitur kreatif seperti video promosi. Akan tetapi terdapat masalah yakni pemerintah telah mengambil langkah tegas dengan menghentikan operasi platform ini. Penutupan TikTok Shop Seller Center oleh pemerintah bisa disebabkan oleh kurangnya izin resmi untuk beroperasi sebagai platform e-commerce. Keputusan ini telah memicu berbagai respons dan reaksi dari berbagai pihak tentang bagaimana respon pengguna terhadap penutupan TikTok Shop Seller Center.

Oleh karena itu, dibutuhkan suatu pendekatan analisis untuk menyimpulkan ulasan terhadap produk secara tepat dan efisien. Analisis sentimen adalah proses otomatis yang berfokus pada pemahaman, ekstraksi, dan pengolahan data teks dengan tujuan untuk mengidentifikasi sentimen yang terkandung dalam pernyataan atau kalimat opini[4]. Melalui analisis sentimen, informasi yang terdapat dalam teks berupa opini atau pandangan terhadap suatu isu atau peristiwa dapat diidentifikasi sebagai bersifat positif atau negatif. Dalam konteks ini, ulasan pengguna pada aplikasi menjadi sumber informasi yang sangat berharga untuk mengevaluasi kinerja aplikasi. Kepuasan pengguna terhadap aplikasi dapat tercermin dalam ulasan dan komentar yang mereka sampaikan di Play Store. Ulasan ini menjadi sumber bagi peneliti untuk menganalisis sentimen terhadap aplikasi. Ulasan teks adalah sebuah jenis teks yang melakukan analisis, penafsiran, atau penilaian terhadap suatu objek khusus yang digunakan sebagai patokan untuk menilai objek tersebut[5].

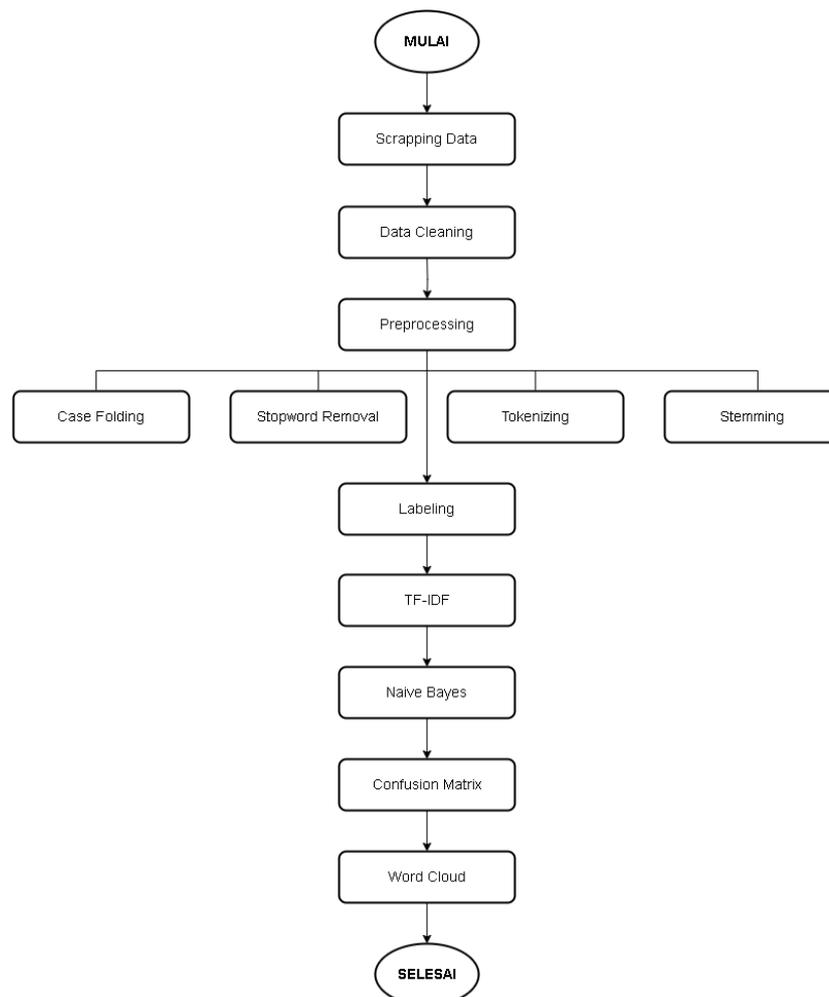
Penelitian ini akan berfokus pada analisis sentimen ulasan aplikasi TikTok Shop Seller Center di Google Play Store dengan menggunakan algoritma Naïve Bayes sebagai metodologi analisis sentimen. Tujuan utama dari penelitian ini adalah untuk menganalisis distribusi sentimen yang terkandung dalam ulasan pengguna terhadap aplikasi ini dan mengidentifikasi apakah ulasan tersebut cenderung bersifat positif atau negatif. Dalam penelitian ini, kami akan mengumpulkan ulasan

pengguna, melakukan tahap preprocessing data, TF-IDF dan mengembangkan model analisis sentimen berbasis algoritma Naïve Bayes. Algoritma Naïve Bayes akan digunakan untuk memprediksi ulasan-ulasan tersebut dan mengkategorikannya ke dalam dua kelompok, yaitu positif dan negatif[6]. Dengan demikian, penelitian ini akan memberikan pemahaman yang kuat tentang bagaimana pengguna secara keseluruhan merespons aplikasi TikTok Shop Seller Center dan mengetahui reputasi aplikasi yang positif atau negatif di kalangan pengguna.

Penelitian sangat penting karena dapat memberikan wawasan yang mendalam tentang respons pengguna terhadap TikTok Shop Seller Center, yang nantinya dapat membantu pengembang dan penjual untuk meningkatkan kualitas layanan mereka dan memenuhi ekspektasi pengguna. Selain itu, penelitian ini juga akan memberikan pemahaman tentang efektivitas Algoritma Naïve Bayes dalam analisis sentimen terhadap aplikasi seluler yang sedang berkembang pesat[7]. Oleh karena itu, penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam pemahaman dan pengembangan aplikasi seluler di era digital saat ini.

B. Metode Penelitian

Untuk alur penelitiannya, tertuang dalam Flowchart dibawah ini:



Gambar 1 Alur penelitian

B.1. Scrapping Data

Data didapatkan melalui proses web scraping yang secara otomatis mengambil data dari situs web tanpa perlu melakukan penyalinan manual. Dengan menggunakan perangkat lunak google-play-scraper untuk mengumpulkan data ulasan terkait dengan aplikasi TikTok Shop Seller Center di Google Playstore. Penggunaan teknik web scraping dalam penelitian ini bertujuan untuk mengakses data ulasan dari berbagai pengguna, yang nantinya akan menjadi sumber data utama dalam analisis sentimen menggunakan metode algoritma Naïve Bayes.

B.2. Data Cleaning

Tahap data cleaning, yang juga sering disebut sebagai tahap pembersihan data, merupakan langkah penting dalam analisis sentimen. Data yang berasal dari Google Playstore sering kali mengandung noise seperti karakter khusus, tautan, atau teks yang kurang relevan. Oleh karena itu, langkah pembersihan data diperlukan untuk meningkatkan kualitas dan ketepatan data. Dengan melakukan pembersihan data, analisis sentimen dapat lebih fokus pada kata kunci yang relevan, dan dapat meningkatkan akurasi dalam memprediksi sentimen.

B.3. Preprocessing

Preprocessing adalah langkah awal dalam pengolahan data untuk analisis sentimen, yang memiliki tujuan untuk menghapus kata-kata atau kalimat yang kurang relevan dan menyusun bentuk kata untuk mengurangi volume pada kosakata. Langkah-langkah pada preprocessing yaitu sebagai berikut:

B.3.1. Case Folding

Case folding merupakan langkah dimana seluruh huruf dalam teks ulasan disesuaikan menjadi huruf kecil atau huruf besar, sehingga seluruh huruf dalam teks memiliki format penulisan yang seragam[8]. Jika tidak dilakukan case folding, kata yang sama yang ditulis dengan casing yang berbeda (seperti "Aplikasi" dan "aplikasi") akan dianggap sebagai kata yang berbeda oleh algoritma yang dapat berdampak atau berpengaruh pada hasil analisis.

B.3.2. Stopword Removal

Stopword removal adalah tahap mengidentifikasi dan menghapus kata-kata umum yang tidak memberikan informasi signifikan dalam analisis sentimen[9]. Istilah "stop words" digunakan karena kata-kata ini sering muncul dalam bahasa alami secara umum dan tidak membawa makna sentimen yang kuat. Tindakan menghilangkan stop words membantu menyederhanakan data dan meningkatkan efisiensi analisis sentimen. Contoh kata-kata stop words mencakup "yang", "adalah", "dan", "di", "itu", serta lainnya.

B.3.3. Tokenizing

Tokenisasi adalah tahap membagi teks atau dokumen menjadi unit-unit yang lebih kecil yang disebut "token"[10]. Dalam analisis teks, token biasanya merujuk pada kata-kata individual dalam teks. Tokenisasi merupakan proses yang penting karena membantu mengubah teks yang panjang menjadi struktur data yang lebih sederhana yang dapat dikelola dan dimengerti oleh komputer. Misalnya, kalimat

"aplikasi tiktok shop seller center sangat bagus!" akan diurai menjadi token-token seperti ["aplikasi", "tiktok", "shop", "seller", "center", "sangat", "bagus"].

B.3.4. Stemming

Stemming merupakan tahap akhir preprocessing dalam penelitian ini yakni data teks yang melibatkan penghilangan akhiran atau awalan kata-kata untuk menghasilkan bentuk dasar atau "akar" kata. Fungsinya adalah untuk mengurangi variasi morfologi dalam kata-kata, maka kata-kata yang memiliki makna serupa tetapi memiliki bentuk yang berbeda dianggap sebagai bentuk yang sama[11].

B.4. Labelling

Labelling atau pelabelan adalah tahap yang penting dalam analisis sentimen yang mencakup pemberian label atau kategori tertentu pada data ulasan untuk mengindikasikan sentimen yang terkandung dalam ulasan tersebut[12]. Dalam konteks penelitian analisis sentimen pada ulasan aplikasi TikTok Shop Seller Center di Google Play Store. Labelling atau pelabelan dilakukan untuk mengelompokkan setiap ulasan pengguna ke dalam beberapa kategori sentimen positif dan sentimen negatif.

B.5. TF-IDF

Memfaatkan representasi TF-IDF adalah cara untuk mengkonversi teks ulasan menjadi vektor numerik yang dapat dimanfaatkan oleh algoritma klasifikasi seperti Naive Bayes[13]. Ini bermanfaat dalam menilai tingkat pentingnya kata-kata dalam setiap ulasan. Pada langkah ini, juga dilakukan proses pembobotan kata, di mana setiap kata diberi bobot yang menggambarkan seberapa pentingnya kata tersebut dalam suatu dokumen. Perhitungan bobot setiap kata dihitung untuk seluruh dokumen dengan tujuan untuk mengidentifikasi dan menghitung jumlah kata yang sering muncul dalam dokumen, dan semakin sering kata tersebut muncul, semakin tinggi bobot atau nilai yang diberikan pada kata tersebut.

B.6. Naïve Bayes

Algoritma klasifikasi Naive Bayes adalah suatu model algoritma yang umumnya digunakan untuk mengklasifikasikan data dengan bantuan perhitungan probabilitas[14]. Perhitungan probabilitas untuk setiap kata dalam dokumen atau dataset dapat menghasilkan dua jenis klasifikasi, yaitu sentimen positif dan negatif, yang berasal dari proses komputasi yang sebelumnya dilakukan pada pembobotan kata TF-IDF. Model Naive Bayes digunakan untuk memprediksi sentimen positif dan negatif berdasarkan vektor TF-IDF dari ulasan. Pada tahap ini data harus dibagi menjadi 2. Penting untuk membagi data menjadi dua bagian, yakni data pelatihan (training data) dan data pengujian (testing data)[15]. Pemisahan ini adalah langkah penting dalam menerapkan algoritma Naive Bayes atau kebanyakan algoritma klasifikasi lainnya. Pemisahan data dilakukan dengan maksud melatih model menggunakan dataset yang ada, lalu menguji sejauh mana model mampu mengaplikasikan pengetahuannya pada dataset yang belum pernah dilihatnya sebelumnya. Data pelatihan digunakan untuk melatih model, sementara data pengujian berguna untuk menilai kinerja model.

B.7. Confusion Matrix

Setelah model Naive Bayes selesai dilatih, evaluasi dilakukan dengan menggunakan confusion matrix. Dalam evaluasi ini, digunakan metrik-metrik seperti akurasi, presisi, recall, dan F1-score yang digunakan untuk menilai sejauh mana model dapat memprediksi sentimen dari ulasan. Saat mengukur performa, terdapat empat konsep yang digunakan untuk mewakili hasil dari proses klasifikasi, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)[16].

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Gambar 2 Tabel confusion matrix

Setelah didapatkan nilai dari keempat istilah tersebut maka setelah itu dapat didapatkan nilai untuk akurasi, presisi, dan recall nya pada rumus berikut:

- Akurasi

Akurasi mengukur seberapa baik model dalam melakukan klasifikasi secara keseluruhan dan sering digunakan dalam masalah klasifikasi yang seimbang. Secara matematis, dapat dirumuskan sebagai berikut:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total}} \quad (1)$$

- Presisi

Presisi adalah metrik yang mengukur sejauh mana prediksi positif model adalah benar. Semakin tinggi nilai presisi, semakin baik model dalam menghindari melakukan prediksi positif yang salah. Secara matematis, dapat dirumuskan sebagai berikut:

$$\text{Presisi} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

- Recall

Recall, juga dikenal sebagai Sensitivitas atau True Positive Rate, adalah metrik yang mengukur sejauh mana model dapat mengidentifikasi semua instance positif yang sebenarnya. Secara matematis, dapat dirumuskan sebagai berikut:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

- F1-Score

F1 Score adalah metrik yang menggabungkan Presisi dan Recall menjadi satu nilai tunggal. Secara matematis, dapat dirumuskan sebagai berikut:

$$F1-Score = \frac{2}{\frac{1}{Presisi} + \frac{1}{Recall}} = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (4)$$

Keterangan:

TP = hasil data dengan nilai prediksi positif dan nilai aktual positif.

TN = hasil data dengan nilai prediksi negatif dan nilai aktual negatif.

FP = hasil data dengan nilai prediksi positif, tetapi nilai aktualnya adalah negatif.

FN = hasil data dengan nilai prediksi negatif, tetapi nilai aktualnya adalah positif.

B.8. Word Cloud

Langkah berikutnya setelah proses evaluasi adalah tahap visualisasi yang melibatkan penggunaan word cloud. Pada tahap ini, digunakan visualisasi dengan word cloud untuk menggambarkan semua sentimen dari seluruh dokumen ulasan, baik sentimen yang bersifat positif maupun sentimen negatif terkait dengan aplikasi TikTok Shop Seller Center.

C. Hasil dan Pembahasan

C.1. Scrapping Data

Data diambil dengan menggunakan teknik web scraping menggunakan Google Colaboratory, dengan menginstal alat bernama google-play-scraper untuk mengambil data ulasan dari Google Playstore. Total data yang berhasil diambil dan yang akan digunakan adalah 5000 data. Hasil pengambilan data mentah melalui proses web scraping dapat dilihat dalam gambar 3. Data yang diperoleh melalui web scraping kemudian akan diorganisir menjadi empat kolom yang akan difilter, yaitu kolom userName, at, score, dan content, guna memudahkan analisis di tahap selanjutnya.

	userName	score	at	content
0	Aline Nander	5	2023-10-04 18:13:26	Untik tiktok seler mesti di perbaiki sistemnya...
1	BTR_INDO ELITE	1	2023-09-26 17:07:04	Saya turunin drastis bintang nya. Dari tadi up...
2	siti dwi	2	2023-10-03 07:52:19	Produk sudah laku ribuan pcs di blokir. Ajuakn...
3	Rehan fajar Rachmansyah	1	2023-10-01 12:40:55	Maksudnya apa ya kunjungan no hp terlalu serin...
4	Franco Rendy	4	2023-08-28 13:23:04	Aplikasi sudah bagus namun yg perlu jadi bahan...

Gambar 3 Hasil web scrapping

Karena hanya membutuhkan kolom content dan score, maka dilakukan filter kolom sehingga menyisakan kolom content dan score untuk menganalisis ulasan ke dalam sentimen positif dan sentimen negatif menjadi pada gambar 4.

	content	score
1077	Bagus dan membuat UMKM. masukan aja.. penarik...	5
435	Ongkirnya Mehong bgt ... dan juga terlalu ba...	5
4958	Ini sangat Bagus karena tiktok asli nya bisa l...	5
426	Lumayan membantu utk pedang online..semoga mas...	5
311	Hidup lagi susah susahnya malah Aplikasi nya D...	5

Gambar 4 Filter data hasil web scrapping

C.2. Data Cleaning

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   content     5000 non-null   object
1   score       5000 non-null   int64
dtypes: int64(1), object(1)
memory usage: 78.2+ KB
```

Gambar 5 Informasi data

Pada gambar 5 terdapat informasi data terdiri dari dua kolom, yaitu kolom content yang memiliki tipe data objek yang umumnya digunakan untuk data teks atau campuran dan kolom score yang memiliki tipe data int64 yang menunjukkan bahwa ini adalah kolom berisi angka bulat. Dalam kedua kolom tersebut, tidak ada nilai yang hilang dalam DataFrame. Selanjutnya, dilakukan tahap text mining atau pembersihan data untuk membersihkan teks dalam DataFrame. Proses ini dilakukan untuk menghilangkan noise, seperti karakter khusus, tautan, atau teks yang tidak relevan. Setelah pembersihan data, DataFrame akan menjadi lebih bersih dengan menghapus karakteristik tertentu yang dianggap tidak diperlukan, sebagaimana ditunjukkan dalam Tabel 1.

Tabel 1 Hasil proses data cleaning

Sebelum Proses Data Cleaning	Setelah Proses Data Cleaning
Pencairan lama giliran udh tgl nya bisa di tarik gak di tambah"n ke daftar tarik. Hadehhh ðŸ™ ðŸ™	Pencairan lama giliran udh tgl nya bisa di tarik gak di tambahn ke daftar tarik Hadehhh

Pada tabel 1 menunjukkan hasil dari ulasan sebelum dan sesudah proses data cleaning menjadi lebih bersih dengan karakteristik-karakteristik tertentu yang dianggap tidak diperlukan seperti tanda baca titik dan karakter. Dengan adanya proses cleaning data untuk memastikan bahwa data yang digunakan adalah akurat, konsisten, dan bebas dari kesalahan atau noise.

C.3. Preprocessing

Preprocessing merupakan langkah awal dalam pengolahan data untuk analisis sentimen dengan tujuan kata-kata atau frasa yang tidak penting serta menormalkan

bentuk kata-kata untuk mengurangi variasi dalam kosa kata. Berikut adalah langkah-langkah dalam proses preprocessing:

C.3.1. Case Folding

Tabel 2 Hasil proses case folding

Sebelum Proses Case Folding	Sesudah Proses Case Folding
Pencairan lama giliran udh tgl nya bisa di tarik gak di tambahn ke daftar tarik Hadehhh	pencairan lama giliran udh tgl nya bisa di tarik gak di tambahn ke daftar tarik hadehhh

Tabel 2 memperlihatkan hasil sebelum dan setelah melewati tahap case folding yang melibatkan perubahan pada huruf kapital. Data sebelum melalui case folding masih memiliki huruf kapital, sementara setelah tahap case folding, semua huruf telah diubah menjadi huruf kecil. Tahapan case folding ini sangat diperlukan untuk mengurangi perbedaan antara huruf besar (uppercase) dan huruf kecil (lowercase) dalam teks, sehingga memungkinkan pengolahan teks menjadi lebih konsisten dan mudah dikelola.

C.3.2. Stopword Removal

Tabel 3 Hasil proses stopword removal

Sebelum Proses Stopword Removal	Sesudah Proses Stopword Removal
pencairan lama giliran udh tgl nya bisa di tarik gak di tambahn ke daftar tarik hadehhh	pencairan giliran udh tgl nya tarik gak tambahn daftar tarik hadehhh

Tabel 3 menunjukkan hasil setelah dilakukan stopword removal, yaitu kata-kata seperti "lama", "bisa", "di" dan "ke" telah dihilangkan karena merupakan kata-kata stopword yang umum dalam bahasa dan tidak memiliki makna khusus dalam konteks ini. Hasil akhir dari penghapusan stopword adalah teks yang lebih ringkas dan lebih terfokus pada kata-kata kunci yang kemungkinan lebih relevan dalam analisis atau proses selanjutnya. Hal ini membantu dalam menyederhanakan teks dan mengurangi kebingungan yang dapat timbul akibat kata-kata yang sering muncul tetapi tidak bersifat penting.

C.3.3. Tokenizing

Tabel 4 Hasil proses tokenizing

pencairan	giliran	udh	tgl	nya	tarik
gak	tambahn	daftar	tarik	hadeh	

Tabel 4 menggambarkan hasil tokenisasi, yang merupakan proses memecah kalimat-kalimat ulasan menjadi unit yang lebih kecil, yaitu kata-kata, sehingga dapat lebih mudah dikelola dan dimengerti oleh komputer.

C.3.4. Stemming

Langkah berikutnya adalah proses stemming, di mana data yang telah diolah pada tahap sebelumnya akan mengalami perubahan kata menjadi bentuk kata dasar. Proses stemming ini melibatkan penggunaan library Sastrawi untuk melakukan pemrosesan bahasa Indonesia dengan mengubah kata-kata dalam bahasa Indonesia ke bentuk dasarnya dengan menghapus awalan dan akhiran. Sastrawi sangat berguna dalam berbagai aplikasi pemrosesan teks dan analisis bahasa alami dalam bahasa Indonesia. Salah satu data yang telah melewati proses stemming dapat dilihat pada tabel 5.

Tabel 5 Hasil proses stemming

Sebelum Proses Stemming	Setelah Proses Stemming
pencairan giliran udh tgl nya tarik gak tambahn daftar tarik hadehhh	cair gilir udh tgl nya tarik gak tambahn daftar tarik hadehhh

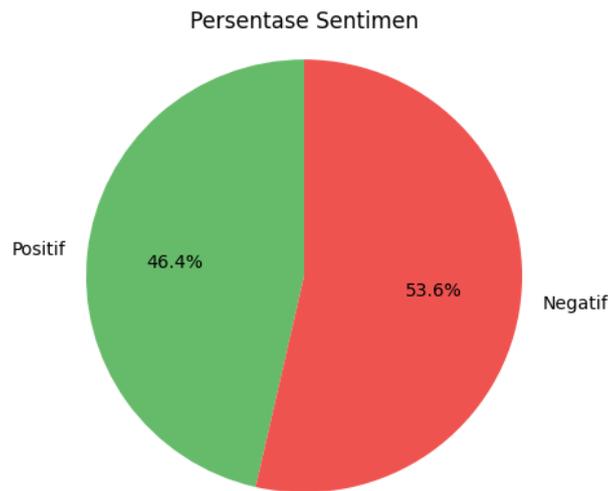
C.4. Labelling

Label diberikan pada data ulasan untuk mengindikasikan sentimen yang terkandung dalam ulasan tersebut dengan mengelompokkan setiap ulasan pengguna ke dalam beberapa kategori sentimen. Proses memberi label ini dapat membantu komputer memahami bagaimana pengguna merasakan atau mengevaluasi produk atau layanan yang diulas. Data yang telah melewati tahap preprocessing akan diberi label seperti pada gambar 6.

Bagus untuk para penjual online & UMKM	5	bagus untuk para penjual online umkm	bagus penjual online umkm	['bagus', 'penjual', 'online', 'umkm']	bagus jual online umkm	Positif
Kapan kami bisa berjualan lagi di tiktok, sedi...	1	kapan kami bisa berjualan lagi di tiktok sedih...	berjualan tiktok sedih banget ga jualan tiktok	['berjualan', 'tiktok', 'sedih', 'banget', 'ga...']	jual tiktok sedih banget ga jual tiktok	Negatif
Apk ngga jelas akun buat apa apa engga di hapu...	1	apk ngga jelas akun buat apa apa engga di hapu...	apk ngga akun engga hapus situ saldonya merugi...	['apk', 'ngga', 'akun', 'engga', 'hapus', 'sit...']	apk ngga akun engga hapus situ saldo rugi guna...	Negatif
tolong tiktok shop dibuka lagi kak.....	5	tolong tiktok shop dibuka lagi kak	tolong tiktok shop dibuka kak	['tolong', 'tiktok', 'shop', 'dibuka', 'kak']	tolong tiktok shop buka kak	Positif
sangat membantu	5	sangat membantu	membantu	['membantu']	bantu	Positif
Mangkanya jgn susah susah . toko tutup pemerin...	1	mangkanya jgn susah susah toko tutup pemerin...	mangkanya jgn susah susah toko tutup pemerinta...	['mangkanya', 'jgn', 'susah', 'susah', 'toko', '....']	mangkanya jgn susah susah toko tutup perintah ...	Negatif

Gambar 6 Hasil proses labelling

Label sentimen diberikan berdasarkan skor yang diberikan oleh pengguna. Skor ini adalah representasi numerik dari penilaian atau ulasan yang diberikan. Pada penelitian ini, kategori dibagi menjadi 2 sentimen yakni sentimen positif dan sentimen negatif. Sentimen positif berasal dari ulasan dengan rating 4 atau 5 sedangkan sentimen negatif berasal dari ulasan dengan rating di bawah 3.



Gambar 7 Diagram pie hasil labelling

Pada gambar 7 menunjukkan hasil labelling dari 5000 data ulasan adalah sentimen positif sebanyak 46.4% dan sentimen negatif sebanyak 53.6%.

C.5. TF-IDF

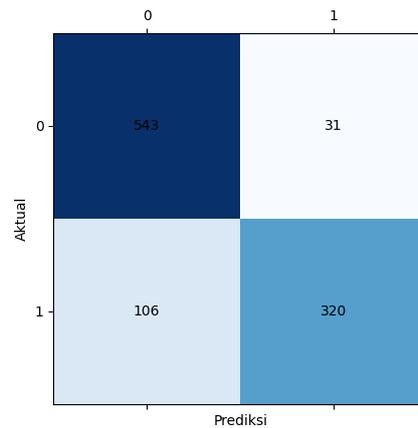
Dalam tahap ekstraksi fitur, library yang diperlukan adalah Scikit-Learn. Scikit-Learn menyediakan serangkaian modul atau fungsi yang dapat digunakan. Salah satu modul atau fungsi yang digunakan dalam tahap ini adalah `TfidfVectorizer`, yang bertugas mengubah kumpulan dokumen data ulasan menjadi matriks fitur TF-IDF. Setelah data mengalami proses TF-IDF, setiap kata dalam setiap baris data akan memiliki bobot numerik. Perhitungan bobot untuk setiap term dilakukan untuk semua dokumen, dengan tujuan mengidentifikasi dan menghitung kata-kata yang kerap muncul pada dokumen. Semakin sering kata tersebut muncul, semakin tinggi bobot atau nilai term tersebut, seperti yang ditunjukkan dalam gambar 8.

	Term	Average TF-IDF Rank
0	bantu	0.049775
1	jual	0.044308
2	bagus	0.034238
3	tiktok	0.032780
4	aplikasi	0.032659
5	produk	0.030509
6	nya	0.029513
7	mudah	0.028922
8	seller	0.027414
9	yg	0.025606
10	gak	0.020573

Gambar 8 Hasil pembobotan TF-IDF

C.6. Naïve Bayes

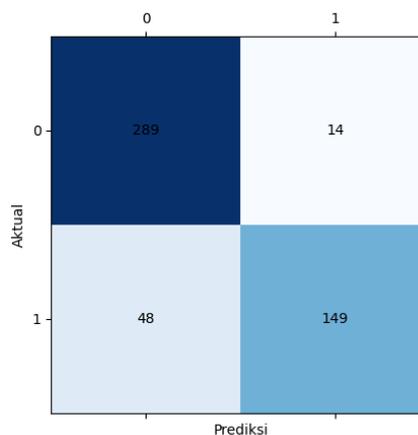
Metode Naive Bayes diterapkan untuk melakukan proses klasifikasi dengan memanfaatkan Multinomial Naïve Bayes yang umum digunakan untuk tugas klasifikasi teks, terutama ketika data memiliki representasi sebagai distribusi multinomial. Pada penelitian ini, akan dilakukan pembagian data sebesar 90 : 10 dan 80 : 20. Dalam pembagian ini, 90% atau 80% dari data digunakan sebagai data pelatihan, sedangkan sisanya sebesar 20% atau 10% digunakan sebagai data pengujian. Setelah pelatihan model menggunakan data pelatihan, langkah berikutnya adalah tahap evaluasi. Evaluasi dilakukan dengan confusion matrix untuk mengukur performa model klasifikasi yang dapat dilihat pada gambar 9 dan 10.



Gambar 9 Visualisasi confusion matrix rasio data 80:20

Dari gambar 9 dapat ditarik beberapa informasi yaitu:

- True Negative (TN) : Data dengan label negatif yang memiliki prediksi benar sebagai label negatif sebanyak 543.
- True Positive (TP) : Data dengan label positif yang memiliki prediksi benar sebagai label positif sebanyak 320.
- False Negative (FN) : Data dengan label positif yang memiliki prediksi salah sebagai label negatif sebanyak 106.
- False Positive (FP) : Data dengan label negatif yang memiliki prediksi salah sebagai label positif sebanyak 31.

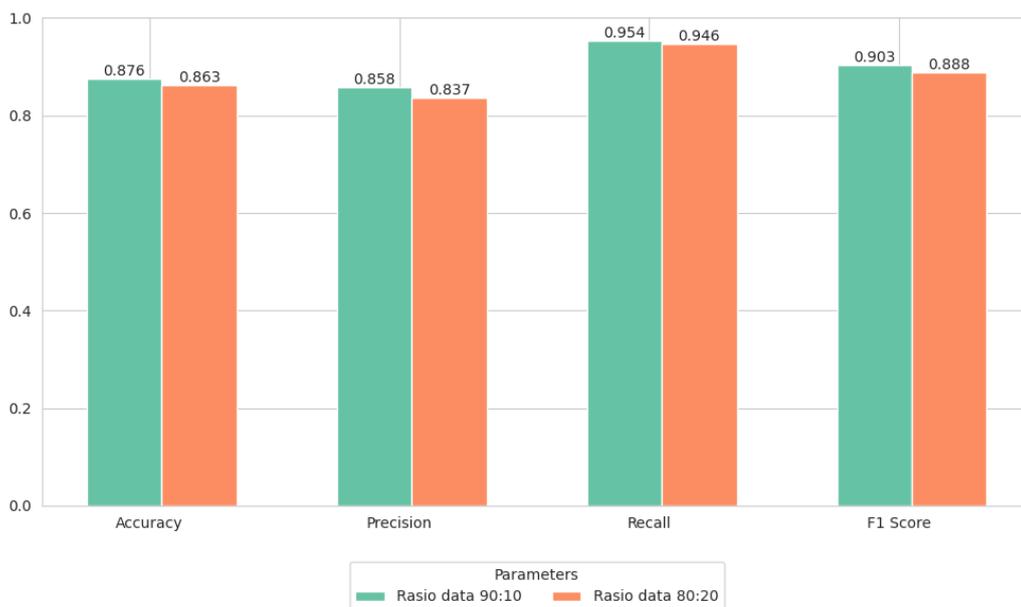


Gambar 10 Visualisasi confusion matrix rasio data 90:10

Dari gambar 10 dapat ditarik beberapa informasi yaitu:

- True Negative (TN) : Data dengan label negatif yang memiliki prediksi benar sebagai label negatif sebanyak 289.
- True Positive (TP) : Data dengan label positif yang memiliki prediksi benar sebagai label positif sebanyak 149.
- False Negative (FN) : Data dengan label positif yang memiliki prediksi salah sebagai label negatif sebanyak 48.
- False Positive (FP) : Data dengan label negatif yang memiliki prediksi salah sebagai label positif sebanyak 14.

Selanjutnya dilakukan proses validasi silang dengan menerapkan algoritma Naïve Bayes Classifier (NBC). Data tersebut akan melewati proses pengujian dan pelatihan, dilanjutkan dengan membangun model menggunakan algoritma NBC untuk mengukur tingkat akurasi, presisi, recall dan skor F1 berdasarkan kinerja algoritma Machine Learning. Penilaian ini didasarkan pada Confusion Matrix. Hasil klasifikasi menggunakan algoritma Naive Bayes dengan penggunaan pembobotan TF-IDF dapat dilihat di bawah ini.

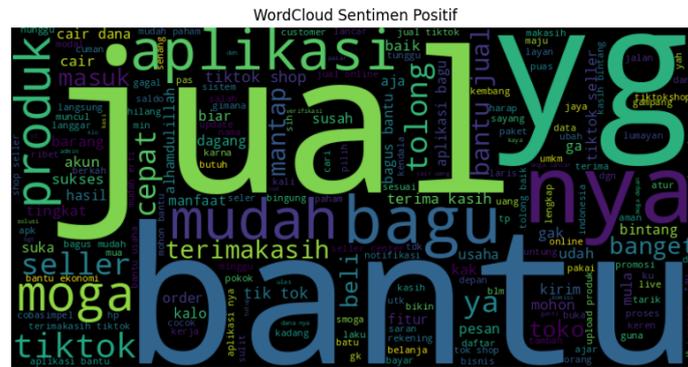


Gambar 11 Hasil metrik evaluasi model

Berdasarkan pada gambar 11 menunjukkan hasil metrik evaluasi dari sebuah model klasifikasi yang menggunakan algoritma Multinomial Naive Bayes (MultinomialNB) untuk melakukan prediksi, didapatkan hasil bahwa rasio data 90:10 lebih baik dibanding 80:20 yaitu didapatkan nilai akurasi 87,6%, presisi 85,8%, recall 95,4% dan F1-Score 90,3%.

C.7. Word Cloud

Tahapan terakhir yaitu adalah tahap visualisasi data menggunakan wordcloud untuk mengidentifikasi kata-kata yang paling kerap muncul dalam setiap jenis sentimen.



Gambar 12 Word cloud sentimen positif



Gambar 13 Word cloud sentimen negatif

Dalam gambar ke 15, terdapat kata-kata yang kerap muncul dalam sentimen positif, antara lain "membantu," "menjual," "aplikasi," "baik," dan "sederhana." Sedangkan di gambar ke 16, terlihat kata-kata yang kerap muncul, antara lain "penjualan," "tidak," "produk," "melanggar," dan "kegagalan."

D. Simpulan

Berdasarkan hasil analisis sentimen dari 5000 data ulasan pengguna aplikasi TikTok Shop Seller Center menunjukkan bahwa mayoritas ulasan memiliki sentimen negatif daripada sentimen positif. Kemudian, pada hasil evaluasi menggunakan confusion matrix mengungkapkan bahwa algoritma Multinomial Naïve Bayes menunjukkan performa yang lebih baik pada pembagian data dengan rasio 90:10 dibandingkan dengan pembagian 80:20 dengan akurasi 87,6%, presisi 85,8%, recall 95,4% dan F1-Score sebesar 90,3%. Data ini memperlihatkan bahwa model algoritma Naïve Bayes mampu mengklasifikasikan sentimen negatif dengan baik. Meskipun mayoritas ulasan cenderung negatif, analisis ini memberikan wawasan penting kepada pengelola aplikasi TikTok Shop Seller Center untuk memahami masalah dan perasaan yang dialami oleh pengguna. Dengan informasi ini, mereka dapat berupaya untuk meningkatkan pengalaman pengguna, mengatasi masalah yang mungkin muncul, dan meningkatkan reputasi aplikasi mereka di mata pengguna.

E. Ucapan Terima Kasih

Peneliti menyampaikan rasa terima kasih kepada Universitas Amikom Yogyakarta beserta tim pendukung yang telah memberikan kontribusi berharga dalam pelaksanaan penelitian ini.

F. Referensi

- [1] M. H. Asnawi, "Perbandingan Algoritma Naïve Bayes, K-NN, dan SVM dalam Pengklasifikasian Sentimen Media Sosial", doi: 10.1234/pns.v10i.85.
- [2] R. Wahyudi *et al.*, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *JURNAL INFORMATIKA*, vol. 8, no. 2, 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [3] S. Saepudin, S. Widiastuti, and C. Irawan, "Sentiment Analysis of Social Media Platform Reviews Using the Naïve Bayes Classifier Algorithm," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 2, pp. 236–243, Jul. 2023, doi: 10.32736/sisfokom.v12i2.1650.
- [4] R. L. Atimi and Enda Esyudha Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *Jurnal Sains dan Informatika*, vol. 8, no. 1, pp. 88–96, Jul. 2022, doi: 10.34128/jsi.v8i1.419.
- [5] R. Tsania, S. A. Putri, D. E. Ratnawati, and D. W. Brata, "Perbandingan Naïve Bayes dan K-Nearest Neighbor untuk Analisis Sentimen Aplikasi Gapura UB Berdasarkan Ulasan Pengguna pada Playstore," 2023. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] D. Sepri, P. Algoritma, N. Bayes, U. Analisis, K. Penggunaan, and A. Bank, "media cetak," *Journal of Computer System and Informatics (JoSYC)*, vol. 2, no. 1, pp. 135–139, 2020.
- [7] M. Daffa Rhajendra¹ and N. Trianasari, "Analisis Sentimen Ulasan Aplikasi Spotify Untuk Peningkatan Layanan Menggunakan Algoritma Naive Bayes Sentiment Analysis of Spotify Application Reviews for Service Improvement Using Naive Bayes Algorithm."
- [8] R. Kosasih and A. Alberto, "Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor," vol. 6, no. 1, 2021, doi: 10.30743/infotekjar.v6i1.3893.
- [9] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter".
- [10] A. U. T. Ama, D. N. Mulya, Y. P. D. Astuti, and I. B. G. Prasadhya, "Analisis Sentimen Customer Feedback Tokopedia Menggunakan Algoritma Naïve Bayes," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, p. 50, Sep. 2022, doi: 10.30865/json.v4i1.4783.
- [11] N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *Jurnal Sistem dan Informatika (JSI)*, vol. 15, no. 1, pp. 27–29, Nov. 2020, doi: 10.30864/jsi.v15i1.332.

-
- [12] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *PETIR*, vol. 15, no. 2, pp. 264–275, Nov. 2022, doi: 10.33322/petir.v15i2.1733.
- [13] G. Darmawan, S. Alam, M. Imam Sulisty, P. Studi Teknik Informatika, S. Tinggi Teknologi Wastukencana Purwakarta, and R. Artikel, "ANALISIS SENTIMEN BERDASARKAN ULASAN PENGGUNA APLIKASI MYPERTAMINA PADA GOOGLE PLAYSTORE MENGGUNAKAN METODE NAÏVE BAYES INFO ARTIKEL ABSTRAK," vol. 2, no. 3, pp. 100–108, 2023, doi: 10.55123.
- [14] O.: Bustami, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENGLASIFIKASI DATA NASABAH ASURANSI."
- [15] L. Rahmawati and D. B. Santoso, "IMPLEMENTASI METODE NAIVE BAYES UNTUK KLASIFIKASI ULASAN APLIKASI E-COMMERCE TOKOPEDIA IMPLEMENTATION OF NAIVE BAYES METHOD FOR CLASSIFICATION OF TOKOPEDIA E-COMMERCE APPLICATION REVIEW," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 6, no. 1, 2023.
- [16] A. Dwiki, A. Putra, and S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," vol. 8, no. 2, 2021, [Online]. Available: <http://jurnal.mdp.ac.id>