



Similarity Report

Metadata

Name of the organization

Universitas Muhammadiyah Sidoarjo

Title

Klasifikasi Ulasan Pelanggan Adika Surya

Author(s)

Coordinator

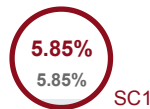
perpustakaan umsidahanin

Organizational unit

Perpustakaan

Record of similarities

SCs indicate the percentage of the number of words found in other texts compared to the total number of words in the analysed document. Please note that high coefficient values do not automatically mean plagiarism. The report must be analyzed by an authorized person.

**25**

The phrase length for the SC 2

4564

Length in words

33471

Length in characters

Alerts

In this section, you can find information regarding text modifications that may aim at temper with the analysis results. Invisible to the person evaluating the content of the document on a printout or in a file, they influence the phrases compared during text analysis (by causing intended misspellings) to conceal borrowings as well as to falsify values in the Similarity Report. It should be assessed whether the modifications are intentional or not.

Characters from another alphabet		0
Spreads		0
Micro spaces		0
Hidden characters		0
Paraphrases (SmartMarks)		22

Active lists of similarities

This list of sources below contains sources from various databases. The color of the text indicates in which source it was found. These sources and Similarity Coefficient values do not reflect direct plagiarism. It is necessary to open each source, analyze the content and correctness of the source crediting.

The 10 longest fragments

Color of the text

NO	TITLE OR SOURCE URL (DATABASE)	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
1	Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee Faqih Ahmad, Gifthera Dwilestari, Asep Saepudin;	23 0.50 %
2	Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine Ahmad Fauzi, Anis Fitri Nur Masruriyah, Tania Puspa Rahayu Sanjaya;	15 0.33 %

3	ANALISIS KEPUTUSAN PEMBELIAN PADA GEN Z TERHADAP DISKON, PROMOSI, DAN KEPERCAYAAN KONSUMEN DI MARKETPLACE SHOPEE Firlyana Muhamad Robby, Aldina Shiratina;	15 0.33 %
4	Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine Rahayu Irma Putri, Fauzi Ahmad, Indra Jamaludin;	14 0.31 %
5	Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine Ahmad Hanif Asyhar,Vina Fitriyana, Dian Candra Rini Novitasari, Lutfi Hakim;	13 0.28 %
6	https://archive.umsida.ac.id/index.php/archive/preprint/download/4816/34394/38819	12 0.26 %
7	https://rama.unimal.ac.id/id/eprint/2237/3/bab%20I.pdf	12 0.26 %
8	Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee Faqih Ahmad, Gifthera Dwilestari,Asep Saepudin;	11 0.24 %
9	Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor (KNN) Hanif Habibie,Atthalla Ikhsan Nuh, Jovandy Adithia;	11 0.24 %
10	https://pdfs.semanticscholar.org/bf72/c797cc40a6fb38bf62dca2085ee6624cb76f.pdf	10 0.22 %
from RefBooks database (4.45 %)		
NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
Source: Paperity		
1	Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine Ahmad Hanif Asyhar,Vina Fitriyana, Dian Candra Rini Novitasari, Lutfi Hakim;	81 (12) 1.77 %
2	Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee Faqih Ahmad, Gifthera Dwilestari,Asep Saepudin;	46 (4) 1.01 %
3	Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine Rahayu Irma Putri, Fauzi Ahmad, Indra Jamaludin;	21 (2) 0.46 %
4	ANALISIS KEPUTUSAN PEMBELIAN PADA GEN Z TERHADAP DISKON, PROMOSI, DAN KEPERCAYAAN KONSUMEN DI MARKETPLACE SHOPEE Firlyana Muhamad Robby, Aldina Shiratina;	15 (1) 0.33 %
5	Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine Ahmad Fauzi, Anis Fitri Nur Masruriyah,Tania Puspa Rahayu Sanjaya;	15 (1) 0.33 %
6	Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor (KNN) Hanif Habibie,Atthalla Ikhsan Nuh, Jovandy Adithia;	11 (1) 0.24 %
7	PENERAPAN METODE NAÏVE BAYES UNTUK MEREKOMENDASIKAN PEKERJAAN YANG SESUAI TERHADAP FRESH GRADUATE Woro Isti Rahayu,Ikrima Ningrumsari Mulyanan, Muhammad Yusril Helmi Setyawan;	7 (1) 0.15 %
8	Sentiment Analysis on Free Lunch & Milk Program Using Naive Bayes Algorithm Hasan Firman Noor,Saputra Ramadani;	7 (1) 0.15 %
from the home database (0.00 %)		
NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
from the Database Exchange Program (0.00 %)		

NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
from the Internet (1.40 %)		
NO	SOURCE URL	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
1	https://pdfs.semanticscholar.org/bf72/c797cc40a6fb38bf62dca2085ee6624cb76f.pdf	16 (2) 0.35 %
2	http://repository.unmuhjember.ac.id/14114/2/b.%20Abstrak.pdf	15 (2) 0.33 %
3	https://rama.unimal.ac.id/id/eprint/2237/3/bab%20I.pdf	12 (1) 0.26 %
4	https://archive.umsida.ac.id/index.php/archive/preprint/download/4816/34394/38819	12 (1) 0.26 %
5	https://archive.umsida.ac.id/index.php/archive/preprint/download/6941/49756/55502	9 (1) 0.20 %

List of accepted fragments (no accepted fragments)

NO	CONTENTS	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
----	----------	---------------------------------------

Customer Review Classification on the MitraShopee Application Using **Support Vector Machine and Naïve Bayes Methods**
 Klasifikasi Ulasan Pelanggan pada Aplikasi MitraShopee **Menggunakan Metode Support Vector Machine dan Naïve Bayes**

Adika Surya Perdana¹⁾, Istian Kriya Almanfaluti^{*,2)}, Alshaf Pebrianggara³⁾

1, 2, 3) **Program Studi Bisnis Digital, Universitas Muhammadiyah Sidoarjo, Indonesia** *Email Penulis Korespondensi: istian.alman@umsida.ac.id

Page | 1

2 | Page

Page | 3

I. Pendahuluan

E-commerce adalah proses jual beli yang berlangsung secara online melalui berbagai platform elektronik, yang mempertemukan konsumen dengan penjual berinteraksi dan melakukan bisnis tanpa harus bertemu secara fisik. Sumber pada KEMENKOPUKM Smesco mencatat terdapat ada sekitar 16.4 juta pelaku UMKM yang telah bergabung di platform digital, menunjukkan pertumbuhan lebih dari 100% sejak awal pandemic. Banyak pemilik usaha baik skala besar maupun kecil yang memindahkan operasi dalam bisnisnya dari konvensional menjadi digital, terutama pada e-commerce .

Gambar SEQ Gambar * ARABIC 1. Data peningkatan e-commerce di Indonesia

Jumlah pengguna di Indonesia jumlah pengguna e-commerce terus meningkat sejak tahun 2020. Pada tahun 2023, tercatat ada sekitar 38.63 juta pengguna, dan diperkirakan angka ini akan terus bertumbuh hingga mencapai 99.1 juta pengguna pada tahun 2029. E-Commerce menyediakan jangkauan pasar yang sangat luas tanpa batasan dan memberikan kemudahan transaksi bagi penjual maupun pembeli. Pembeli tanpa harus keluar rumah untuk membeli suatu barang yang dibutuhkan bahkan akses untuk mencari produk dan membandingkan harga dapat dilakukan dengan lebih mudah.

Gambar SEQ Gambar * ARABIC 2. Jumlah data penipuan transaksi online

Pada Gambar 2. menyajikan data laporan penipuan transaksi online di Indonesia periode tahun 2017 hingga 2022. Terlihat bahwa penipuan transaksi online umum merupakan jenis penipuan yang paling banyak dilaporkan, dengan angka mencapai 405.000 kasus. Dibandingkan dengan penipuan investasi online sebanyak 19.000 kasus dan penipuan jual beli online 12.000 kasus, penipuan transaksi online menunjukkan angka yang jauh lebih tinggi, menandakan perlunya peningkatan kewaspadaan dan upaya pencegahan yang efektif. Dengan banyaknya e-commerce saat ini masyarakat bingung dalam menentukan e-commerce yang benar dipercaya. Sehingga, perlu adanya penilaian yang dapat mengidentifikasi tingkat popularitas e-commerce di Indonesia mengacu pada klasifikasi ulasan dalam Google Play Store yang disebut dalam teks berbentuk pandangan kritik, saran, dan pertanyaan .

Berdasarkan hasil survei “ **Pola Perilaku Masyarakat saat Belanja di Marketplace**” dalam survei yang dilakukan oleh **GoodStats pada tahun 2022, responden menempatkan Shopee sebagai pilihan marketplace** kedua mereka. Dengan meningkatnya pengguna pada aplikasi Shopee ini, akhirnya Shopee berinovasi membuat aplikasi baru yaitu MitraShopee. MitraShopee sendiri merupakan platform yang menawarkan jual beli berbagai produk digital yang tersedia untuk diunduh di Google Play Store. Pada platform Google Play Store terdapat berbagai fitur-fitur yang bisa menunjang bagi pengguna seperti peringkat dan ulasan .

Memberikan ulasan online setelah pembelian menjadi salah satu tahapan penting dalam perjalanan pelanggan. Dengan adanya ulasan, pengguna menjadi lebih mudah dalam memperoleh informasi tentang suatu produk atau layanan, pengguna dapat dengan mudah mempercayai rekomendasi yang telah digunakan oleh pengguna sebelumnya . Belum ditemukan pendekatan yang sistematis dan benar untuk mengidentifikasi apakah sebuah ulasan termasuk dalam kategori kritik, saran, dan pertanyaan. Mengkategorikan ulasan sangat penting bagi pengembang dalam mengambil data dari para ulasan pengguna pada aplikasi guna menutupi keresahan pengguna dalam aplikasi.

Pada Penelitian sebelumnya masih belum ada yang secara spesifik mengkategorikan ulasan pelanggan aplikasi e-commerce ke dalam bentuk kritik,

saran, dan pertanyaan. Padahal pengelompokkan ini penting untuk membantu pengembang memahami kebutuhan dan keluhan pengguna secara lebih mendalam. Sebagian besar penelitian terdahulu masih berfokus pada analisis sentiment di dua kelas, yakni positif dan negatif, tanpa menelusuri lebih jauh bentuk konkret dari ulasan tersebut. Penelitian mengenai klasifikasi ulasan pada aplikasi MitraShopee masih sangat terbatas, sehingga diperlukan pendekatan sistematis untuk mengeksplorasi ulasan pada pengguna secara lebih komprehensif. Dalam konteks ini, belum ada perbandingan yang mendalam terkait performa algoritma Naive Bayes dan Support Vector Machine (SVM) dalam mengelompokkan multi-kelas khusus untuk ulasan aplikasi MitraShopee. Penelitian ini diharapkan dapat menjawab kekosongan tersebut melalui penerapan dua metode klasifikasi untuk mengolah ulasan berbasis data dari Google Play Store secara lebih efektif dan efisien.

Metode analisis sentimen yang digunakan dalam memahami kondisi emosional individu apakah itu positif, atau negatif. Analisis sentiment diterapkan pada aplikasi Shopee, tempat dimana sistem memperoleh dataset ulasan dari website Google Play Store sistem selanjutnya akan mengelompokkan ulasan untuk menilai apakah terdapat sentimen negatif atau positif. Berdasarkan klasifikasi tersebut, pengguna bisa mendapatkan informasi dan mengevaluasi aplikasi Shopee, yang membantu mereka dalam membuat keputusan. Analisis sentimen dalam komentar dan ulasan pada Google Play Store dapat dilakukan menggunakan dua pendekatan populer yaitu Support Vector Machine (SVM) dan Naive Bayes. Ulasan pelanggan pada Google Play Store menjadi penting dalam membantu user memilih platform yang terpercaya. Klasifikasi ulasan pelanggan pada aplikasi MitraShopee dengan penerapan metode Support Vector Machine dan Naive Bayes menunjukkan potensi besar dalam meningkatkan pengalaman pengguna dan memperbaiki layanan e-commerce. Proses pengujian dilakukan dengan membagi data diperuntukkan untuk melatih model, sementara 20% sisanya dimanfaatkan untuk menguji sejauh mana kinerja model yang telah dilatih. Menurut temuan penelitian, algoritma Naive Bayes menunjukkan akurasi sebesar 85%, sementara algoritma Support Vector Machine mendapatkan akurasi 81% dalam analisis sentimen. Dalam analisis sentimen, Naive Bayes terbukti unggul jika dibandingkan dengan metode SVM. Penelitian tentang analisis sentimen sangat diminati karena topik ini memiliki relevansi dan saya tarik yang signifikan. Di bidang ini, telah dilakukan penelitian tentang sentiment masyarakat terhadap PayLater, yang memanfaatkan algoritma Naive Bayes Classifier. Pada penelitiannya diperoleh nilai akurasi dengan menggunakan model confusion matrix menghasilkan tingkat akurasi sebesar 95%. Penelitian selanjutnya tentang studi klasifikasi teks review Tokopedia menggunakan SVM menghasilkan tingkat akurasi 75%.

II. Literature Review

E-Commerce

E-commerce atau perdagangan elektronik, yang merupakan fenomena telah mengubah cara transaksi bisnis di seluruh dunia. E-commerce meliputi proses pembelian, penjualan, dan pertukaran barang serta jasa melalui platform digital, jaringan computer, terutama internet. E-commerce melibatkan transaksi yang melibatkan pembeli dan penjual dengan memanfaatkan teknologi jaringan untuk meningkatkan efisiensi dan efektivitas dalam berbisnis. E-commerce juga mencakup berbagai model transaksi untuk memperluas cakupan pasar dan meningkatkan aksesibilitas bagi konsumen. Perkembangan teknologi informasi, khususnya internet, telah mendorong pertumbuhan e-commerce secara signifikan, memberikan kemudahan untuk transaksi yang lebih cepat dan menghemat biaya.

Klasifikasi Ulasan

Klasifikasi ulasan pelanggan dalam analisis sentiment merupakan pendekatan yang krusial untuk memahami opini dan pengalaman konsumen terhadap produk di platform e-commerce. Proses ini melibatkan pengkategorian ulasan menjadi beberapa kelas, seperti positif, negatif, dan netral, memungkinkan pemilik bisnis untuk mengevaluasi persepsi pelanggan secara lebih efektif. Klasifikasi ulasan tidak hanya membantu dalam memahami reaksi pelanggan tetapi juga berkontribusi pada pengambilan keputusan strategis yang lebih baik bagi perusahaan dalam meningkatkan produk dan layanannya.

Naive Bayes

Sebagai algoritma klasifikasi probalistik, Naive Bayes efektif dan sering digunakan dalam data mining, terutama untuk mengelompokkan teks, analisis sentiment dan deteksi spam. Naive Bayes bekerja berdasarkan dengan asumsi bahwa setiap fitur dalam dataset bersifat independent terhadap fitur lainnya. Meskipun asumsi ini jarang sepenuhnya terpenuhi dalam praktik, pendekatan ini tetap efektif dalam berbagai kasus klasifikasi. Keunggulan Naive Bayes terletak pada kecepatan komputasi dan efisiensi dalam menangani dataset besar dengan jumlah fitur yang tinggi. Naive Bayes dapat diterapkan dalam berbagai bidang, seperti prediksi penyakit, klasifikasi dokumen, dan sistem rekomendasi.

Support Vector Machine (SVM) Support Vector Machine merupakan metode pembagian mesin yang digunakan untuk klasifikasi dan regresi dengan pendekatan berbasis margin maksimum. SVM bekerja melalui pencairan hyperplane optimal untuk dapat dengan tujuan memisahkan kelas dengan margin yang maksimal, sehingga meningkatkan generalisasi sistem. Keunggulan utamanya terdapat dalam hal kemampuannya untuk memproses data berdimensi besar dengan fleksibilitasnya pada menangani kasus nonlinear melalui penggunaan fungsi kernel. Dengan optimisasi parameter yang tepat dan pemilihan kernel yang sesuai, Support Vector Machine tersebut sebagai salah satu algoritma yang populer digunakan pada penelitian dan aplikasi industri.

Term Frequency-Inverse Document (TF-IDF)

TF-IDF adalah sebuah teknik pembobotan yang banyak dimanfaatkan dalam pemrosesan teks dan klasifikasi dokumen. TF-IDF mengintegrasikan dua bagian penting, yaitu Term Frequency yang mengukur berapa sering kata muncul dalam sebuah dokumen, serta Inverse Document Frequency, yang mengukur frekuensi kemunculan kata dalam sebuah koleksi dokumen. Kedua komponen ini bertujuan untuk memberikan nilai lebih pada kata-kata yang sering muncul dalam satu dokumen, namun jarang muncul di dokumen lainnya. Itulah yang dilakukan oleh TF-IDF sangat efektif untuk menghilangkan kata umum yang sering ditemukan di hampir semua file, tetapi tidak memiliki nilai dalam klasifikasi.

III. Metode

Dalam penelitian ini menggunakan 6000 kumpulan data yang mencakup ulasan yang bersumber dari aplikasi e-commerce MitraShopee yang dihosting di platform Google Play Store, yang kemudian dikategorikan menjadi tiga kelompok berbeda yaitu kritik, saran, dan pertanyaan. Pemilihan jumlah data menggunakan metode Time Series diambil dari rentang waktu 09 Februari 2024 sampai dengan 14 April 2025 [12]. Pengambilan data didasarkan pada pertimbangan metodologis dan acuan dari studi-studi terdahulu dalam bidang Natural Language Processing (NLP) dan machine learning. Dalam literatur lain disebutkan bahwa dataset dengan jumlah lebih dari 1.000 hingga 5.000 sampel sudah mencukupi untuk membangun model klasifikasi teks yang valid, dengan catatan data telah melalui tahap pembersihan dan pelabelan yang baik [13]. Dalam kelas kritik mencakup ulasan yang mengekspresikan ketidakpuasan atau masalah yang dihadapi pengguna saat menggunakan aplikasi. Ini bisa berhubungan dengan performa aplikasi, fitur yang tidak berfungsi, atau pengalaman pengguna yang kurang memuaskan. Kelas Saran berisi masukan konstruktif dari pengguna tentang bagaimana aplikasi

dapat ditingkatkan. Pengguna memberi ide atau rekomendasi yang dapat membantu pengembang dalam meningkatkan kualitas aplikasi . Sedangkan, kelas pertanyaan mencakup ulasan yang menunjukkan keingintahuan pengguna tentang fitur atau aspek tertentu dari aplikasi. Ini sering kali berkaitan dengan cara penggunaan aplikasi atau klarifikasi mengenai kebijakan tertentu. Tahap alur penelitian digambarkan pada gambar 3.

Gambar SEQ Gambar * ARABIC 3. Tahapan penelitian

3.1 Pengumpulan Data (Scrapping)

Proses Pengumpulan Data Penyelidikan menggunakan 6000 kumpulan data yang terdiri dari data dari kelas kritik, saran, pertanyaan, lainnya, dan data untuk data testing yang diperoleh dari ulasan aplikasi MitraShopee dimana dapat diakses di Google Play Store melalui google play scrapping. Langkah berikut melibatkan pre-processing. Setelah menyelesaikan pembersihan kumpulan data yang tidak dimurnikan, fase berikutnya memerlukan pelabelan untuk menggambarkan kategori kritik, saran, dan pertanyaan. Setelah ini, prosedur pembobotan Term Frequency-Inverse Document Frequency dijalankan dalam menetapkan signifikan disetiap istilah yang ada pada dokumen. Teknik pengelompokkan diterapkan memanfaatkan algoritma SVM dan Naïve Bayes.

Tabel 1. Contoh hasil pengumpulan data

ID	UserName	Score	At	Comment
1	Penggunaan Google	5	11/17/2022 1:43	Cukup bagus apk nya bpjs
2	Penggunaan Google	5	11/17/2022 1:42	Sangat cepat dan mudah
3	Penggunaan Google	1	11/17/2022 1:42	Bintang setengah ada ga?
...	...	...	...	...
279876	Penggunaan Google	5	2/20/2018 7:32	Mantap gannn

1.
2. 3.2 Pre-processing Data
3. Di bagian preprocessing, kalimat harus disesuaikan untuk menjamin perkembangan data yang lancar ke fase berikutnya. Proses pengolahan teks melibatkan case folding, tokenizing, stopwords serta normalization.
4. 3.2.1 Case Folding
5. Case Folding merupakan sebuah metode untuk menyesuaikan huruf besar menjadi huruf kecil atau dalam bentuk yang konsisten. Proses ini memudahkan pencarian, mengingat tidak semua dokumen teks mengikuti aturan penggunaan huruf kapital yang konsisten .
6.

Tabel 2. Contoh hasil pengumpulan data

Sebelum **Sesudah**
Cukup bagus apk nya bpjs cukup bagus apk nya bpjs
Sangat cepat dan mudah **sangat cepat dan mudah**
Bintang setengah ada ga? bintang setengah ada ga?
Cucokk mantep cucokk mantep
Mantap gannn mantap gannn

1.
2. 3.2.2 Tokenizing
3. Proses ini memcah kalimat menjadi kata-kata dengan membaginya berdasarkan spasi antar kata.

Tabel 3. Contoh data yang telah melalui proses tokenisasi

Sebelum **Sesudah**
cukup bagus apk nya bpjs [cukup, bagus, apk, nya, bpjs]
sangat cepat dan mudah [sangat, cepat, dan, mudah]
bintang setenga ada ga? [bintang, setengah, ada, ga]
cucokk mantep [cucokk, mantep]
mantap gannn [mantap, gannn]

1.
2. 3.2.3 Stopword Removal
3. Stopword removal, juga dikenal sebagai filtering, merupakan Langkah pemilihan kata-kata utama yang muncul dari hasil tokenisasi. Pada Langkah ini, dipilih kata-kata yang mewakili isi dokumen .

Tabel 4. Contoh hasil stopwords

Sebelum **Sesudah** [cukup, bagus, apk, nya, _bpjs]_[bagus, apk, bpjs].
[sangat, cepat, dan, _mudah]_[cepat, mudah]
[bintang, _setengah, ada, ga] [bintang, ga]
[cucokk, mantep] [cucok, mantep]
[mantap, gannn] [mantap]

3.2.4 Normalization

1. Normalisasi dilakukan dengan cara menghapus **kata-kata yang memiliki imbuhan atau tambahan**, serta memperbaiki kata yang salah ejaannya menjadi yang benar .

Tabel 5. Contoh proses normalisasi

Sebelum **Sesudah**
[bagus, apk, bpjs] bagus aplikasi bpjs
[cepat, mudah] cepat mudah
[bintang, ga] bintang ga
[cucok, mantep] cucok mantep

[mantap] mantap

1.

2. 3.3 Pembobotan atau Pelabelan Kata

3. 3.3.1 Term Frequency (TF)

Proses ini menghitung frekuensi kemunculan kata dalam kumpulan data. Frekuensi data dapat dihitung menggunakan rumus:

$$TF(t,d) = \frac{f_{t,d}}{\sum_{t \in D} f_{t,d}} \quad (1)$$

Keterangan :

t = term (kata tertentu yang sedang diproses frekuensinya)

d = dokumen (dokumen dimana kata itu muncul)

1.

2. 3.3.2 Inverse Document Frequency (IDF)

Proses ini menghitung jumlah dokumen yang mengandung istilah kunci tertentu pada dataset. Jumlah dokumen yang berisi kata kunci dapat dihitung dengan rumus berikut: $IDF(t,d) = \frac{1}{\log(D/d)}$ (2)

Keterangan :

N = Jumlah semua dokumen

Df = Jumlah kata pada dokumen

3.3.3 Confusion Matrix

Confusion matrix merupakan sebuah matriks 3x3 yang bertujuan untuk mengevaluasi kinerja model dengan membandingkan prediksi model terhadap kata actual. Confusion matrix menghasilkan nilai akurasi, presisi, recall dan F1 Score dihitung dan ditampilkan dalam format presentase untuk menunjukkan performa model [17]. Tabel berikut menunjukkan confusion matrix:

Tabel 6. Confusion Matrix

Data	Aktual	Data Prediksi	
	Kritik	Saran	Pertanyaan
Kritik	True Kritik	True Saran	False Pertanyaan
Saran	False Kritik	True Saran	False Pertanyaan
Pertanyaan	False Kritik	False Saran	True Pertanyaan

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

Keterangan :

TK = True Kritik

FK = False Kritik

FP = False Pertanyaan

TP = True Pertanyaan

FS = False Sara

TS = True Saran

3.4 Klasifikasi Support Vector Machine (SVM)

Support vector machine merupakan algoritma klasifikasi dimana sangat populer dalam keperluan pengelompokan dan analisis data. Support Vector Machine digunakan untuk analisis data dan pengenalan pola. Salah satu dasar Support Vector Machine, linear classifier, hanya berfokus pada dua kelas, namun umumnya masalah yang ditemukan di dunia nyata bersifat multiclass. Kemudian dikembangkan untuk menangani masalah non-linear yang diselesaikan menggunakan trik kernel dan memetakan data ke ruang yang lebih besar. Algoritma SVM menggunakan hyperplane untuk membedakan antara senitmen positif dan negatif. Metode perhitungan Support Vector Machine:

$$(w \cdot x_i) + b = 0 \quad (4)$$

Data x_i , yang termasuk kelas -1 dapat dirumuskan sebagai berikut

$$(w \cdot x_i + b) \leq -1, y_i = -1 \quad (5)$$

Rumus untuk data x_i pada kelas +1 adalah sebagai berikut

$$(w \cdot x_i + b) \geq 1, y_i = 1 \quad (6)$$

3.5 Naïve Bayes

Algoritma yang diterapkan pada proses pengelompokan ini merupakan Naïve Bayes, tepatnya varian Naïve Bayes Multinomial, yang memang banyak digunakan dalam pemrosesan teks dan pengelompokan data berbasis kata. Naïve Bayes bekerja berdasarkan prinsip probabilitas, di mana Naïve Bayes memprediksi kelas dari sebuah data berdasarkan kemungkinan (probabilitas) suatu kata muncul dalam setiap kelas yang telah ditentukan. Cara perhitungan dengan Naïve Bayes :

$$P(X) = \prod_{i=1}^n P(x_i) \quad (7)$$

Keterangan :

X = Data kelas tidak diketahui

H = Data Hipotesis X adalah kelas yang terpisah

$P(H|X)$ = Probabilitas hipotesis kondisional H

$P(H)$ = Probabilitas hipotesis H

$P(X|H)$ = Probabilitas X

IV. Hasil dan Pembahasan

1. Untuk mendukung penelitian ini, sebanyak 6.000 ulasan dari user aplikasi MitraShopee mengambil langsung dari platform Google Play Store menggunakan Google Play Scraper digunakan sebagai sumber data utama. Seluruh data kemudian diproses dan dianalisis menggunakan bahasa pemrograman Python. Proses pengolahan ini dilakukan di Google Colab karena lebih praktis dan fleksibel, terutama untuk kebutuhan analisis berbasis machine learning.

2.

3. 4.1 Pengumpulan Data

4. Total ada sekitar 6.000 ulasan yang berhasil dikumpulkan dan digunakan sebagai bahan utama untuk dianalisis. Ulasan-ulasan ini mencerminkan

berbagai pengalaman, keluhan, pertanyaan, hingga saran yang diberikan langsung oleh para pengguna aplikasi.

Tabel 7. Isi dataset

No	Rating	Review Text
0	5	bagus, harga kompetitif semoga...
1	5	saangat baik dan bagus
2	5	bisa lebih cepat waktu transaksi...
3	5	mantul
4	5	kemudahan dan kenyamanan...

1. Pengambilan data yang terbaru, agar informasi yang didapat bersifat beragam dan tidak bias terhadap waktu atau tren tertentu. Semua data yang sukses dikumpulkan kemudian disimpan sebagai file CSV, sehingga lebih mudah untuk diproses lebih lanjut dalam tahap analisis.

2.

3. 4.2 Preprocessing Data

4. Proses Preprocessing sebagai langkah pembersihan dan menyiapkan data supaya lebih dapat langsung dimanfaatkan pada pemodelan machine learning. Tahapan ini penting karena data teks mentah biasanya masih mengandung banyak unsur yang tidak relevan atau menyulitkan saat analisis dilakukan.

5. 4.2.1 Case Folding

6. Semua karakter huruf dalam teks ulasan dikonversi ke huruf kecil (lowercase), dalam menyamakan bentuk word, sehingga kata dianggap sama dan tidak diproses dua entitas berbeda.

Tabel 8. Hasil Case Folding

No	Before	After
0	bagus harga kompetitif semoga...	[bagus, harga, kompetitif, ...]
1	Saangat baik dan bagus	[saangat, baik, dan, bagus]
2	Bisa lebih cepat waktu...	[bisa, lebih, cepat, waktu, ...]
3	mantul	[mantul]
4	kemudahan dan kenyamanan...	[kemudahan, dan, kenyamanan...]

1. Pada tahap case folding, seluruh huruf dalam teks ulasan diubah menjadi huruf kecil. Dengan tujuan agar system bisa membaca kata secara konsisten, misalnya “Bagus” dan “bagus” dianggap sama dan tidak dihitung sebagai dua kata berbeda. Ini penting untuk menghindari perhitungan ganda akibat perbedaan penulisan huruf kapital, sehingga analisis data menjadi lebih akurat dan konsisten.

2.

3. 4.2.2 Tokenizing

4. Proses memecah teks Panjang menjadi potongan-potongan kata atau token. Tahap ini untuk mempersiapkan data agar bisa dianalisis per kata.

Tabel 9. Tokenizing

No	Before	After
0	bagus harga kompetitif semoga...	[bagus, harga, kompetitif, ...]
1	Saangat baik dan bagus	[saangat, baik, dan, bagus]
2	Bisa lebih cepat waktu...	[bisa, lebih, cepat, waktu, ...]
3	mantul	[mantul]
4	kemudahan dan kenyamanan...	[kemudahan, dan, kenyamanan...]

1. Proses tokenisasi memudahkan system untuk menganalisis dan mengelola data dalam bentuk unit kata yang lebih terstruktur dan dapat diproses lebih lanjut.

2.

3. 4.2.3 Stopword Removal

4. Tahapan ini berfokus pada penghapusan kata-kata umum yang tidak membawa makna penting dalam analisis dihapus dari teks, untuk mengurangi kebisingan dalam data dan memfokuskan analisis pada kata-kata yang benar-benar bermakna.

5.

Tabel 10. Hasil stopwords removal

No	Before	After
0	[bagus, harga, kompetitif, ...]	[bagus, harga, kompetitif, ...]
1	[saangat, baik, dan, bagus]	[saangat, bagus]
2	[bisa, lebih, cepat, waktu, ...]	[cepat, waktu, transaksinya, ...]
3	[mantul]	[mantul]
4	[kemudahan, dan,]	[kemudahan, kenyamanan, ...]

1. Dengan menghapus kata-kata ini, fokus analisis bisa diarahkan pada kata-kata penting yang benar-benar merepresentasikan isi atau maksud dari ulasan pengguna.

2.

3. 4.2.4 Normalization

4. Langkah normalisasi bertujuan menyamakan bentuk penulisan kata-kata yang memiliki makna yang sama, baik karena singkatan, typo, maupun bahasa gaul akan dinormalisasikan menjadi satu bentuk standar.

Tabel 11. Hasil normalization

No	Before	After
0	[bagus, harga, kompetitif, ...]	[bagus, harga, kompetitif, ...]
1	[saangat, baik, dan, bagus]	[saangat, bagus]
2	[bisa, lebih, cepat, waktu, ...]	[cepat, waktu, transaksinya, ...]

- 3 [mantul] [mantul]
- 4 [kemudahan, dan, ...] [kemudahan, kenyamanan, ...]

1. Normalisasi sangat penting agar sistem tidak keliru dalam memahami makna kata yang sebenarnya dan bisa mengelompokkan ulasan secara lebih tepat.

- 2.
3. 4.3 Visualisasi Wordcloud Sentiment
- 4.
5. Gambar SEQ Gambar * ARABIC 4. Hasil Wordcloud
6. Wordcloud ini menampilkan kata-kata yang dominan dalam ulasan user aplikasi MitraShopee di Google Play Store. kata-kata yang tampil lebih besar menunjukkan bahwa kata tersebut lebih sering digunakan oleh pengguna. Dari visualisasi tersebut. Terlihat jelas bahwa kata “mitra”, “shopee”, dan “aplikasi” mendominasi, yang menunjukkan bahwa pengguna secara aktif menyebutkan nama aplikasi dalam ulasan mereka. Hal ini bisa diartikan bahwa pengguna memiliki fokus pada kualitas dan pengalaman langsung saat menggunakan aplikasi.

- 7.
8. Gambar SEQ Gambar * ARABIC 5. Frekuensi kata
9. Beberapa kata lain yang cukup mencolok seperti “transaksi”, “mudah”, “bagus”, “untung”, “murah”, dan “cepat” mengindikasikan bahwa banyak pengguna merasa puas dengan kemudahan dan keuntungan dalam menggunakan MitraShopee, terutama dari segi transaksi dan harga. Kata-kata positif seperti “mantap”, “alhamdulillah”, dan “puas” juga muncul cukup sering, yang bisa mencerminkan tingkat kepuasan pengguna terhadap layanan yang diberikan.
10. Kemunculan kata seperti “username”, “masuk”, atau “saldo” juga bisa menjadi sinyal adanya masalah teknis atau kendala login yang cukup sering dibicarakan, sehingga dapat menjadi perhatian khusus untuk pengembangan aplikasi kedepannya. Wordcloud memberikan gambaran cepat dan intuitif mengenai topik yang paling sering dibahas pengguna serta nada umum dari ulasan mereka apakah cenderung positif, netral, atau mengandung keluhan.

- 11.
 12. 4.4 Pelabelan Data
- Tabel 12. Hasil normalization

No	Review Text	Label
0	saya kasih bintang	Lainnya
1	mohon struk untuk admin bank dan admin loket...	Saran
2	ni yang nawarin gk jelas langsung ilang biking gw...	Kritik
3	mengapa saldo mitra di tutup...	Pertanyaan
...	...	...
6000	sangat rekomendasi	Lainnya

1. Pada tahap selanjutnya adalah merupakan pelabelan data terhadap ulasa yang telah dikumpulkan dari Google Play Store pada aplikasi MitraShopee. Dari total 6.000 data, setiap ulasan dianalisis secara otomatis menggunakan Google Colab. Dengan memanfaatkan pendekatan berbasis aturan (rule-based-classification), setiap ulasan yang telah dibersihkan dianalisis menggunakan kata kunci tertentu yang telah didefinisikan sebelumnya untuk mengidentifikasi konteks dari isi ulasan dan diklasifikasikan ke dalam tiga kategori utama, yaitu: kritik, saran, dan pertanyaan. Label lainnya diberikan pada ulasan yang bersifat umum, pujian, atau tidak mengandung elemen dari ketiga kategori utama.
2. Dalam membedakan antara kategori kritik, saran, dan pertanyaan pada ulasan pengguna, terdapat beberapa pendekatan yang digunakan baik secara linguistic maupun teknis. Kritik umumnya ditandai dengan ekspresi ketidakpuasan yang mengandung kata-kata bernada negatif seperti “gagal”, “lemot”, atau “jelek”. Label saran sering kali muncul dalam bentuk kalimat usulan atau harapan dengan kata-kata seperti “sebaiknya”, “lebih baik”, “mohon”, atau “diharapkan”, yang terlihat netral hingga positif. Pada label pertanyaan memiliki struktur yang lebih eksplisit, biasanya dimulai dengan kata tanya seperti “apa”, “bagaimana”, atau “kenapa”. Label lainnya digunakan untuk mengelompokkan ulasan yang tidak termasuk ke dalam kategori kritik, saran, atau pertanyaan. Label ini mencakup komentar seperti “terima kasih” atau “bagus”, pernyataan yang singkat atau ambigu, hingga komentar yang tidak relevan. Dalam pendekatan Natural Language Processing (NLP) model klasifikasi dapat mempelajari pola-pola ini melalui representasi teks seperti TF IDF, frekuensi kata kunci, jenis kata yang digunakan, serta nada atau emosi yang terkandung dalam ulasan [22].
3. Dari hasil pelabelan data ulasan terdapat 254 ulasan yang tergolong kritik, 372 saran, 728 pertanyaan, dan 3711 label lainnya. Terlihat bahwa mayoritas ulasan pengguna masih termasuk dalam kategori lainnya, yang berarti banyak pengguna lebih cenderung memberikan testimoni umum atau pujian tanpa menyampaikan masukan spesifik. Namun demikian, ditemukan juga sejumlah ulasan yang memuat kritik dan saran konstruktif, serta beberapa pertanyaan yang menandakan adanya kebutuhan akan informasi lebih lanjut dari pihak pengembang [23].

- 4.
- 5.
- 6.
- 7.
8. 4.5 Evaluasi Model

Tabel 13. Evaluasi model SVM

	Precision	Recall	F1-Score	Support
Kritik	0.94	0.60	0.74	53
Saran	0.76	0.93	0.84	143
Pertanyaan	0.86	0.72	0.78	75
Accuracy	0.81 271			
Macro avg	0.85	0.75	0.79	271
Weighted avg	0.82	0.81	0.80	271

1. Pembagian data dilakukan menggunakan pendekatan Hold-Out Validation dengan proporsi 80% untuk data latih dan 20% untuk data uji [24]. Dari total 1.354 ulasan yang telah diklasifikasikan ke dalam kategori kritik, saran, dan pertanyaan, sebanyak 271 data dialokasikan sebagai data uji.

Pemilihan proporsi ini didasarkan pada praktik umum dalam bidang machine learning, di mana 20% data uji dianggap cukup representative untuk mengukur performa model secara objectif tanpa mengurangi jumlah data latih yang diperlukan untuk membangun model secara optimal.

2. Berdasarkan pengelompokan menggunakan SVM menghasilkan akurasi sebesar 80.8%, yang berarti bahwa dari seluruh data uji yang dipakai (sebanyak 271 ulasan), sekitar 81% dapat diklasifikasikan secara tepat ke dalam tiga kategori yang telah ditentukan, yaitu kritik, saran, dan pertanyaan. Secara lebih rinci, performa model dapat dijelaskan melalui metrik evaluasi.

3. Kategori kritik memiliki nilai precision sebesar 0.94, yang mengindikasikan bahwa model memiliki kinerja baik dalam mengenali ulasan yang sebenarnya tergolong kritik. Namun, nilai recall untuk kategori ini berada di angka 0.60, yang artinya masih ada sejumlah kritik yang tidak terdeteksi dengan baik oleh model [25]. Kategori saran menunjukkan performa paling stabil, dengan nilai precision 0.76 dan recall 0.93. Ini berarti Sebagian besar ulasan berupa saran berhasil diklasifikasikan dengan baik. Nilai F1-Score yang tinggi, yakni 0.84, memperkuat bahwa model cukup konsisten dalam mengenali saran dari pengguna. Kategori pertanyaan memperoleh nilai precision 0.86 dan recall 0.72 ini mengindikasikan bahwa model dapat mengidentifikasi pertanyaan dengan cukup baik, meskipun ada beberapa pertanyaan yang masih keliru diklasifikasikan [26].

4. Performa model Support Vector Machine (SVM) dapat dikatakan cukup optimal dalam memahami dan mengklasifikasikan bentuk komunikasi pengguna dalam ulasan aplikasi MitraShopee. Hasil ini juga menunjukkan bahwa pendekatan klasifikasi berbasis machine learning mampu membantu dalam menyaring masukan dari pengguna secara otomatis dan efisien.

5.

6. Gambar SEQ Gambar * ARABIC 6. Confusion Matrix SVM

7. Pada gambar di atas terlihat bahwa model SVM cukup mampu membedakan antara kategori ulasan. Untuk kelas kritik, dari total 53 data. Sebanyak 32 data berhasil diklasifikasikan dengan benar, sementara 20 data salah prediksi sebagai pertanyaan, dan 1 data salah dikira saran. Ini menunjukkan adanya sedikit kebingungan model dalam membedakan kritik dan pertanyaan, meski sebagian besar prediksi sudah tepat. Untuk kelas saran, performa model sangat baik dengan 54 data terklasifikasi benar, dan hanya 21 yang keliru diprediksi sebagai pertanyaan [27]. Tidak ada data saran yang salah dikira kritik, yang menunjukkan konsistensi prediksi pada kelas ini. Pada kelas pertanyaan, dari total 143 data, model berhasil memprediksi dengan sangat baik sebanyak 133 data terklasifikasi benar, hanya 10 sisanya tersebar salah ke kelas lain.

8.

9.

Tabel 14. Evaluasi model Naïve Bayes

	Precision	Recall	F1-Score	Support
Kritik	1.00	0.08	0.14	53
Saran	0.54	1.00	0.70	143
Pertanyaan	1.00	0.05	0.10	75

Accuracy0.56271

Macro avg0.850.380.32271

Weighted avg0.760.560.43271

1. Pada tahap evaluasi model klasifikasi melalui penerapan algoritma Naïve Bayes, akurasi yang diperoleh mencapai 55.72% dari total 271 data uji. Hasil ini mengindikasikan bahwa lebih dari setengah data berhasil diklasifikasikan dengan benar oleh model, namun masih terdapat kelemahan signifikan dalam mengenali Sebagian kelas. Berdasarkan kategori label, model menunjukkan performa terbaik pada kelas saran. Ini berarti model sangat efektif dalam mengenali hampir semua data yang tergolong sebagai saran. Namun, sebaliknya performa model untuk kategori kritik dan pertanyaan menunjukkan hasil yang kurang memuaskan. Meskipun precision untuk kedua kategori ini tinggi (1.00), hal ini menunjukkan bahwa model hanya memprediksi sangat sedikit data sebagai kritik dan pertanyaan, namun prediksi tersebut benar. Model tidak sensitive dalam mengenali data kritik dan pertanyaan secara menyeluruh [28].

2.

3. Gambar SEQ Gambar * ARABIC 7. Confusion matrix Naïve Bayes

4. Pada gambar di atas terlihat bahwa model ini kurang mampu membedakan antara kategori ulasan. Untuk kelas kritik, hanya 4 data yang benar, dan 49 sisanya justru salah dikira sebagai pertanyaan. Pada kelas saran juga hanya 4 yang benar, sedangkan 71 juga diprediksi salah ke kelas pertanyaan. Meskipun terlihat seolah-olah akurat di kelas pertanyaan (143 benar), tapi ini bukan karena modelnya benar. Akurasinya hanya 56%, dan recall untuk kritik dan saran sangat rendah. Performa Naïve Bayes kurang cocok untuk klasifikasi ulasan MitraShopee yang kelasnya tidak seimbang [29].

5.

6. V. KESIMPULAN

7.

8. Penggunaan algoritma klasifikasi SVM dan Naïve Bayes telah berhasil diimplementasikan dalam mengelompokkan ulasan pengguna aplikasi MitraShopee ke dalam tiga kategori utama, yaitu kritik, saran, dan pertanyaan. Pada penelitian ini menggunakan data training 271 data ulasan, dengan tujuan sebagai upaya mengetahui model yang paling efektif dalam mengolah dan memahami masukan dari pengguna secara otomatis. Berdasarkan hasil evaluasi algoritma SVM mencatatkan performa terbaik dengan capaian akurasi sebesar 81%, precision 94% pada kelas kritik, recall tertinggi 93% pada kelas saran. Hal ini menunjukkan bahwa SVM cukup seimbang dalam mendeteksi tiap kategori dan mampu mengurangi kesalahan klasifikasi secara signifikan. Algoritma Naïve Bayes hanya mencatatkan akurasi sebesar 56%, dengan precision tinggi tetapi recall yang sangat rendah, terutama pada kelas kritik dan pertanyaan. Klasifikasi ini tampak bias terhadap satu kategori, yaitu pertanyaan, yang menyebabkan penurunan kualitas prediksi secara keseluruhan.

9. Dapat disimpulkan bahwa pemilihan algoritma dan proses preprocessing data sangat krusial dalam klasifikasi teks. Setiap model memiliki sensitivitas yang berbeda terhadap karakteristik data, sehingga penyesuaian metode pra-proses seperti tokenisasi, pembersihan teks, dan representasi fitur menjadi hal yang tidak boleh diabaikan. Diharapkan hasil penelitian ini dapat menjadi landasan untuk pengembangan sistem klasifikasi otomatis kedepannya, khususnya untuk menganalisis masukan pengguna dalam platform digital seperti MitraShopee secara lebih cepat dan tepat sasaran.