



ISIAN LAPORAN KEMAJUAN PENELITIAN

Petunjuk: Pengusul hanya diperkenankan mengisi di tempat yang telah disediakan sesuai dengan petunjuk pengisian dan tidak diperkenankan melakukan modifikasi template atau penghapusan di setiap bagian.

JUDUL

Tuliskan judul penelitian

Klasifikasi Aspek Layanan pada Ulasan E-Commerce Menggunakan *Machine Learning*

RINGKASAN

Isian ringkasan penelitian tidak lebih dari 300 kata yang berisi urgensi, tujuan, metode, dan luaran yang ditargetkan.

Perkembangan *e-commerce* di Indonesia meningkatkan persaingan antar platform, sehingga pemahaman terhadap kepuasan pelanggan menjadi sangat penting. Ulasan pelanggan merupakan sumber data yang kaya untuk mengevaluasi kualitas layanan, namun sebagian besar penelitian masih berfokus pada analisis sentimen yang bersifat umum dan belum mampu mengidentifikasi aspek layanan secara spesifik. Selain itu, pendekatan machine learning yang digunakan masih terbatas pada representasi fitur sederhana serta belum mempertimbangkan permasalahan ketidakseimbangan data (*imbalanced data*), sehingga menghasilkan model yang kurang optimal. Penelitian ini bertujuan untuk mengklasifikasikan ulasan *e-commerce* berdasarkan aspek layanan menggunakan metode machine learning dengan integrasi optimasi fitur dan teknik penanganan data tidak seimbang. Tahapan penelitian meliputi: (1) pengumpulan dan pelabelan data ulasan *e-commerce* (Tokopedia dan Shopee), (2) pra-pemrosesan teks, (3) ekstraksi fitur menggunakan TF-IDF dengan variasi unigram dan bigram, (4) penanganan *imbalanced data* menggunakan metode SMOTE, (5) pembangunan model menggunakan beberapa algoritma machine learning, serta (6) evaluasi dan perbandingan kinerja model menggunakan metrik klasifikasi yang relevan. Hasil yang diharapkan adalah model klasifikasi yang mampu mengidentifikasi aspek layanan secara lebih spesifik dan akurat, sehingga dapat memberikan insight bagi pengelola *e-commerce* dalam meningkatkan kualitas layanan. Luaran wajib penelitian ini adalah publikasi pada JITECS (Journal of Information Technology and Computer Science) Vol. 11 No. 3 Tahun 2026 (Sinta 2), serta luaran tambahan berupa Hak Kekayaan Intelektual (HAKI) berbentuk basis data yang merupakan dataset penelitian

KATA KUNCI

Isian kata kunci maksimal 5 kata yang dipisahkan dengan tanda titik koma (;)

Klasifikasi Aspek Layanan; Ulasan *E-Commerce*; *Machine Learning*; TF-IDF; *Imbalanced Data*.

PENDAHULUAN

Pendahuluan penelitian tidak lebih dari 1000 kata yang terdiri dari:

1. Latar belakang dan rumusan permasalahan yang akan diteliti
2. Pendekatan pemecahan masalah
3. *State of the art* dan kebaruan
4. Peta jalan (*road map*) penelitian 5 tahun

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver

1. LATAR BELAKANG DAN RUMUSAN MASALAH

Perkembangan e-commerce di Indonesia meningkatkan persaingan antar platform, sehingga kepuasan pelanggan menjadi faktor utama. Berbagai penelitian menunjukkan bahwa kepuasan dipengaruhi oleh aspek layanan seperti kualitas pelayanan, ketepatan pengiriman, harga, dan kemudahan penggunaan platform (1,2). serta faktor *responsiveness*, *reliability*, dan kejelasan informasi harga (3,4). Hal ini menunjukkan pentingnya evaluasi layanan secara komprehensif.

Namun, sebagian besar penelitian masih berfokus pada analisis sentimen (positif, negatif, netral) yang bersifat umum dan belum mampu mengidentifikasi aspek layanan secara spesifik. Padahal, ulasan pelanggan mengandung informasi penting yang belum dimanfaatkan optimal. Selain itu, pendekatan *machine learning* masih terbatas pada penggunaan TF-IDF tanpa optimasi fitur dan tanpa mempertimbangkan *imbalanced data*, sehingga model cenderung bias dan kurang representative (5).

Oleh karena itu, penelitian ini mengusulkan klasifikasi berbasis aspek layanan pada ulasan e-commerce menggunakan *machine learning* dengan optimasi fitur dan penanganan data tidak seimbang. Penelitian ini diharapkan menghasilkan model yang lebih akurat serta memberikan *insight* spesifik untuk mendukung peningkatan kualitas layanan *e-commerce*.

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah:

- a. Bagaimana performa algoritma *machine learning* dalam mengklasifikasikan ulasan e-commerce berdasarkan aspek layanan?
- b. Bagaimana pengaruh variasi fitur (TF-IDF unigram dan bigram) terhadap kinerja model klasifikasi?
- c. Bagaimana pengaruh teknik penanganan *imbalanced data* terhadap hasil klasifikasi?
- d. Algoritma *machine learning* manakah yang memberikan performa terbaik dalam klasifikasi aspek layanan?

Berdasarkan latar belakang di atas, maka tujuan penelitian ini adalah:

- a. Mengklasifikasikan ulasan e-commerce berbasis aspek layanan.
- b. Menganalisis pengaruh TF-IDF dengan variasi n-gram terhadap kinerja model.
- c. Mengevaluasi dampak penanganan imbalanced data (SMOTE).
- d. Membandingkan performa beberapa algoritma *machine learning*.

2. PENDEKATAN PEMECAHAN MASALAH

Pendekatan pemecahan masalah dalam penelitian ini dilakukan melalui dua tahapan utama, yaitu pendekatan komputasional dan analitis evaluatif.

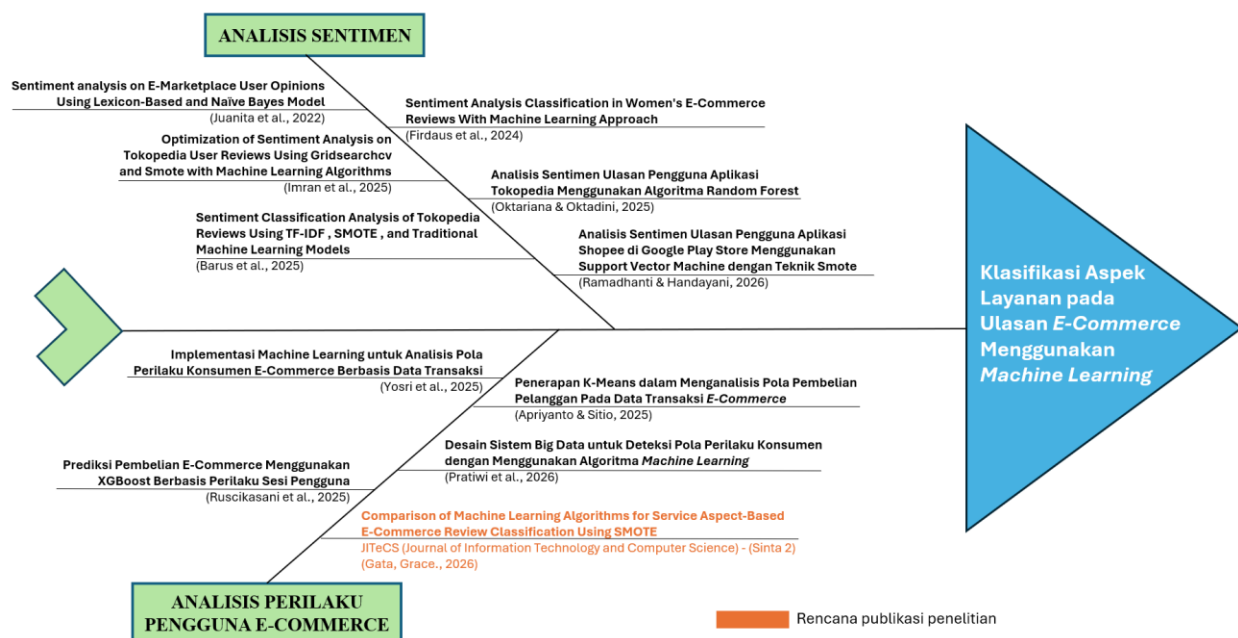
Pendekatan komputasional dilakukan dengan mengumpulkan dan mengolah data ulasan *e-commerce* melalui tahapan prapemrosesan, kemudian direpresentasikan menggunakan TF-IDF (unigram dan bigram). Selanjutnya, data diklasifikasikan ke dalam aspek layanan menggunakan

beberapa algoritma machine learning, dengan optimasi melalui seleksi fitur Chi-Square dan penanganan imbalanced data menggunakan SMOTE.

Pendekatan analitis evaluatif dilakukan dengan membandingkan kinerja model menggunakan metrik accuracy, precision, recall, dan F1-score untuk menentukan model terbaik serta menganalisis pengaruh metode yang digunakan.

3. STATE OF THE ART DAN KEBARUAN

Pada Gambar 1 merupakan *state-of-the art* pada usulan penelitian pengusul terhadap penelitian sebelumnya.



Gambar 1. *State-of-theart* Usulan Penelitian

Secara detail, Tabel 1 menampilkan literatur *review* dari kumpulan publikasi pada topik analisis sentimen yang terdapat pada Gambar 1. Beberapa pendekatan *Machine Learning* (ML) menggunakan beberapa algoritma Naïve Bayes (NB), Support Vector Machine (SVM), KNN (*K-Nearest Neighbors*), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), Gradient Boosting (GB), AdaBoost (AB), XGBoost (XGB).

Tabel 1. Literatur *Review* pada Topik Analisis Sentimen pada ulasan *E-Commerce*

No	Penulis/Tahun	Metode ML	Kelebihan	Keterbatasan
1	Ardiyanto dkk (6) / 2022	Lexicon-Based (TextBlob) + NB	Kombinasi leksikon dan ML	• Hanya fokus sentimen, • Belum klasifikasi aspek layanan • Terbatas pada fitur sederhana
2	Firdaus dkk (7) / 2024	NB, SVM, KNN + ROS/SMOTE	Perbandingan beberapa model +penanganan	
3	Imran dkk (8) / 2025	LR, RF, XGB + SMOTE		

No	Penulis/Tahun	Metode ML	Kelebihan	Keterbatasan
4	Barus dkk (9) / 2025	NB, LR, SVM, DT, RF + TF-IDF + SMOTE	<i>imbalanced</i> data	• Hanya satu algoritma • Belum klasifikasi aspek layanan
5	Oktariana dkk (10) / 2025	RF + TF-IDF	Akurasi tinggi dan efektif untuk data besar	
6	Ramadhanti dkk (11) / 2026	SVM + TF-IDF + SMOTE	SMOTE meningkatkan akurasi secara signifikan	

Berdasarkan Tabel 1, analisis ulasan e-commerce umumnya masih berfokus pada klasifikasi sentimen menggunakan berbagai algoritma machine learning. Beberapa penelitian juga telah mengombinasikan teknik ekstraksi fitur menggunakan TF-IDF serta menerapkan metode penanganan *imbalanced* data seperti SMOTE dan ROS, yang terbukti mampu meningkatkan performa model secara signifikan.

Sebagian besar penelitian membandingkan beberapa algoritma *machine learning* dan menunjukkan bahwa model seperti Random Forest dan SVM memiliki performa yang lebih baik. Namun, penelitian tersebut masih terbatas pada analisis sentimen umum dan belum mengkaji klasifikasi berbasis aspek layanan.

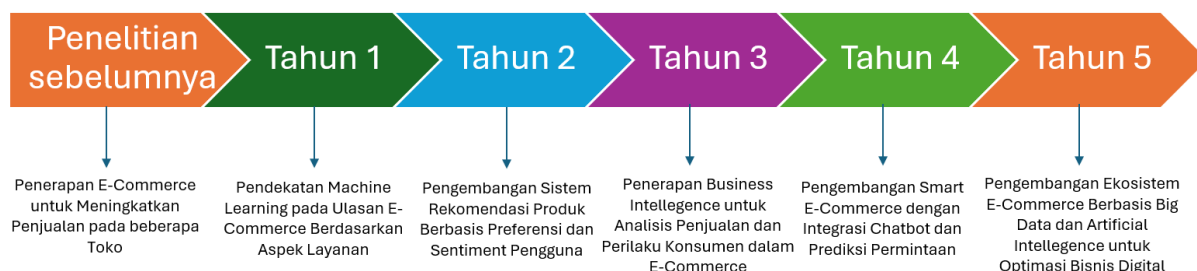
Tabel 2. Literatur *Review* pada Topik Analisis Perilaku Pengguna *E-Commerce*

No	Penulis/Tahun	Metode	Kelebihan	Keterbatasan
1	Yosri dkk (12) / 2025	RF, GB, DT, AB, KNN + SMOTE	Perbandingan beberapa model + penanganan <i>imbalanced</i> data	• Fokus data transaksi, • Tidak berbasis teks, • Tidak analisis ulasan/aspek layanan
2	Apriyanto dkk (13) / 2025	K-Means Clustering	mengidentifikasi segmentasi pelanggan + pola pembelian	
3	Ruscikasani dkk (14) / 2025	XGBoost	mengidentifikasi faktor penting perilaku pengguna dan prediksi konversi	
4	Pratiwi dkk (15) / 2026	Klasifikasi & Klastering + Big Data	Mengolah data besar dan integrasi multi-sumber untuk analisis lebih komprehensif	
4	Klasifikasi Aspek Layanan pada Ulasan <i>E-Commerce</i> Menggunakan <i>Machine Learning</i> (usulan penelitian)	NB, SVM, LR, RF, GB	Perbandingan beberapa model dan penanganan <i>imbalanced</i> + ekstraksi fitur dan seleksi fitur	Analisis teks dan aspek layanan <i>e-commerce</i>

Berdasarkan Tabel 2, penelitian perilaku konsumen e-commerce umumnya menggunakan data transaksional dengan pendekatan machine learning seperti Random Forest, XGBoost, K-Means, serta kombinasi clustering dan classification, termasuk penerapan SMOTE dan feature engineering. Namun, sebagian besar penelitian masih berfokus pada data numerik dan belum memanfaatkan data teks ulasan pelanggan, serta belum mengarah pada identifikasi aspek layanan secara spesifik. Oleh karena itu, berdasarkan Tabel 1 dan 2, penelitian ini mengusulkan klasifikasi aspek layanan pada ulasan e-commerce menggunakan machine learning.

4. PETA JALAN PENELITIAN

Pada Gambar 2 merupakan peta jalan penelitian dari pengusul.



Gambar 2. Peta Jalan Penelitian

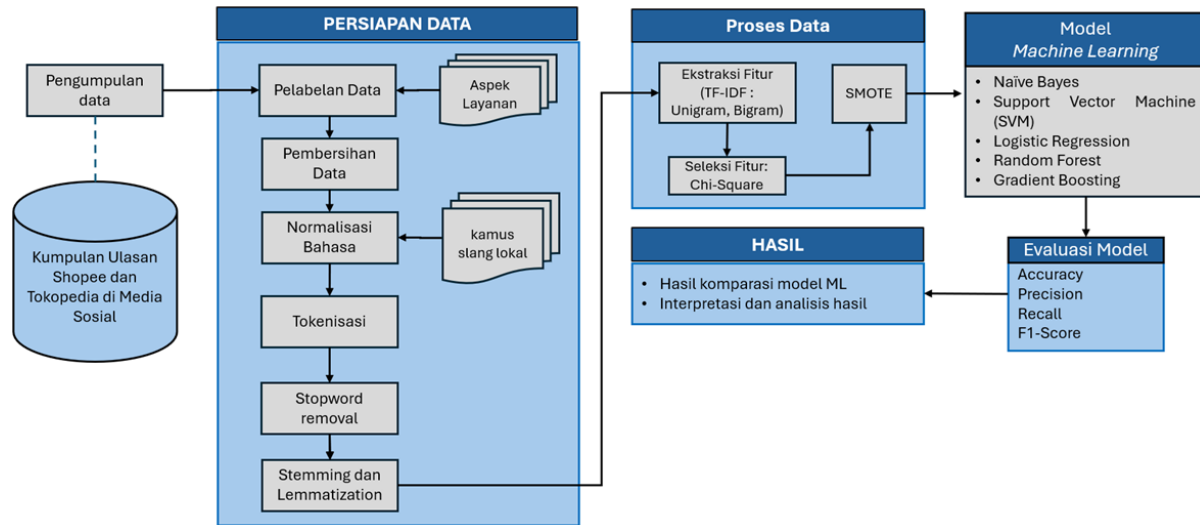
Pada Gambar 2 menampilkan peta jalan penelitian pengusul sesuai dengan riwayat penelitian sebelumnya. Berawal dari tahun 2025, pengusul melakukan penelitian pada topik *E-Commerce* dengan menerapkan aplikasi *e-commerce* untuk meningkatkan penjualan pada beberapa toko seperti toko alat tulis dan aksesoris (16) dan toko penjualan kopi (17). Berdasarkan penelitian tahun 2025, pada proposal ini mengusulkan analisis ulasan *e-commerce* berdasarkan aspek layanan menggunakan pendekatan machine learning. Pada tahun kedua berencana untuk pengembangan sistem rekomendasi produk berbasis preferensi dan sentimen pengguna menggunakan hasil penelitian pada proposal ini.

Berdasarkan penelitian pada tahun kedua, maka pada tahun ketiga melakukan penelitian tentang penerapan *business intelligence* untuk analisis penjualan dan perilaku konsumen dalam e-commerce. Sedangkan pada tahun keempat melakukan pengembangan *smart e-commerce* dengan integrasi chatbot dan prediksi permintaan, melanjutkan penelitian pada tahun ketiga. Pada tahun kelima melakukan pengembangan ekosistem *e-commerce* berbasis big data dan *artificial intelligence* untuk optimasi bisnis digital.

METODE

Isian metode atau cara untuk mencapai tujuan yang telah ditetapkan ditulis tidak melebihi 1000 kata. Bagian ini dapat dilengkapi dengan diagram alir penelitian yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Metode penelitian harus dibuat secara utuh dengan penahapan yang jelas, mulai dari awal bagaimana proses dan luarannya, dan indikator capaian yang ditargetkan.

Pada Gambar 3 menampilkan tahapan penelitian yang diusulkan:



Gambar 3. Tahapan Klasifikasi Aspek Layanan Pada Ulasan *E-Commerce* Menggunakan *Machine Learning*

Berikut ini adalah penjelasan detail dari tahapan pada Gambar 3:

1. Pengumpulan Data

Tahap ini dilakukan dengan mengumpulkan data ulasan pengguna e-commerce melalui proses *crawling*. Penelitian berfokus pada ulasan pelanggan Shopee dan Tokopedia yang diambil dari Google Play Store dan media sosial X (Twitter). Tabel 3 menampilkan contoh ulasan dari kedua platform. Dataset yang digunakan memiliki karakteristik sebagai berikut:

- Jenis data: teks ulasan pelanggan
- Bahasa: Bahasa Indonesia
- Jumlah data: ± 1000 –3000 ulasan

Tabel 3. Contoh Dataset Penelitian

Tanggal	Sumber data	Komentar
25 Maret 2026	Google playstore	padahal saya pelanggan setia shopee,tapi setelah cek out dan buat pesanan nunggu lama banget ekspedisi jemput pesanan ke seller nya,belum lagi di perjalanan itulah yang bikin saya kecewa,maka nya pesanan lama kali di proses nya akhir nya saya batalkan saja pesanannya....
17 Maret 2026	Google playstore	Aplikasinya bagus, buat pesan beli barang pun oke. Masalahnya kenapa saya ngasih bintang 3? karena saya jengkel sama iklan dari Shophee, terlalu maksa. Lagi buka browser, lagi nonton yt, dsb. tiba-tiba ada iklan shopee, tiap ada nih iklan pasti maksa kelempar masuk ke aplikasinya, maksa banget jadi bikin tambah jengkel.
18062020	Aplikasi X	@ShopeeCare @ShopeeID Gada 1 orangpun yang mau di malingin air nya. Hapus tuh fitur SAMPAH!!!! bukan dengan alasan bisa berinteraksi jadi bisa maling!!!!

2. Persiapan data

Tahap ini bertujuan untuk menghasilkan korpus teks yang bersih, terstandar, dan siap dianalisis oleh model *Machine Learning*. Langkah-langkahnya meliputi:

a) Pelabelan Data

Pada tahap pelabelan dilakukan secara manual untuk mengklasifikasikan setiap ulasan ke dalam satu kategori aspek layanan (*single-label classification*). Pada Tabel 4 menampilkan label aspek layanan yang akan digunakan pada penelitian ini.

Tabel 4. Kategori Label Aspek Layanan

Label	Deskripsi	Keyword Bantuan
Pengiriman	Terkait kecepatan/ketepatan pengiriman	kirim, sampai, lama, cepat, resi, kurir, paket rusak
Kualitas Produk	Kondisi/barang	bagus, rusak, cacat
Pelayanan	Respons penjual	Admin ramah, slow respon, seller, iklan, barang hilang, video, live, gratis ongkir, diskon, fast respon
Harga	Nilai harga	murah, mahal

Adapun prosedur pelabelan, menambahkan kolom: **label** pada contoh data yang ditampilkan pada Tahap ini dilakukan dengan mengumpulkan data ulasan pengguna e-commerce melalui proses *crawling*. Penelitian berfokus pada ulasan pelanggan Shopee dan Tokopedia yang diambil dari Google Play Store dan media sosial X (Twitter). Tabel 3 menampilkan contoh ulasan dari kedua platform. Dataset yang digunakan memiliki karakteristik sebagai berikut:

- Jenis data: teks ulasan pelanggan
- Bahasa: Bahasa Indonesia
- Jumlah data: ±1000–3000 ulasan

Tabel 3, kemudian Pada tahap pelabelan dilakukan oleh 2 (dua) annotator. Contoh dataset yang akan diberi label terdapat pada Tabel 5.

Tabel 5. Kategori Label Aspek Layanan

Tanggal	Sumber data	Komentar	Label
25 Maret 2026	Google playstore	padahal saya pelanggan setia shopee,tapi setelah cek out dan buat pesanan nunggu lama banget ekspedisi jemput pesanan ke seller nya,belum lagi di perjalanan itulah yang bikin saya kecewa,maka nya ...	Pengiriman
17 Maret 2026	Google playstore	Aplikasinya bagus, buat pesan beli barang pun oke. Masalahnya kenapa saya ngasih bintang 3? karena saya jengkel sama iklan dari Shophee, terlalu maksa. Lagi buka browser, lagi nonton yt, dsb.	Pelayanan

b) Pembersihan Data

Langkah awal dilakukan dengan menghapus elemen-elemen non-teks yang tidak relevan seperti tanda baca, angka, tautan (URL), simbol, emoji, serta mention dan hashtag. Tujuannya adalah mengurangi gangguan *noise* yang dapat mempengaruhi hasil ekstraksi fitur kata.

c) Normalisasi Bahasa (*Text Normalization*)

Tahapan ini mengkonversi variasi penulisan informal ke bentuk baku Bahasa Indonesia. Proses normalisasi mencakup penyamaan ejaan, penggantian singkatan umum (misalnya: dsb → dan sebagainya, yg → yang, bgt → banget), serta konversi istilah lokal menggunakan kamus slang lokal yang disusun dari korpus media sosial. Normalisasi penting untuk memastikan konsistensi token dan mengurangi variasi kata yang merujuk pada makna sama.

d) Tokenisasi

Setelah teks dinormalisasi, setiap kalimat dipecah menjadi satuan kata (token). Tokenisasi dilakukan menggunakan word tokenizer Bahasa Indonesia agar model dapat mengenali batas kata secara akurat, termasuk untuk kata majemuk.

e) Penghapusan Stopword (*Stopword Removal*)

Tahap ini menghapus kata umum yang kurang informatif, seperti “yang”, “dan”, dan “untuk”, menggunakan daftar stopwords Bahasa Indonesia, sehingga analisis lebih fokus pada kata bermakna.

f) Stemming dan Lematisasi (*Lemmatization*)

Stemming dan lemmatisasi merupakan proses prapemrosesan teks untuk mengubah kata berimbuhan menjadi bentuk dasar menggunakan aturan morfologi Bahasa Indonesia, sehingga kata dengan makna sama tidak dianggap sebagai entitas berbeda.

3. Proses data

Pada tahap ini ada beberapa proses, sebagai berikut:

a) Ekstraksi Fitur

Ekstraksi fitur dilakukan untuk mengubah data teks menjadi representasi numerik menggunakan TF-IDF (18), yang memberikan bobot kata berdasarkan frekuensi dan tingkat kepentingannya dalam korpus. Penelitian ini menggunakan unigram dan bigram, di mana unigram merepresentasikan kata tunggal, sedangkan bigram menangkap pasangan kata untuk mempertimbangkan konteks. Kombinasi keduanya diharapkan menghasilkan fitur yang lebih informatif dalam proses klasifikasi.

b) Seleksi Fitur

Seleksi fitur bertujuan mengurangi dimensi data dan meningkatkan kinerja model. Metode yang digunakan adalah Chi-Square (χ^2) untuk mengukur keterkaitan antara fitur dan kelas (19). Fitur dengan nilai tertinggi dipilih sebagai yang paling relevan sehingga dapat mengurangi *noise*, meningkatkan akurasi, dan mempercepat pelatihan

c) SMOTE

Ketidakseimbangan data (*imbalanced data*) terjadi ketika distribusi kelas tidak merata, sehingga model cenderung bias terhadap kelas mayoritas. Pada ulasan e-commerce, kondisi ini umum terjadi. Untuk mengatasinya, digunakan SMOTE (*Synthetic Minority Over-sampling Technique*) yang menghasilkan data sintetis pada kelas minoritas agar distribusi data lebih seimbang (20).

4. Model Machine Learning

Penelitian ini menggunakan beberapa algoritma *machine learning* untuk klasifikasi ulasan e-commerce, yaitu Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, dan Gradient Boosting. Naïve Bayes merupakan metode probabilistik berbasis Teorema Bayes yang sederhana dan efektif untuk teks (6,9).

SVM bekerja dengan mencari *hyperplane* optimal dan memiliki performa baik pada data berdimensi tinggi (9,11,19). Logistic Regression memodelkan probabilitas kelas dan mudah diinterpretasikan (8,9,12).

Random Forest dan Gradient Boosting merupakan metode *ensemble*; Random Forest menggabungkan banyak *decision tree* untuk meningkatkan akurasi dan mengurangi *overfitting*, (8,9,12). sedangkan Gradient Boosting membangun model secara bertahap untuk memperbaiki kesalahan sebelumnya (12,18).

5. Evaluasi Model

Pada tahap ini, penelitian menyajikan perbandingan kinerja beberapa algoritma *machine learning* dalam mengklasifikasikan ulasan e-commerce berdasarkan aspek layanan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* (21). Hasil evaluasi dianalisis untuk membandingkan kinerja antar model serta mengkaji pengaruh variasi fitur (TF-IDF unigram dan bigram) dan penerapan SMOTE dalam menangani *imbalanced data*.

Berdasarkan analisis tersebut, penelitian mengidentifikasi model dan kombinasi metode yang paling optimal dalam meningkatkan kinerja klasifikasi aspek layanan.

JADWAL PENELITIAN

Jadwal penelitian disusun berdasarkan pelaksanaan penelitian, ditulis dengan mengisi langsung tabel berikut dan diperbolehkan menambahkan baris sesuai banyaknya jenis kegiatan.

No.	Nama Kegiatan	Bulan				
		3	4	5	6	7
1.	Penelusuran pustaka (Analisis Sentimen, <i>E-Commerce</i> , Perilaku Pengguna), mencari ide riset					
2.	Menyusun Proposal Penelitian					
3.	Pengumpulan data di aplikasi X, aplikasi playstore terkait dengan ulasan di <i>E-Commerce</i>					
4.	Persiapan data					
5.	Proses data					
6.	Pemetaan aspek layanan pada ulasan <i>e-commerce</i>					
7.	Membuat Laporan Kemajuan					
8.	Proses data dengan <i>Machine learning</i>					
9.	Evaluasi Beberapa Model <i>Machine Learning</i>					
10.	Penanganan data tidak seimbang (<i>imbalance</i>)					
11.	Interpretasi dan analisis hasil					
12.	Analisis dan visualisasi					
13.	Membuat draft JITECS – Sinta 2					

No.	Nama Kegiatan	Bulan				
		3	4	5	6	7
14.	submit publikasi ke JTeCS					
15.	draft HAKI basis data					
16.	Membuat Laporan Akhir					
17.	Seminar Hasil					

Hasil pelaksanaan penelitian yang telah dicapai meliputi data dan hasil analisis. Seluruh hasil atau capaian yang dilaporkan harus berkaitan dengan tahapan pelaksanaan penelitian sebagaimana direncanakan pada proposal. Penyajian data dapat berupa gambar, tabel, grafik, dan sejenisnya, serta analisis didukung dengan sumber pustaka primer yang relevan dan terkini.

HASIL PELAKSANAAN PENELITIAN

1. Tahap Pengumpulan Data

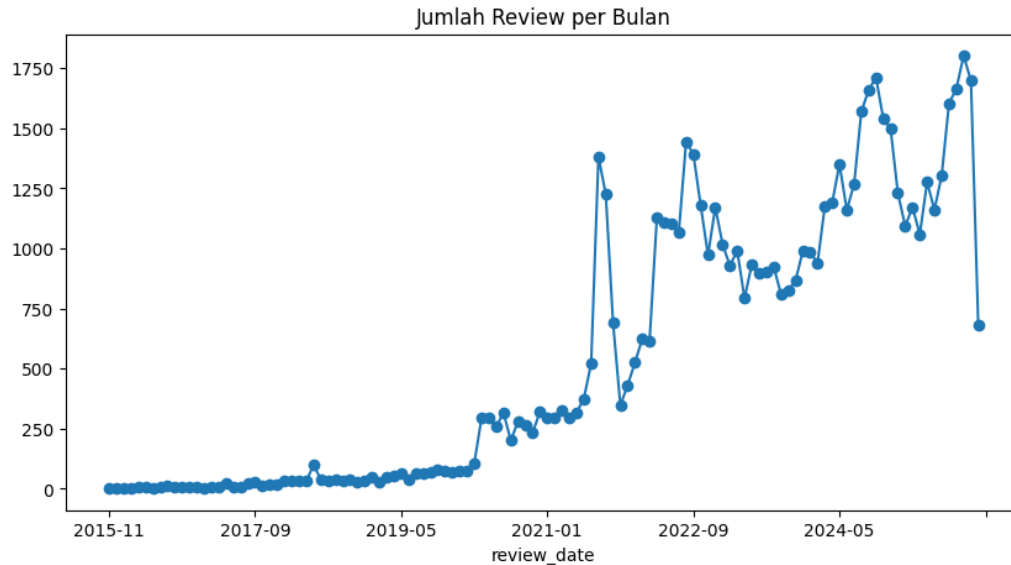
Pada tahap ini melakukan pencarian dataset ulasan pengguna aplikasi *E-Commerce* dengan sumber sebagai berikut:

- Ulasan Tokopedia: <https://www.kaggle.com/datasets/potatokat/tokopedia-reviews>
- Ulasan Shopee: <https://www.kaggle.com/datasets/herudimustobing/review-shopee>

Pada tahap berikut akan dilakukan eksplorasi dataset pada kedua aplikasi.

a. Ulasan Tokopedia

File dataset ulasan Tokopedia terdapat 65.543 komentar memiliki 13 kolom, setelah dilakukan proses penghapusan data duplikasi menjadi 58.068 komentar. Dataset merupakan ulasan dari tanggal 18 November 2015 hingga 12 Desember 2025. Pada Gambar 4 menampilkan frekuensi jumlah ulasan perbulan yang terdapat pada dataset.

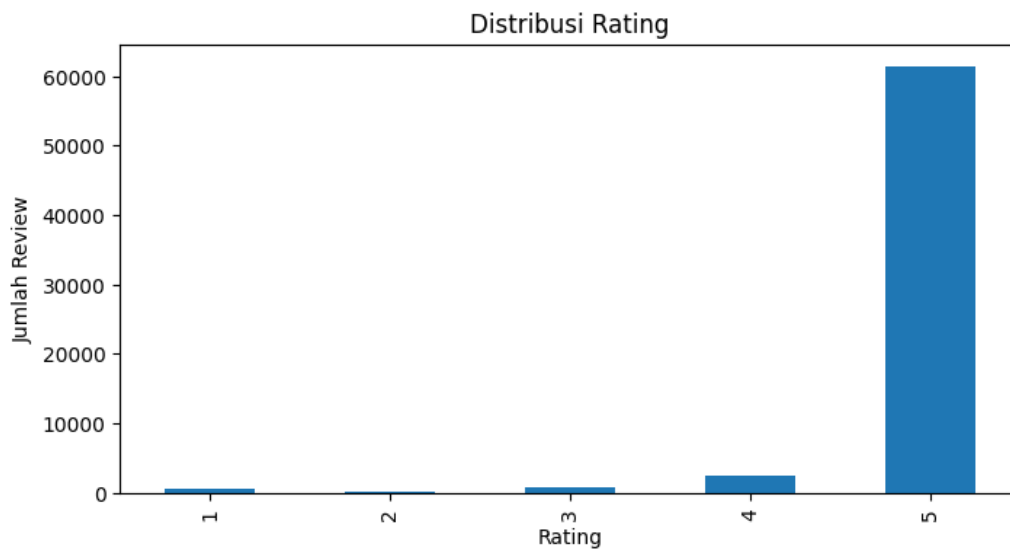


Gambar 4. Jumlah Review per Bulan pada Ulasan Tokopedia

Pada tahapan berikutnya, melakukan eksplorasi pada dataset ulasan di Tokopedia

(1) Distribusi Rating

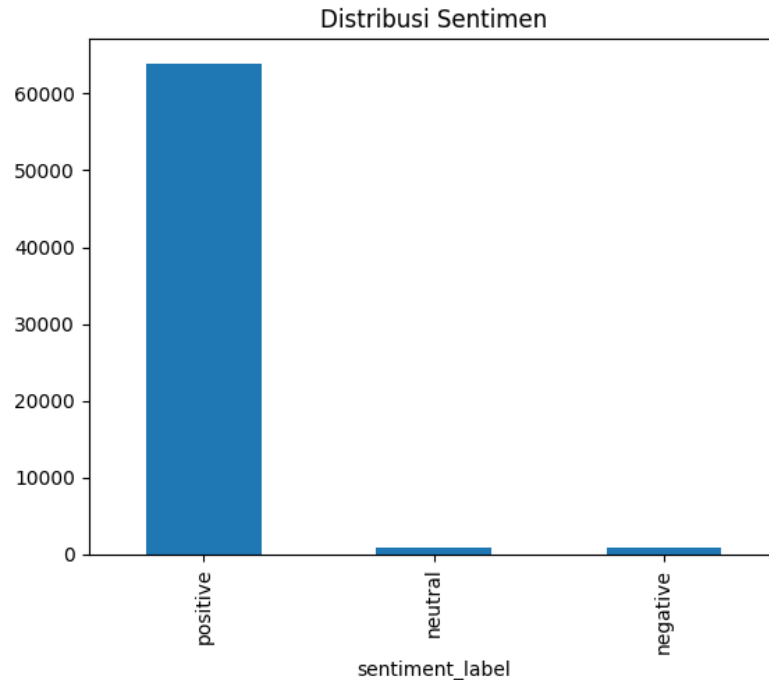
Berdasarkan Gambar 5 menampilkan jumlah ulasan pada Tokopedia berdasarkan rating 1 hingga 5 dari tahun 2015 hingga 2025. Dari tampilan menunjukkan bahwa rating 5 lebih dominan dari rating 1.



Gambar 5. Distribusi Rating pada Ulasan Tokopedia

(2) Distribusi Sentimen pada Ulasan Tokopedia

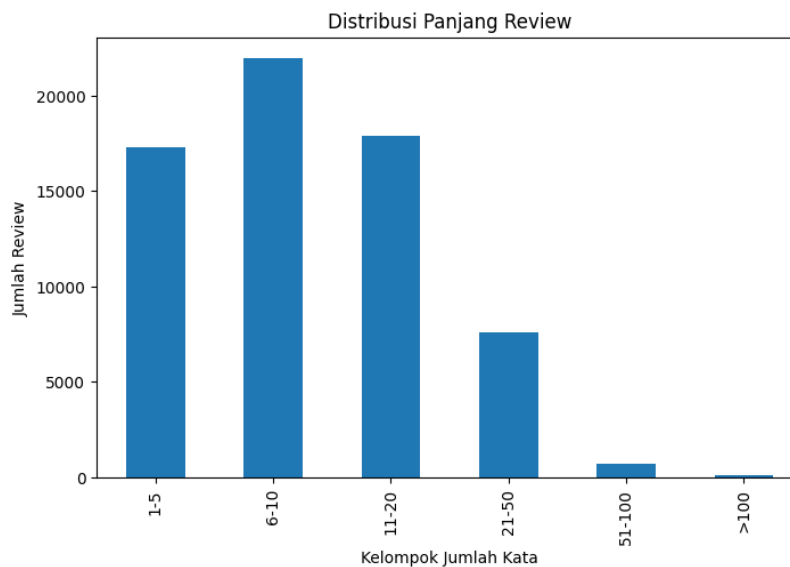
Pada Gambar 6 menampilkan jumlah sentimen positif, netral dan negatif dari kumpulan dataset di ulasan pada aplikasi Tokopedia. Gambar menunjukkan bahwa sentimen positif lebih banyak dari pada sentiment netral dan negatif.



Gambar 6. Distribusi Sentimen pada Ulasan Tokopedia

(3) Distribusi Panjang Ulasan di Aplikasi Tokopedia

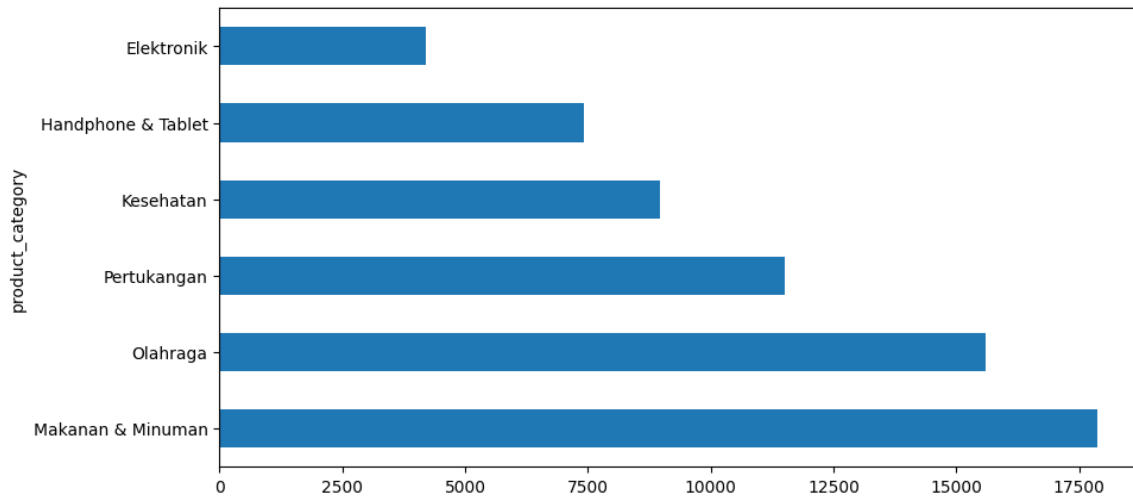
Pada Gambar 7 menampilkan frekuensi jumlah kata yang terdapat pada ulasan di aplikasi Tokopedia. Berdasarkan hasil eksplorasi pada dataset menunjukkan bahwa ulasan dengan jumlah kata minimal 6 hingga 10 kata memiliki frekuensi atau distribusi terbesar di dataset.



Gambar 7. Distribusi Panjang Ulasan di Tokopedia

(4) Distribusi Kategori Produk

Pada Gambar 8 menampilkan frekuensi ulasan berdasarkan kategori produk. Hasil eksplorasi menunjukkan bahwa kategori makanan dan minuman paling banyak menerima ulasan.



Gambar 8. Distribusi Ulasan Berdasarkan Kategori di Tokopedia

2. Persiapan Data

a. Pembersihan Data

Pada tahap ini melakukan tahapan:

- Casefolding, tahapan ini mengubah huruf besar yang terdapat pada 58.068 ulasan di Tokopedia menjadi huruf kecil.
- Bersihkan teks dengan menghapus beberapa hal berikut yaitu (URL, mention, hastags, karakter non-alfabet, tanda baca, spasi kosong, superskrip dan subskrip, emote), hapus baris yang kosong/tidak ada isinya). Proses ini dilakukan pada dataset sebanyak 58.068 baris, dan tidak dilakukan penghapusan untuk angka, karena penelitian ini fokus pada aspek layanan sehingga dikhawatirkan ada ulasan yang berkaitan dengan angka, contohnya “Pengiriman 2 hari”.

- b. Normalisasi, pada tahap ini melakukan normalisasi pada kata-kata yang disingkat, atau istilah. Berikut adalah kumpulan kata yang dinormalisasi.

'yg': 'yang',	'yg': 'yang',	'bgt': 'banget',	'dr': 'dari',	'mantul': 'mantap',
'ga': 'tidak',	'ga': 'tidak',	'bggt': 'banget',	'dgn': 'dengan',	'mantabb': 'mantap',
'gak': 'tidak',	'gak': 'tidak',	'bngt': 'banget',	'dg' : 'dengan',	'ok': 'oke',
'gk': 'tidak',	'gk': 'tidak',	'udh': 'sudah',	'sm': 'sama',	'okee': 'oke',
'ngga': 'tidak',	'ngga': 'tidak',	'sdh': 'sudah',	'sma': 'sama',	'okee': 'oke',
'nggak': 'tidak',	'nggak': 'tidak',	'tp': 'tapi',	'krn': 'karena',	'makasih': 'terima
'tdk': 'tidak',	'tdk': 'tidak',	'tpi': 'tapi',	'trs': 'terus',	kasih',
'recommended': 'rekomendasi',		'n' : 'dan',		
'recomend': 'rekomendasi',		'lg' : 'lagi',		
'fast': 'cepat',		'klu' : 'kalau',		
'brg' : 'barang',		'jg' : 'juga',		
'nga' : 'nggak',		'exp' : 'expired',		
'gaada' : 'nggak ada',		'jd' : 'jadi',		
'nggak' : 'tidak',		'utk' : 'untuk',		
'mksh': 'terima kasih',		'sy' : 'saya',		
'trims': 'terima kasih',		'ori' : 'original',		
'thx': 'terima kasih',				

- c. Tokenisasi. Pada tahap ini melakukan proses tokenisasi dengan mengubah kumpulan kata yang sudah dibersihkan pada tahap sebelumnya dengan cara memisahkan teks menjadi token
- d. Hapus *Stopword* menggunakan *Library* NLTK dan menambah kata-kata yang tidak berhubungan dengan penelitian dan tidak mengandung arti. Berikut adalah kumpulan kata yang ditambahkan pada *Stopword*: 'nya', 'nih', 'aja', 'sih', 'dong', 'deh', 'ya'
- e. Stemming.
- f. Lemmatization

STATUS LUARAN

Uraikan jenis, identitas, dan status ketercapaian setiap luaran yang dijanjikan. Jenis luaran berupa publikasi, perolehan kekayaan intelektual, atau luaran lainnya yang telah dijanjikan pada proposal.			
No.	Kategori Luaran	Jenis Luaran	Target Capaian
1.	Artikel Ilmiah	Jurnal Sinta 2 bernama JITeCS (Journal of Information Technology and Computer Science) Vol. 11 No. 3 Tahun 2026. Penerbit Universitas Brawijaya. https://jitecs.ub.ac.id/index.php/jitecs	Belum diproses
2.	HKI	Basis Data	Belum diproses

DAFTAR PUSTAKA

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver. Sumber pustaka mengutamakan hasil penelitian pada jurnal ilmiah yang terkini (maksimal 5 tahun terakhir). Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka.

1. Aini NF, Darmansyah D, Apriyanto R, Ramdhani R, Prastyo Y. Literature Review: Faktor-Faktor Yang Mempengaruhi Kepuasan Pelanggan Dalam Layanan Pengiriman Barang. *J Manag Innov Entrep.* 2025;2(4):2305–14.
2. Nur A, Garamba Y. Analisis Kepuasan Pelanggan pada E-commerce. *J Ilmu Komput Dan Inform.* 2024;1(2):27–37.
3. Sinarta SY, Haryani E. Analisis Kepuasan Pengguna pada Layanan E-commerce di Indonesia (Systematic Literature Review). *JUSTIN J Sist dan Teknol Inf.* 2025;13(1):135–47.
4. Anhar RA, Ramadhani KA, Andela OM. Pengaruh Kualitas Pelayanan dan Harga Produk terhadap Kepuasan Pelanggan di E-commerce Shopee. *J Ilm Dan Karya Mhs.* 2024;2(4):360–75.
5. Ardiyanto TR, Ramadhayanti A. Pengaruh Ketersediaan Barang dan Kualitas Layanan terhadap Ulasan Pembeli pada Platform E-Commerce Tokoteknopro. *J Manaj Dan Bisnis Ekon.* 2024;2(4):96–104.
6. Juanita S, Adiyarta K, Syafrullah M. Sentiment analysis on E-Marketplace User Opinions Using Lexicon-Based and Naïve Bayes Model. In: *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. Institute of Electrical and

- Electronics Engineers Inc.; 2022. p. 379–82.
7. Firdaus ADA, Rahmawan RD, Mahendra YR, Cahyono HD. Sentiment Analysis Classification in Women's E-Commerce Reviews With Machine Learning Approach. *J Tek Inform* [Internet]. 2024;5(6):1549–59. Available from: <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/2392>
 8. Imran AY, Sanjaya MR, Bayu Wijaya Putra, Gabriel Ekoputra Hartono Cahyadi. Optimization of Sentiment Analysis on Tokopedia User Reviews Using Gridsearchcv and Smote with Machine Learning Algorithms. *INOVTEK Polbeng - Seri Inform*. 2025;10(3):1634–44.
 9. Barus H, Fajri IN, Pristyanto Y. Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF , SMOTE , and Traditional Machine Learning Models. *J Appl Informatics Comput*. 2025;9(5):2552–61.
 10. Oktariana CN, Oktadini NR. Analisis Sentimen Ulasan Pengguna Aplikasi Tokopedia Menggunakan Algoritma Random Forest. *JATI (Jurnal Mhs Tek Inform*. 2025;9(6):10224–31.
 11. Ramadhanti ND, Handayani RN. Analisis Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play Store Menggunakan Support Vector Machine dengan Teknik Smote. *JIP (Jurnal Inform Polinema)*. 2026;12(2):313–21.
 12. Yosri MYKM, Hanni H, Arafah MN, Salzabilah N, Maryam M, Nur IS, et al. Implementasi Machine Learning untuk Analisis Pola Perilaku Konsumen E-Commerce Berbasis Data Transaksi. *IRMEX J Artif Intell Data Sci*. 2025;1(2):36–41.
 13. Apriyanto B, Sitio SLM. Penerapan K-Means dalam Menganalisis Pola Pembelian Pelanggan Pada Data Transaksi E-Commerce. *Bit-Tech (Binary Digit - Technol* [Internet]. 2025;7(3):790–7. Available from: <https://jurnal.kdi.or.id/index.php/bt/article/view/2195/1135>
 14. Ruscikasani SN, Oktalivia RRN, Putra FR, Wahidin AJ, Rahmatullah B, Kurniawati I. Prediksi Pembelian E-Commerce Menggunakan XGBoost Berbasis Perilaku Sesi Pengguna. *J Artif Intell Digit Bus*. 2025;4(4):5666–72.
 15. Pratiwi MP, Natalia EY, Fauzi R. Desain Sistem Big Data untuk Deteksi Pola Perilaku Konsumen dengan Menggunakan Algoritma Machine Learning. *J Desain dan Anal Teknol* [Internet]. 2026;5(1):12–9. Available from: <https://jurnal.kdi.or.id/index.php/bt/article/view/2195/1135>
 16. Muflih FA, Gata G, Usino W, Rusdah R. Perancangan E-Commerce Dengan Content Management System (CMS) Untuk Memperluas Penjualan Toko Nurdin. *Semin Nas Mhs Fak Teknol Inf ke 5* [Internet]. 2024;3(2):1384–93. Available from: <https://senafiti.budiluhur.ac.id/senafiti/article/view/1397/811>
 17. Setyawan I, Hin LL, Anubhakti D, Gata G. Implementasi Electronic Customer Relationship Management Untuk Meningkatkan Kepuasan Dan Layanan Kepada Pelanggan Thani Coffee. *IDEALIS Indones J Inf Syst*. 2024;7(1):51–60.
 18. Karim M, Missen MMS, Umer M, Fida A, Eshmawi AA, Mohamed A, et al. Comprehension of polarity of articles by citation sentiment analysis using TF-IDF and ML classifiers. *PeerJ Comput Sci*. 2022;8:1–21.
 19. Hokijuliandy E, Napitupulu H, Firdaniza F. Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square. *SisInfo* [Internet]. 2023;5(2):40–9. Available from: <https://jurnalunibi.unibi.ac.id/ojs/index.php/SisInfo/article/view/670/581>

20. Taskiran SF, Turkoglu B, Kaya E, Asuroglu T. A comprehensive evaluation of oversampling techniques for enhancing text classification performance. *Sci Rep.* 2025;15(21631):1–20.
21. Vujović ŽD. Classification Model Evaluation Metrics. *Int J Adv Comput Sci Appl.* 2021;12(6):1–8.