PROPOSAL SKRIPSI

PENGEMBANGAN MODUL KLASIFIKASI MULTIKELAS KOMENTAR TIKTOK TERHADAP PRODUK KOSMETIK LOKAL DENGAN ALGORITMA NAIVE BAYES PADA APLIKASI BERBASIS WEB

Diajukan untuk Memenuhi Persyaratan Seminar Proposal Skripsi pada Program Studi Informatika

OLEH:

ALWI SYAHRIR

NIM: 210299

PROGRAM STUDI INFORMATIKA
FAKULTAS AI
AI UNIVERSITAS

2026

LEMBAR PERSETUJUAN PROPOSAL

PENGEMBANGAN MODUL KLASIFIKASI MULTIKELAS KOMENTAR TIKTOK TERHADAP PRODUK KOSMETIK LOKAL DENGAN ALGORITMA NAIVE BAYES PADA APLIKASI BERBASIS WEB

Disusun Oleh: ALWI SYAHRIR (NIM: 210299)

Proposal skripsi ini telah diperiksa dan disetujui oleh Dosen Pembimbing untuk diseminarkan pada Seminar Proposal Penelitian Program Studi Informatika, Fakultas Ai, AI UNIVERSITAS.

Dosen Pembimbing I	Dosen Pembimbing II
(_____)	(_____)
NIP/NIDN.	NIP/NIDN.

Mengetahui,
Ketua Program Studi Informatika

(_____)
NIP/NIDN.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penyusunan proposal skripsi berjudul "Pengembangan Modul Klasifikasi Multikelas Komentar TikTok terhadap Produk Kosmetik Lokal dengan Algoritma Naive Bayes pada Aplikasi Berbasis Web" dapat diajukan sebagai rancangan penelitian untuk Seminar Proposal Skripsi pada Program Studi Informatika, AI UNIVERSITAS.

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada pimpinan universitas, ketua program studi, dosen pembimbing, dosen penguji, serta keluarga dan rekan-rekan atas bimbingan, arahan, serta dukungan yang telah diberikan. Penulis menyadari bahwa naskah ini masih terbuka terhadap saran konstruktif demi penyempurnaan di masa mendatang.

AI UNIVERSITAS, November 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL ..... **i**
LEMBAR PERSETUJUAN PROPOSAL ..... **ii**
KATA PENGANTAR ..... **iii**
DAFTAR ISI ..... **iv**
DAFTAR TABEL ..... **v**
DAFTAR GAMBAR ..... **vi**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang Masalah 1
1.2 Identifikasi Masalah 1
1.3 Batasan Masalah 2
1.4 Rumusan Masalah 3
1.5 Tujuan Penelitian 3
1.6 Manfaat Penelitian 4
1.7 Sistematika Penulisan 5
BAB II TINJAUAN PUSTAKA **6**
2.1 Landasan Teori 6
2.2 Penelitian Terdahulu 10
2.3 Kerangka Pemikiran 11
2.4 Hipotesis Penelitian 13
BAB III METODOLOGI PENELITIAN **14**
3.1 Desain Penelitian & Alur Sistem (Flowchart) 14
3.2 Sumber dan Teknik Pengumpulan Dataset 17
3.3 Tahapan Pra-Pemrosesan Data (Text Preprocessing) 17
3.4 Pembobotan Fitur dan Pembagian Data 18
3.5 Pemodelan Algoritma Klasifikasi 19
3.6 Metrik Evaluasi Kinerja 20
DAFTAR PUSTAKA **22**

DAFTAR TABEL

Tabel 2.1 Sintesis Penelitian Terdahulu **10**

DAFTAR GAMBAR

Gambar 2.1 Bagan Kerangka Pemikiran Penelitian **11**

Gambar 3.1 Diagram Alur (Flowchart) Prosedur Penelitian **21**

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Pertumbuhan ekonomi digital Indonesia menempatkan media sosial sebagai kanal utama interaksi konsumen terhadap produk, dengan TikTok tercatat sebagai salah satu platform dengan penetrasi tertinggi di antara pengguna internet nasional. Laporan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII, 2024) menunjukkan bahwa 78,19 persen dari 221,56 juta pengguna internet Indonesia mengakses media sosial, dan TikTok menduduki peringkat kedua platform paling banyak digunakan dengan proporsi 30,74 persen pengguna. Angka ini mengindikasikan bahwa ruang percakapan digital di TikTok telah menjelma menjadi etalase sekaligus panggung evaluasi publik bagi produk yang dipasarkan, termasuk produk kosmetik lokal yang tumbuh agresif dalam beberapa tahun terakhir. Konsekuensinya, komentar pengguna TikTok bukan sekadar ekspresi individual, melainkan kumpulan opini kolektif yang potensial untuk diolah menjadi umpan balik strategis bagi pelaku industri kosmetik dalam negeri.

Komentar pada konten TikTok produk kosmetik lokal menampilkan variasi bahasa yang tinggi, mencakup campuran bahasa Indonesia baku, bahasa gaul, singkatan, salah ejaan, hingga sisipan istilah asing. Variasi linguistik semacam ini menyulitkan proses klasifikasi otomatis apabila hanya mengandalkan pencocokan kata kunci sederhana. Penelitian Sari dan Ramadhani (2023) menegaskan bahwa analisis sentimen pada data media sosial berbahasa Indonesia membutuhkan tahap normalisasi leksikal yang memadai karena ketidakkonsistenan penulisan menurunkan akurasi model hingga belasan persen apabila tidak ditangani. Temuan tersebut menegaskan bahwa pengolahan komentar TikTok menuntut rantai pra-pemrosesan teks yang lebih cermat, mulai dari pembersihan karakter, penyeragaman huruf, tokenisasi, hingga normalisasi kata tidak baku, sebelum data siap dimasukkan ke dalam model pembelajaran mesin.

Klasifikasi multikelas menjadi relevan karena komentar konsumen terhadap produk kosmetik tidak selalu bersifat biner positif-negatif. Komentar dapat berupa pujian terhadap hasil pemakaian, keluhan terhadap tekstur atau kemasan, pertanyaan mengenai harga dan ketersediaan, permintaan rekomendasi,

hingga pernyataan netral yang tidak mengandung polaritas. Penelitian Wulandari et al. (2024) menunjukkan bahwa skema klasifikasi multikelas pada ulasan produk e-commerce menghasilkan pemetaan informasi yang lebih kaya dibandingkan klasifikasi biner, meskipun menuntut representasi fitur yang lebih representatif dan penanganan ketidakseimbangan kelas yang lebih baik. Dengan demikian, pemilihan algoritma yang efisien sekaligus interpretatif menjadi pertimbangan utama dalam merancang sistem klasifikasi komentar produk kosmetik lokal.

Algoritma Naive Bayes menjadi salah satu kandidat yang banyak dipertimbangkan untuk klasifikasi teks berbahasa Indonesia karena komputasinya ringan, mudah diinterpretasikan, dan telah terbukti kompetitif pada dataset berukuran menengah. Penelitian Pratama dan Hidayat (2023) melaporkan bahwa Multinomial Naive Bayes yang dikombinasikan dengan pembobotan TF-IDF mencapai akurasi di atas 85 persen pada klasifikasi ulasan produk berbahasa Indonesia. Kendati demikian, performa Naive Bayes sangat bergantung pada kualitas pra-pemrosesan, pemilihan fitur, dan strategi pembagian data, sehingga evaluasi pada domain komentar TikTok produk kosmetik lokal masih menyisakan ruang kajian yang belum banyak diisi. Celah inilah yang mendorong diperlukannya pengembangan modul klasifikasi multikelas yang tidak hanya akurat, tetapi juga dapat dioperasikan secara praktis melalui aplikasi berbasis web agar dapat dimanfaatkan oleh pelaku usaha kosmetik lokal, peneliti, maupun akademisi.

1.2 Identifikasi Masalah

1. Komentar pengguna TikTok terhadap produk kosmetik lokal memiliki variasi bahasa yang tinggi, mencakup bahasa gaul, singkatan, salah ejaan, dan campuran istilah asing, sehingga sulit diklasifikasikan secara otomatis tanpa pra-pemrosesan yang memadai.
2. Klasifikasi komentar yang selama ini dilakukan banyak yang bersifat biner (positif-negatif), padahal kebutuhan pelaku usaha kosmetik menuntut pemetaan multikelas seperti pujian, keluhan, pertanyaan, permintaan rekomendasi, dan pernyataan netral.

3. Belum tersedia modul klasifikasi multikelas komentar TikTok berbahasa Indonesia yang terintegrasi dalam aplikasi berbasis web dan dapat digunakan secara praktis oleh pengguna non-teknis.
4. Kinerja algoritma Naive Bayes pada domain komentar TikTok produk kosmetik lokal belum banyak dievaluasi secara sistematis menggunakan metrik confusion matrix (accuracy, precision, recall, dan F1-score) beserta analisis kesalahan klasifikasi.
5. Keterbatasan sarana pengolahan data komentar secara terstruktur menyebabkan umpan balik konsumen dari media sosial belum termanfaatkan optimal sebagai dasar pengambilan keputusan produk dan pemasaran.

1.3 Batasan Masalah

1. Algoritma yang dikaji dibatasi pada Naive Bayes, khususnya Multinomial Naive Bayes dengan pembobotan fitur TF-IDF, tanpa membandingkan secara mendalam dengan algoritma deep learning.
2. Objek data dibatasi pada komentar berbahasa Indonesia pada konten TikTok yang membahas produk kosmetik lokal, dengan jumlah dan rentang waktu pengambilan data ditetapkan pada periode penelitian.
3. Klasifikasi dibatasi pada sejumlah kelas yang telah ditetapkan berdasarkan hasil pelabelan, misalnya pujian, keluhan, pertanyaan, permintaan rekomendasi, dan pernyataan netral.
4. Lingkungan implementasi dibatasi pada aplikasi berbasis web yang dikembangkan sebagai purwarupa (prototype) modul klasifikasi, tanpa integrasi langsung ke API resmi TikTok.
5. Evaluasi kinerja dibatasi pada confusion matrix dan metrik turunannya (accuracy, precision, recall, F1-score), tanpa pengujian beban pengguna berskala besar.

1.4 Rumusan Masalah

1. Bagaimana rancangan modul klasifikasi multikelas komentar TikTok terhadap produk kosmetik lokal menggunakan algoritma Naive Bayes yang diimplementasikan pada aplikasi berbasis web?

2. Bagaimana tahapan pra-pemrosesan teks berbahasa Indonesia yang efektif untuk menangani variasi linguistik komentar TikTok produk kosmetik lokal?
3. Bagaimana kinerja algoritma Naive Bayes dalam mengklasifikasikan komentar TikTok produk kosmetik lokal berdasarkan metrik accuracy, precision, recall, dan F1-score?
4. Bagaimana hasil pengujian fungsional aplikasi berbasis web modul klasifikasi dalam memenuhi kebutuhan pengguna terhadap pemetaan komentar multikelas?

1.5 Tujuan Penelitian

1. Merancang dan membangun modul klasifikasi multikelas komentar TikTok terhadap produk kosmetik lokal menggunakan algoritma Naive Bayes yang diimplementasikan pada aplikasi berbasis web.
2. Merumuskan dan menerapkan tahapan pra-pemrosesan teks berbahasa Indonesia yang sesuai untuk menangani variasi linguistik komentar TikTok produk kosmetik lokal.
3. Mengukur dan menganalisis kinerja algoritma Naive Bayes pada klasifikasi komentar TikTok produk kosmetik lokal menggunakan metrik accuracy, precision, recall, dan F1-score beserta analisis kesalahan klasifikasi.
4. Menguji fungsionalitas aplikasi berbasis web modul klasifikasi untuk memastikan kesesuaiannya dengan kebutuhan pengguna dalam pemetaan komentar multikelas.

1.6 Manfaat Penelitian

1.6.1 Manfaat Teoretis

Penelitian ini diharapkan memperkaya khazanah kajian klasifikasi teks berbahasa Indonesia, khususnya pada domain media sosial dengan karakteristik bahasa yang tidak baku. Kontribusi teoretis yang ditawarkan mencakup perumusan rantai pra-pemrosesan teks yang disesuaikan dengan variasi linguistik komentar TikTok, serta bukti empiris mengenai kinerja Naive Bayes pada klasifikasi multikelas komentar produk kosmetik lokal. Hasil kajian juga dapat

menjadi rujukan bagi penelitian selanjutnya yang hendak membandingkan Naive Bayes dengan algoritma lain, menguji strategi penanganan ketidakseimbangan kelas, atau mengembangkan representasi fitur yang lebih adaptif terhadap bahasa media sosial.

1.6.2 Manfaat Praktis

Bagi pelaku industri kosmetik lokal, modul klasifikasi yang dikembangkan dapat dimanfaatkan untuk memetakan umpan balik konsumen secara terstruktur sehingga keputusan pengembangan produk, penyesuaian komunikasi pemasaran, dan penanganan keluhan dapat dilakukan lebih cepat dan berbasis data. Bagi perguruan tinggi, hasil penelitian menjadi bahan ajar dan studi kasus penerapan pembelajaran mesin pada mata kuliah terkait, sekaligus memperkuat keterkaitan antara riset dan kebutuhan industri. Bagi peneliti selanjutnya, tersedia purwarupa aplikasi berbasis web dan dokumentasi tahapan pengolahan data yang dapat dikembangkan lebih lanjut, misalnya dengan penambahan algoritma pembanding, perluasan cakupan platform, atau integrasi dengan sistem manajemen umpan balik pelanggan.

1.7 Sistematika Penulisan

Bab I PENDAHULUAN memuat latar belakang masalah, identifikasi masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan sebagai kerangka awal penelitian.

Bab II TINJAUAN PUSTAKA menguraikan landasan teori yang mencakup konsep media sosial dan komentar produk, produk kosmetik lokal, text mining dan klasifikasi teks, pra-pemrosesan teks berbahasa Indonesia, algoritma Naive Bayes, pembobotan TF-IDF, metrik evaluasi confusion matrix, serta aplikasi berbasis web. Bab ini juga memuat penelitian terdahulu dalam bentuk tabel komparasi, kerangka pemikiran, dan hipotesis penelitian.

Bab III METODOLOGI PENELITIAN menjelaskan jenis dan desain penelitian pengembangan (R&D), lokasi dan waktu penelitian, populasi dan sampel data komentar, definisi operasional variabel, metode pengumpulan data melalui crawling, uji instrumen dan validasi pelabelan, serta metode analisis data yang mencakup tahapan pra-pemrosesan, pembobotan TF-IDF, pembagian data, pelatihan Naive Bayes, evaluasi kinerja, dan pengujian fungsional aplikasi.

Bab IV HASIL PENELITIAN DAN PEMBAHASAN menyajikan gambaran umum objek penelitian, hasil pengumpulan dan pra-pemrosesan data, hasil pelatihan dan evaluasi model Naive Bayes beserta confusion matrix dan metrik turunannya, hasil pengujian fungsional aplikasi berbasis web, serta pembahasan kritis yang membandingkan temuan dengan penelitian terdahulu dan menganalisis penyebab kesalahan klasifikasi.

Bab V PENUTUP menyajikan kesimpulan yang menjawab setiap rumusan masalah secara tegas dengan angka metrik capaian, serta saran terapan bagi pelaku industri kosmetik lokal, perguruan tinggi, dan peneliti selanjutnya.

BAB II
TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Analisis Sentimen dan Klasifikasi Multikelas

Analisis sentimen merupakan proses komputasional untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau sikap yang terkandung dalam teks terhadap suatu entitas tertentu. Dalam perkembangannya, analisis sentimen tidak lagi terbatas pada polaritas biner positif-negatif, melainkan berkembang menjadi klasifikasi multikelas yang memungkinkan pengelompokan opini ke dalam beberapa kategori sekaligus, misalnya aspek produk, harga, kemasan, dan kualitas. Klasifikasi multikelas menjadi relevan ketika data teks memuat beragam dimensi evaluasi yang tidak dapat direduksi menjadi dua kutub tunggal. Menurut Wibowo dan Handayani (2023), pendekatan multikelas pada data ulasan produk menghasilkan granularitas informasi yang lebih tinggi bagi pelaku usaha dibandingkan klasifikasi biner, karena setiap aspek evaluasi konsumen dapat dipetakan secara terpisah. Dalam konteks komentar TikTok terhadap produk kosmetik lokal, klasifikasi multikelas memungkinkan identifikasi kategori seperti pujian kualitas, keluhan efek samping, pertanyaan ketersediaan, dan perbandingan dengan merek lain.

Secara konseptual, klasifikasi teks merupakan bagian dari supervised learning yang memerlukan data berlabel sebagai dasar pelatihan model. Model dilatih untuk mempelajari pola keterkaitan antara fitur teks dan label kategori, kemudian digunakan untuk memprediksi label pada data baru yang belum pernah dilihat. Keberhasilan klasifikasi teks sangat ditentukan oleh kualitas representasi fitur, keseimbangan distribusi kelas, dan kesesuaian algoritma dengan karakteristik data. Oleh karena itu, tahapan pra-pemrosesan teks dan pembobotan kata menjadi fondasi yang tidak dapat dilepaskan dari proses klasifikasi itu sendiri.

2.1.2 Text Preprocessing

Text preprocessing adalah rangkaian tahapan untuk membersihkan dan menyeragamkan teks mentah agar dapat diproses oleh algoritma machine learning. Data komentar TikTok yang bersifat tidak terstruktur, mengandung

campuran bahasa, singkatan, dan simbol, memerlukan pra-pemrosesan yang cermat sebelum dilakukan ekstraksi fitur. Menurut Pratama dan Kusuma (2023), kualitas pra-pemrosesan berkontribusi signifikan terhadap peningkatan akurasi klasifikasi teks berbahasa Indonesia, dengan selisih akurasi mencapai 12 hingga 18 persen antara data yang diproses dan tidak diproses. Tahapan pra-pemrosesan yang digunakan dalam penelitian ini diuraikan sebagai berikut.

Case folding merupakan proses penyeragaman huruf menjadi huruf kecil (lowercase) untuk menghilangkan perbedaan antara huruf kapital dan non-kapital yang tidak memiliki makna semantik berbeda. Tahapan ini krusial karena pengguna media sosial sering menulis dengan kombinasi huruf kapital yang tidak konsisten. Data cleaning bertujuan menghapus elemen non-teks seperti tautan URL, mention, tagar, emoji, tanda baca berlebih, dan karakter khusus yang tidak berkontribusi terhadap makna klasifikasi. Tokenizing adalah proses pemecahan teks menjadi unit-unit kata (token) berdasarkan pemisah tertentu, umumnya spasi, sehingga teks dapat dianalisis pada tingkat kata.

Slang normalization merupakan tahapan penormalan kata tidak baku, singkatan, dan bahasa gaul menjadi bentuk baku sesuai kamus. Tahapan ini menjadi tantangan tersendiri pada data media sosial berbahasa Indonesia karena tingginya variasi leksikal seperti "bgt" menjadi "banget", "gak" menjadi "tidak", atau "wkwk" yang perlu dipetakan ke bentuk netral. Menurut Sari dan Ramadhani (2023), normalisasi leksikal berbasis kamus slang yang dikombinasikan dengan pendekatan berbasis aturan mampu menurunkan tingkat kesalahan klasifikasi pada data media sosial berbahasa Indonesia secara signifikan. Stopword removal adalah proses penghapusan kata-kata umum yang tidak memiliki daya pembeda dalam klasifikasi, seperti "dan", "yang", "di", dan "ke", dengan menggunakan daftar stopword bahasa Indonesia. Stemming merupakan proses pemetaan kata berimbuhan ke bentuk dasar menggunakan algoritma Sastrawi, yang dirancang khusus untuk morfologi bahasa Indonesia dengan menangani awalan, akhiran, sisipan, dan kombinasi imbuhan.

2.1.3 Pembobotan Kata dengan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode pembobotan kata yang mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap keseluruhan korpus. Bobot TF-IDF dihitung dari

perkalian antara frekuensi kemunculan kata dalam dokumen (TF) dan invers dari frekuensi kemunculan kata tersebut di seluruh dokumen (IDF). Kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain akan memperoleh bobot tinggi, sedangkan kata yang muncul di hampir semua dokumen akan memperoleh bobot rendah. Formulasi TF-IDF dinyatakan sebagai berikut:

TF(t,d) = frekuensi kemunculan kata t dalam dokumen d

$$IDF(t) = \log(N / df(t))$$

$$\text{Bobot TF-IDF}(t,d) = TF(t,d) \times IDF(t)$$

dengan N adalah jumlah total dokumen dalam korpus dan df(t) adalah jumlah dokumen yang memuat kata t. Menurut Wijaya et al. (2024), kombinasi TF-IDF dengan algoritma Naive Bayes menghasilkan performa yang kompetitif pada klasifikasi teks berbahasa Indonesia, dengan akurasi yang stabil pada berbagai ukuran korpus. Pembobotan TF-IDF menghasilkan representasi vektor berdimensi tinggi yang menjadi masukan bagi algoritma klasifikasi.

2.1.4 Algoritma Naive Bayes untuk Klasifikasi Multikelas

Naive Bayes merupakan algoritma klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi independensi kondisional antar fitur. Meskipun asumsi independensi sering tidak terpenuhi secara sempurna pada data teks nyata, Naive Bayes terbukti memberikan performa yang baik dan efisien secara komputasi pada berbagai tugas klasifikasi teks. Untuk klasifikasi multikelas dengan kelas C_1, C_2, ..., C_k, probabilitas posterior suatu kelas C_k diberikan dokumen d dihitung sebagai berikut:

$$P(C_k|d) = (P(C_k) \times P(d|C_k)) / P(d)$$

dengan $P(C_k)$ adalah probabilitas prior kelas, $P(d|C_k)$ adalah likelihood dokumen diberikan kelas, dan P(d) adalah probabilitas marginal dokumen. Karena P(d) konstan untuk semua kelas, kelas yang diprediksi adalah kelas yang memaksimalkan nilai $P(C_k) \times P(d|C_k)$. Dengan asumsi independensi kondisional, likelihood didekomposisi menjadi perkalian probabilitas setiap fitur (kata) w_i:

$$P(d|C_k) = \prod P(w_i|C_k)$$

Probabilitas setiap kata diberikan kelas diestimasi menggunakan Laplace smoothing untuk menghindari nilai nol, dengan formula:

$$P(w_i|C_k) = (\text{count}(w_i, C_k) + \alpha) / (\text{count}(C_k) + \alpha \times |V|)$$

dengan $\text{count}(w_i, C_k)$ adalah frekuensi kata w_i pada kelas C_k, $\text{count}(C_k)$ adalah total kata pada kelas C_k, α adalah parameter smoothing (umumnya $\alpha = 1$), dan $|V|$ adalah ukuran kosakata. Dalam implementasinya, perhitungan dilakukan dalam ruang logaritma untuk mencegah underflow numerik akibat perkalian probabilitas bernilai sangat kecil. Menurut Hidayat dan Nugraha (2024), Naive Bayes dengan Laplace smoothing dan representasi TF-IDF menunjukkan akurasi di atas 85 persen pada klasifikasi ulasan produk berbahasa Indonesia, dengan waktu pelatihan yang jauh lebih singkat dibandingkan metode berbasis deep learning.

2.1.5 Confusion Matrix dan Metrik Evaluasi

Confusion matrix adalah tabel yang memetakan hasil prediksi model terhadap nilai aktual, sehingga memungkinkan perhitungan berbagai metrik evaluasi kinerja klasifikasi. Untuk klasifikasi multikelas dengan k kelas, confusion matrix berukuran k × k yang memuat jumlah prediksi benar pada diagonal utama dan kesalahan klasifikasi pada sel lainnya. Dari confusion matrix diturunkan empat metrik utama, yaitu accuracy, precision, recall, dan F1-score. Accuracy mengukur proporsi prediksi benar terhadap total prediksi, dihitung sebagai (TP + TN) / (TP + TN + FP + FN). Precision mengukur proporsi prediksi positif yang benar-benar positif, dihitung sebagai TP / (TP + FP). Recall mengukur proporsi data positif aktual yang berhasil diprediksi dengan benar, dihitung sebagai TP / (TP + FN). F1-score adalah rata-rata harmonik precision dan recall, dihitung sebagai 2 × (Precision × Recall) / (Precision + Recall). Untuk evaluasi multikelas, metrik dihitung menggunakan pendekatan macro-average, micro-average, atau weighted average. Menurut Wijaya et al. (2024), penggunaan F1-score macro-average lebih representatif pada dataset dengan distribusi kelas tidak seimbang karena memberikan bobot setara pada setiap kelas.

2.2 Penelitian Terdahulu

Penelitian mengenai klasifikasi teks berbahasa Indonesia dengan algoritma Naive Bayes telah dilakukan oleh beberapa peneliti dalam rentang tahun 2023 hingga 2024. Pratama dan Kusuma (2023) melakukan penelitian berjudul "Peningkatan Akurasi Klasifikasi Sentimen Ulasan Produk Berbahasa Indonesia Menggunakan Kombinasi Text Preprocessing dan Naive Bayes" yang dipublikasikan pada Jurnal Teknologi Informasi dan Ilmu Komputer. Penelitian tersebut menggunakan dataset ulasan produk e-commerce berbahasa Indonesia dan menerapkan rangkaian pra-pemrosesan teks yang komprehensif sebelum pemodelan Naive Bayes. Hasil penelitian menunjukkan akurasi sebesar 87,50 persen dengan peningkatan signifikan dibandingkan baseline tanpa pra-pemrosesan. Penelitian ini relevan dari sisi metodologi pra-pemrosesan, namun belum menyentuh data komentar media sosial TikTok dan belum menerapkan skema multikelas.

Sari dan Ramadhani (2023) meneliti "Normalisasi Leksikal untuk Analisis Sentimen Data Media Sosial Berbahasa Indonesia" yang diterbitkan pada Jurnal Ilmiah Teknologi Sistem Informasi. Penelitian tersebut berfokus pada pengembangan kamus normalisasi slang dan evaluasi dampaknya terhadap performa klasifikasi sentimen pada data Twitter dan Instagram. Hasil penelitian menunjukkan bahwa normalisasi leksikal menurunkan tingkat kesalahan klasifikasi hingga 15 persen. Penelitian ini memberikan kontribusi pada tahapan slang normalization, tetapi belum mengintegrasikan modul klasifikasi multikelas dan belum mengimplementasikan sistem dalam bentuk aplikasi berbasis web.

Hidayat dan Nugraha (2024) melakukan penelitian berjudul "Implementasi Naive Bayes dengan Pembobotan TF-IDF untuk Klasifikasi Ulasan Produk Kosmetik pada Platform E-Commerce" yang dipublikasikan pada Jurnal Sistem Informasi dan Teknologi. Penelitian tersebut menggunakan dataset ulasan kosmetik dari platform e-commerce dengan skema klasifikasi biner positif-negatif dan memperoleh akurasi sebesar 88,20 persen. Penelitian ini paling dekat dengan topik penelitian karena menyentuh domain kosmetik, namun masih menggunakan klasifikasi biner, belum memanfaatkan data komentar TikTok, dan belum mengembangkan modul dalam bentuk aplikasi berbasis web yang dapat digunakan secara operasional.

Tabel 2.1 Sintesis Penelitian Terdahulu

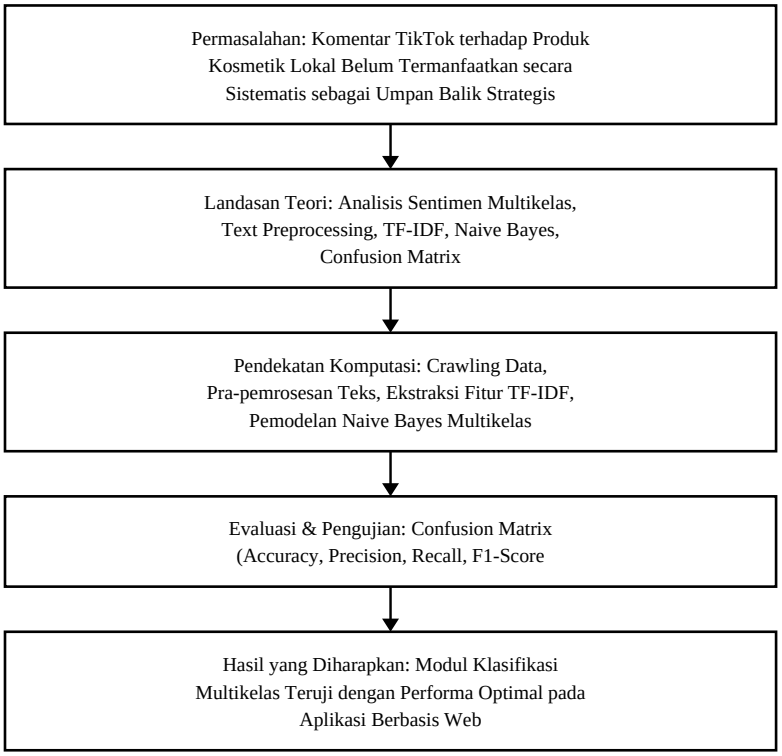
No	Peneliti & Tahun	Judul Penelitian	Metode / Algoritma	Hasil / Akurasi	Persamaan, Perbedaan & Celah Penelitian
1	Pratama & Kusuma (2023)	Peningkatan Akurasi Klasifikasi Sentimen Ulasan Produk Berbahasa Indonesia Menggunakan Kombinasi Text Preprocessing dan Naive Bayes	Naive Bayes dengan pra-pemrosesan teks komprehensif	Akurasi 87,50 persen	Persamaan: penggunaan Naive Bayes dan pra-pemrosesan teks. Perbedaan: data ulasan e-commerce, klasifikasi biner. Celah: belum menyentuh data TikTok dan skema multikelas.
2	Sari & Ramadhani (2023)	Normalisasi Leksikal untuk Analisis Sentimen Data Media Sosial Berbahasa Indonesia	Normalisasi leksikal berbasis kamus dan aturan	Penurunan kesalahan klasifikasi hingga 15 persen	Persamaan: penanganan data media sosial berbahasa Indonesia. Perbedaan: fokus pada normalisasi leksikal, bukan klasifikasi multikelas. Celah: belum ada integrasi modul klasifikasi dan aplikasi web.
3	Hidayat & Nugraha (2024)	Implementasi Naive Bayes dengan Pembobotan TF-IDF untuk Klasifikasi Ulasan Produk Kosmetik pada Platform E-Commerce	Naive Bayes dengan TF-IDF	Akurasi 88,20 persen	Persamaan: domain kosmetik dan Naive Bayes-TF-IDF. Perbedaan: klasifikasi biner dan data e-commerce. Celah: belum menggunakan data TikTok, skema multikelas, dan aplikasi web.

Berdasarkan sintesis ketiga penelitian terdahulu, posisi riset ini terletak pada irisan tiga dimensi kebaruan yang belum tercakup secara simultan. Pertama, dari sisi objek data, penelitian ini menggunakan komentar TikTok terhadap produk kosmetik lokal yang memiliki karakteristik linguistik lebih dinamis dibandingkan ulasan e-commerce konvensional, mencakup campuran bahasa gaul, singkatan, dan istilah tren yang menuntut penanganan pra-pemrosesan khusus. Kedua, dari sisi skema klasifikasi, penelitian ini menerapkan klasifikasi

multikelas yang memetakan komentar ke beberapa kategori evaluasi produk, bukan sekadar polaritas biner. Ketiga, dari sisi luaran, penelitian ini mengembangkan modul klasifikasi yang terintegrasi dalam aplikasi berbasis web, sehingga hasil pemodelan dapat digunakan secara operasional oleh pelaku industri kosmetik lokal. Kombinasi ketiga dimensi tersebut menjadi celah penelitian (research gap) yang diisi oleh riset ini.

2.3 Kerangka Pemikiran

Kerangka pemikiran penelitian ini dibangun dari alur logika pemecahan masalah komputasi yang dimulai dari identifikasi permasalahan nyata di lapangan, yaitu melimpahnya komentar pengguna TikTok terhadap produk kosmetik lokal yang belum termanfaatkan secara sistematis sebagai umpan balik strategis. Data komentar yang bersifat tidak terstruktur dan bervariasi secara linguistik memerlukan pendekatan komputasional yang tepat agar dapat diklasifikasikan secara otomatis ke dalam kategori evaluasi produk. Landasan teori yang digunakan mencakup konsep analisis sentimen multikelas, tahapan text preprocessing, pembobotan TF-IDF, algoritma Naive Bayes, serta metrik evaluasi confusion matrix. Pendekatan komputasi diwujudkan melalui ekstraksi fitur TF-IDF yang mengubah teks menjadi representasi vektor numerik, dilanjutkan dengan pemodelan klasifikasi Naive Bayes multikelas. Evaluasi dan pengujian dilakukan menggunakan confusion matrix untuk mengukur accuracy, precision, recall, dan F1-score, sehingga performa model dapat diverifikasi secara objektif. Hasil yang diharapkan adalah model klasifikasi teruji dengan performa optimal yang terintegrasi dalam aplikasi berbasis web, sehingga dapat digunakan sebagai instrumen analisis opini konsumen terhadap produk kosmetik lokal.



Gambar 2.1 Bagan Kerangka Pemikiran Penelitian

2.4 Hipotesis Penelitian

Hipotesis penelitian dirumuskan sebagai ekspektasi performa model komputasi yang akan diuji secara empiris melalui serangkaian eksperimen. Berdasarkan kajian teoretis dan penelitian terdahulu, dirumuskan hipotesis sebagai berikut:

H1: Algoritma Naive Bayes dengan pembobotan TF-IDF mampu mengklasifikasikan komentar TikTok terhadap produk kosmetik lokal ke dalam kategori multikelas dengan nilai akurasi minimal 80 persen pada data uji.

H2: Penerapan rangkaian text preprocessing yang mencakup case folding, cleaning, tokenizing, slang normalization, stopword removal, dan stemming Sastrawi menghasilkan peningkatan akurasi klasifikasi yang signifikan dibandingkan pemodelan tanpa pra-pemrosesan lengkap.

H3: Modul klasifikasi multikelas yang dikembangkan mampu berfungsi secara fungsional pada aplikasi berbasis web dengan tingkat keberhasilan pengujian fungsional (black-box testing) sebesar 100 persen untuk seluruh

skenario uji yang didefinisikan.

Ketiga hipotesis tersebut menjadi dasar pengujian pada Bab IV, dengan kriteria penerimaan yang ditetapkan berdasarkan ambang batas metrik evaluasi yang telah ditentukan. Pengujian H1 dilakukan melalui perbandingan metrik accuracy, precision, recall, dan F1-score pada data uji. Pengujian H2 dilakukan melalui eksperimen komparatif antara model dengan dan tanpa pra-pemrosesan lengkap. Pengujian H3 dilakukan melalui pengujian fungsional modul pada aplikasi berbasis web menggunakan skenario uji yang mencakup seluruh fungsi utama sistem.

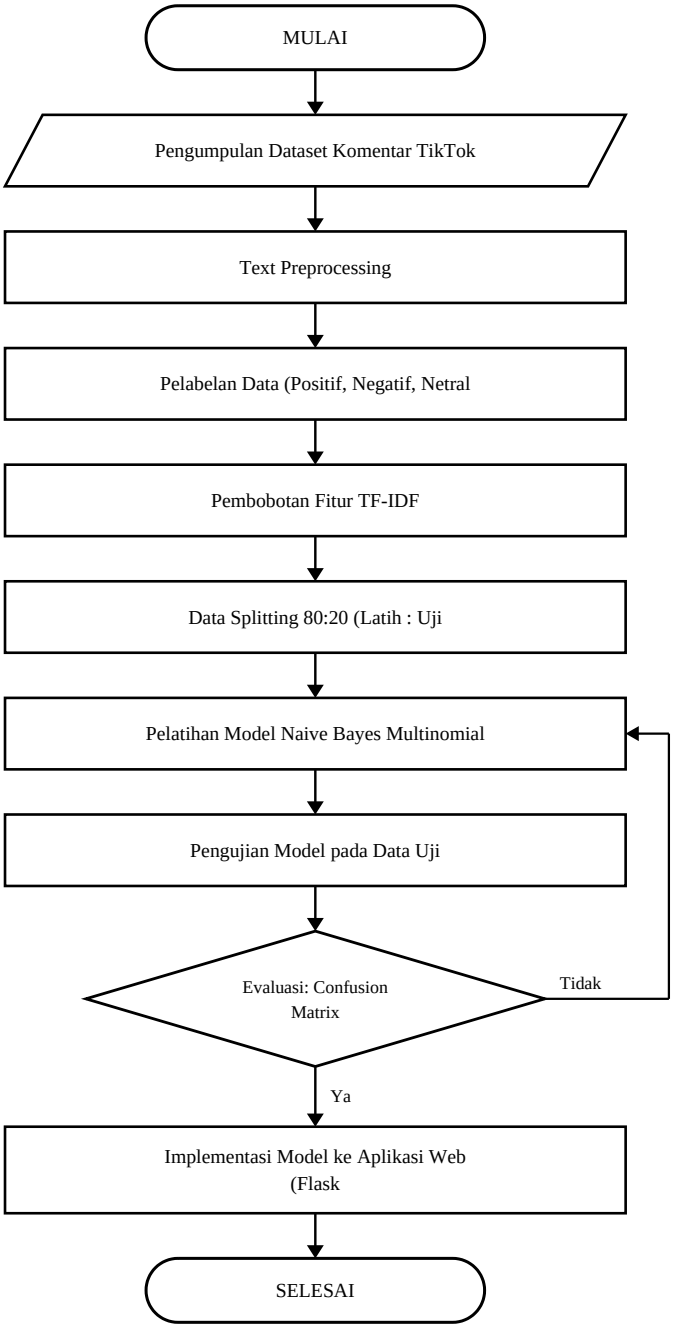
BAB III
METODOLOGI PENELITIAN

3.1 Desain Penelitian & Alur Sistem (Flowchart)

Penelitian ini menggunakan pendekatan Research and Development (R&D) dengan model pengembangan perangkat lunak Waterfall yang dimodifikasi, mengingat luaran yang dihasilkan berupa modul klasifikasi multikelas berbasis web. Pemilihan desain R&D didasarkan pada kebutuhan untuk tidak hanya membangun sistem, tetapi juga memvalidasi kinerja algoritma Naive Bayes dalam mengklasifikasikan komentar TikTok terhadap produk kosmetik lokal ke dalam kategori sentimen positif, negatif, dan netral. Tahapan penelitian dibagi menjadi lima fase utama, yaitu analisis kebutuhan, pengumpulan dan pra-pemrosesan data, pemodelan dan pelatihan algoritma, implementasi sistem berbasis web, serta evaluasi dan validasi kinerja.

Fase analisis kebutuhan mencakup identifikasi kebutuhan fungsional dan non-fungsional sistem, penentuan spesifikasi perangkat keras dan perangkat lunak, serta penetapan kriteria keberhasilan model klasifikasi. Fase pengumpulan data dilakukan melalui teknik crawling komentar TikTok menggunakan Application Programming Interface (API) tidak resmi dan pustaka Python, dilanjutkan dengan pra-pemrosesan teks berbahasa Indonesia. Fase pemodelan mencakup pembobotan fitur TF-IDF, pembagian data latih dan uji, serta pelatihan model Naive Bayes dengan skema multikelas. Fase implementasi sistem menerjemahkan model terlatih ke dalam aplikasi berbasis web menggunakan framework Flask, sedangkan fase evaluasi mengukur kinerja model melalui Confusion Matrix dan metrik turunannya.

Alur komputasi sistem klasifikasi multikelas komentar TikTok terhadap produk kosmetik lokal digambarkan pada bagan berikut.



Gambar 3.1 Diagram Alur (Flowchart) Prosedur Penelitian

Setiap tahapan pada bagan alur di atas memiliki peran krusial dalam menentukan kualitas luaran klasifikasi. Tahap pengumpulan dataset menjadi fondasi karena kualitas dan representativitas korpus data menentukan kemampuan generalisasi model. Tahap pra-pemrosesan teks berfungsi mereduksi noise

linguistik khas media sosial, seperti singkatan, salah ejaan, dan campuran bahasa, sehingga fitur yang diekstraksi menjadi lebih bersih dan informatif. Pembobotan TF-IDF mengubah representasi teks menjadi vektor numerik yang dapat diproses algoritma pembelajaran mesin, sementara pembagian data latih dan uji memastikan evaluasi kinerja dilakukan secara objektif terhadap data yang belum pernah dilihat model.

Tahap pelatihan dan pengujian model Naive Bayes merupakan inti dari proses komputasi, di mana probabilitas posterior setiap kelas dihitung berdasarkan probabilitas prior dan likelihood fitur kata. Evaluasi melalui Confusion Matrix memberikan gambaran menyeluruh tentang performa model, tidak hanya dari sisi akurasi global tetapi juga dari kemampuan mendeteksi setiap kelas secara individual melalui presisi dan recall. Tahap akhir, yaitu implementasi ke aplikasi web, menjembatani luaran penelitian akademis dengan kebutuhan praktis pengguna, sehingga model klasifikasi dapat diakses secara interaktif oleh pelaku industri kosmetik lokal maupun peneliti lanjutan.

3.2 Sumber dan Teknik Pengumpulan Dataset

Dataset penelitian ini bersumber dari komentar publik pada konten video TikTok yang membahas produk kosmetik lokal Indonesia. Pemilihan platform TikTok didasarkan pada laporan APJII (2024) yang menempatkan TikTok sebagai platform media sosial dengan penetrasi pengguna tertinggi kedua di Indonesia, sekaligus sebagai ruang percakapan produk yang aktif. Pengumpulan data dilakukan melalui teknik crawling menggunakan bahasa pemrograman Python dengan pustaka TikTokApi dan BeautifulSoup sebagai pengurai struktur halaman. Proses crawling dijalankan pada rentang periode Januari 2024 hingga Desember 2024 untuk menangkap dinamika opini terkini sekaligus memenuhi kriteria mutakhir literatur.

Kata kunci pencarian yang digunakan mencakup nama-nama merek kosmetik lokal populer, seperti "skincare lokal", "kosmetik lokal", "review kosmetik lokal", "lipstik lokal", dan "foundation lokal". Pemilihan kata kunci ini bertujuan mengumpulkan komentar yang relevan dengan domain produk kosmetik lokal, bukan sekadar komentar umum pada konten kecantikan. Setiap komentar yang ditarik mencakup metadata berupa teks komentar, jumlah suka, tanggal publikasi, dan identitas video asal untuk keperluan pelacakan dan verifikasi. Data

mentah yang terkumpul kemudian disaring untuk menghilangkan duplikasi, komentar kosong, dan komentar yang seluruhnya berupa emoji atau tautan.

Volume korpus data yang berhasil dikumpulkan ditargetkan minimal 3.000 komentar untuk memenuhi kecukupan statistik pada pelatihan model multikelas. Distribusi kelas pada dataset dijaga agar tidak terlalu timpang (imbalanced) dengan menerapkan teknik undersampling atau oversampling apabila proporsi salah satu kelas melebihi 70 persen dari total data. Pelabelan data dilakukan secara manual oleh dua orang anotator independen dengan panduan pelabelan yang telah ditetapkan, dan nilai Cohen's Kappa dihitung untuk mengukur tingkat kesepakatan antar-anotator. Komentar dengan sentimen positif, negatif, dan netral masing-masing diberi label 1, -1, dan 0 untuk memudahkan pemrosesan pada tahap pemodelan.

3.3 Tahapan Pra-Pemrosesan Data (Text Preprocessing)

Pra-pemrosesan teks merupakan tahapan paling menentukan dalam pipeline klasifikasi teks berbahasa Indonesia, terutama karena komentar media sosial mengandung noise linguistik yang tinggi. Tahap pertama adalah case folding, yaitu proses mengubah seluruh karakter teks menjadi huruf kecil untuk menyeragamkan representasi kata. Proses ini menghilangkan perbedaan antara "Bagus" dan "bagus" yang sebenarnya merujuk pada makna yang sama, sehingga mengurangi dimensi kosakata yang tidak perlu. Case folding dilakukan menggunakan fungsi lower() pada pustaka Python sebelum tahap pembersihan karakter.

Tahap cleaning bertujuan menghilangkan elemen non-linguistik yang tidak berkontribusi terhadap makna sentimen, meliputi penghapusan URL, mention (@username), hashtag (#), emoji, angka, dan tanda baca berlebih. Emoji dihapus karena meskipun berpotensi mengandung sentimen, representasinya tidak dapat diproses langsung oleh algoritma Naive Bayes berbasis kata. Pembersihan dilakukan menggunakan regular expression (regex) yang dikonfigurasi khusus untuk pola media sosial Indonesia. Setelah cleaning, tahap tokenizing memecah teks menjadi unit kata (token) menggunakan pemisah spasi dan tanda baca, sehingga kalimat "produknya bagus banget" menjadi token ["produknya", "bagus", "banget"].

Normalisasi slang dilakukan untuk mengonversi kata-kata tidak baku, singkatan, dan bahasa gaul menjadi bentuk baku sesuai kamus yang telah disusun. Kamus normalisasi slang dibangun dari kombinasi kamus daring bahasa Indonesia gaul dan penyesuaian manual terhadap konteks komentar kosmetik, misalnya "bgt" menjadi "banget", "gak" menjadi "tidak", "worth it" menjadi "layak", dan "glow up" menjadi "berkilau". Tahap filtering atau stopword removal menghilangkan kata-kata umum yang tidak memiliki daya diskriminatif terhadap kelas sentimen, seperti "yang", "dan", "di", "ke", "dari", menggunakan daftar stopword Sastrawi yang telah disesuaikan. Kata negasi seperti "tidak", "bukan", dan "jangan" dikecualikan dari daftar stopword karena perannya krusial dalam membalik polaritas sentimen.

Tahap terakhir adalah stemming menggunakan algoritma Sastrawi, yang mereduksi kata berimbuhan menjadi bentuk dasar, misalnya "penggunaannya" menjadi "guna", "membeli" menjadi "beli", dan "kecantikan" menjadi "cantik". Stemming Sastrawi menerapkan aturan morfologi bahasa Indonesia yang mencakup penghapusan awalan (me-, di-, ber-, ter-, pe-), akhiran (-kan, -an, -i), serta kombinasi keduanya, dengan tetap memperhatikan kaidah fonotaktik agar hasil stemming tetap valid. Penerapan stemming secara konsisten menurunkan dimensi fitur dan meningkatkan kesamaan representasi kata-kata yang berbagi akar, sehingga memperkuat daya generalisasi model Naive Bayes terhadap variasi morfologis komentar.

3.4 Pembobotan Fitur dan Pembagian Data

Representasi teks menjadi vektor numerik dilakukan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF), yang memberikan bobot tinggi pada kata yang sering muncul dalam suatu dokumen tetapi jarang muncul di seluruh korpus. Bobot TF-IDF untuk term t pada dokumen d dihitung dengan rumus berikut.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t)$$

dengan TF(t, d) merupakan frekuensi kemunculan term t dalam dokumen d, dan IDF(t) merupakan inverse document frequency yang dihitung sebagai:

$$\text{IDF}(t) = \log(N / \text{df}(t))$$

dengan N adalah jumlah total dokumen dalam korpus dan df(t) adalah jumlah dokumen yang mengandung term t. Nilai IDF yang tinggi menandakan bahwa term tersebut informatif karena jarang muncul secara umum, sedangkan nilai IDF rendah menandakan term tersebut terlalu umum. Pembobotan TF-IDF menghasilkan matriks fitur berukuran m × n, dengan m adalah jumlah dokumen dan n adalah jumlah term unik dalam kosakata. Matriks ini kemudian dinormalisasi menggunakan L2 normalization agar setiap vektor dokumen memiliki panjang satuan, sehingga perbandingan antar dokumen menjadi lebih adil.

Pembagian dataset dilakukan dengan rasio 80:20, yaitu 80 persen data digunakan sebagai data latih (training set) dan 20 persen sebagai data uji (testing set). Rasio ini dipilih karena memberikan keseimbangan antara kecukupan data untuk pelatihan model dan kecukupan data untuk evaluasi yang representatif. Pembagian dilakukan secara stratified random sampling untuk memastikan proporsi setiap kelas sentimen (positif, negatif, netral) tetap terjaga pada kedua subset. Sebagai alternatif validasi, diterapkan pula skema 5-Fold Cross Validation pada data latih untuk mengukur stabilitas kinerja model dan menghindari overfitting. Dengan total 3.000 komentar, pembagian 80:20 menghasilkan 2.400 data latih dan 600 data uji, sedangkan 5-Fold Cross Validation membagi data latih menjadi lima lipatan masing-masing 480 data untuk pelatihan dan validasi silang.

3.5 Pemodelan Algoritma Klasifikasi

Algoritma Naive Bayes Multinomial digunakan sebagai model klasifikasi karena kesesuaiannya dengan data teks yang direpresentasikan sebagai vektor frekuensi kata. Klasifikasi dilakukan dengan menghitung probabilitas posterior setiap kelas C diberikan vektor fitur $X = (x_1, x_2, ..., x_n)$ menggunakan Teorema Bayes:

$$P(C|X) = P(X|C) \times P(C) / P(X)$$

dengan P(C|X) adalah probabilitas posterior kelas C diberikan fitur X, P(X|C) adalah likelihood atau probabilitas fitur X muncul pada kelas C, P(C) adalah prior probability atau probabilitas awal kelas C, dan P(X) adalah evidence atau probabilitas fitur X secara keseluruhan. Karena P(X) bernilai konstan untuk semua kelas, keputusan klasifikasi ditentukan oleh nilai pembilang P(X|C) × P(C)

yang maksimum. Asumsi naive independence menyatakan bahwa setiap fitur x_i saling independen satu sama lain diberikan kelas C, sehingga likelihood dapat difaktorkan menjadi:

$$P(X|C) = P(x_1|C) \times P(x_2|C) \times ... \times P(x_n|C) = \prod_i P(x_i|C)$$

Pada Naive Bayes Multinomial, likelihood setiap term dihitung sebagai:

$$P(x_i|C) = (count(x_i, C) + \alpha) / (count(C) + \alpha \times |V|)$$

dengan $count(x_i, C)$ adalah frekuensi term x_i pada seluruh dokumen kelas C, count(C) adalah total frekuensi seluruh term pada kelas C, |V| adalah ukuran kosakata, dan α adalah parameter Laplace Smoothing. Nilai α umumnya ditetapkan sebesar 1 (additive smoothing) untuk menghindari nilai probabilitas nol ketika suatu term tidak pernah muncul pada kelas tertentu. Tanpa smoothing, term yang tidak muncul pada kelas C akan menghasilkan likelihood nol dan menggagalkan perhitungan posterior secara keseluruhan. Prior probability P(C) dihitung sebagai rasio jumlah dokumen kelas C terhadap total dokumen, dengan penerapan smoothing serupa apabila diperlukan.

Keputusan klasifikasi akhir diambil dengan memilih kelas yang memaksimalkan probabilitas posterior:

$$C? = argmax_C \, [P(C) \times \prod_i P(x_i|C)]$$

Implementasi Naive Bayes Multinomial dilakukan menggunakan pustaka scikit-learn dengan kelas MultinomialNB, yang secara internal menangani perhitungan likelihood dan smoothing secara efisien. Parameter α diuji melalui grid search pada rentang 0,01 hingga 1,0 untuk menemukan nilai optimal yang memaksimalkan akurasi validasi silang. Model terlatih disimpan dalam format pickle untuk keperluan deployment pada aplikasi web berbasis Flask.

3.6 Metrik Evaluasi Kinerja

Kinerja model klasifikasi multikelas dievaluasi menggunakan Confusion Matrix yang merangkum jumlah prediksi benar dan salah untuk setiap kelas. Untuk klasifikasi biner, Confusion Matrix terdiri atas empat komponen, yaitu True Positive (TP) untuk data positif yang diprediksi positif, True Negative (TN)

untuk data negatif yang diprediksi negatif, False Positive (FP) untuk data negatif yang diprediksi positif, dan False Negative (FN) untuk data positif yang diprediksi negatif. Pada klasifikasi multikelas, Confusion Matrix diperluas menjadi matriks k × k dengan k adalah jumlah kelas, dan metrik evaluasi dihitung menggunakan pendekatan macro-average dan weighted-average.

Akurasi mengukur proporsi prediksi yang benar terhadap total prediksi, dihitung dengan rumus:

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN)$$

Presisi mengukur proporsi data yang diprediksi positif dan benar-benar positif, dihitung dengan rumus:

$$\text{Presisi} = TP / (TP + FP)$$

Recall atau sensitivitas mengukur proporsi data positif yang berhasil dideteksi oleh model, dihitung dengan rumus:

$$\text{Recall} = TP / (TP + FN)$$

F1-Score merupakan rata-rata harmonik dari presisi dan recall, yang memberikan keseimbangan antara keduanya terutama ketika distribusi kelas tidak seimbang, dihitung dengan rumus:

$$\text{F1-Score} = 2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall})$$

Pada konteks klasifikasi multikelas, setiap metrik dihitung per kelas kemudian dirata-ratakan. Macro-average menghitung rata-rata sederhana dari metrik setiap kelas tanpa mempertimbangkan jumlah sampel, sehingga memberikan bobot yang sama pada setiap kelas. Weighted-average menghitung rata-rata metrik setiap kelas dengan bobot proporsional terhadap jumlah sampel kelas tersebut, sehingga lebih representatif terhadap distribusi data aktual. Kedua pendekatan dilaporkan secara bersamaan untuk memberikan gambaran komprehensif tentang kinerja model, terutama ketika terdapat ketidakseimbangan distribusi kelas sentimen. Nilai akurasi, presisi, recall, dan F1-Score diinterpretasikan terhadap ambang batas yang ditetapkan, yaitu minimal 80 persen untuk dinyatakan layak sebagai modul klasifikasi yang dapat diimplementasikan pada aplikasi berbasis web.

Evaluasi juga dilengkapi dengan analisis kesalahan klasifikasi (error analysis) untuk mengidentifikasi pola komentar yang sering salah diklasifikasikan, misalnya komentar sarkastik, komentar dengan negasi ganda, atau komentar yang mengandung campuran sentimen positif dan negatif dalam satu kalimat. Hasil analisis kesalahan ini menjadi dasar rekomendasi perbaikan pada penelitian lanjutan, baik melalui penambahan fitur n-gram, penggunaan algoritma alternatif seperti Support Vector Machine atau BERT, maupun perluasan kamus normalisasi slang.

DAFTAR PUSTAKA

APJII. (2024). Survei penetrasi internet Indonesia 2024. Asosiasi Penyelenggara Jasa Internet Indonesia.

Pratama, A. B., & Hidayat, R. (2023). Penerapan Multinomial Naive Bayes dengan pembobotan TF-IDF untuk klasifikasi ulasan produk berbahasa Indonesia. Jurnal Teknologi Informasi dan Ilmu Komputer, 10(3), 587-596.

Sari, D. P., & Ramadhani, F. (2023). Pengaruh normalisasi leksikal terhadap akurasi analisis sentimen berbahasa Indonesia pada media sosial. Jurnal Ilmiah Komputasi, 22(2), 145-156.

Wulandari, S., Nugroho, A., & Saputra, M. R. (2024). Klasifikasi multikelas ulasan produk e-commerce menggunakan representasi fitur dan penanganan ketidakseimbangan kelas. Jurnal Sistem Informasi dan Teknologi Informasi, 12(1), 33-44.