VOL. 01. NO. 03 JANUARI 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

EVALUASI LAYANAN KETENAGAKERJAAN DIGITAL PADA ULASAN APLIKASI SIAPKERJA DI GOOGLE PLAY MENGGUNAKAN NAIVE BAYES

Iskandar[1], Rudi Kurniawan[2], Bani Nurhakim[3], Nining Rahaningsih[4], Willy Prihartono[5].

Program Studi Teknik Informatika[1]
Program Studi Rekayasa Perangkat Lunak[2]
Program Studi Manajemen Informatika[3]
Program Studi Komputerisasi Akuntansi[45]
STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
nandariskandar813@gmail.com,

(*) Corresponding Author : nandariskandar813@gmail.com
Published : 30 Januari 2026

Abstract—*SIAPkerja is a digital employment service application launched by the Indonesian Ministry of Manpower to support job matching, training, and social security programs in a single integrated platform. User reviews on the Google Play Store provide valuable feedback about the quality and effectiveness of this service, but manual analysis is difficult due to the large volume of unstructured text. This study applies a Naive Bayes–based sentiment analysis model to classify SIAPkerja reviews into positive, neutral, and negative categories. The research pipeline includes data collection from the Google Play Store, manual labeling, text preprocessing (cleaning, case folding, tokenization, normalization, stopword removal, and stemming), feature extraction using TF–IDF, handling class imbalance with SMOTE, and model evaluation using 5-fold cross-validation. The dataset consists of 1.359 reviews collected between 1 January 2024 and 30 September 2025. Experimental results show that the Multinomial Naive Bayes model achieves an average accuracy of 82.7% with a macro-averaged F1-score of 0.60. The findings indicate that negative sentiment dominates user reviews, mainly related to login and registration issues, verification and claim processes, and application stability. This study demonstrates that sentiment analysis can be used as a data-driven evaluation tool to support the improvement of digital public employment services in Indonesia*.

Keywords: *Sentiment Analysis, e-Government, Multinomial Naive Bayes, User Reviews, TF–IDF*

Abstrak—SIAPkerja merupakan aplikasi layanan ketenagakerjaan digital yang dikembangkan oleh Kementerian Ketenagakerjaan Republik Indonesia untuk mendukung proses pencarian kerja, pelatihan, serta program jaminan kehilangan pekerjaan dalam satu platform terintegrasi. Ulasan pengguna di Google Play Store menyediakan masukan penting terkait kualitas dan efektivitas layanan, namun sulit dianalisis secara manual karena jumlahnya besar dan berbentuk teks tidak terstruktur. Penelitian ini menerapkan analisis sentimen berbasis algoritma Multinomial Naive Bayes untuk mengklasifikasikan ulasan SIAPkerja ke dalam sentimen positif, netral, dan negatif. Alur penelitian meliputi pengumpulan data ulasan dari Google Play Store, pelabelan manual, preprocessing teks (pembersihan, case folding, tokenisasi, normalisasi, penghapusan stopword, dan stemming), ekstraksi fitur menggunakan TF–IDF, penanganan ketidakseimbangan kelas dengan SMOTE, serta evaluasi model menggunakan skema 5-fold cross-validation. Dataset terdiri dari 1.359 ulasan yang dikumpulkan pada rentang waktu 1 Januari 2024 hingga 30 September 2025. Hasil eksperimen menunjukkan bahwa model Multinomial Naive Bayes mencapai akurasi rata-rata sebesar 82,7% dengan nilai F1-score sebesar 0,60. Temuan ini mengindikasikan bahwa sentimen negatif mendominasi ulasan pengguna, terutama terkait kendala login dan registrasi, proses verifikasi dan klaim, serta stabilitas aplikasi. Penelitian ini menunjukkan bahwa analisis sentimen dapat dimanfaatkan sebagai alat evaluasi berbasis data untuk mendukung peningkatan kualitas layanan ketenagakerjaan digital pemerintah di Indonesia.

Kata Kunci : Analisis Sentimen, e-Government, Multinomial Naive Bayes, Ulasan Pengguna, TF–IDF

INTRODUCTION

Transformasi layanan publik ke kanal digital semakin menguat seiring pemanfaatan aplikasi seluler sebagai media utama interaksi antara pemerintah dan masyarakat. Berbagai kajian menunjukkan bahwa layanan mobile government memungkinkan pemerintah menjangkau pengguna secara lebih luas, fleksibel, dan personal, sekaligus menyediakan basis data yang kaya untuk mengevaluasi kepuasan dan pengalaman pengguna [1], [2]. Dalam konteks tersebut, aplikasi layanan publik tidak hanya dipandang sebagai saluran penyampaian informasi, tetapi juga sebagai instrumen kebijakan yang keberhasilannya perlu dinilai dari sudut pandang pengguna.

Di bidang ketenagakerjaan, Kementerian Ketenagakerjaan Republik Indonesia mengembangkan ekosistem digital SIAPkerja sebagai platform terpadu yang mengintegrasikan layanan informasi lowongan kerja, pelatihan melalui Skillhub, sertifikasi kompetensi, fasilitasi wirausaha, serta program Jaminan Kehilangan Pekerjaan (JKP) dalam satu sistem [3], [4]. Melalui aplikasi ini, pemerintah menargetkan peningkatan akses dan kualitas layanan ketenagakerjaan bagi pencari kerja, pekerja terdampak pemutusan hubungan kerja, serta pelaku usaha. Agar tujuan tersebut tercapai, diperlukan mekanisme evaluasi yang mampu menangkap persepsi dan pengalaman pengguna secara aktual berbasis data.

Umpan balik pengguna yang terekam dalam bentuk ulasan dan penilaian di toko aplikasi, khususnya Google Play Store, telah dimanfaatkan secara luas sebagai sumber data untuk mengevaluasi kualitas layanan, baik pada aplikasi pemerintah maupun layanan publik digital lainnya [2], [5]. Dibandingkan survei konvensional, ulasan aplikasi memiliki kelebihan karena dihasilkan secara sukarela, berkelanjutan, dan dekat dengan waktu pengalaman penggunaan. Analisis sentimen terhadap ulasan tersebut memungkinkan peneliti memetakan kecenderungan persepsi positif, netral, maupun negatif, sekaligus mengidentifikasi aspek teknis dan fungsional yang paling sering dikeluhkan atau diapresiasi oleh pengguna.

Dalam praktiknya, pemrosesan ulasan berbahasa Indonesia menghadapi sejumlah tantangan. Teks ulasan sering kali pendek, padat, dan memuat kata tidak baku, singkatan, bahasa gaul, serta pencampuran istilah asing yang tidak selalu tercakup dalam kamus umum [6]. Selain itu, kehadiran emoji, simbol, dan variasi penulisan kata yang beragam dapat menimbulkan derau (noise) bagi model pembelajaran mesin apabila tidak diatasi dengan tahapan prapemrosesan yang sistematis [7]. Di sisi lain, distribusi kelas sentimen dalam ulasan cenderung tidak seimbang, misalnya ketika ulasan bernada keluhan mendominasi, sehingga model klasifikasi berpotensi lebih sering memprediksi kelas mayoritas jika tidak dilakukan penanganan ketidakseimbangan seperti resampling atau penyesuaian bobot kelas [6], [8].

Berbagai penelitian telah memanfaatkan analisis sentimen berbasis pembelajaran mesin untuk mengkaji kualitas layanan aplikasi pemerintah maupun layanan publik digital yang beroperasi di Indonesia. Saputra et al.[9] menerapkan penambangan teks pada ulasan aplikasi PeduliLindungi di Google Play Store untuk mengelompokkan opini pengguna ke dalam sentimen positif dan negatif. Wulandari dan Hidayanto [2] menggunakan pendekatan serupa untuk mengukur dimensi kualitas layanan aplikasi contact tracing PeduliLindungi melalui topik dan sentimen yang muncul dalam ulasan. Pada ranah kesehatan, beberapa studi membandingkan kinerja algoritma Naive Bayes, Support Vector Machine (SVM), dan model berbasis jaringan saraf untuk menganalisis ulasan aplikasi Mobile JKN, dengan kombinasi fitur TF–IDF, penggantian kata gaul, dan teknik Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas [6], [8], [10]. Di luar sektor kesehatan, analisis sentimen juga telah diterapkan pada berbagai aplikasi layanan publik dan e-government untuk memetakan kepuasan dan keluhan pengguna, umumnya dengan kombinasi algoritma Naive Bayes, SVM, dan fitur TF–IDF [11], [12].

Meskipun demikian, kajian yang secara khusus menyoroti platform ketenagakerjaan digital milik pemerintah masih relatif terbatas. Studi-studi yang ada cenderung berfokus pada aplikasi kesehatan, layanan berita, atau portal e-government generik, sementara layanan ketenagakerjaan seperti SIAPkerja belum banyak diangkat sebagai studi kasus utama [1]. Padahal, SIAPkerja didesain sebagai ekosistem ketenagakerjaan terpadu yang menghubungkan fitur Karirhub, Skillhub, Sertihub, Bizhub, dan JKP dalam satu aplikasi, sehingga persepsi pengguna terhadap stabilitas teknis, kemudahan registrasi, kelengkapan informasi, dan proses klaim menjadi parameter penting keberhasilan kebijakan [3], [4], [13]. Selain itu, banyak penelitian sebelumnya hanya menjelaskan tahapan prapemrosesan pada level umum tanpa mendokumentasikan secara rinci pembangunan kamus normalisasi dan konfigurasi eksperimen, sehingga menyulitkan replikasi dan perbandingan hasil [7].

Berangkat dari kondisi tersebut, penelitian ini berupaya mengisi celah kajian dengan melakukan evaluasi layanan SIAPkerja melalui

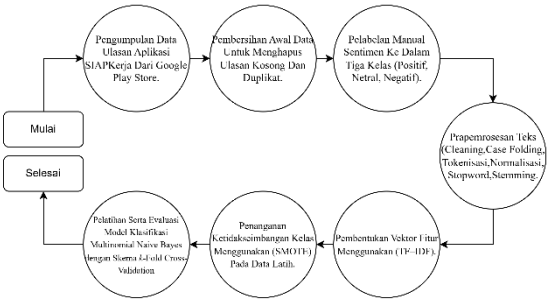
analisis sentimen ulasan pengguna di Google Play Store menggunakan algoritma Multinomial Naive Bayes. Secara spesifik, penelitian ini merumuskan tiga pertanyaan: (1) bagaimana distribusi polaritas sentimen ulasan pengguna terhadap SIAPkerja dalam kategori positif, netral, dan negatif; (2) kata kunci dan tema apa saja yang paling dominan pada ulasan bernada negatif yang dapat menunjukkan aspek layanan yang perlu diperbaiki; dan (3) seberapa baik kinerja model Multinomial Naive Bayes dalam mengklasifikasikan sentimen ulasan setelah menerapkan rangkaian prapemrosesan teks, ekstraksi fitur TF–IDF, serta penyeimbangan kelas dengan SMOTE. Untuk menjawab pertanyaan tersebut, penelitian ini menyusun alur kerja yang terdokumentasi mulai dari pengumpulan data, pelabelan manual sentimen, prapemrosesan teks berbahasa Indonesia, pembangunan model, hingga evaluasi menggunakan skema k-fold cross-validation.

Secara garis besar, penelitian ini memberikan tiga kontribusi utama. Pertama, penelitian ini menghadirkan gambaran empiris mengenai persepsi pengguna terhadap layanan ketenagakerjaan digital nasional berbasis data ulasan aplikasi resmi, sekaligus menampilkan aspek layanan yang paling sering dikritik maupun diapresiasi. Kedua, penelitian ini menyajikan sebuah pipeline prapemrosesan dan pemodelan analisis sentimen yang dapat direplikasi untuk ulasan aplikasi berbahasa Indonesia dengan menggabungkan normalisasi berbasis kamus kustom, representasi fitur TF–IDF, dan SMOTE pada model Multinomial Naive Bayes. Ketiga, penelitian ini menyusun rekomendasi praktis bagi pembuat kebijakan dan pengembang sistem di sektor ketenagakerjaan mengenai pemanfaatan hasil analisis sentimen sebagai bahan pemantauan berkelanjutan dan dasar perbaikan layanan publik digital. Struktur artikel ini disusun sebagai berikut: Bagian 2 menjelaskan kerangka penelitian dan metodologi, Bagian 3 memaparkan hasil dan pembahasan, sedangkan Bagian 4 menyajikan kesimpulan serta arah penelitian selanjutnya.

MATERIALS AND METHODS

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis pembelajaran mesin terawasi (supervised machine learning). Secara umum, alur penelitian disusun dalam beberapa tahap utama, yaitu: (1) pengumpulan data ulasan aplikasi SIAPkerja dari Google Play Store; (2) pembersihan awal data untuk menghapus ulasan kosong dan duplikat; (3) pelabelan manual sentimen ke dalam tiga kelas (positif, netral, negatif); (4) prapemrosesan teks berbahasa Indonesia; (5) pembentukan vektor fitur menggunakan Term Frequency Inverse Document Frequency (TF–IDF) ;(6) penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Oversampling Technique (SMOTE) pada data latih; dan (7) pelatihan serta evaluasi model klasifikasi Multinomial Naive Bayes dengan skema k-fold cross-validation.

Alur tersebut dirancang mengikuti prinsip proses penemuan pengetahuan dari basis data (Knowledge Discovery in Databases/KDD), di mana data mentah yang diperoleh dari toko aplikasi secara bertahap diolah menjadi informasi yang terstruktur dan dapat ditafsirkan. Pada tahap akhir, hasil klasifikasi dan analisis kata kunci digunakan untuk menyusun rekomendasi perbaikan layanan. Secara visual, kerangka penelitian ini dapat digambarkan dalam bentuk diagram alir yang menampilkan urutan tahapan dari pengumpulan data hingga penarikan kesimpulan.



Gambar 1. Alur Penelitian

1. Dataset dan Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi SIAPkerja yang dipublikasikan pada Google Play Store. Pengambilan data dilakukan dengan memanfaatkan skrip Python dan pustaka pihak ketiga yang mendukung proses scraping ulasan aplikasi secara terprogram. Rentang waktu pengambilan data ditetapkan sejak 1 Januari 2024 hingga 30 September 2025, sehingga mencakup periode implementasi dan pemanfaatan SIAPkerja pada fase awal serta perkembangan berikutnya.

Pada tahap pengumpulan, setiap entri ulasan yang diambil umumnya memuat informasi berupa teks ulasan, rating bintang, tanggal publikasi, serta beberapa atribut tambahan seperti nama pengguna dan versi aplikasi. Data mentah hasil scraping kemudian disimpan dalam format berkas teks terstruktur CSV untuk memudahkan proses analisis lanjutan. Jumlah ulasan mentah yang berhasil dikumpulkan pada rentang waktu tersebut adalah 1.359 entri.

Tabel 1. Ringkasan Pengumpulan Data Ulasan Aplikasi SIAPkerja

Sumber data	Rentang waktu pengambilan	Jumlah ulasan mentah	Atribut
Google Play Store aplikasi SIAPkerja	1 Januari 2024 – 30 September 2025	1.359 ulasan	Teks ulasan, rating bintang, tanggal ulasan, nama pengguna

2. Skema Pelabelan Manual Sentimen

Setelah data terkumpul, setiap ulasan terlebih dahulu diberi label sentimen secara manual oleh peneliti. Skema pelabelan yang digunakan membagi ulasan ke dalam tiga kategori, yaitu: (1) sentimen positif, apabila ulasan secara dominan berisi apresiasi, pujian, atau pengalaman penggunaan yang dinilai baik; (2) sentimen negatif, apabila ulasan didominasi keluhan, kekecewaan, atau pengalaman penggunaan yang dinilai buruk; dan (3) sentimen netral, apabila ulasan bersifat informatif, ambigu, atau mengandung campuran unsur positif dan negatif tanpa kecenderungan yang kuat.

Dalam proses pelabelan, rating bintang yang diberikan pengguna digunakan sebagai salah satu sumber informasi pendukung, namun keputusan akhir tetap ditentukan berdasarkan isi teks ulasan. Ulasan dengan rating bintang 1-2 umumnya cenderung dilabeli negatif, rating 4-5 cenderung positif, sedangkan rating 3 ditinjau lebih cermat karena dapat mewakili sentimen netral maupun campuran. Untuk menjaga konsistensi, peneliti menyusun panduan pelabelan sederhana yang berisi contoh kalimat karakteristik untuk masing-masing kelas. Setelah seluruh ulasan selesai dilabeli, dilakukan tahap pembersihan lanjutan untuk menghapus entri duplikat dan ulasan kosong atau tidak relevan. Dari 1.359 ulasan awal, proses ini menghasilkan 1.322 ulasan yang digunakan pada tahap pemrosesan dan pemodelan berikutnya. Distribusi jumlah ulasan pada tiap kelas setelah pelabelan manual dan pembersihan data bersifat tidak seimbang dan dirangkum pada bagian hasil.

Tabel 2. Distribusi Kelas Sentimen Hasil Pelabelan Manual Ulasan SIAPkerja

Kelas Sentimen	Jumlah ulasan	Persentase (%)
Positif	153	11,57
Netral	72	5,45
Negatif	1.097	82,98
Total	1.322	100,00

3. Prapemrosesan Teks

Tahap prapemrosesan (preprocessing) dilakukan untuk meningkatkan kualitas teks sebelum digunakan sebagai masukan model pembelajaran mesin. Mengingat ulasan aplikasi berbahasa Indonesia banyak mengandung kata tidak baku, singkatan, dan variasi penulisan, tahapan prapemrosesan dirancang secara berlapis sebagai berikut:

a. Pembersihan teks (cleaning)**:** menghapus URL, alamat surel, hashtag, mention, angka yang tidak relevan, tanda baca berulang, serta emoji atau simbol tertentu yang tidak dibutuhkan dalam analisis. Pada tahap ini juga dilakukan penggantian karakter berulang (misalnya "baguuuus" menjadi "bagus").
b. Case folding**:** mengubah seluruh huruf menjadi lowercase untuk menghindari pembedaan istilah hanya karena perbedaan kapitalisasi.
c. Tokenisasi**:** memecah teks ulasan menjadi satuan kata (token) dengan memanfaatkan pemisah spasi dan tanda baca, sehingga setiap dokumen direpresentasikan sebagai deretan token.
d. Normalisasi kata tidak baku**:** menerapkan kamus normalisasi kustom yang disusun berdasarkan eksplorasi awal data dan daftar kata tidak baku yang umum muncul dalam ulasan aplikasi. Contohnya, kata "gk", "ga", atau "gak" dinormalisasi menjadi "tidak"; "apk" menjadi "aplikasi"; dan variasi penulisan istilah yang merujuk pada fitur SIAPkerja diseragamkan.
e. Penghapusan stopword**:** menghapus kata-kata umum yang frekuensinya tinggi tetapi kontribusi informatifnya rendah terhadap pembedaan sentimen, seperti kata hubung dan kata tugas tertentu dalam bahasa Indonesia.
f. Stemming**:** mengembalikan setiap kata ke bentuk dasar (lematis) menggunakan stemmer bahasa Indonesia, sehingga variasi imbuhan seperti "mendaftar", "pendaftaran", dan "terdaftar" dipetakan ke bentuk akar yang sama.

4. Ekstraksi Fitur dengan TF–IDF

Setelah prapemrosesan, setiap dokumen ulasan direpresentasikan dalam bentuk vektor numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF–IDF). TF–IDF menggabungkan informasi frekuensi kemunculan kata dalam satu dokumen dengan seberapa jarang kata tersebut muncul pada keseluruhan korpus. Secara konseptual, bobot TF–IDF untuk sebuah kata akan tinggi apabila kata tersebut sering muncul pada dokumen tertentu, tetapi relatif jarang muncul

pada dokumen lain, sehingga kata tersebut dianggap lebih informatif untuk membedakan dokumen.

Dalam penelitian ini, skema TF–IDF diterapkan pada tingkat unigram, sehingga setiap fitur merepresentasikan satu kata dasar hasil stemming. Nilai frekuensi term (TF) dihitung sebagai jumlah kemunculan kata dalam suatu dokumen yang dinormalisasi terhadap panjang dokumen, sedangkan nilai inverse document frequency (IDF) dihitung dengan mempertimbangkan jumlah dokumen yang mengandung kata tersebut. Perkalian antara TF dan IDF menghasilkan matriks fitur berdimensi dokumen × kata yang kemudian menjadi masukan bagi model Multinomial Naive Bayes.

Untuk mengurangi derau, kata yang terlalu jarang muncul dapat dihilangkan dengan menetapkan ambang batas frekuensi minimum. Sebaliknya, kata yang muncul hampir di seluruh dokumen dapat dibatasi kontribusinya agar tidak mendominasi perhitungan bobot. Pengaturan parameter TF-IDF disesuaikan melalui pengujian awal sehingga diperoleh keseimbangan antara kompleksitas model dan kemampuan generalisasi.

5. Penanganan Ketidakseimbangan Kelas dengan SMOTE

Distribusi kelas sentimen pada ulasan SIAPkerja menunjukkan ketidakseimbangan, di mana salah satu kelas umumnya sentimen negatif memiliki jumlah sampel jauh lebih banyak dibandingkan kelas lain. Ketidakseimbangan ini dapat menyebabkan model cenderung memprediksi kelas mayoritas dan mengabaikan pola pada kelas minoritas. Untuk mengurangi bias tersebut, penelitian ini menerapkan teknik Synthetic Minority Oversampling Technique (SMOTE) pada data latih**.**

Secara garis besar, SMOTE bekerja dengan membangkitkan sampel sintetis baru untuk kelas minoritas melalui interpolasi di antara tetangga terdekat dalam ruang fitur. Pada setiap langkah, algoritma memilih satu sampel minoritas, mencari sejumlah k tetangga terdekatnya, kemudian membangkitkan titik baru di sepanjang garis yang menghubungkan sampel tersebut dengan salah satu tetangga. Proses ini diulang hingga jumlah sampel kelas minoritas mendekati jumlah sampel kelas mayoritas. Dalam penelitian ini, SMOTE hanya diterapkan pada data latih setelah transformasi TF–IDF, sementara data uji dibiarkan dalam distribusi aslinya untuk menjaga keaslian evaluasi.

Penerapan SMOTE diintegrasikan menggunakan pustaka imbalanced-learn pada lingkungan Python sehingga parameter seperti sampling strategy dan jumlah tetangga (k-nearest neighbors) dapat dikonfigurasi. Melalui pendekatan ini, diharapkan model Multinomial Naive Bayes memiliki kesempatan yang lebih seimbang untuk mempelajari pola setiap kelas sentimen.

6. Klasifikasi Multinomial Naive Bayes dan Skema Evaluasi

Model utama yang digunakan dalam penelitian ini adalah Multinomial Naive Bayes. Algoritma ini banyak digunakan untuk tugas klasifikasi teks karena mampu menangani fitur berupa frekuensi atau bobot kemunculan kata dan relatif efisien secara komputasi. Inti dari Naive Bayes adalah penerapan teorema Bayes dengan asumsi bahwa fitur-fitur (dalam hal ini kata-kata pada dokumen) bersifat saling bebas bersyarat terhadap kelas. Pada varian multinomial, model memodelkan distribusi kemunculan kata dalam setiap kelas dan menghitung peluang sebuah dokumen termasuk ke dalam suatu kelas berdasarkan produk peluang tiap kata yang muncul.

Dalam implementasinya, matriks fitur TF–IDF hasil dari Subbagian 2.5 digunakan sebagai masukan bagi model Multinomial Naive Bayes. Dataset dibagi menjadi data latih dan data uji dengan proporsi tertentu (misalnya 80%:20%). Model terlebih dahulu dilatih pada data latih yang telah diseimbangkan menggunakan SMOTE, kemudian dievaluasi kinerjanya pada data uji yang tidak mengalami oversampling. Selain itu, untuk memperoleh gambaran yang lebih stabil mengenai kinerja model, penelitian ini menggunakan skema k-fold cross-validation (dengan k = 5), di mana data secara bergantian berperan sebagai data latih dan data uji dalam lima lipatan pembagian.

Metrik evaluasi yang digunakan meliputi akurasi, presisi, recall, dan F1-score. Presisi mengukur proporsi prediksi positif yang benar, recall mengukur proporsi sampel positif yang berhasil dikenali, sedangkan F1-score merupakan rata-rata harmonik antara presisi dan recall. Selain metrik per kelas, penelitian ini juga menghitung rata-rata makro (macro-average) untuk menilai kinerja model secara seimbang pada semua kelas. Confusion matrix disajikan untuk menggambarkan pola kesalahan klasifikasi antar kelas sentimen. Hasil rata-rata dan simpangan baku dari akurasi dan F1-score pada skema k-fold cross-validation digunakan untuk menilai konsistensi kinerja model.

RESULTS AND DISCUSSION

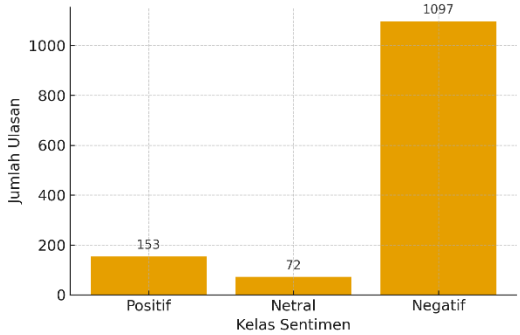
1. Karakteristik Dataset

Bagian ini memaparkan gambaran umum dataset ulasan aplikasi SIAPkerja yang digunakan dalam penelitian. Hasil pembersihan awal terhadap data mentah menunjukkan bahwa dari total 1.359 ulasan yang diperoleh melalui proses scraping Google Play Store, terdapat 37 entri yang harus

dihapus karena bersifat duplikat atau tidak valid (misalnya kosong, hanya berisi simbol, atau tidak relevan dengan konteks aplikasi). Setelah proses pembersihan tersebut, jumlah ulasan yang layak dianalisis adalah 1.322 entri.

Distribusi kelas sentimen dalam 1.322 ulasan bersih menunjukkan ketidakseimbangan yang cukup mencolok. Kelas sentimen negatif mendominasi dataset dengan jumlah 1.097 ulasan (82,98%), sedangkan kelas positif dan netral masing-masing berjumlah 153 ulasan (11,57%) dan 72 ulasan (5,45%). Kondisi ini mengindikasikan bahwa mayoritas pengguna yang menuliskan ulasan di Google Play Store menyampaikan keluhan atau ketidakpuasan terhadap aplikasi SIAPkerja, sementara ulasan yang bernada apresiatif maupun netral relatif lebih sedikit.

Untuk keperluan pemodelan, dataset kemudian dibagi menjadi data latih dan data uji dengan proporsi 80% : 20% menggunakan skema stratified split sehingga proporsi masing-masing kelas tetap terjaga. Pembagian ini menghasilkan 979 ulasan sebagai data latih dan 245 ulasan sebagai data uji. Secara lebih rinci, kelas positif terdiri dari 90 data latih dan 23 data uji, kelas netral terdiri dari 48 data latih dan 12 data uji, sedangkan kelas negatif terdiri dari 841 data latih dan 210 data uji.



Gambar 2. Distribusi kelas sentimen Setelah Pembersihan Data

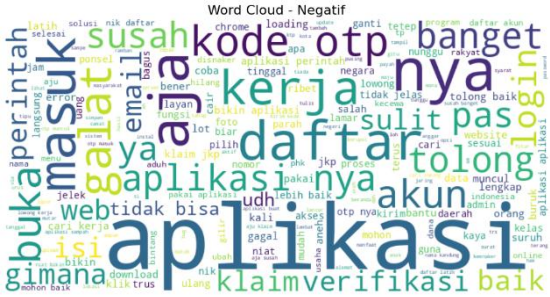
2. Hasil Prapemrosesan Teks dan Distribusi Istilah

Tahap prapemrosesan teks menghasilkan kolom baru yang berisi representasi ulasan dalam bentuk teks bersih yang telah melalui proses cleaning, case folding, tokenisasi, normalisasi kata tidak baku, penghapusan stopword, dan stemming. Secara kualitatif, tahapan ini berhasil mengurangi variasi penulisan yang tidak perlu serta menstandarkan istilah yang sering muncul dalam ulasan. Misalnya, berbagai bentuk penulisan kata "gk", "ga", dan "gak" dipetakan menjadi "tidak", sementara istilah "apk" diseragamkan menjadi "aplikasi". Ulasan yang semula mengandung banyak emoji, pengulangan huruf, dan kombinasi huruf angka menjadi lebih ringkas dan terstruktur setelah dipraproses.

Penerapan TF-IDF pada teks hasil prapemrosesan menghasilkan matriks fitur yang merepresentasikan kontribusi relatif setiap kata terhadap dokumen. Analisis terhadap bobot TF-IDF memperlihatkan bahwa pada kelas positif, istilah seperti "membantu", "terima kasih", "bagus", dan "bermanfaat" cenderung memiliki bobot tinggi. Sebaliknya, pada kelas negatif, kata-kata seperti "error", "gagal", "tidak bisa", "susah", "lemot", dan "keluar sendiri" lebih menonjol. Untuk kelas netral, istilah yang sering muncul antara lain "informasi", "proses", "data", dan "status" yang merefleksikan ulasan bersifat deskriptif atau menanyakan prosedur tanpa ekspresi emosional yang kuat.

Tabel 3. Ulasan Sebelum Dan Sesudah Prapemrosesan Teks

Ulasan sebelum prapemrosesan	Teks setelah prapemrosesan
jkp memang apk yg sangat luar biasa mantap, bagus 👍	jkp memang aplikasi yang sangat luar biasa mantap, bagus
Gk jelas,gk minat membantu app sampah	tidak jelas, tidak minat membantu aplikasi sampah
Bagus, membantu sekali untuk para pencari kerja	bagus, membantu sekali untuk para pencari kerja



Gambar 3. Wordcloud kata-kata dominan pada ulasan aplikasi SIAPkerja menurut kelas sentimen.

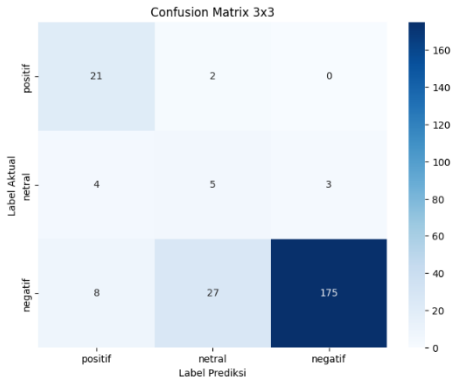
3. Performa Model Multinomial Naive Bayes

Setelah tahapan prapemrosesan dan pembentukan fitur, model Multinomial Naive Bayes dilatih menggunakan data latih yang telah diseimbangkan dengan SMOTE dan dievaluasi pada data uji dengan distribusi asli. Evaluasi pada data uji yang terdiri dari 245 ulasan menunjukkan bahwa model mencapai akurasi sebesar 0,8204 atau sekitar 82,04%. Selain akurasi, penelitian ini juga menghitung nilai Receiver Operating Characteristic–Area Under Curve (ROC–AUC) makro dan Precision–Recall Area Under Curve (PR–AUC) makro. Nilai ROC–AUC makro yang diperoleh adalah 0,8934, sedangkan PR–AUC makro bernilai 0,6863. Nilai ROC–AUC yang mendekati 0,9 menunjukkan bahwa model memiliki kemampuan pemisahan kelas yang cukup baik, sementara nilai PR–AUC menggambarkan keseimbangan antara presisi dan recall pada kondisi distribusi kelas yang tidak seimbang.

Analisis metrik per kelas menunjukkan bahwa model bekerja sangat baik pada kelas negatif, dengan nilai presisi sekitar 0,98, recall 0,83, dan F1-score 0,90 pada 210 sampel uji. Hal ini sejalan dengan dominasi kelas negatif dalam dataset sehingga model memiliki lebih banyak contoh untuk mempelajari pola kata-kata yang berasosiasi dengan keluhan. Pada kelas positif, nilai presisi mencapai 0,64 dengan recall 0,91 dan F1-score 0,75 dari 23 sampel uji, yang mengindikasikan bahwa sebagian besar ulasan positif berhasil dikenali, meskipun masih terdapat beberapa prediksi positif yang keliru. Sementara itu, kinerja model pada kelas netral relatif rendah, dengan presisi 0,15, recall 0,42, dan F1-score 0,22 dari 12 sampel uji. Nilai F1-score pada data uji adalah 0,62, sedangkan F1-score tertimbang (weighted average) mencapai 0,85, yang mencerminkan kinerja model yang baik pada kelas mayoritas tetapi masih terbatas pada kelas minoritas.

Untuk menilai konsistensi kinerja model, penelitian ini juga menerapkan skema 5-fold cross-validation pada data latih. Hasilnya menunjukkan bahwa akurasi rata-rata pada kelima lipatan adalah 0,827 dengan simpangan baku 0,020. Presisi makro rata-rata bernilai 0,575 dengan simpangan baku 0,020, recall makro 0,661 dengan simpangan baku 0,047, dan F1-score 0,596 dengan simpangan baku 0,025. Nilai simpangan baku yang relatif kecil menunjukkan bahwa performa model cukup stabil pada berbagai pembagian data, meskipun perbedaan komposisi kelas netral dan positif pada masing-masing lipatan tetap mempengaruhi variasi metrik.

Confusion matrix pada data uji memperlihatkan bahwa sebagian besar ulasan negatif berhasil diklasifikasikan dengan benar sebagai negatif, sedangkan kesalahan terbesar terjadi pada kelas netral yang sering diprediksi sebagai positif maupun negatif. Hal ini dapat dipahami karena ulasan netral dalam praktiknya sering kali berisi permintaan informasi atau deskripsi proses yang diiringi nada keluhan ringan atau apresiasi terbatas, sehingga secara semantik berada di antara kelas positif dan negatif. Beberapa ulasan positif yang sangat singkat juga terkadang diklasifikasikan sebagai netral ketika model tidak menemukan cukup banyak kata kunci positif yang kuat. Contoh kasus kesalahan klasifikasi yang representatif dapat disajikan dalam Tabel 6 untuk memberikan ilustrasi konkret mengenai keterbatasan model.



Gambar 4. Confusion matrix hasil klasifikasi Pada Data Uji

4. Pembahasan

Dominasi sentimen negatif ini mengindikasikan adanya sejumlah masalah yang dirasakan pengguna dalam memanfaatkan layanan SIAPkerja, baik dari sisi teknis maupun prosedural. Pola kata kunci yang muncul pada kelas negatif, seperti "error", "gagal", "tidak bisa", dan "susah", mengarah pada keluhan terkait kesulitan login, kegagalan proses pendaftaran, masalah verifikasi data, serta gangguan dalam proses klaim dan akses fitur tertentu. Sebaliknya, ulasan positif banyak ditandai dengan kata-kata "membantu", "bagus", dan "bermanfaat" yang biasanya muncul pada pengalaman pengguna yang berhasil mengakses

program bantuan atau merasakan manfaat fitur tertentu.

Dari sisi performa model, akurasi di atas 80% dan nilai F1-score mendekati 0,6 menegaskan bahwa kombinasi prapemrosesan teks, ekstraksi fitur TF–IDF, penyeimbangan data dengan SMOTE, dan algoritma Multinomial Naive Bayes mampu menjadi baseline yang layak untuk tugas klasifikasi sentimen ulasan aplikasi pemerintah. Kinerja yang sangat baik pada kelas negatif dan cukup baik pada kelas positif menunjukkan bahwa model efektif mempelajari pola kata kunci yang jelas merepresentasikan kepuasan maupun ketidakpuasan pengguna. Keterbatasan utama terletak pada kelas netral, yang jumlahnya sedikit dan secara semantik berada di antara kedua kelas lainnya, sehingga memerlukan pendekatan yang lebih halus, misalnya penambahan data pelatihan, pemanfaatan representasi teks yang lebih kaya, atau skema pelabelan ulang yang lebih ketat.

Temuan ini sejalan dengan hasil penelitian lain yang menunjukkan bahwa Naive Bayes dengan TF–IDF sering memberikan kinerja kompetitif sebagai model awal pada tugas analisis sentimen ulasan aplikasi, terutama ketika didukung oleh tahapan prapemrosesan yang baik dan penanganan ketidakseimbangan kelas (misalnya melalui SMOTE). Dalam konteks aplikasi layanan publik digital, dominasi sentimen negatif dan tingginya F1-score pada kelas tersebut memberikan sinyal kuat bagi pengelola layanan untuk memprioritaskan perbaikan pada titik-titik yang paling sering dikeluhkan pengguna. Pada SIAPkerja, hal ini dapat berupa peningkatan stabilitas teknis aplikasi, penyederhanaan proses registrasi dan verifikasi, serta perbaikan alur informasi terkait status pengajuan dan klaim.

Secara praktis, hasil analisis sentimen ini dapat dijadikan salah satu sumber masukan bagi Kementerian Ketenagakerjaan dan tim pengembang SIAPkerja untuk melakukan pemantauan berkala terhadap persepsi pengguna. Integrasi dasbor pemantauan sentimen yang memvisualisasikan tren positif, netral, dan negatif dari waktu ke waktu, beserta kata kunci utama per kategori, berpotensi membantu pengambil kebijakan dalam mengidentifikasi dampak perubahan fitur atau kebijakan terhadap persepsi publik. Ke depan, penelitian lanjutan dapat mengembangkan pendekatan analisis sentimen berbasis aspek (aspect-based sentiment analysis) sehingga dimensi layanan seperti kemudahan penggunaan, keandalan sistem, kejelasan informasi, dan kecepatan respons dapat dievaluasi secara lebih terperinci.

CONCLUSION

Penelitian ini bertujuan untuk mengevaluasi layanan ketenagakerjaan digital SIAPkerja melalui analisis sentimen ulasan pengguna di Google Play Store menggunakan algoritma Multinomial Naive Bayes. Dengan memanfaatkan 1.322 ulasan yang telah melalui proses pembersihan, pelabelan manual ke dalam tiga kelas sentimen, prapemrosesan teks berbahasa Indonesia, pembentukan fitur menggunakan TF–IDF, serta penyeimbangan kelas menggunakan SMOTE pada data latih, penelitian ini menunjukkan bahwa model mampu mencapai akurasi sekitar 82% dengan nilai F1-score mendekati 0,6 dan kinerja yang relatif stabil pada skema 5-fold cross-validation. Hasil ini menegaskan bahwa kombinasi metode yang digunakan dapat berfungsi sebagai baseline yang andal untuk tugas klasifikasi sentimen ulasan aplikasi pemerintah.

Dari sisi konten ulasan, distribusi sentimen memperlihatkan dominasi ulasan bernada negatif yang mencapai lebih dari delapan puluh persen dari seluruh data, sementara ulasan positif dan netral jumlahnya jauh lebih sedikit. Pola kata kunci pada ulasan negatif menunjukkan bahwa keluhan pengguna terutama berkaitan dengan kesulitan login dan registrasi, masalah verifikasi data, gangguan teknis seperti aplikasi error atau keluar sendiri, serta kendala dalam proses klaim dan pemanfaatan fitur tertentu. Sebaliknya, ulasan positif umumnya menyoroti kebermanfaatan aplikasi ketika proses berjalan lancar, misalnya dalam pengurusan program bantuan atau akses layanan ketenagakerjaan. Temuan ini memberikan indikasi kuat bahwa aspek keandalan teknis dan kejelasan prosedur merupakan titik kritis yang perlu diprioritaskan dalam perbaikan layanan SIAPkerja.

Dari perspektif pemodelan, Multinomial Naive Bayes terbukti efektif mengenali pola sentimen negatif dan positif, namun masih menghadapi tantangan dalam membedakan ulasan netral yang secara semantik berada di antara kedua kutub tersebut dan jumlah datanya relatif sedikit. Hal ini membuka ruang pengembangan lebih lanjut, antara lain dengan menambah variasi dan jumlah data pelatihan, membandingkan kinerja dengan algoritma lain seperti Support Vector Machine, regresi logistik, atau model berbasis deep learning, serta menerapkan pendekatan analisis sentimen berbasis aspek untuk memetakan dimensi layanan secara lebih rinci. Selain itu, integrasi hasil analisis sentimen ke dalam dasbor pemantauan yang diperbarui secara berkala dapat membantu pengambil kebijakan dan pengembang sistem dalam memantau dinamika persepsi publik dan

mengevaluasi dampak perubahan kebijakan maupun pembaruan fitur pada aplikasi SIAPkerja.

REFERENCE

[1] Desmal, J. A., And Et Al, "A User Satisfaction Model For Mobile Government Services," *International Journal Of Environmental Research And Public Health, 19, Article 4620*, 2022, Doi: 10.3390/Ijerph19084620.

[2] R. Wulandari And A. N. Hidayanto, "Measuring Contact Tracing Service Quality Using Sentiment Analysis: A Case Study Of Pedulilindungi Indonesia," *Quality \& Quantity*, Vol. 58, Pp. 1409–1424, 2024, Doi: 10.1007/S11135-023-01695-8.

[3] K. K. R. Indonesia, "Menaker Resmikan Portal Layanan Digital Siapkerja Di Bekasi," 2022.

[4] Google, "Siapkerja – Ekosistem Digital Ketenagakerjaan [Aplikasi Seluler]," 2025.

[5] A. Saputra, R. C. S. Hariyono, And N. M. Saraswati, "Analisis Sentimen Pengguna Aplikasi Mypertamina Menggunakan Algoritma Bidirectional Long Short Term Memory," *Jurnal Eksplora Informatika*, Vol. 13, No. 2, Pp. 156–163, 2024, Doi: 10.30864/Eksplora.V13i2.973.

[6] M. Y. A. Qahar, Y. Ruldeviyani, U. N. Mukharomah, M. A. Fidyawan, And R. Putra, "Factor Analysis Influencing Mobile JKN User Experience Using Sentiment Analysis," *International Journal Of Artificial Intelligence*, Vol. 13, No. 2, 2024, Doi: 10.11591/Ijai.V13.I2.Pp1782-1793.

[7] M. Bordoloi And S. K. Biswas, "Sentiment Analysis: A Survey On Design Framework, Applications And Future Scopes," *Artificial Intelligence Review*, Vol. 56, Pp. 12505–12560, 2023, Doi: 10.1007/S10462-023-10442-2.

[8] E. Septiani, T. M. Akhriza, And M. Husni, "Comparison Of The Accuracy Between {Naive Bayes} Classifier And Support Vector Machine Algorithms For Sentiment Analysis In {Mobile JKN} Application Reviews," *Transaction On Informatics And Data Science*, Vol. 1, No. 1, Pp. 21–32, 2024, Doi: 10.24090/Tids.V1i1.12232.

[9] I. Saputra, T. Djatna, R. R. A. Siregar, D. A. Kristiyanti, H. R. Yani, And A. A. Riyadi, "Text Mining Of Pedulilindungi Application Reviews On Google Play Store," *Faktor Exacta*, Vol. 15, No. 2, 2022, Doi: 10.30998/Faktorexacta.V15i2.10629.

[10] P. T. Putra, A. Anggrawan, And H. Hairani, "Comparison Of Machine Learning Methods For Classifying User Satisfaction Opinions Of The Pedulilindungi Application," *Matrik: Jurnal Manajemen, Teknik Informatika, Dan Rekayasa Komputer*, Vol. 22, No. 3, 2023, Doi: 10.30812/Matrik.V22i3.2860.

[11] A. I. Tanggraeni And M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naive Bayes," *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, Vol. 9, No. 2, 2022, Doi: 10.35957/Jatisi.V9i2.1835.

[12] (Alqaryouti *Et Al.*, "Aspect-Based Sentiment Analysis Using Smart Government Review Data," *DOI: Https://Doi*, 2024, Doi: 10.1016/J.Aci.2019.11.003.

[13] F. H. UMSU, "Cari Lowongan Kerja? Berikut Cara Mudah Daftar Akun Siapkerja Kemnaker 2025," 2025.