

Pemetaan Tema Keluhan Pengguna Aplikasi Cek Bansos Menggunakan K-Means dan TF-IDF Berbasis Ulasan Google Play Store

Krisna Mahendra^{1*}, Zidan Syauqi Rahman², Taqwini³, Hendra Supendar⁴, Riza Fahlapi⁵

^{1,2,3,4,5}Sistem Informasi, Teknologi Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ^{1*}krisnamahendra559@gmail.com, ²syauqizidan629@gmail.com, ³taqwini10@gmail.com,

⁴hendra.hds@bsi.ac.id, ⁵riza.rzf@bsi.ac.id

(*Email Corresponding Author: krisnamahendra559@gmail.com)

Received: 30 Mei 2026 | Revision: 30 Mei 2026 | Accepted: 30 Mei 2026

Abstrak

Transformasi layanan sosial digital menuntut aplikasi publik mudah digunakan masyarakat luas Indonesia. Aplikasi Cek Bansos menjadi kanal penting untuk memeriksa bantuan sosial daring. Namun, ulasan pengguna menunjukkan berbagai keluhan teknis dan administratif yang berulang. Penelitian ini bertujuan mengelompokkan ulasan pengguna berdasarkan kemiripan isi teks untuk menemukan tema keluhan dominan. Data diperoleh melalui scraping ulasan Google Play Store pada 23 Mei 2026. Dari 4.588 ulasan mentah, digunakan 3.623 ulasan keluhan dengan rating 1 sampai 3. Tahapan penelitian meliputi preprocessing teks, pembobotan TF-IDF, pemilihan jumlah klaster, dan penerapan algoritma K-Means. Hasil pengujian menunjukkan jumlah klaster terbaik adalah 6 dengan Silhouette Score sebesar 0,0246 dan Davies-Bouldin Index sebesar 6,0238. Tema keluhan yang ditemukan meliputi kendala akun dan masuk akun, fungsi aplikasi, gangguan teknis, serta status data bantuan. Hasil penelitian ini dapat menjadi masukan awal bagi pengelola aplikasi untuk menentukan prioritas perbaikan layanan digital bantuan sosial.

Kata Kunci: Cek Bansos, Google Play Store, K-Means, Klasterisasi, TF-IDF

Abstract

The digital transformation of social services requires public applications that are easy to use by Indonesian citizens. The Cek Bansos application has become an important channel for checking online social assistance. However, user reviews show recurring technical and administrative complaints. This study aims to cluster user reviews based on textual similarity to identify dominant complaint themes. The data were collected by scraping Google Play Store reviews on May 23, 2026. From 4,588 raw reviews, 3,623 complaint reviews with ratings from 1 to 3 were used. The research stages included text preprocessing, TF-IDF weighting, cluster-number selection, and K-Means clustering. The results showed that the best number of clusters was 6, with a Silhouette Score of 0.0246 and a Davies-Bouldin Index of 6.0238. The complaint themes included account and login issues, application functionality, technical failures, and social assistance data status. These findings can provide an initial reference for application managers in prioritizing improvements to digital social assistance services.

Keywords: Check Social Assistance, Clustering, Google Play Store, K-Means, TF-IDF

1. PENDAHULUAN

Transformasi digital yang berlangsung pada berbagai sektor pelayanan publik telah mendorong pemerintah untuk memanfaatkan teknologi informasi sebagai sarana peningkatan kualitas layanan kepada masyarakat[1]. Pemanfaatan aplikasi berbasis mobile menjadi salah satu bentuk implementasi transformasi tersebut karena mampu memberikan akses layanan yang lebih cepat, mudah, dan efisien[2]. Kehadiran layanan digital memungkinkan masyarakat memperoleh informasi secara mandiri tanpa harus datang langsung ke kantor pelayanan[3]. Selain meningkatkan efisiensi proses administrasi, digitalisasi juga menjadi sarana untuk memperluas jangkauan pelayanan publik kepada masyarakat yang berada di berbagai wilayah[4]. Oleh karena itu, keberhasilan implementasi aplikasi layanan publik sangat dipengaruhi oleh tingkat kemudahan penggunaan, keandalan sistem, serta kemampuan aplikasi dalam memenuhi kebutuhan penggunaannya.

Salah satu bentuk layanan publik yang telah mengalami digitalisasi adalah layanan informasi bantuan sosial melalui aplikasi Cek Bansos. Aplikasi ini dikembangkan untuk membantu masyarakat memperoleh informasi terkait status penerima bantuan sosial, melakukan pengecekan data penerima manfaat, menyampaikan usulan penerima bantuan, serta memberikan sanggahan apabila ditemukan ketidaksesuaian data[5]. Peran aplikasi ini menjadi semakin penting karena bantuan sosial merupakan program pemerintah yang secara langsung berkaitan dengan kesejahteraan masyarakat. Melalui aplikasi tersebut, masyarakat diharapkan dapat memperoleh akses informasi yang lebih transparan dan akurat mengenai program bantuan sosial tanpa harus melalui proses birokrasi yang panjang[6].

Meskipun demikian, tingginya jumlah pengguna aplikasi tidak selalu berbanding lurus dengan kualitas pengalaman pengguna yang diperoleh[7]. Berbagai kendala teknis maupun nonteknis masih sering ditemukan selama penggunaan aplikasi. Pengguna dapat mengalami kesulitan ketika melakukan proses pendaftaran akun, masuk ke dalam sistem, melakukan verifikasi data, maupun saat mengakses informasi bantuan sosial yang dibutuhkan[5]. Selain itu, permasalahan terkait stabilitas aplikasi, kompatibilitas perangkat, serta sinkronisasi data juga dapat memengaruhi tingkat

kepuasan pengguna. Apabila permasalahan tersebut tidak segera diidentifikasi dan ditangani, maka dapat menurunkan kepercayaan masyarakat terhadap layanan digital yang disediakan pemerintah.

Salah satu sumber informasi yang dapat dimanfaatkan untuk memahami pengalaman pengguna adalah ulasan yang tersedia pada Google Play Store[8]. Ulasan pengguna mencerminkan pengalaman nyata yang diperoleh selama menggunakan aplikasi sehingga dapat menjadi sumber masukan yang bernilai bagi pengembang maupun pengelola layanan[9]. Berbeda dengan survei konvensional yang memerlukan waktu dan biaya relatif besar, ulasan pengguna tersedia secara terbuka dan terus bertambah seiring meningkatnya jumlah pengguna aplikasi[10]. Informasi yang terkandung dalam ulasan tidak hanya menunjukkan tingkat kepuasan pengguna, tetapi juga menggambarkan berbagai permasalahan yang dialami selama penggunaan aplikasi. Dengan demikian, ulasan pengguna dapat dimanfaatkan sebagai sumber data untuk mengevaluasi kualitas layanan digital secara berkelanjutan.

Namun, pemanfaatan ulasan pengguna sebagai sumber informasi menghadapi beberapa tantangan. Jumlah ulasan yang sangat besar menyebabkan proses analisis secara manual menjadi tidak efektif. Selain itu, ulasan umumnya ditulis menggunakan bahasa informal, singkatan, kesalahan penulisan, maupun variasi istilah yang berbeda untuk menggambarkan masalah yang sama. Kondisi tersebut menyebabkan proses identifikasi tema keluhan menjadi lebih kompleks. Di sisi lain, informasi yang ditampilkan dalam bentuk rating bintang belum mampu menjelaskan secara rinci jenis permasalahan yang dialami pengguna. Beberapa pengguna dapat memberikan rating rendah karena alasan yang berbeda, seperti kegagalan login, aplikasi yang sering mengalami gangguan, kesalahan data bantuan sosial, maupun kendala teknis lainnya. Oleh karena itu, diperlukan pendekatan analisis yang mampu mengelompokkan ulasan berdasarkan kesamaan isi sehingga tema-tema keluhan yang dominan dapat diidentifikasi secara sistematis.

Penelitian mengenai analisis ulasan pengguna aplikasi berbasis mobile telah berkembang pesat seiring meningkatnya pemanfaatan platform digital dalam berbagai sektor layanan[11]. Sebagian besar penelitian memanfaatkan teknik analisis sentimen untuk mengidentifikasi persepsi pengguna terhadap kualitas aplikasi. Pendekatan tersebut umumnya menggunakan algoritma klasifikasi seperti Naive Bayes, Support Vector Machine, Decision Tree, maupun model berbasis deep learning. Hasil penelitian menunjukkan bahwa analisis sentimen mampu memberikan gambaran umum mengenai tingkat kepuasan pengguna berdasarkan distribusi sentimen positif, negatif, dan netral yang terkandung dalam ulasan aplikasi. Meskipun memberikan informasi yang bermanfaat, pendekatan tersebut masih berfokus pada polaritas opini sehingga belum mampu menjelaskan secara detail jenis permasalahan yang menjadi penyebab munculnya sentimen tertentu.

Dalam konteks aplikasi layanan pemerintah, analisis ulasan pengguna telah digunakan untuk mengevaluasi berbagai aplikasi pelayanan publik yang tersedia pada Google Play Store[12]. Hasil penelitian menunjukkan bahwa ulasan pengguna dapat menjadi sumber informasi yang efektif dalam mengidentifikasi kendala penggunaan aplikasi, kualitas layanan digital, serta tingkat penerimaan masyarakat terhadap inovasi teknologi yang diterapkan pemerintah. Informasi yang diperoleh dari ulasan pengguna mampu membantu pengelola layanan dalam memahami kebutuhan pengguna secara lebih objektif dibandingkan pendekatan survei konvensional yang biasanya memiliki jumlah responden terbatas. Oleh karena itu, pemanfaatan data ulasan pengguna menjadi salah satu alternatif yang relevan dalam proses evaluasi dan pengembangan layanan publik berbasis digital.

Sementara itu, penelitian pada bidang text mining menunjukkan bahwa teknik klusterisasi dapat digunakan untuk menemukan pola tersembunyi pada kumpulan dokumen teks tanpa memerlukan proses pelabelan data[13]. Kombinasi metode TF-IDF dan K-Means banyak digunakan karena mampu mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klusterisasi. Pendekatan ini telah diterapkan pada berbagai studi yang berfokus pada analisis pengalaman pengguna aplikasi digital, forum diskusi daring, maupun media sosial. Melalui proses pengelompokan dokumen berdasarkan kemiripan isi, metode tersebut mampu menghasilkan tema-tema utama yang menjadi perhatian pengguna sehingga memberikan informasi yang lebih rinci dibandingkan sekadar klasifikasi sentimen.

Meskipun berbagai penelitian terdahulu telah berhasil memanfaatkan ulasan pengguna untuk melakukan analisis sentimen maupun evaluasi kualitas aplikasi, sebagian besar penelitian masih berfokus pada pengukuran polaritas opini pengguna[14]. Pendekatan tersebut memang mampu menunjukkan kecenderungan persepsi pengguna terhadap suatu aplikasi, tetapi belum dapat menjelaskan secara rinci bentuk permasalahan yang mendasari munculnya sentimen negatif. Akibatnya, informasi yang dihasilkan sering kali kurang mendukung proses pengambilan keputusan yang membutuhkan identifikasi masalah secara spesifik dan operasional.

Selain itu, penelitian yang secara khusus membahas aplikasi Cek Bansos masih didominasi oleh pendekatan klasifikasi sentimen sehingga informasi mengenai tema-tema keluhan pengguna belum banyak dieksplorasi. Padahal, aplikasi Cek Bansos memiliki karakteristik yang berbeda dibandingkan aplikasi komersial karena berkaitan langsung dengan layanan bantuan sosial yang menyangkut kebutuhan masyarakat. Pengelola aplikasi tidak hanya membutuhkan informasi mengenai tingkat kepuasan pengguna, tetapi juga memerlukan pemetaan permasalahan yang lebih detail untuk menentukan prioritas perbaikan layanan[15]. Oleh karena itu, diperlukan pendekatan yang mampu mengelompokkan ulasan berdasarkan kemiripan konteks permasalahan sehingga dapat diketahui secara lebih jelas apakah keluhan pengguna didominasi oleh kendala autentikasi akun, gangguan teknis aplikasi, kesalahan data penerima bantuan, atau permasalahan lainnya.

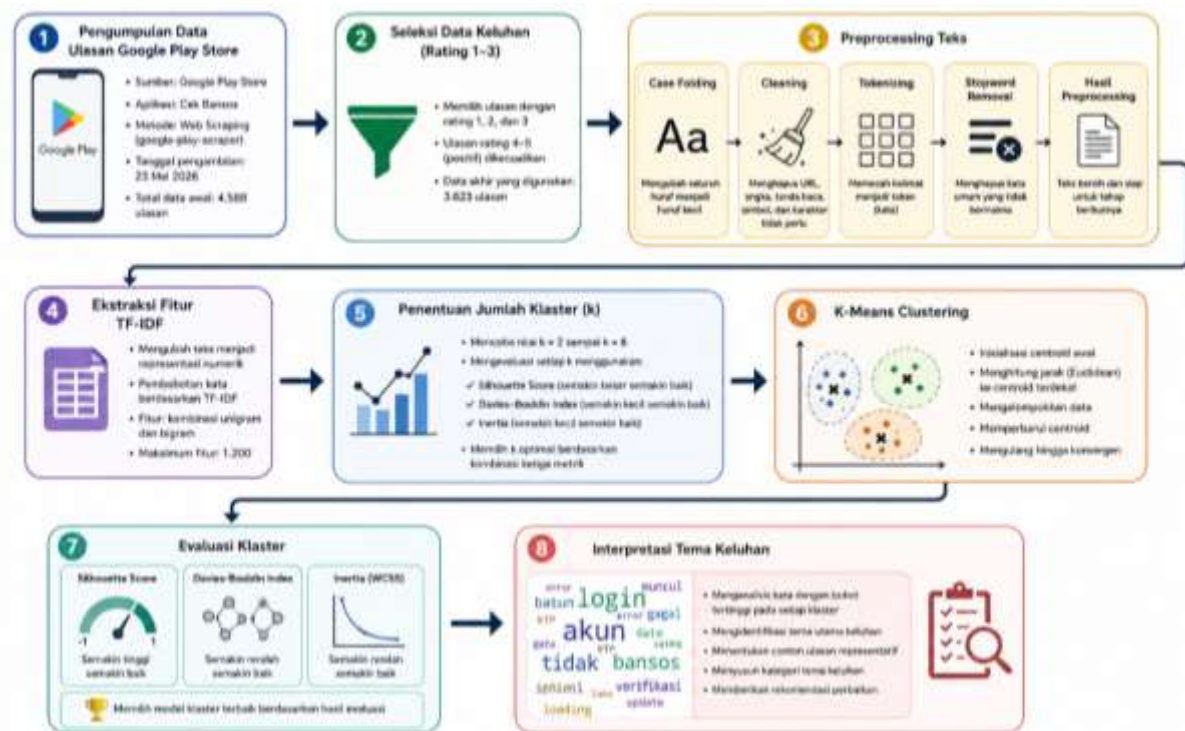
Berdasarkan kondisi tersebut, penelitian ini menawarkan pendekatan pemetaan tema keluhan pengguna aplikasi Cek Bansos melalui kombinasi metode Term Frequency–Inverse Document Frequency (TF-IDF) dan K-Means. Pendekatan ini dipilih karena mampu mengolah data teks dalam jumlah besar serta mengidentifikasi kelompok ulasan

yang memiliki karakteristik serupa. Berbeda dengan penelitian sebelumnya yang berfokus pada klasifikasi sentimen, penelitian ini menitikberatkan pada proses pengelompokan ulasan untuk menemukan tema-tema keluhan yang dominan. Dengan demikian, hasil yang diperoleh tidak hanya menunjukkan adanya ketidakpuasan pengguna, tetapi juga menjelaskan bentuk permasalahan yang menjadi sumber ketidakpuasan tersebut.

Penelitian ini bertujuan untuk mengelompokkan ulasan keluhan pengguna aplikasi Cek Bansos berdasarkan kemiripan isi teks menggunakan metode K-Means dengan representasi fitur TF-IDF. Hasil penelitian diharapkan mampu menghasilkan pemetaan tema keluhan yang lebih komprehensif sehingga dapat menjadi sumber informasi bagi pengelola aplikasi dalam menentukan prioritas perbaikan layanan digital bantuan sosial. Selain memberikan kontribusi praktis bagi pengembangan aplikasi Cek Bansos, penelitian ini juga diharapkan dapat memperkaya kajian text mining khususnya pada analisis ulasan pengguna aplikasi layanan publik berbasis digital..

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan text mining untuk mengidentifikasi tema keluhan pengguna aplikasi Cek Bansos berdasarkan ulasan yang tersedia pada Google Play Store. Metode yang digunakan adalah kombinasi TF-IDF dan K-Means Clustering. Tahapan penelitian meliputi pengumpulan data, seleksi ulasan, preprocessing teks, ekstraksi fitur menggunakan TF-IDF, penentuan jumlah kluster, proses klusterisasi menggunakan K-Means, evaluasi hasil kluster, serta interpretasi tema keluhan. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Tahap pertama penelitian adalah pengumpulan data ulasan pengguna aplikasi Cek Bansos yang tersedia pada Google Play Store. Pengambilan data dilakukan menggunakan pustaka google-play-scraper pada bahasa pemrograman Python. Dataset yang diperoleh terdiri atas berbagai atribut seperti isi ulasan (content), rating (score), tanggal ulasan (at), jumlah pengguna yang memberikan dukungan (thumbsUpCount), dan versi aplikasi (appVersion).

Proses scraping dilakukan pada tanggal 23 Mei 2026 sehingga data yang diperoleh mencerminkan kondisi aplikasi pada periode tersebut. Dari proses pengumpulan data diperoleh sebanyak 4.588 ulasan pengguna yang kemudian digunakan sebagai dataset awal penelitian. Data ulasan dipilih karena merepresentasikan pengalaman nyata pengguna selama menggunakan aplikasi sehingga mampu memberikan gambaran yang lebih objektif mengenai kualitas layanan yang diberikan.

2.2 Seleksi Data Keluhan

Tidak seluruh ulasan digunakan dalam proses analisis. Penelitian ini hanya memfokuskan analisis pada ulasan yang mengandung indikasi keluhan pengguna. Oleh karena itu dilakukan proses seleksi berdasarkan nilai rating yang diberikan pengguna.

Ulasan dengan rating 1 sampai 3 dipilih karena umumnya mencerminkan pengalaman negatif atau ketidakpuasan terhadap aplikasi. Sebaliknya, ulasan dengan rating 4 dan 5 tidak digunakan karena lebih banyak berisi apresiasi atau pengalaman positif. Setelah proses seleksi dilakukan, diperoleh sebanyak 3.623 ulasan yang digunakan sebagai data utama penelitian.

2.3 Preprocessing Teks

Data ulasan yang diperoleh dari Google Play Store masih mengandung berbagai bentuk noise seperti penggunaan huruf kapital yang tidak konsisten, simbol, angka, serta kata-kata yang tidak memiliki makna penting dalam proses analisis. Oleh karena itu dilakukan tahap preprocessing untuk meningkatkan kualitas data teks sebelum diproses lebih lanjut.

Tahap pertama adalah case folding, yaitu mengubah seluruh huruf menjadi huruf kecil agar tidak terjadi perbedaan representasi kata. Tahap berikutnya adalah cleaning, yaitu menghapus URL, angka, tanda baca, simbol, dan karakter yang tidak diperlukan. Selanjutnya dilakukan tokenizing untuk memecah kalimat menjadi kumpulan kata atau token.

Setelah tokenisasi, dilakukan stopword removal untuk menghapus kata-kata umum yang tidak memberikan informasi penting seperti "dan", "yang", "di", dan "ke". Hasil preprocessing berupa kumpulan kata yang lebih bersih dan siap digunakan pada tahap ekstraksi fitur.

2.4 Ekstraksi Fitur Menggunakan TF-IDF

Teks hasil preprocessing tidak dapat langsung diproses oleh algoritma K-Means karena masih berbentuk data tekstual. Oleh karena itu diperlukan proses transformasi menjadi bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF).

Metode TF-IDF memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam suatu dokumen dan tingkat kelangkaannya dalam keseluruhan korpus. Kata yang sering muncul dalam suatu ulasan tetapi jarang ditemukan pada ulasan lain akan memperoleh bobot yang lebih tinggi.

Pada penelitian ini digunakan maksimum 1.200 fitur dengan kombinasi unigram dan bigram untuk menangkap kata tunggal maupun frasa pendek yang sering muncul dalam keluhan pengguna.

2.5 Penentuan Jumlah Kluster

Salah satu parameter penting dalam algoritma K-Means adalah jumlah kluster (k). Penentuan jumlah kluster dilakukan melalui serangkaian eksperimen dengan nilai k antara 2 sampai 6.

Setiap nilai k dievaluasi menggunakan Silhouette Score, Davies-Bouldin Index, dan Inertia. Kombinasi ketiga metrik tersebut digunakan untuk memperoleh jumlah kluster yang mampu menghasilkan pemisahan kelompok data secara optimal sekaligus tetap mudah diinterpretasikan.

2.6 Klusterisasi Menggunakan K-Means

Tahap selanjutnya adalah proses pengelompokan data menggunakan algoritma K-Means. Algoritma ini bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan kedekatan jarak antara data dan centroid kluster.

Proses dimulai dengan inisialisasi centroid awal, dilanjutkan dengan pengelompokan data ke centroid terdekat menggunakan jarak Euclidean. Setelah seluruh data memperoleh kluster, centroid diperbarui berdasarkan rata-rata anggota kluster. Proses tersebut diulang hingga tidak terjadi perubahan signifikan pada posisi centroid atau mencapai kondisi konvergen.

$$J = \sum_{i=1}^k \sum_{j=1}^k \|x_i - c_j\|^2 \quad (1)$$

Melalui proses ini, ulasan yang memiliki kemiripan isi akan berada pada kluster yang sama sehingga tema-tema keluhan dapat diidentifikasi dengan lebih mudah.

2.7 Evaluasi Klaster

Kualitas hasil klasterisasi dievaluasi menggunakan tiga metrik yaitu Silhouette Score, Davies-Bouldin Index, dan Inertia. Silhouette Score digunakan untuk mengukur tingkat kedekatan suatu data terhadap klaster tempat data tersebut berada dibandingkan dengan klaster lainnya.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Semakin tinggi nilai Silhouette Score maka semakin baik kualitas pemisahan klaster yang dihasilkan. Sementara itu, Davies-Bouldin Index digunakan untuk mengukur rasio kedekatan antar anggota klaster dan jarak antarklaster. Nilai Davies-Bouldin Index yang lebih kecil menunjukkan kualitas klaster yang lebih baik.

Evaluasi dilakukan terhadap seluruh konfigurasi jumlah klaster yang diuji sehingga diperoleh nilai k yang paling sesuai untuk karakteristik dataset ulasan pengguna aplikasi Cek Bansos.

2.8 Interpretasi Tema Keluhan

Tahap terakhir adalah interpretasi hasil klasterisasi. Pada tahap ini dilakukan analisis terhadap kata-kata yang memiliki bobot tertinggi pada setiap klaster untuk mengidentifikasi tema utama yang dibahas pengguna. Interpretasi dilakukan dengan mengamati kata kunci dominan, jumlah anggota klaster, serta contoh ulasan yang mewakili masing-masing kelompok. Hasil interpretasi kemudian digunakan untuk menyusun kategori tema keluhan seperti kendala akun dan autentikasi, gangguan teknis aplikasi, masalah data bantuan sosial, maupun kendala operasional lainnya. Tahap ini menjadi bagian penting karena menghasilkan informasi yang dapat dimanfaatkan oleh pengelola aplikasi dalam menentukan prioritas perbaikan layanan digital secara lebih terarah dan berbasis data.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Proses scraping ulasan pengguna aplikasi Cek Bansos melalui Google Play Store menghasilkan sebanyak 4.588 ulasan mentah. Selanjutnya dilakukan proses pembersihan data untuk menghapus duplikasi dan ulasan kosong sehingga diperoleh dataset yang lebih representatif. Penelitian ini hanya memfokuskan analisis pada ulasan yang mengandung indikasi keluhan pengguna, yaitu ulasan dengan rating 1 sampai 3. Setelah proses seleksi dilakukan, diperoleh sebanyak 3.623 ulasan yang digunakan sebagai dataset utama dalam penelitian.

Pemilihan ulasan dengan rating rendah dilakukan karena tujuan penelitian adalah mengidentifikasi tema-tema keluhan yang dialami pengguna aplikasi. Dengan demikian, data yang dianalisis lebih relevan untuk menggambarkan berbagai permasalahan yang muncul selama penggunaan aplikasi Cek Bansos. Dataset yang diperoleh selanjutnya diproses menggunakan tahapan preprocessing dan ekstraksi fitur sebelum dilakukan klasterisasi.

Tabel 1. Ringkasan Dataset Penelitian

Keterangan	Jumlah
Total ulasan hasil scraping	4.588
Ulasan setelah cleaning	4.321
Ulasan rating 1–3	3.623
Jumlah fitur TF-IDF	1.200
Rentang nilai k	2–6

3.2 Hasil Preprocessing Teks

Tahap preprocessing dilakukan untuk meningkatkan kualitas data teks sebelum dilakukan proses klasterisasi. Tahapan yang diterapkan meliputi case folding, cleaning, tokenisasi, dan stopword removal. Proses ini bertujuan untuk menghilangkan berbagai bentuk noise yang dapat memengaruhi kualitas hasil klasterisasi.

Sebagai contoh, ulasan pengguna seperti:

"Saya GAK BISA LOGIN!!! Padahal email sudah benar 123"

setelah melalui preprocessing berubah menjadi:

"gak bisa login email benar"

Hasil preprocessing menunjukkan bahwa data menjadi lebih terstruktur dan memiliki representasi yang lebih konsisten. Kondisi ini membantu proses pembobotan TF-IDF dalam menghasilkan fitur yang lebih relevan terhadap tema keluhan pengguna.

Tabel 2. Contoh Hasil Preprocessing

Tahap	Hasil
Data asli	Saya GAK BISA LOGIN!!! Padahal email sudah benar 123
Case Folding	saya gak bisa login padahal email sudah benar 123
Cleaning	saya gak bisa login padahal email sudah benar
Tokenizing	saya, gak, bisa, login, padahal, email, sudah, benar
Stopword Removal	gak, login, email, benar

3.3 Hasil Evaluasi Jumlah Kluster

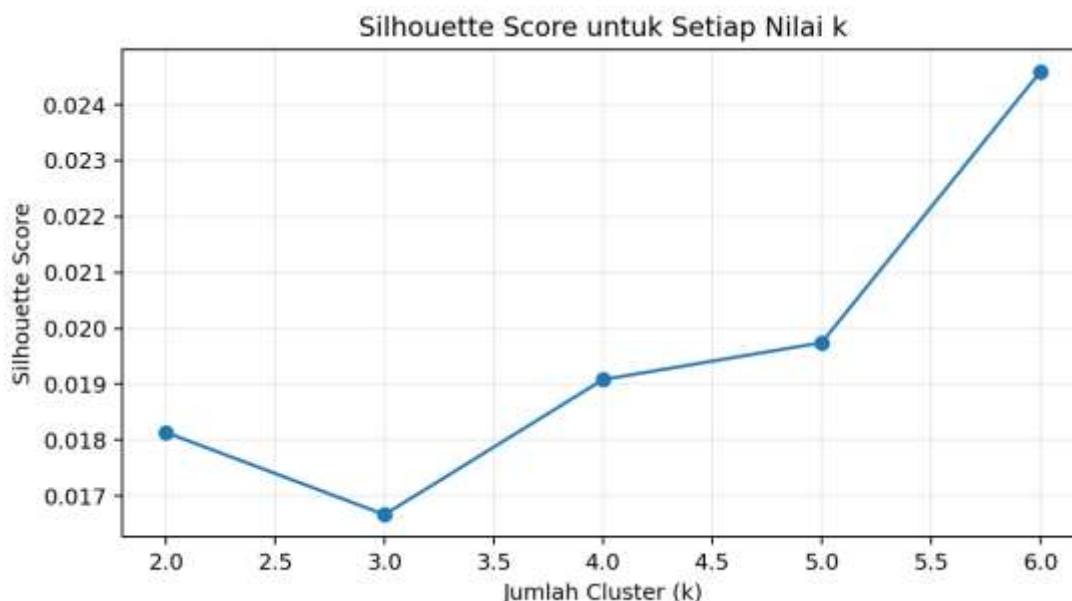
Penentuan jumlah kluster dilakukan dengan menguji beberapa nilai k pada rentang 2 sampai 6. Evaluasi dilakukan menggunakan Silhouette Score, Davies-Bouldin Index, dan Inertia untuk memperoleh konfigurasi kluster yang paling optimal.

Hasil pengujian menunjukkan bahwa nilai k = 6 menghasilkan Silhouette Score tertinggi yaitu 0,0246. Meskipun nilainya relatif rendah, konfigurasi tersebut tetap dipilih karena menghasilkan pemisahan tema yang lebih informatif dibandingkan jumlah kluster lainnya. Selain itu, nilai Davies-Bouldin Index sebesar 6,0238 menunjukkan kualitas pemisahan kluster yang masih dapat diterima untuk karakteristik data ulasan pendek yang memiliki kosakata serupa.

Tabel 3. Evaluasi Jumlah Kluster

k	Silhouette Score	Davies-Bouldin Index	Inertia
2	0,0181	3,9004	3429,19
3	0,0167	7,8401	3400,09
4	0,0191	5,9739	3379,58
5	0,0197	6,6221	3356,12
6	0,0246	6,0238	3321,67

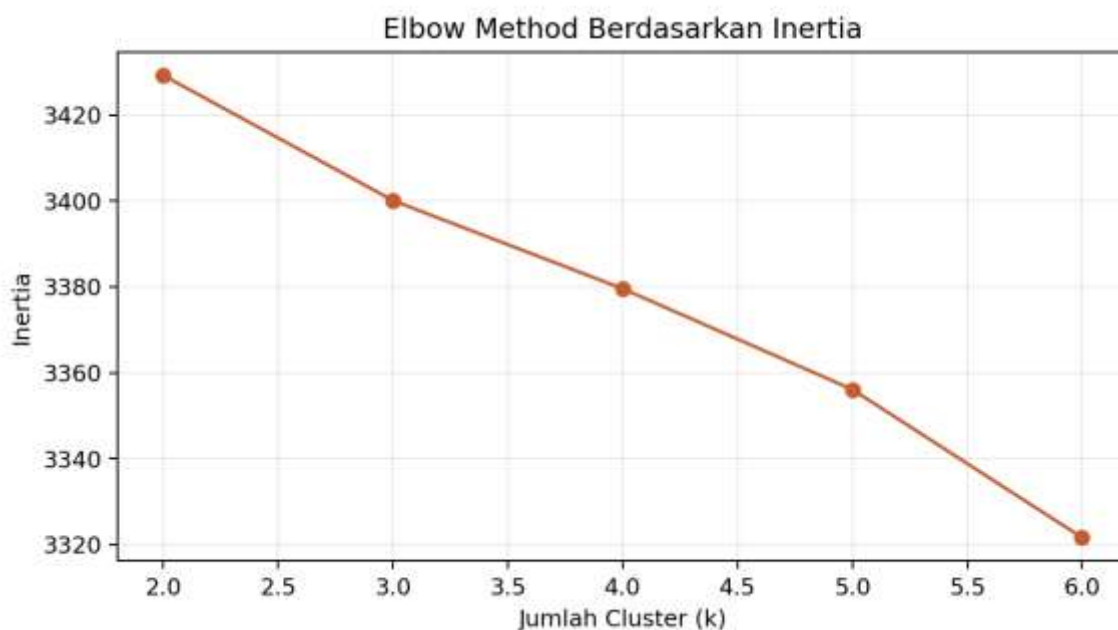
Berdasarkan hasil pengujian pada Tabel 3, dilakukan visualisasi Silhouette Score untuk melihat kecenderungan perubahan kualitas kluster terhadap penambahan jumlah kluster. Grafik pada Gambar 2 menunjukkan bahwa nilai Silhouette Score tertinggi diperoleh ketika jumlah kluster sebanyak enam kelompok.



Gambar 2. Grafik Silhouette Score untuk setiap nilai k

Grafik menunjukkan bahwa peningkatan jumlah klaster tidak selalu menghasilkan kualitas klaster yang lebih baik. Namun pada penelitian ini, nilai $k=6$ memberikan hasil yang paling optimal dibandingkan konfigurasi lainnya sehingga dipilih sebagai jumlah klaster terbaik.

Selain menggunakan Silhouette Score, penelitian ini juga memanfaatkan metode Elbow melalui pengamatan nilai inerti. Grafik Elbow digunakan untuk melihat titik penurunan inerti yang mulai melambat sehingga dapat menjadi indikasi jumlah klaster yang sesuai.



Gambar 3. Grafik Elbow Method berdasarkan inerti

Terlihat bahwa nilai inerti terus menurun seiring bertambahnya jumlah klaster. Namun penurunan tersebut mulai melandai pada rentang nilai k yang lebih tinggi. Oleh karena itu, pemilihan $k=6$ dianggap mampu memberikan keseimbangan antara kualitas klaster dan interpretasi hasil.

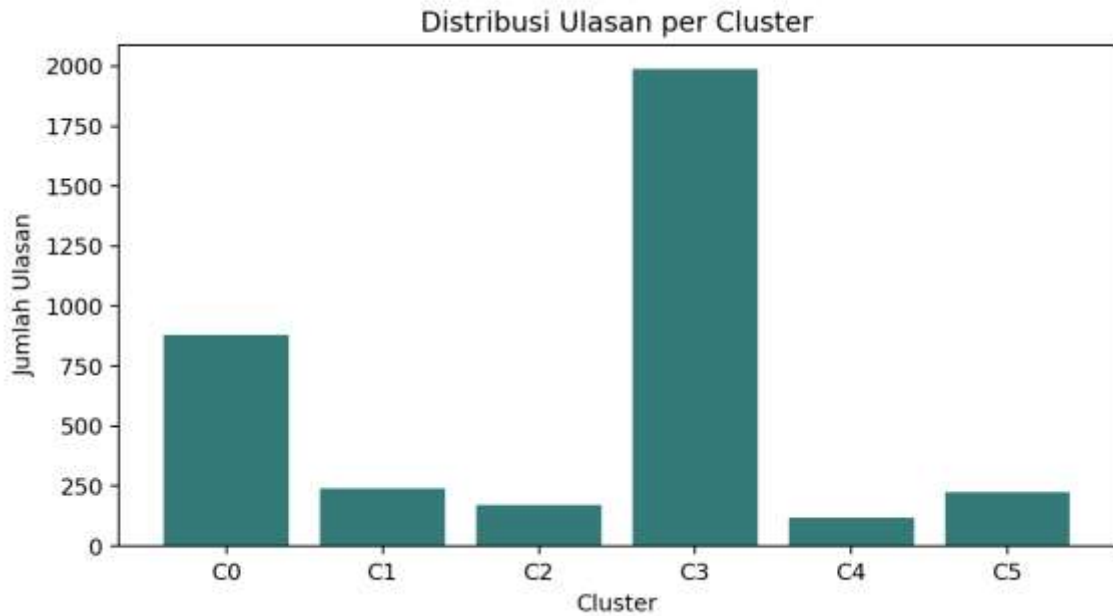
3.4 Hasil Klasterisasi Ulasan Pengguna

Setelah jumlah klaster optimal diperoleh, proses klasterisasi dilakukan menggunakan algoritma K-Means. Hasil klasterisasi menghasilkan enam kelompok utama yang merepresentasikan berbagai jenis keluhan pengguna aplikasi Cek Bansos.

Tabel 4. Hasil Interpretasi Klaster

Klaster	Jumlah	Rating	Tema
C0	880	1,28	Kendala akun dan login
C1	239	1,13	Fungsi aplikasi
C2	171	1,30	Gangguan teknis
C3	1987	1,27	Status bantuan dan data penerima
C4	120	1,32	Tidak dapat dibuka
C5	226	1,42	Susah dibuka

Distribusi jumlah ulasan pada setiap klaster ditampilkan pada Gambar 4 untuk menunjukkan proporsi tema keluhan yang ditemukan pada penelitian ini.

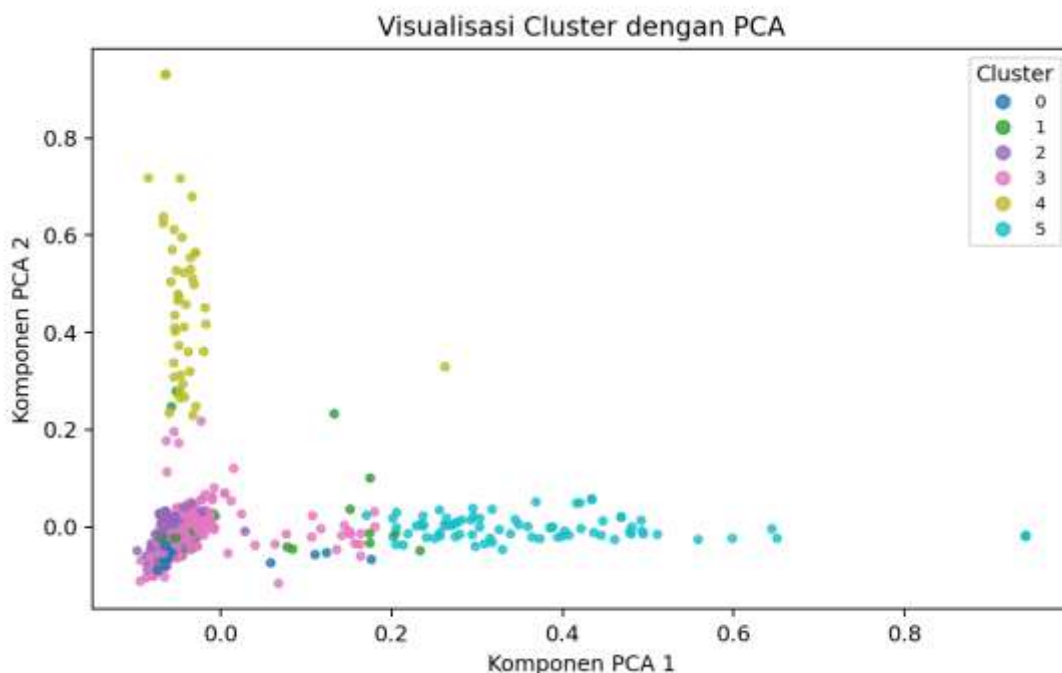


Gambar 4. Distribusi Ulasan pada Setiap Klaster

Hasil menunjukkan bahwa klaster C3 memiliki jumlah anggota terbesar dengan total 1.987 ulasan. Temuan ini mengindikasikan bahwa sebagian besar pengguna mengeluhkan permasalahan yang berkaitan dengan status bantuan sosial dan data penerima bantuan. Sementara itu, klaster akun dan login juga memiliki jumlah anggota yang cukup besar sehingga menunjukkan bahwa proses autentikasi masih menjadi salah satu sumber utama ketidakpuasan pengguna.

3.5 Visualisasi Klaster Menggunakan PCA

Untuk melihat penyebaran data hasil klasterisasi secara visual, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA). Visualisasi ini bertujuan untuk menggambarkan pola pemisahan antar klaster dalam ruang dua dimensi.



Gambar 5. Visualisasi Klaster Menggunakan PCA

Berdasarkan Gambar 5 terlihat bahwa beberapa klaster masih memiliki area yang saling beririsan. Kondisi tersebut sejalan dengan nilai Silhouette Score yang relatif rendah. Namun demikian, pemisahan kelompok masih mampu

menghasilkan tema-tema yang dapat diinterpretasikan dengan baik sehingga tetap memberikan informasi yang bermanfaat bagi pengelola aplikasi.

3.6 Pembahasan

Hasil penelitian menunjukkan bahwa permasalahan utama pengguna aplikasi Cek Bansos tidak hanya berasal dari aspek teknis aplikasi, tetapi juga berkaitan dengan akses akun dan validitas data bantuan sosial. Temuan ini menunjukkan bahwa pengalaman pengguna pada aplikasi layanan publik dipengaruhi oleh kombinasi antara kualitas sistem informasi dan kualitas data yang dikelola. Berbeda dengan aplikasi komersial yang umumnya berfokus pada performa sistem, aplikasi layanan sosial harus mampu menjamin keakuratan informasi yang ditampilkan karena berkaitan langsung dengan kebutuhan masyarakat.

Temuan penelitian ini memperluas hasil penelitian sebelumnya yang berfokus pada analisis sentimen ulasan aplikasi Cek Bansos. Jika penelitian terdahulu hanya menunjukkan dominasi sentimen negatif, penelitian ini berhasil mengidentifikasi bentuk permasalahan yang menjadi sumber sentimen tersebut. Dengan demikian, hasil klusterisasi memberikan informasi yang lebih operasional untuk mendukung proses pengambilan keputusan.

Dari sisi praktis, hasil penelitian dapat menjadi dasar dalam menentukan prioritas pengembangan aplikasi. Kluster akun dan login menunjukkan perlunya perbaikan sistem autentikasi dan pemulihan akun. Kluster gangguan teknis mengindikasikan kebutuhan peningkatan stabilitas aplikasi dan kompatibilitas perangkat. Sementara itu, kluster status bantuan dan data penerima menunjukkan perlunya peningkatan sinkronisasi data serta transparansi informasi kepada masyarakat. Oleh karena itu, penerapan text mining terhadap ulasan pengguna terbukti mampu menghasilkan informasi yang berguna dalam mendukung evaluasi dan peningkatan kualitas layanan digital pemerintah.

4. KESIMPULAN

Penelitian ini berhasil menerapkan metode K-Means Clustering dengan pembobotan TF-IDF untuk memetakan tema keluhan pengguna aplikasi Cek Bansos berdasarkan 3.623 ulasan berating 1 sampai 3 yang diperoleh dari Google Play Store. Hasil pengujian menunjukkan bahwa jumlah kluster terbaik adalah enam kluster dengan nilai Silhouette Score sebesar 0,0246 dan Davies-Bouldin Index sebesar 6,0238. Kluster yang dihasilkan mampu mengidentifikasi beberapa tema keluhan utama, yaitu kendala akun dan login, masalah fungsi aplikasi, gangguan teknis, aplikasi yang sulit atau tidak dapat dibuka, serta permasalahan status bantuan dan data penerima yang menjadi tema keluhan paling dominan. Temuan ini menunjukkan bahwa pendekatan klusterisasi mampu memberikan informasi yang lebih spesifik dibandingkan analisis sentimen karena tidak hanya mengidentifikasi adanya ketidakpuasan pengguna, tetapi juga mengungkap bentuk permasalahan yang mendasarinya. Hasil penelitian dapat dimanfaatkan sebagai dasar dalam menentukan prioritas pengembangan dan perbaikan layanan digital bantuan sosial, khususnya pada aspek autentikasi pengguna, stabilitas aplikasi, serta peningkatan kualitas dan sinkronisasi data bantuan sosial. Untuk penelitian selanjutnya, disarankan menggunakan teknik normalisasi bahasa informal, stemming, word embedding, maupun metode klusterisasi yang lebih adaptif seperti DBSCAN, HDBSCAN, atau pendekatan berbasis transformer guna meningkatkan kualitas pemisahan kluster dan pemahaman konteks semantik ulasan pengguna.

REFERENCES

- [1] M. Firdaus, H. Siregar, A. Wahyudin, Y. A. Hambali, and Rasim, "Perancangan Website Servis Laptop dan Komputer Sebagai Wujud Penerapan Technopreneurship Mahasiswa," *J. Komput. Teknol. Inf. Sist. Inf.*, vol. 4, no. 2, pp. 556–565, 2025, doi: 10.62712/juktisi.v4i2.453.
- [2] N. Alfritri and S. Alfiyah, "Digitalisasi Layanan Keuangan Di BMT UGT Nusantara Capem Batu Urip Melalui Pengembangan Aplikasi Mobile," *J. Penelit. Nusant.*, vol. 1, no. 3, pp. 767–771, 2025.
- [3] L. N. E. Rahma, "Inovasi Layanan Digital Pak Dalman Sebagai Upaya Transformasi Hukum Pelayanan Publik Untuk Mewujudkan Efisiensi Dan ...," *J. Stud. Leg.*, vol. 5, no. 02, 2024, [Online]. Available: <https://studialegalia.ub.ac.id/index.php/studialegalia/article/view/110%0Ahttps://studialegalia.ub.ac.id/index.php/studialegalia/article/download/110/86>
- [4] L. Choirunnisa, T. H. C. Oktaviana, A. A. Ridlo, and E. I. Rohmah, "Peran Sistem Pemerintah Berbasis Elektronik (SPBE) Dalam Meningkatkan Aksesibilitas Pelayanan Publik di Indonesia," *Sosio Yust. J. Huk. dan Perubahan Sos.*, vol. 3, no. 1, pp. 71–95, 2023, doi: 10.15642/sosyus.v3i1.401.
- [5] M. P. Hannah, R. P. Wardani, N. R. Damayanti, and . M., "Rancang Bangun Sistem Informasi Pengelolaan Data Bantuan Sosial Tepat Sasaran Di Kabupaten Ogan Ilir Berbasis Web," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 4, pp. 3221–3232, 2025, doi: 10.29100/jipi.v10i4.6596.
- [6] D. Purbaningrum and Hendri Hermawan Adinugraha, "Transformasi Kebijakan Publik Menyiasati Ketidaktepatan Sasaran Bantuan Sosial Di Desa Pringsurat Untuk Kesejahteraan Masyarakat Yang Lebih Baik," *J. Pembang. dan Kebijak. Publik*, vol. 15, no. 2, pp. 31–44, 2024, doi: 10.36624/jpkp.v15i2.116.
- [7] K. Karina and D. Pibriana, "Penggunaan Metode User Experience Questionnaire Untuk Menganalisis Kualitas Pengalaman Pengguna Aplikasi Myindihome Mobile," *METHOMIKA J. Manaj. Inform. dan Komputerisasi*

- Akunt.*, vol. 7, no. 1, pp. 10–19, 2023, doi: 10.46880/jmika.vol7no1.pp10-19.
- [8] D. Nurwahidah, G. Dwilestari, N. Dienwati Nuris, and R. Narasati, “Analisis Sentimen Data Ulasan Pengguna Aplikasi Google Kelas Pada Google Play Store Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3673–3678, 2024, doi: 10.36040/jati.v7i6.8245.
- [9] S. D. Suryaningsih, A. Andrialdo, D. Hartati, and U. Kalsum, “Pengalaman Pengguna dan Praktik Pengelolaan Keuangan Digital Pada Aplikasi Seabank: Studi Kualitatif Terhadap Ulasan di Google Play Store,” *J. Penelit. Ilm. Multidisipliner*, vol. 2, no. 04, 2026.
- [10] M. A. Yaqin *et al.*, “Survei Ekonomi Berbasis Aplikasi Mobile sebagai Alat Penilaian Tingkat Perekonomian Penduduk,” in *Prosiding Seminar Teknologi Dan Sains*, 2025, pp. 705–712.
- [11] M. Alda, A. A. Syahfitri, A. P. Aji, and F. Abdul Aziz, “Analisis Sistematis Tren Pengembangan Aplikasi Mobile di Indonesia pada Periode Tahun 2024-2025,” *J. Manaj. Inform. Sist. Inf. dan Teknol. Komput.*, vol. 4, no. 2, pp. 632–641, 2025, doi: 10.70247/jumistik.v4i2.223.
- [12] A. Farhan, I. Istiadi, and A. Yuniar Rahman, “Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital Di Google Play Store Dengan Bert,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 3776–3783, 2025, doi: 10.36040/jati.v9i3.13358.
- [13] N. W. Utami and I. G. J. Eka Putra, “Text Minig Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine Similarity,” *J. Inform. Teknol. dan Sains*, vol. 4, no. 3, pp. 255–259, 2022, doi: 10.51401/jinteks.v4i3.1907.
- [14] M. Ridwan Usmana and A. Rais Ruli, “Analisis Sentimen Ulasan Pengguna Terhadap Kualitas Layanan Transportasi Online Pada Aplikasi Kai Access Menggunakan Metode Deep Learning Dengan Algoritma Lstm, Gru, Dan Cnn,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 5, pp. 7546–7555, 2025, doi: 10.36040/jati.v9i5.14744.
- [15] Tofan Tri Nugroho, “Optimalisasi Pelayanan Pelanggan melalui Pemanfaatan Sistem Informasi Manajemen,” *J. Actual Organ. Econ.*, vol. 6, no. 02, pp. 135–141, 2025, doi: 10.65212/jago-e.v6i02.173.