

ANALISIS SENTIMEN BERBASIS ASPEK UNTUK EVALUASI APLIKASI PEMERINTAH INDONESIA MENGGUNAKAN DATASET IGAR

ASPECT-BASED SENTIMENT ANALYSIS FOR EVALUATING INDONESIAN GOVERNMENT APPLICATIONS USING THE IGAR DATASET

Kadek Aryana Dwi Putra^{1*}, Fauzan Hidayatullah², Ni Made Ayu Martiani³

*E-mail: aryanadwiputra@unud.ac.id

¹Perpustakaan dan Sains Informasi, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Udayana

²Pendidikan Luar Sekolah, Fakultas Ilmu Pendidikan, Universitas Negeri Makassar

³Magister Ilmu Manajemen, Fakultas Pascasarjana, Universitas Pendidikan Ganesha

Abstrak

Transformasi layanan publik berbasis aplikasi menuntut pemerintah menyediakan kanal digital yang andal, mudah digunakan, dan mampu menyajikan informasi yang dipercaya pengguna. Namun, evaluasi aplikasi layanan publik sering berhenti pada skor rating atau klasifikasi sentimen umum, sehingga belum menunjukkan aspek layanan yang menjadi sumber kepuasan maupun keluhan. Penelitian ini bertujuan mengevaluasi enam aplikasi pemerintah dan layanan publik Indonesia menggunakan IGAR (Indonesian Government App Review Dataset), yaitu JMO, SATUSEHAT, Mobile JKN, MyPertamina, KAI, dan BMKG. Setelah menghapus 26 ulasan tanpa konten, sebanyak 617.696 ulasan dianalisis. Label sentimen menggunakan label berbasis rating, sedangkan tujuh aspek layanan diekstraksi melalui dictionary-based matching. Kebaruan penelitian terletak pada pemetaan komparatif lintas aplikasi untuk mengidentifikasi aspek layanan yang secara konsisten berasosiasi dengan ketidakpuasan pengguna. Validasi manual terhadap 200 ulasan yang terdeteksi mengandung aspek menghasilkan presisi indikatif sebesar 91,50%; desain validasi tersebut belum mengestimasi recall dan F1-score. Hasil menunjukkan bahwa sentimen positif mendominasi sebesar 54,47%, diikuti sentimen negatif 39,97% dan netral 5,56%. Pada tingkat aspek, performa dan stabilitas sistem serta akses akun dan autentikasi memiliki proporsi negatif tertinggi, masing-masing 82,49% dan 81,69%. Temuan ini menunjukkan bahwa pengembangan aplikasi layanan publik perlu memprioritaskan stabilitas sistem, autentikasi, kualitas data, dan kejelasan alur layanan. Secara ilmiah, pendekatan ini menyediakan kerangka evaluasi yang transparan, interpretatif, dan dapat direplikasi untuk menghubungkan ulasan pengguna dengan prioritas perbaikan layanan digital publik.

Kata kunci: aplikasi layanan publik, analisis sentimen berbasis aspek, ulasan pengguna, e-government, IGAR

Abstract

Application-based public services require governments to provide digital channels that are reliable, usable, and capable of delivering information that users can trust. However, evaluations of public-service applications often stop at rating scores or general sentiment classification, leaving the specific service aspects behind satisfaction and complaints unidentified. This study evaluates six Indonesian government-related and public-service applications using the IGAR (Indonesian Government App Review Dataset): JMO, SATUSEHAT, Mobile JKN, MyPertamina, KAI, and BMKG. After removing 26 reviews without textual content, 617,696 reviews were analyzed. Sentiment labels were derived from user ratings, while seven service aspects were extracted using dictionary-based matching. The novelty of the study lies in cross-application comparative mapping to identify service aspects consistently associated with user dissatisfaction. Manual validation of 200 reviews detected as containing service aspects produced an indicative precision of 91.50%; this validation design did not estimate recall or F1-score. The results show

that positive sentiment dominates at 54.47%, followed by negative sentiment at 39.97% and neutral sentiment at 5.56%. At the aspect level, system performance and stability, together with account access and authentication, recorded the highest negative proportions, at 82.49% and 81.69%, respectively. These findings indicate that public-service application development should prioritize system stability, authentication, data quality, and clearer service flows. Scientifically, this approach provides a transparent, interpretable, and replicable evaluation framework that links user-generated reviews with priorities for improving public digital services.

Keywords: public-service applications, aspect-based sentiment analysis, user reviews, e-government, IGAR

Submitted : 3 Juni 2026 Revised : 16 Juli 2025 Accepted : 21 Juli 2026

1. PENDAHULUAN

Transformasi digital layanan publik telah mendorong pemerintah menggunakan aplikasi sebagai salah satu kanal utama untuk menyediakan informasi, transaksi, dan layanan administratif kepada masyarakat. Dalam konteks tersebut, aplikasi pemerintah tidak hanya berfungsi sebagai perangkat teknis, tetapi juga sebagai antarmuka layanan yang menghubungkan kebutuhan pengguna dengan data, prosedur, dan keputusan pemerintah. Keandalan aplikasi menjadi penting karena gangguan pada proses autentikasi, akses informasi, sinkronisasi data, dan transaksi dapat secara langsung menghambat masyarakat dalam memperoleh layanan yang dibutuhkan. Ulasan pengguna pada Google Play dapat dimanfaatkan sebagai sumber evaluasi karena merekam pengalaman aktual ketika pengguna mengakses, memverifikasi, mencari, dan memproses layanan digital pemerintah.

Dataset IGAR menyediakan korpus ulasan terhadap enam aplikasi pemerintah dan layanan publik Indonesia yang dapat digunakan untuk mengkaji persepsi pengguna terhadap layanan digital, yaitu JMO, SATUSEHAT, Mobile JKN, MyPertamina, KAI, dan BMKG [1], [2]. Ulasan pengguna memiliki nilai analitis karena memuat penilaian terhadap berbagai atribut layanan, seperti kemudahan penggunaan, keandalan sistem, kualitas informasi, transaksi, dan respons layanan [3]. Namun, evaluasi aplikasi pemerintah masih sering bergantung pada skor rating atau klasifikasi sentimen secara umum. Pendekatan tersebut dapat menunjukkan kecenderungan persepsi positif, negatif, atau netral, tetapi belum cukup menjelaskan aspek layanan yang menjadi sumber kepuasan maupun keluhan pengguna.

Analisis sentimen berbasis aspek memungkinkan opini pengguna dihubungkan dengan dimensi tertentu dari produk atau layanan [4], [5]. Pendekatan ini memberikan informasi yang lebih terperinci dibandingkan klasifikasi sentimen pada tingkat dokumen karena dapat menunjukkan aspek yang dibicarakan pengguna beserta kecenderungan sentimennya. Berbagai penelitian menunjukkan bahwa analisis sentimen dapat digunakan untuk memahami persepsi pengguna terhadap aplikasi dan layanan digital, meskipun masih menghadapi tantangan berupa variasi bahasa, ekspresi tidak baku, konteks implisit, sarkasme, dan ketidakseimbangan kelas [6], [7]. Dalam konteks *e-government*, penelitian terdahulu telah menganalisis ulasan aplikasi pemerintah menggunakan *machine learning*, *topic modeling*, *convolutional neural network*, *Word2Vec*, dan pendekatan pembelajaran mendalam lainnya [8], [9], [10]. Meskipun demikian, sebagian penelitian lebih menekankan kinerja model klasifikasi atau identifikasi topik secara umum, sedangkan pemetaan komparatif aspek layanan pada beberapa aplikasi pemerintah Indonesia masih terbatas.

Penelitian ini berbeda dari publikasi IGAR yang berfokus pada penyediaan dataset, deskripsi karakteristik data, distribusi sentimen berbasis rating dan VADER, serta pengujian model klasifikasi sebagai *baseline* [1], [2]. Artikel ini menggunakan IGAR sebagai data sekunder untuk mengevaluasi aplikasi pemerintah Indonesia melalui pemetaan tujuh aspek layanan, yaitu akses akun dan autentikasi, performa dan stabilitas sistem, kualitas data dan informasi, fitur dan transaksi layanan, navigasi dan kemudahan penggunaan, layanan bantuan dan respons, serta pembaruan dan instalasi aplikasi. Kebaruan penelitian tidak terletak pada penciptaan algoritma

ekstraksi aspek baru, tetapi pada penerapan analisis lintas aplikasi untuk mengidentifikasi masalah yang bersifat umum maupun spesifik pada layanan pemerintah dengan fungsi yang berbeda. Pendekatan tersebut relevan dengan kajian keberhasilan sistem informasi dan kualitas layanan mobile government yang menempatkan kualitas sistem, kualitas informasi, kegunaan, aksesibilitas, konsistensi, privasi, dan keamanan sebagai faktor penting dalam pengalaman pengguna [11], [12].

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan mengevaluasi aplikasi pemerintah Indonesia menggunakan analisis sentimen berbasis aspek pada ulasan pengguna dalam dataset IGAR. Secara khusus, penelitian ini menganalisis distribusi sentimen pada setiap aplikasi, mengidentifikasi aspek layanan yang paling sering muncul, menghitung proporsi sentimen negatif pada setiap aspek, serta membandingkan pola keluhan antar-aplikasi. Kontribusi penelitian ini adalah menyediakan pemetaan aspek keluhan yang transparan dan mudah diinterpretasikan sebagai dasar penentuan prioritas perbaikan aplikasi pemerintah. Hasil penelitian diharapkan dapat membantu pengelola layanan digital mengarahkan sumber daya pengembangan pada masalah yang paling memengaruhi akses dan pengalaman pengguna, bukan hanya pada penambahan fitur atau perubahan tampilan aplikasi.

2. METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif deskriptif berbasis analisis teks. Alur penelitian terdiri atas pemilihan *dataset*, pembersihan data, penggunaan label sentimen berbasis *rating*, ekstraksi aspek layanan menggunakan kamus kata kunci, perhitungan distribusi sentimen dan aspek, serta interpretasi hasil dalam konteks evaluasi aplikasi pemerintah. Alur penelitian ditampilkan pada Gambar 1.



Gambar 1. Alur penelitian

2.1 Sumber Data

Data penelitian berasal dari IGAR: Indonesian Government App Review Dataset versi 3 yang tersedia sebagai *dataset* terbuka [2]. Dataset memuat ulasan Google Play terhadap enam aplikasi pemerintah dan layanan publik Indonesia, yaitu JMO, SATUSEHAT, Mobile JKN, MyPertamina, KAI, dan BMKG, dengan rentang waktu ulasan 11 Mei 2012 sampai 26 Maret 2023. IGAR menyediakan dua file utama, yaitu *Rating_labeled.csv* dan *VADER_labeled.csv*. Penelitian ini menggunakan *Rating_labeled.csv* karena label sentimennya diturunkan dari rating yang diberikan pengguna dan tidak bergantung pada hasil terjemahan otomatis. File tersebut memuat nama aplikasi, teks ulasan, terjemahan bahasa Inggris, skor rating, waktu ulasan, versi aplikasi, dan label sentimen. Analisis aspek hanya menggunakan teks asli berbahasa Indonesia pada kolom *content*. Dari 617.722 baris awal, 26 ulasan tanpa konten dihapus karena tidak dapat diproses dalam ekstraksi aspek, sehingga jumlah data yang dianalisis menjadi 617.696 ulasan. Nama aplikasi juga diseragamkan, antara lain *satusehat* menjadi SATUSEHAT, *mobileJKN* menjadi Mobile JKN, dan *pertamina* menjadi MyPertamina.

Tabel 1. Komposisi data setelah pembersihan

Aplikasi	Jumlah ulasan	Proporsi (%)
JMO	200.000	32,38
SATUSEHAT	190.693	30,87
Mobile JKN	122.102	19,77

MyPertamina	96.288	15,59
KAI	7.300	1,18
BMKG	1.313	0,21
Total	617.696	100,00

Komposisi pada Tabel 1 memperlihatkan ketimpangan jumlah ulasan antar-aplikasi. JMO, SATUSEHAT, Mobile JKN, dan MyPertamina mendominasi korpus, sedangkan KAI dan BMKG memiliki jumlah ulasan yang jauh lebih kecil. Oleh karena itu, perbandingan antar-aplikasi pada penelitian ini lebih ditekankan pada proporsi sentimen dan aspek, bukan hanya jumlah absolut ulasan.

2.2 Label Sentimen dan Pembersihan Data

Penelitian ini tidak melakukan anotasi ulang terhadap sentimen setiap ulasan. Label sentimen menggunakan variabel `labelScoreBase` dalam dataset IGAR, yaitu rating 1-2 sebagai negatif, rating 3 sebagai netral, dan rating 4-5 sebagai positif. Dengan demikian, istilah sentimen dalam penelitian ini merujuk pada sentimen berbasis rating, bukan hasil klasifikasi ulang berdasarkan isi teks. Rating diperlakukan sebagai proksi sentimen karena teks ulasan tidak selalu sepenuhnya selaras dengan skor yang diberikan pengguna. Pembersihan teks dilakukan secara terbatas dengan menghapus ulasan tanpa konten, menyeragamkan nama aplikasi, mengubah teks menjadi huruf kecil, dan menormalkan spasi. Stemming dan penghapusan stopword tidak diterapkan karena analisis bergantung pada kemunculan bentuk kata dan frasa tertentu, sedangkan bahasa informal, singkatan, serta variasi ejaan tetap dipertahankan. Variasi yang lazim, seperti error-error dan install-instal, dimasukkan sebagai entri kamus yang berbeda. Karena label sentimen berada pada tingkat ulasan, label yang sama diterapkan pada seluruh aspek yang terdeteksi dalam satu ulasan; penelitian ini tidak mengestimasi polaritas yang berbeda untuk setiap aspek dalam ulasan yang sama.

2.3 Ekstraksi Aspek Layanan

Ekstraksi aspek dilakukan melalui dictionary-based matching. Aspek tidak berasal dari label bawaan IGAR. Kamus disusun secara manual oleh peneliti melalui dua tahap, yaitu menetapkan dimensi layanan berdasarkan kajian konseptual mengenai kualitas sistem dan layanan digital, kemudian menelaah istilah yang berulang dalam ulasan pengguna untuk memilih kata kunci yang mewakili setiap aspek. Proses tersebut menghasilkan tujuh aspek, yaitu akses akun dan autentikasi, performa dan stabilitas sistem, kualitas data dan informasi, fitur dan transaksi layanan, navigasi dan kemudahan penggunaan, layanan bantuan dan respons, serta pembaruan dan instalasi aplikasi. Kata kunci utama setiap aspek disajikan pada Tabel 2. Pencocokan dilakukan secara tidak peka huruf besar-kecil terhadap teks yang telah dinormalisasi. Suatu aspek dinyatakan muncul apabila ulasan memuat sedikitnya satu kata kunci atau frasa yang berkaitan dengan aspek tersebut. Satu ulasan dapat terhubung dengan lebih dari satu aspek sehingga jumlah keseluruhan deteksi aspek dapat melebihi jumlah ulasan yang mengandung aspek.

Validasi kamus dilakukan melalui evaluasi manual terhadap 200 ulasan yang dipilih secara acak dari kumpulan ulasan yang terdeteksi mengandung aspek layanan. Setiap ulasan diperiksa untuk menentukan apakah kata kunci yang terdeteksi benar-benar merujuk pada aspek yang ditetapkan. Sebanyak 183 dari 200 ulasan dinilai sesuai. Karena sampel hanya berasal dari hasil yang terdeteksi positif, ukuran yang dapat dihitung adalah presisi indikatif. Presisi indikatif kamus dihitung dengan persamaan berikut.

$$P_k = (n_\beta / n_v) \times 100\% = (183 / 200) \times 100\% = 91,50\%$$

dengan P_k sebagai presisi indikatif kamus, n_β sebagai jumlah ulasan yang terdeteksi secara benar, dan n_v sebagai jumlah ulasan yang diperiksa secara manual. Nilai 91,50% menunjukkan ketepatan deteksi pada sampel yang terjaring oleh kamus, tetapi tidak mengukur kemampuan kamus menemukan seluruh aspek yang relevan dalam korpus.

Tabel 2. Kamus aspek layanan

Aspek	Kata kunci utama
Akses akun dan autentikasi	<i>login</i> , masuk, daftar, <i>OTP</i> , verifikasi, akun, <i>password</i> , NIK
Performa dan stabilitas sistem	<i>error</i> , eror, <i>bug</i> , <i>crash</i> , <i>force close</i> , lemot, <i>server</i> , gagal, <i>blank</i>
Kualitas data dan informasi	data, sinkron, riwayat, sertifikat, saldo, tagihan, informasi, jadwal
Fitur dan transaksi layanan	bayar, tiket, beli, klaim, antrean, rujukan, subsidi, refund, BBM
Navigasi dan kemudahan penggunaan	menu, fitur, tampilan, UI, mudah, sulit, ribet, bingung, klik
Layanan bantuan dan respons	CS, <i>customer service</i> , respon, komplain, layanan, bantuan, admin, <i>chat</i>
Pembaruan dan instalasi aplikasi	<i>update</i> , versi, perbarui, <i>install</i> , instal, <i>uninstall</i> , <i>download</i> , APK

2.4 Teknik Analisis

Analisis dilakukan dalam empat tahap. Pertama, distribusi sentimen dihitung secara agregat dan pada setiap aplikasi. Persentase sentimen s pada aplikasi a dihitung sebagai berikut.

$$P_{a,s} = (n_{a,s} / N_a) \times 100\%$$

Dengan $n_{a,s}$ sebagai jumlah ulasan dengan sentimen s pada aplikasi a dan N_a sebagai jumlah seluruh ulasan pada aplikasi a . Kedua, setiap ulasan dipindai menggunakan kamus aspek. Jumlah deteksi aspek k dihitung sebagai $D_k = \sum I_{i,k}$, dengan $I_{i,k}$ bernilai 1 apabila ulasan i memuat kata kunci aspek k dan 0 apabila tidak.

Ketiga, proporsi sentimen negatif pada setiap aspek dihitung untuk menentukan aspek yang paling banyak berasosiasi dengan keluhan pengguna.

$$P_k^- = (D_k^- / D_k) \times 100\%$$

Dengan D_k^- sebagai jumlah deteksi aspek k pada ulasan negatif dan D_k sebagai jumlah seluruh deteksi aspek k . Keempat, sebaran aspek dalam ulasan negatif dihitung pada setiap aplikasi untuk membandingkan pola keluhan antaraplikasi.

$$P_{a,k}^- = (n_{a,k}^- / N_a^-) \times 100\%$$

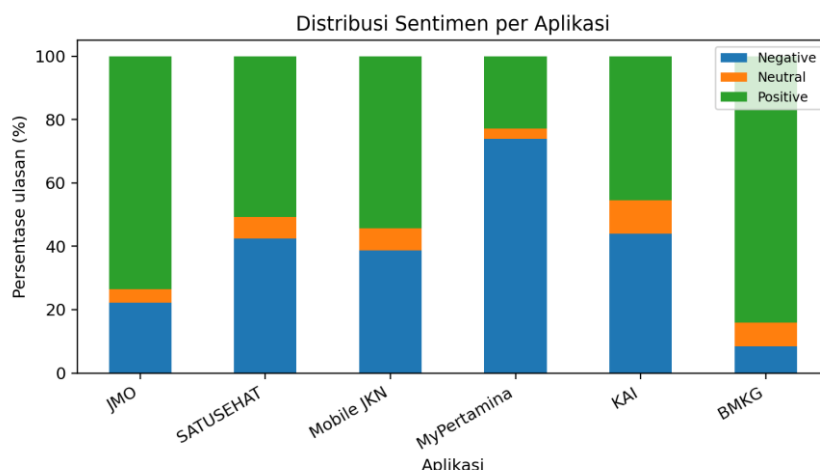
Dengan $n_{a,k}^-$ sebagai jumlah ulasan negatif pada aplikasi a yang terdeteksi mengandung aspek k dan N_a^- sebagai jumlah seluruh ulasan negatif pada aplikasi a . Pengolahan data dilakukan menggunakan Python. Pandas digunakan untuk pembersihan, pengelompokan, dan perhitungan data; *regular expression* digunakan untuk pencocokan kata kunci; sedangkan *Matplotlib* dan *Seaborn* digunakan untuk visualisasi. Hasil selanjutnya diinterpretasikan dalam konteks kualitas sistem, kualitas informasi, kemudahan penggunaan, dan layanan digital pemerintah.

Hubungan antara jenis aplikasi dan kategori sentimen diuji menggunakan Chi-square test of independence pada taraf signifikansi 0,05. Kekuatan hubungan diukur menggunakan Cramer's V, sedangkan arah perbedaan antarsel diperiksa melalui residual Pearson terstandarisasi. Uji Chi-square tidak diterapkan pada tabel aspek dan sentimen karena satu ulasan dapat memuat lebih dari satu aspek sehingga asumsi independensi observasi tidak terpenuhi.

3. HASIL DAN PEMBAHASAN

3.1 Distribusi Sentimen Ulasan Pengguna

Secara agregat, ulasan positif merupakan kategori terbesar dengan 336.430 ulasan atau 54,47% dari total data. Ulasan negatif berjumlah 246.892 atau 39,97%, sedangkan ulasan netral berjumlah 34.374 atau 5,56%. Komposisi ini menunjukkan bahwa persepsi pengguna tidak homogen. Dominasi sentimen positif tidak meniadakan persoalan layanan, karena hampir dua dari setiap lima ulasan berlabel negatif. Proporsi tersebut perlu dibaca sebagai sentimen berbasis rating, bukan sebagai hasil anotasi langsung terhadap makna teks ulasan.



Gambar 2. Distribusi sentimen per aplikasi

Gambar 2 menunjukkan perbedaan tajam antaraplikasi. BMKG memiliki proporsi positif tertinggi sebesar 84,01%, diikuti JMO sebesar 73,51%. Sebaliknya, MyPertamina menunjukkan proporsi negatif paling tinggi, yaitu 73,88%. KAI, Mobile JKN, dan SATUSEHAT memiliki komposisi yang lebih berimbang, meskipun pola sentimennya tetap berbeda. Perbedaan tersebut dapat berkaitan dengan fungsi layanan, tingkat ketergantungan pengguna, kompleksitas proses, dan ekspektasi terhadap setiap aplikasi. Aplikasi yang digunakan untuk memenuhi kebutuhan administratif atau transaksi yang mendesak cenderung menimbulkan ketidakpuasan lebih besar ketika akses atau proses utama gagal. Meskipun demikian, penjelasan ini bersifat interpretatif dan tidak membuktikan hubungan sebab-akibat. Jumlah ulasan BMKG dan KAI yang jauh lebih kecil juga mengharuskan perbandingan antaraplikasi dibaca terutama berdasarkan proporsi, bukan jumlah absolut.

Perbedaan distribusi tersebut diuji menggunakan Chi-square test of independence pada tabel kontingensi enam aplikasi dan tiga kategori sentimen. Hasil pengujian menunjukkan adanya hubungan yang signifikan antara jenis aplikasi dan distribusi sentimen, $\chi^2(10) = 78.927,72$, $p < 0,001$, dengan *Cramer's V* sebesar 0,253. Nilai *Cramer's V* menunjukkan kekuatan hubungan moderat. Dengan ukuran sampel yang sangat besar, signifikansi statistik perlu dibaca bersama ukuran efek; hasil ini menunjukkan bahwa jenis aplikasi berkaitan dengan pola sentimen, tetapi tidak sepenuhnya menentukan sentimen pengguna.

Tabel 3. Hasil uji *Chi-square* antara aplikasi dan sentimen

Pengujian	Chi-square	df	p	Cramer's V	Interpretasi
Aplikasi x sentimen	78.927,72	10	<0,001	0,253	Signifikan; kekuatan moderat

Pemeriksaan terhadap arah perbedaan menunjukkan bahwa ulasan negatif MyPertamina muncul lebih banyak daripada yang diharapkan apabila aplikasi dan sentimen tidak berhubungan, sedangkan ulasan positif JMO dan BMKG muncul lebih banyak daripada nilai harapan. Uji *Chi-square* tidak diterapkan secara langsung pada tabel aspek dan sentimen karena satu ulasan dapat masuk ke beberapa aspek. Kondisi multilabel tersebut menyebabkan observasi antarbaris aspek tidak sepenuhnya independen, sehingga penerapan *Chi-square* sederhana pada jumlah deteksi aspek dapat menghasilkan inferensi yang menyesatkan.

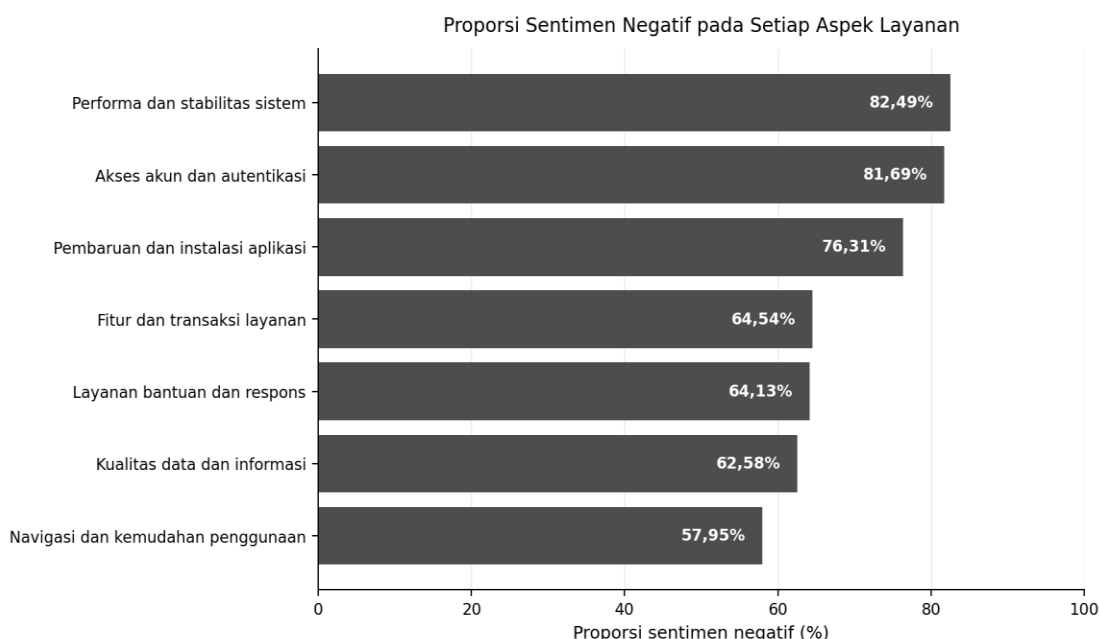
3.2 Sentimen Berdasarkan Aspek Layanan

Dari 617.696 ulasan, sebanyak 228.246 ulasan atau 36,95% memuat sedikitnya satu aspek layanan yang terdeteksi melalui kamus kata kunci. Angka tersebut menunjukkan bahwa lebih dari sepertiga korpus menyebut unsur layanan yang dapat dipetakan secara eksplisit. Sisanya tidak

selalu berarti tidak membahas layanan, karena sebagian ulasan sangat singkat, menggunakan ungkapan implisit, atau memakai istilah yang belum tercakup dalam kamus. Distribusi sentimen berdasarkan aspek layanan ditampilkan pada Tabel 4.

Tabel 4. Distribusi sentimen berdasarkan aspek layanan

Aspek	Jumlah deteksi	Negatif (%)	Netral (%)	Positif (%)
Akses akun dan autentikasi	76.808	81,69	7,61	10,70
Kualitas data dan informasi	67.349	62,58	9,25	28,17
Pembaruan dan instalasi aplikasi	60.297	76,31	8,60	15,09
Navigasi dan kemudahan penggunaan	48.729	57,95	4,95	37,11
Performa dan stabilitas sistem	41.034	82,49	8,20	9,31
Fitur dan transaksi layanan	20.058	64,54	7,08	28,38
Layanan bantuan dan respons	10.110	64,13	8,60	27,27



Gambar 3. Proporsi sentimen negatif pada setiap aspek layanan

Tabel 4 dan Gambar 3 menunjukkan bahwa performa dan stabilitas sistem memiliki proporsi negatif tertinggi sebesar 82,49%, diikuti akses akun dan autentikasi sebesar 81,69%. Kedua aspek tersebut berada pada lapisan dasar penggunaan aplikasi. Gangguan server, aplikasi yang berhenti, atau proses login dan verifikasi yang gagal dapat menghentikan seluruh perjalanan pengguna sebelum layanan utama diakses. Karena itu, dampak gangguan pada kedua aspek lebih luas daripada masalah pada fitur tambahan. Pola tersebut konsisten dengan deskripsi awal dataset IGAR [1], [2], yang menunjukkan dominasi kata negatif seperti susah, ribet, daftar, buka, OTP, dan error. Namun, konsistensi tersebut bukan perbandingan independen karena kedua analisis menggunakan sumber data yang sama.

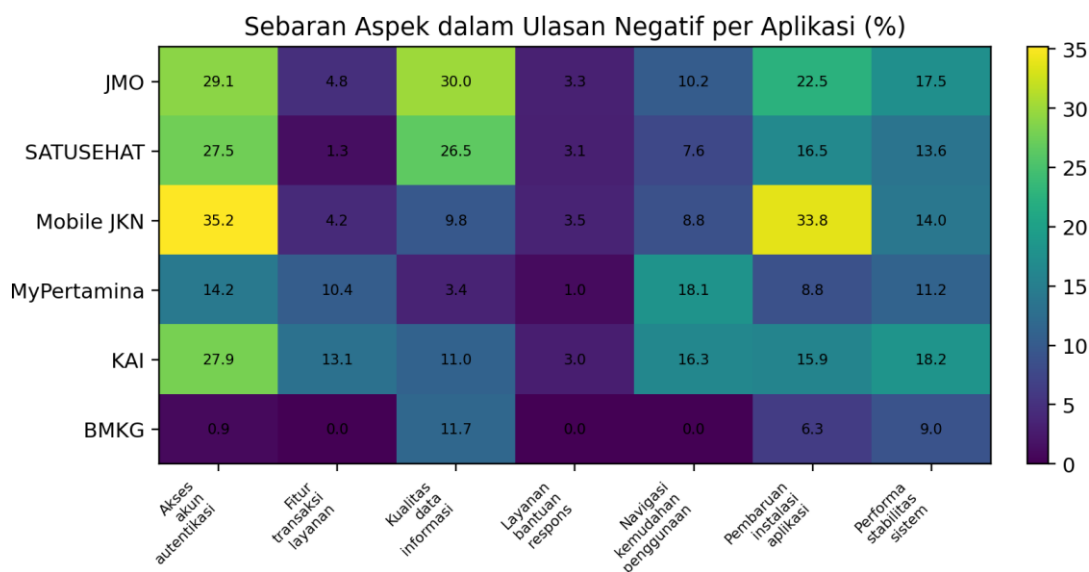
Aspek pembaruan dan instalasi aplikasi memiliki proporsi negatif sebesar 76,31%. Ulasan yang meminta pembaruan berulang, mengalami penurunan fungsi setelah pembaruan, atau gagal membuka versi terbaru menunjukkan bahwa pembaruan tidak selalu dipersepsikan sebagai peningkatan. Kualitas data dan informasi juga memiliki proporsi negatif sebesar 62,58%. Dalam layanan pemerintah, data kepesertaan, saldo, tagihan, jadwal, sertifikat, atau riwayat layanan

memiliki konsekuensi administratif. Ketidaksesuaian atau keterlambatan sinkronisasi tidak hanya memengaruhi kenyamanan, tetapi dapat menurunkan kepercayaan terhadap informasi resmi yang disajikan aplikasi.

Contoh ulasan memperjelas hubungan tersebut. Pengguna KAI menulis, "Pake aplikasi kai yg baru malah sering keluar sendiri dan ada tulisan kaiaccess terhenti", sedangkan pengguna JMO menyatakan, "Mau login aja gk bisa2 padahal sinyal kenceng. Udh versi yg terbaru tetep aja gk bisa di buka." Pengguna Mobile JKN juga mengeluhkan pembaruan berulang dan pemuatan yang lambat. Contoh tersebut menunjukkan bahwa keluhan berkaitan dengan hambatan penyelesaian tugas, bukan hanya ekspresi emosional. Meskipun demikian, proporsi negatif per aspek tidak dapat ditafsirkan sebagai tingkat kegagalan objektif aplikasi. Beberapa kata kunci pada aspek performa, seperti error, gagal, lemot, dan crash, sudah memiliki kecenderungan makna negatif. Dengan demikian, hasil pada Tabel 4 menunjukkan asosiasi kondisional antara kata kunci aspek dan label sentimen, serta dapat dipengaruhi oleh komposisi kamus yang digunakan.

3.3 Aspek Keluhan pada Setiap Aplikasi

Analisis ulasan negatif per aplikasi menunjukkan bahwa sumber keluhan tidak sama. Pada JMO, keluhan paling banyak terkait kualitas data dan informasi sebesar 30,02% dari ulasan negatif, diikuti akses akun sebesar 29,09%. Pada Mobile JKN, akses akun menjadi aspek terbesar sebesar 35,17%, diikuti pembaruan aplikasi sebesar 33,78%. Pada SATUSEHAT, akses akun mencapai 27,53% dan kualitas data 26,47%. Pada KAI, akses akun menjadi aspek paling dominan sebesar 27,93%, sedangkan pada MyPertamina keluhan paling banyak berada pada navigasi dan kemudahan penggunaan sebesar 18,11%. Persentase tersebut menggunakan seluruh ulasan negatif pada masing-masing aplikasi sebagai penyebut. Karena satu ulasan dapat memuat beberapa aspek, persentase antaraspek tidak harus berjumlah 100%.



Gambar 4. Aspek dalam ulasan negatif per aplikasi

Gambar 4 memperlihatkan bahwa pola keluhan berhubungan dengan karakter layanan yang disediakan. Pada SATUSEHAT dan Mobile JKN, autentikasi dan sinkronisasi data menjadi titik kritis karena akses terhadap informasi kesehatan dan kepesertaan bergantung pada identitas pengguna. Pada KAI, kegagalan login atau transaksi dapat langsung menghambat pencarian jadwal dan pembelian tiket. Sementara itu, dominasi aspek navigasi pada MyPertamina dapat mengindikasikan kesulitan pengguna dalam mengikuti alur layanan yang terdiri atas beberapa tahap. Interpretasi tersebut harus dipahami sebagai dugaan yang didukung pola ulasan, bukan bukti kausal mengenai desain teknis internal setiap aplikasi. Temuan ini menunjukkan bahwa

strategi perbaikan perlu disusun berdasarkan fungsi utama dan titik hambatan spesifik, bukan hanya berdasarkan skor rating keseluruhan.

3.4 Perbandingan dengan Penelitian Terdahulu

Hasil penelitian ini memperluas artikel penyedia dataset IGAR [1], [2]. Artikel IGAR berfokus pada pengumpulan data, distribusi label berbasis rating dan VADER, analisis kata dominan, serta penyediaan baseline klasifikasi. Penelitian ini menggunakan korpus yang sama untuk menjawab pertanyaan yang berbeda, yaitu aspek layanan apa yang paling banyak berasosiasi dengan sentimen negatif dan bagaimana pola tersebut berbeda antaraplikasi. Oleh karena itu, kontribusi penelitian ini bukan replikasi klasifikasi sentimen, melainkan penambahan lapisan interpretasi pada tingkat aspek layanan.

Secara metodologis, penelitian pada aplikasi Mitra Darat menunjukkan relevansi analisis sentimen berbasis aspek untuk mengevaluasi aplikasi layanan pemerintah [8]. Penelitian mengenai aplikasi *mobile government* Indonesia melalui *sentiment analysis* dan *topic modeling* juga menempatkan ulasan pengguna sebagai sumber identifikasi masalah layanan [9]. Sementara itu, penggunaan *CNN* dan *Word2Vec* dalam analisis sentimen berbasis aspek mewakili pendekatan berbasis model yang berbeda dari pencocokan kamus pada penelitian ini [10]. Pendekatan model berpotensi menangkap variasi semantik yang lebih luas, sedangkan pendekatan kamus lebih transparan dan mudah ditelusuri, tetapi bergantung pada kelengkapan serta ketepatan kata kunci.

Perbandingan numerik langsung dengan penelitian tersebut tidak dilakukan karena terdapat perbedaan aplikasi, periode data, skema pelabelan sentimen, definisi aspek, dan unit analisis. Membandingkan persentase tanpa kesetaraan metodologis dapat menghasilkan kesimpulan yang menyesatkan. Karena itu, perbandingan dalam penelitian ini ditempatkan pada kesamaan fokus masalah dan perbedaan pendekatan, bukan pada penyandingan angka mentah.

3.5 Implikasi terhadap Evaluasi Aplikasi Pemerintah

Hasil penelitian menunjukkan bahwa sentimen negatif pengguna terkonsentrasi pada aspek dasar, yaitu performa sistem, akses akun, pembaruan aplikasi, dan kualitas data. Dalam kajian keberhasilan *mobile government* berbasis model keberhasilan sistem informasi, kualitas sistem dan kualitas informasi menjadi prasyarat utama sebelum manfaat layanan dapat dirasakan pengguna [11]. Hal ini sejalan dengan konsep kualitas layanan *mobile government* yang menekankan kegunaan, interaksi, konsistensi, informasi, aksesibilitas, serta privasi dan keamanan sebagai dimensi penting pengalaman digital [12]. Dengan demikian, aplikasi yang kaya fitur tetap akan dinilai buruk jika gagal dibuka, lambat, tidak sinkron, atau menyajikan data yang tidak dipercaya.

Dari perspektif penerimaan teknologi, kemudahan penggunaan dan kegunaan yang dipersepsikan memengaruhi sikap pengguna terhadap sistem [13]. Pada layanan pemerintah, faktor tersebut perlu diperluas dengan mempertimbangkan konteks kewajiban administratif, kebutuhan informasi mendesak, dan ketergantungan warga pada data resmi. UTAUT juga menekankan bahwa ekspektasi kinerja, ekspektasi usaha, pengaruh sosial, dan kondisi fasilitasi dapat memengaruhi penerimaan teknologi [14]. Apabila pengguna berulang kali gagal melakukan *login*, memproses klaim, memperoleh tiket, membaca data kesehatan, atau mengakses informasi cuaca, maka aplikasi kehilangan fungsi utamanya sebagai kanal layanan publik.

Temuan ini juga dapat dikaitkan dengan standar kualitas perangkat lunak, terutama aspek *functional suitability*, *reliability*, *interaction capability*, dan *security* [15], [16]. Akses akun dan autentikasi berkaitan dengan keamanan dan kemudahan penggunaan. Performa sistem berkaitan dengan *reliability*. Kualitas data berkaitan dengan *functional suitability* dan akurasi informasi. Dengan demikian, analisis sentimen berbasis aspek dapat menjadi instrumen awal untuk mendeteksi area perbaikan aplikasi pemerintah berdasarkan pengalaman pengguna aktual.

3.6 Keterbatasan Validasi Ekstraksi Aspek

Evaluasi manual terhadap 200 ulasan yang terdeteksi mengandung aspek menunjukkan bahwa 183 ulasan sesuai dengan aspek yang ditetapkan, sedangkan 17 ulasan tidak sesuai. Hasil tersebut setara dengan presisi indikatif 91,50% pada sampel hasil deteksi. Dengan kata lain, sekitar 8,50% sampel terdeteksi tidak mendukung aspek yang dimaksud. Ringkasan validasi yang tersedia tidak memuat catatan per kasus, sehingga penelitian ini tidak menyajikan contoh false positive tertentu sebagai temuan empiris. Menambahkan contoh tanpa catatan anotasi asli akan menghasilkan interpretasi yang tidak dapat diverifikasi.

Secara metodologis, kata yang bersifat umum, seperti data, masuk, layanan, dan fitur, berpotensi digunakan dalam konteks yang berbeda dari aspek yang ditetapkan. Sebaliknya, ulasan yang memakai sinonim, slang, ungkapan implisit, atau sarkasme berpotensi tidak terdeteksi. Pernyataan ini merupakan identifikasi potensi sumber kesalahan berdasarkan karakter pendekatan kamus, bukan hasil kuantifikasi dari 17 kasus yang tidak sesuai.

Desain validasi hanya mengambil ulasan yang telah terdeteksi oleh kamus. Oleh karena itu, data yang tersedia hanya memungkinkan identifikasi hasil deteksi yang benar dan tidak sesuai. Ulasan yang tidak terdeteksi tidak diperiksa secara manual, sehingga false negative dan true negative tidak diketahui. Akibatnya, confusion matrix lengkap, recall, F1-score, dan accuracy keseluruhan tidak dapat dihitung secara sah. Validasi lanjutan perlu mengambil sampel terstratifikasi dari ulasan terdeteksi dan tidak terdeteksi, kemudian melakukan anotasi manual per aspek agar seluruh komponen confusion matrix tersedia.

3.7 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, label sentimen diturunkan dari rating pengguna sehingga tidak selalu identik dengan makna tekstual ulasan. Karena label berada pada tingkat ulasan, penelitian ini tidak membedakan polaritas untuk setiap aspek yang muncul dalam ulasan yang sama. Kedua, ekstraksi aspek berbasis kamus bergantung pada pilihan kata kunci dan belum sepenuhnya menangkap sinonim, konteks implisit, slang, atau sarkasme. Ketiga, distribusi dan periode ulasan berbeda antaraplikasi; karena itu, perbandingan lintas aplikasi bersifat deskriptif dan tidak dapat ditafsirkan sebagai pemeringkatan kualitas pada periode yang sama. Keempat, validasi 200 ulasan hanya mengukur presisi indikatif pada hasil deteksi dan belum menghasilkan confusion matrix lengkap. Kelima, uji Chi-square menunjukkan hubungan antara aplikasi dan sentimen, tetapi tidak menjelaskan penyebab hubungan tersebut. Temuan perlu dipadukan dengan bukti teknis dan penelitian pengguna sebelum digunakan sebagai dasar keputusan pengembangan.

4. KESIMPULAN

Penelitian ini mengevaluasi enam aplikasi pemerintah dan layanan publik Indonesia menggunakan analisis sentimen berbasis aspek terhadap 617.696 ulasan pengguna dalam dataset IGAR. Hasil penelitian menunjukkan bahwa persepsi pengguna tidak cukup dijelaskan melalui sentimen agregat, karena keluhan terkonsentrasi pada aspek layanan tertentu, terutama performa dan stabilitas sistem, akses akun dan autentikasi, pembaruan aplikasi, serta kualitas data. Temuan tersebut menegaskan bahwa kualitas layanan digital publik tidak hanya ditentukan oleh ketersediaan fitur, tetapi juga oleh keandalan sistem, kelancaran autentikasi, ketepatan data, dan kemudahan alur penggunaan.

Kontribusi ilmiah penelitian ini terletak pada penerapan kerangka analisis sentimen berbasis aspek untuk membandingkan pola keluhan pengguna pada beberapa aplikasi pemerintah dan layanan publik dengan fungsi yang berbeda. Pemetaan tujuh aspek layanan memperluas analisis sentimen dari klasifikasi polaritas pada tingkat ulasan menuju identifikasi dimensi layanan yang berasosiasi dengan ketidakpuasan pengguna. Pendekatan berbasis kamus yang digunakan juga memberikan kerangka evaluasi yang transparan, mudah ditelusuri, dan dapat direplikasi atau dikembangkan pada penelitian layanan digital publik lainnya. Dengan demikian, penelitian ini berkontribusi pada pengembangan kajian e-government, kualitas sistem informasi, dan analisis

pengalaman pengguna dengan menghubungkan data ulasan publik dengan prioritas evaluasi layanan.

Secara praktis, hasil penelitian menunjukkan bahwa pengembangan aplikasi layanan publik perlu memprioritaskan penguatan infrastruktur dasar dan fungsi inti, khususnya keandalan server, autentikasi OTP, sinkronisasi basis data, dan kejelasan alur layanan, sebelum memperluas fitur atau melakukan perubahan antarmuka. Penelitian selanjutnya disarankan menggunakan validasi manual yang lebih luas, melibatkan lebih dari satu anotator, memperluas kamus aspek, dan membandingkan pendekatan berbasis kamus dengan model kontekstual seperti transformer agar cakupan serta ketepatan deteksi aspek dapat ditingkatkan.

5. DAFTAR RUJUKAN

- [1] M. Isnan and B. Pardamean, 'IGAR: Indonesian government applications review for sentiment analysis dataset', Data Brief, vol. 66, p. 112708, 2026, doi: 10.1016/j.dib.2026.112708. <https://doi.org/10.1016/j.dib.2026.112708>
PMid:41952845 PMCID:PMC13054429
- [2] I. Mahmud and P. Bens, 'IGAR: Indonesian Government App Review Dataset', 2025. doi: 10.17632/7zryc6k76z.3.
- [3] N. Malik and M. Bilal, 'Natural language processing for analyzing online customer reviews: A survey, taxonomy, and open research challenges', PeerJ Comput. Sci., vol. 10, p. e2203, 2024, doi:10.7717/peerj-cs.2203. <https://doi.org/10.7717/peerj-cs.2203>
PMid:39145232 PMCID:PMC11323031
- [4] Y. C. Hua, P. Denny, J. Wicker, and K. Taskova, 'A systematic review of aspect-based sentiment analysis: domains, methods, and trends', Artif. Intell. Rev., vol. 57, p. 296, 2024, doi:10.1007/s10462-024-10906-z. <https://doi.org/10.1007/s10462-024-10906-z>
- [5] A. D. Gustiavani and M. Muljono, 'Sentiment classification of Indonesian e-government application reviews using advanced learning models', Journal of Applied Informatics and Computing, vol. 10, no. 2, pp. 1662-1673, 2026, doi: 10.30871/jaic.v10i2.12217. <https://doi.org/10.30871/jaic.v10i2.12217>
- [6] S. Zhang, L. Wang, and B. Liu, 'Deep learning for sentiment analysis: A survey', Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 8, no. 4, 2018, doi: 10.1002/widm.1253. <https://doi.org/10.1002/widm.1253>
- [7] A. Wankhade, A. C. S. Rao, and C. Kulkarni, 'A survey on sentiment analysis methods, applications, and challenges', Artif. Intell. Rev., vol. 55, pp. 5731-5780, 2022, doi: 10.1007/s10462-022-10144-1. <https://doi.org/10.1007/s10462-022-10144-1>
PMCID:PMC10164419
- [8] M. Ishaq, D. Lestari, and M. Marchenko, 'Implementation of aspect-based sentiment analysis on the Mitra Darat app user reviews using machine learning', KLIK: Kajian Ilmiah Informatika dan Komputer, vol. 4, no. 6, pp. 2763-2776, 2024, doi: 10.30865/klik.v4i6.1889.
- [9] R. B. Hamid, C. Andriansyah, D. I. Sensuse, S. Lusa, D. Elisabeth, and N. Safitri, 'Sentiment analysis and topic modeling for discovering knowledge in Indonesian mobile government applications', Jurnal Teknik Informatika (JUTIF), vol. 6, no. 5, pp. 3188-3203, 2025, doi: 10.52436/1.jutif.2025.6.5.4991. <https://doi.org/10.52436/1.jutif.2025.6.5.4991>

- [10] N. I. Lestari, S. M. Taib, W. Wibowo, I. A. Aziz, and M. R. Habibi, 'Aspect-Based Sentiment Analysis for Mobile App Review Using Convolutional Neural Network (CNN) and Word2Vec', in 2024 IEEE 7th International Conference on Electrical, Electronics and System Engineering (ICEESE), IEEE, Nov. 2024, pp. 1-6. doi: 10.1109/ICEESE62315.2024.10828541. <https://doi.org/10.1109/ICEESE62315.2024.10828541>
- [11] C. Wang and T. S. H. Teo, 'Online service quality and perceived value in mobile government success: An empirical study of mobile police in China', *Int. J. Inf. Manage.*, vol. 52, p. 102076, 2020, doi:10.1016/j.ijinfomgt.2020.102076. <https://doi.org/10.1016/j.ijinfomgt.2020.102076>
- [12] A. J. Desmal, M. K. Othman, S. Hamid, and A. H. Zolait, 'A user satisfaction model for mobile government services: a literature review', *PeerJ Comput. Sci.*, vol. 8, p. e1074, 2022, doi: 10.7717/peerj-cs.1074. <https://doi.org/10.7717/peerj-cs.1074>
PMid:36091981 PMCID:PMC9455267
- [13] M. N. Huda, 'Analysis the critical factors of m-government service acceptance: An integrating theoretical model between TAM and ECM', *Policy & Governance Review*, vol. 7, no. 2, pp. 109-124, 2023, doi:10.30589/pgr.v7i2.721. <https://doi.org/10.30589/pgr.v7i2.721>
- [14] I. G. L. A. Aprianto, 'Tinjauan Literatur: Penerimaan Teknologi Model UTAUT', *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 2, no. 1, pp. 138-144, 2022, doi:10.24002/konstelasi.v2i1.5377. <https://doi.org/10.24002/konstelasi.v2i1.5377>
- [15] M. D. Mulyawan, I. N. S. Kumara, I. B. A. Swamardika, and K. O. Saputra, 'Kualitas Sistem Informasi Berdasarkan ISO/IEC 25010: Literature Review', *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 1, 2021, doi:10.24843/MITE.2021.v20i01.P02. <https://doi.org/10.24843/MITE.2021.v20i01.P02>
- [16] F. Nooridianah and A. Kurniawati, 'Analisa Kualitas Perangkat Lunak Sistem Informasi Keuangan Mikro (SIKM) Menggunakan ISO 25010', *Jurnal Ilmiah Komputasi*, vol. 20, no. 3, pp. 299-306, 2021. <https://doi.org/10.32409/jikstik.20.3.2742>