E-ISSN: 3025-0501
Vol. 4, No. 5, September 2026
pp. 101-121
DOI: https://doi.org/10.57185/qkc5mc07

Almakki Health Informatics Journal

Analisis Ulasan Pengguna Aplikasi SATUSEHAT Mobile pada Google Play Store Berdasarkan Karakteristik Kualitas ISO/IEC 25010

Rafi Farizki
Politeknik Siber Cerdika Internasional, Indonesia
Corresponding Author: rafifarizki90@gmail.com

Keywords:
SATUSEHAT Mobile, ISO/IEC 25010, software quality, user reviews, health informatics

This is an open access article under the license.CC BY-SA

Copyright © 2026 by Author. Published by Al-Makki Publisher

Abstract

SATUSEHAT Mobile is the primary citizen facing channel within Indonesia's national digital health transformation ecosystem, with more than 50 million downloads on Google Play Store. The application holds an average rating of only 2.2 out of five stars based on 1.12 million reviews. The gap between adoption and satisfaction remains inadequately explained because prior studies generally stop at sentiment polarity classification and do not map user complaints onto a standard software quality model. This study aims to map the content of SATUSEHAT Mobile user reviews onto the ISO/IEC 25010:2023 product quality characteristics and to establish improvement priorities based on the severity level of each characteristic. A descriptive quantitative design with directed content analysis was applied to 2,389 Indonesian language reviews collected from Google Play Store between 19 September 2025 and 20 October 2026. Rule assisted coding was manually verified internally (81.8 percent precision). The results reveal a polarised rating distribution, with 45.5 percent of reviews awarding one star and 39.6 percent awarding five stars. Reliability emerged as the most frequently criticised characteristic, appearing in 656 reviews or 27.5 percent of the corpus, with a severity index of 91.8 percent. Authentication failures involving login, PIN, and verification codes accounted for 71.5 percent of all reliability complaints. Security and safety were rarely mentioned yet carried high severity indices of 77.8 percent and 86.7 percent respectively. The study concludes that application failure is concentrated at the authentication gateway, meaning most users never reach the health functions that justify the application's existence.

Kata Kunci:
SATUSEHAT Mobile, ISO/IEC 25010, kualitas perangkat lunak, ulasan pengguna, informatika kesehatan

Abstrak

Aplikasi SATUSEHAT Mobile merupakan kanal utama masyarakat dalam ekosistem transformasi digital kesehatan nasional, dengan lebih dari 50 juta unduhan di Google Play Store. Meskipun demikian, aplikasi ini hanya memperoleh rata-rata penilaian 2,2 dari lima bintang berdasarkan 1,12 juta ulasan. Kesenjangan antara tingkat adopsi dan tingkat kepuasan tersebut belum terjelaskan secara memadai karena penelitian terdahulu umumnya berhenti pada klasifikasi polaritas sentimen dan tidak memetakan keluhan pengguna ke dimensi kualitas perangkat lunak yang baku. Penelitian ini bertujuan memetakan isi ulasan pengguna SATUSEHAT Mobile ke dalam karakteristik kualitas produk ISO/IEC 25010:2023 dan menentukan prioritas perbaikan berdasarkan tingkat keparahan tiap karakteristik. Penelitian menggunakan desain deskriptif kuantitatif dengan teknik analisis

isi terarah terhadap 2.389 ulasan berbahasa Indonesia yang dikumpulkan dari Google Play Store pada periode 19 September 2025 sampai 20 Oktober 2026. Koding berbantuan aturan diverifikasi secara manual dengan ketepatan 81,8 persen. Hasil penelitian menunjukkan sebaran penilaian yang terpolarisasi, dengan 45,5 persen ulasan berbintang satu dan 39,6 persen berbintang lima. Keandalan menjadi karakteristik yang paling sering dikeluhkan, yaitu pada 656 ulasan atau 27,5 persen korpus, dengan indeks keparahan 91,8 persen. Sebanyak 71,5 persen keluhan keandalan terpusat pada proses autentikasi berupa kegagalan masuk, PIN, dan kode verifikasi. Karakteristik keamanan dan keselamatan jarang disebut namun memiliki indeks keparahan tinggi, yaitu 77,8 persen dan 86,7 persen. Penelitian ini menyimpulkan bahwa kegagalan aplikasi terkonsentrasi pada gerbang autentikasi, sehingga sebagian besar pengguna tidak pernah mencapai fungsi kesehatan yang menjadi alasan keberadaan aplikasi tersebut.

PENDAHULUAN

Transformasi digital kesehatan di Indonesia memasuki fase konsolidasi setelah pemerintah memutuskan untuk menyatukan ratusan aplikasi kesehatan yang sebelumnya berjalan sendiri sendiri ke dalam satu platform nasional (Kementerian Kesehatan Republik Indonesia, 2025b). Kementerian Kesehatan membangun ekosistem SATUSEHAT yang terdiri atas beberapa lapisan pengguna, yaitu SATUSEHAT Mobile untuk masyarakat, SATUSEHAT IndonesiaKu untuk tenaga kesehatan dan kader, SATUSEHAT Data untuk pengambil keputusan, serta SATUSEHAT Rekam Medis untuk fasilitas pelayanan kesehatan. Konsolidasi ini bukan semata persoalan efisiensi teknis, melainkan prasyarat bagi interoperabilitas data kesehatan nasional yang selama ini terfragmentasi di tingkat fasilitas.

Landasan regulasinya diperkuat oleh Peraturan Menteri Kesehatan Nomor 24 Tahun 2022 tentang Rekam Medis (Kementerian Kesehatan Republik Indonesia, 2022), yang mewajibkan seluruh fasilitas pelayanan kesehatan menyelenggarakan rekam medis elektronik dan mengirimkan datanya ke platform nasional. Kewajiban tersebut menggeser posisi pasien dari sekadar objek pencatatan menjadi pemilik dan pengakses riwayat kesehatannya sendiri. Di titik inilah SATUSEHAT Mobile menjalankan fungsi strategis, yaitu sebagai antarmuka tunggal yang memungkinkan warga melihat resume medis, riwayat imunisasi, hasil laboratorium, resep elektronik, dan hasil skrining kesehatan dari berbagai fasilitas dalam satu genggaman.

Aplikasi ini merupakan hasil transformasi PeduliLindungi yang beralih nama dan fungsi sejak 1 Maret 2023 (Kementerian Kesehatan Republik Indonesia, 2025b), sehingga mewarisi basis pengguna sangat besar dari masa penanganan pandemi. Peran aplikasi tersebut kembali meluas ketika pemerintah menjadikannya pintu masuk Program Cek Kesehatan Gratis, yang mengharuskan masyarakat mengunduh aplikasi dan melengkapi data diri untuk memperoleh jadwal pemeriksaan (Kementerian Kesehatan Republik Indonesia, 2025a). Akibatnya, SATUSEHAT Mobile kini tercatat lebih dari 50 juta kali diunduh di Google Play Store (Ministry of Health Republic of Indonesia, 2026) dan secara praktis menjadi perangkat lunak kesehatan dengan jangkauan terluas di Indonesia.

Angka adopsi yang besar tersebut ternyata tidak sejalan dengan tingkat kepuasan penggunanya. Pada 21 Oktober 2026, halaman resmi aplikasi di Google Play Store

menampilkan rata-rata penilaian 2,2 dari lima bintang yang dihimpun dari sekitar 1,12 juta ulasan (Ministry of Health Republic of Indonesia, 2026). Kesenjangan antara skala penggunaan dan skala kepuasan ini menimbulkan persoalan serius dalam konteks informatika kesehatan. Aplikasi yang digunakan puluhan juta warga untuk mengakses data medis pribadi, tetapi dinilai buruk oleh penggunanya sendiri, berisiko menurunkan kepercayaan publik terhadap keseluruhan program transformasi digital kesehatan, bukan hanya terhadap satu produk perangkat lunak.

This pattern is not unique to Indonesia. The World Health Organization (2021) adopted a global strategy on digital health for 2020 to 2025 that positions digital health technologies as a component of national health systems, and more than 50 countries developed COVID contact tracing apps (Garousi et al., 2022). Public health apps nevertheless tend to attract critical evaluation. A mining study of 39,425 reviews of nine European contact tracing apps found that users were generally dissatisfied, with high battery drainage and doubts about whether the apps were really working among the common complaints (Garousi et al., 2022), and a large scale analysis of more than five million reviews of 278 mHealth apps reported that about half of the users who raised account and login issues gave a one star rating (Haggag et al., 2022). A national health application with tens of millions of users and a rating of 2.2 therefore belongs to a wider international pattern in which technical problems, including account and login failures, weigh heavily in public evaluation.

Sejumlah penelitian telah berupaya membaca umpan balik pengguna aplikasi ini. Imanuddin, Adi, dan Gernowo (2023) mengklasifikasikan 25.000 ulasan menggunakan Support Vector Machine dan menemukan dominasi sentimen negatif sebesar 73,4 persen dengan akurasi model 91 persen. Dewi, Dharmendra, dan Setiasih (2023) membandingkan beberapa algoritma klasifikasi dengan teknik penyeimbangan kelas Tomek Links dan melaporkan akurasi tertinggi pada Support Vector Machine. Anggiyanti and Papilaya (2026) combined BERT based sentiment classification, which reached 95 percent accuracy, with BERTopic topic modelling of SATUSEHAT reviews and reported that opinions lean negative and centre on usability and accessibility. Seluruh penelitian tersebut konsisten menunjukkan bahwa persepsi pengguna cenderung negatif, namun berhenti pada tingkat polaritas. Studies of other Indonesian health applications share this focus on polarity. Tarwoto et al. (2025) fine tuned IndoBERT to classify Mobile JKN reviews as positive, negative, or neutral and reported an average accuracy of 97.28 percent, whereas Nurfikri (2022) analysed 1,129 Halodoc reviews through star rating categories and NVivo word frequencies and found that about 74.8 percent of the ratings were positive, which shows that polarity differs across applications and still does not reveal which quality dimension fails.

Pembacaan ulasan toko aplikasi sebagai sumber data rekayasa perangkat lunak sendiri bukan praktik baru. Dąbrowski, Letier, Perini, dan Susi (2022) melalui tinjauan sistematis menunjukkan bahwa analisis ulasan aplikasi telah berkembang menjadi bidang kajian tersendiri dengan beragam tugas analitis, mulai dari klasifikasi, ekstraksi fitur, hingga penentuan prioritas pemeliharaan. Earlier syntheses had already mapped this field (Genc-Nayebi & Abran, 2017; Martin et al., 2017), and a large scale analysis of mHealth app reviews has shown that such reviews can be classified into aspects such as account and login issues and privacy concerns (Haggag et al., 2022). Pemetaan ulasan ke dalam model kualitas baku juga sudah dirintis di luar negeri. Sönmez dan Kekül (2024) mengklasifikasikan 1.681 komentar

pengguna ke dalam metrik kualitas ISO/IEC 25010 menggunakan sejumlah algoritma pembelajaran mesin dan memperoleh akurasi tertinggi 87 persen. Haoues et al. (2023) similarly labelled 1,681 reviews of 86 mHealth apps against ISO/IEC 25010 quality characteristics and compared six machine learning classifiers, which supports applying this standard to health application reviews. Yang belum tersedia adalah penerapan kerangka serupa pada aplikasi kesehatan pemerintah di Indonesia, padahal konteks tersebut memiliki taruhan yang berbeda karena menyangkut akses warga terhadap data medisnya sendiri.

Di sinilah letak celah penelitian yang hendak diisi. Informasi bahwa mayoritas ulasan bernada negatif tidak dapat langsung diterjemahkan menjadi tindakan perbaikan, sebab polaritas tidak menjelaskan dimensi kualitas mana yang sesungguhnya gagal. Keluhan mengenai kegagalan verifikasi nomor induk kependudukan, misalnya, menuntut perbaikan yang sama sekali berbeda dari keluhan mengenai data medis orang lain yang muncul di akun pengguna, meskipun keduanya terklasifikasi sebagai sentimen negatif. Pengelola aplikasi membutuhkan peta yang menunjukkan letak kegagalan pada struktur kualitas perangkat lunak, bukan sekadar proporsi ketidakpuasan.

Standar ISO/IEC 25010 menyediakan kerangka yang dibutuhkan untuk keperluan tersebut. Revisi tahun 2023 menyusun kualitas produk perangkat lunak ke dalam sembilan karakteristik, yaitu kesesuaian fungsional, efisiensi kinerja, kompatibilitas, kapabilitas interaksi, keandalan, keamanan, kemampurawatan, fleksibilitas, dan keselamatan (International Organization for Standardization, 2023). Revisi tersebut mengganti karakteristik usability menjadi kapabilitas interaksi, mengganti portability menjadi fleksibilitas, serta menambahkan keselamatan sebagai karakteristik tingkat pertama. Penambahan terakhir sangat relevan bagi aplikasi kesehatan, karena kesalahan penyajian data medis berpotensi menimbulkan risiko klinis, bukan sekadar ketidaknyamanan.

Penelitian ini menawarkan kebaruan dengan memperlakukan ulasan publik di toko aplikasi sebagai data evaluasi kualitas yang terstruktur, bukan sebagai data sentimen semata. Tujuan penelitian dirumuskan dalam tiga pertanyaan. Pertama, bagaimana sebaran keluhan dan apresiasi pengguna SATUSEHAT Mobile jika dipetakan ke dalam karakteristik kualitas ISO/IEC 25010:2023. Kedua, karakteristik kualitas mana yang memiliki tingkat keparahan tertinggi apabila sebaran tersebut ditabulasikan silang dengan penilaian bintang. Ketiga, prioritas perbaikan apa yang dapat direkomendasikan kepada pengelola aplikasi berdasarkan pemetaan tersebut.

Kontribusi penelitian ini bersifat ganda. Secara metodologis, penelitian menghasilkan skema koding ulasan aplikasi kesehatan berbasis ISO/IEC 25010:2023 yang dilengkapi definisi operasional dan indikator, sehingga dapat digunakan ulang untuk mengevaluasi aplikasi kesehatan lain. Secara praktis, penelitian menghasilkan daftar prioritas perbaikan yang dapat langsung dipakai pengelola aplikasi kesehatan pemerintah dalam menyusun peta jalan pengembangan produk.

METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan desain deskriptif kuantitatif dengan teknik analisis isi terarah sebagaimana dirumuskan Hsieh dan Shannon (2005). Teknik tersebut dipilih karena kategori

analisis tidak dibangun secara induktif dari data, melainkan diturunkan terlebih dahulu dari kerangka teoretis yang telah mapan, yaitu model kualitas produk ISO/IEC 25010:2023. Data kuantitatif berupa frekuensi dan tabulasi silang diperoleh melalui proses koding terhadap data teks ulasan. Content analysis requires that categories and coding units are defined explicitly enough for another analyst to reproduce the coding (Krippendorff, 2019); the coding dictionary is therefore documented in Tabel 1 and in Supplementary Material 1.

Objek dan Sumber Data

Objek penelitian adalah aplikasi SATUSEHAT Mobile dengan identitas paket com.telkom.tracencare yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia. Data primer penelitian berupa ulasan pengguna yang dipublikasikan secara terbuka di Google Play Store. Ulasan tersebut termasuk data publik yang dapat diakses tanpa autentikasi dan tidak melibatkan intervensi terhadap subjek manusia, sehingga penelitian tidak memerlukan persetujuan etik. Nama pengguna yang menyertai ulasan tidak diikutsertakan dalam basis data penelitian maupun dalam laporan ini, dan setiap kutipan ilustratif disajikan tanpa identitas penulisnya.

Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan teknik penambangan ulasan terhadap antarmuka data publik Google Play Store. Parameter pengambilan ditetapkan pada bahasa Indonesia dan wilayah Indonesia dengan urutan ulasan terbaru, sehingga korpus yang terbentuk merepresentasikan kondisi aplikasi pada versi mutakhir. Pengambilan dilakukan pada 21 Oktober 2026 dan menghasilkan 2.400 ulasan yang mencakup periode 19 September 2025 sampai 20 Oktober 2026. Setiap ulasan terekam bersama atribut penilaian bintang, tanggal, versi aplikasi, dan jumlah tanda suka. Metadata aplikasi berupa rata-rata penilaian, jumlah ulasan, dan jumlah unduhan turut direkam sebagai data pendukung. The reviews were retrieved with [name and version of the scraping tool or library, to be completed by the author], with the language set to Indonesian, the country set to Indonesia, the sort order set to newest, and a retrieval count of 2,400 that was fixed by the researcher. Because the retrieval interface returns reviews that contain text and match the language and country filters, whereas the 1.12 million figure on the app page may also include ratings without text, the 2,400 reviews are not a fixed proportion of that total, and spanning 13 months they amount to about 6 reviews per day. The representativeness of the monthly trends is therefore interpreted with caution.

Kriteria Inklusi dan Eksklusi

Ulasan diikutsertakan apabila memuat sekurang kurangnya satu kata bermakna dalam bahasa Indonesia. Ambang panjang minimum tidak diterapkan karena ulasan pendek berisi pujian ringkas merupakan bagian sah dari sebaran opini, dan pengeluarannya akan membuat korpus condong ke arah negatif secara artifisial. Ulasan dikeluarkan apabila hanya berisi emotikon atau karakter tanpa kata, serta apabila merupakan duplikat identik pada ulasan sepanjang delapan kata atau lebih, yang diperlakukan sebagai indikasi spam. Setelah penyaringan, sebanyak 10 ulasan dikeluarkan karena tidak memuat kata bermakna dan 1 ulasan dikeluarkan karena duplikasi, sehingga korpus analitik berjumlah 2.389 ulasan.

Pembersihan dan Penyiapan Data

Tahap penyiapan data mencakup penyeragaman huruf kecil, penghapusan tautan dan karakter bukan huruf, serta normalisasi ragam tulis tidak baku yang lazim dalam ulasan berbahasa Indonesia. Normalisasi diperlukan karena bentuk negasi muncul dalam banyak varian, seperti gabisa, ga bisa, gak bisa, nggak bisa, dan tdk bisa, yang seluruhnya dipetakan ke bentuk baku tidak bisa. Tanpa tahap ini, pencocokan indikator akan kehilangan sebagian besar keluhan yang sebenarnya relevan.

Skema Koding

Skema koding disusun berdasarkan sembilan karakteristik ISO/IEC 25010:2023. Karakteristik kemampurawatan tidak dijadikan kategori aktif karena bersifat internal bagi pengembang dan tidak dapat diamati dari perspektif pengguna akhir. Karakteristik keselamatan dipertahankan secara terbatas untuk menampung ulasan yang melaporkan potensi risiko akibat penyajian data medis yang keliru. Dua kategori tambahan ditetapkan di luar model, yaitu apresiasi umum untuk ulasan positif yang tidak menyebut dimensi kualitas tertentu, dan ketidakpuasan umum untuk ulasan negatif yang tidak menyebut dimensi kualitas tertentu. Kedua kategori tambahan ini mencegah pemaksaan ulasan generik ke dalam karakteristik yang tidak benar benar dirujuk pengguna. Definisi operasional dan indikator setiap kategori disajikan pada Tabel 1. For interpretation, K4 and K5 are read at the sub characteristic level: the sub characteristics most relevant to interaction capability (K4) are user error protection, self descriptiveness, and user assistance, and those most relevant to reliability (K5) are faultlessness, availability, fault tolerance, and recoverability (International Organization for Standardization, 2023). The complete indicator list for each category (coding dictionary) is provided as Supplementary Material 1 so that the coding scheme can be replicated.

Tabel 1. Skema koding ulasan berbasis ISO/IEC 25010:2023

Kode	Karakteristik	Definisi operasional dalam konteks ulasan	Contoh indikator
K1	Kesesuaian fungsional	Fitur tidak tersedia, tidak lengkap, atau memberi hasil yang keliru	tidak ditemukan, tidak bisa diunduh, data tidak muncul, tidak ada pilihan kota
K2	Efisiensi kinerja	Kecepatan respons, konsumsi sumber daya, atau daya tampung aplikasi	lemot, loading lama, berat, boros kuota
K3	Kompatibilitas	Ketidakcocokan dengan perangkat, versi sistem operasi, atau aplikasi lain	tidak support, tidak bisa dipasang di perangkat, bentrok
K4	Kapabilitas interaksi	Kemudahan pemahaman, pengoperasian, navigasi, dan ketersediaan bantuan pengguna (user error protection, self descriptiveness, user assistance)	ribet, membingungkan, mempersulit, tidak ada panduan
K5	Keandalan	Kegagalan, keterhentian, ketersediaan layanan, dan pemulihan setelah gangguan	error, force close, tidak bisa login, PIN salah, OTP tidak masuk, server gangguan

		(faultlessness, availability, fault tolerance, recoverability)	
K6	Keamanan	Kerahasiaan, keutuhan, dan keaslian data serta kekhawatiran privasi	data bocor, privasi, akun diambil alih
K7	Fleksibilitas	Pemasangan, pembaruan, penggantian versi, dan penyesuaian lingkungan	setelah update malah, gagal install, versi lama
K8	Keselamatan	Potensi risiko kesehatan akibat kekeliruan informasi medis dalam aplikasi	riwayat bukan milik saya, tertukar, padahal belum pernah periksa
K9	Apresiasi umum	Ulasan positif tanpa menyebut dimensi kualitas tertentu	bagus, membantu, terima kasih
K10	Ketidakpuasan umum	Ulasan negatif tanpa menyebut dimensi kualitas tertentu	jelek, mengecewakan, tidak jelas

Satu ulasan dapat memuat lebih dari satu karakteristik. Pada kasus demikian, seluruh karakteristik yang muncul dicatat, sehingga unit analisis adalah pernyataan kualitas, bukan ulasan sebagai satu kesatuan. Kategori K9 dan K10 hanya diberikan apabila tidak ada satu pun karakteristik spesifik yang terdeteksi.

Prosedur Koding dan Verifikasi Internal

Koding dilakukan dengan pendekatan berbantuan aturan. Peneliti menyusun daftar indikator untuk setiap karakteristik, menerapkannya pada korpus, lalu memeriksa ulasan yang belum terkode untuk mengidentifikasi ragam ungkapan yang terlewat. Proses tersebut diulang sampai tingkat keterkodean stabil. Pendekatan ini dipilih karena ukuran korpus tidak memungkinkan koding manual menyeluruh dalam rentang waktu penelitian, sekaligus menjamin penerapan kriteria yang konsisten pada seluruh ulasan.

Keabsahan hasil koding diuji melalui pemeriksaan manual berjenjang. Pada putaran pertama, sebanyak 127 ulasan ditarik sebagai sampel terstratifikasi menurut karakteristik yang diberikan, lalu setiap label diperiksa kesesuaiannya dengan isi ulasan. Putaran ini menghasilkan ketepatan keseluruhan 74,0 persen dan mengungkap tiga kesalahan sistematis pada daftar indikator. Pertama, ungkapan susah banget diperlakukan sebagai penanda kapabilitas interaksi, padahal pada korpus ini hampir selalu digunakan untuk melaporkan kegagalan masuk, yang semestinya masuk keandalan. Kedua, frasa hp lama diperlakukan sebagai penanda kompatibilitas perangkat, padahal yang dimaksud pengguna umumnya nomor telepon lama. Ketiga, frasa data pribadi diperlakukan sebagai penanda keamanan, padahal lebih sering muncul pada keluhan ketidakmampuan menyunting data. Daftar indikator kemudian diperbaiki untuk ketiga hal tersebut, ditambah perluasan penangkapan ragam bentuk kata galat.

Pada putaran kedua, sampel terstratifikasi baru sebanyak 110 ulasan ditarik dari hasil koding yang telah diperbaiki, lalu diperiksa dengan prosedur yang sama. Ketepatan keseluruhan meningkat menjadi 81,8 persen, dengan 90 label dinilai sesuai. Perbaikan indikator dihentikan pada putaran ini secara sengaja, karena penyesuaian lanjutan terhadap kesalahan yang tersisa pada sampel verifikasi berisiko menghasilkan aturan yang terlalu disesuaikan dengan sampel tersebut dan belum tentu berlaku pada korpus secara keseluruhan. Perlu dinyatakan secara terbuka bahwa pemeriksaan manual ini dilakukan oleh peneliti yang

sama dengan penyusun daftar indikator, sehingga bersifat verifikasi internal dan bukan uji kesepakatan antarkoder. Konsekuensinya, hasil yang dilaporkan adalah ketepatan koding, bukan koefisien Kappa Cohen sebagaimana lazim dipakai pada analisis isi dengan dua koder independen (Landis & Koch, 1977). Subsampel terstratifikasi untuk koding independen telah disiapkan sebagai bahan penelitian lanjutan. Independent coding by a second coder with Cohen's kappa, and an estimate of recall from a random sample of the 509 uncoded reviews and of the K9 and K10 reviews to count how many specific complaints were missed, are specified as the first priority of follow up work; until these are completed, the precision values reported here should be read as internal verification only.

Teknik Analisis Data

Analisis dilakukan dalam tiga tahap. Tahap pertama menghitung distribusi frekuensi dan proporsi setiap karakteristik kualitas pada seluruh korpus. Tahap kedua menabulasikan silang karakteristik kualitas dengan penilaian bintang untuk mengukur tingkat keparahan. Indeks keparahan dihitung sebagai proporsi ulasan berbintang satu dan dua pada setiap karakteristik terhadap seluruh ulasan yang memuat karakteristik tersebut. Tahap ketiga menyusun matriks prioritas perbaikan dengan memetakan frekuensi kemunculan pada sumbu mendatar dan indeks keparahan pada sumbu tegak, sehingga karakteristik dengan frekuensi tinggi dan keparahan tinggi menempati kuadran prioritas utama. Sebagai analisis tambahan, sebaran keluhan ditelusuri menurut bulan untuk melihat perubahan kualitas yang dipersepsikan sepanjang periode pengamatan. Star ratings reflect the overall assessment of a review and not a single characteristic; in the 157 multi label reviews the same rating is therefore counted for every characteristic assigned, and the severity index should be read with this assumption in mind. A sensitivity analysis restricted to single label reviews is recommended to test the robustness of the severity ranking.

HASIL DAN PEMBAHASAN

Profil Korpus

Korpus analitik berjumlah 2.389 ulasan dengan rata-rata penilaian 2,87 bintang. Sebaran penilaian menunjukkan pola yang sangat terpolarisasi. Sebanyak 1.088 ulasan atau 45,5 persen memberikan satu bintang dan 945 ulasan atau 39,6 persen memberikan lima bintang, sehingga kedua kutub tersebut mencakup 85,1 persen korpus. Penilaian tengah justru sangat langka, yaitu 137 ulasan berbintang dua, 101 ulasan berbintang tiga, dan 118 ulasan berbintang empat. Sebaran selengkapnya disajikan pada Tabel 2.

Tabel 2. Sebaran penilaian bintang pada korpus analitik

Penilaian	Jumlah ulasan	Persentase
1 bintang	1.088	45,5
2 bintang	137	5,7
3 bintang	101	4,2
4 bintang	118	4,9
5 bintang	945	39,6
Total	**2.389**	**100,0**

Pola terbelah ini merupakan temuan pertama yang perlu digarisbawahi. Rata-rata aritmetik sebesar 2,87 sesungguhnya tidak mewakili pengalaman siapa pun, sebab hampir tidak ada pengguna yang menilai aplikasi ini secara moderat. Yang terjadi adalah dua populasi yang berdampingan, yaitu kelompok yang berhasil menggunakan aplikasi dan menyatakan puas, serta kelompok yang gagal pada tahap tertentu dan menyatakan kecewa secara tegas. Pembacaan berbasis rata-rata akan menyembunyikan struktur ini, dan inilah salah satu alasan mengapa pelaporan rating tunggal kurang berguna bagi pengelola aplikasi.

Hasil Koding

Dari 2.389 ulasan, sebanyak 1.880 ulasan atau 78,7 persen berhasil dikode ke dalam sekurang kurangnya satu kategori, menghasilkan total 2.048 label. Sebanyak 157 ulasan memuat lebih dari satu karakteristik kualitas. Sisanya, yaitu 509 ulasan atau 21,3 persen, tidak terkode karena terlalu pendek untuk ditafsirkan atau karena isinya menyoroti penyelenggaraan layanan kesehatan di lapangan, bukan perangkat lunaknya. Sebanyak 1.010 ulasan memuat sekurang kurangnya satu keluhan terhadap dimensi kualitas tertentu.

Ketepatan koding per karakteristik pada putaran verifikasi kedua disajikan pada Tabel 3. Tiga karakteristik mencapai ketepatan penuh, yaitu keandalan, fleksibilitas, dan apresiasi umum, sedangkan keselamatan mendekatinya dengan 92,9 persen. Because each category contains only 3 to 14 verified labels, the per category precision is imprecise: K1, for example, reached 10 of 14 labels (71.4 percent) with a 95 percent Wilson confidence interval of 45.4 to 88.3 percent, and the overall precision of 81.8 percent has an interval of 73.6 to 87.9 percent (Tabel 3). Dua kategori dengan ketepatan terendah adalah kompatibilitas, yang hanya memuat tiga ulasan sehingga tidak layak ditafsirkan secara statistik, dan ketidakpuasan umum. Rendahnya ketepatan pada kategori terakhir perlu dibaca sesuai sifatnya. Kategori tersebut merupakan kategori sisa yang hanya diberikan ketika tidak ada karakteristik spesifik yang terdeteksi, sehingga kesalahannya mencerminkan keterbatasan daya tangkap kategori lain, bukan kesalahan kategori itu sendiri. Ulasan seperti laporan mengenai riwayat pemeriksaan yang muncul tanpa pernah diperiksa, yang semestinya masuk keselamatan, masih lolos ke kategori sisa ini.

Tabel 3. Ketepatan koding per karakteristik pada sampel verifikasi kedua

Kode	Karakteristik	Label diperiksa	Label sesuai	Ketepatan (persen)	Selang kepercayaan 95 persen (Wilson)
K5	Keandalan	14	14	100,0	78,5 sampai 100,0
K7	Fleksibilitas	10	10	100,0	72,2 sampai 100,0
K9	Apresiasi umum	10	10	100,0	72,2 sampai 100,0
K8	Keselamatan	14	13	92,9	68,5 sampai 98,7
K4	Kapabilitas interaksi	14	12	85,7	60,1 sampai 96,0
K2	Efisiensi kinerja	10	8	80,0	49,0 sampai 94,3
K1	Kesesuaian fungsional	14	10	71,4	45,4 sampai 88,3
K3	Kompatibilitas	3	2	66,7	20,8 sampai 93,9
K6	Keamanan	9	6	66,7	35,4 sampai 87,9

K10	Ketidakpuasan umum	12	5	41,7	19,3 sampai 68,0
	Total	**110**	**90**	**81,8**	**73,6 sampai 87,9**

Tabel 4. Frekuensi dan indeks keparahan karakteristik kualitas

Kode	Karakteristik	Frekuensi	Persen korpus	Rerata bintang	Ulasan bintang 1 dan 2	Indeks keparahan
K9	Apresiasi umum	870	36,4	4,68	43	4,9
K5	Keandalan	656	27,5	1,32	602	91,8
K1	Kesesuaian fungsional	207	8,7	1,68	171	82,6
K10	Ketidakpuasan umum	104	4,4	1,11	101	97,1
K2	Efisiensi kinerja	90	3,8	1,67	74	82,2
K4	Kapabilitas interaksi	70	2,9	1,54	59	84,3
K7	Fleksibilitas	24	1,0	1,50	21	87,5
K8	Keselamatan	15	0,6	1,47	13	86,7
K6	Keamanan	9	0,4	1,67	7	77,8
K3	Kompatibilitas	3	0,1	2,33	2	66,7

Tabel 5. Tabulasi silang karakteristik kualitas dengan penilaian bintang

Kode	Karakteristik	1	2	3	4	5	Total
K1	Kesesuaian fungsional	142	29	11	11	14	207
K2	Efisiensi kinerja	64	10	5	4	7	90
K3	Kompatibilitas	2	0	0	0	1	3
K4	Kapabilitas interaksi	55	4	5	0	6	70
K5	Keandalan	543	59	28	7	19	656
K6	Keamanan	7	0	1	0	1	9
K7	Fleksibilitas	19	2	1	0	2	24
K8	Keselamatan	12	1	0	2	0	15
K9	Apresiasi umum	37	6	21	69	737	870
K10	Ketidakpuasan umum	98	3	1	2	0	104

Keandalan sebagai Titik Kegagalan Utama

Keandalan merupakan karakteristik yang paling banyak dikeluhkan, muncul pada 656 ulasan atau 27,5 persen korpus. Jumlah tersebut lebih dari tiga kali lipat karakteristik terbanyak berikutnya. Rerata penilaian pada ulasan bermuatan keluhan keandalan hanya 1,32 bintang, dengan indeks keparahan 91,8 persen. Dengan kata lain, pengguna yang menemui masalah keandalan hampir selalu memberikan penilaian terendah, bukan penilaian menengah.

Penelusuran lebih jauh terhadap isi 656 ulasan tersebut mengungkap bahwa keluhan keandalan tidak tersebar merata, melainkan terkonsentrasi pada satu titik. Sebanyak 469 ulasan atau 71,5 persen menyinggung proses autentikasi, meliputi kegagalan masuk, PIN yang ditolak

meski dinyatakan benar, dan kode verifikasi yang tidak kunjung diterima. Sebanyak 155 ulasan atau 23,6 persen menyinggung validasi identitas berupa penolakan nomor induk kependudukan, nomor telepon, atau surel. Keterhentian aplikasi seperti galat dan penutupan paksa muncul pada jumlah yang sama, yaitu 155 ulasan atau 23,6 persen, sedangkan ketersediaan layanan seperti gangguan server disinggung 41 ulasan atau 6,2 persen. Sebaran subtema disajikan pada Tabel 6.

Tabel 6. Subtema dalam keluhan keandalan

Subtema	Jumlah ulasan	Persentase dari keluhan keandalan
Autentikasi (masuk, PIN, kode verifikasi, sandi)	469	71,5
Validasi identitas (NIK, nomor telepon, surel)	155	23,6
Keterhentian aplikasi (galat, penutupan paksa)	155	23,6
Ketersediaan layanan (server, gangguan, pemeliharaan)	41	6,2

Note. Percentages are calculated from the 656 reliability complaints. They sum to 124.9 percent because one review can contain more than one subtheme.

Temuan ini memiliki implikasi yang melampaui persoalan teknis biasa. Autentikasi adalah gerbang masuk, bukan fitur. Pengguna yang tertahan di gerbang tidak pernah sampai pada resume medis, riwayat imunisasi, hasil laboratorium, atau jadwal pemeriksaan yang justru menjadi alasan keberadaan aplikasi tersebut. Seluruh investasi pada interoperabilitas data kesehatan nasional menjadi tidak berarti bagi warga yang gagal masuk. Dalam kerangka informatika kesehatan, kegagalan ini setara dengan pintu rumah sakit yang terkunci, betapapun lengkapnya layanan di dalamnya.

Ulasan ulasan dalam korpus memperlihatkan pola yang konsisten, yaitu pengguna yang telah mencoba berulang kali, telah menempuh jalur pemulihan PIN, namun tetap menemui penolakan tanpa penjelasan yang dapat ditindaklanjuti. Sebagian pengguna juga melaporkan kebuntuan akibat ketidaksesuaian format, misalnya sistem yang menuntut nomor telepon dua belas digit sementara nomor yang dimiliki hanya sebelas digit. Pesan galat yang diterima umumnya bersifat umum, sehingga pengguna tidak memperoleh petunjuk mengenai tindakan korektif yang harus dilakukan. Dalam kerangka ISO/IEC 25010:2023, persoalan ini bersinggungan dengan subkarakteristik perlindungan pengguna dari kesalahan dan swadeskripsi pada kapabilitas interaksi, bukan hanya dengan keandalan semata. For example, users describe entering a PIN that they are certain is correct and still being rejected, or being asked for a twelve digit phone number when their number has eleven digits. Tabel 7 presents paraphrased examples of these patterns for each reliability subtheme and each quality characteristic, without any user identity.

Tabel 7. Paraphrased examples of review content by reliability subtheme and quality characteristic

Code	Characteristic or subtheme	Paraphrased example

K5	Authentication	The user enters a PIN that is certainly correct and the application still rejects it.
K5	Authentication	The verification code never arrives, even after PIN recovery has been attempted several times.
K5	Identity validation	The registration form rejects the national identity number, phone number, or email address.
K5	Identity validation	The form insists on a twelve digit phone number while the user's number has eleven digits.
K5	Application halts	The application shows an error message or closes by itself when it is opened.
K5	Service availability	The service cannot be reached because of server disruption or maintenance.
K1	Functional suitability	A certificate or document cannot be downloaded, or a city option is missing from the form.
K2	Performance efficiency	The application is slow, loads for a long time, and consumes mobile data.
K4	Interaction capability	The flow is confusing and the generic error message does not tell the user what to do.
K6	Security	The user is worried about data leakage and privacy.
K7	Flexibility	The application stops working after an update, or the installation fails.
K8	Safety	An examination history appears although the user never had that examination, or the data shown does not match what was entered.
K9	General appreciation	A short message of thanks saying that the application helps.
K10	General dissatisfaction	A short negative message that does not name the problem.

Note. The examples are paraphrased summaries of recurring patterns in the reviews and not verbatim quotations; they contain no user names or other identifying details. K3 is omitted because it contains only three reviews.

Karakteristik Berfrekuensi Rendah dengan Keparahan Tinggi

Keamanan dan keselamatan hanya muncul pada 9 dan 15 ulasan, yaitu 0,4 persen dan 0,6 persen korpus. Frekuensi serendah itu mudah ditafsirkan keliru sebagai tanda bahwa kedua dimensi tersebut tidak bermasalah. Penafsiran demikian perlu ditolak karena dua alasan.

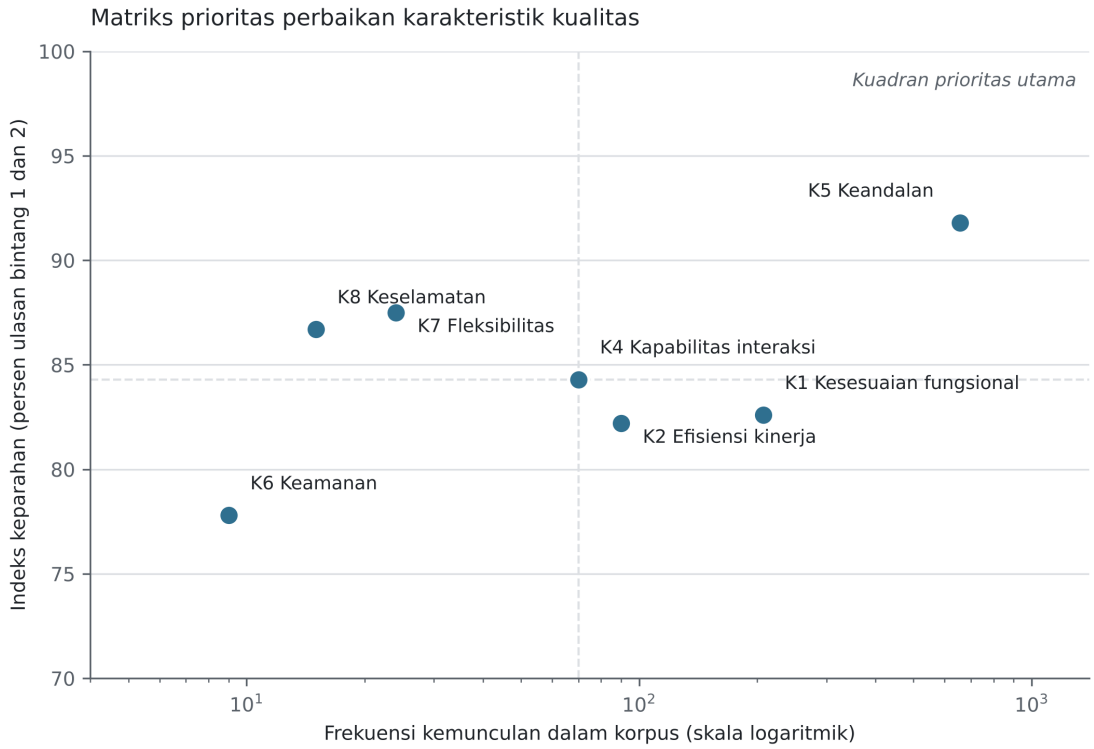
Pertama, indeks keparahan keduanya tergolong tinggi, yaitu 77,8 persen untuk keamanan dan 86,7 persen untuk keselamatan. Ketika dimensi ini disinggung, pengguna hampir selalu memberikan penilaian terendah. Kedua, dan ini lebih penting, kedua dimensi tersebut memiliki karakter yang secara inheren tidak terlaporkan. Pengguna awam tidak memiliki kosakata untuk menggambarkan kelemahan keamanan, dan hanya dapat menyadari kekeliruan data medis apabila kebetulan memeriksa catatan yang bersangkutan. Ulasan dalam korpus ini memuat laporan mengenai hasil pemeriksaan yang muncul padahal pengguna belum pernah menjalani pemeriksaan, serta data yang ditampilkan tidak sesuai dengan yang diisikan. Laporan semacam itu, meskipun sedikit, menandakan persoalan integritas data yang konsekuensinya bersifat klinis.

Dengan demikian, frekuensi rendah pada kedua karakteristik ini sebaiknya dibaca sebagai keterbatasan kanal pelaporan, bukan sebagai bukti ketiadaan masalah. Pengelola aplikasi kesehatan tidak dapat menggantungkan deteksi persoalan integritas data pada ulasan sukarela, melainkan harus menjalankan audit aktif terhadap kecocokan data antara catatan fasilitas dan catatan yang ditampilkan kepada pengguna.

From a policy perspective, these findings connect directly to Law Number 27 of 2022 on Personal Data Protection (Republik Indonesia, 2022), which classifies health data as specific personal data and requires data controllers to notify data subjects in writing when personal data protection fails. A medical record of another person appearing in a user's account is therefore best treated as a potential personal data incident and not only as a software defect. We recommend that the application manager provides a dedicated in app channel for reporting medical record integrity problems, with a documented triage procedure and response time, records each report in an incident register that links it to the record held by the originating health facility, and combines this channel with the active data matching audit described above so that integrity problems are detected without waiting for voluntary reports.

Matriks Prioritas Perbaikan

Pemetaan frekuensi terhadap indeks keparahan menghasilkan matriks prioritas sebagaimana disajikan pada Gambar 1. Keandalan menempati kuadran prioritas utama secara tunggal, dengan frekuensi tertinggi sekaligus keparahan tertinggi di antara seluruh karakteristik aktif. Kesesuaian fungsional, kapabilitas interaksi, dan efisiensi kinerja menyusul pada kuadran frekuensi tinggi dengan keparahan di sekitar median. Keselamatan dan fleksibilitas berada pada kuadran keparahan tinggi dengan frekuensi rendah, sehingga menuntut pemantauan aktif meskipun tidak mendesak secara volume. Kompatibilitas tidak diikutsertakan dalam matriks karena hanya memuat tiga ulasan, jumlah yang terlalu kecil untuk menghasilkan indeks keparahan yang bermakna.



Gambar 1. Matriks prioritas perbaikan karakteristik kualitas berdasarkan frekuensi kemunculan dan indeks keparahan

Gambar 1 dibaca dengan menempatkan setiap karakteristik pada dua sumbu sekaligus. Sumbu mendatar menyatakan frekuensi kemunculan karakteristik dalam korpus dan disajikan dalam skala logaritmik, karena rentang frekuensi antarkarakteristik sangat lebar, yaitu dari 9 sampai 656 ulasan. Tanpa skala logaritmik, seluruh karakteristik berfrekuensi rendah akan berimpit di tepi kiri dan tidak terbaca. Sumbu tegak menyatakan indeks keparahan, yaitu persentase ulasan berbintang satu dan dua pada karakteristik tersebut. Garis putus putus mendatar dan tegak merupakan nilai median kedua sumbu, yang membagi bidang menjadi empat kuadran.

Pembacaan kuadran dilakukan sebagai berikut. Kuadran kanan atas, yaitu frekuensi tinggi dan keparahan tinggi, memuat masalah yang sering terjadi sekaligus berdampak berat, sehingga menjadi prioritas utama. Kuadran kanan bawah memuat masalah yang sering terjadi namun relatif dapat ditoleransi pengguna. Kuadran kiri atas memuat masalah yang jarang dilaporkan tetapi berdampak berat, sehingga menuntut pemantauan aktif meskipun volumenya kecil. Kuadran kiri bawah memuat masalah yang jarang sekaligus ringan, sehingga dapat ditangani paling akhir.

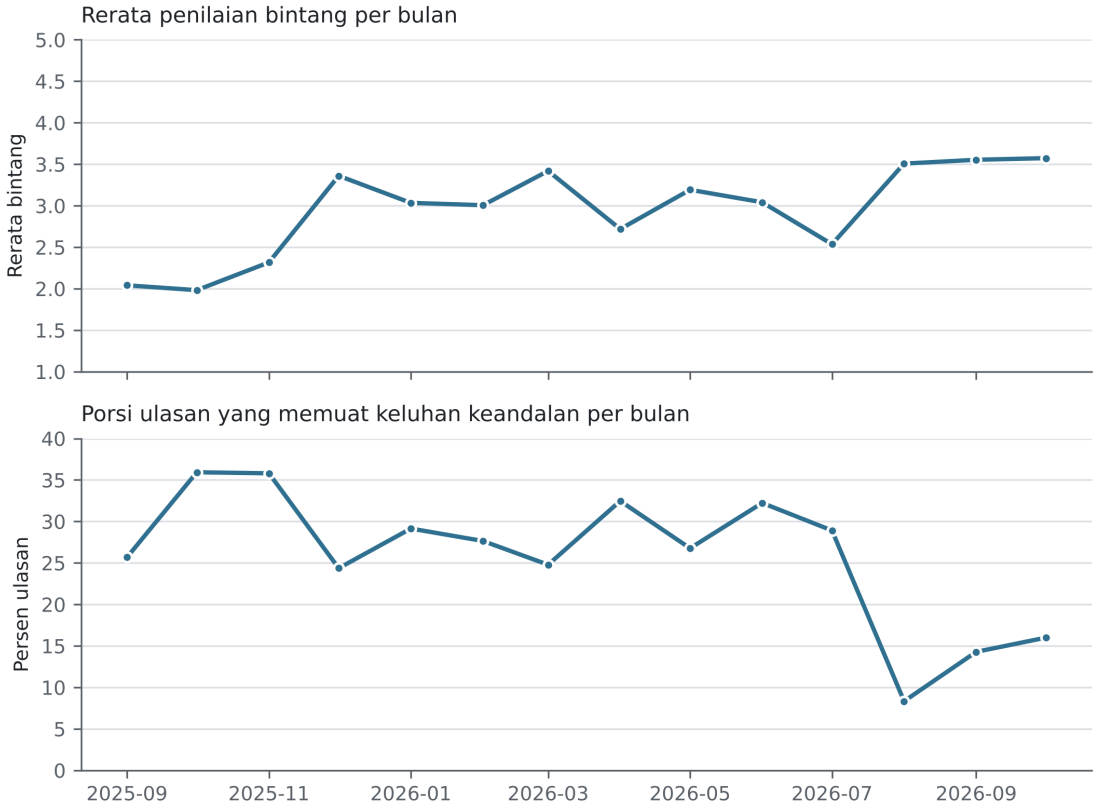
Dalam kasus SATUSEHAT Mobile, keandalan berdiri sendiri di kuadran kanan atas, terpisah jauh dari karakteristik lain pada kedua sumbu. Posisi terisolasi inilah yang menjadi dasar penetapan autentikasi sebagai prioritas pertama. Kesesuaian fungsional, kapabilitas interaksi, dan efisiensi kinerja berada di sisi kanan dengan keparahan di sekitar median. Keselamatan dan fleksibilitas menempati kuadran kiri atas, yang sesuai dengan argumen sebelumnya bahwa frekuensi rendah pada dimensi ini mencerminkan keterbatasan kanal

pelaporan, bukan ketiadaan masalah. Keamanan berada paling kiri dengan frekuensi terendah di antara karakteristik yang dipetakan. Perlu dicatat bahwa karakteristik dengan frekuensi sangat kecil, yaitu keamanan dengan 9 ulasan dan keselamatan dengan 15 ulasan, memiliki indeks keparahan yang rentan berubah oleh sedikit perubahan data, sehingga posisi tegaknya pada gambar ini sebaiknya dibaca sebagai indikasi, bukan sebagai ukuran yang presisi. Kompatibilitas dengan tiga ulasan tidak dipetakan sama sekali atas alasan yang sama.

Berdasarkan matriks tersebut, prioritas perbaikan dapat disusun dalam tiga tingkat. Prioritas pertama adalah perbaikan jalur autentikasi, mencakup penyediaan alternatif masuk yang tidak bergantung pada satu kanal, perbaikan mekanisme pemulihan PIN, toleransi terhadap ragam format nomor telepon, dan penggantian pesan galat umum dengan pesan yang menyebutkan tindakan korektif. Prioritas kedua adalah pemulihan kelengkapan fungsional, terutama pada pengunduhan sertifikat dan dokumen, kelengkapan data wilayah dan fasilitas pada formulir pendaftaran, serta sinkronisasi riwayat pemeriksaan. Prioritas ketiga adalah audit integritas data secara berkala, yang tidak boleh menunggu laporan pengguna.

Perubahan sepanjang Periode Pengamatan

Analisis menurut bulan memperlihatkan perubahan yang searah pada dua indikator, sebagaimana disajikan pada Gambar 2. Rerata penilaian bulanan meningkat dari 2,04 bintang pada September 2025 menjadi 3,57 bintang pada Oktober 2026. Pada saat yang sama, proporsi ulasan yang memuat keluhan keandalan menurun dari 35,9 persen pada Oktober 2025 menjadi 8,3 persen pada Agustus 2026, 14,3 persen pada September 2026, dan 16,0 persen pada Oktober 2026. Both series are reported from September 2025, a partial month of 12 days with 74 reviews, until October 2026, a partial month of 20 days, so that the rating and the reliability complaint series cover the same period; the number of reviews per month is listed in Tabel 8.



Gambar 2. Rerata penilaian bintang dan proporsi ulasan bermuatan keluhan keandalan per bulan, September 2025 sampai Oktober 2026

Gambar 2 disajikan dalam dua panel terpisah yang berbagi sumbu waktu, bukan dalam satu grafik bersumbu ganda. Pemisahan ini disengaja karena kedua indikator memiliki satuan yang berbeda, yaitu skala bintang satu sampai lima pada panel atas dan persentase ulasan pada panel bawah. Menempatkan keduanya pada satu bidang dengan dua sumbu tegak akan menciptakan kesan keterkaitan yang sesungguhnya ditentukan oleh pemilihan rentang sumbu, bukan oleh data.

Panel atas menunjukkan rerata penilaian bulanan yang bergerak naik dari 2,04 bintang pada September 2025 menuju kisaran 3,5 bintang pada tiga bulan terakhir pengamatan. Kenaikan ini tidak berlangsung mulus, melainkan disertai penurunan pada April 2026 dan Juli 2026. Panel bawah menunjukkan proporsi ulasan yang memuat keluhan keandalan, yang bertahan pada kisaran 25 sampai 36 persen sepanjang periode September 2025 hingga Juli 2026, lalu turun tajam menjadi 8,3 persen pada Agustus 2026 dan bertahan di bawah 17 persen setelahnya.

Tabel 8. Monthly number of reviews, mean star rating, and share of reviews with reliability complaints

Month	Reviews (n)	Mean rating	Reliability complaints (percent)
September 2025 (12 days)	74	2.04	[fill]
October 2025	[fill]	[fill]	35.9
November 2025	[fill]	[fill]	[fill]
December 2025	[fill]	[fill]	[fill]

January 2026	[fill]	[fill]	[fill]
February 2026	[fill]	[fill]	[fill]
March 2026	[fill]	[fill]	[fill]
April 2026	[fill]	[fill]	[fill]
May 2026	[fill]	[fill]	[fill]
June 2026	[fill]	[fill]	[fill]
July 2026	377	[fill]	[fill]
August 2026	[fill]	[fill]	8.3
September 2026	[fill]	[fill]	14.3
October 2026 (20 days)	[fill]	3.57	16.0
Total	**2,389**	**2.87**	**27.5**

Note. Cells in brackets are to be completed by the author from the coded dataset.

Dua panel tersebut bergerak berlawanan arah secara umum meskipun tidak monoton, yang memperkuat dugaan bahwa penilaian rendah pada periode awal memang digerakkan oleh persoalan keandalan. Namun penafsiran ini tetap perlu disampaikan dengan hati-hati. Jumlah ulasan per bulan tidak seragam, berkisar dari 74 ulasan pada September 2025 sampai 377 ulasan pada Juli 2026, sehingga proporsi pada bulan dengan ulasan sedikit lebih rentan berfluktuasi. Selain itu, bulan terakhir pengamatan hanya mencakup dua puluh hari pertama Oktober 2026. Lonjakan jumlah ulasan pada Juli 2026 bertepatan dengan periode ramai Program Cek Kesehatan Gratis, yang berarti komposisi penulis ulasan pada bulan tersebut dapat berbeda dari bulan lainnya. A formal test should accompany this descriptive pattern, for example a chi square test for trend of the monthly proportion of reliability complaints, or a comparison of proportions before and after the release of version 8.8 [test statistics and p values to be inserted from the dataset]; until then the pattern is presented as suggestive and not as confirmed.

Perubahan ini bertepatan dengan rangkaian pembaruan aplikasi pada rentang versi 8.8 sampai 8.10, yang menurut catatan rilis mencakup penyegaran tampilan beranda dan penambahan verifikasi mandiri di dalam aplikasi. Penafsiran bahwa pembaruan tersebut memperbaiki keandalan tampak masuk akal, namun harus disampaikan secara berhati-hati. Penjelasan alternatif tetap terbuka, misalnya perubahan komposisi pengguna yang menulis ulasan setelah gelombang pendaftaran Program Cek Kesehatan Gratis mereda, atau perbedaan volume ulasan antarbulan yang memengaruhi kestabilan proporsi. Desain penelitian ini bersifat observasional sehingga tidak dapat menetapkan hubungan sebab akibat antara pembaruan versi dan penurunan keluhan.

Perbandingan dengan Penelitian Terdahulu

Satu temuan pendukung yang patut dicatat datang dari Pratama (2025), yang menganalisis 700 ulasan SATUSEHAT Mobile menggunakan regresi logistik dan menemukan 87 persen ulasan bernada negatif. Yang lebih penting dari angka tersebut adalah kata kata yang paling sering muncul dalam temuannya, yaitu verifikasi, login, bug, dan stuck. Keempatnya berada persis pada wilayah yang dalam penelitian ini teridentifikasi sebagai subtema autentikasi dan keterhentian aplikasi. Dua penelitian yang berbeda korpus, berbeda periode, dan berbeda teknik analisis sampai pada titik kegagalan yang sama, sehingga temuan mengenai

dominasi persoalan autentikasi tidak dapat dianggap sebagai artefak dari satu kali pengambilan data.

Perbandingan dengan penelitian berbasis survei justru memperlihatkan perbedaan yang menarik. Sidabutar dan Ichwani (2024) mengevaluasi pengalaman pengguna SATUSEHAT Mobile dengan kerangka HEART terhadap 100 responden dan melaporkan seluruh variabel berada pada kategori sangat tinggi, dengan rentang 83,50 persen sampai 84,25 persen. Hasil tersebut nyaris berlawanan dengan gambaran yang muncul dari ulasan toko aplikasi. Perbedaan ini kemungkinan besar bersumber dari perbedaan populasi yang terjangkau masing masing metode. Survei menjangkau responden yang berhasil memasang dan menggunakan aplikasi, sehingga secara struktural mengeluarkan pengguna yang gagal pada tahap autentikasi, yaitu kelompok yang justru paling banyak bersuara di kolom ulasan. Implikasinya, evaluasi aplikasi kesehatan publik sebaiknya tidak bersandar pada satu jenis sumber data saja, karena survei dan ulasan toko aplikasi menangkap dua sisi populasi yang berbeda. This argument agrees with research on self selection in online reviews, which shows that reviewers are not a random sample of users because the decision to write a review depends on the experience of the product (Hu et al., 2017); the same logic implies that app store reviews over represent users who failed or who were strongly satisfied, whereas survey samples over represent users who succeeded.

Temuan penelitian ini sejalan dengan Imanuddin, Adi, dan Gernowo (2023) yang melaporkan dominasi sentimen negatif sebesar 73,4 persen, serta dengan Dewi, Dharmendra, dan Setiasih (2023) yang menemukan ketidakseimbangan kelas akibat menumpuknya ulasan negatif. Perbedaannya terletak pada tingkat ketajaman hasil. Penelitian terdahulu menjawab pertanyaan seberapa banyak pengguna tidak puas, sedangkan penelitian ini menjawab pertanyaan pada dimensi mana kualitas itu gagal dan seberapa parah akibatnya. Informasi bahwa 71,5 persen keluhan keandalan terpusat pada autentikasi jauh lebih dapat ditindaklanjuti oleh tim pengembang dibandingkan informasi bahwa mayoritas ulasan bernada negatif.

Keterbatasan Penelitian

Penelitian ini memiliki sejumlah keterbatasan yang perlu dinyatakan secara terbuka. Pertama, korpus hanya mencakup ulasan berbahasa Indonesia pada platform Android, sehingga pengalaman pengguna iOS tidak terwakili. Kedua, ulasan toko aplikasi mengandung bias partisipasi, karena pengguna cenderung menulis ketika mengalami kekecewaan atau kepuasan yang menonjol, bukan ketika mengalami pengalaman biasa saja. Ketiga, sebanyak 21,3 persen ulasan tidak terkode, sebagian karena terlalu pendek dan sebagian karena menyoroti penyelenggaraan layanan di fasilitas kesehatan, bukan perangkat lunaknya. Keempat, karakteristik kemampurawatan tidak dapat dievaluasi dari perspektif pengguna akhir. Kelima, koding berbantuan aturan memiliki keterbatasan daya tangkap terhadap ragam ejaan dan ungkapan yang tidak terdaftar sebagai indikator, serta terhadap ungkapan ironis. Ketepatan koding yang terukur sebesar 81,8 persen berarti kurang lebih satu dari lima label mengandung kekeliruan, dengan konsentrasi kekeliruan pada kategori sisa dan pada karakteristik berfrekuensi kecil. Verifikasi juga dilakukan oleh peneliti yang menyusun aturan koding, sehingga tidak memenuhi syarat independensi antarkoder. Keenam, korpus merupakan 2.400 ulasan terbaru, bukan sampel acak dari seluruh populasi 1,12 juta ulasan, sehingga generalisasi ke seluruh periode penggunaan aplikasi perlu dilakukan secara berhati-hati. Seventh, star

ratings are overall assessments, so in multi label reviews the same rating is counted for each characteristic, and a sensitivity analysis restricted to single label reviews has not yet been performed. Eighth, recall was not measured, so the number of specific complaints that fall into the 509 uncoded reviews or into the residual category K10 is unknown. Ninth, the participation bias noted above is a form of self selection (Hu et al., 2017), and the 2,400 reviews retrieved are not a probability sample of the ratings shown on the app page.

KESIMPULAN

Penelitian ini memetakan 2.389 ulasan pengguna SATUSEHAT Mobile ke dalam karakteristik kualitas produk ISO/IEC 25010:2023 dan menemukan tiga hal pokok. Pertama, penilaian pengguna terpolarisasi secara ekstrem, dengan 85,1 persen ulasan berada pada kutub satu bintang atau lima bintang, sehingga pelaporan berbasis rata-rata menyembunyikan struktur opini yang sebenarnya. Kedua, keandalan merupakan karakteristik yang paling sering dikeluhkan sekaligus paling parah dampaknya, muncul pada 27,5 persen korpus dengan indeks keparahan 91,8 persen, dan 71,5 persen keluhan tersebut terpusat pada proses autentikasi. Ketiga, keamanan dan keselamatan jarang disebut namun memiliki keparahan tinggi, sehingga memerlukan audit aktif alih alih mengandalkan laporan sukarela pengguna. These findings depend on rule assisted coding that has so far been verified only internally (81.8 percent precision), and they should be confirmed through independent coding by a second coder.

Implikasi praktisnya jelas. Perbaikan aplikasi kesehatan nasional sebaiknya dimulai dari gerbang autentikasi, karena kegagalan di titik tersebut membatalkan seluruh nilai yang dibangun di belakangnya. Implikasi metodologisnya, ulasan publik di toko aplikasi terbukti dapat diperlakukan sebagai instrumen evaluasi teknologi kesehatan yang murah, berkelanjutan, dan berskala besar, asalkan dianalisis dengan kerangka kualitas yang baku dan tidak berhenti pada polaritas sentimen. Skema koding yang dihasilkan penelitian ini dapat diterapkan ulang untuk memantau aplikasi kesehatan lain maupun untuk memantau SATUSEHAT Mobile secara berkala sebagai indikator kinerja pengembangan produk.

Penelitian lanjutan disarankan menempuh dua arah. Arah pertama adalah pengujian skema koding ini pada korpus ulasan platform iOS dan pada aplikasi kesehatan pemerintah lainnya untuk menilai keteralihannya. Arah kedua adalah penggabungan analisis ulasan dengan data telemetri aplikasi, sehingga keluhan yang dilaporkan pengguna dapat ditautkan dengan kejadian kegagalan yang terekam sistem.

PENDANAAN

Penulis menyatakan bahwa tidak ada dukungan finansial yang diterima untuk pelaksanaan penelitian ini dan/atau penerbitan artikel ini.

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa penelitian ini dilakukan tanpa adanya hubungan komersial atau finansial yang dapat ditafsirkan sebagai potensi konflik kepentingan.

KONTRIBUSI PENULIS

The author carried out the conceptualization, data collection, data analysis, and writing of this manuscript.

PERNYATAAN PENGGUNAAN KECERDASAN ARTIFISIAL GENERATIF

Penulis menyatakan bahwa kecerdasan artifisial (AI) generatif digunakan dalam penyusunan manuskrip ini. Penulis mengakui penggunaan AI generatif untuk membantu meningkatkan kejelasan bahasa dan memperbaiki struktur tata bahasa manuskrip. Penggunaan tersebut dilakukan dengan mematuhi pedoman etika yang berlaku mengenai penggunaan dan pengungkapan AI generatif dalam penelitian ilmiah. The author made the substantial intellectual contribution to this work. Seluruh isi manuskrip telah ditelaah secara cermat untuk memastikan keakuratannya, dan penulis bertanggung jawab sepenuhnya atas integritas serta isi akhir manuskrip

REFERENSI

Anggiyanti, J. N., & Papilaya, F. S. (2026). Implementasi pemodelan topik BERTopic dan analisis sentimen BERT terhadap ulasan aplikasi Satusehat berbasis natural language processing lanjutan. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 11*(1), 165–177. https://doi.org/10.29100/jipi.v11i1.7633

Dąbrowski, J., Letier, E., Perini, A., & Susi, A. (2022). Analysing app reviews for software engineering: A systematic literature review. *Empirical Software Engineering, 27*, Artikel 43. https://doi.org/10.1007/s10664-021-10065-7

Dewi, I. A. M. C., Dharmendra, I. K., & Setiasih, N. W. (2023). Analisis sentimen review aplikasi Satu Sehat Mobile menggunakan model sampling Tomek Links. *Jurnal Teknologi Informasi dan Komputer, 9*(5). https://doi.org/10.36002/jutik.v9i5.2644

Garousi, V., Cutting, D., & Felderer, M. (2022). Mining user reviews of COVID contact-tracing apps: An exploratory analysis of nine European apps. *Journal of Systems and Software, 184*, 111136. https://doi.org/10.1016/j.jss.2021.111136

Genc-Nayebi, N., & Abran, A. (2017). A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software, 125*, 207–219. https://doi.org/10.1016/j.jss.2016.11.027

Haggag, O., Grundy, J., Abdelrazek, M., & Haggag, S. (2022). A large scale analysis of mHealth app user reviews. *Empirical Software Engineering, 27*(7), Artikel 196. https://doi.org/10.1007/s10664-022-10222-6

Haoues, M., Mokni, R., & Sellami, A. (2023). Machine learning for mHealth apps quality evaluation. *Software Quality Journal, 31*(4), 1179–1209. https://doi.org/10.1007/s11219-023-09630-8

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Hu, N., Pavlou, P. A., & Zhang, J. (2017). On self-selection biases in online product reviews. *MIS Quarterly, 41*(2), 449–471. https://aisel.aisnet.org/misq/vol41/iss2/8

Imanuddin, S. H., Adi, K., & Gernowo, R. (2023). Sentiment analysis on Satusehat application using support vector machine method. *Journal of Electronics, Electromedical Engineering, and Medical Informatics, 5*(3), 143–149. https://doi.org/10.35882/jeemi.v5i3.304

International Organization for Standardization. (2023). *ISO/IEC 25010:2023 Systems and software engineering, Systems and software Quality Requirements and Evaluation (SQuaRE), Product quality model*. ISO.

Kementerian Kesehatan Republik Indonesia. (2022). *Peraturan Menteri Kesehatan Nomor 24 Tahun 2022 tentang Rekam Medis*. Kemenkes RI.

Kementerian Kesehatan Republik Indonesia. (2025a). *Langkah awal program pemeriksaan kesehatan gratis, unduh SATUSEHAT Mobile*. https://kemkes.go.id/id/langkah-awal-program-pemeriksaan-kesehatan-gratis-unduh-satusehat-mobile

Kementerian Kesehatan Republik Indonesia. (2025b). *Kemenkes integrasikan ratusan aplikasi menuju era SATUSEHAT*. Badan Kebijakan Pembangunan Kesehatan. https://www.badankebijakan.kemkes.go.id/kemenkes-integrasikan-ratusan-aplikasi-menuju-era-satusehat/

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Martin, W., Sarro, F., Jia, Y., Zhang, Y., & Harman, M. (2017). A survey of app store analysis for software engineering. *IEEE Transactions on Software Engineering, 43*(9), 817–847. https://doi.org/10.1109/TSE.2016.2630689

Ministry of Health Republic of Indonesia. (2026). *SATUSEHAT Mobile* [Aplikasi seluler]. Google Play Store. https://play.google.com/store/apps/details?id=com.telkom.tracencare

Nurfikri, A. (2022). Sentiment analysis telemedicine apps reviews using NVIVO. *Proceedings, 83*(1), Artikel 4. https://doi.org/10.3390/proceedings2022083004

Pratama, F. (2025). Analisis kepuasan pengguna SATUSEHAT Mobile berdasarkan data ulasan di Google Play Store menggunakan logistic regression. *Jurnal Advance Research Informatika, 4*(1), 11–17. https://doi.org/10.24929/jars.v4i1.4369

Republik Indonesia. (2022). *Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi*. https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022

Sidabutar, S., & Ichwani, A. (2024). Evaluasi user experience aplikasi mobile SATUSEHAT menggunakan HEART metrics. *Jurnal Pepadun, 5*(1), 72–82. https://doi.org/10.23960/pepadun.v5i1.184

Sönmez, S., & Kekül, H. (2024). Multi-class classification of ISO/IEC 25010 software quality metrics using user feedback. *Sivas Cumhuriyet Üniversitesi Mühendislik Fakültesi Dergisi, 2*(2), 214–222. https://doi.org/10.66248/cumfad.1588125

Tarwoto, Nugroho, R., Azka, N., & Graha, W. S. R. (2025). Analisis sentimen ulasan aplikasi Mobile JKN di Google PlayStore menggunakan IndoBERT. *JTIK (Jurnal Teknologi Informasi dan Komunikasi), 9*(2). https://doi.org/10.35870/jtik.v9i2.3340

World Health Organization. (2021). *Global strategy on digital health 2020-2025*. World Health Organization. https://www.who.int/publications/i/item/9789240020924