Journal of Authentic Research
https://journal-center.litpam.com/index.php/jar/index

November 2026 Vol. 5, No. 4
e-ISSN: 2828-3724
pp. 8095-8120

Responsivitas Layanan Publik Digital M-Paspor terhadap Keluhan Pengguna: Analisis Ulasan *Play Store* dengan Perspektif *Agile Government*

[1*]Izzatul Ulya Affifah, [2]Iip Permana

[1] Fakultas Ilmu Sosial, Universitas Negeri Padang, Jalan Prof. Dr. Hamka, Air Tawar, Kota Padang, Sumatera Barat, Indonesia

*Corresponding Author e-mail: affifahulya@gmail.com

Received: September 2026; Revised: October 2026; Published: November 2026

Abstrak

Penelitian ini mengkaji keterkaitan antara suara pengguna dan respons kelembagaan pada layanan publik digital M-Paspor milik Direktorat Jenderal Imigrasi (Ditjenim). Kajian terdahulu pada M-Paspor menilai kepuasan pengguna, tetapi belum menelusuri apakah keluhan yang terakumulasi memperoleh respons yang terdokumentasi. Analisis isi berbantuan komputasi dilakukan terhadap 17.411 ulasan *Google Play Store* dari 18.625 ulasan yang dikumpulkan. Ulasan diklasifikasikan ke dalam empat dimensi *Information Systems Success Model* (ISSM) menggunakan NusaBERT, topiknya dipetakan dengan BERTopic, lalu dibandingkan dengan jejak digital ditjenim dievaluasi melalui kerangka *agile evaluation*. Sentimen negatif terutama muncul pada dimensi *system quality*, *information quality*, dan *user satisfaction*. Keluhan spesifik mencakup 10,2% dari korpus, khususnya terkait kegagalan OTP, kuota yang telah penuh, masalah pembayaran, *timeout*, serta pemeliharaan sistem yang tidak disertai kepastian waktu penyelesaian. Respons yang terlihat secara publik masih bersifat parsial dan reaktif; hanya perbaikan kamera yang secara eksplisit menjawab keluhan pengguna. Selain itu, jeda antara ulasan dan respons berkisar dari dua hingga 1.287 hari, sementara balasan pengembang hanya terdapat pada 14 dari 17.411 ulasan, atau sekitar 0,08%. Kontribusi penelitian ini adalah mengungkap adanya ketidakselarasan antara banyaknya umpan balik pengguna dan respons kelembagaan yang terdokumentasi secara publik. Temuan ini juga membedakan antara ulasan pengguna sebagai potensi sumber sensing dan sensing yang benar-benar dijalankan oleh instansi. Mengingat bukti respons hanya diperoleh dari sumber publik, hasil penelitian ini tidak dapat digunakan untuk menyimpulkan bahwa tidak ada tindak lanjut internal. Demikian pula, sentimen pengulas tidak serta-merta mewakili persepsi seluruh pengguna.

Kata kunci: M-Paspor, layanan publik digital, ISSM, *agile government*, ulasan *Google Play Store*.

How to Cite: Affifah, I U., & Permana, I (2026). Responsivitas Layanan Publik Digital M-Paspor terhadap Keluhan Pengguna: Analisis Ulasan *Play Store* dengan Perspektif *Agile Government*. *Journal of Authentic Research, 5*(4), 8095–8120. https://doi.org/10.36312/jar.v5i4.7082

https://doi.org/10.36312/jar.v5i4.7082

Copyright© 2026, Affifah et al.
This is an open-access article under the CC-BY-SA License.

PENDAHULUAN

Transformasi digital telah mengubah pola penyelenggaraan pelayanan publik oleh pemerintah. Layanan berbasis digital memungkinkan masyarakat mengakses informasi dan layanan secara lebih mudah, cepat, dan efisien, sehingga berpotensi meningkatkan kualitas serta kepuasan pelayanan (Wiranti & Frinaldi, 2023). Dalam kerangka digital government, teknologi digital dipandang sebagai bagian dari strategi modernisasi birokrasi untuk menciptakan nilai publik (OECD, 2014). Oleh karena itu, keberhasilannya tidak hanya diukur dari ketersediaan aplikasi, tetapi juga dari kapasitas tata kelola yang mendukungnya (Harefa & Zebua, 2026; Kasim et al., 2025).

Di Indonesia, arah kebijakan ini diformalkan melalui Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik (SPBE) dan Peraturan Menteri PANRB Nomor 13 Tahun 2017 tentang Sistem Informasi Pelayanan Publik Nasional (Kementerian PANRB, 2017; Republik Indonesia, 2018).

Salah satu bentuk digitalisasi tersebut adalah aplikasi M-Paspor yang dikembangkan oleh Direktorat Jenderal Imigrasi (Ditjenim). Aplikasi ini diluncurkan secara nasional pada 27 Januari 2022 sebagai penyempurnaan dari Aplikasi Pendaftaran Antrean Paspor *Online* (APAPO), yang sebelumnya hanya berfungsi untuk pengambilan nomor antrean (Direktorat Jenderal Imigrasi, 2022; Prameswari et al., 2023). Melalui M-Paspor, pemohon dapat mengajukan permohonan, mengunggah dokumen, membayar PNBP, dan memilih jadwal kedatangan secara mandiri, sehingga di kantor imigrasi pemohon cukup menjalani verifikasi dokumen asli, wawancara, dan perekaman *biometric* (Direktorat Jenderal Imigrasi, 2022). Kebutuhan layanan ini cukup besar; penerbitan paspor meningkat dari 1.018.923 pada 2021 menjadi 3.878.904 pada 2022, mencapai puncak 5.053.315 pada 2023, kemudian 4.838.581 pada 2024 dan 4.033.676 pada 2025 (Direktorat Jenderal Imigrasi, 2026). Sejumlah studi melaporkan bahwa M-Paspor mampu mempercepat proses pengajuan dan mengurangi antrean (Anisa & Dewi, 2024).

Namun, ketersediaan layanan digital tidak otomatis menjamin kepuasan pengguna (Bahri et al., 2025). Kepuasan merupakan indikator utama keberhasilan layanan digital pemerintah karena mencerminkan persepsi warga terhadap kinerja aktual aplikasi (Bahri et al., 2025; Nainggolan & Aqil, 2023). Pada kasus M-Paspor, rating di *Google Play Store* tercatat 1,7 dari 5 pada 2023, 2,2 pada 2024, 2,5 pada 2025, dan 3,1 pada 2026 (*Google Play Store*, 2026). Tren tersebut menunjukkan perbaikan, tetapi nilainya masih tergolong sedang. Penelitian terdahulu melaporkan berbagai kendala, seperti galat sistem, kegagalan unggah dokumen, proses yang lambat, dan kesalahan yang berdampak finansial bagi pengguna (Pamadi et al., 2022; Ramadhani et al., 2025). Persoalan tersebut tidak semata berkaitan dengan kualitas aplikasi, tetapi juga dengan cara penyelenggara menindaklanjuti keluhan yang muncul darinya. Kualitas aplikasi menggambarkan kondisi sistem pada suatu waktu, sedangkan responsivitas menggambarkan tindak lanjut penyelenggara atas umpan balik pengguna; keduanya dapat tidak searah, misalnya ketika aplikasi diperbarui tanpa menjawab keluhan yang paling sering disampaikan. Perbedaan ini memiliki empat konsekuensi. Secara administratif, kegagalan pembayaran atau pengunggahan dokumen dapat menimbulkan kerugian finansial dan menunda pengurusan dokumen yang berbatas waktu (Pamadi et al., 2022; Ramadhani et al., 2025). Dari sisi akses, keterbatasan akses fitur dan respons layanan melalui kanal resmi (Ramadhani et al., 2025), dan kualitas respons layanan terbukti berhubungan dengan kepuasan pengguna M-Paspor (Bahri et al., 2025). Dari sisi transparansi dan akuntabilitas, digitalisasi layanan keimigrasian diharapkan mewujudkan proses yang lebih terbuka dan akuntabel (Prameswari et al., 2023), sehingga kejelasan tindak lanjut atas keluhan menjadi bagian dari tanggung jawab penyelenggara dalam memberikan pertanggungjawaban kepada warga. Umpan balik warga pada platform digital juga dipandang sebagai instrument partisipasi dan indicator legitimasi pemerintah, sehingga tindak lanjut atas keluhan menjadi bagian dari akuntabilitas penyelenggaraan layanan (Bertot et al., 2010; Yulanda & Frinaldi, 2023)

Meskipun demikian, topik mana yang paling dominan dan bagaimana penyelenggara menanggapinya belum dikaji secara sistematis. M-Paspor dipilih sebagai objek penelitian karena tiga alasan. Pertama, layanan ini bersifat nasional dan melayani permintaan yang besar, sehingga kegagalan layanan berdampak pada banyak warga. Kedua, aplikasi ini memindahkan sebagian besar tahapan permohonan ke ranah digital sehingga kualitas aplikasi menentukan akses warga terhadap layanan keimigrasian (Direktorat Jenderal Imigrasi, 2022). Ketiga, ulasan terakumulasi sejak aplikasi diluncurkan pada Januari 2022, sehingga tersedia rentang waktu yang cukup untuk menelusuri keluhan dan tindak lanjut penyelenggara antarversi aplikasi. *Google Play Store* dipilih karena ulasannya terbuka, bertanggal, memuat rating dan versi aplikasi saat ulasan dibuat, serta menyediakan balasan pengembang, sehingga keluhan dapat dikaitkan dengan fase pembaruan dan respons penyelenggara. Ulasan tersebut bersifat sukarela dan berasal dari pengguna *Android*, sehingga tidak mewakili seluruh pengguna; keterbatasan ini dibahas pada bagian Keterbatasan.

Untuk menilai keberhasilan sistem informasi, DeLone dan McLean (2003) merevisi model yang sebelumnya mereka usulkan pada tahun 1992 menjadi *Information Systems Success Model* (ISSM). Model tersebut mencakup enam konstruk, yaitu *system quality*, *information quality*, *service quality*, *use/intention to use*, *user satisfaction*, dan *net benefits*. Dalam penelitian ini, empat konstruk digunakan dengan definisi yang mengacu pada DeLone dan McLean (2003). *System quality* dimaknai sebagai kualitas teknis sistem, *information quality* sebagai kualitas informasi yang dihasilkan sistem bagi pengguna, *service quality* sebagai dukungan yang diberikan penyelenggara layanan kepada pengguna, dan *user satisfaction* sebagai evaluasi pengguna terhadap kinerja sistem secara keseluruhan. Dalam literatur ISSM, service quality umumnya diukur berdasarkan selisih antara harapan pengguna dan pengalaman yang mereka alami. Namun, karena ulasan aplikasi yang dianalisis bersifat singkat dan tidak memuat informasi mengenai harapan pengguna, selisih tersebut tidak dapat dihitung. Oleh karena itu, sebuah ulasan dikodekan sebagai *service quality* apabila pengguna secara eksplisit mengevaluasi dukungan penyelenggara, termasuk respons, bantuan, kanal pengaduan, petugas, atau pemeliharaan. Ketidaksesuaian layanan diidentifikasi berdasarkan pernyataan eksplisit pengguna, misalnya "tidak ada respon" atau "lambat respon". Dengan demikian, *service quality* dalam penelitian ini merujuk pada penilaian pengguna terhadap dukungan penyelenggara yang tampak dalam teks ulasan, bukan pada skor kesenjangan antara harapan dan pengalaman. Kata kunci serta indikator pengodean yang diadaptasi dari Abidin et al. (2023) digunakan semata-mata sebagai penyesuaian praktis untuk analisis ulasan aplikasi, bukan sebagai landasan teoretis.

Model ini telah banyak digunakan pada aplikasi layanan di Indonesia (Agustina & Sutinah, 2019; Amalia & Azizah, 2022; Ernawati et al., 2021; Nusandari et al., 2022), termasuk pada M-Paspor (Liang *et al*., 2024; Varastika et al., 2025). Studi M-Paspor lainnya menelaah kepuasan melalui survei atau wawancara pada satu kantor imigrasi (Anisa & Dewi, 2024; Bahri et al., 2025) atau memanfaatkan ulasan *Play Store* (Maheri et al., 2025; Ramadhani et al., 2025). Penelitian ini berbeda dari studi-studi sebelumnya dalam tiga hal. Pertama, studi berbasis ISSM pada M-Paspor (Liang et al., 2024; Varastika et al., 2025) dan studi kepuasan berbasis survei (Anisa & Dewi, 2024; Bahri et al., 2025) menilai kualitas layanan pada kelompok responden tertentu, sedangkan

penelitian ini memanfaatkan ulasan yang terakumulasi sejak aplikasi diluncurkan sehingga memungkinkan penelusuran antarversi aplikasi. Kedua, studi berbasis ulasan sebelumnya berfokus pada klasifikasi sentimen (Maheri et al., 2025) dan kepuasan pengguna dalam perspektif implementasi *e-government* (Ramadhani et al., 2025), sedangkan penelitian ini memetakan ulasan secara multi-label ke dimensi ISSM dan mengidentifikasi topik dominan pada setiap dimensi. Ketiga, penelitian ini menelusuri sisi penyelenggara, yaitu bagaimana keluhan tersebut ditindaklanjuti.

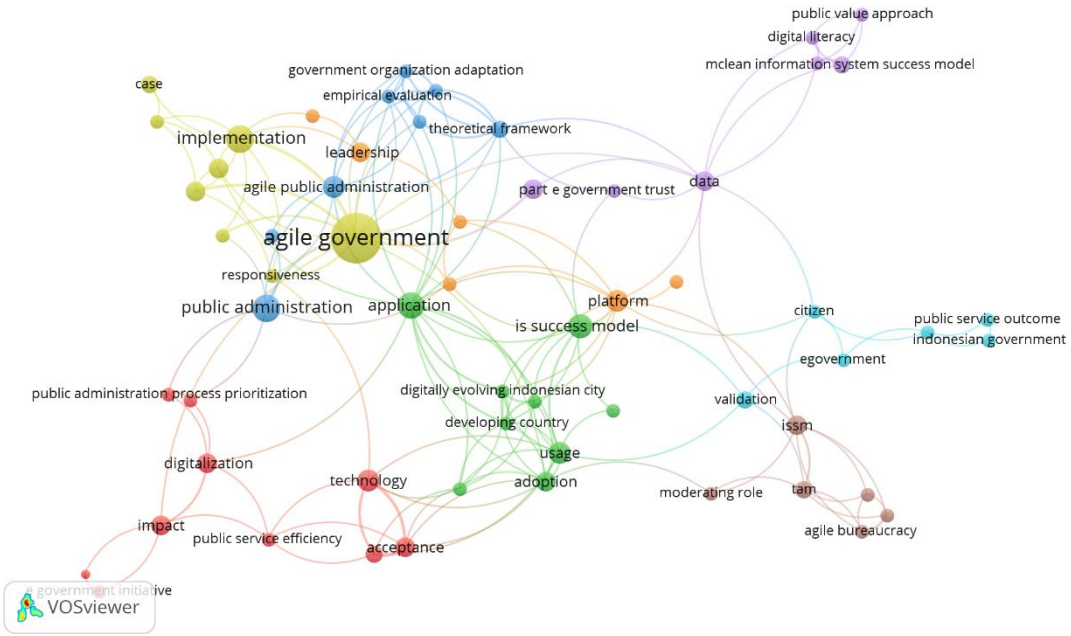
Perbedaan penelitian ini dari empat studi terdekat dapat dirinci pada empat aspek. Dari sisi pertanyaan penelitian, Liang et al. (2024) dan Varastika et al. (2025) menilai kualitas layanan M-Paspor berdasarkan ISSM, Maheri et al. (2025) mengklasifikasikan sentimen ulasan, dan Ramadhani et al. (2025) menilai kepuasan pengguna dalam perspektif implementasi *e-government*, sedangkan penelitian ini menanyakan apakah keluhan yang terakumulasi memperoleh respons yang terdokumentasi. Dari sisi unit analisis, kedua studi berbasis ISSM menggunakan data dari kelompok responden tertentu, sedangkan Maheri et al. (2025) menganalisis 5.934 ulasan tanpa menelusurinya antarversi; penelitian ini menganalisis 17.411 ulasan yang dapat ditelusuri antarversi aplikasi dan membandingkannya dengan jejak digital penyelenggara. Dari sisi pengukuran responsivitas, studi-studi tersebut tidak menelusuri tindak lanjut penyelenggara; Ramadhani et al. (2025) mencatat respons layanan sebagai persepsi pengguna, sedangkan penelitian ini menilai respons yang teramati secara publik melalui empat indikator, yaitu keberadaan, kesesuaian, ketepatan waktu, dan indikasi efektivitas. Dari sisi kontribusi temuan, studi-studi tersebut menghasilkan tingkat kualitas atau kepuasan layanan serta akurasi klasifikasi (misalnya SVM sebesar 80,76% pada Maheri et al., 2025), sedangkan penelitian ini menunjukkan ketidakselarasan antara umpan balik pengguna dan respons kelembagaan yang terdokumentasi secara publik.

Ulasan pengguna merupakan jejak digital pengalaman warga sebagai penerima layanan, sehingga relevan digunakan sebagai sarana evaluasi dalam tata kelola digital yang berpusat pada warga. Volume ulasan yang besar dan tidak terstruktur memerlukan dukungan komputasi. Model bahasa transformer untuk bahasa Indonesia seperti NusaBERT (Wongso et al., 2025) dan pemodelan topik berbasis neural seperti BERTopic (Grootendorst, 2022) memungkinkan ribuan ulasan dikelompokkan secara sistematis, sementara interpretasinya tetap berada pada peneliti (Lim, 2025). Meskipun demikian, ISSM terutama menjelaskan kualitas dan kepuasan, bukan bagaimana penyelenggara menindaklanjuti umpan balik. Celah ini dijembatani oleh konsep *agile government*, yang dalam administrasi publik dipahami sebagai kemampuan pemerintah merespons perubahan kebutuhan public secara efisien (Mergel et al., 2021). Menurut Mergel et al. (2018), ranah evaluasi *agile government* masih jarang dikaji. Penelitian ini mengoperasionalkannya dalam tiga lapis yang dihubungkan oleh satu proposisi kerja. Pertama, ulasan pengguna dipandang sebagai potensi sumber *sensing* yaitu kemampuan mengenali kebutuhan dan keluhan (Chatfield & Reddick, 2017). Kedua, tiga atribut model A.T Kearney (2003, dalam Dahmardeh & Pourshahabi, 2011), yaitu *e-government, customer service, dan performance management,* digunakan sebagai kerangka pengelompokkan keluhan menurut aspek layanan, bukan sebagai ukuran kinerja instansi. Ketiga, *responding*, yaitu tindak lanjut atas keluhan, ditelusuri melalui jejak digital publik penyelenggara. Proposisi kerjanya : keluhan yang terpetakan pada suatu aspek layanan diharapkan

diikuti respons teramati, sesuai, dan tepat waktu; sejauh respons itu tidak teramati, yang dapat disimpulkan hanyalah ketidakselarasan antara umpan balik publik dan respons yang terdokumentasi bukan ketiadaan *sensing* atau tindakan internal.

Untuk memberikan gambaran awal mengenai posisi penelitian ini, dilakukan analisis ko-kemunculan kata kunci menggunakan VOSviewer (van Eck & Waltman, 2010) terhadap literatur tentang *agile government*, ISSM, dan administrasi publik. Hasil analisis ditampilkan pada Gambar 1 sebagai ilustrasi eksploratif, bukan sebagai bukti bahwa penelitian sejenis belum pernah dilakukan. Peta tersebut menunjukkan adanya dua arus penelitian yang relatif terpisah. Klaster *agile government* berfokus pada praktik implementasi, studi kasus, kerangka konseptual, dan adaptasi organisasi pemerintah, sementara klaster ISSM berkaitan dengan penerapan, adopsi, serta penggunaan layanan digital, termasuk dalam konteks kota-kota di Indonesia.

Kedua klaster tersebut terhubung melalui *simpul application*, *platform*, data, dan *public administration*. Di sisi lain, simpul *responsiveness* relatif kecil dan berada di tepi klaster *agile government*. Pola ini hanya menunjukkan bahwa responsivitas terhadap umpan balik pengguna belum tampak menonjol dalam korpus yang dipetakan. Hasil tersebut bergantung pada basis data, kata kunci, dan ambang batas yang digunakan, sehingga tidak dapat ditafsirkan sebagai bukti ketiadaan studi serupa. Penelitian yang paling dekat dengan kajian ini adalah studi berbasis ulasan M-Paspor (Maheri et al., 2025; Ramadhani et al., 2025), yang menitikberatkan pada klasifikasi sentimen dan kepuasan pengguna, tetapi belum menelusuri tindak lanjut yang dilakukan penyelenggara.



Gambar 1. Peta ko-kemunculan kata kunci literatur *agile government*, ISSM, dan administrasi publik sebagai ilustrasi eksploratif (olah data *VOSviewer*)

Berdasarkan perbandingan tersebut, celah yang diisi penelitian ini adalah penelurusan respons penyelenggara terhadap keluhan yang telah dipetakan menurut

dimensi ISSM. Oleh karena itu, penelitian ini bertujuan untuk: (1) mengidentifikasi tingkat kepuasan pengguna M-Paspor beserta hubungannya dengan dimensi *system quality*, *information quality*, dan service quality; (2) mengidentifikasi topik serta sentimen yang dominan pada setiap dimensi ISSM dan menilai keterkaitannya dengan atribut *agile evaluation*, yaitu *e-government*, *customer service*, dan *performance management*; serta (3) menganalisis responsivitas Ditjen Imigrasi terhadap pola keluhan yang disampaikan pengguna. Kebaruan penelitian ini dibedakan menjadi dua. Secara metodologis, penelitian ini memetakan ulasan secara multi-label ke empat konstruk ISSM dan menelusurinya antarfase versi aplikasi. Secara konseptual, penelitian ini membedakan potensi sumber sensing dari responsivitas yang teramati secara publik, sehingga memperlihatkan ketidakselarasan antara suara pengguna dan respons kelembagaan. Penelitian ini diharapkan bermanfaat bagi tiga pihak. Bagi Ditjenim, hasil penelitian memberikan peta keluhan berdasarkan dimensi ISSM dan topik dominan, serta gambaran tindak lanjut yang telah dan belum tercatat, sehingga dapat menjadi masukan bagi penyusunan catatan rilis, komunikasi gangguan, dan pemantauan ulasan. Bagi pembuat kebijakan, khususnya pengelola SPBE dan pelayanan publik, penelitian ini menawarkan cara memanfaatkan ulasan pengguna sebagai pelengkap survei kepuasan masyarakat dalam evaluasi layanan digital. Bagi peneliti selanjutnya, penelitian ini menyediakan rancangan yang memetakan keluhan ke konstruk ISSM, klasifikasi NusaBERT, BERTopic, dan penelusuran jejak digital yang dapat direplikasi dan dikembangkan, misalnya melalui wawancara atau analisis deret waktu antarversi. Lingkup penelitian dibatasi pada ulasan di *Google Play Store* dan empat dimensi ISSM. Dimensi *use/intention to use* dan *net benefits* tidak dioperasionalkan dalam penelitian ini. Pengukuran *use/intention to use* membutuhkan informasi mengenai perilaku penggunaan, misalnya frekuensi penggunaan atau keberlanjutan penggunaan oleh pengguna yang sama. Namun, ulasan aplikasi bersifat anonim, hanya muncul satu kali, dan tidak menyertakan riwayat penggunaan pengguna. Sementara itu, *net benefits* menuntut ukuran dampak pada tingkat individu atau organisasi, seperti penghematan waktu maupun biaya, yang tidak dapat diverifikasi hanya berdasarkan teks ulasan. Oleh karena itu, penelitian ini menggunakan ISSM secara parsial dengan memakai empat dari enam konstruk sebagai kerangka untuk mengelompokkan keluhan pengguna, bukan untuk menguji model secara menyeluruh. Hubungan antar-konstruk dalam ISSM juga tidak diuji. Dengan demikian, penelitian ini tidak menyatakan telah mengevaluasi keberhasilan M-Paspor secara komprehensif.

METODE

Pendekatan Penelitian

Penelitian ini menerapkan analisis isi (*content analysis*) berbantuan komputasi dengan kuantifikasi deskriptif atas teks ulasan pengguna. Pendekatan ini dipilih karena tujuan penelitian adalah mengenali isi dan pola keluhan dalam teks yang tersedia secara terbuka, bukan menguji hubungan antarvariabel. Teknik tersebut juga memungkinkan teks dikodekan ke dalam kategori secara sistematis dan dapat direplikasi (Lim, 2025; Riffe et al., 2024). Kategori awal disusun secara deduktif berdasarkan dimensi ISSM (DeLone & McLean, 2003), lalu diperkaya melalui pembacaan induktif atas topik-topik yang muncul dari BERTopic. Satuan analisis

adalah satu ulasan, dengan konteks waktu dan versi aplikasi pada saat ulasan tersebut dibuat. Mengingat jumlah ulasan mencapai belasan ribu, NusaBERT dan BERTopic digunakan sebagai alat bantu komputasional untuk memperluas pengodean manual ke seluruh korpus. Namun, interpretasi akhir tetap dilakukan oleh peneliti sebagai *expert coder*. Validasi antara analisis kualitatif dan komputasional dirancang melalui tiga jalur. Pertama, hasil pengodean manual menjadi acuan bagi NusaBERT; kinerja model dievaluasi menggunakan akurasi, *macro-F1*, dan AUC-ROC, sementara konsistensi label acuan diperiksa melalui kesepakatan antara dua pelabel menggunakan Cohen's kappa. Kedua, topik yang dihasilkan BERTopic dievaluasi secara komputasional melalui koherensi dan keragaman topik. Topik tersebut kemudian ditafsirkan kembali secara kualitatif dengan mempertimbangkan ulasan representatif, lalu dipetakan ke dimensi ISSM berdasarkan kata kunci pada Tabel 1. Proses ini memungkinkan identifikasi topik yang tidak bermakna atau tidak selaras dengan definisi dimensi. Ketiga, pola kuantitatif antarversi aplikasi, meliputi proporsi ulasan dan sentimen, dibandingkan dengan jejak digital publik serta dibaca bersama ulasan sampel pada setiap klaster keluhan. Dengan demikian, pola yang tampak dalam angka dapat dikonfirmasi atau dibatasi oleh pembacaan teks dan sumber respons yang tersedia. Hasil kuantifikasi berupa frekuensi dan proporsi kemudian diinterpretasikan secara deskriptif. Landasan teori ISSM dan *agile government* telah dijelaskan pada bagian Pendahuluan.

Desain yang digunakan adalah analisis isi berbantuan komputasi dengan kuantifikasi deskriptif yang dijalankan dengan lima: (1) pengumpulan dan pembersihan ulasan, (2) pelabelan manual dan klasifikasi multi-label ke dimensi ISSM, (3) pemodelan topik, (4) pemetaan topik ke atribut *agile evaluation*, dan (5) penelusuran respons penyelenggara. Pendekatan kuantitatif inferensial tidak dipilih karena ulasan bersifat sukarela dan tidak mewakili populasi pengguna, sehingga angka frekuensi dan proporsi hanya digunakan untuk menggambarkan pola, bukan menguji hipotesis. Dengan demikian, kuantifikasi (frekuensi, proporsi, dan sentimen) berfungsi menggambarkan pola, sedangkan pembacaan kualitatif menentukan makna kategori dan menilai kesesuaian respons penyelenggara.

Data dan Pengumpulan Data

Data utama berupa ulasan pengguna M-Paspor di *Google Play Store* yang diambil menggunakan pustaka *google-play-scraper* pada bahasa *Python* dengan parameter bahasa dan negara Indonesia (Dwicahyo & Indah Ratnasari, 2023), sejak aplikasi diluncurkan pada Januari 2022 hingga tahun 2026. Proses ini menghasilkan 18.625 ulasan beserta tanggal, isi, rating, versi aplikasi saat ulasan dibuat (*reviewCreatedVersion*), dan balasan pengembang. Data sekunder untuk menelusuri respons penyelenggara meliputi catatan rilis (*what's new*) tiap versi aplikasi yang dihimpun dari APKPure, siaran pers dan berita resmi Ditjenim, unggahan akun X dan Instagram resmi, FAQ M-Paspor, serta laporan kinerja Ditjenim. Seluruh sumber bersifat terbuka. Nama akun pengulas tidak dicantumkan dalam artikel ini.

Pra-pemrosesan

Ulasan yang mengandung URL dan ulasan duplikat dihapus, sehingga tersisa 17.411 ulasan. Tahap pra-pemrosesan meliputi pembersihan karakter yang tidak diperlukan dan tokenisasi menggunakan tokenizer NusaBERT(Syam et al., 2024). Sebanyak 525 ulasan (sekitar 3,0%) dipilih secara acak sebagai data acuan (*ground truth*) dan dilabeli secara manual ke dalam empat dimensi ISSM secara multi-label, karena satu ulasan dapat memuat beberapa isu sekaligus. Polaritas sentimen ditentukan berdasarkan rating, yaitu rating 1–2 dikategorikan negatif dan 4–5 positif, sedangkan rating 3 dinilai secara semantik. Pelabelan manual mengacu pada prinsip human coding dalam pembelajaran mesin untuk riset administrasi publik (Anastasopoulos & Whitford, 2019) Reliabilitas diuji menggunakan Cohen's kappa (Cohen, 1960) pada pelabelan independen oleh dua pelabel dan diinterpretasikan menurut (Landis & Koch, 1977) Dimensi ISSM dioperasionalkan melalui variabel dan indikator yang mengacu pada (Abidin et al., 2023), sebagaimana pada Tabel 1.

Pelabelan Manual dan Uji Realibilitas

Sebanyak 525 ulasan (±3,0% dari 17.411 ulasan) dipilih secara acak [teknik pengacakan dan seed] sebagai data acuan (*ground truth*). Pelabelan dilakukan secara independen oleh peneliti utama (Pelabel A) dan mahasiswa Ilmu Administrasi Negara angkatan 2022 (Pelabel B) menggunakan panduan yang sama, yang memuat definisi dan indikator setiap dimensi ISSM (Tabel 1). Pelabel B tidak mengetahui label dari Pelabel A. Setiap ulasan dilabeli secara multi-label pada empat dimensi ISSM karena dapat memuat lebih dari satu isu. Sebagai *expert coder*, peneliti utama memahami definisi tiap dimensi sehingga labelnya menjadi acuan untuk melatih dan menguji model, sedangkan label Pelabel B digunakan untuk menguji konsistensi pelabelan. Sentimen ditentukan dari rating: 1–2 negatif, 4–5 positif, dan rating 3 dinilai secara semantik. Prosedur ini mengikuti prinsip *human coding* dalam pembelajaran mesin untuk riset administrasi publik (Anastasopoulos & Whitford, 2019).

Reliabilitas pelabelan diuji menggunakan Cohen's kappa (Cohen, 1960) pada setiap dimensi ISSM dan diinterpretasikan menurut Landis dan Koch (1977), yaitu 0,61–0,80 sebagai substansial dan 0,81–1,00 sebagai hampir sempurna. Validitas isi dijaga melalui buku kode yang mengoperasionalkan dimensi ISSM menurut indikator Abidin et al. (2023) dan pelabelan dilakukan secara independen oleh dua pelabel, yaitu peneliti utama dan seorang mahasiswa Ilmu Administrasi Negara Universitas Negeri Padang. Validitas kinerja klasifikasi dinilai pada data uji menggunakan akurasi, macro-F1, dan AUC-ROC. Hasil uji disajikan pada Tabel 2 di bagian Hasil dan Pembahasan.

Tabel 1. Variabel dan indikator berdasarkan teori ISSM

Variabel	Kata kunci (Abidin et al., 2023)	Penjelasan	Sentimen positif	Sentimen negatif
System quality	Akses, *login*, *error*, *crash*, *loading*, lambat, lemot, *bug*, gangguan, tidak bisa, gagal, aplikasi, *update*, server, *down*, fitur, keamanan, responsif, mudah digunakan	Kualitas teknis sistem informasi yang terbentuk dari integrasi antara *software* dan *hardware*	Aplikasi mudah dimengerti, mudah diakses, tidak ada berkas yang harus diunggah berulang-ulang	Tidak ada fitur *waiting list*, mengakses aplikasi susah karena *overload* pemohon, *upload* data tidak terverifikasi
Information quality	Informasi, petunjuk, panduan, syarat, dokumen, lengkap, jelas, akurat, tidak jelas, membingungkan, status, notifikasi, pemberitahuan, instruksi, prosedur, jadwal, permohonan	Tingkat pemenuhan informasi sistem terhadap kebutuhan dan ekspektasi pengguna; mengevaluasi kualitas *output* sistem informasi	Pilihan kantor imigrasinya banyak, dilengkapi lagi panduannya, berikan peringatan untuk tidak memakai baju warna putih	*App*-nya minim pemberitahuan, tidak bisa bertanya; kuota pendaftaran hanya tersedia pada tanggal 1, selebihnya di-*lock*; infonya kurang jelas

Variabel	Kata kunci (Abidin et al., 2023)	Penjelasan	Sentimen positif	Sentimen negatif
Service quality	Respon, balasan, pengembang, *customer service*, keluhan, solusi, bantuan, pelayanan, tanggapan, lambat respon, tidak ada respon, *maintenance*, pembaruan, perbaikan, antrean, kantor imigrasi	Kesenjangan antara ekspektasi pengguna dan layanan yang diterima	Pelayanannya luar biasa baik, petugasnya cekatan dan ramah, petugasnya *fast* respon	Ketika wawancara petugas tidak melihat tujuan membuat paspor, tetapi langsung menuduh untuk bekerja
User satisfaction	Puas, kecewa, bagus, buruk, membantu, bermanfaat, bintang, rating, terima kasih, menyesal, berguna, mudah, susah, ribet, praktis	Variabel paling krusial dalam menilai keberhasilan sistem informasi; mengukur kepuasan pengguna terhadap performa sistem secara keseluruhan	Memudahkan masyarakat, sangat mudah dan tidak ribet	Sangat kecewa, aplikasi memalukan, aplikasi ga jelas dan ga membantu

Sumber: definisi konstruk mengacu pada Delone dan Mclean (2003); kata kunci diadaptasi dari Abidin et al. (2023), disesuaikan oleh peneliti.

Ulasan berlabel dibagi secara acak (tanpa stratifikasi; *random_state* = 42) menjadi 420 data latih dan 105 data evaluasi (80:20). Tidak digunakan data uji terpisah, sehingga 105 data evaluasi berfungsi sebagai data validasi untuk memilih model terbaik pada setiap akhir epoch dan menentukan ambang keputusan, sekaligus untuk melaporkan kinerja. Model NusaBERT-base (LazarusNLP/NusaBERT-base; Wongso et al., 2025) di-*fine-tune* untuk klasifikasi multi-label dengan lima keluaran biner yang independen, yaitu empat dimensi ISSM dan polaritas sentimen (He & Xia, 2018). Pelatihan menggunakan fungsi *loss binary cross-entropy*, *learning rate* 2e-5, *batch size* 16, 3 *epoch*, *weight decay* 0,01, dan panjang maksimum token 128, sedangkan pengaturan lain menggunakan nilai bawaan Trainer. Ambang keputusan setiap label ditentukan dengan memaksimalkan statistik J Youden (selisih *true positive rate* dan *false positive rate*) pada kurva ROC data evaluasi. Model kemudian digunakan untuk mengklasifikasikan seluruh 17.411 ulasan dengan ambang tersebut / seluruh 17.411 ulasan dengan ambang standar 0,5. Kinerjanya dievaluasi menggunakan *confusion matrix*, *macro-averaged F1-score* (rerata F1 kelas 0 dan kelas 1 pada setiap label), dan AUC-ROC per label.

Pemodelan Topik

Pemodelan topik dilakukan menggunakan BERTopic (Grootendorst, 2022) pada 17.411 ulasan. Setiap ulasan direpresentasikan sebagai vektor menggunakan model sentence embedding bawaan BERTopic, direduksi dimensinya dengan UMAP (*n_neighbors* = 15, *n_components* = 5, *min_dist* = 0,0, *metric* = *cosine*), lalu dikelompokkan dengan HDBSCAN (*min_cluster_size* = 10, *metric* = *euclidean*, *cluster_selection_method* = *excess of mass*). Kata kunci topik diekstraksi dengan c-TF-IDF berbasis *CountVectorizer* tanpa daftar stopword tambahan. Topik dievaluasi menggunakan koherensi topik (Cv) (Röder et al., 2015) dan keragaman topik (*topic diversity*), kemudian dipetakan ke dimensi ISSM berdasarkan kata kunci pada Tabel 1 serta dibaca secara kualitatif bersama contoh ulasan representatif. Jumlah topik direduksi menjadi 47 (termasuk outlier) melalui parameter *nr_topics*, dan dokumen yang tidak masuk klaster mana pun dikelompokkan sebagai outlier (Topik −1) tanpa dialokasikan ulang. Parameter lain menggunakan nilai bawaan BERTopic versi 0.17.4. Parameter utama menggunakan learning rate 2e-5, batch size, 3 epoch, weight decay 0.01. dan panjang maksimum token 128. Pengaturan lain menggunakan nilai bawaan Trainer.

Analisis *Agile Evaluation* dan Responsivitas

Topik dipetakan ke tiga atribut *agile evaluation* model A.T. Kearney yang digunakan (Dahmardeh & Pourshahabi, 2011), yaitu *e-government* (keandalan infrastruktur dan fungsi sistem digital), *customer service* (kemudahan prosedur dan kualitas interaksi layanan), serta *performance management* (pengukuran kinerja untuk perbaikan berkelanjutan). Proses *sensing* diidentifikasi dari pola topik dan sentimen ulasan, sementara *responding* ditelusuri melalui perubahan topik dan sentimen antarversi aplikasi, catatan rilis, serta komunikasi resmi Ditjenim. Hubungan antar-konsep dalam penelitian ini dinyatakan melalui satu proposisi kerja: keluhan pengguna yang dipetakan ke dalam aspek layanan tertentu berpotensi menjadi sumber sensing. Dalam evaluasi layanan yang berkelanjutan, keluhan tersebut idealnya ditindaklanjuti melalui respons yang dapat diamati, relevan dengan isu yang disampaikan, dan diberikan tepat waktu. Tiga atribut model A.T. Kearney sebagaimana diadaptasi oleh Dahmardeh dan Pourshahabi (2011) digunakan semata-mata sebagai kerangka untuk mengelompokkan isu keluhan, yaitu *e-government* yang berkaitan dengan keandalan infrastruktur dan fungsi sistem digital; *customer service* yang berkaitan dengan kemudahan prosedur serta kualitas interaksi layanan; dan *performance management* yang berkaitan dengan pengukuran kinerja untuk mendukung perbaikan berkelanjutan.

Sensing diidentifikasi melalui pola topik dan sentimen yang muncul dalam ulasan pengguna sebagai potensi sumber informasi. Adapun *responding* ditelusuri dengan membandingkan perubahan topik dan sentimen antarversi aplikasi, menelaah catatan rilis, serta memeriksa komunikasi resmi Ditjenim. Penjelasan lebih lanjut mengenai konsep, indikator, data, dan kriteria penarikan kesimpulan disajikan pada Tabel 2.

Tabel 2. Matriks konsep, indikator, data, data dan aturan penarikan kesimpulan

Konsep	Indikator	Data	Aturan penarikan kesimpulan	Yang terukur/tidak terukur
Agile evaluation (Mergel et al., 2018)	Tidak diukur langsung; dioperasionalkan lewat empat baris dibawah	-	Disimpulkan hanya dari keterkaitan antara keluhan yang terpetakan dan respons yang teramati	Terukur: keterkaitan keluhan – respons. Tidak terukur : tingkat kelincahan instansi.
Sensing (potensi sumber)	Pola topik dan sentimen menurut dimensi ISSM dan fase aplikasi	17.411 ulasan; 14.810 ulasan berinformasi versi	Keluhan berulang pada suatu klaster dibaca sebagai informasi yang tersedia untuk dideteksi	Terukur: ketersediaan sinyal keluhan. Tidak terukur : apakah Ditjenim memantau atau memproses ulasan
Atribut Kearney (*e – government, customer service, performance management*)	Klaster topik BERTopic dikelompokkan menurut definisi atribut; ekspresi penilaian umum pengguna (mis. T16) tidak dipetakan ke atribut	Klaster topik dan ulasan representatif	Satu klaster dipetakan ke satu atribut berdasarkan definisi tertulis. Rating rendah pada klaster dibaca sebagai sinyal pengguna, bukan bukti kinerja instansi. *Customer service* dibatasi pada layanan terkait aplikasi; pelayanan tatap muka dicatat terpisah	Terukur: sebaran keluhan menurut aspek layanan. Tidak terukur : kinerja instansi pada atribut tersebut.

Konsep	Indikator	Data	Aturan penarikan kesimpulan	Yang terukur/tidak terukur
Responding (teramati publik)	Keberadaan, kesesuaian, ketepatan waktu, dan efektivitas respons	Catatan rilis, siaran pers, X dan Instagram resmi, FAQ, *live chat,* pemberitaan, LkjIP dan IKM (Januari 2022 – September 2026); balasan pengembang pada 14 dari 17.411 ulasan	Hanya respons yang terbit setelah ulasan sampel dihitung sebagai tindak lanjut. "Belum tercatat" dibaca sebagai keterbatasan bukti, bukan ketiadaan tindakan	Terukur : respons yang dapat diamati warga. Tidak terukur : tindak lanjut, kebijakan, dan proses internal
Responsivitas (penilaian gabungan)	Kategori sesuai, sebagian, sebagian (lemah), belum tercatat per klaster; efektivitas dibaca deskriptif	Tabel 6 dan 7 (setelah *renumbering*)	Dibaca deskriptif tanpa tafsir sebab-akibat; Fase 3 hanya 824 ulasan	Terukur: ketidakselarasan yang teramati. Tidak terukur : hubungan kausal pembaruan aplikasi dan persepsi pengguna

Sumber: olah data dan rumusan peneliti (2026)

Dengan aturan tersebut, yang terukur dalam penelitian ini adalah sebaran keluhan pengguna menurut dimensi ISSM dan aspek layanan, serta ada tidaknya, kesesuaian, ketepatan waktu, dan indikasi efektivitas respons yang teramati secara publik. Penelitian ini tidak mengukur kapabilitas *sensing* yang dioperasikan instansi, kinerja internal pada ketiga atribut, maupun hubungan sebab-akibat antara pembaruan aplikasi dan persepsi pengguna. Pemetaan topik ke atribut A.T. Kearney dilakukan menurut objek yang dikeluhkan, dengan alasan teoretis berikut. Topik dipetakan ke *e-government* apabila keluhan menyangkut keandalan infrastruktur dan fungsi sistem digital yang menjadi prasyarat layanan daring, seperti akses akun, penjadwalan, dan pembayaran, karena atribut ini menilai sejauh mana teknologi digital benar-benar mendukung penyampaian layanan. Topik dipetakan ke customer service apabila keluhan menyangkut kemudahan prosedur dan kualitas interaksi layanan yang diterima pengguna, karena atribut ini menilai kualitas hubungan penyelenggara dengan warga. Topik dipetakan ke *performance management* apabila keluhan menyangkut kinerja aplikasi secara berkelanjutan, yaitu stabilitas, kecepatan, dan pemeliharaan, karena atribut ini berkaitan dengan pengukuran kinerja untuk perbaikan berkelanjutan, dan ulasan negatif atas aspek tersebut merupakan indikator hasil (*outcome*) yang dirasakan pengguna sehingga relevan sebagai masukan pengukuran kinerja. Pemetaan ditentukan oleh isi topik, bukan oleh rating. Rating rendah pada suatu klaster dibaca sebagai penilaian pengguna atas kinerja layanan yang bercampur dengan ekspresi kepuasan, sehingga tidak dibaca sebagai bukti kinerja instansi. Kinerja instansi pada atribut *performance management* dinilai secara terpisah melalui keberadaan dan pemanfaatan pengukuran kinerja oleh penyelenggara, yaitu IKM, laporan kinerja, dan ada tidaknya pemantauan ulasan atau rating (Tabel 7 dan 8).

Penilaian responsivitas dilakukan melalui empat tahap. Pertama, klaster keluhan dominan pada setiap atribut ditentukan berdasarkan hasil pemetaan topik (Tabel 5). Kedua, dari setiap klaster dipilih satu hingga tiga ulasan sampel yang secara eksplisit mengandung keluhan tersebut dan mewakili isinya menurut pembacaan peneliti, disertai tanggal serta versi aplikasi. Ketiga, jejak digital publik ditelusuri melalui catatan rilis aplikasi, siaran pers dan pengumuman situs resmi, unggahan akun X dan Instagram resmi Ditjen Imigrasi, FAQ, kanal live chat, pemberitaan media,

serta laporan kinerja dan IKM pada periode Januari 2022 hingga September 2026, lalu dibandingkan secara kronologis dengan tanggal ulasan sampel sebagai bentuk triangulasi sumber. Respons yang muncul sebelum ulasan sampel tetap dicatat, namun hanya respons yang terbit setelahnya dihitung sebagai tindak lanjut dalam penilaian ketepatan waktu. Keempat, setiap klaster dinilai oleh peneliti menggunakan empat indikator, yaitu keberadaan (ada tidaknya jejak publik yang membahas isu klaster), kesesuaian (sesuai bila respons menjawab isu secara eksplisit; sebagian bila menyentuh isu secara tidak langsung atau hanya sebagian aspek; sebagian (lemah) bila respons bersifat umum seperti *bug fixing*; belum tercatat bila tidak ditemukan jejak), ketepatan waktu (jeda dalam hari dari ulasan sampel ke setiap respons yang tercatat setelahnya, dibaca secara deskriptif), dan efektivitas (indikasi penurunan proporsi ulasan berkata kunci keluhan sejenis pada dua bulan sesudah dibandingkan dua bulan sebelum rilis, serta ada tidaknya keluhan yang sama pada fase berikutnya dan rata-rata klaster sebagai pembanding). Hasil penilaian disajikan pada Tabel 6 dan Tabel 7.

Untuk analisis antarversi, ulasan dipetakan ke fase pembaruan aplikasi berdasarkan *kolom reviewCreatedVersion*: Fase 1 (versi 2.9.3–5.4.7), Fase 2 (6.0.1–6.9.2), dan Fase 3 (7.1.3–7.5.37). Sebanyak 14.810 ulasan (85%) memiliki informasi versi dan dapat dipetakan, sedangkan 2.601 ulasan tanpa informasi versi dikecualikan dari analisis antarversi tetapi tetap digunakan untuk analisis dimensi ISSM dan pemodelan topik. Kategori "belum tercatat" dibaca sebagai keterbatasan bukti, bukan ketiadaan tindakan, dan hasil dibaca secara deskriptif tanpa ditafsirkan sebagai hubungan sebab-akibat.

HASIL DAN PEMBAHASAN

Reliabilitas Pelabelan dan Kinerja Model Klasifikasi

Reliabilitas pelabelan pada 525 ulasan tergolong tinggi (Tabel 2): kappa berada pada kategori hampir sempurna untuk tiga dimensi dan substansial untuk *user satisfaction*. Seluruh ketidaksepakatan bersifat searah: pelabel kedua cenderung lebih inklusif dalam memberikan label dimensi, sehingga perbedaan lebih mencerminkan perbedaan kriteria daripada kebingungan antardimensi. Nilai kappa 1,000 pada sentimen mencerminkan penerapan aturan berbasis rating, bukan penilaian independen.

Tabel 3. Reliabilitas pelabelan dan kinerja NusaBERT pada 105 data uji

Dimensi	Kappa	Akurasi (%)	*Macro*-F1	AUC	Ambang
System quality	0,885	70	0,68	0,794	0,89
Information quality	0,863	73	0,71	0,807	0,29
Service quality	0,868	86	0,84	0,870	0,43
User satisfaction	0,758	67	0,62	0,711	0,69
Sentimen	1,000	90	0,87	0,952	0,30

Catatan: *dihitung sebagai rerata F1 kelas 0 (0,58) dan kelas 1 (0,77). Sumber: olah data peneliti (2026).*

Kinerja model paling tinggi pada *service quality* dan sentimen, namun lebih rendah pada *system quality* dan *user satisfaction*. Dengan demikian, proporsi ulasan

pada dua dimensi terakhir perlu dipahami sebagai estimasi, bukan angka pasti. Temuan di bawah ini dibaca dengan kehati-hatian tersebut.

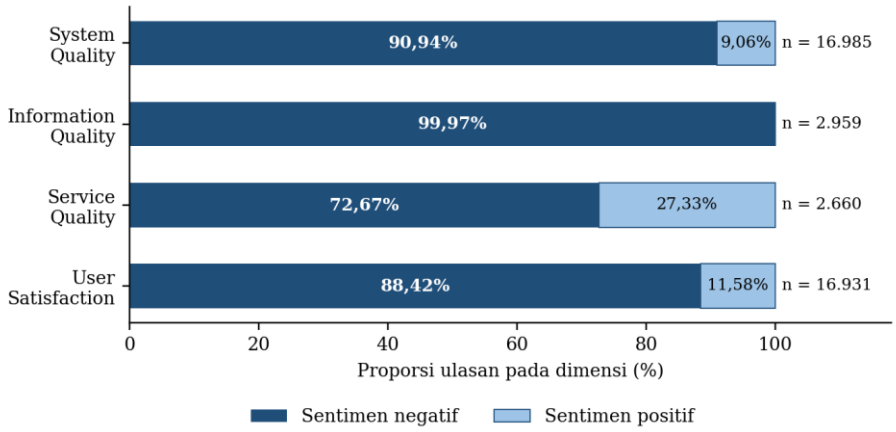
Kepuasan Pengguna dan Distribusi Dimensi ISSM

Rating rata-rata meningkat secara bertahap antarfase pembaruan aplikasi, tetapi mayoritas ulasan tetap negative pada setiap fase (Gambar 3). Meskipun demikian, mayoritas ulasan tetap negatif. Hasil klasifikasi multi-label menunjukkan bahwa ulasan terkonsentrasi pada dimensi *system quality* dan *user satisfaction* (Tabel 3 dan Gambar 2). Persentase pada Tabel 3 tidak berjumlah 100% karena satu ulasan dapat memiliki lebih dari satu label.

Tabel 4. Distribusi ulasan dan sentimen berdasarkan dimensi ISSM (N = 17.411)

Dimensi	Jumlah ulasan	Proporsi dari total (%)	Negatif (%)	Positif (%)
System quality	16.985	97,55	90,94	9,06
Information quality	2.959	17,00	99,97	0,03
Service quality	2.660	15,28	72,67	27,33
User satisfaction	16.931	94,89	88,42	11,58

Sumber : olah data peneliti (2026)



Gambar 2. Proporsi sentimen pada tiap dimensi ISSM

Dimensi *system quality* merupakan aspek yang paling banyak menimbulkan keluhan, dengan 16.985 ulasan atau 97,55% dari total, di mana 90,94% di antaranya bermuatan negatif. Keluhan tersebut meliputi kegagalan pengiriman kode OTP, kuota yang cepat terkuras ketika jadwal pendaftaran dibuka, gangguan sistem yang berulang, serta kinerja aplikasi yang lambat. Temuan ini sejalan dengan Maheri et al. (2025), yang mengidentifikasi sentimen negatif akibat galat sistem dan hambatan dalam proses pendaftaran, serta Ramadhani et al. (2025), yang mencatat kendala teknis dialami sekitar 40–50% pengguna. Kesamaan ini mengindikasikan bahwa keluhan keandalan sistem tidak terbatas pada satu penelitian. Meskipun demikian, masih terdapat 9,06% ulasan pada dimensi ini yang bersentimen positif. Selain itu, penelitian lain tetap menemukan manfaat layanan digital ini, seperti percepatan proses pengajuan dan penurunan antrean (Anisa & Dewi, 2024). Dengan demikian, persoalan utama tampak lebih berkaitan dengan konsistensi kinerja sistem daripada ketiadaan manfaat layanan, setidaknya di kalangan pengulas. Dari sudut pandang administrasi publik, kegagalan pada tahap pendaftaran dan pemilihan jadwal dapat mengubah efisiensi digitalisasi menjadi beban baru bagi warga.

Dimensi *information quality* teridentifikasi pada 2.959 ulasan atau 17,00% dari total ulasan, dengan proporsi sentimen negatif mencapai 99,97%. Keluhan yang paling umum berkaitan dengan pesan galat yang kurang informatif, tidak adanya perkiraan waktu pemulihan ketika terjadi gangguan, kebingungan dalam proses validasi OTP, serta ketentuan unggah foto KTP. Tingginya proporsi ulasan negatif menunjukkan bahwa komentar pada dimensi ini lebih banyak berbentuk kritik, sementara aspek informasi yang memadai relatif jarang mendapatkan apresiasi. Temuan ini sejalan dengan (Ramadhani et al., 2025), yang menemukan inkonsistensi informasi pada sekitar 35% pengguna, serta (Liang et al., 2024), yang menegaskan bahwa *information quality* berpengaruh *user satisfaction* M-Paspor. Batas antara dimensi ini dan *system quality* tidak selalu jelas. Sebagai contoh, pesan *timeout* yang muncul tanpa penjelasan dapat berkaitan dengan *information quality* sekaligus keandalan sistem. Pendekatan multi-label yang digunakan memungkinkan satu ulasan diklasifikasikan ke dalam lebih dari satu dimensi. Dari perspektif administrasi publik, informasi yang jelas selama gangguan merupakan bagian dari transparansi dan akuntabilitas penyelenggara layanan, bukan sekedar kenyamanan pengguna.

Dimensi *service quality* teridentifikasi pada 2.660 ulasan atau 15,28%, menjadikannya dimensi dengan jumlah ulasan paling sedikit di antara empat dimensi yang dianalisis. Meskipun demikian, dimensi ini memiliki proporsi sentimen positif tertinggi, yaitu 27,33%. Keluhan yang muncul meliputi pemeliharaan sistem yang dilakukan tanpa jadwal yang jelas, kesulitan memperoleh kuota yang disertai dugaan adanya percaloan, serta aplikasi yang sulit diakses. Sebaliknya, apresiasi pengguna umumnya ditujukan kepada keramahan dan kecepatan petugas dalam melayani. Temuan ini sejalan dengan (Bahri et al., 2025), yang menemukan bahwa *responsiveness* dan *fulfilment* berpengaruh positif dan signifikan terhadap kepuasan pengguna. Hasil ini juga sejalan dengan (Ramadhani et al., 2025), yang mencatat bahwa sekitar 20% pengguna mengalami rendahnya respons layanan. Perbedaan dengan temuan (Bahri et al., 2025), yang tidak menemukan pengaruh signifikan dari system availability dan contact, kemungkinan disebabkan oleh perbedaan metode penelitian. Survei yang dilakukan pada satu kantor imigrasi memiliki karakter yang berbeda dengan analisis ulasan spontan dari pengguna di tingkat nasional. Selain itu, pujian terhadap "pelayanan" tidak selalu menyebutkan secara spesifik objek yang dimaksud. Oleh karena itu, belum dapat dipastikan apakah apresiasi tersebut ditujukan pada aplikasi atau pada layanan tatap muka di kantor imigrasi. Adapun dugaan percaloan merupakan persepsi pengguna yang belum diverifikasi dalam penelitian ini. Namun, temuan tersebut mengarah pada persoalan keadilan akses, karena pengguna yang gagal memperoleh jadwal melalui kanal resmi berpotensi terdorong menggunakan jasa perantara dengan biaya tambahan.

Dimensi *user satisfaction* teridentifikasi pada 16.931 ulasan atau 94,89% dari total ulasan, dengan 88,42% bermuatan negatif dan 11,58% bermuatan positif. Topik yang paling banyak muncul meliputi jadwal yang penuh, hambatan dalam proses pendaftaran, pesan *timeout*, serta persepsi bahwa aplikasi justru menambah kesulitan bagi pengguna. Hampir seluruh ulasan diberi label *system quality* dan *user satisfaction,* sehingga kedekatan keduanya lebih mencerminkan prevalensi label mencerminkan prevalensilabel daripada hubungan antar dimensi, dan hubungan itu belum diuji secara statistik. Namun, hubungan tersebut dibaca secara deskriptif dan belum diuji sebagai pengaruh secara statistik. Temuan ini sejalan dengan (Liang et al., 2024), yang

menyatakan bahwa kepuasan pengguna dibentuk secara bersamaan oleh *system quality*, *information quality*, dan *service quality*. Dengan demikian, perbaikan pada satu aspek saja belum tentu cukup untuk meningkatkan kepuasan secara menyeluruh. Proporsi sentimen positif pada penelitian ini juga berbeda jauh dengan tingkat kepuasan 60–65% yang dilaporkan (Ramadhani et al., 2025). Perbedaan tersebut dapat dijelaskan oleh karakteristik ulasan *Google Play Store* yang bersifat sukarela dan lebih sering ditulis ketika pengguna mengalami kendala. Selain itu, terdapat ketidaksesuaian antara rating yang diberikan dan isi ulasan (Maheri et al., 2025). Selain itu, isi ulasan dan penilaian numerik dapat berbeda pada sebagian kasus (Maheri et al., 2025); karena sentimen dalam penelitian ini ditentukan dari rating, perbedaan tersebut tidak dapat diperiksa. Indeks Kepuasan Masyarakat (IKM) Ditjen Imigrasi yang mencapai 3,94 dari 4 pada tahun 2024 (Direktorat Jenderal Imigrasi, 2025) juga menunjukkan bahwa survei layanan umum dan rating aplikasi dapat menangkap pengalaman pengguna yang berbeda. Oleh karena itu, pemantauan ulasan pengguna perlu digunakan sebagai pelengkap survei agar evaluasi kinerja layanan digital menjadi lebih komprehensif.

Topik Dominan dan Relevansinya terhadap *Agile Evaluation*

BERTopic memproses sebanyak 17.411 ulasan dan menghasilkan 7.248 dokumen yang tergolong *outlier* (Topik −1), setara dengan 41,6% dari total dokumen. Selain itu, model menghasilkan satu klaster berukuran besar, yaitu Topik 0 dengan 8.383 ulasan, topik 0 dan *outlier* tidak dipetakan ke atribut *agile evaluation* karena kata kunci dominannya bersifat umum sehingga tidak menggambarkan satu isu keluhan serta puluhan topik yang lebih spesifik pada Topik 1 hingga 45. Tabel 4 menyajikan ringkasan topik dominan pada masing-masing dimensi ISSM. Nilai keragaman topik tergolong tinggi pada seluruh dimensi, berkisar antara 0,90 hingga 1,00. Sementara itu, nilai koherensi Cv berada pada kategori sedang, yaitu 0,44 hingga 0,56. Kondisi ini menunjukkan bahwa topik-topik yang dihasilkan relatif terpisah satu sama lain, tetapi kepaduan semantik antarkata penyusun topik masih terbatas. Oleh karena itu, pembacaan kualitatif terhadap contoh ulasan tetap diperlukan untuk memastikan interpretasi topik yang lebih akurat.

Tabel 5. Topik dominan pada tiap dimensi ISSM dan contoh ulasan

Dimensi (Cv; keragaman)	Topik dominan	Contoh ulasan
System quality (0,5618; 0,9000)	T1 validasi email dan OTP T2 pemilihan tanggal kedatangan T3 pembayaran dan *billing* T5 *login* dan akun	"kirim kode otp email tidak kunjung dikirim" "sudah bayar tp ststusnya gak berubah2" "tidak bisa login. Padahal akun nya baru daftar"
Information quality (0,5163; 0,9429)	T7 *server error* T8 galat berulang T11 unggah foto KTP T12 *request timeout*	"setiap habis input data permohonan pasti *error* dan hilang" "tidak bisa diakses, selalu muncul tulisan *request time out*"

Dimensi (Cv; keragaman)	Topik dominan	Contoh ulasan
Service quality (0,4693; 1,0000)	T14 *maintenance* tanpa kepastian T15 kuota dan dugaan percaloan T18 pembukaan kuota T17, T19 apresiasi	"baru install langsung kena info under maintenance" "Aplikasinya sangat membantu dan tidak ada kendala saat mendaftar"
User satisfaction (0,4448; 0,9714)	T22 jadwal penuh T23 *request timeout* T24, T27 aplikasi mempersulit T25 pelayanan ramah	"Susah digunakan. *Request timeout* terus" "Sangat lelet bukan mempermudah malah merasa sulit"

Catatan: T = nomor topik BERTopic. Sumber: olah data peneliti (2026).

Topik-topik tersebut selanjutnya dipetakan ke dalam tiga atribut *agile evaluation* A.T. Kearney, sebagaimana disajikan pada Tabel 6. Pada atribut *e-government*, keluhan pengguna terkonsentrasi pada tiga titik layanan utama, yaitu akses akun, penjadwalan kedatangan, dan pembayaran. Gangguan seperti kegagalan penerimaan OTP, layar putih setelah proses masuk, kuota yang sudah penuh ketika menu jadwal baru dapat diakses, serta status pembayaran yang tidak berubah menghambat tercapainya tujuan digitalisasi antrean yang terbuka, akuntabel, dan efisien (Prameswari et al., 2023). Keluhan tersebut juga sejalan dengan temuan penelitian terdahulu mengenai galat sistem dan kendala dalam mengunggah dokumen (Pamadi et al., 2022; Ramadhani et al., 2025). Pada atribut *customer service*, pengguna mengapresiasi kemudahan pengajuan secara daring serta keramahan petugas. Namun, kekecewaan muncul ketika permohonan tidak diproses meskipun pembayaran telah dilakukan. Sementara itu, pada atribut *performance management*, keluhan pengguna terkonsentrasi pada kinerja aplikasi secara berkelanjutan, yaitu gangguan yang berulang, kinerja yang lambat, dan pemeliharaan tanpa kepastian waktu. Rating rendah pada klaster ini dibaca sebagai indikator ringkas penilaian pengguna atas kinerja tersebut, yang bercampur dengan ekspresi kepuasan, bukan sebagai bukti kinerja instansi.

Tabel 6. Pemetaan topik dominan ke atribut *agile evaluation* A.T. Kearney

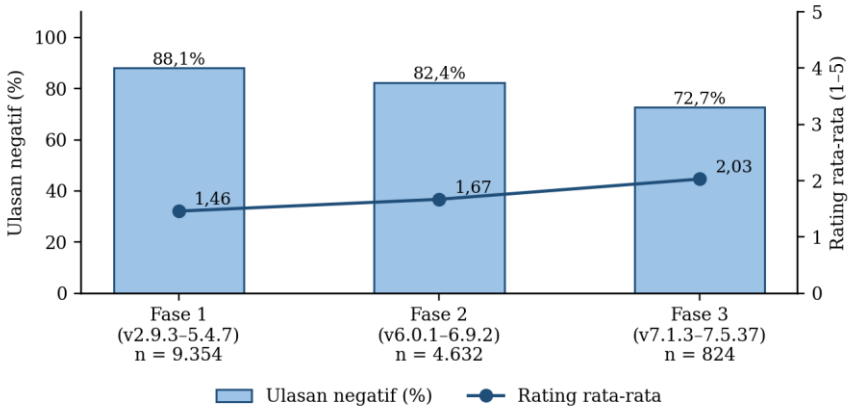
Atribut	Topik	Isu dominan
E-government	T1, T2, T3, T11,T22	Kegagalan OTP dan validasi email; jadwal dan kuota penuh; status pembayaran dan dana hangus, kendala unggah dokumen
Customer service	T17, T25, T74, T81	Apresiasi atas kemudahan pengajuan dan keramahan petugas; kekecewaan ketika permohonan tidak diproses setelah pembayaran
Performance management	T4, T6, T14, T16	Gangguan berulang pada aplikasi pemerintah (rating rata-rata 1,12); aplikasi lambat (1,68); *maintenance* tanpa kepastian; rating rendah (1,89).

Sumber: olah data peneliti (2026). T16 mencerminkan ekspresi penilaian umum pengguna dan dibaca sebagai indikator pendamping.

Responsivitas Ditjenim : Kapabilitas *Sensing* dan *Responding*

Dalam kerangka *agile government,* ulasan pengguna dipandang sebagai potensi sumber *sensing*, yaitu data terbuka yang dapat dimanfaatkan penyelenggara untuk mendeteksi permasalahan layanan. Penelitian ini tidak memperoleh bukti bahwa Ditjenim memantau, menyerap, atau memproses ulasan tersebut sebagai bagian dari proses kelembagaan; indikasi yang teramati terbatas pada balasan pengembang pada 14 dari 17.411 ulasan (0,08%). Dengan demikian, sensing yang dioperasikan instansi tidak dapat dipastikan dari jejak digital publik, dan analisis berikut membaca ulasan sebagai potensi sumber *sensing*. Dari total 17.411 ulasan, sebanyak 14.810 ulasan atau

85% memuat informasi versi aplikasi sehingga dapat dikaitkan dengan fase pembaruan. Sementara itu, 2.601 ulasan tidak mencantumkan informasi versi dan karena itu dikecualikan dari analisis antarversi. Sebagaimana ditunjukkan pada Gambar 3, rating rata-rata mengalami peningkatan, sedangkan persentase ulasan negatif menurun dari satu fase ke fase berikutnya. Namun, interpretasi terhadap pola ini perlu dilakukan secara hati-hati. Jumlah ulasan pada Fase 3 relatif sedikit, yaitu hanya 824 ulasan. Selain itu, pengelompokan dilakukan berdasarkan versi aplikasi, bukan berdasarkan waktu penulisan ulasan. Data yang digunakan juga tidak dapat membuktikan adanya hubungan sebab-akibat antara pembaruan aplikasi dan perubahan persepsi pengguna.



Gambar 3. Ulasan negatif dan rating rata-rata menurut fase pembaruan aplikasi

Catatan rilis menunjukkan bahwa pembaruan dari versi 6.0.1 hingga 7.5.37 terutama mencakup perbaikan bug, penyempurnaan tampilan metode pembayaran, perbaikan proses pemuatan kamera, pembaruan terkait pengenaan PNBP dan masa berlaku paspor, perbaikan alur penggantian paspor serta penjadwalan ulang, dan penyesuaian integrasi proses masuk (*login*) serta verifikasi NIK. Namun, sebagian besar catatan rilis hanya menyebutkan "*bug fixing*" tanpa menjelaskan perbaikan secara rinci. Kronologi ulasan sampel dan respons Ditjenim disajikan pada Tabel 7, rekap empat indikator responsivitas pada Tabel 8, dan jeda waktu divisualisasikan pada Gambar 4.

Tabel 7. Kronologi ulasan sampel dan respons Ditjenim terhadap klaster keluhan

Klaster keluhan	Ulasan sampel (versi/tanggal)	Respons yang tercatat
E-government		
T1 email dan OTP	5.2.1 / 24-10-2022 6.1.0 / 20-01-2024 7.2.0 / 26-08-2025	Catatan rilis umum (*bug fixing*, v7.4.11–7.5.0) v7.5.32 (28-07-2026): penyesuaian integrasi API *login*; v7.5.37 (14-09-2026): penyesuaian verifikasi NIK Tidak ada yang menyebut OTP atau verifikasi *email* secara eksplisit
T2, T22 jadwal dan kuota	5.4.7 / 01-09-2023 6.7.0 / 18-02-2025 7.3.10 / 06-01-2026	Rilis 10-09-2023, diumumkan Humas 26-09-2023: memilih kantor lain bila kuota tersedia, memilih hari dan jam, mengubah jadwal sekali pada H-1 Panduan *reschedule* di Instagram (13-11-2025) Tidak menyebut penambahan kuota

Klaster keluhan	Ulasan sampel (versi/tanggal)	Respons yang tercatat
T3 pembayaran	6.3.0 / 23-10-2024 6.4.0 / 04-11-2024	v6.6.2 (02-12-2024): perubahan tampilan metode pembayaran Tidak menyebut kendala *virtual account* secara eksplisit
T11 unggah dokumen (kamera)	6.2.4 / 01-10-2024 6.3.0 / 29-10-2024 6.4.0 / 23-11-2024	v6.6.3 (05-12-2024): perbaikan *camera loading* Tidak menyebut aplikasi keluar saat unggah dokumen secara eksplisit
Customer service		
T81 pelayanan buruk dan dokumen tambahan	5.4.7 / 14-07-2023 6.1.0 / 10-06-2024	Berita resmi 05-10-2022 dan 03-03-2025; informasi persyaratan tujuan bekerja (24-10-2022); FAQ; kanal *live chat* dan pengaduan Menjelaskan kewenangan petugas meminta dokumen tambahan dan ketentuan PNBP; tidak membahas "SK", biaya, dan pengembalian dokumen
T14 *maintenance* tanpa kepastian	6.1.0 / 30-01-2024	Sebelum sampel: permohonan maaf dan arahan ke *livechat* (siaran pers 08-04-2022; X dan Instagram Februari 2023) Setelah sampel: v6.6.3 (05-12-2024), perbaikan kamera; v6.7.0 (16-12-2024), pembaruan PNBP dan masa berlaku; siaran pers 31-12-2024, pemeliharaan berbatas waktu (peralihan kode kementerian, bukan tanggapan langsung)
Performance management		
T4 gangguan berulang aplikasi pemerintah	5.4.7 / 03-07-2023	Unggahan X (20-06-2024) tentang gangguan PDN; unggahan Instagram (28-06-2024): pemulihan, *Application Recovery Plan*, Satgas, 60% pada 27-06 dan 100% pada 28-06 Catatan rilis v6.0.1–7.5.37; LkjIP dan IKM; tidak ditemukan *balanced scorecard*
T6 aplikasi lambat	5.2.1 / 07-03-2023	Berita pemutakhiran (26-03-2023): cek kuota dan layanan percepatan, tidak terkait kecepatan Siaran pers kinerja kuartal I (22-05-2023): fokus penyempurnaan aplikasi v7.5.37 (14-09-2026): "optimisasi performa"
T16 rating rendah	5.1.0 / 01-02-2023	Unggahan X (03-02-2023) dan Instagram (05-02-2023): informasi pemulihan tanpa estimasi waktu Berita pemutakhiran (26-03-2023); siaran pers kinerja kuartal I (22-05-2023) IKM/LkjIP (3,88/4 pada April–Juni 2024); tidak ada pemantauan rating *Play Store*

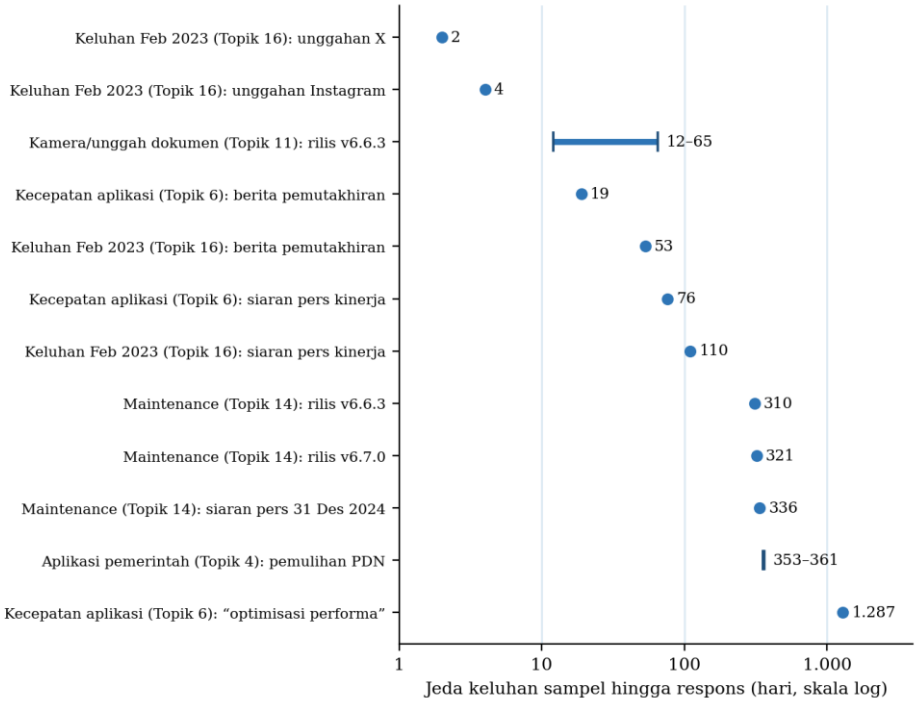
Catatan: Respons yang terbit sebelum ulasan sampel dicatat, tetapi tidak dihitung sebagai tindak lanjut.
Sumber: olah data peneliti (2026).

Tabel 8. Rekap indikator responsivitas Ditjenim per klaster keluhan

Klaster keluhan	Keberadaan	Kesesuaian	Jeda (hari)	Efektivitas
E-government				
T1 email dan OTP	Ada	Sebagian (lemah)	> 365 (rilis 2026)	Keluhan serupa masih muncul pada fase berikutnya
T2, T22 jadwal dan kuota	Ada	Sebagian (lemah)	25 (sampel 2023) 268 (sampel 2025) Sampel 2026: respons terbit lebih dulu	Keluhan serupa muncul kembali pada sampel 2025 dan 2026
T3 pembayaran	Ada	Sebagian (lemah)	28–40	Kendala pemilihan metode pembayaran masih muncul pada April–September 2026
T11 unggah dokumen (kamera)	Ada	Kamera: sesuai; aplikasi keluar saat unggah dokumen: sebagian	12–65	Ada indikasi penurunan: 2,15% (13/605) menjadi 0,28% (4/1.423) pada dua bulan sebelum dan sesudah rilis
Customer service				
T81 pelayanan buruk dan dokumen tambahan	Ada	Sampel 2023: sebagian Sampel 2024: sebagian (lemah)	598 (sampel 2023) 266 (sampel 2024) terhadap berita 03-03-2025	Belum terlihat; keluhan serupa terus bermunculan
T14 *maintenance* tanpa kepastian	Ada	Sebagian (lemah)	310 (v6.6.3) 321 (v6.7.0) 336 (siaran pers)	Belum terlihat; keluhan serupa terus bermunculan

Klaster keluhan	Keberadaan	Kesesuaian	Jeda (hari)	Efektivitas
Performance management				
T4 gangguan berulang aplikasi pemerintah	Ada	Sebagian (lemah)	353 (X) 361 (Instagram)	Rating rata-rata klaster 1,12 (tetap rendah) (tidak ada rilis spesifik)
T6 aplikasi lambat	Ada	Sebagian (lemah)	19 (berita) 76 (siaran pers) 1.287 (v7.5.37)	Rating rata-rata klaster 1,68 (tetap rendah) (tidak ada rilis spesifik)
T16 rating rendah	Ada	Sebagian (lemah)	2 (X) 4 (Instagram) 53 (berita) 110 (siaran pers)	Rating rata-rata klaster 1,89 (tetap rendah) (tidak ada rilis spesifik)

Catatan: Jeda = selisih hari dari ulasan sampel ke setiap respons yang tercatat setelah ulasan sampel. Efektivitas = perbandingan proporsi ulasan berkata kunci sejenis dua bulan sebelum dan sesudah rilis serta ada tidaknya keluhan serupa pada fase berikutnya; rating rata-rata klaster adalah rating per klaster, bukan tren antarfase. Penilaian terbatas pada jejak digital publik, sehingga "belum tercatat" dibaca sebagai keterbatasan bukti, bukan ketiadaan tindakan.
Sumber: olah data peneliti (2026).



Gambar 4. Jeda waktu antara keluhan sampel dan sebagian respons resmi Ditjenim

Berdasarkan empat indikator yang digunakan (Tabel 7), keberadaan respons telah terpenuhi pada ketiga atribut, tetapi tingkat kesesuaiannya umumnya masih berada pada kategori sebagian. Hanya perbaikan fitur kamera pada versi 6.6.3 yang secara langsung menjawab keluhan pengguna. Sementara itu, keluhan terkait OTP, kuota, dan *virtual account* tidak disebutkan secara eksplisit dalam catatan rilis. Ketepatan waktu respons juga bervariasi, mulai dari dua hingga empat hari untuk komunikasi pemulihan gangguan pada Februari 2023, hingga 1.287 hari untuk keluhan mengenai kecepatan aplikasi. Indikasi efektivitas respons masih terbatas. Proporsi ulasan yang menyebutkan kamera atau pengunggahan dokumen menurun dari 2,15% atau 13 dari 605 ulasan pada dua bulan sebelum rilis menjadi 0,28% atau 4 dari 1.423 ulasan pada dua bulan setelah rilis. Namun, perhitungan tersebut didasarkan pada kata kunci dengan jumlah kasus yang relatif kecil, sehingga hasilnya

sebaiknya dipahami sebagai indikasi awal, bukan sebagai bukti yang konklusif. Selain itu, keluhan mengenai pembayaran dan OTP masih muncul pada fase berikutnya. Rating rata-rata pada klaster *performance management* juga tetap rendah, yaitu 1,12, 1,68, dan 1,89.

Berbeda dengan sinyal pengguna pada klaster T4, T6, dan T14, kinerja instansi pada atribut ini dinilai dari pengukuran kinerja yang dilakukan penyelenggara. Pada atribut *performance management*, pengukuran kinerja telah tersedia pada tingkat organisasi melalui IKM, laporan kinerja, dan pelatihan teknologi informasi (Direktorat Jenderal Imigrasi, 2025). Namun, survei IKM mengukur layanan keimigrasian secara umum dan tidak memuat indikator khusus kinerja aplikasi. Tidak ditemukan pula rujukan terhadap rating atau ulasan *Play Store* maupun penerapan *balanced scorecard*. Pemantauan ulasan tampak sangat terbatas, karena hanya 14 dari 17.411 ulasan (0,08%) yang memperoleh balasan pengembang. Secara keseluruhan, ulasan pengguna merupakan sumber sensing yang kaya dan berpotensi besar untuk memantau kondisi layanan. Namun, respons Ditjenim yang dapat diamati melalui kanal publik tampak lebih menitikberatkan pada pemulihan layanan, penyampaian informasi mengenai gangguan, dan penjelasan kebijakan, dibandingkan dengan perbaikan yang spesifik maupun dilakukan secara tepat waktu. Berdasarkan jejak digital yang tersedia secara publik, responsivitas Ditjenim masih bersifat parsial dan cenderung reaktif. Kondisi ini belum sepenuhnya sejalan dengan prinsip agile evaluation, yang menekankan evaluasi yang cepat, berkelanjutan, dan adaptif (Mergel et al., 2018).

Ulasan pengguna dapat dipandang sebagai bentuk evaluasi bottom-up yang melengkapi evaluasi internal berbasis penilaian ahli, sebagaimana tercermin dalam model Dahmardeh dan Pourshahabi (2011). Sebagaimana telah dijelaskan pada bagian Pendahuluan, temuan ini memiliki implikasi terhadap transparansi dan akuntabilitas layanan publik. Respons yang teramati belum memberikan kepastian mengenai penanganan bagi pengguna yang mengalami kegagalan pembayaran atau OTP. Selain itu, pemeliharaan yang tidak disertai jadwal membuat warga sulit merencanakan pengurusan dokumen yang memiliki batas waktu.

Perbedaan Perspektif Pengguna dan Penyelenggara Layanan

Pengguna dan penyelenggara tampaknya menilai kinerja layanan dengan ukuran yang berbeda. Dari sisi pengguna, ulasan menunjukkan sentimen negatif yang dominan pada setiap fase pembaruan, dengan keluhan yang terkonsentrasi pada OTP, kuota, pembayaran, dan pemeliharaan sistem tanpa kepastian waktu. Sebaliknya, penyelenggara memantau kinerja melalui IKM dan laporan kinerja (Direktorat Jenderal Imigrasi, 2025), sementara komunikasi publiknya lebih menitikberatkan pada pemulihan gangguan, penjelasan kebijakan, dan catatan rilis umum (Tabel 6).

Perbedaan ini kemungkinan disebabkan oleh tiga hal. Pertama, IKM mengukur layanan keimigrasian secara umum, termasuk layanan tatap muka, sedangkan ulasan *Google Play Store* bersifat sukarela dan lebih sering ditulis ketika pengguna menghadapi kendala. Kedua, unit penilaiannya berbeda, yaitu layanan secara keseluruhan dibandingkan dengan aplikasi. Ketiga, ulasan muncul secara berkelanjutan, sementara survei dilakukan secara berkala. Oleh karena itu, kedua sumber sebaiknya dipahami sebagai saling melengkapi. Namun, jika pengukuran organisasi tidak mencakup indikator khusus aplikasi, keluhan pengguna belum tentu masuk ke dalam siklus evaluasi formal penyelenggara.

Implikasi Teoritis

Dari sisi ISSM, temuan penelitian ini menunjukkan bahwa model berguna sebagai peta keluhan, meskipun memiliki keterbatasan. Hampir seluruh ulasan terklasifikasi ke dalam *system quality* (97,55%) dan *user satisfaction*, sehingga kedua dimensi tersebut kurang mampu membedakan jenis keluhan, sementara *information quality* hamper seluruhnya bersentimen negatif. Penerapan ISSM dalam penelitian ini bersifat parsial (empat dari enam konstruk) dan deskriptif : hubungan antarkonstruk, yaitu kualitas sistem, informasi, dan layanan yang mendahului kepuasan (DeLone & McLean, 2003), tidak diuji dan tidak diklaim. Temuan ini mengindikasikan bahwa ISSM menjelaskan apa yang dikeluhkan pengguna, tetapi tidak memuat konstruk tentang bagaimana keluhan tersebut di tindaklanjuti penyelenggara. Oleh karena itu, penilaian layanan public digitak berbasis ISSM pelu dilengkapi dengan penelusuran respons terhadap keluhan agar siklus umpan balik dapat digambarkan secara utuh.

Dari perspektif *agile government*, kapasitas *short-term responsiveness*—yakni kemampuan merespons kebutuhan publik melalui pilihan dan suara (*voice*) (Parker & Barlett, 2008, dalam Dahmardeh & Pourshahabi, 2011) menjadi relevan karena suara pengguna tersedia dalam jumlah besar. Berdasarkan bukti yang dapat diamati, penelitian ini mengajukan satu proposisi: pada layanan publik digital, ketersediaan umpan balik publik yang terbuka dan melimpah tidak otomatis diikuti oleh respons terdokumentasi yang relevan dengan keluhan pengguna maupun tepat waktu. Kondisi ini menimbulkan ketidakselarasan antara umpan balik yang diterima dan respons yang diberikan. Proposisi tersebut didukung oleh empat temuan. Pertama, balasan pengembang hanya muncul pada 14 dari 17.411 ulasan, atau sekitar 0,08%. Kedua, hanya perbaikan kamera pada versi 6.6.3 yang secara langsung menjawab keluhan pengguna, sedangkan keluhan mengenai OTP, kuota, dan virtual account tidak disebutkan secara eksplisit dalam catatan rilis. Ketiga, jeda antara ulasan dan respons berkisar dari dua hingga 1.287 hari. Keempat, keluhan dengan karakteristik serupa masih muncul pada fase berikutnya, sebagaimana ditunjukkan pada Tabel 7 dan 8.

Secara teoretis, proposisi ini memperkaya kerangka *agile evaluation* dengan memisahkan ketersediaan umpan balik sebagai potensi sumber sensing dari sensing yang benar-benar dijalankan instansi. Pemisahan ini juga membedakan respons yang teramati secara publik dari tindak lanjut internal. Dengan demikian, evaluasi dapat dilakukan dari luar organisasi menggunakan bukti publik tanpa harus mengasumsikan adanya proses internal tertentu. Namun, proposisi ini bersifat deskriptif dan dibatasi pada satu kasus, sehingga perlu diuji lebih lanjut pada layanan publik digital lain serta dengan data internal penyelenggara. Penerapan atribut Kearney terhadap ulasan publik merupakan bentuk adaptasi karena model tersebut pada awalnya dirancang untuk penilaian internal oleh pakar (Dahmardeh & Pourshahabi, 2011), sementara penelitian ini hanya menelaah jejak digital yang tersedia secara publik. Oleh karena itu, kategori "belum tercatat" tidak serta-merta berarti tindakan tidak dilakukan. Penilaian responsivitas dalam penelitian ini terbatas pada apa yang dapat diamati oleh warga melalui sumber publik.

Keterbatasan Penelitian

Pertama, uji reliabilitas pelabelan hanya dilakukan pada 525 ulasan sampel, bukan pada seluruh korpus. Selain itu, akurasi model untuk dimensi *system quality* sebesar 70% dan *user satisfaction* sebesar 67% tergolong relatif rendah, sehingga proporsi pada kedua dimensi tersebut perlu dipahami sebagai estimasi. Kedua,

sentimen ditentukan berdasarkan aturan rating, sehingga kesepakatan yang sempurna pada pelabelan sentimen tidak mencerminkan penilaian yang dilakukan secara independen. Ketiga, analisis antarversi menggunakan jumlah ulasan yang tidak seimbang antar fase dan tidak dapat membuktikan adanya hubungan sebab-akibat. Keempat, penilaian responsivitas hanya mengandalkan jejak digital yang tersedia secara publik, sehingga tindak lanjut internal yang tidak dipublikasikan tidak dapat ditangkap dalam analisis dan kategori "belum tercatat" tidak berarti tidak ada tindakan. Kelima, data penelitian terbatas pada ulasan pengguna Android yang ditulis secara sukarela, sehingga pengalaman pengguna *iOS* dan pengguna yang tidak menulis ulasan tidak terwakili. Selain itu, dimensi *use/intention to use* dan *net benefits* tidak dioperasionalkan dalam penelitian ini.

KESIMPULAN

Penelitian terhadap 17.411 ulasan M-Paspor menunjukkan bahwa sentimen pengulas didominasi negatif. Sentimen ini ditentukan dari rating ulasan Android yang ditulis secara sukarela, sehingga mencerminkan pengguna yang menulis ulasan dan tidak mewakili kepuasan seluruh pengguna M-Paspor. Keluhan utama berfokus pada kualitas sistem dan kepuasan pengguna, terutama kegagalan OTP, kuota atau jadwal penuh, masalah pembayaran, timeout, serta pemeliharaan sistem tanpa informasi yang jelas, sedangkan kualitas layanan relatif lebih baik dibanding dimensi lainnya. Keluhan tersebut lebih berkaitan dengan ketidakstabilan layanan pada tahap krusial seperti akses akun, penjadwalan, dan pembayaran, serta minimnya komunikasi saat gangguan terjadi. Berdasarkan jejak digital publik, respons Ditjen Imigrasi yang teramati pada ketiga atribut dinilai sebagian dan cenderung reaktif: hanya perbaikan fitur kamera yang secara langsung menjawab keluhan, jeda respons berkisar dari dua hingga 1.287 hari, dan balasan pengembang hanya muncul pada 14 dari 17.411 ulasan (0,08%). Karena bukti respons hanya berasal dari sumber publik, temuan ini menggambarkan respons yang dapat diamati warga dan tidak membuktikan bahwa tindak lanjut internal tidak ada. Masukan pengguna yang tersedia dalam jumlah besar merupakan potensi sumber *sensing*, tetapi respons yang teramati secara publik belum terarah, spesifik, dan tepat waktu. Secara konseptual, penelitian ini menunjukkan bahwa umpan balik publik yang melimpah dan respons yang terdokumentasi dapat tidak selaras, sehingga evaluasi layanan digital berbasis ulasan perlu mencakup penelusuran respons, bukan hanya pemetaan keluhan. Empat konstruk ISSM digunakan untuk memetakan keluhan, dan pemetaan tersebut dibandingkan dengan respons yang teramati melalui kerangka agile evaluation. Temuan ini bersifat deskriptif dan tidak menunjukkan bahwa ketidakselarasan tersebut menyebabkan ketidakpuasan pengguna maupun bahwa pembaruan aplikasi mengubah persepsi pengguna.

Penelitian selanjutnya disarankan menggunakan sumber data yang lebih beragam, seperti ulasan pengguna di iOS atau platform lainnya. Data jejak digital tersebut dapat diperkuat melalui wawancara maupun data internal penyelenggara layanan. Selain itu, analisis deret waktu antarversi aplikasi dapat digunakan untuk menguji hubungan antara respons instansi dan perubahan sentimen pengguna, sementara hubungan antardimensi ISSM perlu diuji secara statistik menggunakan data berlabel dalam jumlah yang lebih besar.

REKOMENDASI

Bagi Ditjen Imigrasi, beberapa langkah yang disarankan meliputi: (1) menyusun catatan rilis yang menjelaskan secara spesifik permasalahan yang telah diperbaiki, misalnya terkait OTP, pembayaran, atau kuota jadwal; (2) mengumumkan pemeliharaan sistem secara proaktif dengan mencantumkan jadwal, penyebab gangguan, dan estimasi waktu pemulihan; (3) melakukan pemantauan serta memberikan balasan terhadap ulasan pengguna secara sistematis, dengan memprioritaskan topik yang paling dominan; (4) menyajikan persyaratan dokumen dan ketentuan biaya secara lebih jelas di dalam aplikasi untuk mencegah ketidaksesuaian antara proses pendaftaran daring dan verifikasi di lokasi; serta (5) menjadikan rating dan ulasan aplikasi sebagai masukan dalam evaluasi kinerja, misalnya dengan menambahkan unsur penilaian aplikasi ke dalam survei kepuasan masyarakat. Bagi penelitian selanjutnya, disarankan untuk memperluas sumber data dengan menyertakan ulasan dari *iOS App Store* atau platform lain, melengkapi jejak digital melalui wawancara atau data internal organisasi, serta memperpanjang rentang waktu penelitian menggunakan analisis deret waktu antarversi guna menguji keterkaitan antara respons instansi dan perubahan sentimen pengguna. Selain itu, penambahan data latih yang telah diberi label juga diperlukan untuk meningkatkan akurasi klasifikasi multi-label.

REFERENSI

Abidin, Zahra, N. S., & Yusuf, A. (2023). Analisis Aplikasi Kredivo Berdasarkan Information System Succes Model (ISSM). *Jurnal Maksipreneur: Manajemen, Koperasi, Dan Entrepreneurship*, *12*(2), 602–616. https://doi.org/10.30588/jmp.v12i2.1193

Agustina, N., & Sutinah, E. (2019). Model Delone dan McLean Untuk Menguji Kesuksesan Aplikasi Mobile Penerimaan Mahasiswa Baru. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, *3*(2), 76–82. https://doi.org/10.30743/infotekjar.v3i2.1008

Amalia, L., & Azizah, A. H. (2022). Evaluasi Kesuksesan Penerapan Sistem Elektronik Kinerja (E-Kinerja) Menggunakan Enhanced Information System Success Model di Kecamatan Benda Tangerang. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *7*(3), 192–210. https://doi.org/10.14421/jiska.2022.7.3.192-210

Anastasopoulos, L. J., & Whitford, A. B. (2019). Machine learning for public administration research, with application to organizational reputation. *Journal of Public Administration Research and Theory*, *29*(3), 491–510. https://doi.org/10.1093/jopart/muy060

Anisa, P. N. K. S., & Dewi, N. L. Y. (2024). Tingkat Keberhasilan Implementasi Aplikasi M-. *Jurnal Astina Mandiri*, *3*(November), 334–339.

Bahri, S., Lumbanraja, P., & Gusrini Rahmadani, V. (2025). Pengaruh Kualitas Layanan Digital (E-Servqual) Terhadap Kepuasan Pengguna Aplikasi M-Paspor (Studi Kasus Kantor Imigrasi Medan). *Jurnal Manajemen Stratejik Dan Simulasi Bisnis (JMASSBI)*, *6*(2), 35–50. https://doi.org/10.25077/mssb.6.2.35-50.2025

Chatfield, A. T., & Reddick, C. G. (2017). Customer agility and responsiveness through big data analytics for public value creation : A case study of Houston 311 on-demand services. *Government Information Quarterly*, 2–10.

https://doi.org/10.1016/j.giq.2017.11.002

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(37), 37–46. https://doi.org/10.1177/001316446002000104

Dahmardeh, N., & Pourshahabi, V. (2011). Agility evaluation in public sector using fuzzy logic. *Iranian Journal of Fuzzy Systems*, *8*(3), 95–111.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, *19*(4), 9–30. https://doi.org/10.1080/07421222.2003.11045748

Direktorat Jenderal Imigrasi. (2022, 23 Januari). *Aplikasi M-Paspor Siap Digunakan di Seluruh Indonesia Mulai 27 Januari 2022*. https://imigrasi.go.id/berita/2022/01/23/aplikasi-m-paspor-siap-digunakan-di-seluruh-indonesia-mulai-27-januari-2022

Direktorat Jenderal Imigrasi. (2025). *Laporan Kinerja Direktorat Jenderal Imigrasi Tahun 2024*. Kementerian Imigrasi dan Pemasyarakatan Republik Indonesia.

Direktorat Jenderal Imigrasi. (2026). *Imigrasi di Era Yuldi Yusman: Rekor PNBP dan Penguatan Penegakan Hukum* (Siaran Pers Nomor SP/IMI/03/2026/06).

Dwicahyo, K., & Indah Ratnasari, C. (2023). Perbandingan Metode Web Scraping Dalam Pengambilan Data: Kajian Literatur. *Automata*, *4*(2), 200–205.

Ernawati, M., Hermaliani, E. H., & Sulistyowati, D. N. (2021). Penerapan DeLone and McLean Model untuk Mengukur Kesuksesan Aplikasi Akademik Mahasiswa Berbasis Mobile. *IKRA-ITH Informatika Jurnal Komputer Dan Informatika*, *5*(1), 58–67.

Google Play Store. (2026). *M-Paspor* [Aplikasi seluler]. https://play.google.com/store/apps/details?id=id.go.imigrasi.paspor_online

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. http://arxiv.org/abs/2203.05794

Harefa, K. A., & Zebua, D. (2026). Transformasi Digital Dan Dampaknya Terhadap Kinerja Organisasi Sektor Publik Di Indonesia. *JIMAKUN: Jurnal Ilmu Manajemen Dan Akuntansi Nusantara*, *2*(1), 237–248.

He, H., & Xia, R. (2018). Joint binary neural network for multi-label learning with applications to emotion classification. *arXiv*. https://doi.org/10.48550/arXiv.1802.00891

Kasim, B., Ferriswara, D., & Haryati, E. (2025). International Journal of Social Science and Digital Transformation In Public Services : A Literature Review On Bureaucratic Innovation And Adaptation In The Era Of Digital Government. *International Journal of Social Science and Humanity*, 0–5.

Kementerian PANRB. (2017). *Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 13 Tahun 2017 tentang Pedoman Penyelenggaraan Sistem Informasi Pelayanan Publik Nasional* (Berita Negara Republik Indonesia Tahun 2017 Nomor 707). https://peraturan.bpk.go.id/Download/123384/PERMENPAN%20NOMOR%2013%20TAHUN%202017.pdf

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data Data for Categorical of Observer Agreement The Measurement. *International Biometric Society*, *33*(1), 159–174.

Liang, S., Siahaan, M., & Jocelyn, J. (2024). Analisis Kesuksesan Aplikasi M-Paspor di Kota Batam dengan Menggunakan Model Delone dan Mclean. *Jurnal Sistem*

Informasi Bisnis, *14*(1), 38–45. https://doi.org/10.21456/vol14iss1pp38-45

Lim, W. M. (2025). What Is Qualitative Research? An Overview and Guidelines. *Australasian Marketing Journal*, *33*(2), 199–229. https://doi.org/10.1177/14413582241264619

Maheri, R., Salisah, F. N., Muttakin, F., & Megawati, M. (2025). Analisis Sentimen Ulasan Aplikasi M-Paspor Menggunakan Naive Bayes dan Support Vector Machine. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, *10*(1), 448–458.

Mergel, I., Ganapati, S., & Whitford, A. B. (2021). Agile: A new way of governing. Public Administration Review, 81(1), 161–165. https://doi.org/10.1111/puar.13202

Mergel, I., Gong, Y., & Bertot, J. (2018). Agile government: Systematic literature review and future research. *Government Information Quarterly*, *35*(2), 291–298. https://doi.org/10.1016/j.giq.2018.04.003

Nainggolan, R. R. E., & Aqil, M. H. (2023). Analisis Faktor-Faktor yang Mempengaruhi Kepuasan Pengguna Aplikasi Pemerintah Kota Pagar Alam. *Jurnal Teknologi Dan Komunikasi Pemerintahan*, *5*(2), 229–249.

Nusandari, K. D., Widayanti, R., Achmad, Y. F., Azizah, A. H., & Santoso, N. A. (2022). Analisis Kesuksesan Pengguna Tangerang Live menggunakan Information System Success Model (ISSM). *Jurnal MENTARI: Manajemen, Pendidikan Dan Teknologi Informasi*, *1*(1), 77–88. https://doi.org/10.34306/mentari.v1i1.131

OECD. (2014). Recommendation of the Council on Digital Government Strategies. In *Public Governance and Territorial Development Directorate: July*.

Pamadi, J. D. P., Dewi, D. S. K., & DJ, E. W. (2022). Evaluasi Implementasi E-government di Kantor Imigrasi Kelas II NON Tempat Pemeriksaan Imigrasi Ponogoro. *Jurnal MODERAT*, *8*(4), 756–768.

Prameswari, A., Zakaria, S., & Centia, S. (2023). Pelayanan Publik Berbasis Electronic Government Melalui Penerapan Aplikasi Mobile Paspor (M-Paspor) di Kantor Imigrasi Kelas I TPI Bandung Tahun 2022. *Jurnal Administrasi Pemerintahan (JANITRA)*, *3*(2), 256–262.

Ramadhani, O. A., Lutfiyah, K. N., Tupen, D. L., Adzzahra, F. C., & Fatkhuri. (2025). Analisis Kepuasan Pengguna terhadap Aplikasi M-Paspor dalam Perspektif Implementasi E-Government di Indonesia. *Jurnal ISO: Jurnal Ilmu Sosial, Politik Dan Humaniora*, *5*(1), 12. https://doi.org/10.53697/iso.v5i1.2622

Republik Indonesia. (2018). *Peraturan Presiden Republik Indonesia Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik* (Lembaran Negara Republik Indonesia Tahun 2018 Nomor 182). https://peraturan.go.id/files/ps95-2018.pdf

Riffe, D., Lacy, S., Watson, B. R., Lacy, S., & Fico, F. G. (2024). Analyzing Media Messages Using Quantitative Content Analysis in Research. In *Analyzing Media Messages* (Fifth). Routledge. https://doi.org/10.4324/9781410613424

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399–408). ACM. https://doi.org/10.1145/2684822.2685324

Syam, A. A., M, G. H., Salim, A., Surianto, D. F., & B, M. F. (2024). Analisis Teknik Preprocessing Pada Sentimen Masyarakat Terkait Konflik ISRAEL-PALESTINA Menggunakan Support Vector Machine. *Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika*, *9*(3), 1464–1472.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer

program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Varastika, D., Wijaya, M. E., Aurelie, M. S. C., Handayani, P. W., & Fitriani, H. (2025). Mobile passport success in indonesia based on the DeLone–McLean information systems success model. *Telematics and Informatics Reports*, *20*, 2–12. https://doi.org/10.1016/j.teler.2025.100251

Wiranti, N. E., & Frinaldi, A. (2023). Meningkatkan Efisiensi Pelayanan Publik dengan Teknologi di Era Digital. *JIM: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, *8*(2), 748–754. https://doi.org/https://doi.org/10.24815/jimps.v8i2.24833

Wongso, W., Setiawan, D. S., Limcorn, S., & Joyoadikusumo, A. (2025). NusaBERT: Teaching IndoBERT to be Multilingual and Multicultural. *Proceedings - International Conference on Computational Linguistics, COLING*, 10–26.