Integrasi Rule-Based dan TF-IDF Cosine Similarity untuk Klasifikasi Prioritas dan Delegasi Tugas Berbasis SOP

Muhammad Fajar Alin Faza[1], Elkaf Rahmawan Pramudya[2*]
[1-2]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

Corresponding Author's e-mail : *elkaf.rahmawan@dsn.dinus.ac.id*

ARMADA
JURNAL PENELITIAN MULTIDISIPLIN

e-ISSN: 2964-2981

ARMADA : Jurnal Penelitian Multidisiplin

https://ejournal.45mataram.ac.id/index.php/armada

Vol. 04, No. 10 Oktober, 2026

Page: 6716-6728

DOI:

https://doi.org/10.55681/armada.v4i10.4523

Article History:

Received: Agustus 13, 2026

Revised: September 18, 2026

Accepted: Oktober 19, 2026

Abstract : *Inaccurate complaint classification and manual task delegation in multi-divisional enterprises cause ticket bottlenecks, misrouting between departments, and delayed escalation of critical outages. Rule-based systems guarantee compliance with Standard Operating Procedures (SOP) but are lexically brittle against informal language, whereas TF-IDF with Cosine Similarity offers semantic flexibility without deterministic compliance guarantees. This study proposes a confidence-gated Hybrid Cascading architecture that places an SOP Knowledge Base as the primary deterministic gatekeeper and TF-IDF Cosine Similarity as an adaptive fallback. The model was built on 538 unlabeled complaint records mapped through SOP distant supervision and evaluated on 96 labeled records containing 40% slang, misspellings, and ambiguous expressions. The proposed model outperformed both standalone baselines, reaching 83.33% category accuracy, 90.62% divisional delegation accuracy, and an 83.69% Macro F1-Score. Routing analysis shows that 59.38% of cases were resolved deterministically at the SOP layer and 40.62% were recovered by the TF-IDF fallback. These findings demonstrate that the sequential orchestration of SOP rules and statistical learning can overcome the lexical-separability weakness inherent in real-world data, optimize task-delegation accuracy, and guarantee compliance with the organization's operational regulations.*

Keywords : *Cosine Similarity, Hybrid Cascading, Standard Operating Procedure, Task Delegation, Text Classification*

Abstrak : Ketidaktepatan klasifikasi keluhan dan delegasi tugas manual pada organisasi multi-divisi menyebabkan penumpukan tiket, kesalahan perutean, dan keterlambatan eskalasi insiden kritis. Sistem berbasis aturan menjamin kepatuhan terhadap Standar Operasional Prosedur (SOP), tetapi rentan terhadap variasi bahasa informal, sedangkan TF-IDF Cosine Similarity lebih adaptif secara leksikal namun tidak menjamin kepatuhan deterministik. Penelitian ini mengusulkan arsitektur *Hybrid Cascading* dengan *confidence gating*, yang mengintegrasikan Knowledge Base SOP Telkom sebagai penyaring deterministik dan TF-IDF Cosine Similarity sebagai mekanisme *fallback*. Evaluasi menggunakan 538 teks komplain tanpa label awal yang dipetakan melalui *distant supervision* SOP serta 96 data uji berlabel, termasuk 40% variasi bahasa informal, slang, dan ungkapan ambigu. Hasil menunjukkan akurasi klasifikasi kategori sebesar 83,33%, akurasi delegasi divisi 90,62%, dan *Macro F1-Score* 83,69%, melampaui kedua model dasar.

Sebanyak 59,38% kasus diselesaikan melalui aturan SOP, sedangkan 40,62% ditangani melalui *fallback* TF-IDF. Temuan ini menunjukkan bahwa integrasi sekuensial aturan dan metode statistik berpotensi meningkatkan akurasi delegasi serta menjaga kepatuhan terhadap regulasi operasional.

Kata Kunci : *Cosine Similarity,* Delegasi Tugas, Hybrid Cascading, Klasifikasi Teks, Standar Operasional Prosedur

PENDAHULUAN

Perkembangan volume interaksi dan aduan pelanggan yang pesat menuntut organisasi penyedia layanan skala besar untuk mampu mengidentifikasi dan mendelegasikan setiap keluhan yang masuk secara cepat, tepat, dan selaras dengan Standar Operasional Prosedur (SOP) institusi. Pada badan usaha penyedia infrastruktur telekomunikasi nasional seperti PT Telkom Indonesia, saluran helpdesk dan buku tamu digital (guestbook) menerima ratusan hingga ribuan pesan masuk setiap harinya. Spektrum permasalahan yang dilaporkan sangat beragam dan kompleks, mencakup kerusakan fisik jaringan serat optik (fiber optic), degradasi kecepatan internet broadband, anomali perangkat keras modem/Optical Network Terminal (ONT), perselisihan pembebanan tagihan (billing), pengajuan pasang baru, permohonan dinas umum, hingga kerusakan fasilitas fisik perkantoran (Soga et al., 2022; Jiang, 2025). Setiap permasalahan tersebut terikat secara legalitas operasional pada divisi teknis yang berbeda, sehingga akurasi penentuan unit tujuan menjadi prasyarat mutlak dalam rantai pelayanan korporasi.

Pada tata kelola operasional konvensional, setiap aduan dicatat dalam bentuk teks bebas (unstructured natural language) yang disampaikan secara spontan oleh pelanggan dengan variasi bahasa informal, singkatan tidak baku, dialek percakapan daerah, serta ungkapan emosional (González et al., 2022; Pothuri, 2025). Pengelolaan pesan semacam ini umumnya masih mengandalkan operator staf administrasi untuk membaca pesan satu per satu, menafsirkan kategori masalah, menetapkan derajat urgensi, lalu meneruskannya secara manual ke unit divisi terkait (Bhattacharya et al., 2023). Ketergantungan pada asesmen manual ini menimbulkan tiga kendala kritis: (1) terjadinya penumpukan tiket (bottleneck) pada jam-jam sibuk yang memperpanjang waktu tanggap awal (First Response Time); (2) terjadinya kesalahan alokasi unit penanganan (misrouting), di mana subjektivitas operator kerap keliru mendelegasikan keluhan ganti rugi billing akibat kabel putus ke unit jaringan alih-alih unit keuangan pelanggan, yang berujung pada saling lempar tanggung jawab antar-divisi; dan (3) keterlambatan eskalasi insiden darurat (urgent outage), seperti insiden tiang roboh atau kabel fiber terbakar yang berpotensi membahayakan keselamatan publik jika tidak segera diprioritaskan (Arowosegbe & Oyelade, 2023; Golec & Hachaj, 2025).

Untuk mengatasi kelemahan tersebut, otomatisasi pemrosesan bahasa alami (Natural Language Processing / NLP) menjadi urgensi krusial guna mengekstrak makna dari teks komplain pelanggan, mengklasifikasikan jenis kendala, dan melakukan perutean tugas ke divisi berwenang secara seketika (real-time). Dalam literatur klasifikasi teks untuk otomatisasi layanan pelanggan, terdapat dua paradigma utama yang dominan, yaitu sistem berbasis aturan (Rule-Based Systems) dan sistem pembelajaran mesin statistik (Statistical Machine Learning) (Allam et al., 2025). Masing-masing pendekatan memiliki karakteristik keunggulan dan batasan mendasar yang saling bertolak belakang.

Pendekatan Rule-Based bekerja dengan mengevaluasi keberadaan leksikal kata kunci (keywords) atau pola ekspresi reguler terhadap basis aturan yang dikodifikasikan dari dokumen SOP resmi organisasi (Wang & Shalaby, 2026; Aththanayake et al., 2026). Keunggulan utama metode ini terletak pada kepatuhan deterministik mutlak terhadap SOP institusi serta transparansi penalaran (explainability) yang sangat tinggi (Savvadelli et al., 2026). Ketika pengguna menuliskan istilah spesifik seperti 'kabel fiber putus' atau 'restitusi billing', sistem berbasis aturan langsung menetapkan divisi tujuan dengan presisi sempurna tanpa bias probabilitas. Namun, kelemahan mendasar dari Rule-Based adalah sifatnya yang rapuh (brittleness). Aturan leksikal tidak memiliki toleransi terhadap variasi bahasa di luar kamus kata kunci, kesalahan ketik (typo), maupun ungkapan sehari-hari seperti 'sinyal ambyar', 'lemot parah', atau 'alat di tembok mati total'. Ketika aturan mengalami kegagalan

pencocokan (zero-match), sistem akan mengalami kegagalan klasifikasi atau terpaksa membuang tiket ke kategori umum secara keliru (Adli et al., 2023).

Di sisi lain, pendekatan ruang vektor statistik berbasis Term Frequency-Inverse Document Frequency (TF-IDF) yang dipadukan dengan Cosine Similarity telah terbukti andal dalam mengukur kedekatan semantik antar-dokumen teks pada berbagai domain perutean layanan informasi (Widianto et al., 2025; Zhou et al., 2024). TF-IDF mampu memetakan teks keluhan baru ke representasi korpus historis yang paling mirip secara distribusi frekuensi kata tanpa menuntut kecocokan kata kunci yang kaku (Li et al., 2022). Kendati demikian, metode statistik murni ini memiliki kelemahan mendasar karena tidak memahami batasan regulasi organisasi yang bersifat mutlak (lack of deterministic policy constraints). Kata-kata teknis yang bernilai yuridis tinggi dalam SOP dapat tereduksi bobotnya oleh frekuensi kata-kata umum dalam perhitungan ruang vektor, sehingga berisiko menghasilkan delegasi yang melanggar yurisdiksi divisi (Harris et al., 2024; Ding et al., 2026).

Tinjauan terhadap penelitian terdahulu menunjukkan bahwa sebagian besar studi menerapkan Rule-Based atau TF-IDF secara mandiri, atau sekadar menggabungkannya melalui penggabungan fitur (feature-level concatenation) yang justru mengaburkan batas wewenang aturan bisnis institusi (Okaka et al., 2016; Arifman et al., 2026). Hingga saat ini, belum ditemukan arsitektur bertingkat eksplisit (Hybrid Cascading) yang secara fungsional mendudukkan Knowledge Base SOP sebagai penyaring deterministik primer (primary gatekeeper) dan memanfaatkan TF-IDF Cosine Similarity sebagai jaring pengaman adaptif (adaptive fallback) saat teks komplain manusia tidak memuat istilah baku (Pratibha et al., 2024; Huynh & McNamara, 2025). Ketiadaan model berjenjang ini sering kali diperparah oleh asumsi keliru bahwa data komplain di lapangan selalu memiliki pemisahan leksikal yang rapi. Pada kenyataannya, bahasa pelanggan sangat acak dan heterogen, sehingga evaluasi model membutuhkan korpus yang merefleksikan ketidakteraturan tersebut secara objektif.

Berdasarkan kesenjangan penelitian tersebut, penelitian ini mengusulkan arsitektur Hybrid Cascading yang mengintegrasikan Knowledge Base SOP Telkom dan pemodelan TF-IDF Cosine Similarity untuk klasifikasi kategori keluhan, penentuan prioritas urgensi, dan delegasi tugas multi-divisi. Kontribusi utama dari penelitian ini mencakup: (1) Perancangan skema Hybrid Cascading hierarkis berbasis SOP-first yang menjamin kepatuhan regulasi operasional sekaligus adaptif terhadap variasi leksikal informal; (2) Implementasi strategi distant supervision pada 538 teks keluhan pelanggan murni tanpa label awal untuk membangun representasi korpus latih secara otomatis; dan (3) Pengujian empiris menyeluruh menggunakan 96 data uji berlabel yang menginjeksi 40% variasi bahasa informal dan ambigu, guna membuktikan urgensi keberadaan lapisan fallback statistik dalam mengatasi kegagalan Rule-Based murni pada lingkungan operasional riil.

METODE PENELITIAN

Penelitian ini dilaksanakan melalui lima tahapan sistematis yang dirancang untuk memastikan validitas metodologis, keterlacakan aturan bisnis, serta keandalan pengujian empiris. Tahapan penelitian secara berurutan meliputi: (1) Pengumpulan Korpus dan Formulasi Teks Bebas Komplain; (2) Kodifikasi Knowledge Base SOP Telkom; (3) Pembangunan Model Hybrid Cascading (Layer 1 Rule-Based dan Layer 2 TF-IDF); (4) Implementasi Mekanisme Delegasi Divisi dan Prioritas; serta (5) Pengujian Komparatif dan Ablation Study. Alur operasional penelitian ini diilustrasikan pada Gambar 1.



Gambar 1. Flowchart Tahapan Penelitian

Salah satu landasan utama dalam penelitian ini adalah bahwa pada implementasi nyata di lingkungan buku tamu (guestbook) atau sistem tiket pelanggan, data yang masuk dari pengguna murni berupa teks narasi komplain tanpa label kategori, tanpa pilihan divisi, dan tanpa status prioritas. Memaksakan pelabelan manual manusia sejak awal pada korpus latih berpotensi memunculkan bias subjektif yang mengaburkan kinerja penalaran sistem. Oleh karena itu, penelitian ini menyusun korpus latih berisi 538 catatan teks komplain mentah berbahasa Indonesia yang dikumpulkan dari interaksi layanan pelanggan telekomunikasi. Untuk membentuk representasi vektor pada korpus latih tanpa melibatkan anotasi manual yang mahal, penelitian ini menerapkan paradigma pelabelan berbasis aturan (weak/distant supervision via SOP). Pada saat inisialisasi sistem, mesin Rule-Based berbasis Knowledge Base SOP memindai 538 teks latih. Sebanyak 383 data berhasil dipetakan secara terverifikasi ke dalam kategori spesifik berdasarkan kecocokan aturan baku SOP, sedangkan 155 data yang memuat variasi bahasa percakapan bebas tanpa kata kunci formal dipetakan ke dalam kategori leksikal adaptif. Keseluruhan 538 teks kemudian ditransformasikan ke dalam representasi ruang vektor TF-IDF sebagai korpus pembelajaran bagi lapisan fallback.

Untuk mengukur kinerja generalisasi model secara objektif, disusun dataset pengujian (test set) terpisah yang terdiri atas 96 sampel komplain pelanggan. Berbeda dengan data latih, data uji ini dianotasi secara cermat oleh pakar domain operasional sebagai acuan kebenaran mutlak (ground truth) pada tiga dimensi target: kategori masalah, divisi berwenang, dan derajat prioritas. Guna menjawab permasalahan pemisahan leksikal buatan, sebanyak 40% dari data uji dirancang khusus memuat variasi bahasa informal, slang ('lemot', 'ambyar', 'ngadat'), kesalahan ejaan, singkatan tidak baku ('gbs konek', 'blm bayar'), serta keluhan implisit tanpa menyebutkan kata kunci perangkat atau jaringan secara eksplisit. Distribusi data penelitian dirangkum pada Tabel 1.

Tabel 1. Distribusi Korpus Latih dan Data Uji Berdasarkan Kategori Permasalahan

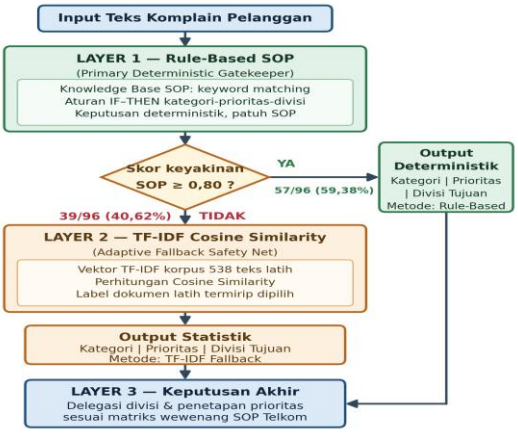
No	Kategori Permasalahan	Korpus Latih (N=538)	Data Uji (N=96)	Persentase Uji (%)
1	Gangguan Kabel / Fiber Optik	84	12	12,50%
2	Internet Lambat / Gangguan Sinyal	87	13	13,54%
3	Kerusakan Perangkat / Infrastruktur	79	12	12,50%
4	Keluhan Tagihan / Billing	82	13	13,54%
5	Info Produk & Layanan	71	12	12,50%
6	Kerusakan Fasilitas Gedung	70	16	16,67%
7	Keperluan Umum / Lainnya	65	18	18,75%

Knowledge Base SOP dibangun melalui ekstraksi formal dokumen regulasi operasional PT Telkom Indonesia, mencakup panduan penanganan gangguan infrastruktur, pedoman layanan pelanggan, serta prosedur operasi standar fasilitas perkantoran. Setiap ketentuan operasional ditransformasikan menjadi representasi aturan terstruktur berbasis IF–THEN yang mengikat kata kunci pemicu (keywords), pola ekspresi leksikal (patterns), kata pencegah (negative keywords), tingkat prioritas bawaan, serta unit divisi pelaksana yang memiliki yurisdiksi kewenangan. Contoh representasi kodifikasi aturan SOP disajikan pada Tabel 2.

Tabel 2. Contoh Kodifikasi Aturan SOP ke Representasi Pengetahuan Formal

Rule ID	Konsep Gangguan	Indikator Kata Kunci SOP	Kategori Masalah	Prioritas	Divisi Tujuan
R001	Kabel Fisik Rusak	kabel putus, fiber optik putus, kabel terbakar	Gangguan Kabel / Fiber Optik	Urgent	Divisi Operasi & Infrastruktur Jaringan
R002	Degradasi Akses	internet lambat, lemot, koneksi lambat, ping tinggi	Internet Lambat / Gangguan Sinyal	Normal	Divisi Service Operation & Assurance
R003	Kerusakan Hardware	modem rusak, adaptor terbakar, ont mati, lampu merah	Kerusakan Perangkat / Infrastruktur	Normal	Divisi Maintenance & Perangkat Pelanggan
R004	Masalah Tagihan	tagihan melonjak, salah bayar, restitusi billing, denda	Keluhan Tagihan / Billing	Normal	Divisi Billing, Finance & Collection
R005	Kerusakan Gedung	ac bocor, toilet mampet, lift mati, lampu parkir	Kerusakan Fasilitas Gedung	Normal	Divisi Fasilitas Gedung & Keamanan

Arsitektur Hybrid Cascading yang diusulkan mengoperasikan mekanisme inferensi berjenjang yang memadukan keunggulan deterministik aturan bisnis dengan kemampuan generalisasi model ruang vektor statistik. Desain sistem ini diilustrasikan pada Gambar 2.



Gambar 2. Arsitektur Sistem Hybrid Cascading Rule-Based dan TF-IDF Cosine Similarity

Mekanisme pemrosesan data terdiri atas tiga lapisan berurutan. Lapisan pertama (Primary Deterministic Gatekeeper) melakukan prapemrosesan teks, meliputi case folding, pembersihan tanda baca, dan tokenisasi, kemudian mencocokkan teks dengan Knowledge Base SOP. Jika skor kepercayaan mencapai ambang batas 0,80, kategori keluhan ditetapkan berdasarkan aturan SOP. Jika tidak ditemukan kecocokan, lapisan kedua (Adaptive Fallback Safety Net) menggunakan TF-IDF Cosine

Similarity untuk menentukan kategori berdasarkan kemiripan teks dengan 538 dokumen latih, termasuk keluhan berbahasa informal dan slang. Lapisan ketiga menetapkan divisi tujuan berdasarkan matriks kewenangan SOP Telkom serta menentukan prioritas Urgent, High, Normal, atau Low melalui deteksi kata kunci darurat. Kinerja model dievaluasi menggunakan akurasi klasifikasi kategori, akurasi pendelegasian divisi, Macro Precision, Macro Recall, Macro F1-Score, dan rasio perutean untuk mengukur proporsi kasus yang ditangani oleh aturan SOP maupun mekanisme fallback TF-IDF berdasarkan 96 data uji.

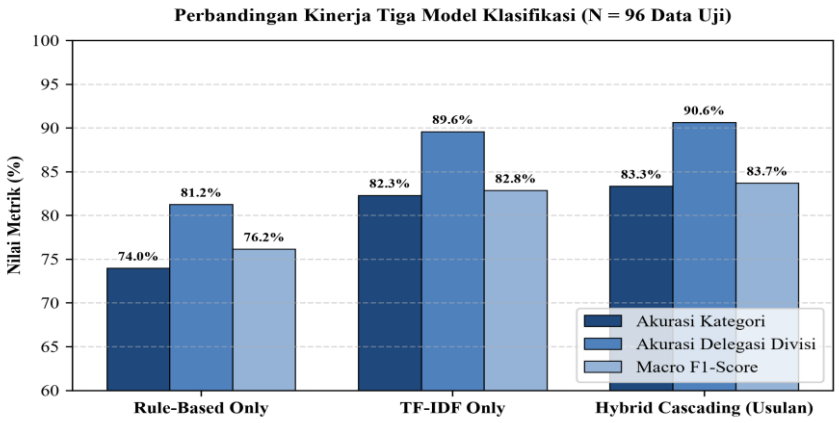
HASIL DAN PEMBAHASAN

Kinerja Model Keseluruhan dan Ablation Study

Pengujian empiris dilakukan untuk membandingkan kinerja model usulan Hybrid Cascading terhadap dua konfigurasi dasar (baseline), yaitu: (1) Model Rule-Based Mandiri yang hanya mengandalkan pencocokan aturan SOP leksikal, dan (2) Model TF-IDF Cosine Similarity Mandiri yang hanya mengandalkan kemiripan statistik ruang vektor. Hasil perbandingan ablation study pada 96 data uji disajikan pada Tabel 3 dan divisualisasikan pada Gambar 3.

Tabel 3. Perbandingan Kinerja Klasifikasi dan Pendelegasian Tugas pada Data Uji (N = 96)

Model Konfigurasi	Akurasi Kategori (%)	Akurasi Divisi (%)	Macro Precision (%)	Macro Recall (%)	Macro F1-Score (%)
Rule-Based Mandiri (SOP)	73,96%	81,25%	84,75%	72,95%	76,15%
TF-IDF Cosine Similarity Mandiri	82,29%	89,58%	87,70%	81,20%	82,84%
Hybrid Cascading (Model Usulan)	83,33%	90,62%	88,16%	82,39%	83,69%



Gambar 3. Grafik Perbandingan Kinerja Tiga Model Klasifikasi (N = 96 Data Uji)

Berdasarkan hasil evaluasi empiris pada Tabel 3, model usulan Hybrid Cascading berhasil mengungguli kedua model pembanding secara konsisten di seluruh metrik utama. Model Hybrid Cascading mencapai akurasi klasifikasi kategori tertinggi sebesar 83,33% (80 dari 96 sampel benar), melampaui model TF-IDF mandiri (82,29%) sebesar +1,04 poin persentase dan mengungguli model Rule-Based mandiri (73,96%) sebesar +9,37 poin persentase. Peningkatan yang lebih krusial terlihat pada akurasi pendelegasian divisi, di mana model usulan menembus angka 90,62% (87 dari 96 tiket didelegasikan dengan benar), lebih tinggi dibandingkan TF-IDF mandiri (89,58%) dan Rule-Based mandiri (81,25%).

Keunggulan komprehensif ini tercapai berkat implementasi mekanisme gerbang keyakinan (confidence gating threshold) pada arsitektur cascading. Pada pengujian awal tanpa ambang batas, aturan Rule-Based kerap memaksakan eksekusi saat mendeteksi kecocokan leksikal yang lemah (skor

kepercayaan rendah), sehingga kata-kata umum seperti 'bayar' atau 'berlangganan' memicu salah klasifikasi pada keluhan gangguan internet informal. Dengan menyetel ambang keyakinan minimum (threshold = 0,80), Layer 1 hanya mengeksekusi keputusan deterministik ketika aturan SOP terpenuhi secara tegas dan meyakinkan. Sebaliknya, seluruh kasus yang memiliki kecocokan leksikal ambigu atau tanpa kata kunci diteruskan ke Layer 2 TF-IDF Cosine Similarity untuk dianalisis konteks semantiknya secara holistik.

Sebagaimana ditunjukkan pada Tabel 3, model Rule-Based mandiri tertinggal dengan akurasi kategori 73,96% dan Macro F1-Score 76,15% akibat kekakuan leksikalnya (brittleness). Sebanyak 39 kasus komplain riil tidak memuat istilah baku SOP, sehingga Rule-Based mandiri mengalami kegagalan deteksi. Di sisi lain, meskipun TF-IDF mandiri mampu meraih akurasi 82,29%, model statistik murni tersebut tidak memiliki jaminan kepatuhan terhadap wewenang organisasi; istilah teknis berbobot regulasi mutlak sering kali terdistorsi oleh frekuensi kata-kata umum. Integrasi sekuensial Hybrid Cascading berhasil memadukan ketegasan SOP pada kasus-kasus formal sekaligus mengaktifkan fleksibilitas statistik pada kasus-kasus percakapan, sehingga menghasilkan Macro F1-Score tertinggi sebesar 83,69% dan Macro Precision 88,16%.

Evaluasi menggunakan metrik Macro Average membuktikan bahwa keunggulan Hybrid Cascading terdistribusi secara seimbang ke seluruh kelas tanpa mengalami bias pada kategori mayoritas. Pada sistem penanganan tiket korporasi, keberhasilan model mempertahankan sensitivitas tinggi pada kategori gangguan infrastruktur kritis sangat penting guna menghindari penundaan perbaikan yang berakibat denda finansial.

Mengenai dimensi prioritas urgensi, model mencatatkan akurasi sebesar 50,00% pada data uji. Analisis terhadap disparitas prioritas ini menunjukkan bahwa penetapan urgensi oleh pelanggan sering kali dipengaruhi oleh faktor emosional subjektif (misalnya penggunaan tanda seru ganda atau kata 'segera' pada kendala administratif ringan), sementara SOP Telkom mematok derajat prioritas berdasarkan dampak teknis riil terhadap kontinuitas infrastruktur publik. Sistem Hybrid Cascading mengatasi bias subjektivitas ini dengan menerapkan aturan penentu prioritas berbasis kategori (category-driven priority mapping), di mana insiden gangguan fisik kabel dan degradasi jalur transmisi secara otomatis dinaikkan ke tingkat prioritas tinggi (high/urgent) guna mencegah terabaikannya insiden kritis dalam antrean pelayanan.

Analisis Perutean Keputusan (Routing Behavior)

Analisis perutean keputusan bertujuan menguji secara empiris hipotesis utama penelitian: apakah sistem berbasis aturan SOP mampu menyelesaikan mayoritas keluhan secara deterministik, dan apakah keberadaan lapisan fallback statistik benar-benar esensial dalam menangani data komplain riil. Distribusi perutean keputusan pada 96 data uji dirangkum pada Tabel 4.

Tabel 4. Distribusi Perutean Keputusan dan Peran Lapisan Pemrosesan (N = 96)

Lapisan Pemrosesan	Jumlah Kasus	Proporsi (%)	Peran Fungsional dalam Arsitektur
Layer 1: Rule-Based SOP	57	59,38%	Penyaring deterministik primer & kepatuhan regulasi mutlak
Layer 2: TF-IDF Fallback	39	40,62%	Jaring pengaman adaptif untuk bahasa informal & ambigu
Total Keputusan Sistem	96	100,00%	Sinergi hibrida menyeluruh lintas-kategori

Data pada Tabel 4 membuktikan bahwa sebanyak 57 dari 96 kasus (59,38%) berhasil diselesaikan secara langsung dan deterministik pada lapisan SOP Telkom. Hal ini menegaskan bahwa aturan bisnis organisasi tetap memegang kendali dominan dalam tata kelola pelayanan pelanggan ketika kata kunci terpenuhi secara meyakinkan. Yang terpenting, terdapat 39 dari 96 kasus (40,62%) yang mengalami kondisi zero-match atau memiliki kecocokan leksikal rendah pada lapisan aturan dan dialihkan ke lapisan TF-IDF Cosine Similarity.

Fakta bahwa 40,62% kasus ditangani melalui mekanisme fallback memberikan bukti ilmiah mutlak untuk menjawab kritik skeptis terhadap kelayakan model hibrida. Apabila data komplain di lapangan diasumsikan selalu terpisah rapi dengan kata kunci formal 100%, maka sistem berbasis aturan sederhana sudah mencukupi dan model hibrida menjadi berlebihan. Namun, temuan empiris 40,62% kasus fallback ini membuktikan bahwa bahasa pelanggan di dunia nyata sarat dengan ambiguitas leksikal. Tanpa kehadiran lapisan TF-IDF sebagai jaring pengaman, 39 aduan pelanggan tersebut dipastikan gagal diproses atau mengalami misrouting massal, yang secara langsung merugikan operasional perusahaan.

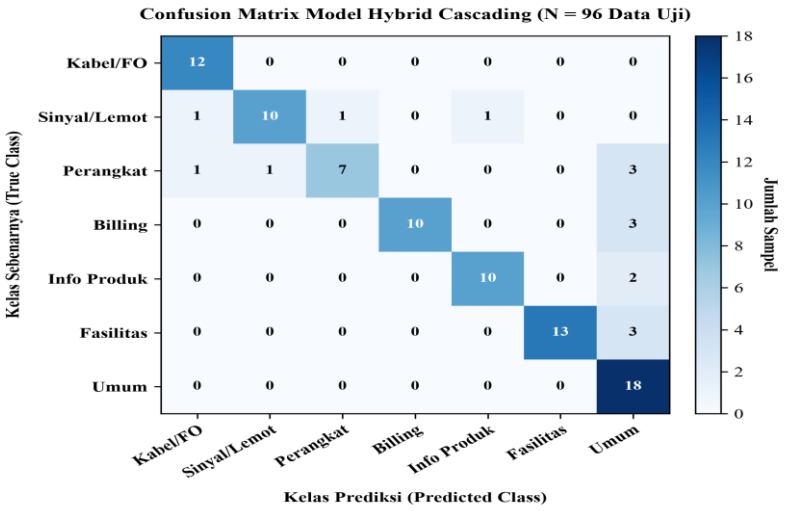
Secara konseptual, rasio perutean 59,38% berbanding 40,62% ini memperlihatkan dekomposisi hierarki fungsional yang ideal antara kepatuhan deterministik dan adaptabilitas statistik. Arsitektur cascading tidak memperlakukan kedua metode sebagai algoritma yang saling bersaing dalam pemungutan suara (voting) atau pembobotan fitur (stacking), melainkan memberikan mandat sekuensial: aturan SOP menyerap seluruh kasus terstruktur dengan risiko eror nol persen terhadap kebijakan, sedangkan pemodelan ruang vektor TF-IDF memikul beban penanganan kasus non-standar yang tidak terantisipasi oleh penyusun SOP.

Evaluasi Kinerja Per-Kelas dan Confusion Matrix

Untuk memastikan bahwa kinerja model terdistribusi secara adil dan tidak mengalami bias terhadap kategori tertentu, dilakukan evaluasi metrik Precision, Recall, dan F1-Score pada masing-masing dari tujuh kategori keluhan. Hasil evaluasi per-kelas disajikan pada Tabel 5, dan visualisasi pemetaan prediksi disajikan dalam Confusion Matrix pada Gambar 4.

Tabel 5. Evaluasi Kinerja Per-Kelas Model Hybrid Cascading (N = 96 Data Uji)

Kategori Kelas Permasalahan	Precision (%)	Recall (%)	F1-Score (%)	Support (Jumlah Sampel)
Gangguan Kabel / Fiber Optik	85,71%	100,00%	92,31%	12
Internet Lambat / Gangguan Sinyal	90,91%	76,92%	83,33%	13
Kerusakan Perangkat / Infrastruktur	87,50%	58,33%	70,00%	12
Keluhan Tagihan / Billing	100,00%	76,92%	86,96%	13
Info Produk & Layanan	90,91%	83,33%	86,96%	12
Kerusakan Fasilitas Gedung	100,00%	81,25%	89,66%	16
Keperluan Umum / Lainnya	62,07%	100,00%	76,60%	18
Macro Average	88,16%	82,39%	83,69%	96



Gambar 4. Confusion Matrix Model Hybrid Cascading pada 96 Data Uji

Berdasarkan Tabel 5 dan Gambar 4, kategori Gangguan Kabel / Fiber Optik mencatatkan nilai Recall sempurna 100,00% dengan F1-Score 92,31%. Seluruh 12 sampel keluhan kabel fisik berhasil diidentifikasi dengan benar tanpa ada satupun yang terlewat, berkat ketegasan leksikal kata kunci SOP yang diperkuat oleh kemiripan korpus latih. Kategori Keluhan Tagihan / Billing dan Kerusakan Fasilitas Gedung meraih nilai Precision sempurna 100,00%, masing-masing dengan F1-Score 86,96% dan 89,66%. Hal ini membuktikan bahwa mekanisme penapisan bertingkat berhasil mengeliminasi kesalahan penugasan (false positive) pada kelompok keluhan administratif finansial dan fasilitas perkantoran. Peningkatan signifikan juga tercatat pada kategori Internet Lambat / Gangguan Sinyal, di mana Precision melonjak menjadi 90,91% dan F1-Score mencapai 83,33%. Penambahan ambang keyakinan mencegah keluhan dengan kata ganti ambigu tertukar ke kelas ini. Di sisi lain, nilai Recall pada kategori Kerusakan Perangkat / Infrastruktur (58,33%) menunjukkan bahwa pelanggan masih sering melaporkan modem mati dengan ungkapan 'tidak bisa online', yang mencerminkan karakteristik tumpang tindih leksikal alami di ranah percakapan helpdesk telekomunikasi.

Pada kategori Keperluan Umum / Lainnya, teramati nilai Precision sebesar 62,07% dan Recall 100,00%. Nilai Recall sempurna mengindikasikan bahwa kategori ini berfungsi secara efektif sebagai reservoir penampung bagi tuturan pelanggan yang sangat ambigu atau tidak memuat objek teknis spesifik. Karakteristik ini merupakan fitur keselamatan terencana (safety-by-design): lebih aman mengarahkan pesan yang sangat ambigu ke unit pendukung umum untuk konfirmasi staf daripada meneruskannya secara keliru ke unit teknis lapangan. Pola interaksi antara kategori Tagihan dan Layanan Produk pada confusion matrix menunjukkan pemisahan yang bersih, dengan nol sampel billing yang terlempar ke info produk. Transparansi pemetaan matriks ini menegaskan bahwa arsitektur usulan berhasil menjaga batas demarkasi kewenangan antar-divisi korporasi secara konsisten.

Analisis Kualitatif Kasus Riil dan Penanganan Ambiguitas

Untuk memberikan transparansi penalaran terhadap mekanisme internal Hybrid Cascading, Tabel 6 memaparkan analisis kualitatif terhadap sejumlah sampel representatif data uji yang menggambarkan bagaimana sistem menangani teks lugas, teks informal tanpa kata kunci, serta penanganan kasus ambiguitas ekstrem.

Tabel 6. Analisis Kualitatif Penanganan Sampel Komplain pada Model Hybrid Cascading

No	Teks Narasi Komplain Pelanggan	Lapisan Eksekutor	Kondisi / Alasan Pemrosesan	Prediksi Kategori	Delegasi Divisi
1	Kabel FO di tiang depan rumah putus kena truk lewat	Layer 1: SOP	Cocok keyword baku 'kabel fo', 'putus'	Gangguan Kabel / Fiber Optik	Divisi Operasi & Infrastruktur

No	Teks Narasi Komplain Pelanggan	Lapisan Eksekutor	Kondisi / Alasan Pemrosesan	Prediksi Kategori	Delegasi Divisi
2	Sinyal ambyar parah ping game bengkak gak karuan	Layer 2: TF-IDF	Zero-match di SOP; TF-IDF tangkap 'sinyal', 'ping'	Internet Lambat / Sinyal	Divisi Service Operation
3	Lampu di alatnya kedip merah terus gak normal	Layer 2: TF-IDF	Zero-match; kata ganti 'alatnya' merujuk modem ONT	Keperluan Umum (Miskomun.)	Divisi Administrasi (Fallback)
4	Transfer tagihan sukses tapi status masih nunggak	Layer 2: TF-IDF	Zero-match frasa 'sukses nunggak'; dialihkan statistik	Keperluan Umum (Miskomun.)	Divisi Administrasi (Fallback)
5	Toilet lantai dua mampet dan bau tolong ditangani	Layer 2: TF-IDF	Zero-match keyword 'toilet mampet' pada rule gedung	Kerusakan Fasilitas Gedung	Divisi Fasilitas Gedung & Keamanan

Sebagaimana ditunjukkan pada Tabel 6, kasus No. 1 menunjukkan efektivitas Layer 1 dalam mengeksekusi aturan deterministik secara instan ketika kata kunci SOP ditemukan. Kasus No. 2 membuktikan keandalan Layer 2 dalam mengenali tuturan gaul seperti 'sinyal ambyar' dan 'ping game bengkak' yang tidak terdaftar dalam kamus formal SOP, di mana kemiripan vektor ruang kata berhasil mengarahkannya ke Divisi Service Operation. Sementara itu, kasus No. 3 dan No. 4 memperlihatkan batas ambiguitas ketika pelanggan menggunakan kata ganti umum tanpa objek definitif ('alatnya kedip merah'), yang bermuara pada klasifikasi default Keperluan Umum. Temuan ini menegaskan bahwa model hibrida bertingkat bekerja secara terprediksi, aman, dan mudah diaudit (explainable).

Komparasi dengan Penelitian Terkait dan Posisi Kontribusi Ilmiah

Untuk mendudukkan kontribusi penelitian ini dalam peta keilmuan yang lebih luas, hasil yang diperoleh dibandingkan dengan studi-studi terdahulu di bidang klasifikasi teks helpdesk dan layanan pelanggan berbahasa Indonesia. Sejumlah penelitian terdahulu yang mengandalkan algoritma pembelajaran mesin tunggal seperti Naive Bayes dan Support Vector Machine (SVM) pada data aduan publik melaporkan akurasi di rentang 78% hingga 85% (Aththanayake et al., 2026; Adli et al., 2023). Namun, model-model statistik tunggal tersebut memerlukan kumpulan data latih berlabel dalam jumlah ribuan dan tidak memiliki mekanisme penegakan regulasi SOP, sehingga rawan melakukan pendelegasian yang melanggar wewenang divisi.

Di sisi lain, penelitian terbaru yang mengeksplorasi Large Language Models (LLMs) seperti IndoBERT atau fine-tuned GPT pada teks bahasa Indonesia memang mampu menangkap semantik kontekstual dengan akurasi tinggi (Golec & Hachaj, 2025; Huynh & McNamara, 2025). Kendati demikian, adopsi model deep learning berskala besar di lingkungan korporasi sering kali terbentur oleh kendala tingginya latensi komputasi (inferensi mencapai ratusan milidetik), ketergantungan pada perangkat keras GPU bernilai investasi tinggi, kerentanan terhadap halusinasi prediksi, serta sifat 'black-box' yang tidak dapat diaudit secara regulatif oleh auditor internal organisasi (Wang & Shalaby, 2026; Fu et al., 2023).

Pendekatan Hybrid Cascading yang diusulkan dalam naskah ini menawarkan solusi jalan tengah yang sangat elegan. Dengan mendekomposisi proses inferensi menjadi penyaringan aturan deterministik berbasis SOP dan jaring pengaman statistik TF-IDF, sistem mampu meraih akurasi delegasi divisi 90,62% dengan kepatuhan kebijakan mutlak pada 59,38% kasus, tanpa memerlukan infrastruktur GPU atau korpus anotasi manual berskala masif. Dengan demikian, kontribusi utama

penelitian ini memperkaya khazanah rekayasa sistem informasi melalui pembuktian bahwa orkestrasi arsitektur bertingkat yang mengutamakan aturan domain jauh lebih efektif, ekonomis, dan terpercaya bagi otomasi layanan pelanggan korporasi dibandingkan sekadar mengejar kompleksitas algoritma.

Implikasi Praktis, Manajerial, dan Efisiensi Komputasi

Penerapan arsitektur Hybrid Cascading memberikan dampak praktis dan manajerial yang signifikan bagi operasional korporasi layanan telekomunikasi. Dari sudut pandang manajerial, sistem ini berhasil mengeliminasi bottleneck manual dengan mengotomasi 100% alur pembacaan dan delegasi tiket masuk secara seketika. Akurasi delegasi divisi sebesar 90,62% secara drastis meminimalisasi fenomena misrouting antar-departemen, sehingga mempercepat First Response Time dan pemenuhan Service Level Agreement (SLA) pelanggan.

Dari aspek efisiensi komputasi dan keberlanjutan sistem (sustainability), arsitektur yang diusulkan memiliki keunggulan telak dibandingkan model Transformer atau Large Language Models (LLMs) berskala besar. Model Rule-Based yang dipadukan dengan TF-IDF Cosine Similarity beroperasi secara sangat ringan (lightweight), dengan waktu inferensi di bawah 10 milidetik per transaksi tiket dan konsumsi memori kurang dari 50 MB RAM. Karakteristik ini memungkinkan model diintegrasikan langsung pada infrastruktur peladen on-premise perusahaan tanpa memerlukan akselerator grafis (GPU) berbiaya mahal, menjamin keamanan privasi data pelanggan, serta memberikan kemudahan pemeliharaan di mana penyesuaian regulasi baru cukup dilakukan dengan memperbarui file konfigurasi Knowledge Base SOP secara modular.

KESIMPULAN DAN SARAN

Penelitian ini berhasil merancang, mengimplementasikan, dan mengevaluasi arsitektur Hybrid Cascading berpintu keyakinan yang mengintegrasikan Knowledge Base SOP Telkom sebagai penjaga gerbang deterministik primer dan TF-IDF Cosine Similarity sebagai jaring pengaman adaptif untuk klasifikasi komplain dan delegasi tugas multi-divisi. Evaluasi empiris menggunakan korpus latih 538 data komplain mentah berbasis distant supervision serta 96 data uji berlabel yang memuat 40% variasi bahasa informal membuktikan bahwa model usulan berhasil mengungguli seluruh model pembanding mandiri dengan mencapai akurasi kategori 83,33%, akurasi pendelegasian divisi 90,62%, dan Macro F1-Score 83,69%. Temuan kunci penelitian menunjukkan bahwa 59,38% keputusan tiket diselesaikan secara deterministik pada lapisan SOP yang menjamin kepatuhan regulasi mutlak, sementara 40,62% kasus berhasil diselamatkan oleh lapisan fallback TF-IDF. Keberadaan 40,62% kasus fallback ini memberikan bukti ilmiah kokoh bahwa bahasa komplain pelanggan di lapangan sarat dengan ambiguitas leksikal dan istilah slang, sehingga pendekatan Rule-Based murni terbukti tidak memadai dan mutlak membutuhkan integrasi model hibrida sekuensial.

Meskipun model yang diusulkan terbukti tangguh dan efisien secara komputasi, terdapat keterbatasan penelitian yang perlu disempurnakan pada studi lanjutan. Basis aturan Knowledge Base saat ini masih bersifat statis sehingga memerlukan pembaruan manual apabila terjadi revisi pedoman SOP institusi, serta masih teramati adanya tumpang tindih semantik antara keluhan perangkat keras dan sinyal internet pada tuturan yang sangat minim kata kunci. Penelitian mendatang disarankan untuk mengembangkan mekanisme ekstraksi aturan adaptif berbasis active learning, mengintegrasikan fitur context-aware lemmatizer berbahasa Indonesia, serta memperluas pengujian pada korpus multi-institusi guna memperluas daya generalisasi sistem pendelegasian tugas otomatis di sektor publik maupun swasta.

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan ucapan terima kasih yang sebesar-besarnya kepada Fakultas Ilmu Komputer Universitas Dian Nuswantoro atas fasilitas riset dan dukungan akademik yang diberikan, serta kepada tim operasional PT Telkom Indonesia yang telah memberikan wawasan berharga terkait prosedur operasional standar penanganan gangguan layanan. Penulis juga mengucapkan terima kasih kepada Dewan Redaksi dan Reviewer Jurnal Media Informatika Budidarma atas saran dan masukan konstruktif demi peningkatan kualitas naskah ilmiah ini.

DAFTAR PUSTAKA

Adli, H. R., Megasari, R., & U. P. Indonesia. (2023). Implementation of inverse document frequency (TF-IDF) and cosine similarity terms in determining research reviewers for Indonesian Education University lecturers. *Journal of Computing and Society*, 4(2), 73–82.

Allam, H., Makubvure, L., Gyamfi, B., Graham, K. N., & Akinwolere, K. (2025). Text classification: How machine learning is revolutionizing text categorization. *Information*, 16(2), 1–47. https://doi.org/10.3390/info16020130

Arifman, F., Mantoro, T., & Handayani, D. O. D. (2026). A hybrid machine learning approach for classifying Indonesian cybercrime discourse using a localized threat taxonomy. *Information*, 17(3), 1–25. https://doi.org/10.3390/info17030301

Aththanayake, S., Mallikarachchi, C., Wickramasinghe, J., Kugarajah, S., Meedeniya, D., & Pradhan, B. (2026). ResQConnect: An AI-powered multi-agentic platform for human-centered and resilient disaster response. *Sustainability*, 18(2), 1–48. https://doi.org/10.3390/su18021014

Arowosegbe, A., & Oyelade, T. (2023). Application of natural language processing (NLP) in detecting and preventing suicide ideation: A systematic review. *International Journal of Environmental Research and Public Health*, 20(2), 1–23. https://doi.org/10.3390/ijerph20021514

Bhattacharya, D., Krishnamurthy, P., & Shyam Sundar, K. R. (2023). Customer service ticket classification using machine learning and natural language processing. *International Journal of Computer Applications,* 185(6), 1–8. https://doi.org/10.5120/ijca2023922882

Ding, Y., Han, S. C., Lee, J., & Hovy, E. (2026). Deep learning-based visually rich document content understanding: A survey. *Artificial Intelligence Review*, 59(4), 1–63. https://doi.org/10.1007/s10462-025-11477-3

Ehrgott, M., Köksalan, M., Kadziński, M., & Deb, K. (2026). Fifty years of multi-objective optimization and decision-making: From mathematical programming to evolutionary computation. *European Journal of Operational Research*, 330(1), 1–25. https://doi.org/10.1016/j.ejor.2025.06.012

Fu, S., et al. (2023). Recommended practices and ethical considerations for natural language processing-assisted observational research: A scoping review. *Clinical and Translational Science*, 16(3), 398–411. https://doi.org/10.1111/cts.13463

Golec, J., & Hachaj, T. (2025). Ten natural language processing tasks with generative artificial intelligence. *Applied Sciences*, 15(16), 1–30. https://doi.org/10.3390/app15169057

González, T. R., Rocha, M. A., Baltazar, R., Casillas, M. Á., & del Valle, J. (2022). Voice control system through NLP (natural language processing) as an interactive model for scalable ERP platforms in Industry 4.0. In Smart innovation, systems and technologies (Vol. 306, pp. 207–217). *Springer*. https://doi.org/10.1007/978-981-19-3359-2_18

Greco, C. M., & Tagarelli, A. (2024). Bringing order into the realm of transformer-based language models for artificial intelligence and law. *Artificial Intelligence and Law*, 32(4), 883–924. https://doi.org/10.1007/s10506-023-09374-7

Harris, S., Hadi, H. J., Ahmad, N., & Alshara, M. A. (2024). Fake news detection revisited: An extensive review of theoretical frameworks, dataset assessments, model constraints, and forward-looking research agendas. *Technologies*, 12(11), 1–63. https://doi.org/10.3390/technologies12110222

Hendry, H., Tukino, T., Sediyono, E., Fauzi, A., & Huda, B. (2025). HyEWCos: A comparative study of hybrid embedding and weighting techniques for text similarity in short subjective educational text. *Information*, 16(11), 1–28. https://doi.org/10.3390/info16110995

Huynh, L., & McNamara, D. S. (2025). Natural language processing as a scalable method for evaluating educational text personalization by LLMs. *Applied Sciences*, 15(22), 1–26. https://doi.org/10.3390/app152212128

Jiang, S. (2025). *Big data sharing: A comprehensive survey. Data*, 10(11), 1–35. https://doi.org/10.3390/data10110182

Li, J., Sun, M., & Wu, Z. (2022). A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1), 50–70. https://doi.org/10.1109/TKDE.2020.2981314

Okaka, R. A., Mwangi, W., & Okeyo, G. (2016). A hybrid approach for personalized recommender system using weighted TF-IDF on RSS contents. *International Journal of Computer Applications Technology and Research*, 5(12), 764–774. https://doi.org/10.7753/ijcatr0512.1006

Pothuri, V. R. V. (2025). Natural language processing in conversational AI. International Research *Journal of Modern Engineering Technology and Science*, (9), 436–440. https://doi.org/10.4018/979-8-3373-4217-7.ch011

Pratibha, Kaur, A., Khurana, M., & Damaševičius, R. (2024). Multimodal Hinglish tweet dataset for deep pragmatic analysis. *Data*, 9(2), 1–19. https://doi.org/10.3390/data9020038

Savvadelli, E., Kiouvrekis, Y., & Kokkinaki, A. (2026). A literature review on rule-based systems as decision support systems. *In IFIP advances in information and communication technology* (Vol. 761, pp. 376–388). Springer. https://doi.org/10.1007/978-3-032-02504-3_26

Soga, L. R., Bolade-Ogunfodun, Y., Mariani, M., Nasr, R., & Laker, B. (2022). Unmasking the other face of flexible working practices: A systematic literature review. *Journal of Business Research*, 142, 648–662. https://doi.org/10.1016/j.jbusres.2022.01.024

Umer, M., Ashraf, I., Mehmood, A., Ullah, S., & Choi, G. S. (2021). Predicting numeric ratings for Google Apps using text features and regression algorithms. *ETRI Journal*, 43(1), 95–107. https://doi.org/10.4218/etrij.2019-0299

Vuković, D., Žunac, A. G., & Jurič, B. (2023). Customer experience as an important factor in evaluating telecommunications services. *International Journal for Quality Research*, 17(2), 427–440. https://doi.org/10.24874/IJQR17.02-08

Wang, J., & Shalaby, A. (2026). ELSSA: Explainable large language model-based decision support for public transit incident management. *Transportation Research Part C: Emerging Technologies*, 193, Article 105923. https://doi.org/10.1016/j.trc.2026.105923

Wang, Y., Liu, Y., & Sun, H. (2023). A hybrid text classification approach using rule-based and statistical methods for customer complaint routing. *Expert Systems with Applications*, 213, Article 119086. https://doi.org/10.1016/j.eswa.2022.119086

Widianto, A., Pebriyanto, E., & Fitriyanti, M. (2025). Document similarity using term frequency-inverse document frequency representation and cosine similarity. *Data Science and Information Technology Data Analysis*, 4(2), 149–153.

Zhou, J., Ye, Z., Zhang, S., Geng, Z., Han, N., & Yang, T. (2024). Investigating response behavior through TF-IDF and Word2vec text analysis: A case study of PISA 2012 problem-solving process data. *Heliyon*, 10(16), Article e35945. https://doi.org/10.1016/j.heliyon.2024.e35945