



Klasifikasi Prioritas Pengaduan Pelayanan Publik Menggunakan *Natural Language Processing* dan *Machine Learning*

Suherman¹, Ismail²

Program Studi Teknik Informatika, Universitas Lamappapoleonro¹

Program Studi Sistem Informasi, Universitas Lamappapoleonro²

Jl. Salotungo No. 62, Watansoppeng, Soppeng, Sulawesi Selatan, Indonesia^{1, 2}

suherman@unipol.ac.id*¹, ismail@unipol.ac.id²

Kata Kunci :

Pengaduan publik; NLP; Machine Learning; Klasifikasi prioritas; TF-IDF

ABSTRAK

Penelitian ini mengklasifikasikan prioritas pengaduan pelayanan publik menggunakan Natural Language Processing dan Machine Learning pada 470 pengaduan Kantor Desa Gattareng Toa. Label Tinggi, Sedang, dan Rendah ditetapkan melalui penilaian dua pakar dan konsensus. Teks diproses melalui pembersihan, normalisasi, tokenisasi, dan TF-IDF, kemudian diuji menggunakan Naive Bayes, Logistic Regression, SVM, dan Random Forest. Evaluasi lima fold terkelompok berdasarkan pola inti menunjukkan Logistic Regression menghasilkan accuracy 88,94% dan Macro-F1 87,13%, dengan 30 pengaduan Tinggi salah diklasifikasikan sebagai Sedang. Penyesuaian ambang meningkatkan recall Tinggi dari 68,42% menjadi 80,00%, menurunkan kasus terlewat menjadi 19, tetapi meningkatkan peringatan keliru dari 11 menjadi 24. Temuan menunjukkan pentingnya evaluasi ketahanan pola dan kompromi risiko; keluaran model dimaksudkan sebagai alat bantu verifikasi petugas, bukan keputusan otomatis.

Keywords:

Public complaints; NLP; Machine learning; Priority classification; TF-IDF

ABSTRACT

This study classifies public-service complaint priorities using natural language processing and machine learning on 470 complaints from Gattareng Toa Village Office. High, Medium, and Low labels were established through two expert assessments and consensus. Complaint texts underwent cleaning, normalization, tokenization, and TF-IDF representation before comparison of Naive Bayes, Logistic Regression, SVM, and Random Forest. Five-fold core-pattern-grouped evaluation found Logistic Regression achieved 88.94% accuracy and 87.13% macro-F1, but misclassified 30 High-priority complaints as Medium. Adjusting the High-priority threshold improved High-class recall from 68.42% to 80.00%, reducing missed cases to 19 while increasing false alerts from 11 to 24. These findings highlight the need for pattern-aware validation and explicit risk trade-offs. The proposed classifier supports human complaint triage rather than replacing administrative decisions.

---Jurnal JISTI @2026---

PENDAHULUAN

Pengaduan masyarakat merupakan mekanisme partisipasi yang memungkinkan pemerintah mengenali gangguan layanan, mencatat kebutuhan warga, dan menetapkan tindak lanjut yang dapat dipertanggungjawabkan. Di lingkungan pemerintahan desa, pengaduan sering disampaikan melalui bahasa sehari-hari yang ringkas, tidak seragam, dan memuat beberapa persoalan sekaligus. Ketika seluruh aduan diperiksa dengan urutan kedatangan tanpa analisis urgensi, laporan yang menyangkut keselamatan atau layanan dasar berisiko tertunda. Transformasi pengelolaan pengaduan karena itu memerlukan alat bantu yang mampu membaca isi laporan dan mengelompokkannya secara konsisten, sambil mempertahankan tanggung jawab keputusan pada aparatur desa. Jiang et al. (2023) memetakan pemanfaatan Natural Language Processing (NLP) dalam pemerintahan, sedangkan Esperança et al. (2025) menggarisbawahi perlunya pengenalan pola semantik dari aduan sektor publik.



Pemrosesan bahasa alami memungkinkan teks tidak terstruktur diubah menjadi representasi numerik untuk analisis. Pada konteks Indonesia, Madyatmadja et al. (2022) membangun kerangka analitik kontekstual untuk laporan warga, dan Madyatmadja et al. (2023) membandingkan berbagai algoritma pada platform pengaduan Tangerang. Purba et al. (2024) melaporkan penggunaan Logistic Regression untuk mendeteksi pengaduan pada media sosial; Pradananto dan Junaidi (2026) menggunakan TF-IDF dan Decision Tree untuk kategori aduan Laporan Gubernur. Kajian-kajian tersebut memberikan dasar penerapan klasifikasi teks, tetapi tugas membedakan jenis/topik pengaduan tidak identik dengan menentukan tingkat prioritas penanganannya.

Pemanfaatan NLP untuk layanan pengaduan semakin beragam. Dharmendra et al. (2026) menguji arsitektur deep learning pada pengaduan Denpasar, Febyani et al. (2026) mengklasifikasikan kategori pengaduan Wargaku menggunakan representasi teks, dan Prayogo et al. (2026) mengevaluasi validitas laporan masyarakat melalui pembelajaran mendalam. Venequenn dan Surahmat (2026) memanfaatkan Naive Bayes pada skala kelurahan. Sementara itu, Rakhimzhanov et al. (2025) membandingkan representasi berbasis embedding pada pengaduan transportasi publik, serta Wibowo (2024) mendiskusikan implikasi manajerial klasifikasi pengaduan. Berbagai studi tersebut memperkuat relevansi Machine Learning untuk layanan publik, namun perbedaan definisi kelas, wilayah, dan protokol evaluasi membatasi perbandingan langsung antarpelitian.

Penentuan prioritas menambah dimensi keputusan karena tidak semua kesalahan prediksi mempunyai dampak serupa. Vairetti et al. (2024) mengintegrasikan analisis teks dan pengambilan keputusan multikriteria untuk pemeringkatan keluhan, sedangkan Lidiawaty et al. (2025) mengembangkan pendekatan urgensi aduan lalu lintas berbahasa Indonesia. Sabilla dan Dhenabayu (2026) mengevaluasi klasifikasi dan skor prioritas pada layanan NYC 311; Santoso et al. (2026) menunjukkan relevansi pendekatan cost-sensitive pada klasifikasi intent pengaduan pemerintahan yang tidak seimbang. Studi Ojo et al. (2024) juga menempatkan analisis teks sebagai dasar memprioritaskan persoalan layanan kesehatan. Penelitian tersebut menunjukkan bahwa ketepatan klasifikasi dan konsekuensi penanganan harus dinilai bersama, bukan hanya berdasarkan akurasi agregat.

Penelitian dalam domain layanan kesehatan memperlihatkan pentingnya memanfaatkan narasi bebas sebagai informasi operasional. Li et al. (2025) mengembangkan sistem klasifikasi keluhan pasien, sedangkan Feizollah et al. (2025) melalui tinjauan pemetaan menunjukkan bahwa pemanfaatan umpan balik tekstual harus terkait secara jelas dengan tujuan layanan. Prinsip tersebut relevan untuk pengaduan desa: label Tinggi, Sedang, dan Rendah tidak cukup dipahami sebagai kategori statistik, tetapi harus dapat diterjemahkan menjadi tindakan verifikasi yang sesuai dengan wewenang dan kapasitas petugas. Anadas dan Suryono (2025) juga memperlihatkan karakteristik analisis pendapat publik pada kanal layanan pemerintah, walaupun sentimen tidak dapat dianggap setara dengan urgensi laporan.

Dari sisi teknis, kualitas representasi dan validasi sangat memengaruhi kesimpulan. Chai (2023) menjelaskan bahwa pemilihan teknik pembersihan, tokenisasi, dan normalisasi bergantung pada karakter data serta kebutuhan analitik; Semary et al. (2024) dan Malik dan Bilal (2024) mengulas kontribusi ekstraksi fitur dan variasi metode NLP. Lebih jauh, Kapoor dan Narayanan (2023) menunjukkan bahwa kebocoran data dapat menghasilkan evaluasi Machine Learning yang terlalu optimistis. Pengaduan yang menggunakan pola kalimat serupa dapat muncul pada pelatihan dan pengujian walaupun tidak memiliki teks identik. Karena itu, pengujian berbasis kelompok pola ini perlu melengkapi pembagian acak agar ketahanan terhadap keluhan dengan pola berbeda dapat dinilai.

Persoalan yang diangkat pada penelitian ini berasal dari data pengaduan Kantor Desa Gattareng Toa, Kabupaten Soppeng, yang menurut keterangan peneliti dikumpulkan melalui observasi lapangan dan dikonfirmasi dengan narasumber serta pakar. Tersedia 470 catatan dengan tiga kelas prioritas. Dua penilai menghasilkan 431 kesepakatan awal, sedangkan 39 perbedaan diselesaikan melalui konsensus



bersama pakar; seluruh keputusan akhir mengikuti label Penilai 1. Dataset memuat 394 teks unik, tetapi pemeriksaan pola narasi mengidentifikasi 59 kelompok pola inti. Kondisi ini menuntut evaluasi yang membedakan kemampuan mengenali pengulangan kalimat dari kemampuan menangani pola aduan yang belum diamati dalam pelatihan.

Penelitian Suherman dan Ismail (2026) sebelumnya mengkaji reduksi kesalahan klasifikasi penerima bantuan sosial berbasis data sosial-ekonomi. Studi saat ini memperluas ranah penerapan Machine Learning ke pengelolaan pengaduan pelayanan desa, dengan unit analisis berupa teks dan dampak keputusan berupa prioritas verifikasi. Kontribusi penelitian difokuskan pada evaluasi empat algoritma konvensional yang ringan, pengendalian kemiripan antardata pelatihan dan pengujian, serta analisis pergeseran risiko ketika ambang pengaduan prioritas Tinggi disesuaikan. Komisarenko dan Kull (2025) menekankan pentingnya mempertimbangkan biaya kesalahan yang tidak simetris; dalam penelitian ini nilai bobotnya tetap diperlakukan sebagai asumsi eksperimen, bukan ketentuan resmi desa.

Berdasarkan latar belakang tersebut, penelitian bertujuan: (1) membandingkan kinerja Multinomial Naive Bayes, Logistic Regression, Linear SVM, dan Random Forest untuk tiga tingkat prioritas pengaduan; (2) mengevaluasi ketahanan prediksi menggunakan pemisahan berdasarkan pola inti keluhan; dan (3) mengukur kompromi antara pengurangan pengaduan Tinggi yang terlewat dan bertambahnya laporan non-Tinggi yang diprioritaskan berlebihan melalui pembobotan kelas atau penyesuaian ambang. Kebaruan yang diajukan bersifat aplikatif-evaluatif, yaitu penggabungan validasi pola baru dan analisis konsekuensi kesalahan untuk konteks pengaduan desa, bukan penciptaan algoritma klasifikasi baru.

METODE PENELITIAN

Penelitian menggunakan desain kuantitatif eksperimental pada teks pengaduan pelayanan publik Kantor Desa Gattareng Toa. Unit analisis adalah satu catatan pengaduan, sedangkan variabel target merupakan prioritas akhir Tinggi, Sedang, atau Rendah berdasarkan konsensus pakar. Alur kerja mengikuti audit data, validasi label, prapemrosesan NLP, pembentukan fitur, perbandingan model, uji generalisasi berdasarkan pola inti, serta analisis perubahan biaya kesalahan sebagaimana Gambar 1.



Validasi terkelompok berdasarkan 59 pola inti • pemilihan model hanya pada data latih • keputusan akhir diverifikasi manusia

Gambar 1. Alur penelitian NLP–Machine Learning dan evaluasi risiko

Sumber: Rancangan penelitian menggunakan Python (2026).

a. Dataset dan validasi label

Dataset berjumlah 470 observasi dan seluruh ID pengaduan bersifat unik. Kelas prioritas ditetapkan oleh dua penilai; terdapat 431 kesepakatan awal dan 39 ketidaksesuaian yang diselesaikan melalui diskusi/konsensus dengan pakar. Label akhir mempertahankan keputusan musyawarah. Pengaduan yang berisi teks identik tidak otomatis dihapus karena dapat merepresentasikan kejadian berbeda; seluruh observasi dipertahankan untuk evaluasi, tetapi kedekatan teks dikendalikan saat pembagian data.



Tabel 1. Distribusi Label Prioritas Pengaduan

Prioritas	Jumlah	Persentase	Makna operasional
Tinggi	95	20,21%	Mendesak; potensi risiko keselamatan, kesehatan, atau layanan esensial
Sedang	235	50,00%	Mengganggu pelayanan dan memerlukan tindak lanjut terjadwal
Rendah	140	29,79%	Saran atau persoalan dengan tingkat urgensi lebih rendah
Total	470	100,00%	Label akhir berdasarkan konsensus pakar

Sumber: Data observasi lapangan Kantor Desa Gattareng Toa (2026).

b. Prapemrosesan dan ekstraksi fitur

Tahapan NLP mencakup normalisasi Unicode, text cleaning terhadap tanda baca/spasi dan pola informasi personal, case folding, normalisasi singkatan berbasis kamus, serta tokenisasi kata. Kata negasi seperti “tidak”, “belum”, dan “bukan” dipertahankan karena berpotensi menentukan makna keluhan. Stemming dan stopword removal tidak diterapkan pada konfigurasi utama. Teks ditransformasikan melalui TF-IDF unigram–bigram dengan $\text{min_df} = 2$, $\text{max_df} = 0,95$, $\text{max_features} = 5.000$, dan normalisasi L2; kosakata dan bobot IDF selalu dipelajari hanya dari partisi latih untuk menghindari kebocoran informasi (Chai, 2023; Semary et al., 2024).

c. Algoritma dan desain validasi

Empat algoritma dievaluasi, yakni Multinomial Naive Bayes, Logistic Regression, Linear Support Vector Machine, dan Random Forest. Prosedur awal memakai pembagian kelompok teks identik 80:20 (376 latih, 94 uji) dan validasi silang lima fold dalam data latih. Karena kesamaan teks tinggi, uji ketahanan utama menggunakan nested StratifiedGroupKFold dengan lima fold luar dan tiga fold dalam; kelompok dibentuk dari pola inti aduan setelah penghilangan frasa pembuka/penutup yang berulang. Pengaturan ini menghasilkan 59 kelompok pola inti, sehingga kelompok yang sama tidak dibagi antara fold latih dan uji. Pemilihan parameter, TF-IDF, dan penyesuaian ambang dilaksanakan pada data latih setiap fold; prediksi yang dilaporkan merupakan out-of-fold (Kapoor & Narayanan, 2023).

d. Penguatan model dan metrik

Model terpilih dibandingkan pada konfigurasi baseline, pembobotan kelas, penyesuaian ambang keputusan prioritas Tinggi, dan kombinasi keduanya. Penyetelan dilakukan di validasi silang internal, sedangkan evaluasi menggunakan accuracy, macro-F1, recall dan precision kelas Tinggi, confusion matrix, serta biaya kesalahan. Skema ilustratif menetapkan biaya 5 bagi pengaduan aktual Tinggi yang terlewat, 2 bagi prediksi Tinggi yang ternyata bukan Tinggi, dan 1 bagi kesalahan lainnya. Biaya ini diasumsikan untuk analisis sensitivitas dan belum divalidasi menjadi aturan resmi Kantor Desa Gattareng Toa (Komisarenko & Kull, 2025; Santoso et al., 2026).

HASIL DAN PEMBAHASAN

1. Karakteristik Dataset

Seluruh 470 pengaduan memiliki ID dan label akhir. Kategori Sedang merupakan kelas terbanyak (235; 50,00%), disusul Rendah (140; 29,79%) dan Tinggi (95; 20,21%). Proporsi tersebut menunjukkan risiko model menekankan kelas dominan apabila hanya dioptimalkan menggunakan accuracy. Pada pemeriksaan label terdapat 39/470 perbedaan awal (8,30%) di antara dua penilai. Setelah diskusi, seluruh kasus memiliki konsensus final, tetapi ketidaksepakatan awal tetap dicatat untuk menjaga jejak penilaian.

Sebanyak 394 teks berbeda secara persis; 76 merupakan pengulangan di luar kemunculan pertama. Ketika frasa pembuka dan penutup yang berpola sama dipisahkan, diperoleh 59 pola inti pengaduan. Jumlah kelompok efektif jauh lebih kecil dibanding jumlah observasi sehingga interpretasi performa harus menilai ketergantungan linguistik dan potensi ketidaksetaraan aduan baru. Ke-59 kelompok



tersebut digunakan pada validasi terkelompok, bukan sebagai bukti bahwa keseluruhan variasi bahasa warga desa telah terwakili.

Tabel 2. Ringkasan Audit Data dan Pelabelan

Komponen pemeriksaan	Hasil
Jumlah catatan / ID unik	470 / 470
Teks unik / kemunculan berulang	394 / 76
Pola inti pengaduan	59 kelompok
Kesepakatan awal dua penilai	431 (91,70%)
Perbedaan yang diselesaikan melalui konsensus	39 (8,30%)
Kelas Tinggi / Sedang / Rendah	95 / 235 / 140

Sumber: Audit dataset penelitian Oleh TIM Peneliti (2026).

2. Hasil Prapemrosesan dan eksperimen pembagian awal

Prapemrosesan mempertahankan keseluruhan 470 catatan, dengan pembagian latih–uji sebanyak 376 dan 94 observasi. Dalam pembagian tersebut, TF-IDF menghasilkan 1.107 fitur unigram dan bigram yang dipelajari hanya dari data latih. Kinerja data uji tampak sangat tinggi: tiga algoritma mencapai accuracy dan macro-F1 100%, sedangkan Random Forest mencapai 98,94% accuracy. Hasil ini bersifat diagnostik, karena audit kemiripan menemukan 87 dari 94 teks uji memiliki kemiripan karakter sekurang-kurangnya 0,80 dengan salah satu teks pelatihan. Oleh karena itu skor awal bukan dasar tunggal menyatakan kemampuan generalisasi model.

Tabel 3. Performa Empat Algoritma pada Pembagian Awal 80:20

Algoritma	CV latih Macro-F1	Accuracy uji	Macro-F1 uji
Naive Bayes	97,62%	100,00%	100,00%
Logistic Regression	99,68%	100,00%	100,00%
Linear SVM	99,05%	100,00%	100,00%
Random Forest	94,24%	98,94%	98,75%

Sumber: Pengolahan data menggunakan Python/Google Colab (2026).

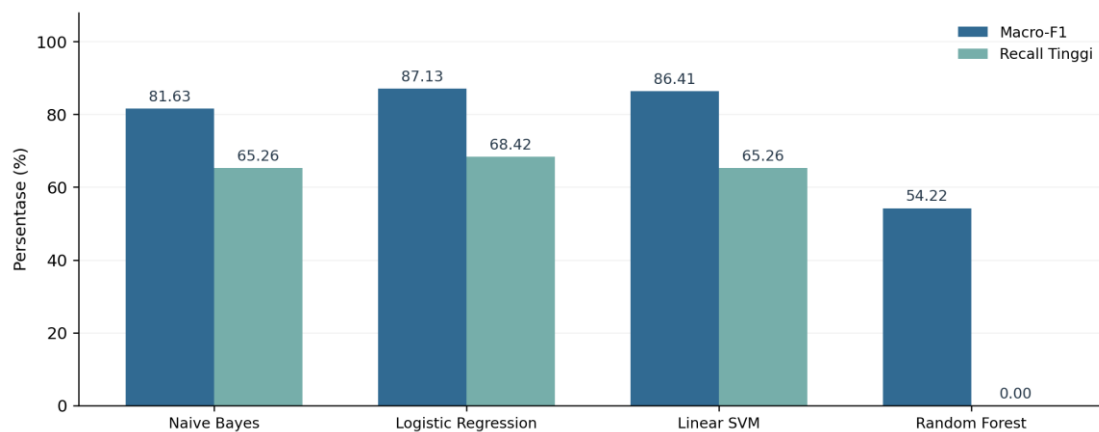
3. Evaluasi generalisasi terhadap pola inti baru

Ketika kelompok pola inti dipisahkan menggunakan nested cross-validation, terjadi penurunan performa pada semua model. Logistic Regression tetap memiliki macro-F1 tertinggi sebesar 87,13% dan accuracy 88,94%, diikuti Linear SVM (macro-F1 86,41%), Naive Bayes (81,63%), dan Random Forest (54,22%). Perubahan besar antara Tabel 3 dan Tabel 4 memperlihatkan bahwa kemiripan antarnaskah dapat membuat kinerja awal tampak terlalu tinggi. Evaluasi berdasarkan pola inti lebih relevan sebagai uji ketahanan internal, meskipun belum menggantikan pengujian prospektif pada pengaduan yang benar-benar baru.

Tabel 4. Kinerja Out-of-Fold pada Validasi Berbasis Pola Inti

Algoritma	Accuracy	Macro-F1	Recall Tinggi
Logistic Regression	88,94%	87,13%	68,42%
Linear SVM	88,30%	86,41%	65,26%
Multinomial Naive Bayes	83,40%	81,63%	65,26%
Random Forest	72,13%	54,22%	0,00%

Sumber: Nested StratifiedGroupKFold 5 × 3; Python/Google Colab (2026).



Gambar 2. Macro-F1 dan recall kelas Tinggi pada validasi pola inti

Sumber: Hasil implementasi menggunakan Python (2026).

4. Analisis kesalahan klasifikasi prioritas Tinggi

Confusion matrix Logistic Regression baseline mengungkap 30 dari 95 pengaduan aktual Tinggi diprediksi Sedang, tanpa kasus Tinggi yang bergeser langsung ke Rendah. Dengan demikian, recall Tinggi hanya 68,42%, walaupun macro-F1 tergolong baik. Pada kelas Sedang, sebanyak 11 contoh dikategorikan Tinggi, sedangkan terdapat sejumlah kesalahan menuju Rendah. Pola ini memperlihatkan bahwa batas antara Tinggi dan Sedang merupakan daerah keputusan tersulit, terutama ketika laporan menggunakan kata-kata umum tanpa penanda eksplisit tentang ancaman keselamatan, gangguan kesehatan, atau terhentinya layanan penting.

Tabel 5. Confusion Matrix Logistic Regression Baseline (Aktual × Prediksi)

Aktual	Rendah	Sedang	Tinggi	Total
Rendah	138	2	0	140
Sedang	9	215	11	235
Tinggi	0	30	65	95
Total	147	247	76	470

Sumber: Hasil implementasi menggunakan Python (2026).

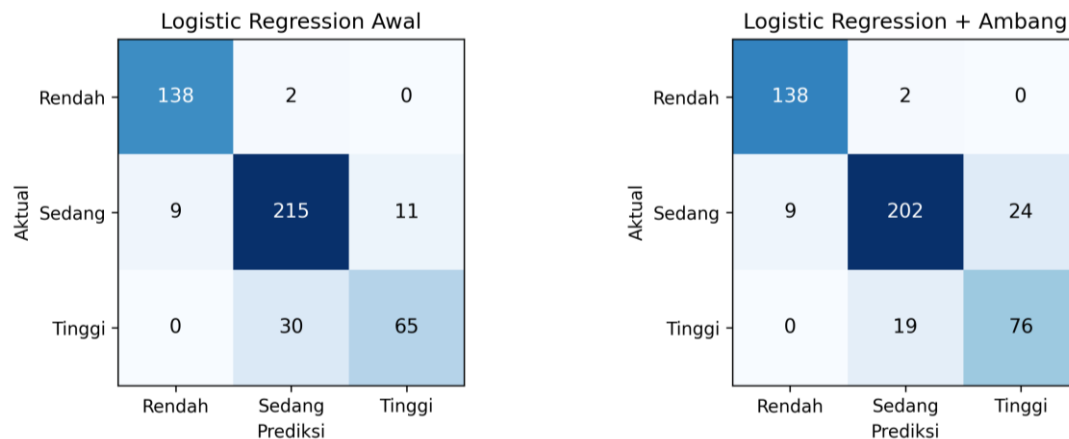
5. Penguatan klasifikasi melalui pembobotan dan penyesuaian ambang

Empat konfigurasi Logistic Regression dievaluasi pada prosedur validasi pola inti yang sama. Pembobotan risiko mengurangi salah klasifikasi Tinggi→Sedang dari 30 menjadi 21, sedangkan penyesuaian ambang menjadi 19. Kombinasi bobot dan ambang menurunkannya menjadi 16 kasus serta menghasilkan recall Tinggi 83,16%, tetapi meningkatkan false alert menjadi 38. Strategi penyesuaian ambang tanpa kombinasi menghasilkan macro-F1 tertinggi secara numerik (87,44%), recall Tinggi 80,00%, dan false alert sebanyak 24. Dengan demikian, model tidak mengalami perbaikan tanpa biaya; peningkatan sensitivitas menambah beban pemeriksaan pengaduan non-Tinggi.

Tabel 6. Perbandingan Penguatan Model Logistic Regression

Strategi	Macro-F1	Recall Tinggi	Tinggi → Sedang	False alert Tinggi
Baseline	87,13%	68,42%	30	11
Bobot kelas	87,00%	77,89%	21	30
Ambang Tinggi	87,44%	80,00%	19	24
Bobot + ambang	86,72%	83,16%	16	38

Sumber: Hasil implementasi menggunakan Python (2026).



Gambar 3. Perubahan confusion matrix baseline dan penyesuaian ambang Tinggi

Sumber: Hasil implementasi menggunakan Python (2026).

6. Evaluasi risiko keputusan dan beban verifikasi

Dengan bobot ilustratif 5:2:1, biaya kesalahan baseline mencapai 38,94 per 100 pengaduan, sedangkan penyesuaian ambang menurunkannya secara numerik menjadi 32,77. Namun, total kesalahan tidak menurun: baseline menghasilkan 52 kesalahan, sedangkan strategi ambang menghasilkan 54 kesalahan. Perubahan itu disebabkan pertukaran antara berkurangnya pengaduan mendesak yang terlewat dan bertambahnya pengaduan yang diperingatkan sebagai Tinggi. Ketika biaya verifikasi petugas tinggi atau kapasitas penanganan terbatas, strategi yang paling tepat dapat berbeda. Oleh sebab itu pengaturan ambang harus disesuaikan dengan kebutuhan desa, bukan dipilih semata-mata berdasarkan skor pengujian.

Tabel 7. Biaya Risiko Ilustratif pada 470 Pengaduan

Strategi	Total kesalahan	Biaya/100 aduan	Prediksi Tinggi
Baseline	52	38,94	76
Bobot kelas	55	35,96	104
Ambang Tinggi	54	32,77	100
Bobot + ambang	58	34,04	117

Sumber: Hasil implementasi menggunakan Python (2026).

Analisis bootstrap berkelompok memperkirakan perbaikan recall Tinggi dengan strategi ambang sebesar 4,07–22,08 poin persentase pada rentang persentil 95%. Sebaliknya, rentang perubahan macro-F1 melintasi nol (sekitar –3,48 sampai +4,08 poin persentase), sebagaimana rentang perubahan biaya risiko. Temuan ini menunjukkan adanya bukti internal tentang peningkatan sensitivitas kelas Tinggi, tetapi belum cukup untuk memastikan peningkatan menyeluruh pada pengaduan masa depan. Pengujian dengan catatan yang dikumpulkan setelah pengembangan model masih diperlukan.

7. Pembahasan dan implikasi penelitian

Keunggulan numerik Logistic Regression pada data berbentuk TF-IDF konsisten dengan temuan empiris klasifikasi teks oleh Purba et al. (2024), Febyani et al. (2026), dan Madyatmadja et al. (2023), meskipun perbedaan dataset dan definisi label menghalangi perbandingan angka secara langsung. Kinerja Random Forest yang menurun tajam pada pola baru menunjukkan bahwa kompleksitas model tidak otomatis meningkatkan ketahanan. Temuan ini memperkuat perhatian Kapoor dan Narayanan (2023) terhadap bias evaluasi ketika informasi yang berkaitan erat tersebar di partisi data.

Dibandingkan penelitian prioritas aduan yang menggabungkan model transformer dan metode keputusan (Vairetti et al., 2024), pendekatan ini menekankan model ringan dengan verifikasi evaluasi berbasis pola inti dan aturan keputusan yang sederhana. Kajian Lidiawaty et al. (2025) menunjukkan



bahwa penanda konteks bahaya dapat dipakai untuk menilai urgensi, sementara Santoso et al. (2026) dan Komisarenko dan Kull (2025) mengingatkan bahwa pengaturan biaya perlu sesuai kepentingan dan ketidakpastian keputusan. Hasil Suherman dan Ismail (2026) tentang klasifikasi bantuan sosial juga menempatkan kesalahan prediksi sebagai persoalan praktis; pada pengaduan desa, dimensi risikonya diarahkan pada kemungkinan laporan mendesak terlambat diverifikasi.

Keterbatasan penelitian meliputi ukuran sampel dan keragaman teks yang relatif terbatas, yakni hanya 59 pola inti pada 470 catatan; potensi ketergantungan narasi yang belum seluruhnya terwakili oleh pengelompokan heuristik; belum adanya uji prospektif lintas periode; serta ketiadaan evaluasi waktu respons dan dampak implementasi di kantor desa. Temuan administratif pada audit tanggal dan dokumentasi pelabelan perlu dilampirkan dalam arsip penelitian, tanpa mengubah observasi asli tanpa bukti. Di samping itu, bobot risiko 5:2:1 merupakan ilustrasi analitik, belum hasil penetapan resmi pemangku kepentingan. Karena itu model disarankan sebagai sistem bantu triase untuk verifikasi oleh aparat, bukan penentu final prioritas pelayanan secara otonom.

SIMPULAN DAN SARAN

Penelitian ini mengembangkan dan mengevaluasi klasifikasi tiga tingkat prioritas terhadap 470 pengaduan pelayanan publik Kantor Desa Gattareng Toa menggunakan prapemrosesan NLP, TF-IDF, serta empat algoritma Machine Learning. Validasi awal 80:20 memberikan nilai sangat tinggi pada sebagian model, tetapi evaluasi lebih ketat berdasarkan 59 kelompok pola inti menunjukkan Logistic Regression sebagai kandidat terbaik dengan accuracy 88,94%, macro-F1 87,13%, dan recall prioritas Tinggi 68,42%. Sebanyak 30 aduan Tinggi keliru dikategorikan Sedang; penyesuaian ambang mampu menurunkannya menjadi 19 dan meningkatkan recall Tinggi menjadi 80,00%, dengan konsekuensi false alert meningkat dari 11 menjadi 24 serta jumlah kesalahan total dari 52 menjadi 54. Skema biaya ilustratif menunjukkan potensi pengurangan risiko, tetapi belum membuktikan perbaikan operasional pada aduan baru. Dengan demikian, kontribusi utama penelitian terletak pada validasi berbasis pola dan analisis pertukaran risiko, sedangkan implementasi membutuhkan penetapan kebijakan prioritas bersama aparat desa, validasi prospektif, dan pengawasan manusia terhadap seluruh rekomendasi model.

DAFTAR PUSTAKA

- Anadas, S., & Suryono, R. R. (2025). Analisis sentimen publik terhadap program Lapor Mas Wapres periode 2024 di media sosial X menggunakan metode Naive Bayes dan Support Vector Machine (SVM). *Jurnal Pendidikan dan Teknologi Indonesia*. <https://doi.org/10.52436/1.jpti.934>
- Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, 29(3), 509–553. <https://doi.org/10.1017/S1351324922000213>
- Dharmendra, I. K., Wijaya, I. M. P. P., Putra, I. M. A. W., Atmojo, Y. P., & Pratiwi, L. P. S. (2026). Classifying public complaints in Denpasar: A comparative study of CNN, RNN, LSTM, and stacking deep learning models. *Jurnal Teknik Informatika (JUTIF)*, 7(1), 411–430. <https://doi.org/10.52436/1.jutif.2026.7.1.4153>
- Esperança, M., Freitas, D., Paixão, P. V., Marcos, T. A., Martins, R. A., & Ferreira, J. C. (2025). Proactive complaint management in public sector informatics using AI: A semantic pattern recognition framework. *Applied Sciences*, 15(12), 6673. <https://doi.org/10.3390/app15126673>
- Febyani, A. D., Nuryana, I. K. D., & Angresti, N. D. (2026). Public complaint text classification in the Wargaku application using natural language processing. *Journal of Education Technology and Information System*, 2(2), 145–158. <https://doi.org/10.26740/jetis.v2i02.43497>
- Feizollah, A., Lin, C.-Y., O'Malley, L., Thompson, W., Listl, S., & Byrne, M. (2025). The use of natural language processing to interpret unstructured patient feedback on health services: Scoping review. *Journal of Medical Internet Research*, 27, e72853. <https://doi.org/10.2196/72853>



- Jiang, Y., Pang, P. C.-I., Wong, D., & Kan, H. Y. (2023). Natural language processing adoption in governments and future research directions: A systematic review. *Applied Sciences*, 13(22), 12346. <https://doi.org/10.3390/app132212346>
- Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 100804. <https://doi.org/10.1016/j.patter.2023.100804>
- Komisarenko, V., & Kull, M. (2025). Cost-sensitive classification with cost uncertainty: Do we need surrogate losses? *Machine Learning*, 114, Article 132. <https://doi.org/10.1007/s10994-024-06634-8>
- Li, X., Shu, Q., Kong, C., Wang, J., Li, G., Fang, X., Lou, X., & Yu, G. (2025). An intelligent system for classifying patient complaints using machine learning and natural language processing: Development and validation study. *Journal of Medical Internet Research*, 27, e55721. <https://doi.org/10.2196/55721>
- Lidiawaty, B. R., Suryani, E., & Vinarti, R. A. (2025). Traffic urgency model: A novel approach to prioritize complaint texts using enhanced named entity recognition for Bahasa Indonesia cases. *IEEE Access*, 13, 57720–57733. <https://doi.org/10.1109/ACCESS.2025.3556423>
- Madyatmadja, E. D., Yahya, B. N., & Wijaya, C. (2022). Contextual text analytics framework for citizen report classification: A case study using the Indonesian language. *IEEE Access*, 10, 31432–31444. <https://doi.org/10.1109/ACCESS.2022.3158940>
- Madyatmadja, E. D., Sianipar, C. P. M., Wijaya, C., & Sembiring, D. J. M. (2023). Classifying crowdsourced citizen complaints through data mining: Accuracy testing of k-nearest neighbors, random forest, support vector machine, and AdaBoost. *Informatics*, 10(4), 84. <https://doi.org/10.3390/informatics10040084>
- Malik, N., & Bilal, M. (2024). Natural language processing for analyzing online customer reviews: A survey, taxonomy, and open research challenges. *PeerJ Computer Science*, 10, e2203. <https://doi.org/10.7717/peerj-cs.2203>
- Ojo, A., Rizun, N., Walsh, G., Isazad Mashinchi, M., Venosa, M., & Rao, M. N. (2024). Prioritising national healthcare service issues from free text feedback: A computational text analysis and predictive modelling approach. *Decision Support Systems*, 181, 114215. <https://doi.org/10.1016/j.dss.2024.114215>
- Pradananto, S., & Junaidi, M. (2026). Klasifikasi kategori aduan masyarakat pada aplikasi Laporan Gubernur menggunakan TF-IDF dan Decision Tree. *Jurnal Teknik Mesin, Industri, Elektro dan Informatika*, 5(2). <https://doi.org/10.55606/jtmei.v5i2.6176>
- Prayogo, I. W., Kurniawati, Y., Hidayati, I. E., Khairunnisa, A., & Ruldeviyani, Y. (2026). Text classification of public complaint validity with deep learning approaches. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 15(2), 115–123. <https://doi.org/10.22146/jnteti.v15i2.24003>
- Purba, M., Asri, S. D., Ayumi, V., Salamah, U., & Iryani, L. (2024). Klasifikasi dataset teks pengaduan masyarakat terhadap pemerintah di sosial media menggunakan Logistic Regression. *JSAI (Journal Scientific and Applied Informatics)*, 7(1), 78–83. <https://doi.org/10.36085/jsai.v7i1.6447>
- Rakhimzhanov, D., Belginova, S., & Yedilkhan, D. (2025). Automated classification of public transport complaints via text mining using LLMs and embeddings. *Information*, 16(8), 644. <https://doi.org/10.3390/info16080644>
- Sabilla, P., & Dhenabayu, R. (2026). Comparative complaint text classification using SVM, CNN, and BERT with priority scoring for NYC 311 service requests. *Jurnal Riset Informatika*, 8(4), 576–590. <https://doi.org/10.34288/jri.v8i4.589>
- Santoso, D. A., Indrayani, H., Widyatmoko, K., Alzami, F., & Iswahyudi, I. (2026). A systematic comparison of cost-sensitive methods for long-tail intent classification in Indonesian government complaint routing. *Statistics, Optimization & Information Computing*, 16(4), 2928–2956. <https://doi.org/10.19139/soic-2310-5070-4091>
- Semary, N. A., Ahmed, W., Amin, K., Pławiak, P., & Hammad, M. (2024). Enhancing machine learning-based sentiment analysis through feature extraction techniques. *PLOS ONE*, 19(2), e0294968. <https://doi.org/10.1371/journal.pone.0294968>



-
- Suherman, & Ismail. (2026). Optimasi model machine learning untuk reduksi kesalahan klasifikasi penerima BLT berbasis data sosial-ekonomi. *Jurnal Ilmiah Sistem Informasi dan Teknik Informatika (JISTI)*, 9(1), 83–98. <https://doi.org/10.57093/jisti.v9i1.407>
- Vairetti, C., Aránguiz, I., Maldonado, S., Karmy, J. P., & Leal, A. (2024). Analytics-driven complaint prioritisation via deep learning and multicriteria decision-making. *European Journal of Operational Research*, 312(3), 1108–1118. <https://doi.org/10.1016/j.ejor.2023.08.027>
- Venequenn, Z., & Surahmat, A. (2026). Implementation of the Naive Bayes algorithm for classification of public service complaints in e-government at Kunciran Indah Tangerang. *Scientific Journal of Information System*, 4(1), 38–43. <https://doi.org/10.70429/sjis.v4i1.334>
- Wibowo, G. D. (2024). Artificial intelligence-based classification of public complaints to enhance public service efficiency. *ESENSI: Jurnal Manajemen Bisnis*, 27(1). <https://doi.org/10.56943/esensi.v27i1.32>