

Klasifikasi Otomatis Keluhan Masyarakat Kota XYZ Berbasis IndoBERT untuk Mendukung Manajemen Layanan Publik

Darmansah^{1*}, Rifky Akbar Vetian², Koko Handoko³, Muhammad Rasid Ridho⁴, Novri Adhiatma⁵

^{1,4,5}Program Studi Sistem Informasi, Fakultas Teknik dan Komputer, Universitas Putera Batam, Indonesia

^{2,3}Program Studi Teknik Informatika, Fakultas Teknik dan Komputer, Universitas Putera Batam, Indonesia

¹darmansah071@gmail.com, ²rifky.vetian@puterabatam.ac.id, ³kokohandoko01@gmail.com, ⁴rasid517@gmail.com, ⁵ad.novri@gmail.com

Abstrak.

Tujuan: Pengelolaan keluhan masyarakat dalam jumlah besar memerlukan proses klasifikasi yang efektif agar keluhan dapat dikelompokkan sesuai bidang layanan dan diteruskan kepada unit terkait. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model IndoBERT dalam mengklasifikasikan keluhan masyarakat ke dalam lima kategori, yaitu infrastruktur, administrasi kependudukan, kesehatan, kebersihan, dan transportasi.

Metode/Design/Pendekatan: Penelitian menggunakan 5.000 data keluhan masyarakat Kota XYZ dari SP4N-LAPOR pada periode 2021 - 2026 yang dibagi secara stratified menjadi 80% data training, 10% validation, dan 10% testing. Tahapan penelitian meliputi preprocessing, tokenisasi, fine-tuning IndoBERT, dan evaluasi menggunakan accuracy, precision, recall, F1-score, serta confusion matrix. Kinerja IndoBERT juga dibandingkan dengan model baseline Support Vector Machine (SVM) dan Naive Bayes menggunakan fitur TF-IDF.

Hasil/Temuan: Hasil penelitian menunjukkan bahwa IndoBERT memperoleh accuracy sebesar 91% dan macro F1-score sebesar 90,6%, lebih tinggi dibandingkan SVM dengan accuracy 85% dan Naive Bayes sebesar 80%. Confusion matrix menunjukkan bahwa sebagian besar keluhan berhasil diklasifikasikan dengan benar, meskipun masih terdapat kesalahan pada kategori yang memiliki kemiripan konteks, terutama kebersihan dengan infrastruktur dan transportasi.

Kebaharuan/Originalitas/Nilai: Nilai penelitian ini terletak pada penerapan IndoBERT untuk mengelompokkan keluhan masyarakat berbahasa Indonesia berdasarkan konteks teks serta memetakan hasil klasifikasi ke unit layanan terkait. Hasil penelitian dapat menjadi dasar pendukung pengelolaan dan analisis pengaduan masyarakat secara otomatis, meskipun penerapan langsung pada alur operasional pelayanan masih memerlukan pengujian lebih lanjut.

Keywords: IndoBERT, Text Classification, Public Complaints, Public Services, Natural Language Processing, Service Classification.

Abstract.

Purpose: Managing a large volume of public complaints requires an effective classification process so that complaints can be categorized according to service areas and directed to the relevant units. This study aims to develop and evaluate an IndoBERT model for classifying public complaints into five categories: infrastructure, population administration, healthcare, sanitation, and transportation.

Method/Design/Approach: The study used 5,000 public complaint data from XYZ City from SP4N-LAPOR for the 2021-2026 period, stratified into 80% training data, 10% validation data, and 10% testing data. The research stages included preprocessing, tokenization, fine-tuning IndoBERT, and evaluation using accuracy, precision, recall, F1-score, and confusion matrix. IndoBERT performance was also compared with baseline Support Vector Machine (SVM) and Naive Bayes models using the TF-IDF feature.

Results/Findings: The results showed that IndoBERT achieved an accuracy of 91% and a macro F1-score of 90.6%, outperforming SVM with an accuracy of 85% and Naive Bayes with 80%. The confusion matrix indicated that most complaints were correctly classified, although misclassification occurred in categories with similar contexts, particularly sanitation, infrastructure, and transportation.

Novelty/Originality/Value: The contribution of this study lies in applying IndoBERT to classify Indonesian-language public complaints based on contextual information and mapping the classification results to relevant service units. These findings provide a basis for supporting the automated organization and analysis of public complaints, although further evaluation is required to assess its direct implementation in operational public service workflows.

Keywords: IndoBERT, Text Classification, Public Complaints, Public Services, Natural Language Processing, Service Classification.

Article history:

Received, 2026-09-24

Revised, 2026-10-03

Accepted, 2026-10-08

^{1*}Corresponding author.

Darmansah

Email addresses: darmansah071@gmail.com

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

PENDAHULUAN

Perkembangan layanan pemerintahan berbasis digital menyebabkan masyarakat semakin mudah menyampaikan keluhan, aspirasi, dan laporan terkait pelayanan publik melalui berbagai kanal digital [1], [2], [3]. Data pengaduan yang dihasilkan umumnya berbentuk teks tidak terstruktur dengan karakteristik bahasa yang beragam, seperti penggunaan bahasa informal, singkatan, kesalahan penulisan, dan variasi istilah. Meningkatnya jumlah pengaduan dapat menyediakan informasi penting mengenai permasalahan yang dihadapi masyarakat dalam memperoleh pelayanan publik. Namun, proses pengelompokan pengaduan secara manual membutuhkan waktu dan sumber daya yang relatif besar. Oleh karena itu, diperlukan pendekatan otomatis yang mampu mengklasifikasikan data keluhan secara cepat dan konsisten sehingga informasi yang terkandung di dalamnya dapat dimanfaatkan secara lebih efisien untuk mendukung pemantauan dan evaluasi pelayanan publik [4], [5], [6]. Pemrosesan bahasa alami (*Natural Language Processing/NLP*) dan pembelajaran mendalam telah banyak digunakan untuk mengolah teks berbahasa Indonesia. Salah satu model yang banyak diterapkan adalah *Bidirectional Encoder Representations from Transformers* (BERT), yang mampu memanfaatkan konteks kata berdasarkan hubungan antarkata dalam suatu teks. Untuk bahasa Indonesia, IndoBERT dikembangkan melalui pelatihan pada korpus berbahasa Indonesia dan dapat diterapkan pada berbagai tugas NLP, termasuk klasifikasi teks. Kemampuan model dalam menangkap konteks memungkinkan IndoBERT digunakan untuk mengolah teks dengan struktur dan variasi penggunaan bahasa yang beragam [7], [8], [9]. Dalam konteks pengaduan masyarakat, kemampuan tersebut relevan karena informasi mengenai permasalahan pelayanan tidak selalu disampaikan menggunakan istilah yang seragam.

Penelitian terdahulu menunjukkan bahwa IndoBERT telah banyak digunakan dalam pengolahan dan klasifikasi teks berbahasa Indonesia. Penelitian [10] membandingkan BERT, IndoBERT, dan CNN-LSTM untuk klasifikasi opini publik mengenai vaksin COVID-19, sedangkan penelitian [11] membandingkan LSTM dan IndoBERT untuk identifikasi hoaks pada *Twitter*. Sedangkan [12] menggunakan IndoBERT untuk mengklasifikasikan umpan balik mahasiswa, sementara Ciam dan Panjaitan menerapkan IndoBERT untuk analisis sentimen terhadap layanan BPJS. Selain itu, [13] menggunakan *fine-tuned* IndoBERT untuk analisis sentimen berbasis aspek pada ulasan hotel di Indonesia. Penelitian-penelitian tersebut menunjukkan bahwa IndoBERT dapat digunakan untuk memahami konteks dan melakukan klasifikasi teks pada berbagai domain.

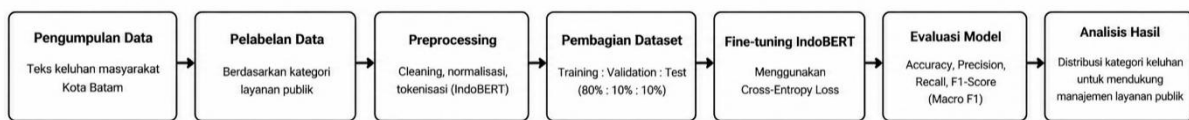
Meskipun demikian, penelitian terdahulu tersebut memiliki fokus yang berbeda dengan klasifikasi keluhan masyarakat berdasarkan kategori layanan publik. Sebagian penelitian berfokus pada sentimen, opini, hoaks, atau umpan balik terhadap suatu layanan, sehingga keluaran klasifikasi terutama menggambarkan polaritas atau karakteristik tertentu dari teks. Sementara itu, penelitian ini berfokus pada pengelompokan keluhan berdasarkan jenis permasalahan atau kategori layanan publik yang menjadi objek pengaduan. Perbedaan ini penting karena satu keluhan dapat mengandung informasi mengenai permasalahan layanan tertentu tanpa harus merepresentasikan sentimen positif atau negatif. Dengan demikian, klasifikasi berdasarkan kategori layanan dapat memberikan informasi yang lebih langsung mengenai jenis layanan atau permasalahan yang paling banyak dilaporkan masyarakat. Kategori keluhan dalam penelitian ini disusun berdasarkan jenis layanan publik yang menjadi objek pengaduan, seperti infrastruktur, administrasi pemerintahan, lingkungan, transportasi, fasilitas umum, kesehatan, dan layanan sosial. Pengelompokan tersebut relevan dengan konteks pelayanan publik karena setiap kategori merepresentasikan bidang layanan yang secara langsung berkaitan dengan kebutuhan dan aktivitas masyarakat. Kategori tersebut juga memungkinkan data pengaduan yang semula berbentuk teks tidak terstruktur diubah menjadi informasi terstruktur berdasarkan bidang permasalahan. Dengan adanya pengelompokan tersebut, distribusi pengaduan pada setiap kategori dapat dihitung dan digunakan untuk menggambarkan pola permasalahan pelayanan publik pada wilayah penelitian.

Kota XYZ merupakan wilayah perkotaan dengan aktivitas masyarakat dan jenis pelayanan publik yang beragam. Kondisi tersebut menghasilkan data pengaduan dengan karakteristik dan topik yang bervariasi, baik yang berkaitan dengan infrastruktur, administrasi pemerintahan, lingkungan, transportasi, fasilitas umum, kesehatan, maupun layanan sosial. Karakteristik lokal tersebut menjadi konteks penting dalam penerapan model klasifikasi karena istilah, pola penyampaian, dan jenis permasalahan yang muncul dalam pengaduan dapat berbeda antarwilayah. Oleh karena itu, penerapan IndoBERT pada data pengaduan Kota XYZ diarahkan untuk menguji kemampuan model dalam mengelompokkan keluhan berdasarkan kategori layanan publik sekaligus menggambarkan distribusi pengaduan pada masing-masing kategori. Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model IndoBERT dalam mengklasifikasikan keluhan masyarakat Kota XYZ berdasarkan kategori layanan publik. Evaluasi dilakukan dengan mengukur performa klasifikasi menggunakan metrik yang sesuai, seperti *accuracy*, *precision*, *recall*, dan *F1-score*, sehingga kemampuan model dalam membedakan setiap kategori dapat diketahui secara kuantitatif. Selain itu, penelitian ini menganalisis distribusi jumlah keluhan pada setiap kategori untuk mengidentifikasi kategori layanan yang memiliki frekuensi pengaduan relatif tinggi maupun rendah pada data penelitian. Hasil klasifikasi dan analisis distribusi tersebut diharapkan dapat menghasilkan informasi terstruktur mengenai karakteristik pengaduan

masyarakat yang dapat digunakan sebagai bahan pendukung dalam pemantauan dan evaluasi pelayanan publik di Kota XYZ. Dengan demikian, kontribusi penelitian ini terletak pada penerapan dan evaluasi IndoBERT pada data keluhan masyarakat Kota XYZ dengan skema kategori yang merepresentasikan bidang layanan publik, serta analisis distribusi hasil klasifikasi untuk memberikan gambaran mengenai pola pengaduan pada wilayah penelitian. Kontribusi tersebut ditempatkan dalam konteks data dan kategori layanan yang digunakan, sehingga tidak mengklaim bahwa metode IndoBERT secara umum merupakan pendekatan yang baru, melainkan mengevaluasi penerapannya pada karakteristik data pengaduan masyarakat di wilayah penelitian.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan menerapkan *Natural Language Processing* (NLP) dan model *Bidirectional Encoder Representations from Transformers* (BERT) yang telah diadaptasi untuk bahasa Indonesia, yaitu IndoBERT. Penelitian bertujuan membangun model klasifikasi otomatis untuk mengelompokkan keluhan masyarakat Kota XYZ berdasarkan kategori layanan publik. Tahapan penelitian meliputi pengumpulan data, pelabelan data, *preprocessing*, pembagian dataset, fine-tuning IndoBERT, evaluasi performa model, dan analisis hasil klasifikasi:



Gambar 1 Diagram alir penelitian

Gambar tersebut menggambarkan proses penelitian dari data mentah hingga diperoleh model klasifikasi dan informasi mengenai distribusi kategori keluhan masyarakat. Tahapan pertama yang dilakukan pada penelitian ini adalah pengumpulan data. Dataset penelitian terdiri atas 5.000 data keluhan masyarakat Kota XYZ yang diperoleh dari SP4N-LAPOR pada periode 2021 - 2026.

Data yang digunakan berupa teks atau deskripsi keluhan. Data kemudian diseleksi dengan menghapus data kosong, duplikat, tidak relevan, dan tidak memiliki informasi teks yang memadai. Data yang mengandung informasi pribadi juga dianonimkan sebelum digunakan. Setelah proses seleksi dan pembersihan, kemudian Skema label ditentukan berdasarkan karakteristik permasalahan yang ditemukan dalam dataset dan dapat disesuaikan dengan struktur layanan publik di Kota XYZ. Kategori keluhan yang digunakan dalam penelitian ditunjukkan pada Tabel 1:

Tabel 1. Pembagian Kategori

Kode	Kategori Keluhan
C1	Infrastruktur
C2	Administrasi Kependudukan
C3	Kesehatan
C4	Kebersihan
C5	Transportasi

Pelabelan dilakukan oleh minimal dua anotator berdasarkan kategori permasalahan utama pada setiap keluhan. Perbedaan label diselesaikan melalui diskusi untuk memperoleh label akhir, sedangkan konsistensi anotasi dapat diukur menggunakan Cohen's Kappa. Kemudian tahap *preprocessing* dilakukan untuk membersihkan dan menyeragamkan teks, meliputi penghapusan data kosong dan duplikat, normalisasi karakter, serta spasi dan format teks.

Pembersihan dilakukan secara terbatas agar informasi kontekstual yang penting bagi klasifikasi tetap dipertahankan. Untuk mencegah data leakage, pembentukan fitur TF-IDF hanya dilakukan menggunakan data training, kemudian diterapkan pada data validation dan testing. Selanjutnya sebanyak 5.000 data keluhan dibagi menjadi data training, validation, dan testing dengan proporsi 80%, 10%, dan 10%. Pembagian dilakukan secara stratified berdasarkan kategori keluhan untuk menjaga proporsi kelas pada masing-masing subset. Dengan jumlah dataset sebanyak 5.000 data, pembagian menghasilkan dataset seperti tabel dibawah ini:

Tabel 2. Pembagian Dataset

Dataset	Proporsi	Jumlah Data
Training	80%	4.000
Validation	10%	500
Testing	10%	500
Total	100%	5.000

Data training digunakan untuk melatih model, data validation untuk menentukan konfigurasi model, dan data testing untuk mengevaluasi performa akhir. Untuk mencegah data *leakage*, setiap data hanya berada pada satu subset, dan data testing tidak digunakan dalam proses pelatihan maupun pemilihan konfigurasi model. Model IndoBERT digunakan sebagai *pre-trained language* model dan disesuaikan dengan karakteristik data keluhan melalui proses *fine-tuning*. Proses pembelajaran model menggunakan fungsi *Cross-Entropy Loss* yang dirumuskan sebagai persamaan (1) berikut ini:

$$\mathcal{L}_{CE} = - \sum_{k=1}^K y_k \text{Log} (\hat{y}_k) \quad (1)$$

dengan (L) merupakan nilai *loss*, (K) jumlah kelas, (y_k) label aktual pada kelas ke- (k) , dan ($\hat{y}_k$) probabilitas prediksi model pada kelas ke- (k) . Konfigurasi seperti *learning rate*, *batch size*, jumlah *epoch*, dan *maximum sequence length* ditentukan melalui eksperimen pada data validasi. Evaluasi model dilakukan menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Karena jumlah data pada setiap kategori keluhan berpotensi tidak seimbang, *Macro F1-score* digunakan sebagai salah satu metrik utama. Nilai *Macro F1-score* diperoleh dengan menghitung rata-rata *F1-score* dari seluruh kelas, yang dirumuskan sebagai persamaan (2) berikut:

$$F1_{marco} = \frac{1}{K} \sum_{k=1}^K F1_k \quad (2)$$

dengan (K) merupakan jumlah kategori dan ($F1_k$) merupakan nilai *F1-score* pada kategori ke- (k) . Selain itu, confusion matrix digunakan untuk mengetahui pola kesalahan klasifikasi antar kategori. Hasil pengujian kemudian dianalisis untuk mengetahui distribusi kategori keluhan masyarakat Kota XYZ dan mengidentifikasi pola permasalahan layanan publik yang paling banyak dilaporkan. Hasil klasifikasi digunakan sebagai informasi pendukung dalam pemantauan dan evaluasi layanan publik. Dengan demikian, penelitian tidak hanya menghasilkan model klasifikasi berbasis IndoBERT, tetapi juga menghasilkan informasi terstruktur mengenai kategori keluhan masyarakat yang dapat dimanfaatkan untuk mendukung manajemen layanan publik.

HASIL DAN PEMBAHASAN

Dataset penelitian terdiri atas 5.000 data keluhan masyarakat Kota XYZ yang diperoleh dari kanal pengaduan publik selama periode Januari–Desember 2025. Setiap data terdiri atas teks keluhan dan label kategori layanan. Berdasarkan proses pelabelan, data dikelompokkan ke dalam lima kategori, yaitu infrastruktur, administrasi kependudukan, kesehatan, kebersihan, dan transportasi. Adapun distribusi dataset yang di gunakan seperti yang terlihat pada tabel dibawah ini:

Tabel 3. Distribusi Dataset Berdasarkan Kategori

Kategori	Jumlah Data	Persentase
Infrastruktur	1.350	27,0%
Administrasi Kependudukan	1.050	21,0%
Kesehatan	850	17,0%
Kebersihan	900	18,0%
Transportasi	850	17,0%
Total	5.000	100%

Kategori infrastruktur memiliki jumlah data terbanyak, yaitu 1.350 data (27%), sedangkan kategori kesehatan dan transportasi masing-masing berjumlah 850 data (17%). Perbedaan jumlah data antar kategori menunjukkan adanya variasi distribusi kelas yang perlu diperhatikan dalam proses pelatihan dan evaluasi model. Selanjutnya, dataset dibagi secara stratified dengan proporsi 80% untuk data training, 10% untuk validation, dan 10% untuk testing. Pembagian tersebut menghasilkan 4.000 data training, 500 data validation, dan 500 data testing, dengan proporsi setiap kategori tetap terjaga pada masing-masing subset. Tahap preprocessing dilakukan untuk

mengurangi variasi penulisan pada data keluhan, terutama yang disebabkan oleh singkatan, bahasa informal, kesalahan ejaan, dan penggunaan simbol. Proses normalisasi dilakukan secara terbatas agar informasi semantik yang diperlukan oleh IndoBERT tetap dipertahankan. Berikut Adalah hasil dari pre preprocessing seperti tabel dibawah ini:

Tabel 4. Hasil Preprocessing

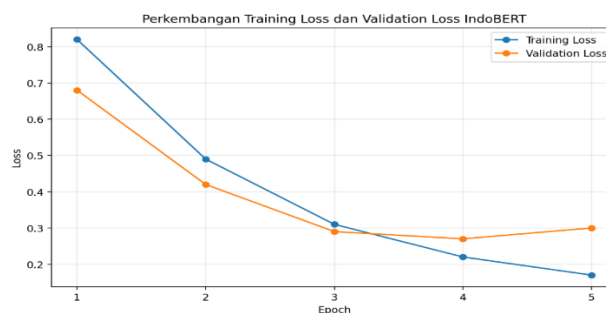
Teks Awal	Setelah Preprocessing
"Jln. ABC rusak parah!!! Tolong segera diperbaiki "	"jalan abc rusak parah tolong segera diperbaiki"
"KTP saya blm selesai2 sejak bulan lalu"	"ktp saya belum selesai sejak bulan lalu"

Hasil preprocessing menunjukkan bahwa variasi penulisan dapat dikurangi tanpa menghilangkan informasi utama dari keluhan. Teks hasil preprocessing selanjutnya digunakan sebagai masukan pada tahap tokenisasi dan klasifikasi. Kemudian Data dibagi secara stratified berdasarkan kategori keluhan menjadi 4.000 data training, 500 data validation, dan 500 data testing. Data training digunakan untuk proses fine-tuning, data validation digunakan untuk menentukan konfigurasi model, sedangkan data testing hanya digunakan pada tahap evaluasi akhir. IndoBERT digunakan sebagai model pre-trained yang kemudian dilakukan fine-tuning menggunakan data keluhan masyarakat. Konfigurasi pelatihan ditunjukkan pada Tabel dibawah ini:

Tabel 5. Konfigurasi dan Pelatihan IndoBERT

Parameter	Nilai
Model	IndoBERT
Jumlah Epoch	5
Batch Size	16
Learning Rate	2×10^{-5}
Maximum Sequence Length	128
Optimizer	AdamW
Data Training	80%
Data Validation	10%
Data Testing	10%
Loss Function	Cross Entropy Loss

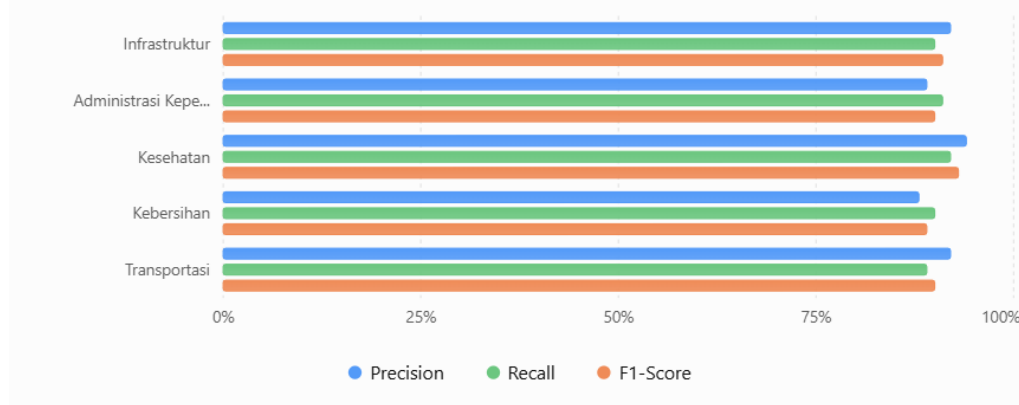
Pemilihan learning rate sebesar 2×10^{-5} dilakukan untuk menjaga proses pembaruan bobot model agar tidak terlalu besar selama proses *fine-tuning*. Sementara itu, maximum *sequence length* sebesar 128 digunakan untuk membatasi panjang masukan teks sehingga proses pelatihan dapat dilakukan secara efisien. Teks keluhan yang memiliki panjang lebih dari batas tersebut akan dipotong, sedangkan teks yang lebih pendek akan dilakukan padding sesuai kebutuhan model. Perubahan nilai *training loss* dan *validation loss* digunakan untuk mengamati proses pembelajaran model pada setiap *epoch*. Adapun hasil dari *training loss* dan *validation loss* dapat dilihat pada gambar dibawah ini:



Gambar 2 Training Loss Dan Validation Loss

Berdasarkan grafik pelatihan, *training loss* menurun dari 0,82 pada epoch pertama menjadi 0,17 pada epoch kelima. Penurunan tersebut menunjukkan bahwa model semakin mampu menyesuaikan parameter terhadap pola pada data *training*. Sementara itu, *validation loss* menurun dari 0,68 pada epoch pertama menjadi 0,27 pada epoch keempat. Pada epoch kelima, *validation loss* meningkat menjadi 0,30. Kondisi tersebut menunjukkan bahwa epoch keempat menghasilkan nilai *validation loss* terendah, sedangkan penambahan satu epoch berikutnya tidak

memberikan perbaikan pada data validasi. Berdasarkan pertimbangan tersebut, model pada *epoch* keempat dipilih sebagai model terbaik apabila pemilihan model didasarkan pada nilai *validation loss* terendah. Penggunaan *epoch* kelima sebagai model akhir perlu dijelaskan kembali apabila eksperimen aktual memang menggunakan *epoch* kelima. Pemilihan model sebaiknya tidak hanya didasarkan pada training loss, tetapi juga mempertimbangkan performa pada data validation agar potensi *overfitting* dapat dikendalikan. Berikut Adalah hasil dari Evaluasi Model yang dapat dilihat gambar dibawah ini:



Gambar 3 Hasil Evaluasi Model

Berdasarkan hasil evaluasi, model IndoBERT memperoleh *accuracy* sebesar 91% dan *macro F1-score* sebesar 90,6% pada data pengujian. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi pada sebagian besar data keluhan dengan baik. Nilai *precision* sebesar 91% menunjukkan bahwa sebagian besar prediksi kategori yang dihasilkan model sesuai dengan label sebenarnya. Sementara itu, *recall* sebesar 90,4% menunjukkan bahwa sebagian besar keluhan pada masing-masing kategori berhasil ditemukan oleh model.

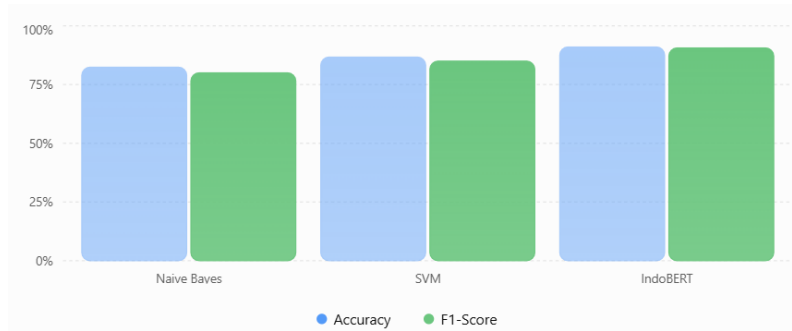
Perbedaan nilai *precision* dan *recall* pada beberapa kategori menunjukkan bahwa kemampuan model tidak sepenuhnya sama untuk setiap jenis keluhan. Kategori kesehatan memperoleh performa relatif tinggi dengan *F1-score* sebesar 93%, sedangkan kategori kebersihan memperoleh *F1-score* sebesar 89%. Perbedaan tersebut dapat dipengaruhi oleh karakteristik data. Keluhan yang memiliki kata dan konteks yang spesifik cenderung lebih mudah diklasifikasikan. Kemudian untuk mengetahui secara lebih rinci kemampuan model IndoBERT dalam membedakan setiap kategori keluhan Masyarakat, maka di lakukan *Confusion matrix*. Adapun hasil dari *Confusion matrix* seperti tabel dibawah ini:

Tabel 6. *Confusion Matrix* Model IndoBERT

Label Aktual	Infrastruktur	Administrasi	Kesehatan	Kebersihan	Transportasi	Total
Infrastruktur	243	3	1	8	5	260
Administrasi	4	232	2	1	1	240
Kesehatan	2	3	250	1	2	258
Kebersihan	12	1	2	225	10	250
Transportasi	7	2	1	11	229	250

Berdasarkan confusion matrix, sebagian besar data berada pada diagonal utama, yang menunjukkan bahwa IndoBERT mampu mengklasifikasikan keluhan dengan baik. Prediksi benar masing-masing kategori adalah 250 data kesehatan, 243 infrastruktur, 232 administrasi, 225 kebersihan, dan 229 transportasi. Kesalahan klasifikasi masih terjadi, terutama pada kategori kebersihan yang diprediksi sebagai infrastruktur dan transportasi. Hal ini dapat disebabkan oleh keluhan yang memiliki lebih dari satu konteks, seperti sampah di jalan atau halte yang kotor. Oleh karena itu, *confusion matrix* tidak hanya menunjukkan prediksi yang benar, tetapi juga pola kesalahan antar-kategori yang dapat menjadi dasar analisis lebih lanjut.

Berdasarkan hasil pengujian, kesalahan klasifikasi paling banyak terjadi pada kategori kebersihan yang diprediksi sebagai infrastruktur dan transportasi. Hal ini disebabkan oleh adanya kemiripan konteks, seperti keluhan “sampah menumpuk di sepanjang jalan” yang berkaitan dengan kebersihan dan infrastruktur. Meskipun demikian, sebagian besar data berhasil diklasifikasikan dengan benar. Hasil ini menunjukkan bahwa IndoBERT mampu mengenali konteks keluhan, meskipun masih terdapat beberapa data yang sulit dibedakan antar-kategori. Untuk mengetahui kinerja IndoBERT, dilakukan perbandingan dengan model *baseline*. Model *baseline* yang digunakan dalam penelitian ini adalah *Naive Bayes* dan *Support Vector Machine* (SVM) dengan representasi fitur TF-IDF. Berikut adalah hasil perbandingan ditunjukkan pada gambar dibawah ini:



Gambar 4 Hasil Perbandingan

Berdasarkan hasil pengujian, IndoBERT memperoleh *accuracy* 91% dan *F1-score* 90,60%, lebih tinggi dibandingkan SVM sebesar 85% dan *Naive Bayes* sebesar 80%. Perbedaan ini berkaitan dengan representasi teks, di mana TF-IDF berfokus pada frekuensi kata, sedangkan IndoBERT mempertimbangkan konteks kata dalam kalimat. Hal tersebut membuat IndoBERT lebih mampu menangani variasi istilah dan struktur kalimat pada data pengaduan. Namun, pemilihan model tetap perlu mempertimbangkan kebutuhan komputasi dan karakteristik dataset. Hasil klasifikasi IndoBERT dapat dimanfaatkan untuk membantu manajemen layanan publik dengan mengelompokkan keluhan masyarakat secara otomatis berdasarkan kategori. Penerapan hasil klasifikasi ditunjukkan pada tabel dibawah ini:

Tabel 7. Hasil Klasifikasi Keluhan dan Pemetaan Unit Terkait

Keluhan Masyarakat	Hasil Klasifikasi	Unit Terkait
"Jalan banyak berlubang dan belum diperbaiki."	Infrastruktur	Unit/Pengelola Infrastruktur
"KTP saya belum selesai sejak bulan lalu."	Administrasi	Unit Administrasi Kependudukan
"Sampah belum diangkut selama beberapa hari."	Kebersihan	Unit Kebersihan
"Obat di puskesmas sedang kosong."	Kesehatan	Unit Kesehatan
"Bus tidak beroperasi sesuai jadwal."	Transportasi	Unit Transportasi

Berdasarkan tabel tersebut, Pemetaan tersebut menunjukkan salah satu bentuk pemanfaatan hasil klasifikasi untuk mengorganisasi informasi pengaduan berdasarkan bidang layanan. Namun, penelitian ini belum menguji penerapan model secara langsung pada alur operasional pengaduan atau mengukur dampaknya terhadap waktu penyelesaian dan kualitas pelayanan. Oleh karena itu, hasil penelitian diposisikan sebagai informasi pendukung untuk pengelompokan dan analisis data pengaduan, bukan sebagai bukti bahwa penggunaan model secara langsung meningkatkan kualitas pelayanan publik.

KESIMPULAN

Berdasarkan hasil penelitian, IndoBERT mampu melakukan klasifikasi keluhan masyarakat ke dalam lima kategori, yaitu infrastruktur, administrasi kependudukan, kesehatan, kebersihan, dan transportasi. Dengan menggunakan 5.000 data keluhan dan pembagian data secara stratified, model IndoBERT memperoleh *accuracy* sebesar 91% dan *macro F1-score* sebesar 90,6%. Hasil *confusion matrix* menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar, meskipun masih terdapat kesalahan terutama pada kategori kebersihan yang memiliki kemiripan konteks dengan infrastruktur dan transportasi. Dibandingkan dengan model baseline berbasis TF-IDF, yaitu SVM dan *Naive Bayes*, IndoBERT memperoleh performa yang lebih tinggi, dengan *accuracy* masing-masing sebesar 85% dan 80%. Hal ini menunjukkan bahwa kemampuan IndoBERT dalam mempertimbangkan konteks kata dapat membantu menangani variasi bahasa pada data keluhan masyarakat. Hasil klasifikasi juga dapat dimanfaatkan untuk mengelompokkan keluhan berdasarkan bidang layanan dan memetakan keluhan kepada unit terkait. Namun, penelitian ini belum menguji penerapan model secara langsung dalam alur operasional pengaduan maupun dampaknya terhadap kualitas pelayanan. Dengan demikian, model IndoBERT dalam penelitian ini dapat digunakan sebagai pendukung proses pengelompokan dan analisis data keluhan masyarakat.

REFERENSI

- [1] R. Amelia, D. Darmansah, N. S. Prastiwi, and M. E. Purbaya, "Impementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 338, 2022, doi: 10.30865/jurikom.v9i2.3895.
- [2] D. D. Darmansah, "Perancangan Sistem Informasi Pengolahan Jadwal Mata Pelajaran Siswa Secara

- Online Di Smpn 31 Padang Berbasis Web,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 3, pp. 451–465, 2020, doi: 10.35957/jatisi.v7i3.490.
- [3] K. Zainurrohman, D. Y. Kristiyanto, and D. Darmansah, “Evaluasi Usability Website Sistem Informasi Pelayanan Perizinan Banyumas (SIPANJIMAS) Menggunakan Metode System Usability Scale (SUS),” *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 510, 2022, doi: 10.30865/json.v3i4.4141.
- [4] I. F. Rozi, V. N. Wijayaningrum, and N. Khozin, “Klasifikasi Teks Laporan Masyarakat Pada Situs Laport! Menggunakan Recurrent Neural Network,” *Sistemasi*, vol. 9, no. 3, p. 633, 2020, doi: 10.32520/stmsi.v9i3.977.
- [5] Ignatius Wisnu Prayogo, Yulia Kurniawati, Isnina Eva Hidayati, Amelia Khairunnisa, and Yova Ruldeviyani, “Text Classification of Public Complaint Validity With Deep Learning Approaches,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 15, no. 2, pp. 115–123, 2026, doi: 10.22146/jnteti.v15i2.24003.
- [6] F. Handayani and S. Pribadi, “Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110,” *J. Tek. Elektro*, vol. 7, no. 1, pp. 19–24, 2019, doi: 10.15294/jte.v7i1.8585.
- [7] M. H. S. Adani, A. Rausanfitra, and T. Mustaqim, “Comparative Analysis of High School Student and AI-Generated Essays Using IndoBERT and Linguistic Features,” *Sci. J. Informatics*, vol. 12, no. 3, pp. 525–538, 2025, doi: 10.15294/sji.v12i3.27732.
- [8] H. U. Gunawan and F. J. Kaunang, “Analisis Sentimen Multi-Aspek terhadap Ulasan Pengguna Aplikasi M-Pajak Menggunakan Model IndoBERT,” *Techno.Com*, vol. 25, no. 1, pp. 141–152, 2026, doi: 10.62411/tc.v25i1.15761.
- [9] A. Fatihaturrahmah and K. D. Tania, “Review: A Hybrid Approach of Aspect-Based Sentiment Analysis and Knowledge Extraction for Evaluating Security Perceptions in Digital Payment Applications,” *Sci. J. Informatics*, vol. 12, no. 4, pp. 635–650, 2025, doi: 10.15294/sji.v12i4.31557.
- [10] A. Elhan, M. K. D. Hardhienata, Y. Herdiyeni, S. H. Wijaya, and J. Adisantoso, “Analisis Sentimen Pengguna Twitter terhadap Vaksinasi COVID-19 di Indonesia menggunakan Algoritme Random Forest dan BERT,” *J. Ilmu Komput. dan Agri-Informatika*, vol. 9, no. 2, pp. 199–211, 2022, doi: 10.29244/jika.9.2.199-211.
- [11] M. Ikram, K. Sinapoy, Y. Sibaroni, and S. S. Prasetyowati, “Comparison of LSTM and IndoBERT Method,” *J. Resti*, vol. 5, no. 158, pp. 2–6, 2023, doi: 10.29207/resti.v7i3.4830.
- [12] F. Indriani, R. A. Nugroho, M. R. Faisal, and D. Kartini, “Comparative Evaluation of IndoBERT, IndoBERTweet, and mBERT for Multilabel Student Feedback Classification,” *J. RESTI*, vol. 8, no. 6, pp. 748–757, 2024, doi: 10.29207/resti.v8i6.6100.
- [13] S. Apriliani, A. Erfina, and C. Warman, “Fine-Tuned IndoBERT for Aspect-Based Sentiment Analysis of Indonesian Five-Star Hotel Reviews,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 4, pp. 437–445, 2025, doi: 10.32736/sisfokom.v14i4.2491.