Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal.171- 179
DOI : https://doi.org/10.47233/jsit.v6i1.4319
ISSN : **2807-7393**

Optimalisasi *Chatbot Helpdesk* TIK Berbasis *Seq2Seq Encoder BiLSTM* dan *Retrieval*

Retno Reynaldi Lakoro[a], Syahrial[b], Widya Eka Pranata[c] , Mohamad Ilyas Abas[d]
[a]Ilmu Komputer, Sains dan Ilmu Komputer, Universitas Muhammadiyyah Gorontalo, retno.lakoro@gmail.com
[b]Ilmu Komputer, Sains dan Ilmu Komputer, Universitas Muhammadiyyah Gorontalo, syahrial@umgo.ac.id
[c]Ilmu Komputer, Sains dan Ilmu Komputer, Universitas Muhammadiyyah Gorontalo, widyapranata@umgo.ac.id
[d]Ilmu Komputer, Sains dan Ilmu Komputer, Universitas Muhammadiyyah Gorontalo, ilyasabas@umgo.ac.id

Submitted: 19-01-2026, Reviewed: 20-01-2026, Accepted 29-01-2026

Abstract
The Information and Communication Technology (ICT) helpdesk service at the Bone Bolango Regency Communication and Information Agency still faces problems such as limited service hours, dependence on staff availability, and inconsistent answers to recurring questions. These conditions result in inconsistent response times, potentially reducing service effectiveness and making it difficult to meet ICT information needs quickly and consistently within the local government environment. This study aims to evaluate and optimize the performance of ICT helpdesk chatbots in understanding the context of user questions, classifying service categories, and selecting relevant and consistent answers in the closed public service domain. The approach used is a quantitative method with an experimental design, in which the chatbot model was developed using a hybrid Sequence-to-Sequence (Seq2Seq) architecture with a Bidirectional Long Short-Term Memory (BiLSTM) encoder and attention mechanism, as well as a retrieval-based response approach in selecting answers. The dataset consisted of 7,192 question-answer pairs that were manually compiled based on real cases of ICT services and grouped into four service categories, while the analysis technique was carried out through model performance evaluation using accuracy, precision, recall, F1-score, top-3 accuracy metrics, and confusion matrix analysis to assess the stability and accuracy of the classification. The results show that the model is capable of achieving classification accuracy close to 99% with a high F1-score and stable top-3 accuracy on the test data, proving that the encoder-centric BiLSTM approach with a retrieval mechanism effectively produces relevant, consistent, and controlled responses for application in domain-specific public service helpdesk chatbots.
Keywords: *BiLSTM, chatbot, IT helpdesk, retrieval, Seq2Seq*

Abstrak
Layanan helpdesk Teknologi Informasi dan Komunikasi (TIK) di Dinas Komunikasi dan Informatika Kabupaten Bone Bolango masih menghadapi permasalahan berupa keterbatasan jam layanan, ketergantungan pada ketersediaan petugas, serta inkonsistensi jawaban terhadap pertanyaan yang bersifat berulang. Kondisi tersebut menyebabkan waktu respons yang tidak seragam, berpotensi menurunkan efektivitas pelayanan, serta menyulitkan pemenuhan kebutuhan informasi TIK secara cepat dan konsisten pada lingkungan pemerintahan daerah. Penelitian ini bertujuan untuk mengevaluasi dan mengoptimalkan kinerja chatbot helpdesk TIK dalam memahami konteks pertanyaan pengguna, mengklasifikasikan kategori layanan, serta memilih jawaban yang relevan dan konsisten pada domain layanan publik tertutup. Pendekatan yang digunakan adalah metode kuantitatif dengan desain eksperimen, di mana model chatbot dikembangkan menggunakan arsitektur hibrid Sequence-to-Sequence (Seq2Seq) dengan encoder Bidirectional Long Short-Term Memory (BiLSTM) dan mekanisme attention, serta pendekatan retrieval-based response dalam pemilihan jawaban. Dataset terdiri atas 7.192 pasangan pertanyaan–jawaban yang disusun secara manual berdasarkan kasus nyata layanan TIK dan dikelompokkan ke dalam empat kategori layanan, sedangkan teknik analisis dilakukan melalui evaluasi kinerja model menggunakan metrik akurasi, precision, recall, F1-score, top-3 accuracy, serta analisis confusion matrix untuk menilai kestabilan dan ketepatan klasifikasi.
Hasil penelitian menunjukkan bahwa model mampu mencapai akurasi klasifikasi mendekati 99% dengan nilai F1-score yang tinggi dan top-3 accuracy yang stabil pada data uji, sehingga membuktikan bahwa pendekatan encoder-centric BiLSTM dengan mekanisme retrieval efektif menghasilkan respons yang relevan, konsisten, dan terkontrol untuk diterapkan pada chatbot helpdesk layanan publik berbasis domain tertutup.
Kata Kunci: BiLSTM, chatbot, helpdesk TIK, retrieval, Seq2Seq

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license

PENDAHULUAN

Pemerintah Indonesia secara berkelanjutan mendorong penerapan Sistem Pemerintahan Berbasis Elektronik (SPBE) sebagai kerangka utama transformasi digital untuk meningkatkan efektivitas tata kelola pemerintahan serta kualitas pelayanan publik. Pada tingkat pemerintah daerah, implementasi SPBE tidak hanya dimaknai sebagai digitalisasi proses administrasi, tetapi juga sebagai upaya strategis untuk menghadirkan layanan publik yang lebih responsif, transparan, dan mudah diakses oleh masyarakat maupun Organisasi Perangkat Daerah (OPD). Di Kabupaten Bone Bolango, komitmen terhadap penerapan SPBE

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal.171- 179
DOI : https://doi.org/10.47233/jsit.v6i1.4319
ISSN : **2807-7393**

secara formal diatur melalui Peraturan Bupati Nomor 23 Tahun 2023 tentang Penyelenggaraan Sistem Pemerintahan Berbasis Elektronik, yang menjadi landasan bagi pemanfaatan teknologi informasi dan komunikasi dalam mendukung kinerja pemerintahan daerah secara terintegrasi [1].

Sejalan dengan kebijakan tersebut, berbagai penelitian menunjukkan bahwa implementasi SPBE dan transformasi digital pemerintahan memberikan dampak positif terhadap efisiensi kerja, percepatan waktu layanan, serta peningkatan konsistensi informasi yang diterima oleh pengguna layanan. Studi pada Dinas Komunikasi dan Informatika Kota Palu melaporkan bahwa pemanfaatan sistem berbasis elektronik mampu meningkatkan kualitas pelayanan dan mengurangi ketergantungan terhadap proses manual [2]. Selain itu, transformasi inovasi pelayanan publik menuju pemerintahan digital dipandang sebagai prasyarat penting dalam menjawab meningkatnya tuntutan masyarakat terhadap layanan yang cepat, akurat, dan tersedia secara berkelanjutan [3]. Meskipun demikian, efektivitas implementasi SPBE sangat bergantung pada kesiapan sistem pendukung, khususnya pada unit layanan teknis yang berinteraksi langsung dengan pengguna.

Dalam struktur pemerintahan daerah, Dinas Komunikasi dan Informatika (Diskominfo) memiliki peran strategis sebagai pengelola infrastruktur TIK, pengembang sistem informasi, serta penyelenggara layanan helpdesk teknis bagi OPD. Layanan helpdesk TIK berfungsi sebagai titik kontak utama dalam menangani gangguan aplikasi, jaringan, hosting dan subdomain, serta permintaan informasi teknis lainnya. Penelitian mengenai manajemen service desk pada instansi pemerintah menunjukkan bahwa layanan helpdesk yang terkelola dengan baik berkontribusi langsung terhadap kontinuitas layanan dan stabilitas operasional sistem pemerintahan berbasis elektronik [4].

Namun, dalam praktiknya, layanan helpdesk TIK di banyak pemerintah daerah masih didominasi oleh mekanisme manual yang mengandalkan komunikasi melalui pesan singkat, panggilan telepon, atau media informal lainnya. Kondisi ini menyebabkan keterbatasan jam layanan, ketergantungan tinggi pada ketersediaan petugas, serta potensi inkonsistensi jawaban terhadap pertanyaan yang bersifat berulang. Ketika volume permintaan meningkat, waktu respons menjadi tidak seragam dan berpotensi menghambat proses penyelesaian gangguan. Permasalahan tersebut menunjukkan adanya kebutuhan akan sistem pendukung yang mampu memberikan respons awal secara cepat, konsisten, dan terstandar untuk mendukung implementasi SPBE secara optimal.

Sebagai solusi terhadap permasalahan tersebut, pemanfaatan chatbot berbasis kecerdasan buatan mulai banyak dikaji dalam konteks layanan publik. Penelitian sebelumnya melaporkan bahwa chatbot mampu meningkatkan ketersediaan layanan daring, mempercepat waktu respons, serta menjaga konsistensi informasi pada layanan online [5]. Pada instansi pemerintahan, chatbot helpdesk berbasis artificial intelligence juga terbukti efektif dalam mendukung layanan TIK publik dengan mengurangi beban kerja petugas dan meningkatkan efisiensi penanganan pertanyaan rutin [6]. Meskipun demikian, efektivitas chatbot sangat ditentukan oleh arsitektur model dan mekanisme pengendalian jawaban yang digunakan.

Dalam pengembangan chatbot berbasis pemrosesan bahasa alami, pendekatan pembelajaran mendalam telah menunjukkan keunggulan dibandingkan metode berbasis aturan atau pencocokan kata sederhana. Model Long Short-Term Memory (LSTM) mampu memodelkan ketergantungan sekuens teks dengan lebih baik, sementara Bidirectional Long Short-Term Memory (BiLSTM) memiliki keunggulan tambahan dalam menangkap konteks kalimat dari dua arah sehingga menghasilkan representasi semantik yang lebih kaya [7], [8]. Studi perbandingan juga menunjukkan bahwa BiLSTM cenderung memberikan performa yang lebih stabil dibandingkan LSTM satu arah pada berbagai tugas pemrosesan bahasa alami [9].Selain itu, arsitektur Sequence-to-Sequence (Seq2Seq) dengan mekanisme attention banyak digunakan pada sistem tanya jawab untuk memodelkan hubungan antara input dan output secara adaptif. Pendekatan ini memungkinkan model memfokuskan perhatian pada bagian input yang relevan saat membentuk representasi jawaban. Penelitian pada sistem tanya jawab berbasis bahasa Indonesia menunjukkan bahwa kombinasi Seq2Seq dan BiLSTM mampu meningkatkan pemahaman konteks pertanyaan secara signifikan [10], [11]. Namun, penggunaan chatbot generatif murni pada lingkungan layanan publik tertutup memiliki risiko menghasilkan jawaban di luar domain pengetahuan, informasi yang tidak tervalidasi, serta potensi halusinasi jawaban yang dapat berdampak pada kesalahan informasi dan keamanan data [12].

Sebaliknya, pendekatan chatbot berbasis information retrieval, seperti Term Frequency–Inverse Document Frequency (TF-IDF) dan cosine similarity, masih banyak digunakan pada layanan helpdesk karena stabil dan terkontrol. Penelitian terdahulu menunjukkan bahwa metode TF-IDF mampu memberikan respons yang relevan pada domain terbatas, namun performanya sangat bergantung pada kesesuaian kosakata dan kualitas pra-pemrosesan teks [13]. Penelitian lain juga melaporkan bahwa chatbot helpdesk berbasis TF-IDF memiliki keterbatasan dalam menangkap konteks semantik kalimat yang kompleks, sehingga kurang adaptif terhadap variasi redaksi pertanyaan pengguna [14]. Kondisi ini menunjukkan adanya celah penelitian untuk mengembangkan pendekatan hibrid yang menggabungkan pemahaman konteks berbasis pembelajaran mendalam dengan kestabilan pemilihan jawaban berbasis retrieval.

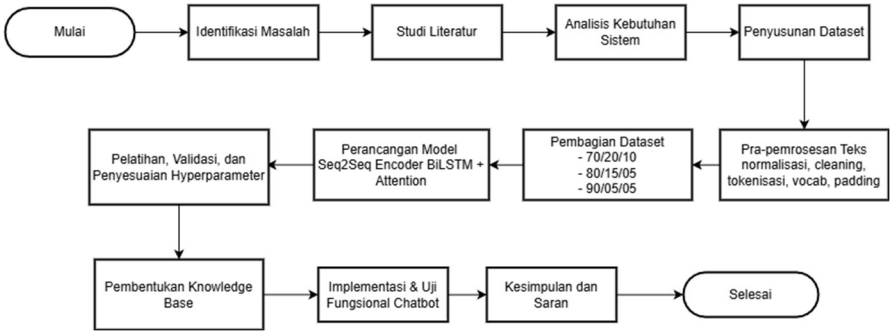
Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal.171- 179
DOI : https://doi.org/10.47233/jsit.v6i1.4319
ISSN : **2807-7393**

Berdasarkan tinjauan tersebut, hingga saat ini belum banyak penelitian yang secara khusus mengembangkan dan mengevaluasi chatbot helpdesk layanan TIK pada lingkungan pemerintah daerah menggunakan pendekatan encoder-centric Seq2Seq dengan BiLSTM dan mekanisme attention yang dikombinasikan dengan retrieval-based response, khususnya pada domain layanan publik tertutup. Oleh karena itu, penelitian ini menawarkan **novelty** berupa penerapan arsitektur Seq2Seq encoder-centric BiLSTM dengan attention, di mana decoder hanya digunakan pada tahap pelatihan untuk memperkuat pembelajaran representasi, sedangkan pada tahap inferensi sistem sepenuhnya mengandalkan embedding encoder untuk klasifikasi dan pencarian jawaban yang aman dan terkontrol.

Ruang lingkup penelitian ini dibatasi pada domain layanan helpdesk TIK Diskominfo Kabupaten Bone Bolango, dengan jenis pertanyaan yang mencakup kebijakan pemerintah, hosting dan subdomain, layanan aplikasi, serta layanan jaringan. Tujuan penelitian ini adalah mengevaluasi kinerja chatbot helpdesk secara kuantitatif dalam mengklasifikasikan kategori layanan dan memilih jawaban yang relevan, yang diukur menggunakan metrik akurasi, precision, recall, F1-score, top-k accuracy, serta analisis waktu respons sistem. Pemilihan metrik tersebut merujuk pada kajian evaluasi model klasifikasi dan sistem kecerdasan buatan yang menekankan pentingnya pengukuran performa secara komprehensif dan terukur [15], [16]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi praktis dan ilmiah dalam pengembangan chatbot helpdesk yang andal, aman, dan sesuai dengan kebutuhan implementasi SPBE di lingkungan pemerintah daerah.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental untuk mengembangkan dan mengevaluasi chatbot helpdesk layanan Teknologi Informasi dan Komunikasi (TIK) pada lingkungan pemerintahan daerah. Pendekatan ini dipilih untuk memungkinkan pengujian kinerja sistem secara terukur melalui metrik evaluasi kuantitatif, sekaligus pengujian secara praktis melalui integrasi sistem ke dalam aplikasi web chatbot, sebagaimana pendekatan chatbot helpdesk pada instansi pemerintahan juga banyak dikaji pada penelitian sebelumnya [6]. Selain itu, pendekatan eksperimental dilakukan melalui pengujian konfigurasi dan parameter model secara sistematis untuk memperoleh performa terbaik berdasarkan indikator kinerja yang terukur, sejalan dengan penelitian terdahulu yang menekankan pentingnya evaluasi eksperimental dalam menentukan konfigurasi optimal pada sistem berbasis kecerdasan buatan [17]. Proses penelitian disusun secara berurutan mulai dari penyusunan dataset, pra-pemrosesan teks, perancangan arsitektur model, pelatihan dan validasi, pengujian pada data uji, hingga uji fungsional sistem [5].



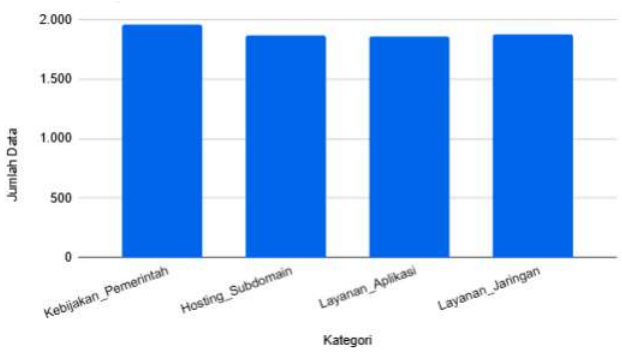
Gambar 1. Diagram Alur Penelitian

Berdasarkan Gambar 1, tahapan penelitian diawali dengan pengumpulan dan penyusunan dataset pertanyaan–jawaban layanan TIK. Dataset kemudian diproses melalui tahapan pra-pemrosesan teks agar siap digunakan dalam model pembelajaran mendalam. Tahap selanjutnya adalah perancangan arsitektur model chatbot berbasis Seq2Seq encoder BiLSTM dengan mekanisme attention, yang dilanjutkan dengan proses pelatihan dan validasi model menggunakan beberapa skenario pembagian data. Setelah model terbaik diperoleh, pengujian dilakukan pada data uji, kemudian model diintegrasikan ke dalam aplikasi web chatbot untuk dilakukan uji fungsional [6].

Dataset penelitian terdiri atas 7.192 pasangan pertanyaan–jawaban yang disusun secara manual berdasarkan kasus nyata layanan helpdesk TIK di Dinas Komunikasi dan Informatika Kabupaten Bone Bolango. Data dikumpulkan dari dokumentasi internal layanan, catatan penanganan gangguan, serta prosedur layanan resmi yang berlaku. Setiap pasangan pertanyaan–jawaban diberi label kategori layanan secara manual ke dalam empat kategori, yaitu kebijakan pemerintah, hosting dan subdomain, layanan aplikasi, dan layanan jaringan. Proses pelabelan dilakukan secara konsisten dan diverifikasi secara internal untuk memastikan

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal.171- 179
DOI : https://doi.org/10.47233/jsit.v6i1.4319
ISSN : **2807-7393**

kesesuaian antara pertanyaan, jawaban, dan kategori layanan, selaras dengan pengembangan sistem layanan informasi berbasis tanya-jawab yang juga menekankan konsistensi data dan struktur informasi [18].



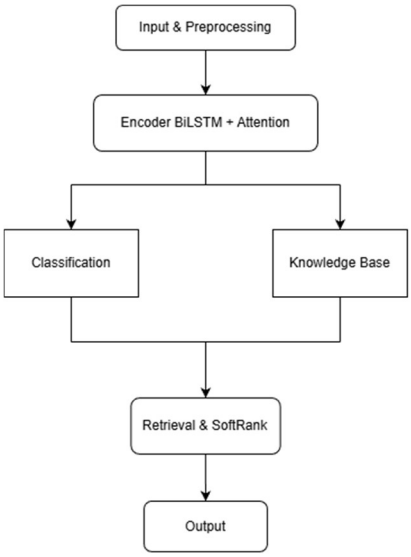
Gambar 2. Distribusi Dataset Berdasarkan Kategori Layanan

Berdasarkan Gambar 2, jumlah data pada setiap kategori layanan relatif seimbang, sehingga model memiliki peluang yang proporsional untuk mempelajari pola pertanyaan dari masing-masing kategori dan mengurangi potensi bias akibat ketimpangan kelas.

Sebelum digunakan dalam proses pelatihan, dataset melalui tahapan pra-pemrosesan teks yang meliputi normalisasi huruf, pembersihan karakter non-informatif, tokenisasi berbasis kata, pembentukan kosakata, serta padding untuk menyeragamkan panjang sekuens. Tahapan pra-pemrosesan ini bertujuan mengurangi variasi kata yang tidak relevan serta menstabilkan proses pembelajaran model, sejalan dengan kajian yang menekankan pengaruh preprocessing terhadap kualitas representasi teks [19], [20].

Dataset kemudian dibagi ke dalam data latih, data validasi, dan data uji menggunakan teknik stratified random split dengan beberapa skenario rasio, yaitu 70:20:10, 80:15:5, dan 90:5:5. Pembagian data dilakukan menggunakan nilai random seed yang tetap untuk menjamin reprodusibilitas eksperimen serta mencegah terjadinya data leakage antara data latih, validasi, dan data uji. Strategi pemisahan data ke dalam set pelatihan, validasi, dan pengujian digunakan untuk memastikan evaluasi yang objektif serta menjaga generalisasi model, sebagaimana praktik evaluasi pada studi klasifikasi berbasis pembelajaran mendalam yang juga memisahkan data menjadi train–validation–test dan menilai kinerja menggunakan confusion matrix [21].

Arsitektur chatbot dikembangkan menggunakan pendekatan Seq2Seq encoder-centric, dengan encoder Bidirectional Long Short-Term Memory (BiLSTM) dan mekanisme attention sebagai komponen utama dalam membangun representasi konteks pertanyaan. Pemilihan BiLSTM didasarkan pada kemampuannya memodelkan konteks dua arah sehingga representasi sekuens lebih kaya dibanding model satu arah, sebagaimana ditunjukkan pada penerapan BiLSTM untuk klasifikasi intent chatbot [8]. Selain itu, penggunaan kerangka Seq2Seq dan BiLSTM pada sistem tanya-jawab juga telah digunakan pada penelitian terdahulu sehingga relevan sebagai dasar perancangan arsitektur [10].



Gambar 3. Arsitektur Model Seq2Seq BiLSTM Retrieval

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal.171- 179
DOI : https://doi.org/10.47233/jsit.v6i1.4319
ISSN : **2807-7393**

Gambar 3 menunjukkan bahwa encoder BiLSTM berfungsi sebagai inti sistem dalam menghasilkan embedding kontekstual dari pertanyaan pengguna. Decoder Seq2Seq digunakan pada tahap pelatihan dengan mekanisme teacher forcing untuk memperkuat pembelajaran representasi encoder, sebagaimana umum digunakan pada pengembangan chatbot berbasis Seq2Seq [11]. Proses pelatihan dilakukan menggunakan optimizer Adam dengan learning rate sebesar 2×10^{-4}, batch size 64, jumlah epoch 100, dropout 0,6, serta fungsi loss berupa cross-entropy untuk mengoptimalkan kinerja model.

Pada tahap inferensi, pertanyaan pengguna dipetakan ke dalam ruang vektor menggunakan encoder BiLSTM. Representasi vektor tersebut kemudian dibandingkan dengan seluruh vektor pertanyaan dalam basis pengetahuan menggunakan cosine similarity sebagai ukuran kemiripan semantik. Jawaban dipilih dari pasangan pertanyaan–jawaban dengan nilai kemiripan tertinggi yang memenuhi ambang batas tertentu. Apabila nilai kemiripan berada di bawah ambang batas yang ditetapkan, sistem tidak memaksakan jawaban dan memberikan respons alternatif berupa permintaan klarifikasi, sehingga mengurangi risiko jawaban yang tidak relevan atau berada di luar domain layanan.

Untuk memperkuat posisi pendekatan yang diusulkan, kinerja model dalam penelitian ini dibandingkan secara konseptual dengan pendekatan chatbot helpdesk berbasis information retrieval klasik yang telah banyak diterapkan pada penelitian terdahulu, khususnya metode TF-IDF dan Cosine Similarity. Penelitian oleh Setiawan dan Adnyana menunjukkan bahwa chatbot helpdesk berbasis TF-IDF dan Cosine Similarity mampu memberikan respons yang relevan dengan tingkat question recognition dan F1-score yang moderat, namun performanya sangat bergantung pada ukuran dataset, kualitas pra-pemrosesan, serta kesesuaian kosakata antara pertanyaan pengguna dan basis pengetahuan [14]. Temuan serupa mengenai penggunaan TF-IDF dan cosine similarity pada chatbot layanan informasi juga dilaporkan pada studi lain [13].

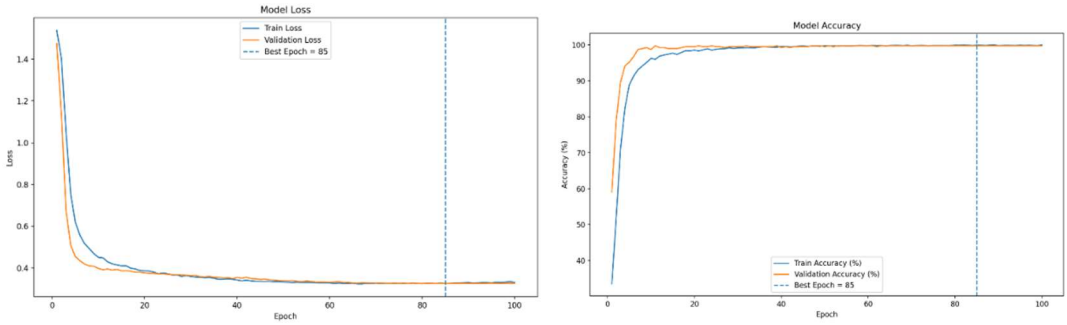
Evaluasi kinerja model dilakukan pada data uji menggunakan metrik accuracy, precision, recall, F1-score, top-3 accuracy, serta analisis confusion matrix untuk mengamati pola kesalahan klasifikasi. Penggunaan metrik klasifikasi tersebut mengacu pada kajian tentang ukuran performa yang umum digunakan untuk evaluasi model klasifikasi [16], serta prinsip evaluasi yang menekankan kesesuaian metrik dengan konteks penerapan sistem [15]. Seluruh eksperimen dijalankan pada lingkungan komputasi berbasis server menggunakan bahasa pemrograman Python dan framework PyTorch [22], [23]. Selain evaluasi kuantitatif, dilakukan uji fungsional pada aplikasi web chatbot dengan menggunakan contoh pertanyaan layanan nyata untuk memastikan sistem mampu memberikan respons yang relevan dan sesuai dengan kebutuhan operasional helpdesk [6].

HASIL DAN PEMBAHASAN

Hasil Penelitian

Model chatbot helpdesk layanan TIK dikembangkan dan dievaluasi melalui tahapan pelatihan, validasi, dan pengujian akhir menggunakan tiga skenario pembagian data, yaitu 70:20:10, 80:15:5, dan 90:5:5. Pendekatan yang digunakan bersifat eksperimental berbasis sistem, di mana data observasi diperoleh dari hasil pengujian kinerja model serta pengamatan terhadap perilaku sistem.

Performa model selama proses pelatihan diamati melalui perubahan nilai akurasi dan loss pada data pelatihan dan validasi.



Gambar 4. Grafik Loss Pelatihan dan Validasi 90/5/5

Berdasarkan Gambar 4, terlihat bahwa proses pelatihan berlangsung secara konvergen. Akurasi pelatihan dan validasi meningkat secara konsisten, sementara nilai loss menurun hingga mencapai titik optimal. Pola ini menunjukkan bahwa model mampu mempelajari representasi konteks pertanyaan secara stabil tanpa indikasi overfitting yang signifikan.

Ringkasan hasil pelatihan pada epoch terbaik untuk masing-masing eksperimen pembagian data disajikan

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal.171- 179
DOI : https://doi.org/10.47233/jsit.v6i1.4319
ISSN : **2807-7393**

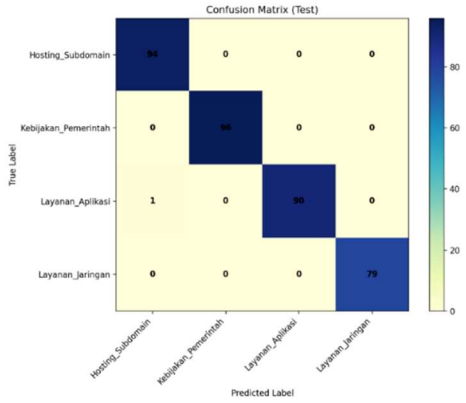
pada Tabel 1. Epoch terbaik ditentukan berdasarkan nilai *validation loss* terendah pada setiap skenario.

Tabel 1. Hasil eksperimen berdasarkan skenario pembagian data

Skenario Split	Epoch Terbaik	Train Loss	Train Accuracy (%)	Validation Loss	Validation Accuracy (%)
70 : 20 : 10	85	0,3341	99,72	0,3347	99,46
80 : 15 : 5	99	0,3304	99,91	0,3231	99,54
90 : 5 : 5	85	0,3240	99,91	0,3222	99,74

Evaluasi pada data validasi menunjukkan bahwa seluruh skenario menghasilkan akurasi di atas 99%. Distribusi nilai precision, recall, dan f1-score untuk masing-masing kategori layanan memperlihatkan kinerja yang konsisten antar kelas. Pada skenario 70:20:10, nilai f1-score untuk keempat kategori berada pada rentang 0,99–1,00 dengan total data validasi sebanyak 1.438 data. Epoch terbaik dipilih berdasarkan nilai validation loss terendah pada masing-masing skenario split data.

Analisis lebih lanjut dilakukan menggunakan confusion matrix yang ditampilkan pada Gambar 5.



Gambar 5. Confusion Matrix Klasifikasi Kategori Layanan pada Data Uji

Sebagian besar data berhasil diklasifikasikan dengan benar ke dalam kategori layanan yang sesuai. Kesalahan klasifikasi ditemukan dalam jumlah yang sangat terbatas dan tidak terkonsentrasi pada satu kategori tertentu.

Ringkasan akurasi hasil pengujian pada data uji disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Akhir Model Pada Data Uji

Skenario Split	Jumlah Data Uji	Akurasi Data Uji (%)
70 : 20 : 10	720	99,17
80 : 15 : 5	360	99,44
90 : 5 : 5	360	99,72

Hasil pada Tabel 2 menunjukkan bahwa model mampu mempertahankan performa yang tinggi pada seluruh skenario pembagian data. Skenario 90:5:5 menghasilkan akurasi tertinggi, sejalan dengan hasil pelatihan dan validasi.

Selain evaluasi numerik, dilakukan pengujian fungsional chatbot helpdesk TIK untuk memastikan sistem berjalan sesuai dengan tujuan perancangan. Pengujian ini dilakukan dengan memberikan sejumlah pertanyaan layanan TIK yang merepresentasikan kasus nyata, seperti permohonan reset akun aplikasi, gangguan layanan jaringan, serta permintaan informasi terkait hosting dan subdomain OPD. Setiap pertanyaan diuji untuk mengamati kemampuan sistem dalam mengklasifikasikan kategori layanan dan memilih jawaban yang sesuai dari basis pengetahuan melalui mekanisme retrieval.

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal.171- 179
DOI : https://doi.org/10.47233/jsit.v6i1.4319
ISSN : **2807-7393**



Gambar 6. Tampilan hasil uji fungsional chatbot helpdesk TIK

Gambar 6 menampilkan contoh antarmuka chatbot saat menerima pertanyaan pengguna dan menampilkan hasil klasifikasi kategori layanan beserta jawaban yang dipilih berdasarkan nilai kemiripan semantik tertinggi. Hasil pengujian menunjukkan bahwa chatbot mampu memberikan respons yang sesuai dengan konteks pertanyaan dan menampilkan jawaban terstandar dari basis pengetahuan layanan TIK, serta tidak menghasilkan respons di luar domain layanan yang telah ditetapkan.

Berdasarkan observasi internal selama uji fungsional, staf layanan helpdesk menilai bahwa chatbot membantu mempercepat pemberian respons awal terhadap pertanyaan berulang dan mengurangi ketergantungan pada pencarian manual dokumen layanan. Temuan ini menunjukkan bahwa penerapan chatbot tidak hanya efektif secara kuantitatif, tetapi juga relevan secara operasional dalam mendukung aktivitas layanan helpdesk TIK.

Pembahasan

Hasil pengujian menunjukkan bahwa perbedaan proporsi pembagian data memberikan pengaruh terhadap kinerja model chatbot helpdesk layanan TIK. Pada ketiga skenario pembagian data yang diuji, seluruh model mampu mencapai tingkat akurasi di atas 99%. Namun demikian, terdapat perbedaan performa yang konsisten, di mana skenario 90:5:5 menunjukkan kinerja terbaik dibandingkan skenario 70:20:10 dan 80:15:5, baik pada tahap validasi maupun pengujian akhir. Pengujian dilakukan menggunakan satu kali pelatihan untuk setiap skenario dengan penentuan epoch terbaik berdasarkan *validation loss*, sebagaimana umum dilakukan pada penelitian chatbot berbasis dataset domain tertutup.

Performa terbaik pada skenario 90:5:5 dapat dijelaskan oleh dominasi data latih yang lebih besar dibandingkan data validasi dan data uji. Proporsi data latih yang tinggi memungkinkan encoder BiLSTM mempelajari representasi sekuens pertanyaan secara lebih komprehensif, terutama dalam menangkap variasi linguistik dan konteks layanan TIK. Hal ini sejalan dengan karakteristik model berbasis pembelajaran mendalam yang sangat bergantung pada ketersediaan data latih untuk membangun representasi fitur yang stabil dan diskriminatif.

Jika dibandingkan antar skenario, skenario 70:20:10 menunjukkan nilai validation loss yang relatif lebih tinggi. Hal ini mengindikasikan bahwa meskipun jumlah data validasi lebih besar, kapasitas data latih pada skenario tersebut belum sepenuhnya optimal untuk memaksimalkan kemampuan generalisasi model. Skenario 80:15:5 memperlihatkan peningkatan performa, namun masih berada di bawah skenario 90:5:5. Dengan demikian, dapat disimpulkan bahwa peningkatan proporsi data latih berkontribusi langsung terhadap penurunan validation loss dan peningkatan akurasi model.

Untuk memvalidasi keunggulan pendekatan yang diusulkan, hasil penelitian ini dibandingkan secara konseptual dengan pendekatan chatbot helpdesk berbasis information retrieval dan pembobotan kata, seperti model TF-IDF dan Cosine Similarity yang umum digunakan pada penelitian sebelumnya [24]. Pendekatan tersebut mengukur kemiripan pertanyaan berdasarkan representasi vektor kata dan kesamaan statistik antar dokumen, sehingga efektivitasnya sangat bergantung pada kesesuaian kosakata dan frekuensi kata dalam dataset [24]. Pada konteks layanan helpdesk, pendekatan berbasis TF-IDF dan Cosine Similarity telah dilaporkan mampu memberikan respons yang relevan, namun masih memiliki keterbatasan dalam menangkap konteks sekuensial dan variasi redaksi pertanyaan pengguna.

Analisis terhadap confusion matrix menunjukkan bahwa kesalahan klasifikasi yang terjadi bersifat sangat terbatas dan tidak terdistribusi secara merata pada seluruh kategori. Kesalahan utama ditemukan pada pertanyaan kategori Layanan_Aplikasi yang diprediksi sebagai Hosting_Subdomain. Hal ini dapat dijelaskan

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal.171- 179
DOI : https://doi.org/10.47233/jsit.v6i1.4319
ISSN : **2807-7393**

oleh adanya irisan semantik antar kategori layanan, di mana permasalahan aplikasi sering kali berkaitan langsung dengan aspek hosting atau infrastruktur server. Temuan ini menunjukkan bahwa kesalahan klasifikasi bukan disebabkan oleh kegagalan model dalam memahami konteks, melainkan oleh kedekatan domain antar kategori layanan.

Meskipun akurasi yang diperoleh tergolong sangat tinggi, hasil penelitian ini perlu dipahami dengan mempertimbangkan potensi bias dataset. Dataset yang digunakan merupakan dataset domain tertutup yang disusun dari layanan resmi dengan kategori yang telah terdefinisi secara jelas. Kondisi ini berpotensi meningkatkan performa model karena variasi konteks di luar domain layanan tidak diikutsertakan. Oleh karena itu, klaim kinerja model dibatasi pada konteks layanan helpdesk TIK dan tidak digeneralisasikan secara langsung ke domain lain tanpa proses adaptasi tambahan.

Selain evaluasi kuantitatif, pengujian fungsional chatbot memberikan gambaran mengenai kesiapan sistem untuk digunakan secara operasional. Pengujian dilakukan dengan mengamati kemampuan chatbot dalam mengidentifikasi kategori layanan, menampilkan jawaban yang sesuai dengan standar layanan, serta menjaga konsistensi respons terhadap variasi redaksi pertanyaan pengguna. Hasil pengujian menunjukkan bahwa chatbot mampu memenuhi seluruh indikator tersebut, sehingga pendekatan encoder dan mekanisme retrieval yang digunakan tidak hanya efektif secara numerik, tetapi juga relevan untuk penggunaan nyata pada layanan helpdesk TIK.

Secara keseluruhan, hasil pembahasan ini menunjukkan bahwa kombinasi arsitektur Seq2Seq encoder BiLSTM, mekanisme attention, dan pendekatan retrieval-based response memberikan keunggulan yang signifikan dalam konteks chatbot layanan pemerintahan. Pendekatan ini mampu menghasilkan sistem yang akurat, stabil, dan terkontrol, sesuai dengan kebutuhan layanan publik yang menuntut konsistensi dan keandalan informasi.

KESIMPULAN

Penelitian ini berhasil mengembangkan chatbot helpdesk layanan TIK berbasis arsitektur Seq2Seq dengan encoder BiLSTM, mekanisme attention, dan pendekatan retrieval-based response. Hasil pengujian menunjukkan bahwa model mencapai kinerja terbaik pada skenario pembagian data 90:5:5 dengan akurasi data uji sebesar 99,72% serta nilai precision, recall, dan f1-score yang konsisten pada seluruh kategori layanan, sehingga mampu mengklasifikasikan pertanyaan layanan TIK secara akurat pada domain tertutup.

Pendekatan encoder-centric terbukti efektif dalam memperkuat representasi konteks pertanyaan tanpa ketergantungan pada proses generatif saat inferensi, sementara mekanisme retrieval memastikan konsistensi jawaban sesuai standar layanan resmi. Secara praktis, sistem berpotensi meningkatkan efisiensi layanan helpdesk melalui percepatan waktu respons dan pengurangan beban kerja petugas terhadap pertanyaan berulang.

Penelitian ini memiliki keterbatasan pada penggunaan dataset lokal dengan cakupan domain yang sempit, sehingga kinerja model berpotensi dipengaruhi oleh bias dataset. Penelitian selanjutnya disarankan untuk melakukan evaluasi pembanding dengan metode baseline lain, memperluas dataset ke domain yang lebih beragam, serta melibatkan pengujian pengguna dalam skala yang lebih luas.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Dinas Komunikasi dan Informatika Kabupaten Bone Bolango sebagai lokasi pelaksanaan penelitian serta Universitas Muhammadiyah Gorontalo sebagai institusi akademik yang memberikan dukungan pembimbingan dan fasilitas akademik.

DAFTAR PUSTAKA

[1] B. B. Bolango, “Peraturan Bupati Bone Bolango Nomor 23 Tahun 2023 tentang Penyelenggaraan Sistem Pemerintahan Berbasis Elektronik di Lingkungan Pemerintah Kabupaten Bone Bolango,” 2023, *Pemerintah Kabupaten Bone Bolango, Bone Bolango*.

[2] Muhammad Fadhil Ramli, “IMPLEMENTASI SISTEM PEMERINTAHAN BERBASIS ELEKTRONIK TERHADAP PENINGKATAN PELAYANAN DI DINAS KOMUNIKASI DAN INFORMATIKA KOTA PALU PROVINSI SULAWESI TENGAH”.

[3] M. S. J. Sangaji and J. Irianto, “Transformasi Inovasi Pelayanan Publik menuju Pemerintahan Digital,” *Jejaring Adm. Publik*, vol. 17, no. 1, pp. 54–70, 2025, doi: 10.20473/jap.v17i1.72708.

[4] A. Pamurjadi and B. P. P, “Analisis dan Rancangan Manajemen Service Desk di Instansi Pemerintah (Studi Kasus: Badan Litbang Pertanian),” *J. IPTEKKOM J. Ilmu Pengetah. Teknol. Inf.*, vol. 16, no. 1, p. 69, 2014, doi: 10.33164/iptekkom.16.1.2014.69-88.

[5] M. Muliyono and S. Sumijan, “Identifikasi Chatbot dalam Meningkatkan Pelayanan Online Menggunakan Metode Natural Language Processing,” *J. Inform. Ekon. Bisnis*, vol. 3, pp. 142–147, 2021, doi: 10.37034/infeb.v3i4.102.

[6] K. Kartini, M. Malabay, and R. Widayanti, “Pemanfaatan Artificial Intelligence Dalam Implementasi Chatbot

ISSN : **2807-7393**

Helpdesk untuk Mendukung Layanan TIK Publik pada Instansi Pemerintahan," *Bull. Comput. Sci. Res.*, vol. 5, no. 5, pp. 1235–1248, 2025, doi: 10.47065/bulletincsr.v5i5.769.

[7] M. E. Rianto, M. Maulidiansyah, and A. Tholib, "Implementasi AI Chatbot Sebagai Support Assistant Website Universitas Nurul Jadid Menggunakan Algoritma Long Short-Term Memory (LSTM)," *J. Electr. Eng. Comput.*, vol. 6, no. 1, pp. 267–275, 2024, doi: 10.33650/jeecom.v6i1.8556.

[8] Y. Cahyadi, S. Redjeki, A. Almagrib, B. Satriani, and N. Naufal, "Bidirectional Long Short-Term Memory Model for Intent Classification in Customer Service Chatbot," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 1, p. 296, 2025, doi: 10.30811/jaise.v5i1.6520.

[9] R. Yafik and M. Azhari, "Analisis Perbandingan Metode LSTM dan BiLSTM untuk Prediksi Harga Saham Menggunakan Alpha Vantage," vol. 4, no. 3, pp. 1549–1557, 2026.

[10] K. Sintiya, F. A. Bachtiar, and B. D. Setiawan, "Penerapan Arsitektur Seq2Seq dan Bidirectional Long Short-Term Memory pada Sistem Tanya Jawab Fakultas Ilmu Komputer Universitas Brawijaya," *... Teknol. Inf. dan Ilmu ...*, vol. 9, no. 8, pp. 1–10, 2025, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15253%0Ahttps://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/15253/6775

[11] M. Nafi' Baliyamalkan, R. Setya Perdana, and P. P. Adikara, "Pembangkitan Respons pada Model Seq2seq Chatbot Berbahasa Indonesia dengan Multimodal Learning (Intensi dan Entitas)," vol. 1, no. 1, pp. 2548–964, 2017, [Online]. Available: http://j-ptiik.ub.ac.id

[12] F. I. Haekal, R. Setya Perdana, and P. P. Adikara, "Sistem Tanya Jawab Closed-Domain terhadap Dokumen Fatwa menggunakan Retrieval Augmented Generation dan Large Language Model," vol. 9, no. 5, pp. 2548–964, 2025, [Online]. Available: http://j-ptiik.ub.ac.id

[13] B. Shaib Pramono, M. Alwi Nur Fathihah, M. Azizul Dzikri, and I. Maulana Cakra Dwi Noto, "Penerapan TF-IDF dan Cosine Similarity dalam Chatbot VirtualAsisten: Studi Kasus Layanan Informasi Pendaftaran SekolahMenengah Pertama," vol. 4, no. 1, pp. 150–158, 2025.

[14] G. H. Setiawan and I. M. B. Adnyana, "Improving Helpdesk Chatbot Performance with Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity Models," *J. Appl. Informatics Comput.*, vol. 7, no. 2, pp. 252–257, 2023, doi: 10.30871/jaic.v7i2.6527.

[15] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Sci. Rep.*, vol. 12, no. 1, pp. 1–9, 2022, doi: 10.1038/s41598-022-09954-8.

[16] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[17] S. Syahrial and R. Lamusu, "Pembentukan Pola Desain Motif Karawo Gorontalo Menggunakan K-Means Color Quantization dan Structured Forest Edge Detecion," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, pp. 625–634, 2021, doi: 10.25126/jtiik.2021834491.

[18] M. I. R. Dinar Nur Safitri, "Rancang Bangun Penyedia Layanan Informasi Pelayanan," *J. Ilmu Komput. dan Desain Komun. Vis.*, vol. 6, no. 2, pp. 1–10, 2021.

[19] A. E. Budiman and A. Widjaja, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 3, pp. 475–488, 2020, doi: 10.28932/jutisi.v6i3.2892.

[20] D. Apriliani, S. F. Handayani, and I. T. Saputra, "Implementasi Natural Language Processing (NLP) Dalam Pengembangan Aplikasi Chatbot Pada SMK YPE Nusantara Slawi Implementation of Natural Language Processing (NLP) in the Development of a Chatbot Application at SMK YPE Nusantara Slawi," *Techno.COM*, vol. 22, no. 4, pp. 1037–1047, 2023.

[21] M. I. Abas, S. Syarif, I. Nurtanio, and Z. Tahir, "Comparison of Convolutional Neural Network Methods for the Classification of Maize Plant Diseases," 2024. doi: 10.26594/register.v10i1.3656.

[22] M. R. S. Alfarizi, M. Z. Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, "Use of Python as a Programming Language for Machine Learning and Deep Learning," *Sci. Work Students Belief Monoth. (KARIMAH TAUHID)*, vol. 2, no. 1, pp. 1–6, 2023.

[23] I. Iswahyudi, D. Hindarto, and H. Santoso, "PyTorch Deep Learning for Food Image Classification with Food Dataset," *Sinkron*, vol. 8, no. 4, pp. 2651–2661, 2023, doi: 10.33395/sinkron.v8i4.12987.

[24] M. Mustaqim *et al.*, "Pengembangan Chatbot Untuk Layanan Pimpinan Daerah Muhammadiyah Kota Yogyakarta Menggunakan Metode Rule-based," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 2, pp. 1–15, 2024, doi: 10.28946/lexl.v5i2.2351.