BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 929-938
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.728

Question Answering System Zakat dengan Metode Long Short-Term Memory (LSTM)

Moch Apip Tanuwijaya*, Jumadi, Eva Nurlatifah

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sunan Gunung Djati, Bandung, Indonesia
Email: 1,*afiftanuwijaya@icloud.com, 2jumadi@uinsgd.ac.id, 3evanurlatifah@uinsgd.ac.id
Email Penulis Korespondensi: afiftanuwijaya@icloud.com

Abstrak−Zakat merupakan salah satu pilar penting dalam sistem keuangan Islam yang berfungsi sebagai instrumen redistribusi kekayaan. Namun, belum tersedia sistem Question Answering System (QAS) berbahasa Indonesia yang mampu memberikan jawaban otomatis dan kontekstual terkait zakat. Penelitian ini bertujuan mengembangkan sistem QAS zakat berbasis Long Short-Term Memory (LSTM) yang terintegrasi dengan platform Telegram. Dataset dikumpulkan dari buku panduan resmi BAZNAS dan diproses melalui tokenisasi, padding, dan label encoding. Arsitektur model terdiri dari layer embedding, dua layer LSTM bertingkat (dengan return_sequences, dropout, dan recurrent dropout), serta dua layer dense bertingkat (200 dan 100 unit) dengan dropout tambahan sebelum output softmax. Model dilatih menggunakan Adam optimizer (learning rate 0.003), batch size 24, dan 100 epoch. Evaluasi dilakukan menggunakan confusion matrix dengan hasil akurasi validasi sebesar 93%, precision 0.94, recall 0.93, dan F1-score 0.92 (weighted average). Sistem diintegrasikan ke Telegram Bot API dengan respons di bawah dua detik dan mampu menangani ratusan label kelas secara stabil. Sistem ini menunjukkan potensi dalam mendukung edukasi zakat digital dan dapat dikembangkan lebih lanjut dalam ekosistem Islamic Finance Technology dan Digital Religious Education.

Kata Kunci: Question Answering System; LSTM; NLP; Zakat; Telegram Bot API; Islamic Finance Technology; Digital Religious Education

Abstract−Zakat is a fundamental pillar of Islamic finance that serves as a mechanism for wealth redistribution. However, there is currently no Indonesian-language Question Answering System (QAS) capable of automatically and contextually responding to zakat-related queries. This study aims to develop a zakat-focused QAS using a Long Short-Term Memory (LSTM) model integrated into the Telegram platform. The dataset was compiled from the official BAZNAS zakat guidebook and processed through tokenization, padding, and label encoding. The model architecture consists of an embedding layer, two stacked LSTM layers (with return sequences, dropout, and recurrent dropout), followed by two dense layers (200 and 100 units) with additional dropout layers before the softmax output. The model was trained using the Adam optimizer (learning rate 0.003), a batch size of 24, and 100 epochs. Evaluation was conducted using a confusion matrix, resulting in a validation accuracy of 93%, with a precision of 0.94, recall of 0.93, and F1-score of 0.92 (weighted average). The system was deployed via the Telegram Bot API and demonstrated response times under two seconds, with stable performance across hundreds of question labels. This work contributes to the advancement of digital zakat education and presents a scalable solution that can be further extended within the ecosystem of Islamic Finance Technology and Digital Religious Education.

Keywords: Question Answering System; LSTM; NLP; Zakat; Telegram Bot API; Islamic Finance Technology; Digital Religious Education

1. PENDAHULUAN

Zakat adalah salah satu rukun Islam yang memiliki peranan penting dalam aspek sosial dan ekonomi umat Muslim [1]. Sebagai kewajiban, umat Islam diperintahkan untuk menyisihkan sebagian harta mereka dengan tujuan untuk membersihkan kekayaan yang dimiliki, di mana dalam harta tersebut terdapat hak-hak orang yang membutuhkan. Zakat berfungsi sebagai sarana untuk mendistribusikan kekayaan dari individu kepada kelompok yang berhak menerimanya [2]. Menurut Undang-Undang Nomor 23 Tahun 2011 tentang Pengelolaan Zakat, zakat memiliki potensi besar untuk mengurangi kemiskinan dan kesenjangan sosial di Indonesia. Berdasarkan data dari Badan Amil Zakat Nasional (Baznas), potensi zakat di Indonesia mencapai lebih dari 300 triliun rupiah [3]. Pada tahun 2022, pengumpulan zakat mencapai Rp22,475 triliun, yang kemudian disalurkan kepada 33,9 juta jiwa mustahik [4]. Artinya, potensi besar tersebut belum sepenuhnya dapat dioptimalkan. Salah satu penyebabnya adalah keterbatasan akses informasi zakat yang praktis dan mudah dipahami masyarakat.

Sebagian besar masyarakat masih mengalami kesulitan dalam memahami hukum, jenis, dan tata cara zakat, terutama dari sumber resmi yang tersedia. Selain itu, minimnya sistem tanya jawab zakat berbahasa Indonesia yang interaktif turut memperburuk aksesibilitas informasi. Beberapa sistem layanan digital yang telah tersedia, seperti chatbot resmi BAZNAS, masih bersifat *rule-based*, dengan jawaban yang kaku dan terbatas pada skenario tertentu. Hal ini menyebabkan respons yang diberikan tidak selalu relevan dengan maksud pertanyaan pengguna. Dengan demikian, terdapat kebutuhan akan sistem berbasis kecerdasan buatan yang mampu menjawab pertanyaan tentang zakat secara fleksibel dan kontekstual, khususnya dalam bahasa Indonesia.

QAS dirancang untuk menghasilkan jawaban langsung dan singkat berdasarkan pertanyaan yang diajukan oleh pengguna dalam bahasa alami [5]. Sistem ini menggunakan kecerdasan buatan (AI) untuk membuat mesin berpikir dan berperilaku cerdas, di mana perangkat lunak cerdas yang ada di dalamnya memungkinkan mesin untuk memahami konteks dan memberikan jawaban yang relevan sehingga dapat menghemat waktu dan usaha pengguna dalam mencari informasi, jika dibandingkan dengan metode pencarian tradisional yang memerlukan eksplorasi dokumen secara menyeluruh. QAS juga sering dilengkapi dengan kalimat pendukung sebagai referensi jawaban yang diberikan [6]. Sistem ini merupakan bagian dari pengembangan *Natural Language Processing* (NLP), yaitu teknologi yang memungkinkan

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 929-938
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.728

interaksi antara manusia dan komputer menggunakan bahasa manusia [7]. Berbeda dengan mesin pencari biasa, QAS memberikan jawaban langsung tanpa perlu membuka banyak sumber informasi. Pengguna dapat berinteraksi dengan sistem seperti dalam sebuah percakapan, menjadikannya lebih praktis dan mudah digunakan [8].

Dalam konteks ini, QAS berbasis AI dan NLP sangat relevan untuk mempermudah pencarian informasi zakat secara efisien dan efektif [9]. Oleh sebab itu, penelitian ini memilih algoritma *Long Short-Term Memory* (LSTM) sebagai metode *deep learning* yang tepat untuk menangani pemrosesan data berurutan [10]. Secara khusus, LSTM memiliki kemampuan dalam menangkap hubungan jangka panjang antar kata dalam kalimat, yang menjadikannya efektif dalam memahami konteks pertanyaan dan jawaban. Berbeda dengan metode NLP tradisional seperti TF-IDF atau *rule-based* yang tidak mempertimbangkan urutan kata, LSTM mempertahankan informasi dari urutan sebelumnya. Sementara arsitektur modern seperti *Transformer* dan BERT memang menawarkan akurasi lebih tinggi pada berbagai tugas NLP, model-model tersebut memerlukan sumber daya komputasi yang jauh lebih besar. Dalam konteks penelitian ini yang berfokus pada sistem tanya jawab berbasis teks dengan implementasi ringan dan efisien (misalnya pada platform Telegram), LSTM menjadi pilihan yang lebih tepat dan praktis karena mampu memberikan hasil yang akurat dengan kompleksitas yang lebih rendah

Berbagai penelitian telah dikembangkan dalam bidang QAS. Sistem QAS berbasis hadis yang memanfaatkan *merger retriever*, LangChain, dan model Gemini-pro menunjukkan kepuasan pengguna yang tinggi, dengan tingkat penilaian "Sangat Setuju" sebesar 88,5% dan skor evaluator LangChain sebesar 91% [11]. Penelitian lain mengembangkan QAS berbasis web menggunakan LangChain dan LLM, dengan data dari buku fikih empat madzhab, menghasilkan skor F1 hingga 81% BERTScore dan 56% ROUGE [12]. Pendekatan *passage retrieval* berbasis FAISS untuk bahasa Indonesia menggunakan model IndoBERT-QA berhasil meningkatkan akurasi dari 65% menjadi 72,5% setelah pengoptimalan panjang karakter *passag*e [13]. Sementara itu, model BERT untuk QAS menghasilkan F1 *Score* 0,6 dan *Exact Match* 0,5, mencerminkan kemampuan dalam memberikan jawaban yang relevan namun masih terbatas pada skenario data kecil [14].

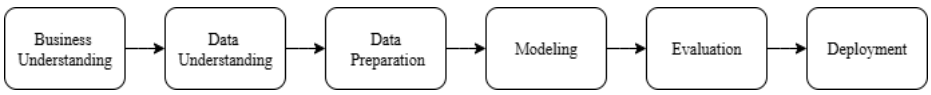
LSTM banyak digunakan dalam pemrosesan data sekuensial. Prediksi harga saham menggunakan LSTM menunjukkan hasil akurat dengan MAPE sebesar 0,71% dan RMSE 40,85 [15]. Untuk klasifikasi teks berita, model LSTM mencapai akurasi 88,75% dan F1-score 88,76%, membuktikan kemampuannya dalam memahami konteks teks [16]. Dalam analisis sentimen Twitter tentang AI, LSTM berhasil meraih akurasi 74,25% dan F1-*score* 80,69% [17]. Sementara pada QAS bahasa Sunda, model LSTM mencapai akurasi training sebesar 97% dari 12.853 data, namun keakuratan menjawab pertanyaan masih tergantung kesesuaian dengan data latih [5]. Hasil-hasil tersebut menunjukkan bahwa berbagai model QAS seperti BERT dan LLM telah berhasil diterapkan pada domain hadis dan fikih. Namun, penerapan model LSTM dalam sistem tanya jawab, khususnya pada topik zakat, masih belum banyak dilakukan. Penelitian ini berupaya untuk mengatasi keterbatasan yang ada dengan mengembangkan sistem QAS zakat berbasis LSTM menggunakan dataset dari buku resmi Baznas. Penelitian ini diharapkan dapat memberikan solusi yang efektif dan relevan dalam menjawab pertanyaan seputar zakat.

LSTM banyak diaplikasikan dalam penelitian QAS karena kemampuannya dalam memproses data sekuensial, baik sebagai input maupun output, serta merupakan bentuk pengembangan dari Recurrent Neural Network (RNN) [18]. LSTM dirancang untuk mengatasi keterbatasan model sebelumnya dalam mengelola urutan data panjang, sehingga lebih optimal dalam memahami konteks percakapan atau pertanyaan kompleks [19]. Dengan struktur internal berupa gate dan cell, LSTM mampu mengatur aliran informasi dengan lebih baik [20], di mana setiap unit terdiri dari forget gate, input gate, dan output gate [21]. Karena kemampuannya mengingat informasi dalam urutan panjang, LSTM efektif untuk diterapkan pada sistem tanya jawab berbasis bahasa alami, termasuk untuk menjawab pertanyaan seputar zakat yang seringkali membutuhkan pemahaman konteks menyeluruh.

Penelitian ini bertujuan untuk mengembangkan sebuah *question answering system* menggunakan algoritma *long short-term memory* yang akan menyediakan informasi terkait zakat dalam bahasa Indonesia. Sistem ini dirancang untuk menggantikan proses pencarian informasi dari Buku Standar Laboratorium Manajemen Zakat (Baznas) yang sebelumnya dilakukan secara manual, menjadi sistem otomatis berbasis question answering. Dengan pendekatan ini, diharapkan pengguna dapat memperoleh informasi dengan cepat, tepat, dan relevan tanpa harus membaca seluruh dokumen.

2. METODOLOGI PENELITIAN

Penelitian ini mengimplementasikan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang merupakan pendekatan standar yang digunakan untuk mengembangkan solusi terhadap masalah yang berkaitan dengan data *mining*. [22]. Pemilihan metode CRISP-DM dalam penelitian ini didasarkan pada kemampuan metodologi tersebut untuk menyediakan kerangka yang sistematis dan terstruktur dalam setiap tahapan pengolahan data, mulai dari pemahaman masalah hingga implementasi akhir. Pendekatan ini memungkinkan proses pengembangan QAS berbasis LSTM dapat berjalan secara terarah dan terukur, serta memastikan hasil akhir dapat diintegrasikan dengan baik ke dalam platform chat seperti Telegram. Diagram alur penelitian ditampilkan pada Gambar 1.



Gambar 1. Alur Metode Penelitian

Menurut Gambar 1, tahap awal *business understanding*, dilakukan untuk memahami kebutuhan akan sistem QAS Zakat. Pada tahap *data understanding*, dianalisis isi buku panduan zakat Baznas untuk menyiapkan data yang relevan. Selanjutnya, *data preparation* mencakup augmentasi data dan *preprocessing teks* hingga membentuk pasangan pertanyaan dan jawaban. Tahap modeling melibatkan pelatihan model LSTM. Setelah itu, dilakukan evaluasi kinerja model menggunakan confusion matrix seperti akurasi, presisi, recall dan F1-score pada data validasi. Terakhir, pada tahap deployment, model diintegrasikan ke dalam platform Telegram untuk memudahkan akses pengguna.

2.1 Business Understanding

Tahapan ini bertujuan untuk memahami kebutuhan dan tujuan dari sistem yang akan dikembangkan. QAS Zakat dirancang untuk membantu masyarakat mendapatkan informasi seputar zakat secara cepat dan akurat. Tujuan utamanya adalah menciptakan sistem berbasis kecerdasan buatan yang dapat memberikan jawaban otomatis atas pertanyaan terkait zakat, yang bersumber dari referensi resmi BAZNAS, serta dapat diakses secara mudah melalui platform digital seperti Telegram.

2.2 Data Understanding

Pada tahap ini, dilakukan identifikasi terhadap struktur dan isi data yang akan digunakan dalam penelitian ini. Dataset yang digunakan berbentuk file CSV (Comma-Separated Values), yang terdiri dari dua komponen utama, yaitu *question* (pertanyaan) dan *answer* (jawaban). Dataset ini diperoleh dari buku *Standar Laboratorium Manajemen Zakat* terbitan BAZNAS, yang memuat penjelasan mengenai hukum zakat, jenis zakat, penerima zakat, serta ketentuan lainnya.

Dataset yang digunakan berjumlah 1159 pertanyaan dan jawaban, yang memungkinkan analisis terhadap berbagai jenis pertanyaan umum yang sering muncul dalam konteks zakat. Kualitas data ini sangat penting untuk penelitian ini. Oleh karena itu, dilakukan verifikasi kebenaran setiap jawaban yang ada dalam dataset dengan merujuk pada sumber-sumber yang dapat dipertanggungjawabkan, seperti literatur hukum dan panduan resmi dari BAZNAS. Proses verifikasi ini dilakukan manual terhadap referensi yang ada.

Selain itu, perlu dicatat bahwa dataset ini mungkin mengandung *class imbalance*, di mana beberapa kategori pertanyaan atau jawaban lebih banyak muncul dibandingkan yang lain. Hal ini akan dianalisis lebih lanjut untuk memastikan bahwa model yang dikembangkan dapat menangani ketidakseimbangan kelas dan memberikan prediksi yang akurat pada semua kategori pertanyaan.

2.3 Data Preparation

Pada tahap *data preparation*, langkah pertama yang dilakukan adalah analisis data untuk memahami karakteristik dan kualitas dataset yang akan digunakan dalam pelatihan model LSTM. Proses analisis ini bertujuan untuk memeriksa distribusi, panjang, dan variasi data agar memastikan dataset siap digunakan untuk pelatihan. Setelah analisis, tahap selanjutnya adalah augmentasi data dengan menggunakan teknik *synonym replacement*, di mana variasi pertanyaan dibuat dengan mengganti kata-kata dengan sinonim dan mengubah struktur kalimat, guna meningkatkan keragaman data. Kemudian, dilakukan tokenisasi untuk mengonversi kalimat menjadi urutan token yang bisa diproses oleh model. Proses berikutnya adalah *padding*, yang bertujuan untuk menyamakan panjang input agar dapat diproses secara *batch* oleh model. Selanjutnya, dilakukan *encoding* untuk mengubah token menjadi representasi numerik berdasarkan indeks dalam kamus kata. Terakhir, data dibagi menjadi *train-validation split* untuk memisahkan data pelatihan dan pengujian model dengan rasio tertentu.

2.4 Modeling

Tahapan modeling bertujuan untuk membangun dan melatih model LSTM agar mampu memahami pola dan konteks dari pertanyaan berbasis teks serta menghasilkan jawaban yang relevan. Model LSTM dipilih karena kemampuannya dalam menangani data sekuensial dan mempertahankan informasi dari urutan kata yang panjang, yang sangat cocok untuk tugas NLP. Sebelum pelatihan dimulai, data teks yang telah melalui tahap augmentasi dan tokenisasi dikonversi menjadi bentuk numerik menggunakan representasi token, sehingga dapat diproses oleh jaringan saraf. *Tokenization layer* digunakan untuk memetakan setiap kata menjadi angka atau token yang dapat dikenali oleh model.

Model LSTM dirancang dengan beberapa lapisan, termasuk input layer, satu atau lebih *hidden* LSTM layers, dan *output layer* yang menghasilkan prediksi berupa label yang relevan. Parameter model, seperti jumlah neuron, jumlah epoch, learning rate, dan ukuran *batch*, ditentukan melalui eksperimen untuk mendapatkan hasil yang optimal. Proses pelatihan dilakukan menggunakan data latih, sedangkan data validasi digunakan untuk mengevaluasi performa model dalam menjawab pertanyaan yang belum pernah dilihat sebelumnya.

2.5 Evaluation

Pada tahapan *Evaluation*, performa model dievaluasi menggunakan data validasi. Beberapa metrik yang digunakan dalam evaluasi antara lain akurasi, precision, recall, F1-score, dan confusion matrix. Metrik ini digunakan untuk mengukur sejauh mana model mampu mengklasifikasikan pertanyaan zakat dengan tepat dan memberikan prediksi yang akurat sesuai kategori yang benar. Evaluasi ini bertujuan untuk memastikan bahwa model tidak hanya bekerja secara teoritis, tetapi juga memiliki kinerja yang andal sebelum diintegrasikan ke dalam sistem tanya-jawab zakat berbasis Telegram dan digunakan oleh pengguna secara langsung. Metrik ini sangat berguna, terutama ketika menghadapi masalah

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 929-938
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.728

ketidakseimbangan kelas (class imbalance), di mana beberapa kelas lebih sering muncul dibandingkan dengan kelas lainnya. Dengan menggunakan metrik ini dapat mengevaluasi kinerja model secara lebih holistik dan memastikan bahwa model tidak hanya akurat, tetapi juga mampu mengklasifikasikan setiap kelas dengan baik.

2.6 Deployment

Tahapan *deployment* dilakukan untuk mengimplementasikan model yang telah dilatih ke dalam lingkungan yang dapat diakses langsung oleh pengguna. Dalam penelitian ini, QAS Zakat diintegrasikan ke dalam platform Telegram dengan memanfaatkan Telegram Bot API. Model LSTM yang telah dilatih berfungsi sebagai *backend chatbot*, yang memproses input pertanyaan dari pengguna dan memberikan respons berdasarkan klasifikasi yang dihasilkan oleh model.

Proses *deployment* dimulai dengan menyimpan model dalam format HDF5 (.h5) yang dapat dimuat ulang, kemudian menghubungkannya dengan skrip bot berbasis Python. Ketika pengguna mengirimkan pertanyaan melalui Telegram, sistem akan melakukan *preprocessing* terhadap input teks, lalu mengirimkannya ke model untuk diklasifikasikan. Berdasarkan prediksi yang dihasilkan, sistem akan menampilkan jawaban yang relevan dari dataset yang telah disiapkan sebelumnya.

Melalui integrasi ini, pengguna dapat berinteraksi secara langsung dengan sistem melalui antarmuka percakapan, tanpa perlu membuka dokumen atau melakukan pencarian manual. *Deployment* ini bertujuan untuk mempermudah akses informasi zakat dengan cara yang cepat, akurat, dan efisien.

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Permasalahan utama dalam penelitian ini adalah terbatasnya akses informasi zakat yang mudah, cepat, dan akurat bagi masyarakat umum serta minimnya sistem tanya jawab zakat berbahasa Indonesia yang interaktif. Banyak individu yang kebingungan dalam memahami ketentuan zakat, seperti jenis-jenis zakat, syarat wajib, perhitungan, serta siapa saja yang berhak menerima zakat. Hal ini menyebabkan kurangnya literasi zakat di masyarakat, yang pada akhirnya dapat berdampak pada tidak optimalnya penghimpunan dan distribusi zakat di Indonesia.

Penelitian ini bertujuan untuk menawarkan solusi digital berbasis QAS dengan menggunakan LSTM yang dapat memberikan jawaban cepat dan relevan terhadap pertanyaan-pertanyaan seputar zakat. Dengan mengintegrasikan sistem ini ke dalam *platform* Telegram, diharapkan layanan informasi zakat dapat diakses secara praktis dan efisien, sesuai dengan kebiasaan komunikasi digital masyarakat saat ini.

3.2 Data Understanding

Dataset yang digunakan dalam penelitian ini disusun dalam format CSV, dengan dua komponen utama yaitu question dan *answer*. Kolom *question* berisi berbagai pertanyaan dalam bahasa alami yang sering diajukan terkait zakat, sementara kolom *answer* berisi jawaban yang sesuai untuk setiap pertanyaan tersebut. Setiap pasangan *question* dan *answer* mencerminkan interaksi dalam sistem tanya jawab zakat yang dirancang.

Sumber utama data diambil dari Buku Standar Laboratorium Manajemen Zakat (BAZNAS), yang mencakup berbagai topik zakat, seperti zakat fitrah, zakat mal, perhitungan nisab, mustahik, dan tata cara pembayaran. Dataset ini dikembangkan dengan menyusun dan mengembangkan berbagai variasi pertanyaan yang mungkin diajukan oleh pengguna seputar zakat.

Dalam data tersebut, terdapat berbagai tipe karakter. Karakter huruf digunakan dalam kata dan kalimat pertanyaan atau jawaban, sedangkan karakter angka digunakan dalam data numerik seperti nisab dan jumlah zakat. Karakter spasi memisahkan kata, dan karakter simbol (misalnya tanda tanya dan titik) memberikan struktur serta makna tambahan dalam kalimat.

Dataset ini terdiri dari total 1159 pasangan pertanyaan dan jawaban, yang mencakup berbagai variasi pertanyaan terkait topik zakat. Pembagian data dalam dataset ini meliputi beberapa kategori jawaban yang merepresentasikan berbagai jenis pertanyaan, seperti informasi mengenai zakat fitrah, zakat mal, perhitungan nisab, dan lainnya. Selain itu, proses augmentasi data dilakukan dengan menambah variasi pertanyaan melalui penggunaan sinonim dan variasi kalimat, yang memungkinkan jumlah data bertambah dan meningkatkan keragaman dalam dataset, sehingga model dapat belajar dari berbagai bentuk pertanyaan yang mungkin diajukan oleh pengguna.

Pemahaman terhadap struktur dan sebaran data ini sangat penting sebelum melanjutkan ke tahap preprosesing dan pemodelan, karena memastikan distribusi data yang merata dan representatif terhadap kebutuhan sistem tanya jawab berbasis zakat.

3.3 Data Preparation

Pada tahap data p*reparation*, dataset yang telah diperoleh melalui proses data *understanding* kemudian diproses dengan melakukan *t*ext pre-processing agar dapat digunakan dalam pelatihan model. Berikut adalah langkah-langkah utama dalam *text pre-processing*:

a. Analisis Dataset
   Pada tahap pertama dalam persiapan data, dilakukan analisis mendalam terhadap dataset yang akan digunakan untuk melatih sistem QAS Zakat dengan menggunakan metode LSTM. Proses analisis dataset bertujuan untuk memahami

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 929-938
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.728

karakteristik dan kualitas data yang ada, yang kemudian akan membantu dalam pengambilan keputusan untuk tahap persiapan berikutnya. Berikut adalah hasil dari analisis dataset:

1. Jumlah dan Keunikan Data
   Dataset ini terdiri dari 1.159 pertanyaan, dengan 1.146 pertanyaan unik dan 230 jawaban unik. Hal ini menunjukkan bahwa dataset memiliki variasi yang cukup besar dalam jenis pertanyaan dan jawaban terkait zakat. Panjang rata-rata pertanyaan adalah 50,2 karakter, sementara panjang rata-rata jawaban adalah 294,8 karakter. Dengan variasi panjang yang cukup besar, model perlu dipersiapkan untuk menangani perbedaan panjang teks ini.
2. Distribusi Kelas Jawaban
   Analisis distribusi jawaban dilakukan dengan menghitung jumlah sampel untuk setiap kelas jawaban. Dataset ini memiliki 230 kelas jawaban, dengan rata-rata sampel per kelas sekitar 5 sampel. Beberapa kelas memiliki jumlah sampel yang lebih sedikit, yaitu minimal 4 sampel per kelas, sementara kelas lainnya memiliki jumlah sampel yang lebih banyak, dengan maksimal 10 sampel. Analisis ini penting untuk memahami apakah terdapat kelas yang dominan atau jika distribusi data tidak merata.
3. Balancing Data
   Mengingat adanya beberapa kelas dengan jumlah sampel yang sangat sedikit, dilakukan penyaringan untuk memastikan bahwa setiap kelas memiliki jumlah sampel yang cukup. Kelas dengan kurang dari 3 sampel dihapus dari dataset untuk memastikan distribusi kelas yang lebih seimbang. Setelah proses balancing, dataset tetap memiliki 230 kelas jawaban, dengan distribusi yang lebih merata di antara kelas-kelas tersebut.

b. Augmentasi Dataset

Augmentasi data dilakukan untuk meningkatkan keragaman dan jumlah dataset dengan menciptakan variasi pertanyaan yang relevan. Proses ini melibatkan penggantian kata-kata dalam pertanyaan dengan sinonim yang sesuai, seperti mengganti kata "apa" dengan "apakah" atau "jelaskan", dan penggantian istilah zakat dengan variasi seperti "kewajiban zakat" atau "ibadah zakat". Selain itu, struktur kalimat juga dimodifikasi untuk menghasilkan variasi dalam cara pertanyaan diajukan, seperti mengubah "Apa itu zakat fitrah?" menjadi
"Tolong bantu, zakat fitrah itu bagaimana?". Augmentasi ini dilakukan dengan rasio augment_ratio = 0.8, yang berarti sekitar 80% variasi tambahan untuk setiap pertanyaan dalam dataset asli. Hasil dari augmentasi ini menggandakan jumlah data dari 1.159 pertanyaan menjadi 3.479 pertanyaan, sehingga memberikan model berbagai variasi yang memperkaya pemahaman terhadap pertanyaan tentang zakat. Dataset yang telah diperbesar dan divariasikan ini akan digunakan untuk melatih model LSTM, sehingga model lebih mampu menangani beragam variasi pertanyaan yang mungkin diajukan oleh pengguna.

Tabel 1. Contoh Variasi Pertanyaan setelah Augmentasi

Pertanyaan Asli	Variasi 1	Variasi 2
Apa itu zakat fitrah?	Bisakah dijelaskan zakat fitrah?	Saya ingin tahu zakat fitrah?
Bagaimana cara bayar zakat mal?	Bagaimana cara menunaikan zakat mal?	Apa prosedur bayar zakat mal?
Siapa yang berhak menerima zakat?	Penerima zakat itu siapa?	Yang berhak menerima zakat siapa saja?

Tabel 1, menunjukkan contoh pertanyaan asli beserta variasinya setelah augmentasi. Proses augmentasi ini berhasil menghasilkan variasi kalimat dengan penggunaan sinonim dan perubahan struktur kalimat, yang diharapkan dapat meningkatkan kemampuan model dalam memahami berbagai variasi pertanyaan terkait zakat.

c. Tokenization

Pada tahap Tokenization, teks pertanyaan dalam dataset diubah menjadi urutan angka yang dapat diproses oleh model (LSTM). Tokenizer digunakan untuk memecah teks menjadi kata-kata dan menggantikan setiap kata dengan angka yang sesuai dalam kamus. Kamus ini dibangun berdasarkan frekuensi kata yang terdapat dalam dataset. Dalam hal ini, Tokenizer diinisialisasi dengan parameter num_words=5000, yang membatasi kamus hanya pada 5.000 kata yang paling sering muncul dalam dataset, dan parameter oov_token='<OOV>' digunakan untuk menangani kata-kata yang tidak ditemukan dalam kamus (Out-of-Vocabulary).
Proses dimulai dengan memanggil *fit_on_texts*() untuk membangun kamus dari kolom *question* dalam dataset. Setelah itu, *texts_to_sequences*() digunakan untuk mengonversi setiap kalimat pertanyaan menjadi urutan angka, yang merepresentasikan setiap kata dalam pertanyaan berdasarkan posisi kata tersebut dalam kamus.

d. Padding

Setelah proses tokenisasi, langkah selanjutnya adalah *Padding*, yang bertujuan untuk memastikan bahwa semua urutan angka memiliki panjang yang sama. Karena panjang pertanyaan dalam dataset bervariasi, pad_sequences() digunakan untuk menambahkan padding pada urutan yang lebih pendek dari panjang urutan terpanjang, atau memangkas urutan yang lebih panjang dari batas yang telah ditentukan.
Dalam implementasi ini, panjang urutan terpanjang dihitung menggunakan maxlen = max(len(s) for s in seqs). Kemudian, *pad_sequences*() memastikan bahwa semua urutan memiliki panjang yang konsisten dengan parameter maxlen. *Padding* dilakukan pada akhir urutan (dengan *padding='post'*), yang berarti bahwa angka nol (0) ditambahkan di bagian belakang urutan yang lebih pendek. Untuk urutan yang lebih panjang dari panjang yang telah ditentukan,

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 929-938
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.728

proses *truncating* akan memangkasnya di bagian belakang (dengan *truncating='post'*). Langkah ini memastikan bahwa input data memiliki panjang yang seragam, sehingga model dapat memprosesnya secara konsisten.

e. Encoding

Pada tahap *encoding*, label jawaban yang ada dalam dataset diubah menjadi format numerik menggunakan LabelEncoder. LabelEncoder mengonversi kategori jawaban yang bersifat teks menjadi label numerik, yang memungkinkan model untuk mengolah variabel kategori dalam bentuk numerik. Setiap kategori jawaban yang unik diberikan label integer mulai dari 0 hingga jumlah kelas yang ada dalam dataset.

Metode fit_transform() diterapkan pada kolom *answer* untuk mengubah setiap jawaban menjadi label numerik yang sesuai. Proses ini menghasilkan *array label* yang digunakan sebagai target dalam pelatihan model. Setelah label jawaban dienkode, variabel num_classes dihitung untuk mengetahui jumlah kelas yang berbeda dalam dataset. Selain itu, ukuran kamus (*vocabulary size*) dihitung menggunakan len(tokenizer_q.word_index) + 1, yang memberikan informasi tentang berapa banyak kata unik yang dikenali oleh Tokenizer selama proses tokenisasi.

f. Data Split (*Train-Validation*)

Pada tahap Split Data, dataset dibagi menjadi dua bagian utama yaitu data pelatihan (train) dan data validasi (validation). Pembagian ini bertujuan untuk melatih model pada data pelatihan dan menguji kinerjanya pada data validasi. Untuk memastikan proporsi kelas dalam data pelatihan dan validasi tetap konsisten dengan distribusi kelas dalam dataset asli, digunakan teknik stratified split. Dengan stratified split, pembagian data dilakukan sedemikian rupa sehingga proporsi masing-masing kelas dalam data pelatihan dan validasi tetap sama. Teknik ini sangat penting ketika bekerja dengan dataset yang tidak seimbang, di mana beberapa kelas mungkin memiliki jumlah sampel yang lebih sedikit.

Namun, jika ada kelas dengan terlalu sedikit sampel, misalnya hanya ada satu sampel dalam kelas tersebut, maka ValueError dapat terjadi. Dalam kasus ini, fallback dilakukan ke random split, di mana data dibagi secara acak tanpa mempertimbangkan distribusi kelas. Pembagian data dilakukan dengan menggunakan train_test_split() dari pustaka scikit-learn, dengan parameter test_size=0.2, yang berarti 20% data digunakan untuk validasi, sementara 80% sisanya digunakan untuk pelatihan. Hasil dari pembagian ini adalah dua set data: X_train dan X_val untuk fitur, serta y_train dan y_val untuk label.

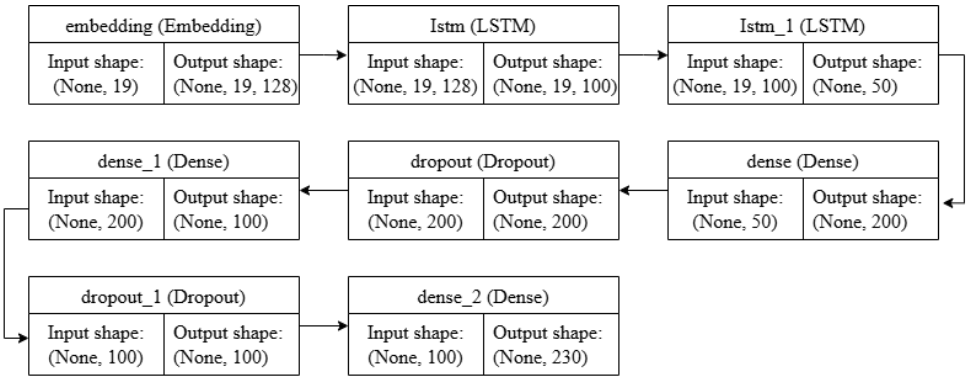
Tabel 2. Rasio Train-Validation Split (80:20)

Dataset	Jumlah Data	Train (80%)	Validation (20%)
Total Data	3.479	2.783	696

Berdasarkan Tabel 2, setelah melalui proses split data hasilnya menunjukkan bahwa data pelatihan terdiri dari 2.783 sampel, data validasi terdiri dari 696 sampel. Pembagian ini memungkinkan evaluasi yang lebih akurat terhadap performa model dan membantu mendeteksi potensi masalah seperti overfitting.

3.4 Modeling

Pada tahap Modeling, model Long Short-Term Memory (LSTM) dibangun untuk menangani tugas *question answering system* zakat. Model ini menggunakan arsitektur yang dirancang untuk mengoptimalkan akurasi, dengan lapisan-lapisan yang memungkinkan model untuk memahami konteks pertanyaan dan menghasilkan jawaban yang tepat. Berikut adalah arsitektur model yang dihasilkan pada Gambar 2.



Gambar 2. Layer LSTM

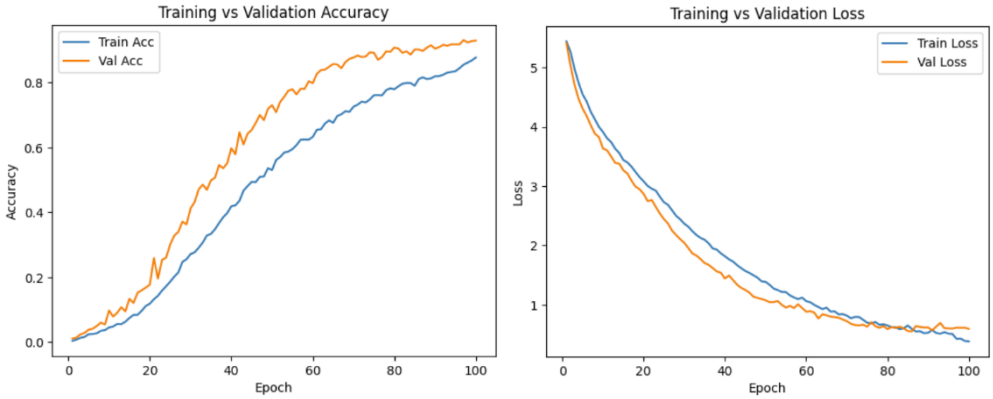
Berdasarkan Gambar 2, model dimulai dengan lapisan Embedding, yang mengonversi urutan token menjadi vektor berdimensi 128. Vektor-vektor ini memberikan representasi yang lebih baik dari kata-kata dalam kalimat. Lapisan LSTM pertama dengan 100 unit digunakan untuk memproses urutan kata dalam kalimat dan menangkap hubungan jangka panjang antar kata. Lapisan kedua LSTM dengan 50 unit diikuti oleh lapisan Dense untuk memperluas kapasitas model dalam belajar lebih banyak fitur. Setiap lapisan LSTM dilengkapi dengan Dropout untuk mengurangi overfitting dan meningkatkan generalisasi model.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 929-938
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.728

Setelah lapisan LSTM, model melanjutkan dengan dua lapisan Dense, yang pertama dengan 200 unit dan yang kedua dengan 100 unit, untuk meningkatkan kemampuan model dalam belajar dari data lebih dalam. Lapisan Dropout tambahan diterapkan setelah lapisan Dense untuk memastikan regularisasi yang lebih baik.

Lapisan output menggunakan Dense dengan softmax activation untuk menghasilkan output probabilitas bagi masing-masing kelas jawaban. Ini memungkinkan model untuk memilih jawaban yang paling tepat berdasarkan input pertanyaan yang diberikan.

Setelah model dibentuk dengan arsitektur Long Short-Term Memory (LSTM), tahap pelatihan dilakukan dengan membagi data menjadi dua bagian, yaitu 80% untuk pelatihan dan 20% untuk validasi. Proses pelatihan menggunakan optimizer Adam dengan nilai *learning rate* sebesar 0.003, batch size sebanyak 24, dan dijalankan selama 100 epoch. Untuk menghindari overfitting dan menjaga kestabilan proses pelatihan, digunakan mekanisme EarlyStopping dan ReduceLROnPlateau. Berdasarkan hasil pelatihan, model berhasil mencapai akurasi pelatihan tertinggi sebesar 87,72% dan akurasi validasi tertinggi sebesar 93,08%. Nilai akurasi validasi yang lebih tinggi menunjukkan bahwa model tidak hanya mampu mempelajari pola dari data pelatihan, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data baru yang belum pernah dilihat sebelumnya.

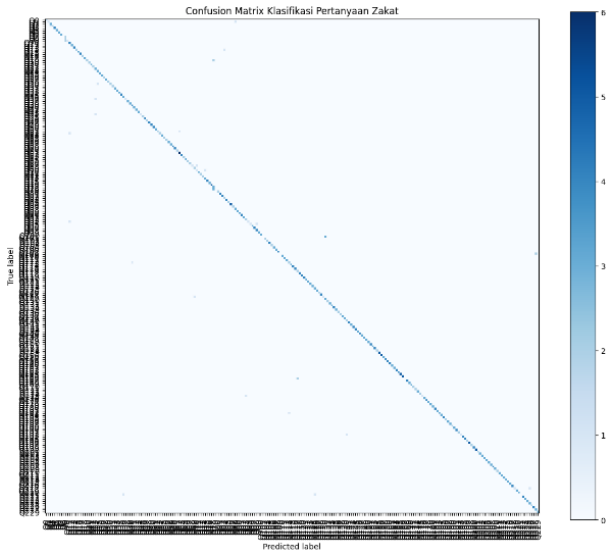


Gambar 3. Grafik Akurasi dan Loss Training - Validation

Pada Gambar 3, ditampilkan kurva akurasi selama proses pelatihan dan validasi. Kurva tersebut menunjukkan tren peningkatan yang stabil, di mana akurasi pelatihan dan validasi saling mengikuti secara konsisten. Tidak ditemukan indikasi overfitting, yang ditandai dengan tidak adanya jarak signifikan antara kedua kurva tersebut. Hal ini menunjukkan bahwa strategi pelatihan yang diterapkan cukup efektif dalam mengoptimalkan kinerja model.

3.5 Evaluation

Setelah proses pelatihan model selesai, langkah berikutnya adalah tahap evaluasi. Pada tahap ini, evaluasi dilakukan menggunakan confusion matrix untuk menghitung nilai akurasi, precision, recall, dan F1-score. Tahap evaluasi ini bertujuan untuk mengukur performa model ketika diuji menggunakan data validasi yang belum pernah dipelajari sebelumnya. Dalam proses ini, confusion matrix menjadi alat evaluasi utama dalam menilai sejauh mana prediksi model sesuai dengan label sebenarnya.



Gambar 4. Confusion Matrix klasifikasi Zakat

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 929-938
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.728

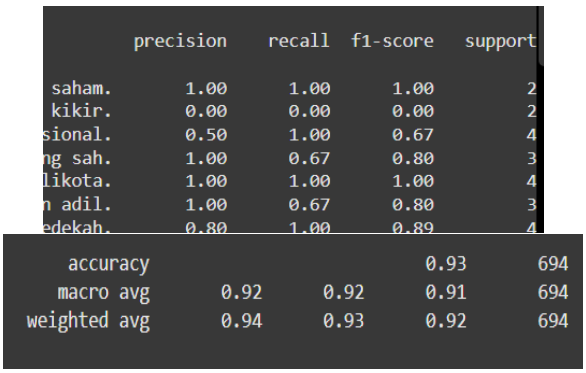
Gambar 4, menunjukkan bahwa sebagian besar prediksi model berada pada diagonal utama, menandakan bahwa model mampu mengklasifikasikan data dengan cukup akurat. Namun, karena jumlah kelas sangat besar, label-label pada sumbu X dan Y tidak terbaca secara jelas. Oleh karena itu, visualisasi digunakan untuk menunjukkan pola distribusi klasifikasi secara umum, bukan untuk analisis per kelas secara mendetail. Hasil dari keempat metrik tersebut dapat dilihat pada Gambar 5.

Pada Gambar 6, classification report menunjukkan performa model dalam bentuk ringkasan metrik evaluasi. Model mencapai akurasi keseluruhan sebesar 93%, dengan nilai precision, recall, dan F1-score rata-rata tertimbang masing-masing sebesar 0.94, 0.93, dan 0.92. Metrik ini dihitung secara otomatis menggunakan fungsi classification_report() dari *scikit-learn*, dan mempertimbangkan seluruh distribusi kelas yang tidak seimbang. Hasil ini menunjukkan bahwa model mampu melakukan klasifikasi dengan baik dan konsisten meskipun variasi antar kelas cukup besar.

Pengujian model juga dilakukan terhadap sejumlah pertanyaan acak. Dari hasil pengujian, model mampu memberikan respons yang sesuai pada sebagian besar pertanyaan. Beberapa contoh ditampilkan pada Tabel 3.

Tabel 3. Hasil Pengujian Implementasi Model

Pertanyaan	Jawaban	Keterangan
apa yang dimaksud dengan zakat?	Secara bahasa, zakat berasal dari kata zaka yang berarti bersih, tumbuh, berkembang, dan berkah. Secara istilah...	Berhasil
Zakat ada berapa macam?	Zakat terbagi menjadi dua jenis yaitu Zakat Fitrah dan Zakat Mal	Berhasil
Apa itu zakat fitrah?	Zakat fitrah, juga disebut zakat nafs (jiwa), adalah zakat jiwa yang wajib ditunaikan oleh setiap muslim...	Berhasil
Apa itu zakat mal?	Zakat mal adalah zakat atas harta benda seperti emas, perak, uang, perdagangan, hasil pertanian, peternakan...	Berhasil
Sampai kapan batas bayar zakat fitrah?	Batas waktu pembayaran zakat fitrah adalah sebelum dilaksanakannya salat Hari Raya Idul Fitri. Idealnya, zakat fitrah...	Berhasil



	precision	recall	f1-score	support
saham.	1.00	1.00	1.00	2
kikir.	0.00	0.00	0.00	2
sional.	0.50	1.00	0.67	4
ng sah.	1.00	0.67	0.80	3
likota.	1.00	1.00	1.00	4
n adil.	1.00	0.67	0.80	3
edekah.	0.80	1.00	0.89	4
accuracy			0.93	694
macro avg	0.92	0.92	0.91	694
weighted avg	0.94	0.93	0.92	694

Gambar 5. Classification Report Zakat

Berdasarkan Tabel 3, percobaan pada 5 pertanyaan hasil menunjukkan bahwa model memiliki kemampuan menjawab pertanyaan-pertanyaan zakat secara relevan dan akurat. Secara keseluruhan, performa model cukup memuaskan.

3.6 Deployment

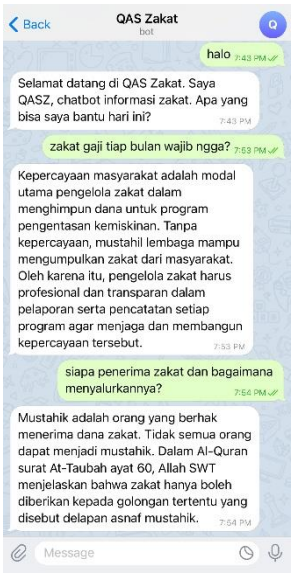
Tahapan *deployment* dilakukan untuk mengimplementasikan model yang telah dilatih ke dalam platform yang dapat diakses langsung oleh pengguna. Dalam penelitian ini, QAS Zakat diintegrasikan ke dalam platform Telegram menggunakan Telegram Bot API. Model LSTM yang telah dilatih berfungsi sebagai backend chatbot, yang memproses input pertanyaan dari pengguna dan memberikan respons yang relevan berdasarkan klasifikasi yang dihasilkan oleh model.

Proses *deployment* dimulai dengan menyimpan model yang telah dilatih dalam format HDF5 (.h5), yang kemudian dimuat kembali menggunakan skrip Python yang menghubungkan model dengan bot. Pengguna mengirimkan pertanyaan melalui Telegram, dan sistem melakukan *preprocessing* pada teks input, termasuk tokenisasi menggunakan tokenizer yang disimpan dalam tokenizer.pkl. Input yang telah diproses kemudian diklasifikasikan menggunakan model yang telah dilatih. Setelah prediksi klasifikasi dihasilkan, sistem mengembalikan jawaban yang relevan kepada pengguna berdasarkan kategori yang diprediksi oleh model, yang diterjemahkan melalui label encoder. Jawaban tersebut kemudian disampaikan langsung kepada pengguna melalui antarmuka percakapan di Telegram.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 5, August 2025 | Hal 929-938
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i5.728

Gambar 6. Demo Qas Zakat Pertama

Berdasarkan Gambar 6, Demo interaksi pengguna dengan Qas Zakat di Telegram. Pengguna mengajukan pertanyaan seputar zakat, dan bot memberikan jawaban relevan berdasarkan model QAS berbasis LSTM. Antarmuka percakapan ini memungkinkan pengguna untuk mengakses informasi zakat dengan mudah, menghindari kebutuhan untuk membuka dokumen atau pencarian manual, serta memberikan respons yang akurat secara real-time. Namun, dalam implementasinya, sistem masih memiliki beberapa keterbatasan yang perlu dicatat seperti yang terdapat pada Gambar 7.



Gambar 7. Demo Qas Zakat Kedua

Berdasarkan Gambar 7, Salah satu kendala utama adalah model LSTM menunjukkan performa yang menurun ketika menghadapi pertanyaan yang ambigu, terlalu umum, atau menggunakan gaya bahasa tidak baku, oleh karena itu pengembangan tetap diperlukan, khususnya melalui perluasan dan variasi dataset untuk mencakup konteks pertanyaan yang lebih kompleks dan ragam bahasa pengguna

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem *Question Answering System* (QAS) zakat berbasis Long Short-Term Memory (LSTM) dengan akurasi evaluasi sebesar 93%, precision 0.94, recall 0.93, dan F1-score 0.92 (weighted average). Sistem diintegrasikan ke dalam platform Telegram dan mampu memberikan jawaban otomatis terhadap pertanyaan seputar zakat secara real-time. Meskipun performa model cukup baik, keterbatasan masih ditemukan, terutama dalam menjawab pertanyaan dengan gaya bahasa tidak baku seperti “zakat gaji tiap bulan wajib nggak?” atau pertanyaan kompleks seperti “siapa penerima zakat dan bagaimana menyalurkannya?”. Model juga menunjukkan kelemahan pada kelas minoritas akibat distribusi dataset yang tidak seimbang. Oleh karena itu, pengembangan selanjutnya dapat

mencakup integrasi model bahasa besar (LLM) untuk pertanyaan kompleks, augmentasi dataset dengan pertanyaan kolokial berbahasa Indonesia, serta penambahan modul parsing untuk menangani pertanyaan multi-inten, agar sistem lebih adaptif terhadap keragaman ekspresi pengguna.

REFERENCES

[1] N. Latifah, Paujiah, and H. Pronixca, "Analisis Peran Zakat Dalam Pembangunan Ekonomi," *Islamologi: Jurnal Ilmiah Keagamaan*, vol. 1, no. 2, pp. 470–480, 2024.

[2] N. Khamimah and Baidhowi, "Optimalisasi Zakat Melalui Baznas Untuk Mendukung Jaminan Sosial Kesehatan Dengan Prinsip Syariah," *Causa: Jurnal Hukum dan Kewarganegaraan*, vol. 13, no. 11, pp. 21–30, 2025, doi: 10.6679/tz3k6p18.

[3] Humas BAZNAS RI, "Optimalkan Potensi Zakat, BAZNAS Dorong Pentingnya Dukungan UPZ di Lembaga Pemerintahan - BAZNAS." Accessed: Jun. 25, 2025. [Online]. Available: https://baznas.go.id/news-show/Optimalkan_Potensi_Zakat,_BAZNAS_Dorong_Pentingnya_Dukungan_UPZ_di_Lembaga_Pemerintahan/2063

[4] *OUTLOOK ZAKAT INDONESIA 2024*. Pusat Kajian Strategis BAZNAS, 2024. Accessed: Jun. 25, 2025. [Online]. Available: https://www.puskasbaznas.com/publications/books/1857-buku-outlook-zakat-indonesia-2024

[5] R. Kurniawan, T. I. Ramadhan, and R. Hartono, "Implementasi Sistem Question Answering Menggunakan Metode Long Short Term Memory (Lstm) Pada Studi Kasus Bahasa Sunda," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 7570–7578, 2024, doi: 10.36040/jati.v8i4.10237.

[6] A. Setiawan, O. N. Pratiwi, and R. Y. Fa'rifah, "Question Answering System Dalam Bentuk Chatbot Pada Platform Line Untuk Mata Pelajaran Sejarah SMA/MA Dengan Menggunakan Algoritma Levenshtein Distance," *eProceedings of Engineering*, vol. 8, no. 5, 2021.

[7] G. F. Avisyah, I. J. Putra, and S. S. Hidayat, "Open Artificial Intelligence Analysis using ChatGPT Integrated with Telegram Bot," *Jurnal ELTIKOM: Jurnal Teknik Elektro, Teknologi Informasi Dan Komputer*, vol. 7, no. 1, pp. 60–66, 2023, doi: 10.31961/eltikom.v7i1.724.

[8] A. Marsadualan, H. Harmastuti, and J. Triyono, "Rancang Bangun Aplikasi Tanya Jawab Mengenai Ist Akprind Yogyakarta Berbasis Mobile Menggunakan Algoritma Boyer Moore," in *Seminar Nasional Inovasi Sains Teknologi Informasi Komputer*, 2023, pp. 226–238.

[9] R. Arora, P. Singh, H. Goyal, S. Singhal, and S. Vijayvargiya, "Comparative Question Answering System based on Natural Language Processing and Machine Learning," in *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 373–378. doi: 10.1109/ICAIS50930.2021.9396015.

[10] P. B. Wintoro, H. Hermawan, M. A. Muda, and Y. Mulyani, "Implementasi Long Short-Term Memory pada Chatbot Informasi Akademik Teknik Informatika Unila," *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, vol. 12, no. 1, p. 68, 2022, doi: 10.36448/expert.v12i1.2593.

[11] R. Saputra, "Penerapan Merger Retriever pada Question Answering System Hadits," *SATIN–Sains dan Teknologi Informasi*, vol. 10, no. 1, pp. 24–35, 2024, doi: 10.33372/stn.v10i1.1117.

[12] S. Rahayu, N. S. Harahap, S. Agustian, and P. Pizaini, "Penerapan Teknologi LangChain pada Question Answering System Fikih Empat Madzhab: Application of Langchain Technology to the Fiqh Question Answering System of Four Madhhab," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 974–983, 2024, doi: 10.57152/malcom.v4i3.1397.

[13] T. I. Ramadhan, A. Supriatman, and T. R. Kurniawan, "Passage Retrieval untuk Question Answering Bahasa Indonesia Menggunakan BERT dan FAISS," *Jurnal Algoritma*, vol. 21, no. 2, pp. 156–163, 2024, doi: 10.33364/algoritma/v.21-2.2100.

[14] J. S. Wibowo, H. Februariyanti, and H. Listiyono, "Model Penjawab Pertanyaan Otomatis Berdasarkan Peringkat Relevansi Kalimat Menggunakan Model BERT," *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 5, no. 3, pp. 1100–1108, 2024, doi: 10.30645/kesatria.v5i3.427.g423.

[15] A. Rosyd, A. I. Purnamasari, and I. Ali, "Penerapan metode long short term memory (LSTM) dalam memprediksi harga saham PT Bank Central Asia," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 501–506, 2024, doi: 10.36040/jati.v8i1.8440.

[16] C. N. Daiman, A. Y. Rahman, and F. Nudiyansyah, "Klasifikasi Teks Berita Breaking News Di Manggarai Menggunakan Long Short Term Memory (LSTM)," *Jurnal Mnemonic*, vol. 7, no. 2, pp. 170–174, 2024, doi: 10.36040/mnemonic.v7i2.9939.

[17] A. Azrul, A. I. Purnamasari, and I. Ali, "Analisis sentimen pengguna Twitter terhadap perkembangan artificial intelligence dengan penerapan algoritma Long Short-Term Memory (LSTM)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 413–421, 2024, doi: 10.36040/jati.v8i1.8416.

[18] A. Silvanie and R. Subekti, "Aplikasi Chatbot Untuk Faq Akademik Di Ibi-K57 Dengan Lstm Dan Penyematan Kata," *JIKO (Jurnal Informatika dan Komputer)*, vol. 5, no. 1, pp. 19–27, 2022, doi: 10.33387/jiko.v5i1.3703.

[19] Y. A. Pradana, I. Cholissodin, and D. Kurnianingtyas, "Analisis sentimen pemindahan ibu kota Indonesia pada media sosial Twitter menggunakan metode LSTM dan Word2Vec," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 5, pp. 2389–2397, 2023.

[20] P. Choudhary and S. Chauhan, "An intelligent chatbot design and implementation model using long short-term memory with recurrent neural networks and attention mechanism," *Decision Analytics Journal*, vol. 9, p. 100359, 2023, doi: 0.1016/j.dajour.2023.100359.

[21] R. Luthfiansyah and B. Wasito, "Penerapan Teknik Deep Learning (Long Short Term Memory) dan Pendekatan Klasik (Regresi Linier) dalam Prediksi Pergerakan Saham BRI," *Jurnal Informatika dan Bisnis*, vol. 12, no. 2, pp. 42–54, 2023, doi: 10.46806/jib.v12i2.1059.

[22] Y. Singgalen, "Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, pp. 1551–1562, Jul. 2023, doi: 10.30865/mib.v7i3.6461.