

Implementasi Chatbot Telegram untuk Layanan FAQ Penerimaan Santri Baru di Pondok Pesantren Darunnajah Berbasis Natural Language Processing

Amanda Khairiyah Nasution^{1*}, Tutik Lestari², Supriadi Panggabean³

¹⁻³ Program Studi Sistem dan Teknologi Informasi, Universitas Darunnajah, Indonesia

*Penulis Korespondensi: amandakhairiyah@gmail.com

Abstract. Information services for New Student Admission (Penerimaan Santri Baru/PSB) at Pondok Pesantren Darunnajah Jakarta still rely on manual message replies, causing response delays outside working hours due to the high volume of repetitive questions. This study aims to design and build a Telegram chatbot based on Natural Language Processing (NLP) for the PSB Frequently Asked Questions (FAQ) service, to test its functional feasibility and intent-matching accuracy, and to analyze its role in service efficiency. The method used is Research and Development (R&D), adapting the Borg and Gall model into six stages. The product was built using Python, the Telegram Bot API, and the Sastrawi library. The intents.json knowledge base contains 21 intent categories, 356 question patterns, and 21 main responses. Intent determination uses root-word token overlap matching preceded by five text preprocessing stages: case folding, character cleaning, tokenization, stopword removal, and stemming. Evaluation was conducted through black box testing, intent category testing, and user assessment via a Likert-scale questionnaire completed by 23 respondents. Black box testing achieved 100% conformity (13 of 13 scenarios), intent category testing 100% (21 of 21 categories), and user assessment 100% (805 of 805) with an average of 5.00 in the excellent category. However, 41.18% of open-question respondents found specific topics unanswered due to knowledge-base limitations. The chatbot is concluded to be functionally feasible and well-accepted for improving PSB FAQ service efficiency.

Keywords: chatbot; Frequently Asked Questions; Natural Language Processing; new student admission; Telegram

Abstrak. Layanan informasi Penerimaan Santri Baru (PSB) di Pondok Pesantren Darunnajah Jakarta masih bertumpu pada balasan pesan manual, sehingga menimbulkan keterlambatan respons di luar jam kerja akibat tingginya pertanyaan berulang. Penelitian ini bertujuan merancang dan membangun chatbot Telegram berbasis Natural Language Processing (NLP) untuk layanan Frequently Asked Questions (FAQ) PSB, menguji kelayakan fungsional serta akurasi pencocokan maksud (intent), dan menganalisis perannya terhadap efisiensi layanan. Metode yang digunakan adalah Research and Development (R&D) dengan memodifikasi model Borg dan Gall menjadi enam tahap. Produk dibangun menggunakan bahasa pemrograman Python, Telegram Bot API, dan pustaka Sastrawi. Basis pengetahuan intents.json memuat 21 kategori intent, 356 pola pertanyaan, dan 21 respons utama. Penentuan intent menggunakan pencocokan irisan kata dasar (token overlap) setelah melalui lima tahap prapemrosesan teks, yaitu case folding, pembersihan karakter, tokenisasi, stopword removal, dan stemming. Evaluasi dilakukan melalui black box testing, pengujian kategori intent, serta penilaian pengguna via angket skala Likert oleh 23 responden. Hasil pengujian menunjukkan black box testing memperoleh kesesuaian 100% (13 dari 13 skenario), pengujian kategori intent 100% (21 dari 21 kategori), dan penilaian pengguna 100% (805 dari 805) dengan rata-rata 5,00 pada kategori sangat baik. Kendati demikian, 41,18% responden pertanyaan terbuka menemukan pertanyaan spesifik yang belum terjawab akibat keterbatasan basis pengetahuan. Disimpulkan bahwa chatbot Telegram ini layak secara fungsional maupun penerimaan pengguna untuk membantu efisiensi layanan FAQ PSB..

Kata kunci: chatbot; Frequently Asked Questions; Natural Language Processing; penerimaan santri baru; Telegram

1. LATAR BELAKANG

Perkembangan teknologi informasi pada satu dasawarsa terakhir mengubah cara lembaga pendidikan berkomunikasi dengan masyarakat. Layanan informasi yang dahulu bertumpu pada tatap muka dan surat-menyurat kini bergeser ke saluran digital yang menuntut kecepatan dan ketersediaan sepanjang waktu. Salah satu cabang teknologi yang berkembang pesat dan menopang pergeseran tersebut adalah kecerdasan buatan (artificial intelligence), khususnya penerapannya dalam bentuk chatbot, yaitu program yang berinteraksi dengan pengguna melalui teks maupun suara dan memberikan tanggapan berdasarkan kata kunci atau konteks pertanyaan. Selama lima tahun terakhir tren penggunaan chatbot meningkat pesat karena kemudahan penggunaan, keterjangkauan akses, dan luasnya bidang penerapannya, sehingga banyak organisasi mengadopsinya untuk mengotomatiskan layanan agar pertanyaan dapat dijawab cepat bahkan di luar jam kerja (Pratama & Setiawan, 2023; Furqan, 2023).

Pondok Pesantren Darunnajah Jakarta merupakan lembaga pendidikan Islam berasrama yang diminati masyarakat. Seiring bertambahnya jumlah pendaftar, kebutuhan akan layanan informasi Penerimaan Santri Baru (PSB) yang cepat dan responsif semakin meningkat, meliputi syarat pendaftaran, biaya, alur pendaftaran, jadwal tes, fasilitas, kegiatan santri, lokasi, rute, dan kontak panitia. Pengamatan terhadap layanan yang berjalan menunjukkan penyampaian informasi PSB masih bertumpu pada portal pendaftaran daring, brosur, dan pesan langsung kepada panitia melalui WhatsApp. Saluran pesan langsung menjadi yang paling banyak digunakan sekaligus paling terbatas: petugas tidak selalu dapat menanggapi pertanyaan dengan segera karena keterbatasan jam layanan dan sumber daya manusia. Persoalan ini diperberat oleh sifat pertanyaan yang berulang, sehingga waktu petugas habis untuk pekerjaan yang sebenarnya dapat diotomatiskan (Alpariji, 2024).

Penerapan teknologi di lingkungan pesantren menghadapi tantangan tersendiri. Sebagian kalangan memandang adopsi teknologi modern berpotensi mengurangi kualitas interaksi antarmanusia yang menjadi inti pendidikan berbasis nilai. Oleh karena itu, penerapan kecerdasan buatan di pesantren perlu dilakukan secara berhati-hati dan bertahap, dimulai dari teknologi sederhana yang tidak menuntut infrastruktur besar sehingga manfaatnya dapat dinilai sebelum investasi diperbesar. Dalam kerangka itu,

Natural Language Processing (NLP) menjadi pilihan yang relevan karena memungkinkan sistem mengolah masukan teks pengguna sehingga maksud pertanyaan dapat dikenali dan jawaban yang sesuai dikirimkan. Penerapan NLP pada chatbot terbukti mampu meningkatkan efisiensi penanganan pertanyaan berulang di institusi pendidikan (Dewi & Kurniawan, 2024; Ardabili, 2025).

Telegram dipilih sebagai kanal layanan karena dapat diakses melalui telepon genggam maupun komputer, tidak menuntut antarmuka yang rumit, serta menyediakan Bot API yang terbuka dan tidak berbayar. Dengan demikian pengembangan dapat dilakukan secara mandiri oleh lembaga tanpa biaya berlangganan dan tanpa memindahkan data pertanyaan calon santri kepada penyedia layanan pihak ketiga. Beberapa penelitian terdahulu telah mengembangkan chatbot serupa, tetapi umumnya bergantung pada mesin NLP pihak ketiga seperti Wit.AI (Ai, 2022) atau kerangka kerja berbasis pembelajaran mesin yang menuntut pelatihan model seperti Rasa (Irfani dkk., 2022) dan Dialogflow (Irawan & Parapat, 2024). Ketergantungan tersebut menimbulkan biaya, kebutuhan data berlabel, serta perpindahan data pengguna ke luar lembaga. Celah inilah yang menjadi urgensi penelitian: dibutuhkan chatbot berbiaya rendah, dapat dikelola mandiri, dan menjaga data pengguna tetap berada dalam kendali lembaga.

Berdasarkan uraian tersebut, penelitian ini bertujuan merancang dan membangun chatbot Telegram berbasis NLP untuk layanan FAQ PSB di Pondok Pesantren Darunnajah, menguji kelayakan fungsional dan akurasi pencocokan maksud melalui pengujian sistem, serta menganalisis persepsi pengguna dan peran produk terhadap efisiensi layanan informasi, khususnya pada pertanyaan yang bersifat berulang. Hasil yang diperoleh diharapkan tidak hanya bermanfaat bagi Pondok Pesantren Darunnajah, tetapi juga menjadi rujukan bagi lembaga pendidikan lain yang hendak mengadopsi teknologi serupa

2. KAJIAN TEORITIS

Chatbot atau conversational agent adalah program komputer yang dirancang untuk menirukan percakapan manusia melalui teks atau suara sehingga memungkinkan interaksi otomatis antara pengguna dan sistem (Caldarini dkk., 2022). Dalam layanan informasi, platform pesan instan seperti Telegram dan WhatsApp sering digunakan sebagai antarmuka utama karena aksesibilitasnya yang tinggi dan tersedianya Application

Programming Interface (API) yang mendukung otomatisasi balasan (Santoso & Pratami, 2023). Berdasarkan cara menghasilkan jawaban, chatbot dibedakan menjadi retrieval based, yaitu jawaban diambil dari kumpulan tanggapan yang telah disusun, dan generative based, yaitu jawaban dibangkitkan oleh model bahasa. Berdasarkan ruang lingkupnya, chatbot dibedakan menjadi closed domain yang hanya melayani topik tertentu dan open domain yang melayani topik apa pun. Produk pada penelitian ini termasuk retrieval based dan closed domain.

Natural Language Processing (NLP) merupakan pendekatan terkomputerisasi untuk menganalisis teks. Arsitektur chatbot berbasis NLP umumnya memuat komponen Natural Language Understanding (NLU) untuk mengenali maksud (intent recognition) dan entitas (entity extraction) dari pesan masuk sebelum diteruskan ke pengelola dialog guna menghasilkan respons yang tepat (Wibowo & Susanti, 2023). Pada layanan dengan ruang lingkup tertutup dan kategori terbatas, pendekatan berbasis aturan masih memadai sebagai tahap awal (Rizqi, 2025).

Prapemrosesan Teks Bahasa Indonesia

Prapemrosesan teks (text preprocessing) merupakan tahapan awal yang krusial karena masukan pengguna umumnya berupa teks tidak terstruktur yang mengandung kata tidak baku, variasi kapitalisasi, angka, dan tanda baca (Al Farisi dkk., 2024). Pada teks Bahasa Indonesia, prapemrosesan umumnya melibatkan lima tahapan berurutan (IJSA, 2025): case folding untuk menyeragamkan huruf menjadi huruf kecil; pembersihan teks untuk membuang karakter non-alfabet; tokenizing untuk memecah kalimat menjadi token kata; stopword removal untuk menghapus kata umum berfrekuensi tinggi namun minim informasi; dan stemming untuk mengembalikan kata berimbuhan ke bentuk dasarnya. Karena Bahasa Indonesia memiliki aturan aglutinasi yang kompleks, stemming umumnya mengacu pada algoritma Nazief & Adriani atau Tala. Pustaka Sastrawi pada Python merupakan implementasi terbuka yang memanfaatkan algoritma tersebut dengan bantuan kamus kata dasar (Nugroho & Wibowo, 2023; Suryani & Wijaya, 2024).

Intent, Pattern, dan Algoritma Pencocokan

Pada chatbot berbasis pengambilan jawaban, maksud pertanyaan direpresentasikan sebagai intent yang memiliki sekumpulan pattern (variasi kalimat bermaksud sama) dan

response (jawaban yang dikirimkan). Algoritma paling sederhana untuk menentukan intent adalah pencocokan leksikal berbasis irisan kata dasar (token overlap), yaitu menghitung banyaknya kata dasar yang sama antara pertanyaan pengguna dan setiap pola, kemudian memilih pola dengan kesamaan tertinggi. Algoritma ini tidak memerlukan pelatihan model, ringan dijalankan, dan keputusannya dapat ditelusuri, tetapi tidak menimbang kepentingan kata dan tidak mengenali sinonim. Alternatif yang lebih maju adalah pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) yang dipadukan cosine similarity untuk mengukur kemiripan makna antar teks (Ramadhan & Yulia, 2024). Pemilihan algoritma bergantung pada ketersediaan data berlabel, luas ruang lingkup, dan sumber daya lembaga.

Telegram Bot API dan FAQ

Telegram menyediakan Bot API, yaitu antarmuka pemrograman terbuka yang memungkinkan pengembang membuat akun bot dan menghubungkannya dengan program sendiri; setiap bot dikenali melalui token rahasia sebagai kunci akses. Terdapat dua cara bot menerima pesan: long polling, yaitu program secara berkala menanyakan pesan baru kepada peladen Telegram sehingga mudah diterapkan tanpa alamat publik tetapi lebih boros sumber daya; dan webhook, yaitu peladen Telegram mengirim pesan langsung ke alamat aplikasi melalui HTTPS sehingga lebih hemat dan cepat tetapi menuntut alamat publik bersertifikat (Umam dkk., 2025; Utomo & Arifin, 2024). Adapun Frequently Asked Questions (FAQ) adalah kumpulan pertanyaan yang sering diajukan disertai jawaban yang telah disiapkan; pemaduan chatbot dengan basis data FAQ menjadikan tanggapan bersifat interaktif dan seketika, sekaligus meningkatkan efisiensi layanan dan pengalaman pengguna (Bastian & Saputra, 2023).

Penelitian yang Relevan dan Posisi Penelitian

Sejumlah penelitian terdahulu berkaitan dengan penerapan chatbot pada layanan informasi. Ai (2022) mengembangkan chatbot Telegram layanan informasi akademik dengan mesin NLP pihak ketiga Wit.AI; Irfani dkk. (2022) membangun chatbot Telegram akademik menggunakan framework Rasa yang menuntut pelatihan model; dan Irawan & Parapat (2024) mengembangkan chatbot pelayanan masyarakat berbasis Dialogflow. Ketiganya memiliki persamaan dalam pemanfaatan chatbot sebagai media layanan informasi otomatis, tetapi berbeda dengan penelitian ini pada tiga hal. Pertama, objek

layanan: penelitian terdahulu berada pada ranah akademik dan administrasi publik, sedangkan penelitian ini pada ranah penerimaan santri baru di lembaga pendidikan Islam berasrama yang isi, ragam penanya, dan pola musimannya berbeda. Kedua, ketergantungan pada pihak ketiga: penelitian ini menyusun sendiri rangkaian prapemrosesan dan pencocokan maksud sehingga data pengguna tetap dalam kendali lembaga dan tanpa biaya berlangganan. Ketiga, cakupan pengujian: penelitian ini memadukan pengujian black box, pengujian seluruh kategori maksud, dan penilaian pengguna dalam satu rangkaian.

3. METODE PENELITIAN

Penelitian ini menggunakan metode Research and Development (R&D) karena tujuan utamanya menghasilkan produk teknologi berupa chatbot Telegram sekaligus menguji kelayakan dan fungsionalitasnya (Hidayat & Pratama, 2023; Rizqi, 2025). Langkah R&D standar Borg & Gall meliputi sepuluh tahapan; dengan pertimbangan keterbatasan waktu, biaya, dan sumber daya, tahapan tersebut disederhanakan menjadi enam tahap utama, yaitu analisis kebutuhan, perencanaan produk, pengembangan produk awal, uji coba terbatas, revisi produk, dan penetapan produk akhir (Al Farisi dkk., 2024; Mustafa dkk., 2024). Penyederhanaan ini lazim dilakukan pada pengembangan perangkat lunak berlingkup terbatas, dengan konsekuensi penelitian berfokus sampai penetapan produk akhir dan belum mencakup diseminasi massal.

Penelitian dilaksanakan pada Februari–Agustus 2026 di kantor Penerimaan Santri Baru Pondok Pesantren Darunnajah Jakarta. Prosedur pengembangan diawali dengan analisis kebutuhan untuk mengenali jenis informasi yang paling sering dibutuhkan, bersumber dari brosur resmi, riwayat pesan WhatsApp layanan PSB, komentar media sosial, dan wawancara panitia. Kebutuhan informasi kemudian dikelompokkan menjadi kategori maksud, disusun pola pertanyaan dan jawaban untuk setiap kategori, serta dirancang arsitektur sistem dan alur pemrosesan teks. Pada pengembangan produk awal, basis pengetahuan disimpan dalam berkas intents.json, program dibangun dengan Python dan dihubungkan ke akun bot melalui Telegram Bot API, disertai penerapan fungsi prapemrosesan dan pencarian jawaban. Produk diuji melalui skenario percakapan dan dicoba oleh sejumlah pengguna, lalu direvisi berdasarkan temuan untuk ditetapkan sebagai produk akhir. Alat dan bahan yang digunakan meliputi komputer jinjing, telepon

pintar, Python 3, pustaka Telegram Bot API dan Sastrawi, Visual Studio Code, Google Forms, berkas intents.json, serta brosur resmi PSB dan riwayat pesan layanan.

Uji coba dirancang untuk menilai kesesuaian fungsi dan kelayakan produk melalui tiga cara yang saling melengkapi. Black box testing menilai apakah fungsi utama berjalan sesuai rancangan dengan membandingkan keluaran nyata dan keluaran yang diharapkan tanpa memeriksa kode. Pengujian kategori intent memastikan seluruh kategori pada basis pengetahuan dapat dipanggil dan menghasilkan jawaban. Penilaian pengguna menilai penerimaan terhadap kemudahan, kejelasan, akurasi, dan manfaat produk melalui angket berskala Likert dan pertanyaan terbuka. Subjek uji coba ditentukan secara purposive, yaitu pihak yang memahami isi informasi PSB sehingga mampu menilai ketepatan jawaban, terdiri atas ketua panitia, admin, staf, petugas layanan informasi, guru, dan wali santri, dengan jumlah 23 responden yang berada pada rentang lazim uji coba terbatas.

Instrumen utama adalah angket skala Likert lima tingkat (1 = sangat tidak setuju hingga 5 = sangat setuju) yang memuat tujuh butir berskala dan dua pertanyaan terbuka. Butir disusun berdasarkan aspek kemudahan penggunaan, kemudahan navigasi, akurasi jawaban, kejelasan penyajian, kelengkapan media, kinerja pengenalan bahasa, dan manfaat terhadap layanan. Data kuantitatif dianalisis dengan statistika deskriptif. Persentase kelayakan dihitung dengan rumus:

$$\text{Persentase kelayakan} = (\text{Total skor yang diperoleh} \div \text{Skor maksimal}) \times 100\%$$

Skor maksimal satu butir diperoleh dari jumlah responden dikalikan skor tertinggi, sedangkan skor maksimal seluruh instrumen diperoleh dari jumlah butir dikalikan skor maksimal satu butir. Hasil persentase ditafsirkan ke dalam lima kategori: sangat baik (81–100%), baik (61–80%), cukup (41–60%), kurang (21–40%), dan sangat kurang (0–20%); produk dinyatakan layak apabila berada pada kategori baik atau sangat baik. Data kualitatif dari pertanyaan terbuka dianalisis melalui reduksi data, pengelompokan tema, dan penarikan kesimpulan sebagai dasar penyusunan rencana perbaikan produk.

4. HASIL DAN PEMBAHASAN

Analisis Kebutuhan dan Basis Pengetahuan FAQ

Kebutuhan informasi pengguna dikelompokkan menjadi tiga rumpun, yaitu informasi administrasi pendaftaran (alur pendaftaran, syarat berkas, cara pembayaran,

jadwal tes), informasi akademik dan kehidupan santri (jenjang pendidikan, kegiatan harian, materi tes, fasilitas, ekstrakurikuler), serta informasi lokasi dan komunikasi (alamat, rute, prosedur kunjungan, kontak panitia). Setiap kebutuhan dipetakan pada satu intent dengan satu respons utama agar jawaban yang diterima pengguna bersifat konsisten. Basis pengetahuan disusun dalam berkas intents.json yang, berdasarkan hasil identifikasi, terdiri atas 21 kategori intent, 356 pola pertanyaan, dan 21 respons utama. Setiap intent memuat tiga unsur, yaitu tag sebagai label kategori, patterns sebagai variasi pertanyaan, dan responses sebagai jawaban. Struktur ini memisahkan data dari kode sehingga penambahan atau perbaikan FAQ dapat dilakukan tanpa mengubah logika program. Rekapitulasi kategori intent disajikan pada Tabel 1.

Tabel 1. Rekapitulasi kategori intent pada dataset intents.json

No	Tag Intent	Jumlah Pola	Ringkasan Respons
1	salam	70	Salam pembuka dan tawaran bantuan informasi PSB
2	penutup	36	Ucapan terima kasih dan penutup percakapan
3	profil_pesantren	20	Profil lembaga pendidikan Islam berasrama
4	lokasi	19	Alamat lengkap dan nomor telepon pesantren
5	kunjungan_lokasi	7	Prosedur, alamat, dan jadwal survei/kunjungan
6	rute	17	Rute dari tol lingkar luar, bandara, dan Kebayoran Lama
7	jenjang_pendidikan	22	Jenjang pendidikan dari PAUD hingga MA berakreditasi A
8	kegiatan_harian	6	Jadwal kegiatan santri pukul 04.00 hingga malam
9	perbedaan_pesantren	13	Kekhasan kurikulum terpadu dan pembinaan santri
10	biaya	22	Rincian estimasi biaya pendaftaran dan biaya bulanan
11	cara_pembayaran	6	Langkah pembayaran melalui ATM dan internet banking
12	waktu_dan_alur_daftar	19	Alur pendaftaran daring melalui portal resmi
13	syarat_pendaftaran	11	Persyaratan umum dan dokumen yang disiapkan
14	pendaftaran_mutasi	6	Dokumen tambahan untuk santri mutasi/pindahan
15	materi_tes_masuk	10	Materi ujian masuk setiap jenjang
16	fasilitas	17	Masjid, perpustakaan, laboratorium, dan sarana lain

No	Tag Intent	Jumlah Pola	Ringkasan Respons
17	penanganan_anak	12	Pola pembinaan bersama wali kamar dan pengasuhan
18	ekstrakurikuler	16	Kegiatan wajib dan kegiatan pilihan santri
19	brostur	8	Tautan unduh dan gambar brosur resmi
20	kontak_panitia	7	Nomor kontak panitia PSB
21	alumni	12	Sebaran alumni pada perguruan tinggi dalam/luar negeri
Jumlah	356	21 intent dan 21 respons utama	

Sumber: hasil identifikasi berkas intents.json (2026)

Arsitektur Sistem dan Algoritma NLP

Produk berbentuk program yang dijalankan pada komputer dan terhubung dengan akun bot Telegram melalui Telegram Bot API, sehingga pengguna cukup memakai aplikasi Telegram yang umum tanpa memasang aplikasi khusus. Arsitektur berjalan pada lingkungan lokal dengan metode komunikasi long polling. Alur kerja dimulai ketika pengguna mengirim pertanyaan melalui Telegram; pesan diteruskan Telegram Bot API menuju bot server Python yang mengeksekusi modul prapemrosesan teks (*preprocess_text*) dan modul pencocokan (*cari_jawaban*), lalu mengambil respons yang paling sesuai dari *intents.json* dan mengirimkannya kembali kepada pengguna. Selain menerima pertanyaan bebas, sistem menyediakan enam tombol menu untuk topik yang paling sering ditanyakan serta fitur pengiriman berkas gambar brosur ketika kata “brostur” terdeteksi.

Perlu ditegaskan bahwa algoritma yang diterapkan adalah pendekatan berbasis aturan dengan pencocokan leksikal, yaitu penentuan intent melalui penghitungan irisan kata dasar (*token overlap*) antara pertanyaan pengguna dan pola di dalam dataset. Sistem ini secara tegas tidak menggunakan model klasifikasi statistik (seperti Naive Bayes atau SVM), representasi vektor kata (seperti Word2Vec atau TF-IDF), maupun model transformer (seperti BERT); tidak ada proses pelatihan model dan tidak ada bobot yang dipelajari dari data, sehingga seluruh keputusan sistem dapat ditelusuri langsung dari isi dataset. Pemilihan pendekatan ini didasarkan pada tiga pertimbangan: ruang lingkup layanan bersifat tertutup dengan kategori terbatas; belum tersedia dataset pertanyaan berlabel yang memadai karena riwayat percakapan PSB belum terdokumentasi digital; serta pendekatan ini dapat dijalankan mandiri tanpa mengirim data pengguna ke pihak ketiga. Penerapan NLP terletak pada rangkaian prapemrosesan teks berbahasa Indonesia

dan mekanisme penentuan maksud berbasis kata dasar. Kelima tahap prapemrosesan dirangkum pada Tabel 2.

Tabel 2. Tahapan prapemrosesan teks dan fungsinya

No	Tahap	Operasi yang Dijalankan	Alasan
1	Case folding	Seluruh huruf diubah menjadi huruf kecil.	Menyamakan perlakuan terhadap perbedaan kapitalisasi.
2	Pembersihan karakter	Karakter selain huruf (angka, tanda baca, lambang) dihapus.	Menghilangkan derau yang tidak menentukan maksud pertanyaan.
3	Tokenisasi	Kalimat dipecah menjadi token kata tunggal berdasarkan spasi.	Menyiapkan satuan terkecil yang dibandingkan.
4	Stopword removal	Token yang termasuk daftar kata umum dihapus (mis. yang, ke, di, dan).	Menyisakan kata yang lebih menentukan makna.
5	Stemming	Setiap token dikembalikan ke bentuk kata dasar memakai pustaka Sastrawi.	Menyatukan pendaftaran, mendaftar, dan daftar menjadi satu bentuk.

Sumber: hasil pengembangan produk (2026)

Rangkaian yang sama diterapkan pada pertanyaan pengguna dan pada setiap pola dataset, sehingga keduanya dibandingkan dalam bentuk himpunan kata dasar yang setara; masukan “Berapa biaya masuk santri baru ya?”, misalnya, menghasilkan himpunan {berapa, biaya, masuk, santri, baru, ya} yang dicocokkan dengan pola dataset untuk memilih intent berskor irisan tertinggi. Perlu dicatat bahwa daftar stopwords pada produk awal masih terbatas pada sepuluh kata, sehingga sejumlah kata umum yang tidak informatif belum tersaring dan berpotensi menambah kecocokan semu; hal ini menjadi salah satu catatan keterbatasan yang dibahas kemudian.

Hasil Pengujian Fungsional

Pengujian black box disusun berdasarkan fitur utama, yaitu perintah awal, menu tombol, pencarian jawaban FAQ, pengiriman brosur, dan penanganan pertanyaan di luar cakupan. Dari tiga belas skenario yang diuji, seluruhnya memperoleh status berhasil sehingga persentase kesesuaian fungsional mencapai 100 persen. Ringkasan hasil disajikan pada Tabel 3. Perlu digarisbawahi bahwa hasil ini membuktikan berjalannya fitur pada masukan yang diuji, bukan bahwa sistem mampu memahami seluruh ragam pertanyaan pengguna.

Tabel 3. Ringkasan hasil pengujian black box

No	Masukan Uji	Keluaran yang Diharapkan	Status
1	/start	Menampilkan salam pembuka dan menu tombol.	Berhasil
2	Berapa biaya masuk	Memberikan rincian estimasi biaya.	Berhasil
3	Apa saja syarat pendaftaran	Memberikan daftar syarat dan dokumen.	Berhasil
4	Dimana lokasi	Memberikan alamat lengkap pesantren.	Berhasil

No	Masukan Uji	Keluaran yang Diharapkan	Status
5	Bagaimana rutenya	Memberikan rute menuju Darunnajah.	Berhasil
6	Apa saja fasilitas	Memberikan informasi fasilitas pesantren.	Berhasil
7	Kegiatan ekstrakurikuler	Memberikan daftar kegiatan wajib dan pilihan.	Berhasil
8	minta brosur	Memberi tautan brosur dan mengirim berkas gambar.	Berhasil
9	halo	Memberikan respons salam pembuka.	Berhasil
10	terima kasih	Memberikan respons penutup.	Berhasil
11	anak hiperaktif	Memberikan respons tentang pembinaan santri.	Berhasil
12	jadwal turnamen bola	Menyatakan pertanyaan belum dikenali.	Berhasil
13	12345!!!	Menyatakan pertanyaan kurang jelas.	Berhasil

Sumber: hasil pengujian (2026)

Pengujian kategori intent dilakukan dengan mengambil satu contoh pertanyaan yang mewakili setiap intent. Seluruh 21 kategori menghasilkan respons yang sesuai, sehingga tingkat keberhasilannya 100 persen. Angka ini perlu dibaca dengan hati-hati: masukan uji diambil dari pola yang telah ada di dalam dataset, sehingga pengujian ini bersifat verifikasi bahwa struktur dataset terbaca dan seluruh kategori dapat dipanggil, bukan pengujian kemampuan menggeneralisasi pertanyaan baru. Pengujian generalisasi memerlukan rancangan tersendiri dengan pertanyaan di luar pola serta metrik accuracy, precision, recall, dan F1-score.

Hasil Penilaian Pengguna

Rekapitulasi jawaban 23 responden atas tujuh butir berskala disajikan pada Tabel 4. Seluruh responden memilih skor lima pada setiap butir, sehingga total skor yang diperoleh sama dengan skor maksimal, yaitu 805 dari 805 atau 100 persen dengan rata-rata 5,00 pada kategori sangat baik.

Tabel 4. Rekapitulasi skor penilaian pengguna pada setiap butir

No	Aspek yang Dinilai	Rerata	Persentase	Kategori
1	Kemudahan akses dan penggunaan (P1)	5,00	100%	Sangat baik
2	Kemudahan navigasi menu tombol (P2)	5,00	100%	Sangat baik
3	Akurasi jawaban terhadap informasi resmi (P3)	5,00	100%	Sangat baik
4	Kejelasan bahasa dan format jawaban (P4)	5,00	100%	Sangat baik
5	Manfaat fitur pengiriman brosur (P5)	5,00	100%	Sangat baik
6	Pengenalan variasi kalimat dan kata kunci (P6)	5,00	100%	Sangat baik
7	Manfaat terhadap beban kerja staf PSB (P7)	5,00	100%	Sangat baik

No	Aspek yang Dinilai	Rerata	Persentase	Kategori
Jumlah keseluruhan (805/805)	5,00	100%	Sangat baik	

Sumber: hasil penilaian pengguna melalui Google Forms (2026)

Selain butir berskala, dua pertanyaan terbuka memberikan gambaran yang tidak tertangkap skala kuantitatif. Pada pertanyaan mengenai topik yang belum terjawab, dari 17 responden yang mengisi, sepuluh orang (58,82%) menyatakan seluruh pertanyaannya terjawab, sedangkan tujuh orang (41,18%) menyebutkan setidaknya satu topik yang belum tersedia. Rincian topik di luar cakupan disajikan pada Tabel 5. Pada pertanyaan terbuka kedua, dua belas responden memberikan apresiasi tanpa usulan teknis, sedangkan delapan responden lainnya menyampaikan usulan operasional, di antaranya penyediaan opsi eskalasi ke petugas manusia, peningkatan kepekaan terhadap sinonim, perluasan kanal ke WhatsApp, dukungan multibahasa, dan permintaan penerapan nyata.

Tabel 5. Topik pertanyaan yang berada di luar cakupan dataset

No	Topik yang Ditanyakan	Frekuensi	Usulan Penanganan
1	Jumlah santri aktif saat ini	2	Menambah intent jumlah_santri, diperbarui tiap tahun ajaran.
2	Kapasitas dan jumlah penghuni kamar asrama	1	Menambah intent kapasitas_asrama.
3	Menu makan dan konsumsi santri	1	Menambah intent konsumsi_santri.
4	Prestasi dan kejuaraan yang diraih	1	Menambah intent prestasi.
5	Sistem organisasi santri	1	Menambah intent organisasi_santri.
6	Permintaan foto fasilitas	1	Menambah berkas gambar pada respons intent fasilitas.
Jumlah	7	Berasal dari 7 responden yang berbeda.	

Sumber: hasil analisis pertanyaan terbuka (2026)

Pembahasan

Rumusan pertama menyangkut perancangan dan pembangunan chatbot. Hasil penelitian mengonfirmasi bahwa rancangan sistem berhasil dibangun dengan mengintegrasikan enam komponen utama, yakni pengguna Telegram, Telegram Bot API, bot server Python, modul prapemrosesan teks, algoritma pencocokan intent, dan basis pengetahuan intents.json. Keberhasilan arsitektur dibuktikan melalui pengujian black box pada 13 skenario dan verifikasi 21 kategori intent yang seluruhnya menghasilkan luaran secara presisi. Efektivitas ini bersumber dari pemisahan antara data basis pengetahuan dan logika program, sehingga pembaruan variasi pertanyaan dapat dilakukan tanpa

mengubah baris kode—fleksibilitas yang krusial bagi lembaga yang belum memiliki tim pengembang khusus. Temuan ini sejalan dengan penelitian terdahulu yang memanfaatkan Telegram sebagai antarmuka layanan informasi (Ai, 2022; Puspita, 2026), namun berbeda karena dibangun mandiri tanpa mesin NLP pihak ketiga sehingga data pengguna tetap dalam kendali lembaga.

Rumusan kedua menyangkut peran produk terhadap efisiensi layanan. Sistem mampu merespons pertanyaan baku di dalam dataset secara otomatis tanpa keterlibatan petugas dan beroperasi 24 jam; seluruh responden menyetujui bahwa sistem efektif memangkas beban kerja staf. Namun, persentase peningkatan efisiensi belum dapat diukur secara eksak karena tidak tersedia data metrik operasional seperti perbandingan volume pesan manual sebelum-sesudah atau durasi balasan, mengingat prototipe belum memiliki mekanisme pencatatan (logging) dan belum berjalan di peladen produksi. Karena itu klaim efisiensi bersifat indikatif-potensial. Efisiensi optimal pun hanya berlaku pada pertanyaan umum bernilai tetap; untuk verifikasi dokumen, pengecekan pembayaran, atau keputusan administratif, peran panitia tetap tidak tergantikan, sehingga fitur eskalasi ke petugas menjadi kebutuhan mendesak.

Rumusan ketiga menyangkut persepsi dan pengalaman pengguna. Hasil kuesioner mencatat kepuasan sempurna (100%, rata-rata 5,00). Namun keseragaman nilai tertinggi menimbulkan tiga catatan metodologis: terjadi ceiling effect sehingga instrumen tidak lagi membedakan tingkat penilaian; simpangan baku bernilai nol sehingga Cronbach's Alpha tak dapat dihitung; serta potensi pengisian ganda tidak dapat sepenuhnya disingkirkan. Selain itu, mayoritas responden merupakan pihak internal panitia PSB yang memadai untuk memvalidasi akurasi isi tetapi belum sepenuhnya mewakili profil calon santri atau wali awam. Evaluasi kualitatif memperjelas gambaran ini: 41,18% responden pertanyaan terbuka masih menemukan pertanyaan spesifik yang belum terjawab. Enam topik yang belum tercakup bersifat out of scope—informasi yang belum dimasukkan, bukan kesalahan pencocokan—sehingga perbaikan paling berdampak adalah penambahan intent baru dan pengayaan variasi pola. Angka 100% dengan demikian merupakan indikasi penerimaan, bukan ukuran ketelitian sistem yang sesungguhnya perlu diuji melalui pertanyaan di luar pola.

5. KESIMPULAN DAN SARAN

Chatbot Telegram untuk layanan FAQ penerimaan santri baru di Pondok Pesantren Darunnajah berhasil dibangun menggunakan Python, Telegram Bot API, dan pustaka Sastrawi, dengan basis pengetahuan intents.json yang memuat 21 kategori intent, 356 pola pertanyaan, dan 21 respons utama, di mana maksud pertanyaan ditentukan melalui pencocokan irisan kata dasar yang didahului lima tahap prapemrosesan. Produk dinyatakan layak secara fungsional karena pengujian black box menghasilkan kesesuaian pada tiga belas dari tiga belas skenario dan pengujian kategori intent pada dua puluh satu dari dua puluh satu kategori, dengan kegagalan yang ditemukan bersifat out of scope, bukan kesalahan pencocokan. Persepsi dan pengalaman pengguna berada pada kategori sangat baik dengan total skor 805 dari 805 (100%) dan rata-rata 5,00 dari 23 responden. Namun demikian, angka tersebut merupakan indikasi penerimaan, bukan ukuran ketelitian sistem, dan produk masih berstatus prototipe pada lingkungan lokal sehingga memerlukan pemindahan ke lingkungan produksi sebelum dipakai sebagai layanan resmi.

Berdasarkan kesimpulan dan keterbatasan tersebut, disarankan agar sistem dipindahkan ke server produksi dengan komunikasi webhook agar aktif 24 jam secara aman, basis pengetahuan dimigrasikan ke basis data berantarmuka pengelola (dashboard) sehingga panitia non-teknis dapat memperbaruinya sendiri, serta diaktifkan fungsi pencatatan pesan dan jalur eskalasi otomatis ke petugas untuk pertanyaan khusus. Dari sisi metode pemrosesan bahasa, algoritma pencocokan perlu ditingkatkan ke arah kemiripan makna (semantic similarity), misalnya TF-IDF dengan cosine similarity atau model berbasis IndoBERT, disertai perluasan variasi pola dan penambahan topik yang belum terakomodasi. Penelitian lanjutan disarankan mengukur kinerja sistem secara objektif menggunakan metrik accuracy, precision, recall, dan F1-score dengan pertanyaan di luar pola, serta melibatkan calon santri dan wali sebagai responden utama agar hasil evaluasi lebih mewakili sasaran layanan. Keterbatasan utama penelitian ini terletak pada status produk sebagai prototipe lokal, komposisi responden yang didominasi pihak internal, serta instrumen yang mengalami ceiling effect.

UCAPAN TERIMA KASIH

Artikel ini merupakan bagian dari skripsi penulis pada Program Studi Sistem dan Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Darunnajah Jakarta. Penulis menyampaikan terima kasih kepada para dosen pembimbing atas bimbingan dan koreksinya, serta kepada panitia dan staf layanan Penerimaan Santri Baru Pondok Pesantren Darunnajah Jakarta yang telah membantu proses pengumpulan data dan pengujian produk.

DAFTAR REFERENSI

- Ai, W. (2022). Perancangan dan pembuatan aplikasi chatbot Telegram untuk layanan informasi akademik Fakultas Vokasi Universitas Negeri Surabaya metode Natural Language Processing menggunakan Wit.AI. *Jurnal Manajemen Informatika*, 12(2), 1–10.
- Al Farisi, F. A., Perdana, R. S., & Adikara, P. P. (2024). Implementasi chatbot berbasis Natural Language Processing menggunakan pustaka Sastrawi untuk layanan informasi. *Journal of Computer Science and Research*, 2(2), 101–112.
- Alpariji, S. (2024). *Perancangan chatbot WhatsApp dengan teknologi Natural Language Processing untuk layanan informasi akademik*. Repositori Universitas Nusa Putra.
- Ardabili, M. M. (2025). *Pembuatan chatbot berbasis Natural Language Processing (NLP) sebagai sarana komunikasi menggunakan Dialogflow dan integrasi Telegram*. Repositori STT Terpadu Nurul Fikri.
- Bastian, A., & Saputra, E. (2023). Perancangan chatbot informasi akademik berbasis Natural Language Processing menggunakan platform Telegram. *Jurnal Rekayasa Sistem dan Teknologi Informasi*, 7(2), 145–154.
- Caldarini, G., Jaf, S., & McGarry, K. (2022). A literature survey of recent advances in chatbots. *Information*, 13(1), 41. <https://doi.org/10.3390/info13010041>
- Dewi, R. K., & Kurniawan, A. (2024). Evaluasi kinerja chatbot Telegram berbasis NLP dalam layanan informasi FAQ pendidikan tinggi. *Jurnal Informatika dan Sistem Informasi*, 10(1), 78–89.
- Furqan, M. (2023). Implementasi chatbot dalam sistem layanan publik berbasis NLP. *Jurnal Teknologi dan Riset Komputer*, 8(3), 102–110.
- Hidayat, A., & Pratama, R. (2023). Metodologi penelitian pengembangan (R&D) dalam bidang teknologi informasi dan rekayasa perangkat lunak. *Jurnal Edukasi dan Teknologi Informasi*, 5(1), 34–45.
- Hidayat, R., & Nurjanah, S. (2024). Penerapan model intent classification dan entity recognition pada chatbot WhatsApp layanan institusi. *Jurnal Ilmiah Komputer dan Informatika (JIKI)*, 12(2), 201–212.
- Indonesian Journal of Statistics and Its Applications. (2025). Exploring rule-based approach and text preprocessing techniques on Indonesian text. *Indonesian Journal of Statistics and Its Applications*, 9(1), 45–58.

- Irawan, D. S., & Parapat, D. A. S. H. (2024). Pengembangan dan implementasi chatbot pelayanan masyarakat berbasis Dialogflow di Kelurahan Gebang Putih Surabaya. *El-Mujtama: Jurnal Pengabdian Masyarakat*, 4(2), 486–493. <https://doi.org/10.47467/elmutjama.v5i1.6262>
- Irfani, M. H., Mair, Z. R., Cahyani, S., & Ikhwan, M. (2022). Implementasi chatbot Telegram layanan informasi akademik Universitas Indo Global Mandiri menggunakan framework Rasa Open Source. *Jurnal Sistem dan Teknologi Informasi*, 2(2), 122–134.
- Mustafa, P. S., Permadi, A. G., & Masdar, M. (2024). *Metodologi penelitian dan pengembangan (Research and Development) dalam bidang pendidikan dan teknologi*. Deepublish.
- Nugroho, A. S., & Wibowo, A. (2023). Perbandingan algoritma stemming Nazief-Adriani dan Tala dalam pemrosesan teks Bahasa Indonesia pada sistem informasi. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 10(3), 589–598.
- Pratama, I. G. A., & Setiawan, B. (2023). Otomatisasi pelayanan FAQ menggunakan chatbot WhatsApp API berbasis Natural Language Processing. *Jurnal Sistem Informasi dan Teknologi Komputer*, 8(1), 33–42.
- Puspita, R. D. (2026). Pengembangan chatbot Telegram berbasis NLP untuk layanan informasi akademik. *Jurnal Inovasi TIK dan Komputasi*, 5(2), 112–121.
- Rahadian, E. (2025). Pengembangan aplikasi chatbot WhatsApp, Telegram, dan website laporan untuk instansi pemerintah berbasis Natural Language Processing. *Jurnal Manajemen Informatika & Teknologi*, 6(1), 54–71.
- Ramadhan, F., & Yulia, Y. (2024). Analisis efektivitas algoritma cosine similarity dan TF-IDF pada chatbot FAQ lembaga pendidikan. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 8(4), 512–521.
- Rizqi, R. M. (2025). Integrasi algoritma NLP untuk peningkatan kecerdasan chatbot: Kajian literatur dan analisis perkembangan terkini. *Jurnal Riset Teknik Komputer*, 2(4), 157–171.
- Santoso, H., & Pratami, A. (2023). Integrasi bot Telegram dan Natural Language Processing untuk layanan customer service otomatis. *Jurnal Teknologi Informasi dan Rekayasa Komputer*, 4(3), 210–219.
- Suryani, L., & Wijaya, T. (2024). Pemanfaatan pustaka Sastrawi dan algoritma Levenshtein Distance pada chatbot FAQ berbahasa Indonesia. *Jurnal Algoritma dan Pemrograman*, 11(2), 99–108.
- Susanto, E. N. E., & Akbar, M. (2026). Implementasi chatbot WhatsApp berbasis NLP untuk otomasi layanan pelanggan. *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, 6(1), 12–24.
- Umam, A. K., Wijayanti, E., & Chamid, A. A. (2025). Penerapan NLP pada chatbot Telegram untuk layanan informasi otomatis. *Bulletin of Computer Science Research*, 5(4), 329–339.
- Utomo, P., & Arifin, M. (2024). Penggunaan webhook dan Node.js dalam pembangunan chatbot WhatsApp untuk layanan administrasi. *Jurnal Sistem Komputer dan Inteligensia Artificial*, 3(1), 45–56.
- Wibowo, M. A., & Susanti, E. (2023). Arsitektur chatbot berbasis NLP menggunakan framework Rasa untuk asisten virtual FAQ. *Jurnal Rekayasa Perangkat Lunak Indonesia*, 7(1), 15–26.