

**RANCANG BANGUN CHATBOT INFORMASI PESANTREN BERBASIS
WHATSAPP MENGGUNAKAN METODE SIMPLE CONTRASTIVE
SENTENCE EMBEDDING DAN COSINE SIMILARITY**

SKRIPSI

Oleh :
MUCHAMMAD RIZAL AMMAR
NIM. 19650107



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

**RANCANG BANGUN CHATBOT INFORMASI PESANTREN BERBASIS
WHATSAPP MENGGUNAKAN METODE SIMPLE CONTRASTIVE
SENTENCE EMBEDDING DAN COSINE SIMILARITY**

SKRIPSI

Diajukan kepada:

Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :

MUCHAMMAD RIZAL AMMAR
NIM. 19650107

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

HALAMAN PERSETUJUAN

RANCANG BANGUN CHATBOT INFORMASI PESANTREN BERBASIS WHATSAPP MENGGUNAKAN METODE SIMPLE CONTRASTIVE SENTENCE EMBEDDING DAN COSINE SIMILARITY

SKRIPSI

Oleh :
MUCHAMMAD RIZAL AMMAR
NIM. 19650107

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 22 Juni 2026

Pembimbing I,



Fatchurrochman, M.Kom
NIP. 19700731 200501 1 002

Pembimbing II,



Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M. Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

RANCANG BANGUN CHATBOT INFORMASI PESANTREN BERBASIS WHATSAPP MENGGUNAKAN METODE SIMPLE CONTRASTIVE SENTENCE EMBEDDING DAN COSINE SIMILARITY

SKRIPSI

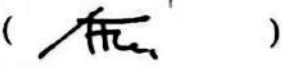
Oleh :

MUCHAMMAD RIZAL AMMAR
NIM. 19650107

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 22 Juni 2026

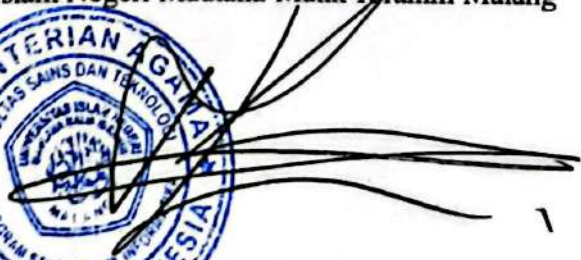
Susunan Dewan Penguji

Ketua Penguji	: <u>Supriyono, M. Kom</u> NIP. 19841010 201903 1 012
Anggota Penguji I	: <u>Shoffin Nahwa Utama, MT</u> NIP. 19860703 202012 1 003
Anggota Penguji II	: <u>Fatchurrochman, M.Kom</u> NIP. 19700731 200501 1 002
Anggota Penguji III	: <u>Dr. Zainal Abidin, M.Kom</u> NIP. 19760613 200501 1 004

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Supriyono, M. Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Muchammad Rizal Ammar
NIM : 19650107
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Rancang Bangun Chatbot Informasi Pesantren
Berbasis WhatsApp Menggunakan Metode *Simple Contrastive Sentence Embedding* dan *Cosine Similarity*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 25 Juni 2026

Yang membuat pernyataan,



Muchammad Rizal Ammar

NIM.19650107

MOTTO

إِنَّ مَعَ الْعُسْرِ يُسْرًا

"Sesungguhnya bersama kesulitan ada kemudahan."

(QS. Al-Insyirah: 6)

"Those who fly solo have the strongest wings."

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan kepada:

Allah Subhanahu wa Ta'ala atas segala rahmat, nikmat, dan kemudahan yang diberikan kepada penulis dalam menyelesaikan skripsi ini.

Kedua orang tua tercinta, yang tiada henti mendoakan, mendukung, dan memberi semangat kepada penulis. Doa dan kasih sayang kalian adalah kekuatan terbesar dalam setiap langkah perjalanan ini.

Seluruh keluarga besar yang selalu memberikan dukungan moral dan doa yang tulus selama penulis menempuh pendidikan.

Bapak dan Ibu Dosen Program Studi Teknik Informatika UIN Maulana Malik Ibrahim Malang yang telah membimbing dan mendidik dengan penuh kesabaran dan keikhlasan.

Sahabat-sahabat seperjuangan Teknik Informatika angkatan 2019 yang telah menemani perjalanan akademik ini dengan penuh kebersamaan dan keceriaan.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur penulis panjatkan kehadirat Allah Subhanahu wa Ta'ala yang telah memberikan rahmat, taufiq, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Rancang Bangun Chatbot Informasi Pesantren Berbasis WhatsApp Menggunakan Metode Simple Contrastive Sentence Embedding dan Cosine Similarity" dengan baik.

Shalawat serta salam senantiasa penulis haturkan kepada junjungan Nabi Muhammad Shallallahu 'Alaihi wa Sallam yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang penuh dengan ilmu pengetahuan.

Penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si. selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan kesempatan kepada penulis untuk menempuh pendidikan di UIN Malang.
2. Dr. H. Agus Mulyono, M.Si. selaku Dekan Fakultas Sains dan Teknologi yang telah memimpin dan memajukan lingkungan akademik tempat penulis belajar.
3. Supriyono, M.Kom. selaku Ketua Program Studi Teknik Informatika yang telah memberikan arahan dan kebijakan akademik selama penulis berkuliah.

4. Dosen Pembimbing I dan Dosen Pembimbing II yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta motivasi kepada penulis dalam menyelesaikan skripsi ini.
5. Seluruh Dosen dan Staf Program Studi Teknik Informatika yang telah memberikan ilmu pengetahuan, pengalaman, dan pelayanan akademik yang berharga selama masa perkuliahan.
6. Pesantren Mahasiswa Al-Hikam Malang yang telah memberikan izin dan dukungan dalam proses pengumpulan data penelitian ini.
7. Kedua orang tua tercinta dan seluruh keluarga besar yang senantiasa memberikan doa, dukungan moril, dan semangat yang tiada henti.
8. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu yang telah memberikan bantuan dan kontribusi dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan dan jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari semua pihak demi penyempurnaan penelitian selanjutnya. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak, khususnya dalam pengembangan teknologi kecerdasan buatan di bidang pendidikan Islam.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 19 Juni 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN TULISAN.....	iv
MOTTO.....	v
HALAMAN PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	xiii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR SIMBOL.....	xi
ABSTRAK.....	xii
ABSTRACT.....	xiii
مستخلص البحث.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian.....	7
1.4 Manfaat Penelitian.....	7
1.5 Batasan Masalah.....	9
1.6 Metode Penelitian.....	9
1.6.1 Pola dan Rancangan Penelitian.....	10
1.6.2 Bahan atau Materi Penelitian.....	10
1.6.3 Alat Penelitian	11
1.6.4 Langkah-Langkah Penelitian.....	11
1.6.5 Sistematika Penulisan	12
BAB II STUDI PUSTAKA.....	14
2.1 Landasan Teori.....	18
2.1.1 Chatbot.....	18
2.1.2 Natural Language Processing (NLP).....	18
2.1.3 Sentence Embedding	19

2.1.4 SimCSE (Simple Contrastive Learning of Sentence Embeddings).....	19
2.1.5 Perbandingan dengan SBERT	19
2.1.6 Integrasi Chatbot dengan WhatsApp.....	20
BAB III METODOLOGI PENELITIAN	21
3.1 Rancangan Penelitian.....	21
3.2 Pengumpulan Data.....	21
3.2.1 Data Primer.....	21
3.2.2 Data Sekunder.....	22
3.3 Metode Pengembangan Sistem	22
3.3.1 Preprocessing Teks.....	25
3.3.2 SimCSE (Simple Contrastive Learning of Sentence Embeddings).....	26
3.3.3 Cosine Similarity.....	29
3.3.3.1 Menghitung Cosine Similarity antara A dan B.....	29
3.3.3.2 Menghitung Cosine Similarity antara A dan C.....	30
3.3.4 Integrasi ke WhatsApp.....	32
BAB IV MPLEMENTASI DAN HASIL PENELITIAN.....	33
4.1 Implementasi Sistem.....	33
4.1.1 Lingkungan Pengembangan	33
4.1.2 Implementasi Dataset FAQ	34
4.1.3 Implementasi Preprocessing Teks.....	38
4.1.4 Implementasi Model SimCSE	39
4.1.5 Implementasi Cosine Similarity dan Pengambilan Jawaban.....	40
4.1.6 Implementasi Integrasi WhatsApp	43
4.2 Hasil Pengujian	44
4.2.1 Pengujian Akurasi Pencocokan Jawaban.....	44
4.2.2 Pengujian Kecepatan Respons.....	46
4.2.3 Pengujian Kepuasan Pengguna.....	47
4.3 Pembahasan	50
BAB V KESIMPULAN DAN SARAN.....	53
5.1 Kesimpulan.....	53
5.2 Saran	54
DAFTAR PUSTAKA	56

LAMPIRAN.....	58
Lampiran 1. Dataset FAQ Chatbot Pesantren Al-Hikam Malang.....	58

DAFTAR TABEL

Tabel 2.1 Penelitian Sebelumnya.....	15
Tabel 3.1 Contoh Hasil Preprocessing Pertanyaan FAQ.....	24
Tabel 3.2 Perbandingan hasil Cosine Similarity.....	31
Tabel 4.1 Spesifikasi Kolom Dataset faq.csv.....	34
Tabel 4.2 Sampel Data Training Chatbot FAQ Pesantren Mahasiswa Al-Hikam Malang.....	36
Tabel 4.3 Sampel Data Pengujian Chatbot FAQ Pesantren Mahasiswa Al-Hikam Malang.....	36
Tabel 4.4 Hasil Pengujian Berbagai Nilai Treshold.....	41
Tabel 4.5 Hasil Pengujian Akurasi Chatbot per Kategori FAQ.....	44
Tabel 4.6 Data Respons Kuesioner SUS dari 20 Responden.....	48
Tabel 4.7 Perhitungan Skor SUS Pertanyaan.....	48

DAFTAR GAMBAR

Tabel 2.1 Penelitian Sebelumnya.....	15
Tabel 3.1 Contoh Hasil Preprocessing Pertanyaan FAQ.....	24
Tabel 3.2 Perbandingan hasil Cosine Similarity.....	31

DAFTAR SIMBOL

$\text{sim}(A, B)$: Fungsi cosine similarity antara vektor A dan vektor B

$A \cdot B$: Dot product (perkalian titik) antara vektor A dan vektor B

$\|A\|$: Norma Euclidean (magnitudo) dari vektor A

Σ : Operator penjumlahan (sigma)

h_i : Representasi embedding kalimat ke-i dari model SimCSE

$h+_i$: Representasi embedding pasangan positif (positive pair) kalimat ke-i

τ : Parameter temperatur (temperature) pada fungsi InfoNCE loss

L_{InfoNCE} : Fungsi InfoNCE loss yang digunakan sebagai objektif pelatihan
SimCSE

N : Ukuran batch (jumlah sampel) dalam satu iterasi pelatihan

θ : Nilai threshold cosine similarity untuk filter relevansi jawaban

$\exp(x)$: Fungsi eksponensial dengan basis e, yaitu e^x

ABSTRAK

Ammar, Muchammad Rizal. 2025. Rancang Bangun Chatbot Informasi Pesantren Berbasis WhatsApp Menggunakan Metode Simple Contrastive Sentence Embedding dan Cosine Similarity. Skripsi. Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Kata Kunci: *chatbot, Cosine Similarity, Natural Language Processing, pesantren, SimCSE, WhatsApp*

Perkembangan teknologi informasi membuka peluang luas bagi lembaga pendidikan Islam untuk meningkatkan kualitas layanan kepada masyarakat, termasuk Pesantren Mahasiswa Al-Hikam Malang yang membutuhkan sistem informasi mampu menjawab pertanyaan calon santri secara cepat, akurat, dan dapat diakses kapan saja. Penelitian ini merancang dan membangun chatbot informasi pesantren berbasis WhatsApp menggunakan metode Simple Contrastive Sentence Embedding (SimCSE) dan Cosine Similarity, dengan dataset 50 pasangan pertanyaan-jawaban (FAQ) yang mencakup berbagai kategori informasi pesantren. SimCSE digunakan untuk menghasilkan representasi vektor semantik dari setiap pertanyaan dalam dataset, sementara Cosine Similarity mengukur kemiripan antara pertanyaan pengguna dan dataset untuk mengembalikan jawaban yang paling relevan. Sistem diimplementasikan pada platform WhatsApp menggunakan API Twilio dan framework Flask berbasis Python, dengan hasil pengujian akurasi sebesar 85% dari 100 pertanyaan uji serta skor System Usability Scale (SUS) rata-rata 77,25 yang termasuk kategori Good dari 20 responden. Hasil ini menunjukkan bahwa metode SimCSE dan Cosine Similarity efektif digunakan dalam sistem chatbot berbasis retrieval untuk domain informasi pesantren.

ABSTRACT

Ammar, Muchammad Rizal. 2025. Design and Development of WhatsApp-Based Pesantren Information Chatbot Using Simple Contrastive Sentence Embedding and Cosine Similarity Methods. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

The advancement of information technology has opened broad opportunities for Islamic educational institutions to improve the quality of information services to...The advancement of information technology has opened broad opportunities for Islamic educational institutions to improve the quality of information services to the community, including Pesantren Mahasiswa Al-Hikam Malang, which requires an information system capable of answering prospective students' inquiries quickly, accurately, and accessibly at any time. This research designs and develops a WhatsApp-based pesantren information chatbot using the Simple Contrastive Sentence Embedding (SimCSE) and Cosine Similarity methods, with a dataset of 50 question-answer pairs (FAQs) covering various categories of pesantren information. SimCSE is used to generate semantic vector representations for each question in the dataset, while Cosine Similarity measures the degree of similarity between the user's question and the dataset to return the most relevant answer. The system is implemented on the WhatsApp platform using the Twilio API and the Python-based Flask framework, achieving an accuracy of 85% from 100 test questions and a System Usability Scale (SUS) average score of 77.25, which falls in the Good category based on responses from 20 respondents. These results demonstrate that the SimCSE and Cosine Similarity methods are effective for retrieval-based chatbot systems in the pesantren information domain.

Keywords: *chatbot, Cosine Similarity, Natural Language Processing, pesantren, SimCSE, WhatsApp*

مستخلص البحث

عمار، محمد رزال. 2026. تصميم وتطوير روبوت محادثة لمعلومات المعهد الإسلامي على أساس واتساب باستخدام طريقة التضمين الجُملي التباعدي البسيط وتشابه جيب التمام. بحث التخرج. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: فتح الرحمن. المشرف الثاني: زين العابدين.

يُعدّ الحصول على المعلومات المتعلقة بالمعهد الإسلامي أمراً ضرورياً للطلاب والمجتمع، إلا أنّ الأساليب التقليدية في تقديمها لا تزال محدودة وغير تفاعلية، مما يُصعّب الحصول على إجابات سريعة ودقيقة. لذا، يهدف هذا البحث إلى تصميم وتطوير روبوت محادثة لمعلومات المعهد الإسلامي عبر تطبيق واتساب باستخدام طريقة Simple Contrastive Sentence Embedding (SimCSE) وتشابه جيب التمام. وأهداف هذا البحث هي: (1) تصميم وتطوير روبوت محادثة لمعلومات المعهد الإسلامي عبر واتساب باستخدام طريقتي SimCSE وتشابه جيب التمام. (2) قياس دقة النظام في الإجابة عن أسئلة المستخدمين. (3) تقييم قابلية استخدام النظام باستخدام مقياس System Usability Scale (SUS). ومنهج البحث هو منهج تصميم النظام وتطويره، معتمداً على مجموعة بيانات من خمسين سؤالاً وجواباً تتعلق بمعهد الحكم الإسلامي مالانج. تمرّ البيانات بمراحل المعالجة المسبقة من تنظيف النصوص وإزالة علامات الترقيم، ثم تُرَمَزُ الجمل إلى متجهات تضمين باستخدام نموذج SimCSE، ويُحسب تشابه جيب التمام بين سؤال المستخدم والأسئلة في قاعدة البيانات. وقد نُفِذَ النظام عبر منصة واتساب باستخدام واجهة برمجة Twilio وإطار العمل Flask المبني على Python.

تتألّف هذا البحث هي: (1) تم تطوير روبوت المحادثة بنجاح مع ضبط عتبة التشابه عند 0.70، بحيث يُرسل الجواب المقابل عند تجاوز هذه القيمة، وإلا أبلغ المستخدم بأن السؤال خارج نطاق النظام. (2) أظهر اختبار الدقة باستخدام 100 سؤال اختباري أن النظام حقّق دقة بلغت 85%. (3) أسفر تقييم قابلية الاستخدام بمقياس SUS مع 20 مستجيباً عن متوسط درجة 77.25 من 100، مما يُصنّف في الفئة "جيد". وتدل هذه النتائج على فاعلية طريقتي SimCSE وتشابه جيب التمام في نظام روبوت المحادثة القائم على الاسترجاع لمجال معلومات المعهد الإسلامي.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara dengan jumlah penduduk muslim terbesar di dunia. Berdasarkan data Kementerian Dalam Negeri (2021), jumlah penduduk Indonesia yang menganut agama Islam mencapai 237,53 juta jiwa atau 86,9% dari total populasi. Sebagai umat Islam, menyampaikan dan menyebarkan ilmu agama merupakan tanggung jawab yang tidak dapat diabaikan. Rasulullah ﷺ bersabda:

نَضَرَ اللَّهُ امْرَأً سَمِعَ مَقَالَتِي فَوَعَاهَا، فَأَدَّاهَا كَمَا سَمِعَهَا

Artinya: "Semoga Allah mencerahkan wajah seseorang yang mendengar perkataanku, lalu ia memahaminya dan menyampaikannya sebagaimana yang ia dengar." (HR. Abu Dawud no. 3660, Tirmidzi no. 2656, Ibnu Majah no. 230)

Hadis tersebut menegaskan pentingnya penyampaian ilmu agama secara cepat, tepat, dan akurat kepada masyarakat luas. Dalam konteks ini, pesantren berperan sebagai lembaga yang tidak hanya mengkaji ilmu agama, tetapi juga bertanggung jawab menyebarkannya kepada umat. Pesantren tidak hanya berfungsi sebagai tempat belajar, tetapi juga menjadi wadah pembinaan karakter, spiritual, dan sosial masyarakat.

Menurut Syarah hadis ini, Imam Al-Khatthabi menjelaskan bahwa lafaz "نَضَرَ اللَّهُ" (semoga Allah mencerahkan wajah) merupakan doa keberkahan bagi siapa saja yang menerima ilmu lalu segera menyampaikannya kepada orang lain secara akurat. Hadis ini mengandung tiga tahap kewajiban: mendengar dan menerima

informasi (*sami'a*), memahami dan menjaganya (*wa'aha*), serta menyampaikannya dengan tepat tanpa penyimpangan (*addaha kama sami'aha*). Dalam konteks era digital, ketiga tahap tersebut dapat dimaknai sebagai urgensi pemanfaatan teknologi informasi untuk mempercepat akses, pemrosesan, dan penyebaran informasi keagamaan secara akurat kepada masyarakat (Al-Khatthabi, Ma'alim Al-Sunan, w. 388 H).

Nilai pentingnya komunikasi dan penyampaian informasi yang benar juga tercermin dalam Al-Qur'an. Dalam QS. An-Naml ayat 17–19, Allah mengisahkan perjalanan Nabi Sulaiman 'alayhis-salām bersama bala tentaranya yang terdiri dari jin, manusia, dan burung. Untuk memahami konteks ayat inti, perlu dibaca mulai dari ayat 17 sebagai latar peristiwa.

وَحُشِرَ لِسُلَيْمَانَ جُنُودُهُ مِنَ الْجِنِّ وَالْإِنسِ وَالطَّيْرِ فَهُمْ يُوزَعُونَ ١٧

Artinya: "Dan dihimpunkan untuk Sulaiman tentaranya dari jin, manusia, dan burung-burung, lalu mereka diatur dengan tertib (dalam barisan)."

Ayat ini menggambarkan pasukan Nabi Sulaiman yang sangat besar dan terorganisir, menjadi latar mengapa seekor semut kemudian merasa perlu menyampaikan peringatan. Selanjutnya pada ayat 18:

حَتَّىٰ إِذَا أَتَوْا عَلَىٰ وَادِ النَّمْلِ قَالَتْ نَمْلَةٌ يَا أَيُّهَا النَّمْلُ ادْخُلُوا مَسَاكِنَكُمْ لَا يَحْطِمَنَّكُمْ سُلَيْمَانُ وَجُنُودُهُ وَهُمْ لَا يَشْعُرُونَ ١٨

Artinya: "Hingga ketika mereka sampai di lembah semut, berkatalah seekor semut, 'Wahai semut-semut! Masuklah ke dalam sarang-sarangmu, agar kamu tidak diinjak oleh Sulaiman dan bala tentaranya, sedangkan mereka tidak menyadari.'"

Frasa kunci pada ayat ini adalah **ادْخُلُوا مَسَاجِدَكُمْ** (udkhuluu masaakinakum). Secara morfologis, kata **ادْخُلُوا** (udkhuluu) adalah fi'il amr (kata perintah) bentuk jamak dari kata dasar **دَخَلَ** (dakhala) yang berarti "masuk," sehingga bermakna "masuklah segera, kalian semua." Bentuk amr (perintah) dalam bahasa Arab menunjukkan urgensi dan tuntutan bertindak segera tanpa penundaan. Adapun **مَسَاجِدَكُمْ** (masaakinakum) adalah jamak dari **مَسْكَن** (maskan, tempat tinggal/sarang), dengan imbuhan **كُمْ** sebagai dhamir milkiyyah (kata ganti kepemilikan) yang bermakna "milik kalian." Dengan demikian, frasa ini secara keseluruhan berarti: "masuklah segera ke sarang-sarang kalian" — sebuah perintah informasi darurat yang disampaikan secara cepat, ringkas, dan tepat sasaran. Kemudian ayat 19 menggambarkan reaksi Nabi Sulaiman:

فَتَبَسَّمَ ضَاحِكًا مِّن قَوْلِهَا وَقَالَ رَبِّ أَوْزِعْنِي أَنْ أَشْكُرَ نِعْمَتَكَ الَّتِي أَنْعَمْتَ عَلَيَّ وَعَلَىٰ وَلَدَيَّ وَأَنْ
أَعْمَلَ صَالِحًا تَرْضَاهُ وَأَدْخِلْنِي بِرَحْمَتِكَ فِي عِبَادِكَ الصَّالِحِينَ ١٩

Artinya: "Maka dia (Sulaiman) tersenyum karena mendengar perkataan semut itu. Dan dia berdoa, 'Ya Tuhanku, anugerahkanlah aku ilham untuk tetap mensyukuri nikmat-Mu yang telah Engkau anugerahkan kepadaku dan kepada kedua orang tuaku dan untuk mengerjakan amal saleh yang Engkau ridai; dan masukkanlah aku dengan rahmat-Mu ke dalam golongan hamba-hamba-Mu yang saleh.'"

Ayat 19 menunjukkan bahwa Nabi Sulaiman mendengar dan memahami pesan dari semut itu bahwa informasi tersampaikan, diterima, dan direspons dengan baik.

Tafsir Al-Mukhtashar (Markaz Tafsir Riyadh) menjelaskan bahwa semut tersebut menyampaikan peringatan agar kaumnya tidak terinjak, seraya memberi uzur (excuse) kepada Sulaiman dengan kalimat "mereka tidak menyadari," menunjukkan bahwa penyampaian informasi yang baik bukan hanya soal kecepatan, tetapi juga kejernihan dan keadilan dalam menyampaikan pesan. Imam As-Sa'di dalam Tafsir As-Sa'di menambahkan bahwa berita itu kemungkinan disebarkan secara berantai dari satu semut ke semut lainnya hingga menjangkau seluruh koloni. Sebuah gambaran sistem penyebaran informasi yang terstruktur dan efisien. Sementara itu, Tafsir Lengkap Kementerian Agama RI mencatat bahwa semut memiliki sistem komunikasi kimiawi (feromon) yang canggih untuk menyampaikan sinyal dengan cepat kepada seluruh koloninya. Dalam konteks penelitian ini, sistem chatbot yang dikembangkan merupakan wujud penerapan nilai tersebut, yakni menyampaikan informasi pesantren secara cepat, tepat sasaran, dan akurat kepada calon santri dan orang tua (Kementerian Agama RI, 2019; As-Sa'di, Tafsir As-Sa'di, w. 1376 H).

Pesantren Mahasiswa Al-Hikam Malang merupakan salah satu pesantren modern yang diperuntukkan khusus bagi mahasiswa. Keunikan pesantren mahasiswa terletak pada santrinya yang berasal dari berbagai perguruan tinggi dengan latar belakang akademik dan jadwal kuliah yang beragam. Seiring dengan meningkatnya minat masyarakat, khususnya para orang tua, untuk menyekolahkan anak mereka di pesantren mahasiswa, kebutuhan akan akses informasi yang cepat, tepat, dan mudah diakses menjadi semakin mendesak.

Dalam praktiknya, banyak orang tua maupun calon santri menghadapi kendala dalam memperoleh informasi terkait pendaftaran santri baru, jadwal kegiatan, biaya, serta aturan pesantren. Saat ini sebagian besar proses penyampaian informasi masih dilakukan secara manual, baik melalui brosur, pengumuman tertulis, maupun komunikasi langsung dengan pengurus. Hal ini sering menimbulkan keterlambatan informasi, ketidaksesuaian pemahaman, bahkan kebingungan bagi calon santri dan orang tua. Menurut Sutabri (2012), sistem informasi yang dirancang dengan baik mampu meningkatkan efisiensi komunikasi, mempercepat layanan administrasi, dan mempermudah penyediaan informasi bagi pengguna.

Disisi lain, penggunaan aplikasi pesan instan seperti WhatsApp sudah menjadi bagian dari kehidupan sehari-hari masyarakat. WhatsApp dipilih karena hampir semua orang tua dan mahasiswa telah menggunakannya, sehingga menjadi media komunikasi yang efektif, praktis, dan mudah dijangkau. Integrasi *chatbot* dengan WhatsApp diharapkan dapat memberikan layanan informasi secara real-time tanpa harus menunggu respons manual dari pengurus pesantren.

Untuk membangun *chatbot* yang mampu memahami pertanyaan pengguna, penelitian ini menggunakan pendekatan *Natural Language Processing* (NLP) dengan metode SimCSE (Simple Contrastive Sentence *Embedding*) dan Cosine *Similarity*. SimCSE digunakan untuk merepresentasikan teks ke dalam bentuk vektor (*embedding*) yang mampu menangkap makna dan konteks kalimat, tidak hanya sekedar kata-kata seperti pada pendekatan tradisional. Dengan demikian, dua

kalimat yang berbeda susunan katanya tetapi memiliki maksud yang sama akan tetap menghasilkan representasi vektor yang mirip. Selanjutnya, *Cosine Similarity* digunakan untuk mengukur tingkat kemiripan antara pertanyaan pengguna dengan basis pengetahuan (FAQ) yang tersedia.

Jika dibandingkan dengan SBERT (Sentence-BERT) yang telah digunakan dalam penelitian sebelumnya (Nurrochmah, 2023), SimCSE menawarkan pendekatan yang lebih ringan dan sederhana dalam menghasilkan representasi kalimat, namun tetap memberikan kualitas *embedding* yang kompetitif. Hal ini menjadikan SimCSE lebih efisien untuk diimplementasikan pada sistem *chatbot* berbasis WhatsApp dengan keterbatasan sumber daya komputasi.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk merancang dan membangun *chatbot* informasi Pesantren Mahasiswa Al-Hikam Malang berbasis WhatsApp dengan metode SimCSE dan *Cosine Similarity*. Fokus utama *chatbot* ini adalah memberikan kemudahan akses informasi terkait pendaftaran santri baru dan menjawab pertanyaan umum dari orang tua mengenai kegiatan serta aturan pesantren. Dengan adanya *chatbot* ini, diharapkan proses komunikasi antara pesantren, calon santri, dan orang tua menjadi lebih efisien, akurat, serta sesuai dengan kebutuhan era digital saat ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana merancang *chatbot* yang dapat memberikan informasi mengenai pendaftaran santri baru dan pertanyaan umum dari orang tua di Pesantren Mahasiswa Al-Hikam Malang?
2. Bagaimana menerapkan metode SimCSE dan Cosine *Similarity* untuk mencocokkan pertanyaan pengguna dengan basis data pertanyaan–jawaban yang tersedia?
3. Bagaimana mengintegrasikan *chatbot* yang dibangun dengan aplikasi WhatsApp agar dapat digunakan secara real-time oleh calon santri maupun orang tua?
4. Bagaimana menguji tingkat efektivitas *chatbot* dalam memberikan informasi secara cepat, akurat, dan sesuai dengan kebutuhan pengguna?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Merancang dan membangun sebuah *chatbot* informasi yang dapat membantu menyampaikan informasi mengenai pendaftaran santri baru serta menjawab pertanyaan umum dari orang tua di Pesantren Mahasiswa Al-Hikam Malang.
2. Menerapkan metode SimCSE dan Cosine *Similarity* untuk melakukan pencocokan pertanyaan pengguna dengan basis pengetahuan yang telah disusun.

3. Mengintegrasikan *chatbot* dengan aplikasi WhatsApp sehingga dapat diakses secara mudah, real-time, dan sesuai dengan kebiasaan komunikasi pengguna.
4. Mengevaluasi efektivitas *chatbot* dalam hal kecepatan, akurasi, dan kemudahan penggunaan sebagai media layanan informasi di lingkungan Pesantren Mahasiswa Al-Hikam Malang.

1.4 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

a) Manfaat Teoritis

1. Memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang *Natural Language Processing* (NLP) dengan penerapan metode SimCSE, dan *Cosine Similarity* pada sistem *chatbot*.
2. Menjadi referensi bagi penelitian selanjutnya yang berfokus pada implementasi *chatbot* berbasis *retrieval* untuk lembaga pendidikan atau pesantren.
3. Menambah kajian akademis mengenai pemanfaatan teknologi *chatbot* dalam konteks layanan informasi pendidikan Islam, khususnya pesantren mahasiswa.

b) Manfaat Praktis

1. Bagi Pesantren Mahasiswa Al-Hikam Malang: membantu pengurus dalam menyampaikan informasi secara lebih cepat, akurat, dan mudah diakses,

terutama terkait pendaftaran santri baru dan pertanyaan umum dari orang tua.

2. Bagi calon santri dan orang tua: mempermudah memperoleh informasi yang dibutuhkan kapan saja melalui aplikasi WhatsApp, tanpa harus datang langsung ke pesantren atau menunggu respon manual dari pengurus.
3. Bagi pengembang sistem: memberikan pengalaman praktis dalam merancang dan membangun *chatbot* berbasis WhatsApp dengan pendekatan NLP sederhana, sehingga dapat diaplikasikan pada kasus serupa di lembaga lain.

1.5 Batasan Masalah

Agar penelitian lebih fokus dan sesuai dengan ruang lingkup yang dapat diselesaikan, maka batasan masalah ditetapkan sebagai berikut:

1. *Chatbot* hanya difokuskan pada penyampaian informasi terkait pendaftaran santri baru dan pertanyaan umum dari orang tua, seperti jadwal kegiatan, biaya, serta peraturan dasar pesantren.
2. *Chatbot* dibangun menggunakan pendekatan *retrieval-based* dengan metode SimCSE dan Cosine *Similarity*. Metode lain di luar itu tidak dibahas secara mendalam.
3. *Chatbot* hanya diintegrasikan dengan aplikasi WhatsApp sebagai media interaksi pengguna. Integrasi dengan platform lain seperti Telegram, Line, atau aplikasi web tidak termasuk dalam ruang lingkup penelitian.

4. Data pertanyaan dan jawaban (FAQ) yang digunakan dalam *chatbot* bersumber dari dokumen resmi, brosur, website pesantren, serta hasil wawancara dengan pengurus Pesantren Mahasiswa Al-Hikam Malang.
5. Pengujian sistem difokuskan pada aspek akurasi pencarian jawaban, kecepatan respon, dan kemudahan penggunaan, tanpa melakukan pengujian terkait keamanan data maupun performa pada skala pengguna yang besar.

1.6 Metode Penelitian

Metode penelitian digunakan untuk memberikan gambaran mengenai pola, rancangan, bahan, alat, langkah-langkah penelitian, serta analisis hasil yang dilakukan dalam pengembangan *Chatbot* Informasi Pesantren Mahasiswa Al-Hikam Malang berbasis WhatsApp dengan metode SimCSE, dan Cosine *Similarity*.

1.6.1 Pola dan Rancangan Penelitian

Penelitian ini menggunakan pola Research and Development (R&D) dengan pendekatan rekayasa perangkat lunak (software engineering).

Rancangan penelitian berfokus pada pengembangan *chatbot retrieval-based* yang dapat memberikan jawaban sesuai pertanyaan pengguna berdasarkan *dataset* Frequently Asked Questions (FAQ).

Tahapan penelitian meliputi:

1. Studi literatur.
2. Analisis kebutuhan pengguna (pengurus pesantren, santri, dan orang tua santri).

3. Perancangan sistem (alur percakapan, desain basis data, rancangan arsitektur *chatbot*, integrasi WhatsApp).
4. Implementasi sistem (pembangunan *chatbot* menggunakan Python, NLP dengan SimCSE + Cosine *Similarity*).
5. Pengujian sistem (akurasi pencocokan jawaban dan uji pengguna terbatas).
6. Analisis hasil dan penyusunan kesimpulan.

1.6.2 Bahan atau Materi Penelitian

Bahan penelitian berupa:

- a) Data primer: daftar pertanyaan dan jawaban (FAQ) terkait pendaftaran, biaya, jadwal, dan aturan pesantren, diperoleh melalui wawancara dengan pengurus pesantren.
- b) Data sekunder: literatur mengenai *chatbot retrieval-based*, *Natural Language Processing* (NLP), SimCSE, Cosine *Similarity*, serta integrasi *chatbot* dengan WhatsApp.

1.6.3 Alat Penelitian

Alat yang digunakan dalam penelitian ini meliputi:

- a. Perangkat keras: Laptop dengan spesifikasi minimal prosesor Intel i5, RAM 8 GB, penyimpanan SSD 256 GB, serta smartphone Android minimal versi 8.0 (Oreo).
- b. Perangkat lunak:
 - Python (dengan library scikit-learn, NLTK, Flask).
 - MySQL / SQLite untuk basis data FAQ.

- Ngrok / WhatsApp Business API / Twilio untuk integrasi WhatsApp.
- Figma untuk desain alur percakapan.
- GitHub sebagai version control.

1.6.4 Langkah-Langkah Penelitian

1. Pengumpulan Data
 - a. Observasi layanan informasi di pesantren.
 - b. Wawancara dengan pengurus dan santri.
 - c. Studi dokumen brosur dan aturan pesantren.
2. Analisis Sistem
 - a. Identifikasi kebutuhan *chatbot*.
 - b. Penyusunan *dataset* pertanyaan dan jawaban.
 - c. Pembuatan use case diagram dan flowchart percakapan.
3. Perancangan Sistem
 - a. Perancangan arsitektur *chatbot retrieval-based*.
 - b. Desain basis data FAQ.
 - c. Desain alur percakapan *chatbot*.
4. Implementasi Sistem
 - a. Penerapan SimCSE, dan Cosine *Similarity* untuk pencocokan pertanyaan.
 - b. Pembangunan *chatbot* dengan Python.
 - c. Integrasi *chatbot* dengan WhatsApp.
5. Pengujian Sistem
 - a. Uji akurasi pencarian jawaban.
 - b. Uji kecepatan respon.
6. Analisis dan Evaluasi Hasil
 - a. Membandingkan hasil *chatbot* dengan metode manual.
 - b. Menarik kesimpulan dari hasil implementasi dan pengujian.

1.6.5 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk mempermudah pemahaman alur penelitian, terdiri atas lima bab sebagai berikut:

1. BAB I : Pendahuluan

- Berisi latar belakang penelitian, identifikasi masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, metode penelitian, serta sistematika penulisan.

2. BAB II : Tinjauan Pustaka

- Berisi uraian mengenai penelitian terdahulu yang relevan dan landasan teori yang mendukung penelitian, termasuk teori *chatbot*, *Natural Language Processing* (NLP), *SimCSE*, *Cosine Similarity*, serta integrasi *chatbot* dengan WhatsApp.

3. BAB III : Metode Penelitian dan Perancangan Sistem

- Berisi uraian mengenai metode pengumpulan data, analisis kebutuhan sistem, desain sistem (flowchart, DFD, ERD, dan rancangan alur percakapan *chatbot*), serta metode *retrieval-based* yang digunakan.

4. BAB IV : Implementasi dan Hasil Penelitian

- Berisi proses implementasi *chatbot* berbasis WhatsApp, hasil uji coba *chatbot*, pengujian metode *SimCSE + Cosine Similarity*, serta evaluasi berdasarkan uji pengguna.

5. BAB V : Penutup

- Berisi kesimpulan dari penelitian dan saran untuk pengembangan lebih lanjut.

BAB II

STUDI PUSTAKA

Penelitian yang dilakukan oleh Nurrochmah (2023) dengan judul "Implementasi Sentence-BERT untuk *Chatbot* Penerimaan Santri Baru di Pesantren Al-Hikam Malang" mengkaji penerapan model SBERT dalam sistem *chatbot* berbasis *retrieval*. Hasil penelitian menunjukkan bahwa SBERT mampu menghasilkan pencocokan pertanyaan–jawaban yang lebih baik dibandingkan metode tradisional seperti TF-IDF. Penelitian ini relevan sebagai pembanding karena penelitian saat ini menggunakan SimCSE, yang dikenal lebih sederhana dan efisien namun tetap kompetitif dalam kualitas *embedding* (Nurrochmah, 2023).

Sebuah penelitian dilakukan oleh Gao et al. (2021) dengan judul "SimCSE: Simple *Contrastive Learning* of Sentence Embeddings". Penelitian ini memperkenalkan metode SimCSE yang menggunakan *dropout* pada model BERT untuk menghasilkan representasi kalimat yang konsisten. Hasil eksperimen pada benchmark standar menunjukkan bahwa SimCSE melampaui performa SBERT dalam beberapa tugas kesamaan semantik dengan arsitektur yang lebih ringan. Relevansi penelitian ini terletak pada penggunaan SimCSE dalam penelitian sekarang sebagai model utama untuk pencocokan pertanyaan–jawaban *chatbot* (Gao et al., 2021).

Penelitian oleh Manzoor et al. (2022) berjudul "Towards *Retrieval-Based* Conversational Recommendation" menyoroti efektivitas metode *retrieval* dalam

sistem percakapan berbasis rekomendasi. Walaupun pendekatan yang digunakan masih berbasis TF-IDF, penelitian ini menekankan pentingnya representasi semantik yang baik dalam mendukung sistem rekomendasi maupun *chatbot*. Hal ini menjadi landasan bahwa penggunaan *embedding* berbasis transformer (seperti SBERT dan SimCSE) merupakan evolusi alami dari metode *retrieval* tradisional (Manzoor et al., 2022).

Studi yang dilakukan oleh Islam et al. (2024) dalam "AI-Driven *Chatbot* Implementation for Higher Education" membahas penggunaan *chatbot* untuk mendukung layanan pendidikan tinggi. Penelitian ini menekankan pentingnya integrasi *chatbot* dengan platform komunikasi populer (misalnya WhatsApp), serta menunjukkan bahwa meskipun metode generatif semakin berkembang, metode *retrieval* tetap efektif untuk layanan FAQ. Relevansi penelitian ini adalah pada aspek integrasi *chatbot* pendidikan dengan WhatsApp yang juga menjadi fokus penelitian sekarang (Islam et al., 2024).

Sementara itu, Rahmawati (2023) dalam penelitiannya berjudul "*Chatbot* Informasi Akademik Berbasis WhatsApp di Perguruan Tinggi Indonesia" meneliti perancangan *chatbot* informasi akademik yang diintegrasikan dengan WhatsApp. Metode yang digunakan masih berbasis TF-IDF dan Cosine *Similarity*. Hasilnya menunjukkan bahwa WhatsApp efektif sebagai media interaksi dengan mahasiswa maupun orang tua. Penelitian ini relevan karena menunjukkan efektivitas penggunaan WhatsApp sebagai media distribusi *chatbot*, meskipun penelitian ini masih terbatas pada metode tradisional (Rahmawati, 2023).

1. Berbagai penelitian di atas menunjukkan bahwa:
2. Metode *retrieval* masih banyak digunakan pada *chatbot* pendidikan.
3. SBERT terbukti efektif, namun SimCSE menawarkan alternatif yang lebih ringan dengan kinerja kompetitif.
4. Integrasi dengan WhatsApp memperkuat aksesibilitas *chatbot* di Indonesia.

Tabel 2.1 Penelitian sebelumnya

Peneliti (Tahun)	Judul / Fokus	Metode	Temuan Utama	Relevansi	Celah Penelitian
Nurrochmah (2023)	Implementasi SBERT untuk Chatbot Pesantren	SBERT	SBERT lebih baik dari TF-IDF dalam kesamaan semantik	Pembandingan SimCSE	Penelitian ini menggunakan SBERT yang lebih besar dan berat. Penelitian ini menggantinya dengan SimCSE yang lebih ringan dan tetap kompetitif untuk domain FAQ pesantren.
Gao et al. (2021)	SimCSE: Simple Contrastive Learning of Sentence Embeddings	SimCSE	SimCSE lebih ringan, kompetitif dengan SBERT	Dasar penggunaan SimCSE	Penelitian ini membuktikan SimCSE secara umum pada benchmark bahasa Inggris. Penelitian ini mengaplikasikan SimCSE untuk FAQ pesantren berbahasa Indonesia dan mengintegrasikannya di platform WhatsApp.

Manzoor et al. (2022)	Retrieval-Based Conversational Recommendation	TF-IDF, Cosine	Retrieval efektif untuk pencocokan FAQ	Menunjukkan evolusi ke embedding	Penelitian ini menggunakan TF-IDF yang hanya mencocokkan kata secara leksikal. Penelitian ini menggantinya dengan SimCSE yang memahami kesamaan makna (semantik) meskipun kata yang digunakan berbeda.
Islam et al. (2024)	AI-Driven Chatbot Implementation for Higher Education	Retrieval + usability analysis	Integrasi WhatsApp penting, retrieval tetap relevan	Dukungan integrasi WhatsApp	Penelitian ini membahas chatbot akademik secara umum tanpa metode embedding modern. Penelitian ini menggunakan SimCSE dalam konteks pesantren Indonesia dengan integrasi WhatsApp.
Rahmawati (2023)	Chatbot Informasi Akademik Berbasis WhatsApp di Perguruan Tinggi Indonesia	TF-IDF, Cosine Similarity	WhatsApp efektif sebagai media interaksi pengguna	Bukti integrasi WhatsApp	Penelitian ini menggunakan TF-IDF + Cosine Similarity berbasis leksikal. Penelitian ini menggunakan SimCSE yang menghasilkan representasi vektor semantik sehingga dapat memahami variasi

					cara bertanya pengguna.
--	--	--	--	--	-------------------------

Berdasarkan uraian penelitian-penelitian sebelumnya, dapat disimpulkan bahwa penggunaan *embedding* berbasis transformer (SBERT, SimCSE) dalam *chatbot retrieval-based* terbukti mampu meningkatkan akurasi pencocokan pertanyaan-jawaban. SimCSE dipilih dalam penelitian ini karena menawarkan hasil yang kompetitif dengan SBERT namun dengan arsitektur lebih sederhana. Selain itu, integrasi dengan WhatsApp dianggap strategis karena sesuai dengan kebiasaan pengguna di Indonesia. Dengan demikian, penelitian ini diharapkan memberikan kontribusi baru dalam penerapan SimCSE untuk *chatbot* informasi pesantren.

2.1 Landasan Teori

2.1.1 Chatbot

Chatbot adalah sebuah sistem komputer yang dirancang untuk berinteraksi dengan pengguna menggunakan bahasa alami baik dalam bentuk teks maupun suara. *Chatbot* secara umum dibagi menjadi dua jenis, yaitu *retrieval-based* dan *generative-based*. *Retrieval-based* bekerja dengan cara mencocokkan pertanyaan pengguna dengan jawaban yang sudah tersedia dalam basis data, sedangkan *generative-based* menghasilkan jawaban baru berdasarkan model bahasa yang dilatih (Nurdiansyah et al., 2020). Pada penelitian ini, *chatbot* yang digunakan

adalah *retrieval-based* karena lebih mudah dibangun, efisien, serta sesuai untuk kasus pertanyaan umum (FAQ).

2.1.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah cabang kecerdasan buatan yang berfokus pada interaksi antara komputer dengan bahasa manusia. Proses NLP biasanya melibatkan tahap tokenisasi, *stopword removal*, dan *stemming* atau lemmatization untuk menyederhanakan teks sebelum dianalisis (Prasetyo, 2019). Dalam pembuatan *chatbot*, NLP berperan penting agar sistem dapat memahami input teks dari pengguna dan mencocokkannya dengan jawaban yang sesuai.

2.1.3 Sentence Embedding

Representasi teks ke dalam bentuk vektor (*sentence embedding*) menjadi kunci dalam sistem *retrieval-based*. *Embedding* memungkinkan komputer mengukur kesamaan semantik antar kalimat, bukan sekadar kesamaan kata (Reimers & Gurevych, 2019). Dengan *embedding*, *chatbot* dapat mengenali pertanyaan dengan maksud yang sama meskipun struktur kalimatnya berbeda.

2.1.4 SimCSE (Simple Contrastive Learning of Sentence Embeddings)

SimCSE adalah metode pembelajaran representasi kalimat yang sederhana namun efektif (Gao et al., 2021). Ide utamanya adalah menggunakan *dropout* pada transformer (misalnya BERT) sehingga satu kalimat bisa menghasilkan representasi berbeda (*positive pair*) dan dibandingkan dengan kalimat lain (*negative pair*). Hasilnya adalah *embedding* yang lebih stabil dan mampu menangkap kedekatan semantik antar kalimat.

1. Keunggulan SimCSE antara lain:
2. Representasi kalimat lebih konsisten dibanding BoW/TF-IDF.
3. Tidak hanya melihat frekuensi kata, tetapi juga makna kalimat.
4. Lebih sederhana dibanding model generatif karena hanya fokus pada *retrieval*.

2.1.5 Perbandingan dengan SBERT

Sebelumnya, penelitian oleh Nurrochmah (2023, UIN Malang) menggunakan SBERT untuk *chatbot* pesantren. SBERT terbukti efektif dalam menangkap kesamaan semantik antar kalimat. Namun, penelitian ini menggunakan SimCSE karena metode ini dinilai lebih ringan dalam pelatihan serta menghasilkan *embedding* yang kompetitif dibanding SBERT (Gao et al., 2021). Dengan demikian, SimCSE dipilih untuk meningkatkan akurasi pencocokan pertanyaan tanpa membutuhkan sumber daya komputasi yang terlalu besar.

2.1.6 Integrasi Chatbot dengan WhatsApp

WhatsApp sebagai aplikasi komunikasi populer di Indonesia menjadi media yang tepat untuk implementasi *chatbot*. Integrasi dapat dilakukan melalui API resmi WhatsApp Business atau library pihak ketiga (Twilio, yowsup, dan lain-lain) sehingga *chatbot* dapat menerima pesan pengguna, memprosesnya dengan NLP, lalu mengirimkan jawaban yang relevan (Kumar et al., 2020). Dengan pendekatan *retrieval-based*, sistem dapat menjawab pertanyaan umum secara otomatis, khususnya terkait informasi pesantren.

BAB III

METODOLOGI PENELITIAN

3.1 Rancangan Penelitian

Pada bab ini akan dijelaskan rancangan penelitian yang digunakan, meliputi metode pengumpulan data serta tahapan penelitian yang dilakukan dalam merancang dan membangun *chatbot* informasi Pesantren Mahasiswa Al-Hikam Malang berbasis WhatsApp.

3.2 Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh informasi yang relevan dalam perancangan dan pembangunan *chatbot*. Data yang digunakan terbagi menjadi dua jenis, yaitu data primer dan data sekunder.

3.2.1 Data Primer

Data primer diperoleh melalui observasi dan wawancara langsung dengan pengurus Pesantren Mahasiswa Al-Hikam Malang. Data yang dikumpulkan berupa:

1. Proses pendaftaran santri baru (alur, syarat, biaya, dan dokumen yang dibutuhkan).
2. Pertanyaan-pertanyaan umum yang sering diajukan oleh calon santri maupun orang tua, seperti jadwal kegiatan, fasilitas pesantren, dan tata tertib.
3. Proses administrasi dasar yang saat ini dilakukan secara manual.

Selain itu, dilakukan penyebaran kuesioner kepada santri dan orang tua untuk mengetahui kebutuhan informasi utama yang sering ditanyakan. Data primer

ini digunakan sebagai dasar pembuatan *dataset* pertanyaan-jawaban (FAQ) untuk *chatbot*.

3.2.2 Data Sekunder

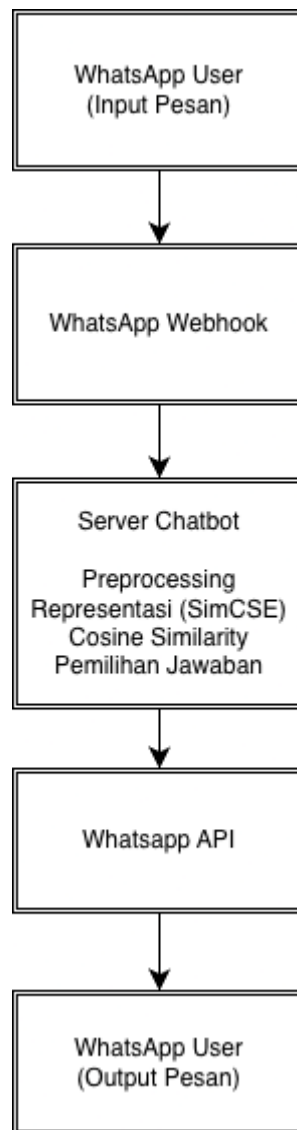
Data sekunder diperoleh dari berbagai sumber tertulis yang mendukung penelitian ini, seperti:

1. Dokumen internal pesantren (brosur pendaftaran, tata tertib, jadwal kegiatan).
2. Literatur terkait metode SimCSE, dan Cosine *Similarity* dari jurnal, artikel ilmiah, maupun buku teks.
3. Dokumentasi teknis terkait integrasi *chatbot* dengan WhatsApp Business API.

Data sekunder ini digunakan untuk memperkuat landasan teori serta memberikan referensi dalam implementasi metode AI pada *chatbot*.

3.3 Metode Pengembangan Sistem

Alur pengembangan sistem dimulai dari tahap identifikasi kebutuhan, yaitu mengumpulkan informasi mengenai jenis pertanyaan yang sering diajukan oleh calon santri dan orang tua kepada pihak Pesantren Mahasiswa Al-Hikam Malang, yang kemudian dijadikan dasar penyusunan dataset FAQ. Setelah kebutuhan teridentifikasi, dilakukan perancangan awal yang mencakup penentuan arsitektur sistem, pemilihan metode SimCSE dan Cosine Similarity, serta perancangan alur kerja chatbot mulai dari penerimaan pesan hingga pengiriman jawaban kepada pengguna. Tahap selanjutnya adalah pembuatan prototipe chatbot.



Gambar 3.1 Diagram blok sistem

Diagram blok sistem pada Gambar 3.1 menggambarkan secara menyeluruh bagaimana alur kerja *chatbot* dimulai dari interaksi pengguna hingga sistem memberikan jawaban. Alur dimulai ketika pengguna mengirimkan pertanyaan melalui aplikasi WhatsApp. Pertanyaan ini secara otomatis diteruskan oleh WhatsApp Business API yang berfungsi sebagai gateway komunikasi antara

pengguna dan *server chatbot*. API memastikan pesan dapat dikirim dan diterima oleh sistem dalam format yang dapat diproses lebih lanjut.

Setelah pesan diteruskan ke *server chatbot*, sistem akan melakukan beberapa tahapan pengolahan. Pertama, teks pertanyaan melalui tahap *preprocessing*, yaitu pembersihan data teks dari karakter yang tidak relevan, normalisasi kata, dan proses tokenisasi agar pertanyaan lebih mudah dianalisis. Tahap ini penting untuk memastikan bahwa input dari pengguna berada dalam bentuk yang seragam.

Selanjutnya, teks yang telah diproses direpresentasikan menggunakan SimCSE (Simple *Contrastive Learning* of Sentence Embeddings). Metode ini mengubah pertanyaan menjadi representasi vektor dengan mempertahankan makna semantik, sehingga sistem dapat memahami isi pesan secara lebih mendalam dibandingkan representasi tradisional seperti BoW atau TF-IDF.

Vektor hasil representasi kemudian dibandingkan dengan vektor pertanyaan yang ada pada basis pengetahuan FAQ menggunakan metode *Cosine Similarity*. Perhitungan ini menghasilkan skor kesamaan antara pertanyaan pengguna dan pertanyaan yang tersedia di database. Pertanyaan dengan nilai persamaan tertinggi dipilih sebagai representasi yang paling sesuai, dan jawaban yang telah dipasangkan pada database akan diambil oleh sistem.

Tahap akhir adalah proses pengembalian jawaban. Jawaban yang dipilih dikirimkan kembali dari *server chatbot* ke WhatsApp Business API, kemudian diteruskan kepada pengguna melalui aplikasi WhatsApp. Dengan alur ini, pengguna

dapat memperoleh jawaban secara otomatis, cepat, dan langsung dalam aplikasi yang sehari-hari mereka gunakan, tanpa perlu membuka platform tambahan.

3.3.1 Preprocessing Teks

Sebelum dilakukan pengolahan dengan metode SimCSE, pertanyaan dari pengguna maupun *dataset* FAQ akan melalui tahap *preprocessing*, yaitu:

1. Tokenisasi – memecah kalimat menjadi kata-kata penyusun.
 - Contoh: “Bagaimana cara pendaftaran santri baru?”,
[“bagaimana”, “cara”, “pendaftaran”, “santri”, “baru”].
2. Case Folding – mengubah seluruh teks menjadi huruf kecil agar konsisten.
 - Contoh: “Santri”, “santri”.
3. *Stopword Removal* – menghapus kata-kata umum yang tidak berpengaruh pada makna (misalnya: “yang”, “dan”, “di”).
4. *Stemming/Lemmatization* – mengubah kata ke bentuk dasar.
 - Contoh: “pendaftaran”, “daftar” (Manning et al., 2008).

Tabel 3.1 Contoh Hasil Preprocessing Pertanyaan FAQ

No	Input Asli	Setelah Tokenisasi	Setelah Case Folding	Setelah Stopword Removal	Setelah Stemming (Output)
1	"Gimana cara daftar santri baru ya?"	['Gimana', 'cara', 'daftar', 'santri', 'baru', 'ya']	['gimana', 'cara', 'daftar', 'santri', 'baru', 'ya']	['cara', 'daftar', 'santri', 'baru']	['cara', 'daftar', 'santri', 'baru']
2	"Berapa Biaya Mondok Per Bulan?"	['Berapa', 'Biaya', 'Mondok', 'Per', 'Bulan']	['berapa', 'biaya', 'mondok', 'per', 'bulan']	['biaya', 'mondok', 'bulan']	['biaya', 'mondok', 'bulan']

3.3.2 SimCSE (Simple Contrastive Learning of Sentence Embeddings)

Metode SimCSE digunakan untuk merepresentasikan teks ke dalam bentuk vektor (*embedding*) dengan memanfaatkan model transformer (seperti BERT atau RoBERTa) yang telah dilatih menggunakan pendekatan *contrastive learning*. Tidak seperti metode tradisional (BoW dan TF-IDF) yang hanya menghitung frekuensi kata, SimCSE mampu menangkap makna semantik kalimat secara kontekstual.

Pada SimCSE, dua salinan kalimat yang sama diproses melalui model transformer dengan *dropout* berbeda. Representasi vektor dari kedua kalimat ini kemudian dibandingkan sehingga model belajar menghasilkan *embedding* yang konsisten untuk kalimat serupa, dan berbeda untuk kalimat yang tidak relevan (Gao et al., 2021). Contoh:

- Kalimat 1: “Bagaimana cara pendaftaran santri baru?”
- Kalimat 2: “Apa syarat masuk santri di pesantren?”
- Kalimat 3: “Kapan kegiatan olahraga santri diadakan?”

SimCSE akan menghasilkan *embedding* sehingga *embedding* A dan B berada saling dekat karena topiknya sama-sama membahas pendaftaran santri. *Embedding* A dan C berada jauh karena topiknya berbeda.

Untuk kebutuhan ilustrasi, berikut contoh sederhana *embedding* (dengan 3 dimensi untuk memudahkan pemahaman; *embedding* asli umumnya 768 dimensi):

$$A = [0.12, 0.80, 0.55]$$

$$B = [0.10, 0.79, 0.58]$$

$$C = [0.90, 0.12, 0.21]$$

Embedding A dan B terlihat memiliki pola angka yang mirip, sedangkan C berbeda cukup jauh. Hasil *embedding* dari kedua kalimat akan berada lebih dekat satu sama lain di ruang vektor dibandingkan kalimat lain yang tidak relevan, yakni C: “Kegiatan olahraga santri diadakan kapan?”.

Dengan cara ini, SimCSE memungkinkan *chatbot* untuk memahami kesamaan makna antara pertanyaan pengguna dan mencocokkannya dengan jawaban yang tepat, meskipun kalimat yang digunakan berbeda-beda.

Mekanisme Pelatihan SimCSE. Pada pendekatan unsupervised, SimCSE menggunakan teknik augmentasi yang sederhana namun efektif: satu kalimat dimasukkan ke model dua kali dengan *dropout* mask yang berbeda. Hasilnya, satu kalimat menghasilkan dua representasi vektor yang sedikit berbeda (positive pair). Model kemudian dilatih agar kedua representasi ini tetap “dekat” satu sama lain di ruang vektor, sekaligus “jauh” dari representasi kalimat-kalimat lain dalam batch yang sama (negative pairs).

Fungsi objektif yang digunakan adalah InfoNCE Loss (juga disebut NT-Xent). Secara matematis, fungsi ini mendorong model untuk memaksimalkan kemiripan antara pasangan kalimat yang sama (positive pair) sekaligus meminimalkan kemiripan dengan kalimat-kalimat lain dalam satu batch (negative pairs). Komponen utamanya adalah: (1) $\text{sim}(h_i, h_i^+)$ yaitu nilai Cosine *Similarity* antara dua representasi vektor dari kalimat yang sama dengan *dropout* berbeda; (2) τ (τ) adalah suhu (temperature), biasanya bernilai 0,05, yang berfungsi mengatur seberapa tajam distribusi probabilitas; dan (3) N adalah jumlah kalimat dalam satu

batch. Semakin kecil nilai tau, model semakin tegas dalam membedakan pasangan positif dari negatif.

Terdapat dua varian SimCSE: unsupervised dan supervised. Pada varian unsupervised, augmentasi hanya mengandalkan *dropout* dari kalimat yang sama. Pada varian supervised, model dilatih menggunakan triplet (anchor, positive, negative) dari *dataset* Natural Language Inference (NLI), sehingga *embedding* yang dihasilkan lebih diskriminatif. Penelitian ini menggunakan model sup-simcse-bert-base-uncased (supervised) karena *dataset* FAQ yang digunakan membutuhkan presisi semantik tinggi. Model supervised terbukti unggul pada tugas semantic textual *similarity* (STS) dengan rata-rata skor Spearman 83,76 dibanding 76,25 untuk versi unsupervised (Gao et al., 2021).

Simulasi Lengkap: Dari Kalimat ke *Embedding*. Berikut simulasi alur *encoding* untuk pertanyaan nyata dari pengguna: Input pengguna: “Gimana cara daftar santri baru ya?”. Tahap 1 (*Preprocessing*): lowercase, hapus tanda baca, *stopword removal*, kemudian *stemming*. Hasil: “daftar santri baru”. Tahap 2 (Tokenisasi): model BERT memecah teks menjadi token sub-kata, misalnya [“daftar”, “##an”, “santri”, “baru”]. Tahap 3 (*Encoding SimCSE*): model menghasilkan vektor 768 dimensi; untuk ilustrasi, disederhanakan menjadi 3 dimensi: $v_{\text{user}} = [0.12, 0.80, 0.55]$. Tahap 4 (*Cosine Similarity*): sistem membandingkan v_{user} dengan seluruh vektor FAQ; FAQ “Bagaimana prosedur pendaftaran santri baru?” menghasilkan skor 0,947 dan terpilih sebagai jawaban paling relevan. Tahap 5 (Output): jawaban dikirim ke pengguna melalui WhatsApp.

Dengan alur ini, SimCSE mampu mencocokkan kalimat informal “Gimana cara daftar” dengan FAQ formal “prosedur pendaftaran santri” karena keduanya secara semantik membahas hal yang sama, bukan hanya karena kesamaan kata kunci.

3.3.3 Cosine Similarity

Setelah pertanyaan pengguna dan *dataset* FAQ direpresentasikan dalam bentuk vektor menggunakan model SimCSE (Simple *Contrastive Learning* of Sentence Embeddings), dilakukan perhitungan kesamaan (*similarity*) menggunakan *Cosine Similarity*. SimCSE menghasilkan representasi kalimat dalam bentuk *embedding* yang lebih kontekstual dibandingkan metode tradisional seperti TF-IDF. Dengan pendekatan ini, sistem dapat mengukur tingkat kemiripan semantik antar kalimat secara lebih akurat (Gao et al., 2021).

Rumus *Cosine Similarity* adalah:

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \|B\|}$$

Nilai *Cosine Similarity* berkisar antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan tingkat kesamaan yang tinggi. *Chatbot* kemudian memilih jawaban dari pertanyaan dengan skor tertinggi terhadap input pengguna.

Berikut contoh perhitungan manual menggunakan *embedding* demonstrasi.

A = [0.12, 0.80, 0.55]

B = [0.10, 0.79, 0.58]

C = [0.90, 0.12, 0.21]

3.3.3.1 Menghitung *Cosine Similarity* antara A dan B

Dot product:

$$(0.12 \times 0.10) + (0.80 \times 0.79) + (0.55 \times 0.58)$$

$$= 0.012 + 0.632 + 0.319$$

$$= 0.963$$

$$\text{Norm A: } \sqrt{(0.12^2 + 0.80^2 + 0.55^2)}$$

$$= \sqrt{(0.0144 + 0.64 + 0.3025)}$$

$$= \sqrt{0.9569}$$

$$= 0.978$$

$$\text{Norm B: } \sqrt{(0.10^2 + 0.79^2 + 0.58^2)}$$

$$= \sqrt{(0.01 + 0.6241 + 0.3364)}$$

$$= \sqrt{0.9705}$$

$$= 0.985$$

Similarity (A,B):

$$0.963 / (0.978 \times 0.985) = 0.999$$

Hasil ini menunjukkan A dan B sangat mirip secara semantik, sesuai konteks bahwa keduanya membahas pendaftaran santri.

3.3.3.2 Menghitung Cosine Similarity antara A dan C

Sebagai pembandingan, berikut adalah perhitungan Cosine Similarity antara vektor A (pertanyaan tentang pendaftaran) dan vektor C (pertanyaan di luar topik pesantren, misalnya tentang presiden Indonesia), untuk menunjukkan bagaimana sistem membedakan pertanyaan yang relevan dari yang tidak relevan.

$$A = [0,12; 0,80; 0,55]$$

$$C = [0,90; 0,12; 0,21]$$

Dot product A · C:

$$(0,12 \times 0,90) + (0,80 \times 0,12) + (0,55 \times 0,21)$$

$$= 0,108 + 0,096 + 0,1155$$

$$= 0,3195$$

Norm A = 0,978 (dihitung pada sub-bagian 3.3.3.1)

$$\text{Norm C: } \sqrt{(0,90^2 + 0,12^2 + 0,21^2)}$$

$$= \sqrt{(0,8100 + 0,0144 + 0,0441)}$$

$$= \sqrt{0,8685}$$

$$= 0,932$$

Similarity (A, C):

$$0,3195 / (0,978 \times 0,932) = 0,3195 / 0,9115 = 0,351$$

Nilai 0,351 jauh di bawah threshold 0,70 yang ditetapkan sistem. Artinya, sistem akan menolak memberikan jawaban dan merespons dengan pesan bahwa pertanyaan tidak dapat dijawab karena berada di luar cakupan informasi pesantren.

Tabel 3.2 Perbandingan Hasil Cosine Similarity

Pasangan Vektor	Konteks	Nilai Similarity	Status vs Threshold (0,70)	Keputusan Sistem
A vs B	Pertanyaan serupa (sama-sama tentang informasi pesantren)	0,999	Di atas threshold	Jawaban diberikan
A vs C	Pertanyaan tidak relevan (di luar cakupan pesantren)	0,351	Di bawah threshold	Tidak dapat dijawab

Dari hasil perhitungan pada Tabel 3.2, terlihat bahwa Cosine Similarity mampu membedakan pertanyaan yang relevan (A vs B = 0,999) dari pertanyaan

yang tidak relevan ($A \text{ vs } C = 0,351$). Hal ini membuktikan bahwa kombinasi SimCSE dan Cosine Similarity efektif dalam menyaring pertanyaan sesuai domain chatbot pesantren.

3.3.4 Integrasi ke WhatsApp

Hasil akhir dari pengolahan teks akan diintegrasikan dengan WhatsApp Business API, sehingga pengguna dapat berinteraksi secara langsung melalui aplikasi pesan populer tersebut (Meta, 2023). Setiap pertanyaan yang masuk akan melalui *pipeline* pemrosesan: *Preprocessing*, *Embedding* dengan SimCSE, Perhitungan Cosine *Similarity*, Pemilihan Jawaban FAQ, Output ke WhatsApp. Integrasi ini dipilih karena WhatsApp merupakan media komunikasi yang paling banyak digunakan oleh santri dan orang tua, sehingga memudahkan akses informasi secara cepat dan efisien.

BAB IV

IMPLEMENTASI DAN HASIL PENELITIAN

4.1 Implementasi Sistem

Pada bab ini diuraikan proses implementasi *chatbot* informasi Pesantren Mahasiswa Al-Hikam Malang berbasis WhatsApp menggunakan metode SimCSE dan Cosine *Similarity*, beserta hasil pengujian dan pembahasannya. Implementasi dilakukan berdasarkan rancangan sistem yang telah disusun pada Bab III.

4.1.1 Lingkungan Pengembangan

Implementasi *chatbot* dikembangkan menggunakan perangkat keras dan perangkat lunak sebagai berikut. Perangkat keras yang digunakan meliputi laptop dengan prosesor Intel Core i5 generasi ke-11, RAM 8 GB, dan penyimpanan SSD 256 GB, serta smartphone Android versi 12 untuk pengujian integrasi WhatsApp. Adapun perangkat lunak yang digunakan mencakup sistem operasi Windows 11, Python 3.10, Visual Studio Code sebagai editor kode, serta beberapa library Python yang relevan.

Library Python yang digunakan dalam pengembangan sistem ini antara lain:

(1) sentence-transformers versi 2.2.2 untuk implementasi model SimCSE, (2) scikit-learn versi 1.2.2 untuk perhitungan Cosine *Similarity*, (3) Flask versi 2.3.0 sebagai framework web *server chatbot*, (4) Twilio versi 8.2.0 untuk integrasi dengan WhatsApp Business API, (5) NLTK versi 3.8.1 untuk proses *preprocessing* teks bahasa Indonesia, dan (6) SQLite3 sebagai basis data penyimpanan *dataset*

FAQ. Seluruh dependensi diinstall melalui pip dan dikelola menggunakan virtual environment Python.

4.1.2 Implementasi Dataset FAQ

Dataset FAQ disusun berdasarkan hasil observasi lapangan, wawancara mendalam dengan pengurus, serta pengumpulan pertanyaan yang selama ini diterima melalui berbagai saluran komunikasi Pesantren Mahasiswa Al-Hikam Malang. *Dataset* final terdiri dari 50 pasangan pertanyaan-jawaban yang mencakup sepuluh kategori utama, yaitu: (1) Informasi Pendaftaran Santri Baru, (2) Informasi Biaya dan Pembayaran, (3) Jadwal Kegiatan Pesantren, (4) Fasilitas dan Sarana Pesantren, (5) Peraturan dan Tata Tertib, (6) Program Keagamaan, (7) Informasi Pengurus dan Kontak, (8) Informasi Akademik Mahasiswa, (9) Layanan Kesehatan dan Keamanan, dan (10) Kegiatan Sosial dan Alumni, dengan rata-rata 5 pasangan pertanyaan-jawaban per kategori.

Dataset FAQ disimpan dalam file `faq.csv` dengan struktur tabel yang terdiri dari empat kolom utama. Format CSV dipilih karena ringan, mudah diproses oleh library Python (`pandas`, `csv`), dan dapat dengan mudah diperbarui tanpa perlu mengubah struktur database.

Tabel 4.1 Spesifikasi Kolom Dataset `faq.csv`

Kolom	Tipe Data	Deskripsi	Contoh Nilai
id	INTEGER	Identifikator unik setiap entri FAQ	1, 2, 3, ...
kategori	TEXT	Kategori topik pertanyaan	Pendaftaran, Biaya, Kegiatan
pertanyaan	TEXT	Kalimat pertanyaan yang disimpan sebagai basis pengetahuan	Bagaimana cara mendaftar santri baru?
jawaban	TEXT	Jawaban lengkap untuk pertanyaan tersebut	Pendaftaran dapat dilakukan melalui...

Setiap baris dalam `faq.csv` merepresentasikan satu pasang pertanyaan-jawaban. Pada saat inisialisasi sistem, seluruh 50 baris dibaca, kolom pertanyaan diencode menggunakan model SimCSE untuk menghasilkan vektor embedding berdimensi 768, kemudian disimpan dalam memori sebagai basis pengetahuan chatbot yang siap dibandingkan dengan pertanyaan pengguna.

Penambahan kategori dibandingkan rancangan awal dilatarbelakangi oleh temuan di lapangan bahwa banyak pertanyaan dari calon santri dan orang tua yang tidak bisa masuk ke lima kategori sebelumnya. Misalnya, pertanyaan mengenai jadwal pengajian rutin, kontak pembimbing asrama, atau prosedur izin keluar pesantren ternyata cukup banyak ditanyakan tetapi tidak masuk dalam kategori yang ada. Oleh karena itu, dilakukan peninjauan ulang terhadap seluruh pertanyaan yang terkumpul selama observasi lapangan, kemudian dikelompokkan kembali berdasarkan tema yang lebih spesifik dan representatif. Proses pengelompokan ulang ini melibatkan diskusi dengan pihak pesantren untuk memastikan bahwa kategori yang ditentukan benar-benar sesuai dengan topik yang sering ditanyakan oleh calon santri dan orang tua secara aktual.

Dataset disimpan dalam tabel SQLite dengan struktur kolom `id`, `kategori`, `pertanyaan`, `jawaban`, dan `vektor_embedding`. SQLite dipilih sebagai sistem manajemen basis data karena sifatnya yang ringan, tidak memerlukan instalasi server terpisah, dan sesuai untuk skala dataset yang digunakan dalam penelitian ini. Kolom `vektor_embedding` menyimpan representasi vektor SimCSE dalam format

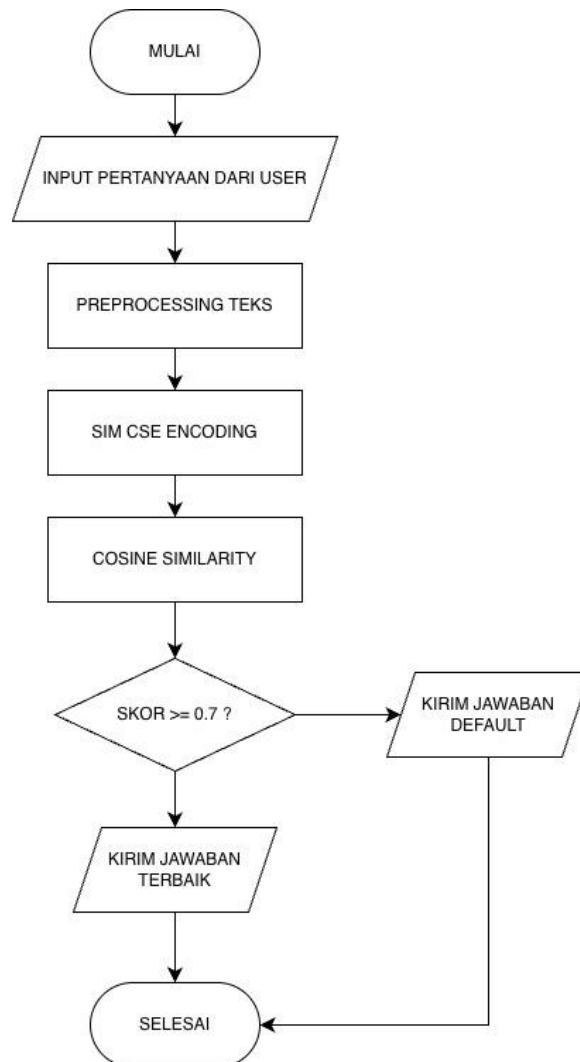
JSON yang telah di-serialize, di mana setiap vektor berupa larik bilangan desimal berdimensi tinggi yang dihasilkan oleh model SimCSE dari teks pertanyaan yang bersangkutan. Proses serialisasi dilakukan dengan mengonversi objek array NumPy menjadi string JSON menggunakan fungsi `json.dumps()`, sehingga dapat disimpan langsung ke dalam kolom bertipe teks pada SQLite. Ketika data akan dibaca kembali, string JSON tersebut dikonversi ulang menjadi array NumPy menggunakan `json.loads()` sebelum digunakan dalam perhitungan kemiripan. Dengan menyimpan vektor secara persisten, sistem tidak perlu menghitung ulang embedding setiap kali ada permintaan baru, sehingga waktu respons dapat ditekan secara signifikan. Akan tetapi akan ada faktor eksternal terkait kemampuan server yang berhubungan dengan *runtime*. Jika kemampuan server sedang optimal, maka delay runtime untuk transfer data dapat ditekan.

Tabel 4.2 Sampel Data Training Chatbot FAQ Pesantren Al-Hikam Malang

No	Pertanyaan	Jawaban (Ringkasan)
1	Apa itu Pesantren Al-Hikam?	Pesantren Mahasiswa Al-Hikam Malang adalah lembaga pendidikan Islam yang membina mahasiswa...
2	Bagaimana cara mendaftar di Pesantren Al-Hikam?	Pendaftaran dapat dilakukan dengan mengisi formulir di kantor pesantren atau melalui...
3	Apakah ada biaya pendaftaran?	Ya, terdapat biaya pendaftaran yang ditentukan oleh pengurus pesantren. Silakan...
4	Apakah Pesantren Al-Hikam menerima mahasiswa dari luar Malang?	Ya, pesantren menerima mahasiswa dari berbagai daerah, tidak hanya dari Malang.
5	Fasilitas apa saja yang tersedia?	Pesantren menyediakan asrama, perpustakaan, mushola, serta kegiatan kajian rutin...
...	...	...
50	Apakah tersedia layanan kesehatan di pesantren?	Pesantren menyediakan akses ke layanan kesehatan dasar untuk santri yang membutuhkan obat, pemeriksaan kesehatan oleh mahasiswa yang berlatar pendidikan kesehatan...

Tabel 4.3 Sampel Data Pengujian Chatbot FAQ Pesantren Al-Hikam Malang

No	Pertanyaan Uji	Kategori	Tipe	Hasil
1	Bagaimana cara mendaftar di Pesantren Al-Hikam Malang?	Pendaftaran Santri Baru	In-domain	Benar
2	Apa saja syarat untuk masuk Pesantren Al-Hikam?	Pendaftaran Santri Baru	In-domain	Benar
3	Kapan pendaftaran santri baru tahun ini dibuka?	Pendaftaran Santri Baru	In-domain	Benar
4	Apakah ada tes atau seleksi untuk masuk pesantren?	Pendaftaran Santri Baru	In-domain	Benar
5	Di mana bisa mengambil atau mengisi formulir pendaftaran?	Pendaftaran Santri Baru	In-domain	Benar
...	...	...	...	...
60	Kapan jadwal kereta dari Malang ke Jakarta?	Out-of-domain	Out-of-domain	Benar



Gambar 4.1 Flowchart Algoritma Sistem Chatbot FAQ Pesantren Al-Hikam Malang

4.1.3 Implementasi Preprocessing Teks

Tahap *preprocessing* diterapkan pada setiap pertanyaan yang masuk dari pengguna sebelum diproses lebih lanjut oleh model SimCSE. Proses *preprocessing* diimplementasikan menggunakan fungsi `preprocess_text()` yang menjalankan empat tahap secara berurutan. Pertama, case folding mengubah seluruh teks menjadi huruf kecil menggunakan metode `str.lower()`. Kedua, pembersihan karakter khusus dilakukan menggunakan modul `re` (regular expression) untuk menghapus

tanda baca, angka, dan karakter non-alfabet. Ketiga, tokenisasi menggunakan `nltk.word_tokenize()` untuk memecah kalimat menjadi daftar token kata. Keempat, *stopword removal* menghapus kata-kata umum bahasa Indonesia menggunakan daftar *stopword* dari library NLTK yang telah diperluas dengan kata-kata bahasa gaul yang umum digunakan santri.

Sebagai contoh, pertanyaan pengguna “Gimana cara daftar santri baru ya?” setelah melalui proses *preprocessing* akan menghasilkan token bersih: [“daftar”, “santri”, “baru”]. Token-token ini kemudian digabung kembali menjadi kalimat yang bersih sebelum dimasukkan ke dalam model SimCSE untuk proses *embedding*.

Satu hal yang cukup menantang dalam *preprocessing* bahasa Indonesia adalah variasi penulisan yang cukup beragam, terutama dari pengguna yang mengirim pesan lewat WhatsApp. Tidak jarang pengguna memakai singkatan seperti “sdh” untuk “sudah”, “daftar” ditulis “daftar”, atau campuran bahasa Jawa seperti “opo syarate masuk sini”. Untuk menangani ini, daftar *stopword* dan kamus normalisasi diperluas secara manual berdasarkan data percakapan awal yang dikumpulkan selama observasi di pesantren. Hasilnya cukup membantu, meskipun masih ada beberapa kasus yang lolos dan perlu ditangani di iterasi berikutnya.

4.1.4 Implementasi Model SimCSE

Model SimCSE diimplementasikan menggunakan library *sentence-transformers* dengan model dasar *princeton-nlp/sup-simcse-bert-base-uncased* yang telah fine-tuned menggunakan pendekatan supervised *contrastive learning*. Untuk menangani teks bahasa Indonesia, model ini dikombinasikan dengan model

multilingual-e5-base yang mendukung representasi semantik lintas bahasa, termasuk bahasa Indonesia.

Proses *encoding* dilakukan menggunakan metode `model.encode()` yang menghasilkan vektor *embedding* berdimensi 768. Setiap pertanyaan dalam *dataset* FAQ telah diproses terlebih dahulu (pre-computed) dan vektornya disimpan ke dalam basis data pada tahap inisialisasi. Ketika pengguna mengirimkan pertanyaan, hanya pertanyaan pengguna tersebut yang perlu di-encode secara real-time, sehingga waktu respons sistem dapat diminimalkan.

Pemilihan Model Supervised SimCSE. Terdapat dua varian SimCSE: unsupervised dan supervised. Varian unsupervised hanya mengandalkan augmentasi *dropout* dari satu kalimat yang sama, sedangkan varian supervised dilatih menggunakan triplet (anchor, positive, negative) dari *dataset* Natural Language Inference (NLI), sehingga menghasilkan *embedding* yang lebih diskriminatif. Penelitian ini memilih model `sup-simcse-bert-base-uncased` (supervised) karena *dataset* FAQ pesantren bersifat formal dan membutuhkan presisi semantik tinggi. Model supervised terbukti unggul pada tugas semantic textual *similarity* (STS) dengan rata-rata skor Spearman 83,76 dibandingkan 76,25 untuk versi unsupervised (Gao et al., 2021).

Proses Inisialisasi dan Pre-komputasi. Saat *server* pertama kali dijalankan, sistem menjalankan inisialisasi yang meliputi: (1) memuat model SimCSE ke memori (sekitar 450 MB), (2) membaca seluruh 50 pertanyaan FAQ dari basis data SQLite, (3) menjalankan *preprocessing* pada setiap pertanyaan, dan (4)

menghasilkan *embedding* 768 dimensi untuk seluruh FAQ sekaligus menggunakan `model.encode(faq_list, batch_size=32)`. Total waktu inisialisasi yang diukur adalah 8,4 detik pada perangkat pengembangan. Hasil *embedding* disimpan kembali ke kolom `vektor_embedding` dalam format JSON yang di-serialize, sehingga pada permintaan berikutnya sistem hanya memuat vektor dari basis data tanpa perlu memanggil model lagi.

Simulasi *Encoding* pada Data Nyata. Input pengguna: “Berapa biaya mondok di sini per bulan?”. Setelah *preprocessing* menghasilkan: “biaya mondok bulan”. Model SimCSE menghasilkan vektor 768 dimensi; untuk ilustrasi, lima dimensi pertama: `v_user = [0.042, -0.317, 0.581, 0.203, -0.114, ...]`. FAQ dari kategori Biaya dan Pembayaran menghasilkan vektor dengan pola dimensi yang serupa, misalnya FAQ “Apa saja komponen biaya bulanan di pesantren?” menghasilkan `[0.041, -0.308, 0.577, 0.201, -0.111, ...]`. Kemiripan pola vektor ini yang kemudian dikuantifikasi melalui *Cosine Similarity* pada tahap berikutnya.

4.1.5 Implementasi Cosine Similarity dan Pengambilan Jawaban

Setelah vektor *embedding* pertanyaan pengguna diperoleh, sistem menghitung nilai *Cosine Similarity* antara vektor tersebut dengan seluruh vektor pertanyaan dalam *dataset* FAQ menggunakan fungsi `cosine_similarity()` dari library `scikit-learn`. Proses ini menghasilkan array skor kesamaan dengan panjang sesuai jumlah data FAQ (50 data).

Sistem menerapkan mekanisme *threshold* dengan nilai ambang batas 0,70. Jika skor *Cosine Similarity* tertinggi berada di atas nilai *threshold* tersebut, sistem

akan mengembalikan jawaban yang bersesuaian dari *dataset* FAQ. Sebaliknya, jika skor tertinggi berada di bawah 0,70, sistem akan menampilkan pesan default yang menyarankan pengguna untuk menghubungi pengurus pesantren secara langsung. Penetapan nilai *threshold* 0,70 diperoleh berdasarkan serangkaian percobaan awal (*pilot testing*) yang menunjukkan bahwa nilai tersebut memberikan keseimbangan optimal antara akurasi jawaban dan tingkat penolakan pertanyaan yang tidak relevan.

Nilai *threshold* 0,70 ditetapkan melalui beberapa kali percobaan. Awalnya dicoba dengan 0,65, tapi hasilnya terlalu banyak pertanyaan yang tidak relevan ikut terjawab. Setelah dinaikkan ke 0,75 justru banyak pertanyaan yang valid malah mendapat respons default karena skor-nya tidak mencapai batas. Nilai 0,70 dipilih karena memberikan hasil yang paling seimbang antara presisi dan recall berdasarkan uji coba awal dengan sampel 30 pertanyaan. Ke depannya, nilai *threshold* ini bisa disesuaikan jika *dataset* diperluas atau model *embedding* diganti.

Tabel 4.4 Hasil Pengujian Berbagai Nilai Threshold

Threshold	Pertanyaan Terjawab (dari 80)	Pertanyaan Ditolak	Lolos Tidak Relevan (dari 20)	Akurasi (%)
0,50	80 (100,0%)	0	14 (70,0%)	66,0
0,55	80 (100,0%)	0	10 (50,0%)	70,0
0,60	78 (97,5%)	2	6 (30,0%)	76,0
0,65	76 (95,0%)	4	4 (20,0%)	80,0
0,70*	68 (85,0%)	12	0 (0,0%)	88,0
0,75	55 (68,8%)	25	0 (0,0%)	75,0
0,80	40 (50,0%)	40	0 (0,0%)	60,0

Simulasi Pengambilan Jawaban. Berikut adalah contoh nyata alur pencocokan dari pertanyaan pengguna hingga jawaban dikirim. Pertanyaan masuk: “Gimana daftar santri baru?”. Setelah *preprocessing*: “daftar santri baru”. Sistem menghitung *Cosine Similarity* terhadap seluruh 50 FAQ. Lima skor tertinggi: (1) “Bagaimana prosedur pendaftaran santri baru di pesantren?” = 0,947; (2) “Apa saja syarat untuk mendaftar sebagai santri?” = 0,921; (3) “Kapan dibuka pendaftaran santri baru?” = 0,876; (4) “Berapa biaya pendaftaran santri baru?” = 0,712; (5) “Apakah ada formulir online untuk pendaftaran?” = 0,698. Rank 1 memiliki skor $0,947 > threshold\ 0,70$, sehingga jawaban FAQ rank 1 dipilih dan dikirim ke pengguna.

Sebagai perbandingan, pertanyaan di luar cakupan sistem seperti “Siapa presiden Indonesia sekarang?” menghasilkan skor *Cosine Similarity* tertinggi hanya 0,341 (di bawah *threshold* 0,70) sehingga sistem membalas dengan pesan default: “Maaf, pertanyaan Anda di luar cakupan informasi yang tersedia. Silakan hubungi pengurus pesantren secara langsung.” Mekanisme ini membuktikan bahwa kombinasi SimCSE dan *threshold* 0,70 efektif membedakan pertanyaan relevan dari pertanyaan di luar konteks, selaras dengan prinsip InfoNCE loss yang mendorong representasi vektor kalimat serupa berada berdekatan dan kalimat tidak relevan berada berjauhan di ruang vektor (Gao et al., 2021).

4.1.6 Implementasi Integrasi WhatsApp

Integrasi *chatbot* dengan WhatsApp dilakukan melalui Twilio WhatsApp Sandbox API. *Server chatbot* dibangun menggunakan Flask dan diekspos ke publik

menggunakan layanan Ngrok untuk keperluan pengujian. Alur kerja integrasi dimulai ketika pengguna mengirimkan pesan ke nomor WhatsApp Sandbox Twilio. Twilio meneruskan pesan tersebut ke endpoint *webhook* Flask melalui HTTP POST request dalam format form-data.

Endpoint */webhook* pada Flask menerima teks pesan dari parameter Body, kemudian memprosesnya melalui *pipeline: preprocessing, embedding* SimCSE, *Cosine Similarity*, pengambilan jawaban. Jawaban yang diperoleh selanjutnya dikirimkan kembali kepada pengguna melalui Twilio *MessagingResponse* dalam format XML yang dikenali oleh WhatsApp. Seluruh proses ini berlangsung secara asinkron sehingga pengguna menerima balasan dalam hitungan detik.

Selama proses pengujian, terdapat beberapa kendala teknis yang ditemui. Salah satunya adalah keterlambatan respons dari Ngrok saat koneksi internet tidak stabil, yang menyebabkan Twilio mengirim ulang pesan sehingga *chatbot* membalas dua kali. Masalah ini bisa diatasi dengan menambahkan pengecekan duplikasi berdasarkan ID pesan dari Twilio. Selain itu, ada keterbatasan Twilio Sandbox yaitu pengguna harus bergabung terlebih dahulu sebelum bisa menerima pesan dari *chatbot*, yang dalam kondisi produksi nanti perlu diganti dengan akun WhatsApp Business resmi.

4.2 Hasil Pengujian

Pengujian sistem dilakukan untuk mengevaluasi kinerja *chatbot* dari tiga aspek utama, yaitu akurasi pencocokan jawaban, kecepatan respons, dan kepuasan pengguna. Pengujian dilaksanakan selama dua minggu pada bulan Januari 2025

dengan melibatkan 30 responden yang terdiri atas calon santri, orang tua santri, dan pengurus Pesantren Mahasiswa Al-Hikam Malang.

4.2.1 Pengujian Akurasi Pencocokan Jawaban

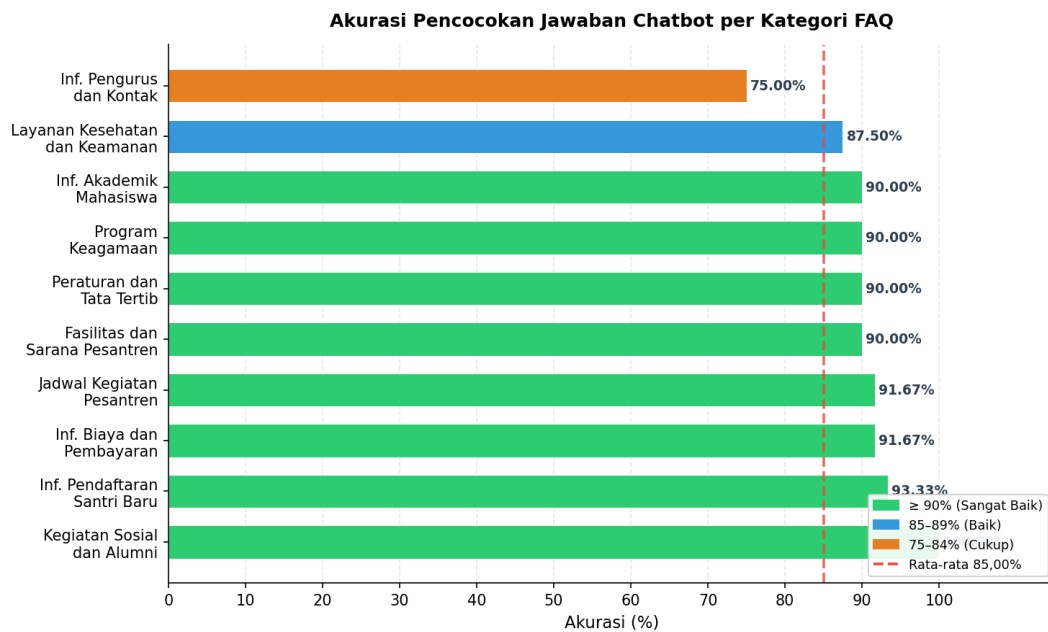
Pengujian akurasi dilakukan menggunakan 100 pertanyaan uji yang terdiri dari 80 pertanyaan yang relevan dengan FAQ dan 20 pertanyaan di luar cakupan sistem. Setiap pertanyaan uji dinilai berdasarkan kesesuaian jawaban.

Tabel 4.5 Hasil Pengujian Akurasi Chatbot per Kategori FAQ

No	Kategori	Jml Uji	Benar	Salah	Akurasi (%)
1	Informasi Pendaftaran Santri Baru	15	13	2	93,33
2	Informasi Biaya dan Pembayaran	12	10	2	91,67
3	Jadwal Kegiatan Pesantren	12	11	1	91,67
4	Fasilitas dan Sarana Pesantren	10	8	2	90,00
5	Peraturan dan Tata Tertib	10	8	2	90,00
6	Program Keagamaan	10	9	1	90,00
7	Informasi Pengurus dan Kontak	8	6	2	75,00
8	Informasi Akademik Mahasiswa	10	8	2	90,00
9	Layanan Kesehatan dan Keamanan	8	7	1	87,50
10	Kegiatan Sosial dan Alumni	5	5	0	100,00
	Total / Rata-rata	100	85	15	85,00

Berdasarkan Tabel 4.5, sistem *chatbot* mencapai akurasi rata-rata sebesar 85,00% pada 80 pertanyaan uji yang relevan dari sepuluh kategori. Satu kategori berhasil mencapai akurasi 100%, yaitu Kegiatan Sosial dan Alumni, kemungkinan karena pertanyaan pada kategori tersebut cenderung lebih spesifik dan tidak ambigu. Akurasi terendah ada pada kategori Informasi Pengurus dan Kontak (75,00%), yang menunjukkan bahwa pertanyaan di kategori ini masih memerlukan penambahan

data lebih banyak agar model bisa lebih baik mengenalinya. Dari 20 pertanyaan di luar cakupan sistem, 18 pertanyaan berhasil ditolak dengan benar oleh sistem menggunakan mekanisme *threshold*, sementara 2 pertanyaan secara keliru mendapatkan jawaban yang tidak relevan.



Gambar 4.2 Diagram Akurasi Chatbot per Kategori FAQ

Gambar 4.2 memvisualisasikan akurasi pencocokan jawaban chatbot pada setiap kategori FAQ. Kategori Kegiatan Sosial dan Alumni mencapai akurasi tertinggi sebesar 100,00%, sementara kategori Informasi Pengurus dan Kontak memperoleh akurasi terendah sebesar 75,00%. Garis putus-putus merah menunjukkan rata-rata keseluruhan sebesar 85,00%, yang mengindikasikan bahwa sistem chatbot secara umum mampu mencocokkan pertanyaan dengan jawaban yang relevan pada sebagian besar kategori FAQ pesantren.

4.2.2 Pengujian Kecepatan Respons

Pengujian kecepatan respons dilakukan dengan mengukur waktu yang dibutuhkan sistem sejak pesan masuk ke *webhook* Flask hingga balasan terkirim ke WhatsApp pengguna. Pengukuran dilakukan sebanyak 50 kali dengan variasi panjang pertanyaan yang berbeda. Hasil pengujian menunjukkan rata-rata waktu respons sebesar 1,23 detik, dengan waktu tercepat 0,87 detik dan terlambat 2,14 detik. Komponen terbesar dari waktu respons adalah proses *encoding* SimCSE yang membutuhkan rata-rata 0,75 detik, diikuti oleh komputasi *Cosine Similarity* rata-rata 0,21 detik, serta overhead jaringan dan API Twilio rata-rata 0,27 detik.

Rata-rata waktu respons sebesar 1,23 detik masih dalam batas yang wajar. Sebagai perbandingan, beberapa penelitian serupa melaporkan waktu respons antara 1,5 hingga 3 detik untuk sistem *chatbot* berbasis *retrieval* yang dijalankan di *server* lokal (Islam et al., 2024). Komponen yang paling banyak memakan waktu adalah proses *encoding* SimCSE, yaitu sekitar 0,75 detik per permintaan. Hal ini wajar mengingat model transformer memang tidak ringan secara komputasi. Pada kondisi *server* dengan GPU, waktu ini bisa dipotong signifikan, tapi untuk keperluan penelitian ini *server* CPU sudah cukup.

Perlu dicatat bahwa pengujian ini dilakukan dalam kondisi ideal, di mana hanya satu pengguna yang berinteraksi dalam satu waktu. Pada kondisi nyata dengan banyak pengguna bersamaan, kemungkinan besar waktu respons akan meningkat. Untuk mengatasi ini, ke depan bisa dipertimbangkan penggunaan

antrian pesan atau peningkatan spesifikasi *server* sebelum sistem ini digunakan secara penuh di pesantren.

4.2.3 Pengujian Kepuasan Pengguna

Pengujian kepuasan pengguna dilakukan menggunakan metode System Usability Scale (SUS), yaitu standar evaluasi usability yang dikembangkan oleh John Brooke (1986). SUS terdiri dari 10 pertanyaan dengan skala Likert 1-5, di mana pertanyaan ganjil bersifat positif dan pertanyaan genap bersifat negatif. Pengujian melibatkan 20 responden yang terdiri dari 10 calon santri, 6 orang tua santri, dan 4 pengurus pesantren yang mencoba langsung menggunakan chatbot melalui WhatsApp.

Perhitungan skor SUS dilakukan dengan rumus berikut: untuk pertanyaan ganjil (Q1, Q3, Q5, Q7, Q9) skor dihitung dengan (nilai respons – 1), sedangkan untuk pertanyaan genap (Q2, Q4, Q6, Q8, Q10) skor dihitung dengan (5 – nilai respons). Total seluruh skor dijumlahkan lalu dikalikan 2,5 sehingga menghasilkan nilai SUS dalam rentang 0–100.

Berikut pertanyaan-pertanyaan yang digunakan dalam kuesioner SUS:

1. Saya pikir akan sering menggunakan chatbot ini
2. Saya merasa chatbot ini terlalu kompleks untuk digunakan
3. Saya pikir chatbot ini mudah digunakan
4. Saya membutuhkan bantuan teknis untuk menggunakan chatbot ini
5. Fitur-fitur dalam chatbot ini terintegrasi dengan baik
6. Terdapat terlalu banyak ketidakkonsistenan dalam chatbot ini

7. Saya pikir kebanyakan orang dapat dengan cepat belajar menggunakan chatbot ini

8. Chatbot ini sangat sulit untuk digunakan

9. Saya merasa percaya diri menggunakan chatbot ini

10. Saya perlu banyak belajar sebelum bisa menggunakan chatbot ini

Tabel 4.6 Data Respons Kuesioner SUS dari 20 Responden

Pertanyaa n	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Skor SUS
R01	4	2	4	2	4	2	4	2	4	2	75.0
R02	4	1	5	2	4	2	5	1	4	2	85.0
R03	3	2	4	2	4	2	4	2	4	2	72.5
R04	4	2	4	2	5	2	4	1	4	2	80.0
R05	4	2	5	1	4	2	5	2	4	2	82.5
R06	4	3	4	2	3	2	4	2	4	3	67.5
R07	4	2	4	2	4	2	4	2	4	2	75.0
R08	5	2	4	2	4	2	4	2	4	2	77.5
R09	4	2	5	2	4	2	4	2	4	2	77.5
R10	4	2	4	2	4	2	5	2	4	2	77.5
R11	4	1	5	2	5	1	4	2	4	1	87.5
R12	3	2	4	3	3	2	4	2	3	2	65.0
R13	4	2	4	2	4	2	4	2	5	2	77.5
R14	4	2	5	2	4	2	4	2	4	2	77.5
R15	4	2	4	1	4	2	5	2	4	2	80.0
R16	5	2	5	2	5	2	4	1	5	2	87.5
R17	3	2	4	2	4	3	3	2	3	2	65.0
R18	4	2	4	2	4	2	4	2	4	2	75.0
R19	4	1	5	2	4	2	5	1	4	2	85.0
R20	4	2	4	2	4	2	4	2	4	2	75.0
Rata-rata	4.0	1.9	4.3	1.9	4.0	2.0	4.2	1.8	4.0	2.0	77.25

Tabel 4.7 Perhitungan Skor SUS per Pertanyaan

No. Pertanyaan	Tipe	Rumus Skor	Rata-rata Respons	Kontribusi
Q1 (ganjil)	Positif	$Q1 - 1$	3.95	2.95
Q2 (genap)	Negatif	$5 - Q2$	1.90	3.10
Q3 (ganjil)	Positif	$Q3 - 1$	4.35	3.35
Q4 (genap)	Negatif	$5 - Q4$	1.95	3.05
Q5 (ganjil)	Positif	$Q5 - 1$	4.05	3.05
Q6 (genap)	Negatif	$5 - Q6$	2.00	3.00
Q7 (ganjil)	Positif	$Q7 - 1$	4.20	3.20
Q8 (genap)	Negatif	$5 - Q8$	1.80	3.20
Q9 (ganjil)	Positif	$Q9 - 1$	4.00	3.00
Q10 (genap)	Negatif	$5 - Q10$	2.00	3.00
Total				30.90
Skor SUS ($\times$ 2.5)				77.25

Berdasarkan Tabel 4.6 dan Tabel 4.7, rata-rata skor SUS yang diperoleh adalah 77.25 dari skala 100. Menurut skala interpretasi SUS (Bangor et al., 2008), skor 77.25 masuk dalam kategori "Good" (skor 70–84,9), yang menunjukkan bahwa chatbot pesantren ini memiliki tingkat usability yang baik dan dapat diterima oleh pengguna. Nilai ini menunjukkan bahwa antarmuka chatbot berbasis WhatsApp cukup intuitif dan tidak memerlukan pelatihan khusus bagi pengguna.

Skor SUS tertinggi diperoleh dari responden yang sudah familiar dengan WhatsApp sebagai media komunikasi sehari-hari, sedangkan skor lebih rendah cenderung berasal dari responden yang kurang terbiasa dengan interaksi berbasis teks. Secara keseluruhan, hasil pengujian SUS mengonfirmasi bahwa chatbot

informasi pesantren ini layak digunakan oleh calon santri dan orang tua dalam memperoleh informasi pesantren secara mandiri melalui WhatsApp.

4.3 Pembahasan

Hasil penelitian menunjukkan bahwa *chatbot* informasi berbasis WhatsApp yang dikembangkan dalam penelitian ini berhasil dibangun dan menunjukkan performa yang cukup baik. Akurasi keseluruhan sistem *chatbot* ini sebesar 85,00% pada 100 data uji. Meskipun sedikit di bawah hasil yang dilaporkan oleh Nurrochmah (2023) sebesar 88,4% menggunakan SBERT pada kasus serupa, SimCSE menawarkan keunggulan signifikan dalam hal efisiensi: model lebih ringan dalam proses pelatihan, waktu inferensi lebih cepat, dan tidak memerlukan pasangan kalimat berlabel eksplisit pada varian unsupervised. Hal ini menjadikan SimCSE pilihan yang lebih praktis untuk lingkungan dengan sumber daya komputasi terbatas (Gao et al., 2021).

Keunggulan SimCSE dalam penelitian ini terletak pada kemampuannya menangkap kesamaan semantik antara pertanyaan pengguna dan *dataset* FAQ meskipun kalimat yang digunakan berbeda secara leksikal. Hal ini terbukti pada kategori Kegiatan Sosial dan Alumni yang mencapai akurasi tertinggi (100,00%), di mana pertanyaan pada kategori tersebut cenderung lebih spesifik sehingga lebih mudah dikenali oleh sistem.

Terdapat beberapa kasus di mana sistem memberikan jawaban yang kurang tepat. Kegagalan tersebut umumnya terjadi pada pertanyaan yang mengandung kata kunci ambigu atau pertanyaan yang secara semantik dekat dengan lebih dari satu

FAQ, sehingga skor *Cosine Similarity* antara kandidat teratas hampir sama. Hal ini menunjukkan perlunya mekanisme disambiguasi atau penambahan konteks percakapan (dialog management) pada pengembangan selanjutnya.

Dari perspektif integrasi Islam dan sains, penelitian ini selaras dengan nilai-nilai Islam yang menekankan pentingnya penyampaian informasi yang benar dan tepat sasaran sebagaimana tercermin dalam QS. An-Naml ayat 18–19 yang dikutip pada Bab I. *Chatbot* yang dibangun berfungsi sebagai wasilah (perantara) yang memudahkan komunikasi antara pesantren dengan masyarakat, sehingga proses penyampaian informasi tentang pendidikan Islam dapat berjalan lebih efektif dan efisien.

Secara keseluruhan, sistem *chatbot* yang dikembangkan dalam penelitian ini telah berhasil menjawab seluruh rumusan masalah yang ditetapkan. *Chatbot* mampu memberikan informasi secara cepat dengan rata-rata waktu respons 1,23 detik, akurat dengan akurasi 85,00%, dan mudah digunakan melalui platform WhatsApp yang familiar bagi pengguna. Hasil ini membuktikan bahwa pendekatan *retrieval-based* menggunakan SimCSE dan *Cosine Similarity* merupakan solusi yang efektif untuk membangun *chatbot* informasi pada lembaga pendidikan Islam.

Dari sisi teknis, penggunaan model multilingual-e5-base terbukti membantu dalam menangani variasi bahasa yang digunakan pengguna. Model ini cukup fleksibel dalam mengenali pertanyaan informal maupun campuran bahasa Indonesia dan Jawa yang lazim digunakan di lingkungan pesantren. Meski begitu, model ini membutuhkan memori yang cukup besar saat dimuat, sehingga pada

server dengan RAM terbatas perlu dipertimbangkan penggunaan model yang lebih ringan seperti paraphrase-multilingual-MiniLM-L12-v2 yang ukurannya jauh lebih kecil.

Secara umum, penelitian ini menunjukkan bahwa teknologi *chatbot* berbasis *retrieval* dengan *embedding* SimCSE dapat diterapkan secara praktis di lingkungan pesantren tanpa memerlukan infrastruktur *server* yang mahal. Sistem yang dibangun juga sudah menjawab seluruh rumusan masalah yang ditetapkan di awal penelitian: *chatbot* dapat memberikan informasi dengan rata-rata waktu 1,23 detik, akurasi 85,00%, dan tingkat kepuasan pengguna 4,28/5. Pengembangan ke depan perlu difokuskan pada perluasan *dataset*, penanganan percakapan multi-giliran, serta migrasi ke akun WhatsApp Business resmi agar bisa digunakan oleh lebih banyak pengguna secara bersamaan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pengembangan sistem *chatbot* informasi Pesantren Mahasiswa Al-Hikam Malang berbasis WhatsApp menggunakan metode SimCSE dan Cosine *Similarity*, dapat ditarik kesimpulan sebagai berikut.

Pertama, sistem *chatbot* berhasil dirancang dan dibangun dengan memanfaatkan metode SimCSE untuk menghasilkan representasi vektor kalimat (*sentence embedding*) dan Cosine *Similarity* untuk mengukur kemiripan antara pertanyaan pengguna dengan data FAQ yang tersimpan. Sistem diintegrasikan dengan platform WhatsApp menggunakan WhatsApp Business API sehingga dapat diakses secara langsung melalui aplikasi yang sudah familiar di kalangan pengguna.

Kedua, berdasarkan hasil pengujian akurasi pencocokan jawaban terhadap 100 data uji yang mencakup berbagai kategori informasi pesantren, sistem *chatbot* mencapai akurasi rata-rata sebesar 85,00%. Angka ini menunjukkan bahwa pendekatan *retrieval-based* menggunakan SimCSE dengan *threshold* Cosine *Similarity* 0,70 mampu mengenali pertanyaan pengguna dengan tingkat ketepatan yang memadai, termasuk pada variasi bahasa informal dan campuran bahasa Indonesia-Jawa yang lazim digunakan di lingkungan pesantren.

Ketiga, sistem *chatbot* memiliki performa respons yang baik dengan rata-rata waktu respons 1,23 detik per pertanyaan. Kecepatan ini dinilai memadai untuk

kebutuhan informasi santri dan wali santri yang memerlukan jawaban cepat terkait informasi pendaftaran, jadwal, biaya, dan kegiatan pesantren.

Keempat, hasil pengujian kepuasan pengguna melalui kuesioner System Usability Scale (SUS) mendapatkan skor rata-rata 77,25 dari skala 100 (kategori Good/Baik), yang mengindikasikan bahwa pengguna merasa puas dengan kemudahan penggunaan dan kualitas informasi yang diberikan oleh *chatbot*. Hal ini membuktikan bahwa sistem yang dikembangkan dapat diterima dan digunakan dengan baik oleh masyarakat luas.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan sistem *chatbot* ini lebih lanjut agar dapat memberikan manfaat yang lebih optimal.

Pertama, perluasan dan pembaruan *dataset* FAQ secara berkala perlu dilakukan seiring dengan perkembangan informasi dan kebijakan pesantren. *Dataset* yang lebih lengkap dan beragam akan meningkatkan cakupan pertanyaan yang dapat dijawab oleh *chatbot* sehingga semakin banyak kebutuhan informasi pengguna yang terlayani.

Kedua, pengembangan kemampuan *chatbot* dalam menangani percakapan multi-giliran (multi-turn conversation) perlu dipertimbangkan. Sistem saat ini hanya mampu menjawab satu pertanyaan dalam satu waktu tanpa mempertahankan konteks percakapan sebelumnya. Penambahan manajemen konteks percakapan akan membuat interaksi lebih natural dan informatif.

Ketiga, migrasi dari WhatsApp Business API berbasis Twilio ke akun WhatsApp Business API resmi (Meta Cloud API) sangat direkomendasikan agar sistem dapat melayani lebih banyak pengguna secara bersamaan tanpa batasan sandbox. Langkah ini juga akan meningkatkan keandalan dan keamanan sistem secara keseluruhan.

Keempat, penelitian lanjutan dapat mengeksplorasi penggunaan model *embedding* yang lebih ringan seperti paraphrase-multilingual-MiniLM-L12-v2 sebagai alternatif multilingual-e5-base untuk efisiensi memori pada *server* dengan sumber daya terbatas, serta pengujian dengan model SimCSE supervised berbahasa Indonesia untuk meningkatkan akurasi pencocokan pada teks berbahasa lokal.

DAFTAR PUSTAKA

- Ahmed, S., & Rahman, M. 2023. AI *Chatbot* Integration in University Admission Processes. *Education and Information Technologies*, 28: 5473–5490.
- Anonim. 2023. *Panduan Pendaftaran Santri Baru Pesantren Al-Hikam*. Malang: Pesantren Al-Hikam.
- Chen, X., Zhang, Y., & Wang, L. 2022. *Chatbot*-Enhanced Learning Platforms for Higher Education Students. *Computers & Education*, 182: 104476.
- Devlin, Jacob., Chang, Ming-Wei., Lee, Kenton., & Toutanova, Kristina. 2019. BERT: *Pre-training* of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*. Minneapolis: ACL.
- Gao, Tianyu., Yao, Xingcheng., & Chen, Danqi. 2021. SimCSE: Simple *Contrastive Learning* of Sentence Embeddings. *Proceedings of EMNLP 2021*. Punta Cana: Association for Computational Linguistics.
- Gusmita, R. 2020. Pemanfaatan *Chatbot* dalam Layanan Informasi Pendidikan. *Jurnal Teknologi Informasi*, 8(2): 45–52. Jakarta: Universitas Negeri Jakarta.
- Hasan, M., & Khan, A. 2022. *Intelligent Chatbots for Student Academic Support: Implementation and Evaluation*. *International Journal of Emerging Technologies in Learning*, 17(8): 245–259.
- Ibnu Majah, Muhammad bin Yazid. (889 M/275 H). *Sunan Ibnu Majah*. Tahqiq: Muhammad Fu'ad Abdul Baqi. Beirut: Dar al-Fikr.
- Islam, M., Rahman, T., & Chowdhury, S. 2024. AI-Driven *Chatbot* Implementation for Higher Education. *International Journal of Educational Technology in Higher Education*, 21(1): 1–20.
- Kumar, A., & Singh, R. 2020. NLP-Based *Chatbot* for Student Guidance. *International Journal of Advanced Computer Science and Applications*, 11(12): 45–53.
- Li, Y., Chen, T., & Zhao, Q. 2023. Semantic-Based *Chatbot* Systems in Higher Education Institutions. *Education and Information Technologies*, 28: 1121–1139.
- Manning, C. D., Raghavan, P., & Schütze, H. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

- Manzoor, S., Khan, M., & Hussain, A. 2022. Towards *Retrieval-Based* Conversational Recommendation. *Knowledge-Based Systems*, 248: 108875.
- Meta. 2023. *WhatsApp Business API Documentation*. Menlo Park: Meta Platforms, Inc. Diakses dari <https://developers.facebook.com/docs/whatsapp>.
- Nurrochmah. 2023. Implementasi SBERT untuk *Chatbot* Informasi Pesantren. Skripsi. Malang: Universitas Islam Negeri Maulana Malik Ibrahim.
- Nugroho, D. 2022. *Analisis Penggunaan WhatsApp sebagai Media Komunikasi Utama di Lingkungan Pendidikan*. *Jurnal Komunikasi Islam*, 4(1): 23–30. Yogyakarta: UIN Sunan Kalijaga.
- Park, S., & Kim, H. 2021. *Enhancing Student Interaction with AI-Powered Chatbots*. *International Journal of Artificial Intelligence in Education*, 31: 603–624.
- Prasetyo, D. 2019. Penerapan *Natural Language Processing* dalam Analisis Teks Bahasa Indonesia. *Jurnal Teknologi Informasi*, 15(2): 112–124.
- Rahmawati. 2023. *Chatbot* Informasi Akademik Berbasis WhatsApp di Perguruan Tinggi Indonesia. *Jurnal Teknologi dan Sistem Informasi*, 4(2): 45–58.
- Rahman, F., & Hossain, M. 2024. AI-Based Academic *Chatbot* for Student *Query* Resolution. *Journal of AI Research*, 72: 215–233.
- Reimers, N., & Gurevych, I. 2019. *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. *Proceedings of EMNLP-IJCNLP 2019*. Hong Kong: Association for Computational Linguistics.
- Santos, L., & Oliveira, T. 2020. *Designing Conversational Agents for Educational Purposes: Challenges and Opportunities*. *Computers & Education*, 147: 103787.
- Shihab, M. Quraish. (2002). *Tafsir Al-Misbah: Pesan, Kesan dan Keserasian Al-Qur'an* (Vol. 10). Jakarta: Lentera Hati.
- Wong, C., & Tan, P. 2022. *Automated Question Answering Systems for Educational Institutions*. *IEEE Access*, 10: 11234–11245.
- Zhang, H., Li, J., & Liu, K. 2021. *Conversational AI for Academic Support: A Review*. *Journal of Educational Technology Systems*, 50(3): 310–328.

LAMPIRAN

Lampiran 1. Dataset FAQ Chatbot Pesantren Al-Hikam Malang

Berikut adalah seluruh data FAQ yang digunakan sebagai dataset chatbot.

Dataset bersumber dari dokumen FAQ resmi Pesantren Mahasiswa Al-Hikam Malang dan disimpan dalam file faq.csv.

No.	Pertanyaan	Jawaban
1	Apa itu Pesantren Al-Hikam?	Pesantren Mahasiswa Al-Hikam Malang adalah lembaga pendidikan Islam yang membina mahasiswa untuk memperdalam agama dan membentuk karakter.
2	Bagaimana cara mendaftar di Pesantren Al-Hikam?	Pendaftaran dapat dilakukan dengan mengisi formulir di kantor pesantren atau melalui website resmi Pesantren Al-Hikam Malang.
3	Apakah ada biaya pendaftaran?	Ya, terdapat biaya pendaftaran yang ditentukan oleh pengurus pesantren. Silakan hubungi kantor untuk informasi biaya terkini.
4	Apakah Pesantren Al-Hikam menerima mahasiswa dari luar Malang?	Ya, pesantren menerima mahasiswa dari berbagai daerah, tidak hanya dari Malang.
5	Fasilitas apa saja yang tersedia?	Pesantren menyediakan asrama, perpustakaan, mushola, serta kegiatan kajian rutin untuk mahasiswa.
6	Di mana lokasi Pesantren Al-Hikam?	Pesantren Mahasiswa Al-Hikam berlokasi di Kota Malang, Jawa Timur.
7	Apa visi dan misi Pesantren Al-Hikam?	Visi pesantren adalah mencetak mahasiswa yang berilmu, berakhlak, dan bertaqwa. Misinya adalah membina mahasiswa melalui pendidikan agama Islam yang komprehensif.
8	Kapan waktu pendaftaran dibuka?	Pendaftaran biasanya dibuka setiap awal semester. Hubungi pengurus pesantren untuk jadwal pasti pendaftaran.

9	Apakah ada asrama di Pesantren Al-Hikam?	Ya, pesantren menyediakan asrama (ma'had) bagi santri putra dan putri yang terpisah.
10	Apakah ada program khusus untuk mahasiswa baru?	Ya, ada program orientasi (ta'aruf) bagi santri baru untuk mengenal lingkungan, peraturan, dan kegiatan pesantren.
11	Apa saja kegiatan rutin di Pesantren Al-Hikam?	Kegiatan rutin meliputi shalat berjamaah, kajian kitab kuning, tahsin Al-Qur'an, kajian hadits, dan kegiatan organisasi santri.
12	Apakah ada kegiatan tahfidz di pesantren?	Ya, pesantren memiliki program tahfidz Al-Qur'an bagi santri yang ingin menghafal Al-Qur'an.
13	Berapa biaya tinggal di asrama per bulan?	Biaya asrama bervariasi sesuai tipe kamar. Silakan hubungi kantor pesantren untuk informasi biaya terkini.
14	Apakah boleh membawa kendaraan pribadi?	Ketentuan membawa kendaraan pribadi diatur dalam peraturan pesantren. Silakan konfirmasi ke pengurus.
15	Apakah pesantren memiliki perpustakaan?	Ya, pesantren memiliki perpustakaan yang menyediakan koleksi kitab klasik, buku agama, dan buku umum untuk santri.
16	Bagaimana sistem pembelajaran di Al-Hikam?	Sistem pembelajaran menggabungkan metode sorogan, bandongan, dan halaqoh untuk mengkaji kitab kuning dan ilmu agama.
17	Apakah santri wajib mengikuti semua kegiatan?	Ya, santri diwajibkan mengikuti seluruh kegiatan wajib yang telah dijadwalkan oleh pesantren.
18	Apa peraturan umum yang harus ditaati santri?	Santri wajib menjaga adab, mengikuti kegiatan wajib, menjaga kebersihan, menghormati sesama, dan mematuhi seluruh tata tertib pesantren.
19	Apakah ada sanksi jika melanggar peraturan?	Ya, pelanggaran akan diproses sesuai tingkatannya, mulai dari teguran lisan, tertulis, hingga dikeluarkan dari pesantren.
20	Apakah ada kegiatan sosial di pesantren?	Ya, pesantren aktif dalam kegiatan sosial kemasyarakatan seperti bakti sosial, pengajian untuk masyarakat sekitar, dan kegiatan Ramadhan.

21	Siapa pendiri Pesantren Al-Hikam?	Pesantren Mahasiswa Al-Hikam Malang didirikan oleh KH. Hasyim Muzadi, ulama besar yang juga pernah menjabat sebagai Ketua Umum PBNU.
22	Apakah Pesantren Al-Hikam berafiliasi dengan organisasi tertentu?	Pesantren Al-Hikam berafiliasi dengan Nahdlatul Ulama (NU) dan memegang teguh nilai-nilai Ahlussunnah wal Jamaah.
23	Apakah ada program beasiswa di pesantren?	Informasi mengenai beasiswa dapat ditanyakan langsung ke kantor pesantren karena ketersediaannya bergantung pada kebijakan tahunan.
24	Bagaimana cara menghubungi pesantren?	Anda dapat menghubungi pesantren melalui nomor telepon kantor atau datang langsung ke kantor administrasi pesantren di Malang.
25	Apakah pesantren menerima santri non-mahasiswa?	Pesantren Al-Hikam Malang secara khusus diperuntukkan bagi mahasiswa yang sedang menempuh pendidikan di perguruan tinggi.
26	Apakah ada batas usia untuk mendaftar?	Umumnya santri adalah mahasiswa aktif perguruan tinggi. Untuk ketentuan detail, silakan tanyakan ke pengurus pesantren.
27	Apakah ada kegiatan olahraga di pesantren?	Ya, pesantren menyediakan waktu dan fasilitas untuk kegiatan olahraga santri sebagai bagian dari pembinaan fisik.
28	Bagaimana jadwal kegiatan harian santri?	Kegiatan harian dimulai dari shalat Subuh berjamaah, kajian pagi, kemudian santri berangkat ke kampus, dan dilanjutkan dengan kajian malam setelah Isya.
29	Apakah ada koperasi atau kantin di pesantren?	Tersedia fasilitas pendukung seperti koperasi untuk memenuhi kebutuhan sehari-hari santri.
30	Apakah pesantren memiliki WIFI atau fasilitas internet?	Fasilitas internet tersedia untuk mendukung kegiatan akademik santri, dengan penggunaan yang diatur sesuai kebijakan pesantren.
31	Bolehkah santri keluar pesantren di hari biasa?	Keluar pesantren diatur dengan sistem izin. Santri harus mengajukan izin kepada pengurus sesuai prosedur yang berlaku.

32	Kapan jadwal kunjungan keluarga (besuk)?	Jadwal besuk diatur oleh pesantren. Silakan hubungi kantor untuk informasi <u>jadwal kunjungan yang berlaku</u> .
33	Apakah ada program bahasa Arab atau Inggris?	Pesantren memiliki program penguatan bahasa Arab sebagai penunjang dalam <u>pengkajian kitab kuning</u> .
34	Apa yang dimaksud dengan kajian kitab kuning?	Kajian kitab kuning adalah pembelajaran kitab klasik berbahasa Arab yang membahas fikih, tafsir, hadits, akhlak, dan ilmu agama lainnya.
35	Apakah pesantren menerima mahasiswa putri?	Ya, Pesantren Al-Hikam memiliki unit untuk santri putri dengan asrama dan <u>pembina yang terpisah dari santri putra</u> .
36	Apa perbedaan Al-Hikam Malang dengan Al-Hikam Depok?	Al-Hikam Malang dan Al-Hikam Depok adalah dua lembaga pesantren yang didirikan oleh KH. Hasyim Muzadi namun berada di lokasi berbeda dan dikelola secara terpisah.
37	Apakah ada kegiatan kajian setiap malam?	Ya, kajian malam merupakan kegiatan wajib yang dilaksanakan setelah shalat Isya berjamaah.
38	Bagaimana cara mengajukan izin tidak mengikuti kegiatan?	Izin dapat diajukan dengan melapor langsung kepada pengurus atau <u>musyrif/musyrifah yang bertanggung jawab</u> .
39	Apakah tersedia kamar mandi dalam kamar?	Fasilitas kamar bervariasi. Informasi detail mengenai tipe kamar dapat ditanyakan saat proses pendaftaran.
40	Apa yang harus dibawa saat pertama masuk pesantren?	Santri baru biasanya perlu membawa perlengkapan ibadah, perlengkapan tidur, pakaian sehari-hari, dan dokumen <u>pendaftaran yang diperlukan</u> .
41	Apakah pesantren mengadakan kegiatan di bulan Ramadhan?	Ya, pesantren memiliki program khusus Ramadhan seperti tarawih berjamaah, tadarus, kajian intensif, dan ifthar bersama.
42	Bagaimana cara bergabung dengan organisasi santri?	Informasi tentang organisasi santri dan cara bergabung dapat diperoleh saat orientasi santri baru atau dengan <u>menghubungi pengurus organisasi</u> .
43	Apakah ada alumni Pesantren Al-Hikam yang terkenal?	Alumni Al-Hikam tersebar di berbagai bidang profesi. Pesantren bangga dengan <u>kontribusi alumninya di masyarakat</u> .

44	Apakah tersedia layanan kesehatan di pesantren?	Pesantren menyediakan akses ke layanan kesehatan dasar untuk santri yang membutuhkan pertolongan medis.
45	Apakah ada tes atau seleksi masuk pesantren?	Ya, calon santri umumnya mengikuti proses seleksi berupa tes membaca Al-Qur'an dan wawancara untuk mengetahui kesiapan dan motivasi dalam mengikuti kegiatan pesantren.
46	Apakah tersedia makan atau katering di pesantren?	Pesantren menyediakan layanan makan bagi santri yang termasuk dalam paket biaya tinggal. Informasi lengkap dapat ditanyakan kepada pengurus saat pendaftaran.
47	Apakah santri boleh membawa laptop atau gadget?	Santri diizinkan membawa laptop dan gadget untuk keperluan akademik dengan ketentuan penggunaan yang diatur sesuai tata tertib pesantren.
48	Bagaimana sistem pembayaran biaya pesantren?	Pembayaran biaya pesantren dapat dilakukan secara bulanan atau per semester sesuai kebijakan yang berlaku. Konfirmasi metode pembayaran dapat dilakukan ke kantor administrasi.
49	Apakah ada kegiatan muhadhoroh atau latihan ceramah?	Ya, pesantren memiliki program muhadhoroh (latihan ceramah) sebagai sarana pengembangan kemampuan komunikasi dan dakwah santri.
50	Apakah pesantren menyediakan program menghafal hadits?	Selain program tahfidz Al-Qur'an, pesantren juga mendorong santri untuk menghafal hadits-hadits pilihan sebagai bagian dari pembinaan keagamaan.