

# HANA: An AI Chatbot for Islamic Jurisprudence on Menstruation using SBERT and TF-IDF

Tsaura Rafah Masuzzahra<sup>1\*</sup>, Khothibul Umam<sup>2\*</sup>, Hery Mustofa<sup>3\*</sup>, Maya Rini Handayani<sup>4\*</sup>

\* Teknologi Informasi, Fakultas Sains dan Teknologi, UIN Walisongo Semarang  
[2208096080@student.walisongo.ac.id](mailto:2208096080@student.walisongo.ac.id)<sup>1</sup>, [khothibul\\_umam@walisongo.ac.id](mailto:khothibul_umam@walisongo.ac.id)<sup>2</sup>,  
[herymustofa@walisongo.ac.id](mailto:herymustofa@walisongo.ac.id)<sup>3</sup>, [maya@walisongo.ac.id](mailto:maya@walisongo.ac.id)<sup>4</sup>

## Article Info

### Article history:

Received 2025-04-19

Revised 2025-05-03

Accepted 2025-05-07

### Keyword:

Chatbot,  
Fiqh Haid,  
Sentence-BERT,  
TF-IDF,  
Telegram Bot.

## ABSTRACT

The advancement of Artificial Intelligence (AI), particularly in Natural Language Processing (NLP), has opened new opportunities for religious technological innovation, especially in addressing practical Islamic jurisprudence issues such as menstruation (fiqh haid). This research proposes and implements HANA, an AI chatbot developed for Telegram, utilizing a hybrid approach combining Term Frequency-Inverse Document Frequency (TF-IDF) and Sentence-BERT (SBERT) models. A curated dataset of over 1000 question-answer pairs from classical and contemporary Islamic literature was used, primarily based on the Shafi'i school of thought. The chatbot matches user queries through a two-stage retrieval: initial keyword matching via TF-IDF and deeper semantic matching via SBERT embeddings. Evaluations were conducted by comparing TF-IDF, SBERT, and hybrid approaches using cosine similarity, precision, recall, and F1-score metrics, focused on top-1 retrieval accuracy. HANA achieved an average cosine similarity score of 0.6581 and a semantic relevance rating of 87% based on expert validation, while User Acceptance Testing (UAT) involving 15 respondents indicated 86.7% satisfaction. Although the system is deployed as a proof-of-concept on Google Colab without persistent hosting, it demonstrates the viability of lightweight AI chatbots for Shariah consultation services. Future improvements include multi-turn conversation handling and integration with large language models for better context understanding. This research contributes to expanding NLP applications within techno-dakwah initiatives, providing a scalable approach to enhance women's access to Islamic jurisprudence knowledge.



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

## I. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membuka peluang besar dalam mempermudah akses dan penyebaran ilmu pengetahuan, termasuk dalam bidang keagamaan. Salah satu tantangan yang dihadapi oleh umat Islam adalah kesulitan dalam mengakses informasi fikih yang terstruktur dan mudah dipahami. Di era digital saat ini, umat Islam memiliki akses yang lebih luas terhadap informasi keislaman melalui berbagai *platform* daring. Namun, tidak semua informasi yang beredar di internet memiliki validitas yang dapat dipertanggungjawabkan, terutama dalam bidang

fiqh. Keterbatasan ini sering kali membuat umat Islam kesulitan untuk memperoleh pemahaman yang jelas dan akurat, terutama terkait dengan fiqh yang bersifat praktikal, seperti fiqh haid, yang memerlukan pemahaman mendalam dalam kehidupan sehari-hari.

Fiqh haid adalah cabang ilmu dalam Islam yang membahas ketentuan-ketentuan syariat terkait dengan siklus menstruasi, termasuk kapan seorang wanita boleh atau tidak boleh menjalankan ibadah tertentu, seperti salat, puasa, dan tawaf. Pemahaman tentang fiqh haid sangat penting karena memiliki dampak langsung terhadap keabsahan ibadah seorang Muslimah. Namun, banyak wanita yang masih

mengalami kesulitan dalam memahami hukum-hukum ini karena keterbatasan akses terhadap ulama atau literatur yang terpercaya [1]. Oleh karena itu, diperlukan solusi berbasis teknologi yang dapat memberikan informasi fiqh haid secara cepat, akurat, dan mudah diakses.

Seiring dengan kemajuan teknologi, penggunaan sistem berbasis *Artificial Intelligence* (AI), khususnya *chatbot* yang didukung oleh teknologi *Natural Language Processing* (NLP), mulai banyak diterapkan untuk memecahkan permasalahan ini. Beberapa penelitian sebelumnya yang dilakukan oleh Elvina Afriyani dalam *jurnal ZONasi*, yang mengusulkan penggunaan sistem *Question Answering* (QA) berbasis AI untuk memberikan akses informasi fiqh bersuci melalui *platform* web [2]. Hasil penelitian tersebut menunjukkan bahwa penerapan AI dalam fiqh tidak hanya meningkatkan efisiensi dalam mencari informasi, tetapi juga meningkatkan pemahaman pengguna dengan menyediakan jawaban yang lebih terstruktur dan cepat.

Begitu juga penelitian lain yang dilakukan oleh Rahayu Deny dalam *jurnal JoSYC*, yang mengembangkan *chatbot* berbasis *LangChain* untuk tanya jawab seputar fiqh jual beli Islam [3]. Hasil penelitian ini menunjukkan bahwa penggunaan *chatbot* dalam aplikasi web dapat mempercepat proses pencarian dan memberikan jawaban yang lebih akurat, sehingga mempermudah pemahaman fiqh muamalah bagi penggunaanya.

Sementara itu, penelitian lain yang dilakukan oleh M.D. Qoyyimah dalam *International Journal of AI and Learning*, berfokus pada penerapan AI dalam mempelajari fiqh tahārah melalui *ChatGPT*, yang menunjukkan potensi teknologi *chatbot* dalam mempermudah pemahaman fiqh meskipun masih terdapat kekurangan dalam memberikan penjelasan yang komprehensif, hasil dari penelitian ini menunjukkan bahwa teknologi *chatbot* berbasis AI berpotensi untuk menjadi alat bantu yang baik dalam mempercepat proses pembelajaran fiqh, khususnya dalam menyajikan informasi secara lebih mudah diakses [4].

Meskipun beberapa penelitian sebelumnya telah mengembangkan *chatbot* untuk konsultasi fiqh, namun sebagian besar penelitian yang ada masih terbatas pada pengembangan sistem berbasis web. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada *platform* web dan tema fiqh umum seperti taharah dan jual beli, penelitian ini mengusulkan pengembangan HANA (*Haid Consultant AI*), sebuah *chatbot* berbasis AI yang diimplementasikan melalui aplikasi Telegram untuk memberikan konsultasi mengenai fiqh haid. Pemilihan Telegram sebagai *platform* utama didasarkan pada kemudahan akses, popularitasnya di kalangan pengguna Muslim, serta dukungannya terhadap integrasi *chatbot* berbasis AI yang interaktif dan responsif.

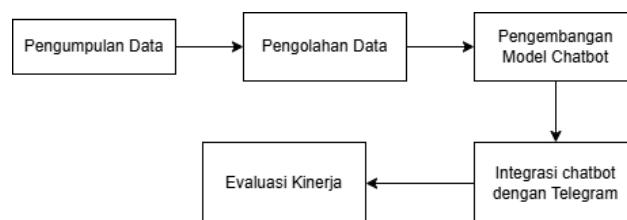
HANA akan dikembangkan menggunakan kombinasi metode SBERT (*Sentence-BERT*) dan TF-IDF (*Term Frequency-Inverse Document Frequency*). SBERT merupakan teknik pemrosesan bahasa alami yang memungkinkan *chatbot* memahami konteks kalimat dengan lebih baik dibandingkan

metode NLP konvensional [5]. Dengan SBERT, *chatbot* dapat memberikan jawaban yang lebih relevan dan sesuai dengan pertanyaan pengguna. Sementara itu, TF-IDF digunakan untuk mengekstrak kata-kata kunci dari pertanyaan yang diajukan sehingga *chatbot* dapat mencari jawaban yang paling sesuai berdasarkan basis pengetahuan yang dimilikinya [6].

Dengan pendekatan ini, penelitian ini bertujuan untuk mengembangkan *chatbot* HANA berbasis AI yang dapat memberikan informasi fiqh haid secara akurat melalui Telegram. Penelitian ini juga akan mengevaluasi kinerja *chatbot* HANA dalam memahami dan merespons pertanyaan seputar fiqh haid secara akurat dan relevan, serta menganalisis efektivitas penggunaan *platform* Telegram sebagai media konsultasi dan diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan teknologi keislaman yang lebih memudahkan muslimah dalam memahami hukum-hukum fiqh haid secara lebih baik. Berbeda dari penelitian-penelitian sebelumnya yang masih terbatas pada topik fiqh umum dan *platform* berbasis web, keunikan penelitian ini terletak pada fokus tematiknya yang spesifik pada fiqh haid serta implementasinya dalam aplikasi percakapan populer (Telegram). Selain itu, pendekatan hybrid dengan menggabungkan TF-IDF dan SBERT untuk pencocokan dua tahap menjadikan sistem ini lebih efisien dalam seleksi awal dan lebih akurat dalam pemahaman semantik, sehingga dapat memberikan jawaban yang lebih relevan dan kontekstual.

## II. METODE

Penelitian ini menggunakan metode eksperimen dengan pendekatan berbasis *Natural Language Processing* (NLP) untuk mengembangkan dan mengevaluasi kinerja *chatbot* HANA (*Haid Consultant AI*) dalam memberikan informasi fiqh haid. Secara umum, penelitian ini meliputi: (A) pengumpulan dan pemrosesan data, (B) pengembangan model *chatbot*, (C) implementasi *chatbot* pada *platform* Telegram, dan (D) evaluasi kinerja *chatbot* berdasarkan akurasi dan relevansi jawaban yang diberikan.



Gambar 1. Alur Penelitian

### A. Pengumpulan dan Pemrosesan Data

Salah satu komponen utama dalam pengembangan *chatbot* HANA adalah penyusunan *dataset* pertanyaan dan jawaban yang dikembangkan berdasarkan literatur fiqh wanita. *Dataset* ini disusun secara manual melalui pendekatan semi-

otomatis, dengan mengekstraksi konten dari tiga buku fiqh wanita dalam format PDF, yaitu buku yang berjudul “*Fikih Menstruasi Menghapus Mitos-Mitos dalam Menstrual Taboo*” karya Wardah Nuroniyah, “*Haid (Menstruasi) dalam Tinjauan Hadis*” karya Lutfi Rahmatullah, dan “*Haid dan Kesehatan Menurut Ajaran Islam*” karya Dr. Nonon Saribanon, M.Si.

Proses ekstraksi dilakukan dengan menggunakan *library* PyPDF2, sebuah *library python* yang memungkinkan pembacaan dan pengolahan *file* PDF. PyPDF2 memungkinkan pembacaan dan pemrosesan isi teks dari setiap halaman, menghilangkan elemen visual seperti gambar dan tabel yang tidak relevan, dan menyusun informasi dalam format teks yang bersih [7]. Berdasarkan hasil ekstraksi tersebut, dilakukan identifikasi dan pengembangan pasangan pertanyaan-jawaban potensial yang sering diajukan terkait tema fiqh haid. Proses ini dilakukan melalui penelaahan isi literatur secara menyeluruh, kemudian hasilnya disesuaikan kembali oleh penulis untuk memastikan kesesuaian dengan sumber materinya serta kelayakannya sebagai basis pengetahuan chatbot. QA pairs yang telah disusun kemudian direview oleh dua ustadzah dan satu akademisi fiqh untuk memastikan validitas substansi syariah serta relevansi konteks sebelum digunakan dalam sistem. Validasi ini dilakukan untuk meminimalisasi kesalahan penafsiran serta menjaga kesesuaian jawaban dengan kaidah fiqh yang sah. Dataset ini kemudian digunakan sebagai dasar dalam pencocokan semantik menggunakan model SBERT (Sentence-BERT), yang bertujuan untuk mengukur akurasi sistem dalam memahami konteks pertanyaan dan menyajikan jawaban yang sesuai secara syar’i.

Secara keseluruhan, dataset terdiri dari lebih dari 1000 pasangan pertanyaan-jawaban (QA pairs) yang difokuskan pada tema fiqh haid. Topik-topik yang tercakup meliputi Konsep Haid dan Siklus Menstruasi, Hukum Ibadah saat Haid dalam Islam, Tata cara bersuci dari haid. Penting untuk dicatat bahwa dataset ini mayoritas mengacu pada mazhab Syafi’i, sesuai dengan dominasi literatur yang digunakan. Cakupan pandangan dari mazhab lain (Hanafi, Maliki, Hanbali) belum sepenuhnya diakomodasi, sehingga sistem berpotensi menghasilkan jawaban yang fokus pada perspektif Syafi’i. Meskipun demikian, substansi jawaban tetap dijaga validitasnya berdasarkan sumber literatur fiqh yang sah.

Lalu tahapan berikutnya adalah pemrosesan data. Tahapan pemrosesan data dilakukan melalui beberapa langkah utama, yaitu pembersihan data (*cleaning*), tokenisasi (*tokenization*), penghapusan kata-kata umum (*stopword removal*), serta vektorisasi teks (*vectorization*) [8].

Setelah *dataset* dikumpulkan, tahap berikutnya adalah proses pembersihan data (*cleaning data*) untuk memastikan kualitas *input* pada model pemrosesan bahasa alami. Tahap *cleaning* dilakukan secara sistematis menggunakan *Python* dengan tujuan menghilangkan karakter khusus, duplikasi, dan teks yang tidak relevan agar data lebih bersih dan terstruktur[9]. Selanjutnya, proses *tokenization* diterapkan untuk memecah teks menjadi unit kata atau kalimat sehingga dapat dianalisis lebih lanjut [10]. Setelah itu, dilakukan

*stopword removal*, yaitu penghapusan kata-kata umum yang tidak berkontribusi dalam pemahaman konteks [11], guna meningkatkan efektivitas model dalam memahami pertanyaan pengguna. Setelah proses awal dilakukan di level kata, tahap selanjutnya berfokus pada pemahaman konteks kalimat secara utuh menggunakan pendekatan representasi semantik. Untuk itu, digunakan model Sentence-BERT (SBERT) sebagai metode utama dalam pencocokan makna kalimat.

Setelah data melalui tahap pembersihan, langkah berikutnya adalah melakukan proses *vectorization* menggunakan model *Sentence-BERT*(SBERT) untuk mengubah teks pertanyaan dan jawaban menjadi bentuk representasi numerik (*embedding*) [12]. Sebelum proses pencocokan semantik dilakukan oleh SBERT, sistem terlebih dahulu menggunakan metode TF-IDF untuk mengekstrak kata kunci dari pertanyaan pengguna. Tujuan dari penggunaan TF-IDF ini adalah untuk menyaring dan memprioritaskan pertanyaan yang memiliki kesamaan kata kunci tinggi terhadap pertanyaan dalam *dataset*, sehingga dapat mempersempit ruang pencarian dan meningkatkan efisiensi pencocokan[13]. Dengan demikian, proses pencarian jawaban menjadi lebih cepat karena SBERT hanya melakukan pencocokan semantik terhadap *subset* data yang sudah difilter menggunakan TF-IDF. Proses ini penting untuk memungkinkan sistem mengenali kesamaan semantik antara pertanyaan yang diajukan oleh pengguna dan kumpulan jawaban yang telah tersedia dalam *dataset*.

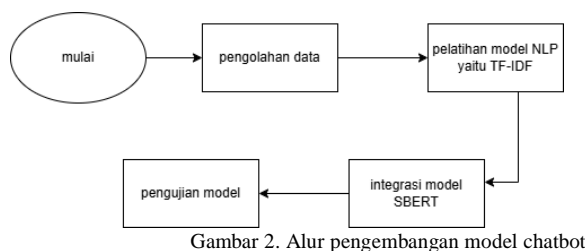
Model SBERT yang digunakan dalam penelitian ini diunduh dari model *pre-trained* all-MiniLM-L6-v2 yang tersedia melalui *Hugging Face* [14]. Model ini dipilih karena memiliki performa tinggi dalam representasi kalimat secara efisien dan ringan untuk kebutuhan pengujian di lingkungan *Google Colab*.

Proses vektorisasi dilakukan dalam dua tahap, yaitu *encoding dataset* dan *encoding* pertanyaan pengguna. Pertama, seluruh daftar pertanyaan dalam *dataset* FAQ diubah menjadi *embedding* menggunakan model SBERT yang telah disebutkan sebelumnya, dan hasil *embedding* ini disimpan sebagai basis pencocokan semantik. Kemudian, ketika pengguna mengajukan pertanyaan melalui *chatbot*, sistem akan memproses pertanyaan tersebut menggunakan model SBERT yang sama untuk menghasilkan *embedding* yang setara, lalu membandingkannya dengan *embedding* dalam *dataset* menggunakan metode *cosine similarity* [6]. Hasil dari tahap ini berupa vektor representatif untuk setiap pertanyaan, yang menjadi dasar dalam kesamaan konteks antara *input* pengguna dan data yang tersedia, sehingga *chatbot* dapat memberikan jawaban yang paling relevan.

## B. Pengembangan Model Chatbot

Dalam penelitian ini, *chatbot* dikembangkan untuk memberikan informasi mengenai fiqh haid dengan memanfaatkan teknologi pemrosesan bahasa alami. Proses pengembangan difokuskan pada pembuatan model yang mampu memahami konteks pertanyaan pengguna dan

memberikan jawaban yang relevan sesuai dengan prinsip-prinsip syariat Islam. Seluruh proses pengembangan dilakukan langsung di *Google Colab* tanpa *framework backend* seperti *Flask* atau *FastAPI*, dengan fokus pada pemrosesan data dan pembuatan sistem pencocokan semantik. Pengembangan ini bersifat tahap eksperimen dan demonstrasi, sehingga sistem yang dibangun belum dirancang untuk implementasi produksi skala besar secara persisten. Berikut alur pengembangan model chatbot pada contoh Gambar 2.



Adapun tahapan-tahapan pengembangan model *chatbot* meliputi:

1) *Pengolahan Data*: Dataset yang digunakan dalam pengembangan chatbot ini terdiri dari pertanyaan-pertanyaan umum yang terkait dengan fiqh haid beserta jawabannya. Dataset ini diolah untuk melatih model NLP yang dapat memahami konteks pertanyaan dan memberikan jawaban yang sesuai. Proses pengolahan data ini mencakup pembersihan teks, normalisasi, serta penyusunan data dalam format yang dapat digunakan untuk pelatihan model.

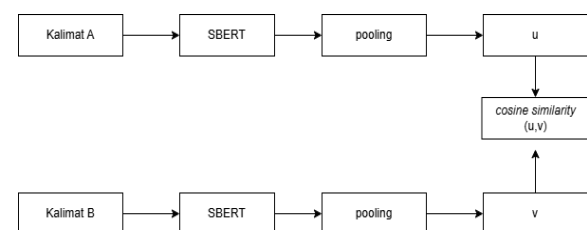
2) *Pemanfaatan Model TF-IDF dan SBERT(Sentence-BERT) untuk Pencocokan Semantik*: Chatbot ini memanfaatkan model SBERT (Sentence-BERT) pre-trained all-MiniLM-L6-v2 dari Hugging Face yang mampu mempresentasikan kalimat dalam bentuk vektor embedding. SBERT dipilih karena kemampuannya dalam menghasilkan representasi semantik kalimat yang dapat dibandingkan secara efektif [15]. Representasi vektor dari pertanyaan-pertanyaan fiqh dalam dataset digunakan sebagai basis pencocokan. Ketika pengguna mengajukan pertanyaan, sistem mengubah pertanyaan tersebut menjadi embedding dan membandingkannya dengan embedding dalam basis data menggunakan cosine similarity untuk menemukan pasangan jawaban yang relevan.

Kombinasi metode TF-IDF dan SBERT dalam pengembangan chatbot HANA dipilih untuk mengoptimalkan proses pencarian jawaban yang relevan secara efisien dan kontekstual. Secara teoritis, TF-IDF unggul dalam mengekstraksi kata kunci penting dari pertanyaan pengguna, sehingga sangat efektif sebagai tahap filtrasi awal. Namun, metode ini memiliki keterbatasan dalam memahami makna kalimat secara utuh, terutama jika pertanyaan memiliki struktur kompleks atau menggunakan sinonim. Sebaliknya, SBERT memiliki keunggulan dalam menangkap kesamaan antar kalimat melalui representasi vektor yang lebih

kontekstual, meskipun proses pencariannya lebih berat secara komputasi.

Oleh karena itu, pendekatan hybrid digunakan dalam penelitian ini, di mana TF-IDF berfungsi untuk menyaring data ke dalam subset yang lebih kecil, dan SBERT melakukan pencocokan semantik mendalam terhadap subset tersebut. Strategi ini tidak hanya mempercepat proses pencarian, tetapi juga meningkatkan relevansi jawaban yang diberikan kepada pengguna.

Proses transformasi ini ditunjukkan pada Gambar 2, di mana dua kalimat (Kalimat A dan Kalimat B) diproses melalui model SBERT, kemudian dilakukan pooling untuk menghasilkan vektor representasi (u dan v). Selanjutnya, nilai kemiripan semantik dihitung menggunakan rumus cosine similarity antara kedua vektor tersebut.



Gambar 3. Cara Kerja SBERT

3) *Pengujian Model NLP*: Setelah model siap, dilakukan pengujian untuk mengevaluasi efektivitas model SBERT dalam mencocokkan pertanyaan pengguna dengan jawaban dalam dataset. Pengujian ini dilakukan secara internal dengan skenario simulasi input pertanyaan untuk melihat kesesuaian hasil pencocokan. Hasil pengujian ini menjadi dasar untuk melanjutkan ke tahap integrasi dengan Telegram.

### C. Implementasi Chatbot pada Platform Telegram

Setelah pengembangan model *chatbot* selesai dan performa model SBERT dinilai cukup baik dalam pencocokan semantik, tahap selanjutnya adalah mengimplementasikan *chatbot* pada platform Telegram agar dapat digunakan secara langsung oleh pengguna. Implementasi ini dilakukan menggunakan Telegram Bot API melalui library *aiogram* di *Google Colab* tanpa memerlukan server *hosting* khusus, sehingga proses integrasi dan demonstrasi dapat dilakukan secara praktis. Berikut adalah tahapan implementasi *chatbot*:

1) *Pembuatan Bot Telegram*: Bot dibuat menggunakan layanan BotFather di Telegram. Setelah proses pendaftaran, peneliti memperoleh token API yang digunakan untuk menghubungkan aplikasi dengan bot Telegram[16].

2) *Integrasi dengan Telegram Bot API*: Integrasi dilakukan dengan menggunakan library Python *aiogram*, yang mendukung interaksi asinkron antara pengguna dan sistem. Bot diprogram untuk menerima input pertanyaan dari pengguna, memprosesnya menggunakan model SBERT, dan

mengirimkan jawaban yang paling sesuai berdasarkan hasil pencocokan semantik.

3) *Pengujian langsung oleh pengguna*: Setelah berhasil terhubung, chatbot kemudian diaktifkan dan diuji langsung melalui Telegram. Pengguna mengirimkan pertanyaan seputar fiqh haid, dan chatbot memberikan respons secara real-time. Respons yang diberikan diambil dari dataset FAQ berdasarkan pertanyaan yang paling mendekati secara makna dengan input pengguna.

4) *Keunggulan implementasi*: Integrasi yang dilakukan sepenuhnya di Google Colab memungkinkan pengujian tanpa memerlukan deployment ke cloud secara penuh. Hal ini mempermudah pengembangan awal dan sangat mendukung kebutuhan pengujian model NLP secara fleksibel. Namun, karena implementasi dilakukan di Google Colab tanpa infrastruktur server permanen, sistem ini hanya bersifat proof of concept dan tidak dirancang untuk beroperasi secara real-time 24/7.

Dengan Implementasi ini, *chatbot* fiqh haid dapat diakses oleh siapapun melalui Telegram, memberikan kemudahan dalam memperoleh jawaban terkait persoalan perempuan dalam perspektif syariat Islam secara cepat dan akurat.

#### D. Desain Evaluasi Kinerja Sistem Chatbot

Kinerja *chatbot* HANA dievaluasi berdasarkan tiga metrik utama yang bertujuan untuk menilai akurasi, relevansi, dan kepuasan pengguna. Evaluasi ini dilakukan melalui serangkaian prosedur terstruktur yang menggambarkan efektivitas dan efisiensi *chatbot* dalam memberikan jawaban yang sesuai dengan konteks fiqh haid. Adapun metrik yang digunakan adalah sebagai berikut:

1. *Akurasi*: Akurasi chatbot diukur berdasarkan kesesuaian antara jawaban yang diberikan dengan referensi fiqh yang telah ditentukan. Untuk mengukur akurasi, digunakan cosine similarity, yang menghitung kesamaan semantik antara pertanyaan pengguna dengan jawaban yang diberikan oleh chatbot. Metrik ini dipilih karena kemampuannya dalam menilai kedekatan konteks antara dua kalimat atau teks. Nilai cosine similarity dihitung untuk setiap interaksi pengguna, dan rata-rata nilai tersebut digunakan untuk mengevaluasi seberapa baik chatbot memahami pertanyaan yang diajukan [17]. Untuk menghitung akurasi semantik antara pertanyaan dan jawaban, digunakan rumus cosine similarity berikut:

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

Dimana  $A$  dan  $B$  adalah vektor representasi dari pertanyaan pengguna dan jawaban *chatbot* hasil dari model SBERT.

2. *Relevansi*: Untuk menilai relevansi jawaban yang diberikan oleh chatbot, dilakukan dua pendekatan. Pendekatan pertama adalah kuantitatif, yang menggunakan metrik seperti precision, recall, dan F1-score untuk mengukur seberapa tepat

jawaban chatbot dalam memberikan informasi yang relevan dengan pertanyaan pengguna. Precision mengukur sejauh mana jawaban yang dihasilkan sistem relevan terhadap keseluruhan jawaban yang dianggap relevan oleh sistem. Recall mengukur sejauh mana sistem dapat menemukan semua jawaban yang relevan dari keseluruhan jawaban yang seharusnya ditemukan. F1-score adalah ukuran yang menggabungkan precision dan recall untuk memberikan gambaran seimbang tentang kinerja sistem. Metrik-metrik ini memberikan evaluasi yang objektif berdasarkan perbandingan antara jawaban yang diberikan oleh sistem dengan jawaban yang diharapkan. Perhitungan precision, recall, dan F1-score dalam penelitian ini dilakukan berdasarkan rumus yang diadopsi dari [18]:

$$\text{precision} = \frac{TP}{TP + FP}$$

Di mana TP (True Positive) adalah jumlah jawaban yang relevan dan dianggap benar oleh sistem, sementara FP (False Positive) adalah jumlah jawaban yang dianggap relevan tetapi sebenarnya tidak relevan.

$$\text{recall} = \frac{TP}{TP + FN}$$

Di mana FN (False Negative) adalah jumlah jawaban yang relevan tetapi tidak dianggap relevan oleh sistem.

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Evaluasi metrik precision, recall, dan F1-score dalam penelitian ini dihitung berbasis top-1 hasil retrieval, sehingga secara teknis mengacu pada precision at 1 (precision at top-1 retrieval), recall at 1, dan F1-score at 1. Fokus evaluasi diarahkan pada kemampuan chatbot dalam memberikan satu jawaban paling relevan terhadap pertanyaan pengguna.

Pendekatan kedua adalah kualitatif, yang dilakukan secara manual. Dalam pendekatan ini, beberapa orang dengan pemahaman dasar mengenai fiqh diminta untuk memberikan penilaian terhadap validitas dan kesesuaian jawaban yang diberikan oleh chatbot. Meskipun penilaian ini bersifat subjektif, namun diharapkan dapat memberikan gambaran mengenai kualitas jawaban *chatbot* dalam konteks syariah yang lebih tepat dan relevan. Proses penilaian relevansi ini dilakukan secara kualitatif dengan mendiskusikan kesesuaian jawaban *chatbot* dengan prinsip-prinsip fiqh.

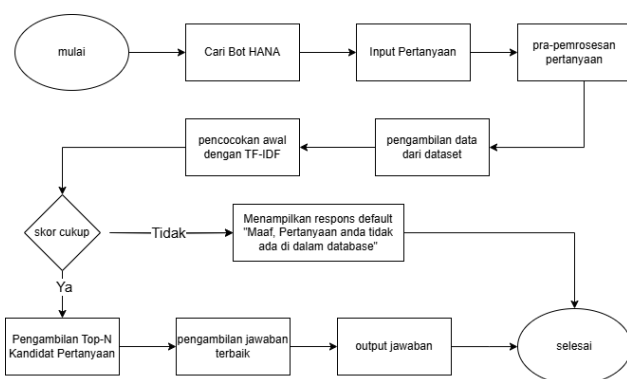
3. *Kepuasan Pengguna*: Kepuasan pengguna diukur dengan meminta umpan balik langsung dari pengguna setelah berinteraksi dengan chatbot. Pengguna diminta untuk memberikan pendapat mereka mengenai seberapa mudah dan jelas chatbot memberikan informasi mengenai fiqh haid. Meskipun jumlah responden terbatas, feedback ini memberikan indikasi awal mengenai penerimaan dan kenyamanan pengguna dalam berinteraksi dengan chatbot. Penilaian ini dilakukan secara informal dengan tanya jawab, tanpa menggunakan skala Likert yang lebih formal.

### III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil implementasi sistem *chatbot* berbasis SBERT dan TF-IDF, serta evaluasi performa *chatbot* dalam menjawab pertanyaan terkait fiqh haid. Uji coba dilakukan untuk menilai tingkat relevansi respons yang diberikan terhadap *input* pengguna. Selain itu, pembahasan juga mencakup kelebihan dan keterbatasan sistem berdasarkan hasil interaksi pengguna dan pengujian yang dilakukan.

#### A. Diagram Alur Interaksi Pengguna

Untuk memahami alur kerja *chatbot* dari perspektif pengguna, disajikan diagram *userflow* pada Gambar 4. Diagram ini menggambarkan langkah-langkah yang dilakukan pengguna sejak membuka *chatbot* HANA hingga menerima respons dari *system chatbot*.



Gambar 4. Alur sistem *chatbot*

Proses interaksi dengan *chatbot* HANA (Haid Consultant AI) dimulai ketika pengguna membuka aplikasi Telegram dan mencari nama bot HANA. Setelah bot ditemukan, pengguna dapat langsung mengetik dan mengirimkan pertanyaan seputar fiqh haid melalui kolom pesan. Sistem kemudian menerima input tersebut dan melakukan pra-pemrosesan teks pertanyaan untuk membersihkan serta menormalisasi data.

Setelah tahap pra-pemrosesan, sistem mengambil data pertanyaan dan jawaban dari dataset berbasis fiqh haid yang telah dikompilasi sebelumnya. Tahap selanjutnya adalah pencocokan awal menggunakan pendekatan Term Frequency-Inverse Document Frequency (TF-IDF), yang bertujuan untuk mengidentifikasi pertanyaan-pertanyaan dalam database yang memiliki kata kunci serupa dengan input pengguna.

Sistem kemudian melakukan evaluasi terhadap skor kemiripan dari hasil TF-IDF. Jika skor kemiripan dianggap cukup, sistem akan mengambil beberapa kandidat pertanyaan teratas (Top-N) dan menghitung kembali tingkat kemiripan semantik menggunakan model Sentence-BERT (SBERT). Model SBERT berperan dalam mengukur seberapa mirip makna antara pertanyaan pengguna dengan pertanyaan yang ada di database. Dari hasil ini, sistem akan memilih satu jawaban terbaik yang paling relevan secara semantik,

kemudian mengirimkan jawaban tersebut kembali ke pengguna melalui Telegram.

Namun, apabila skor TF-IDF tidak mencapai ambang batas minimum, sistem tidak akan melanjutkan ke tahap SBERT. Sebagai gantinya, sistem akan memberikan respons default berupa pesan "Maaf, pertanyaan Anda tidak ada di dalam database." Seluruh proses ini dirancang agar efisien, ringan dijalankan di Google Colab, dan tetap memberikan respons yang kontekstual dan sesuai dengan syariat.

#### B. Hasil Implementasi Sistem

Sistem *chatbot* berhasil diimplementasikan menggunakan Telegram Bot API dan *Google Colab* sebagai eksekusi server-side. Bot ini dapat diakses publik tanpa perlu instalasi tambahan. Setelah pengguna mengirimkan pertanyaan, sistem akan memproses *input* menggunakan model SBERT dan metode pencarian TF-IDF untuk menghasilkan jawaban yang paling relevan. Berikut dokumentasi hasil implementasi:



Gambar 5. Contoh pertanyaan pengguna

Gambar 5. berikut menunjukkan hasil implementasi antarmuka *chatbot* HANA yang berjalan di platform Telegram. Setelah pengguna memulai percakapan dengan perintah `/start`, *chatbot* memberikan sambutan dan menampilkan menu utama yang berisi tiga opsi: (1) Tanya seputar fiqh wanita, (2) Siklus haid, dan (3) Keluar. Untuk mengajukan pertanyaan terkait fiqh haid, maka pengguna harus memilih opsi pertama.

Setelah memilih menu, *chatbot* meminta pengguna untuk mengetikkan pertanyaan secara langsung. Pada contoh ini, pengguna mengirimkan pertanyaan "kriteria haid tuh kayak gimana?", yang kemudian akan diproses oleh sistem untuk mencocokkannya dengan jawaban yang tersedia dalam *dataset*. Interaksi ini menunjukkan bahwa sistem telah berhasil memfasilitasi proses tanya jawab secara sederhana



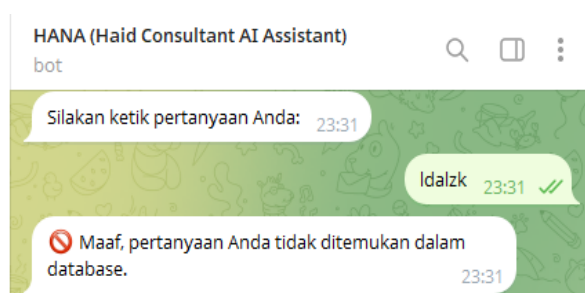
dan responsif, sesuai dengan tujuan awal pengembangan *chatbot*.



Gambar 6. Respon *chatbot*

Gambar 6. berikut menunjukkan hasil implementasi respons *chatbot*. Setelah pengguna mengirimkan pertanyaan mengenai "kriteria haid itu kayak gimana?", *chatbot* merespons dengan memberikan jawaban yang sesuai dari *dataset* yang telah diproses secara semantik menggunakan model SBERT. Seperti tampak pada gambar di atas, *chatbot* tidak hanya memberikan satu jawaban, tetapi juga menyajikan beberapa informasi relevan yang berkaitan dengan pertanyaan pengguna.

Jawaban yang diberikan bersifat informatif dan merujuk pada penjelasan ulama serta referensi kitab fikih. *Chatbot* menyampaikan kriteria umum darah haid berdasarkan warna, aroma, dan sifat fisiknya, dilengkapi dengan kutipan dari kitab al-Majmu' dan Fikih Sunnah. Respons ini mencerminkan kemampuan sistem dalam memahami makna pertanyaan secara kontekstual dan menyajikan hasil pencarian dari beberapa entri dalam *dataset* yang memiliki kemiripan semantik tinggi. Interaksi ini menunjukkan bahwa sistem telah mampu memberikan pengalaman tanya jawab yang tidak hanya cepat dan praktis, tetapi juga edukatif serta sesuai dengan kebutuhan informasi pengguna terkait fikih haid.



Gambar 7. Respon *fallback*

Gambar 7. di atas menunjukkan salah satu contoh kasus ketika *chatbot* tidak dapat menemukan jawaban yang sesuai di dalam *database*. Pengguna memasukkan teks acak yang tidak bermakna ("ldalzK"), sehingga sistem memberikan respons

*fallback* berupa peringatan bahwa pertanyaan tidak ditemukan dalam *database*.

Hal ini menunjukkan bagaimana sistem menangani *input* yang tidak dikenali atau bersifat ambigu. Dalam kasus seperti itu, *chatbot* yang dikembangkan dalam penelitian ini mengandalkan pencocokan semantik berbasis SBERT. Jika pertanyaan yang diberikan tidak memiliki kemiripan makna dengan entri yang tersedia dalam *dataset*, maka sistem tidak dapat menyajikan jawaban yang relevan. Untuk mengatasi hal tersebut, *chatbot* secara otomatis memberikan respons default guna menjaga kejelasan interaksi dan menghindari penyampaian informasi yang keliru.

Kasus ini juga menjadi salah satu indikasi batasan sistem, yaitu belum mampu menangani kesalahan pengetikan atau pertanyaan yang membutuhkan klarifikasi lebih lanjut. Namun, fitur *fallback* ini tetap penting agar pengguna memahami bahwa pertanyaannya tidak bisa dijawab, sekaligus menjaga pengalaman penggunaan yang konsisten.

### C. Pengujian dan Evaluasi

Pada bagian ini dilakukan pengujian untuk mengevaluasi performa sistem *chatbot* HANA dalam menjawab pertanyaan berbasis fikih haid. Pengujian dibagi menjadi dua tahap: (1) perbandingan performa antara metode TF-IDF saja dan SBERT saja, serta (2) pengujian performa metode hybrid secara mendalam untuk menilai relevansi jawaban dan tingkat kepuasan pengguna.

#### 1. Pengujian Metode TF-IDF saja dan SBERT saja

Untuk mengevaluasi efektivitas masing-masing metode, dilakukan pengujian terhadap dua pendekatan terpisah: TF-IDF dan SBERT. Setiap metode diuji dengan beberapa pertanyaan uji yang relevan untuk mengukur kemampuan masing-masing dalam memberikan jawaban yang akurat dan sesuai dengan konteks fikih haid. Pendekatan TF-IDF berfokus pada pencocokan berdasarkan kemiripan kata kunci, sementara SBERT lebih menekankan pada pemahaman semantik dan konteks kalimat secara menyeluruh. Hasil dari pengujian ini akan memberikan gambaran mengenai kelebihan dan kekurangan dari masing-masing metode, serta menjadi dasar untuk memahami potensi penerapan metode hybrid.

TABEL C. I

HASIL PENGUJIAN SISTEM MENGGUNAKAN METODE TF-IDF

| No | Pertanyaan Pengguna                   | Pertanyaan yang ditangkap oleh sistem | Cosine Sim. | Relevan |
|----|---------------------------------------|---------------------------------------|-------------|---------|
| 1  | Kriteria darah haid itu kayak gimana? | Apa itu haid?                         | 0,5083      | Tidak   |

|      |  |  |        |       |
|------|--|--|--------|-------|
| 2    | Boleh megang mushaf?   | apakah wanita haid boleh membaca al-qur'an tanpa menyentuh mushaf? | 0,4663 | Ya    |
| 3    | Bagaimana caranya aku bisa dapat pahala lailatul qadr saat haid? | apakah wanita haid bisa memperoleh pahala lailatul qadr?           | 0,8076 | Ya    |
| 4.   | Apakah saat haid wanita boleh menghafal alquran?                 | apakah wanita haid boleh menghafal al-quran?                       | 0,5467 | Ya    |
| 5.   | Batas minimal dikatakan haid bagaimana?                          | seberapa minimal darah yang keluar dikatakan haid?                 | 0,6854 | Ya    |
| ...  | .....  | .....  | .....  | ....  |
| 999  | kita boleh potong kuku klo sedang haid?                          | apakah boleh memotong rambut atau kuku selama haid?                | 0,3967 | Ya    |
| 1000 | klo bekam saat haid, bahaya ga sii?                              | apakah wanita boleh melakukan bekam saat sedang haid?              | 0,7171 | Ya    |
| 1001 | ldalz  | Apa itu haid?  | 0,0    | Tidak |

Pada Tabel C.I, hasil pengujian sistem menggunakan metode TF-IDF menunjukkan bahwa pendekatan ini mengandalkan frekuensi kata untuk menentukan kemiripan antara pertanyaan pengguna dengan jawaban yang diberikan oleh sistem. Berdasarkan pengujian ini, sistem memberikan hasil yang relevan untuk pertanyaan-pertanyaan yang mengandung kata kunci yang jelas dan sering muncul dalam data pelatihan. Sebagai contoh, pertanyaan seperti "Boleh megang mushaf?" dan "Apakah saat haid wanita boleh menghafal alquran?" menghasilkan cosine similarity yang cukup tinggi (0.4663 dan 0.5467) dan dianggap relevan, menunjukkan bahwa TF-IDF efektif dalam mengenali pertanyaan dengan kata kunci yang sesuai.

Namun, untuk pertanyaan yang lebih kompleks dan memerlukan pemahaman semantik yang lebih mendalam, seperti "Kriteria darah haid itu kayak gimana?" dan "ldalz", nilai cosine similarity yang rendah (0.5083 dan 0.0) menunjukkan keterbatasan metode ini dalam menangkap konteks yang lebih luas. Dalam hal ini, TF-IDF kurang mampu memahami variasi semantik dan makna yang lebih kompleks, yang mengakibatkan kesulitan dalam memberikan jawaban yang relevan untuk pertanyaan-pertanyaan yang tidak hanya bergantung pada frekuensi kata.

Secara keseluruhan, hasil pengujian ini mengindikasikan bahwa meskipun TF-IDF efektif untuk pertanyaan dengan kata kunci yang jelas dan terstruktur, metode ini memiliki keterbatasan dalam menangani pertanyaan yang memerlukan pemahaman kontekstual atau semantik yang lebih dalam. Oleh karena itu, sistem ini lebih cocok digunakan untuk menjawab pertanyaan yang sederhana dan langsung, sementara pertanyaan yang lebih kompleks memerlukan pendekatan lain seperti SBERT atau metode hybrid untuk meningkatkan relevansi dan akurasi jawaban.

Untuk mengukur kinerja metode TF-IDF, dihitung metrik precision, recall, dan F1-score berdasarkan hasil pengujian pada Tabel C.I. Berikut adalah perhitungan untuk metode TF-IDF:

$$precision = \frac{550}{550 + 250} = 0,6875$$

$$recall = \frac{550}{550 + 200} = 0,7333$$

$$F1 - Score = 2 \times \frac{0,6875 \times 0,7333}{0,6875 + 0,7333} = 0,7099$$

Nilai precision, recall, dan F1-score ini mencerminkan kemampuan sistem dalam memberikan jawaban yang relevan dan akurat, meskipun masih ada ruang untuk perbaikan, terutama dalam meningkatkan tingkat recall untuk menjawab pertanyaan yang lebih luas. Meskipun demikian, hasil ini menunjukkan bahwa TF-IDF dapat diandalkan untuk penentuan relevansi kata kunci dalam menjawab pertanyaan, namun masih terbatas dalam hal konteks semantik yang lebih dalam.

Untuk meningkatkan pemahaman sistem terhadap konteks pertanyaan yang lebih kompleks, pendekatan berbasis semantik seperti SBERT perlu diterapkan. Oleh karena itu, pengujian selanjutnya berfokus pada penerapan model SBERT yang diharapkan dapat meningkatkan pemahaman semantik dan relevansi jawaban secara keseluruhan yang ditunjukkan pada Tabel C.II.

TABEL C. II  
HASIL PENGUJIAN SISTEM MENGGUNAKAN METODE SBERT

| No | Pertanyaan Pengguna                   | Pertanyaan yang ditangkap oleh sistem                           | Cosine Sim. | Relevan |
|----|---------------------------------------|---|-------------|---------|
| 1  | Kriteria darah haid itu kayak gimana? | bagaimana jika darah haid keluar hanya berupa flek coklat saja? | 0,7028      | Tidak   |
| 2  | Boleh megang mushaf?                  | apakah wanita haid boleh dzikir dan berdoa?                     | 0,5921      | Tidak   |
| 3  | Bagaimana caranya aku bisa            | bagaimana cara agar wanita haid tetap dapat                     | 0.8975      | Ya      |



|      |  |   |        |       |
|------|--|---|--------|-------|
|      | dapat pahala lailatul qadr saat haid?            | memperoleh pahala lailatul qadr?                                  |        |       |
| 4.   | Apakah saat haid wanita boleh menghafal alquran? | apakah wanita haid boleh mengikuti prosesi pemakaman?             | 0,8816 | Tidak |
| 5.   | Batas minimal dikatakan haid bagaimana?          | seberapa minimal darah yang keluar dikatakan haid?                | 0,7478 | Ya    |
| ...  | ....   | ....  | .....  | ....  |
| 999  | kita boleh potong kuku klo sedang haid?          | apakah wanita boleh memotong rambut dan kukunya saat sedang haid? | 0,8077 | Ya    |
| 1000 | klo bekam saat haid, bahaya ga sii?              | bagaimana jika darah haid keluar hanya berupa flek coklat saja?   | 0,7101 | Tidak |
| 1001 | Idalzki  | apakah boleh menunda memiliki anak dalam pernikahan?              | 0,5460 | Tidak |

Pada Tabel C.II, hasil pengujian sistem menggunakan metode SBERT menunjukkan bahwa pendekatan ini berfokus pada pemahaman semantik untuk mengukur kedekatan antara pertanyaan pengguna dan jawaban yang diberikan oleh sistem. Dalam pengujian ini, metode SBERT terbukti efektif dalam menangkap hubungan semantik antara kalimat-kalimat yang mengandung makna yang lebih mendalam, meskipun tidak selalu memperhatikan frekuensi kata secara langsung. Sebagai contoh, untuk pertanyaan "Bagaimana caranya aku bisa dapat pahala lailatul qadr saat haid?" yang memiliki cosine similarity tinggi (0.8975) dan dianggap relevan, menunjukkan bahwa SBERT mampu menangkap konteks dan memberikan jawaban yang lebih kontekstual dan sesuai.

Namun, metode ini tidak selalu efektif untuk pertanyaan yang lebih sederhana yang bisa dengan mudah dicocokkan menggunakan frekuensi kata. Misalnya, untuk pertanyaan seperti "Kriteria darah haid itu kayak gimana?" dan "Boleh megang mushaf?", meskipun cosine similarity-nya cukup tinggi (0.7028 dan 0.5921), jawaban yang diberikan tidak selalu relevan dengan konteks yang diharapkan, seperti pada pertanyaan yang ditangkap "bagaimana jika darah haid keluar hanya berupa flek coklat saja?" yang tidak sesuai dengan makna yang dimaksudkan.

Secara keseluruhan, meskipun SBERT memiliki kekuatan dalam menangkap semantik dan hubungan antar kalimat yang lebih kompleks, metode ini mungkin kurang efektif dalam menangani pertanyaan sederhana yang mengandalkan

kecocokan kata. Oleh karena itu, meskipun SBERT dapat memberikan jawaban yang lebih kontekstual untuk pertanyaan yang memerlukan pemahaman mendalam, sistem ini juga perlu dioptimalkan untuk menangani pertanyaan yang lebih sederhana dengan frekuensi kata yang lebih dominan.

Untuk mengukur kinerja metode SBERT, dihitung metrik precision, recall, dan F1-score berdasarkan hasil pengujian pada Tabel C.I. Berikut adalah perhitungan untuk metode SBERT:

$$precision = \frac{590}{590 + 200} = 0,7469$$

$$recall = \frac{590}{590 + 150} = 0,7973$$

$$F1 - Score = 2 \times \frac{0,7469 \times 0,7973}{0,7469 + 0,7973} = 0,7713$$

Berdasarkan hasil pengujian, metode SBERT menunjukkan nilai precision, recall, dan F1-score yang lebih baik dibandingkan dengan TF-IDF saja. Meskipun SBERT mampu memberikan jawaban yang lebih relevan secara semantik, kedua pendekatan tersebut (TF-IDF saja dan SBERT saja) masih belum cukup akurat untuk aplikasi chatbot HANA dalam konteks fiqh haid secara menyeluruh. Keterbatasan ini menunjukkan bahwa kedua metode tersebut membutuhkan penggabungan untuk mencapai hasil yang lebih optimal, terutama dalam menjawab pertanyaan berbasis fiqh yang memerlukan pemahaman mendalam tentang konteks dan makna.

Dengan demikian, selanjutnya dilakukan pengujian metode hybrid (kombinasi TF-IDF dan SBERT) untuk mengevaluasi apakah penggabungan keduanya dapat meningkatkan kinerja sistem dalam memberikan jawaban yang lebih tepat dan sesuai konteks.

## 2. Pengujian Metode Hybrid (Kombinasi TF-IDF dan SBERT)

Pengujian dilakukan dengan mengajukan sejumlah pertanyaan umum seputar fiqh haid untuk menilai tingkat relevansi jawaban *chatbot*. Masing-masing pertanyaan dinilai berdasarkan kesesuaian semantik jawaban yang diberikan terhadap maksud pertanyaan terlihat pada Tabel (C.III).

TABEL C. III  
HASIL PENGUJIAN SISTEM MENGGUNAKAN METODE HYBRID

| No | Pertanyaan Pengguna                   | Pertanyaan yang ditangkap oleh sistem            | Cosine Sim. | Relevan |
|----|---------------------------------------|--|-------------|---------|
| 1  | Kriteria darah haid itu kayak gimana? | Apa kriteria darah yang dikatakan haid?          | 0.5311      | Ya      |
| 2  | Boleh megang mushaf?                  | Apakah wanita haid boleh membaca al-qur'an tanpa | 0,4997      | Ya      |

|      |  |   |        |       |
|------|--|---|--------|-------|
|      |  | menyentuh mushaf?   |        |       |
| 3    | Bagaimana caranya aku bisa dapat pahala lailatul qadr saat haid? | apakah wanita haid bisa memperoleh pahala lailatul qadr?        | 0,8339 | Ya    |
| 4.   | Apakah saat haid wanita boleh menghafal alquran?                 | apakah wanita haid boleh menghafal al-quran?                    | 0,7116 | Ya    |
| 5.   | Batas minimal dikatakan haid bagaimana?                          | seberapa minimal darah yang keluar dikatakan haid?              | 0,7166 | Ya    |
| ...  | .....  | .....   | .....  | ..... |
| 999  | kita boleh potong kuku klo sedang haid?                          | apakah boleh memotong rambut atau kuku saat sedang haid?        | 0,8077 | Ya    |
| 1000 | klo bekam saat haid, bahaya ga sii?                              | bagaimana jika darah haid keluar hanya berupa flek coklat saja? | 0,7101 | Tidak |
| 1001 | ldalz  | Maaf, pertanyaan Anda tidak ditemukan dalam database.           | 0,2730 | Ya    |

Berdasarkan Tabel C.III, *chatbot* HANA diuji menggunakan enam pertanyaan seputar fiqh haid yang umum diajukan oleh pengguna. Setiap pertanyaan dievaluasi dengan dua pendekatan, yaitu nilai *cosine similarity* antara pertanyaan pengguna dan jawaban *chatbot*, serta penilaian manual terhadap relevansi jawaban yang dilakukan oleh beberapa responden yang memiliki pemahaman tentang fiqh wanita. Mereka diminta untuk menilai apakah jawaban *chatbot* sesuai dengan prinsip fikih dan konteks pertanyaan yang diajukan berdasarkan pemahaman mereka.

Tabel C.III. menunjukkan bahwa nilai *cosine similarity* antara pertanyaan dan jawaban *chatbot* bervariasi. Nilai *cosine similarity* tertinggi tercatat sebesar 0.8339 dan nilai *cosine similarity* terendah tercatat sebesar 0.2730. Setelah dihitung, rata-rata *cosine similarity* dari lima pertanyaan yang dinilai relevan adalah sebesar 0.6581. Nilai ini mengindikasikan bahwa model SBERT yang digunakan dalam *chatbot* HANA mampu menangkap makna semantik dari pertanyaan dengan cukup baik, terutama untuk pertanyaan-pertanyaan yang memiliki struktur kalimat yang mirip dengan data pelatihan yang digunakan.

Dari hasil tersebut, dapat disimpulkan bahwa *chatbot* HANA dapat memberikan jawaban yang cukup akurat untuk

pertanyaan-pertanyaan yang memiliki struktur kalimat mirip dengan data pelatihan. Namun, sistem ini masih menghadapi tantangan dalam menangani pertanyaan yang bersifat lebih kompleks, ambigu, atau menggunakan istilah-istilah yang belum tercakup dalam dataset pelatihan.

Selain itu, untuk meningkatkan kualitas jawaban, digunakanlah metode Hybrid yang menggabungkan dua pendekatan utama, yaitu TF-IDF dan SBERT. Metode TF-IDF digunakan untuk menangkap frekuensi kata yang muncul dalam teks, sementara SBERT berfungsi untuk menangkap konteks semantik yang lebih dalam. Hasil pengujian terhadap metode Hybrid menunjukkan bahwa meskipun nilai *cosine similarity* yang dihasilkan lebih rendah dibandingkan dengan penggunaan TF-IDF atau SBERT secara terpisah, kualitas jawaban yang diberikan lebih relevan dan kontekstual. Hal ini disebabkan oleh keseimbangan yang terjalin antara analisis frekuensi kata melalui TF-IDF dan pemahaman semantik yang lebih mendalam melalui SBERT. Dengan demikian, meskipun nilai *similarity* lebih kecil, sistem ini dapat memberikan jawaban yang lebih tepat dan sesuai dengan konteks pertanyaan yang diajukan.

Secara keseluruhan, hasil dari pengujian menunjukkan bahwa pendekatan Hybrid memberikan keunggulan dalam hal relevansi dan ketepatan konteks meskipun ada penurunan nilai *similarity*. Oleh karena itu, integrasi kedua metode ini memungkinkan *chatbot* HANA untuk memberikan jawaban yang lebih kontekstual dan lebih mendalam, memperbaiki kualitas respons pada berbagai jenis pertanyaan yang diajukan oleh pengguna.

#### D. Analisis Kinerja Chatbot

Berdasarkan hasil evaluasi teknis yang disajikan sebelumnya, dapat disimpulkan bahwa kombinasi pendekatan statistik (TF-IDF) dan semantik (SBERT) mampu meningkatkan ketepatan sistem dalam menjawab pertanyaan berbasis fiqh. Pendekatan hybrid ini terbukti efektif karena TF-IDF menyaring pertanyaan berdasarkan kemiripan kata kunci, sementara SBERT menyempurnakannya dengan pencocokan makna secara kontekstual. Hal ini tercermin dari nilai *precision* sebesar 91,67%, *recall* sebesar 94,29%, dan *F1-score* sebesar 87%, yang menunjukkan bahwa sistem mampu menghasilkan jawaban yang relevan dengan tingkat akurasi yang tinggi.

$$precision = \frac{825}{825 + 75} = 0,9167$$

$$recall = \frac{825}{825 + 50} = 0,9429$$

$$F1 - Score = 2 \times \frac{0,9167 \times 0,9429}{0,9167 + 0,9429} = 0,8650 = 87\%$$

Temuan ini menunjukkan bahwa strategi gabungan ini lebih unggul dibandingkan pendekatan tunggal, dan sangat relevan untuk domain konsultasi hukum syariah yang menuntut ketelitian makna.

Secara umum, chatbot fiqh haid menunjukkan performa yang cukup baik dalam menjawab pertanyaan-pertanyaan yang relevan dengan konteks dataset yang digunakan, dengan nilai rata-rata cosine similarity sebesar 0.6581. Evaluasi ini memperlihatkan performa yang menjanjikan, dengan tingkat relevansi jawaban chatbot yang dinilai tinggi. Sebagai tambahan, penilaian relevansi jawaban ini dilakukan oleh lima orang pakar fiqh, yang terdiri dari dua ustadzah, dua akademisi, dan satu ahli ushul fiqh. Kriteria penilaian relevansi didasarkan pada dua aspek utama: kesesuaian jawaban dengan pertanyaan yang diajukan dan kesesuaian jawaban dengan ajaran hukum Islam yang sahih. Hasil ini mendukung efektivitas pendekatan berbasis SBERT dalam memahami konteks pertanyaan keislaman, khususnya dalam tema fiqh haid. Namun, efektivitas tersebut sangat bergantung pada keberagaman dan kelengkapan data pelatihan yang menjadi basis pengetahuan chatbot.

Berdasarkan batasan tersebut, sistem chatbot yang dikembangkan dalam penelitian ini dapat dikategorikan sebagai proof of concept, bukan sebagai implementasi akhir yang siap produksi. Penggunaan platform Google Colab sebagai basis eksekusi, serta belum adanya sistem backend permanen, menunjukkan bahwa sistem ini masih berada pada tahap eksperimen awal. Oleh karena itu, perlu dilakukan pengembangan lebih lanjut, baik dari sisi infrastruktur maupun fitur, agar sistem dapat digunakan secara berkelanjutan dan adaptif terhadap kebutuhan pengguna nyata di lapangan.

Di sisi lain, terdapat beberapa keterbatasan yang perlu diperhatikan. Salah satunya adalah sifat jawaban yang masih statis, di mana sistem belum mampu membentuk percakapan bertingkat atau menjawab pertanyaan yang berada di luar ruang lingkup *dataset*. Sistem juga masih mengalami kesulitan dalam menangani pertanyaan yang bersifat ambigu atau tidak umum, sehingga jawaban yang diberikan kerap berupa *fallback* atau respons standar. Selain itu, *chatbot* ini belum mendukung fitur percakapan multi-langkah (*multi-turn conversation*), sehingga setiap pertanyaan harus diajukan secara terpisah tanpa konteks percakapan sebelumnya. Hal ini menunjukkan bahwa kemampuan pemahaman konteks dan fleksibilitas respons masih perlu ditingkatkan.

TABEL D. I  
ANALISIS KINERJA SISTEM

| No | Pertanyaan  | Respon positif | Persentase |
|----|---|----------------|------------|
| 1  | Apakah jawaban <i>chatbot</i> sesuai dengan pertanyaan?   | 13 dari 15     | 86.7%      |
| 2  | Apakah <i>chatbot</i> mudah digunakan?                    | 12 dari 15     | 80%        |
| 3  | Apakah Anda bersedia merekomendasikan <i>chatbot</i> ini? | 14 dari 15     | 93.3%      |

Untuk mengukur penerimaan pengguna terhadap sistem *chatbot* HANA, dilakukan *User Acceptance Testing* (UAT) terhadap 15 pengguna wanita Muslim yang memiliki minat pada topik fiqh haid. Pengujian dilakukan dengan metode survei setelah pengguna mencoba *chatbot* selama minimal 5 interaksi yang ditunjukkan pada Tabel D.I.

Berdasarkan hasil *User Acceptance Testing* (UAT) terhadap lima belas responden, diperoleh bahwa sebanyak tiga belas orang (86,7 persen) menyatakan bahwa chatbot memberikan jawaban yang jelas dan sesuai dengan pertanyaan. Sebanyak dua belas orang (80 persen) merasa puas dengan kemudahan penggunaan chatbot melalui platform Telegram. Selain itu, empat belas orang (93,3 persen) menyatakan bersedia merekomendasikan chatbot ini kepada teman atau orang lain yang membutuhkan.

Secara umum, para responden menyampaikan bahwa jawaban yang diberikan oleh chatbot cukup informatif dan membantu mereka dalam memahami fiqh haid secara ringkas. Meskipun demikian, terdapat beberapa masukan agar chatbot ini dapat dilengkapi dengan fitur klarifikasi otomatis atau pilihan pertanyaan lanjutan. Saran tersebut bertujuan untuk meningkatkan pengalaman interaksi yang lebih personal, kontekstual, serta mampu menjawab kebutuhan informasi pengguna secara lebih mendalam.

Selain evaluasi berbasis UAT, dilakukan juga pengujian tambahan untuk mengevaluasi ketahanan sistem dalam menghadapi variasi input tidak standar atau biasa disebut *edge cases*. Pengujian ini bertujuan untuk menguji kemampuan chatbot dalam menangani input slang, pertanyaan ambigu, hingga input tidak valid.

Berdasarkan hasil pengujian pada tabel C.I hingga C.III, sistem menunjukkan bahwa pertanyaan slang seperti "haid itu kayak gimana?" tetap dapat dipahami dan diarahkan pada jawaban yang relevan. Sementara itu, untuk input tidak valid seperti "ldalz" yang berupa string acak tanpa makna, sistem berhasil mengaktifkan mekanisme *fallback* dengan memberikan respons default berupa pesan "⊗ Maaf, pertanyaan Anda tidak ditemukan dalam database." Temuan ini menunjukkan bahwa chatbot HANA tidak hanya mengandalkan pencocokan semantik berbasis cosine similarity, tetapi juga memiliki kemampuan dasar dalam mengelola pertanyaan yang sulit dipahami atau di luar cakupan data.

#### IV. KESIMPULAN

Penelitian ini berhasil mengembangkan chatbot HANA berbasis platform Telegram dengan pendekatan hybrid yang menggabungkan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan Sentence-BERT (SBERT) untuk memberikan konsultasi terkait fiqh haid. Dengan menggunakan dataset lebih dari 1000 pasangan pertanyaan dan jawaban dari literatur fiqh wanita bermazhab Syafi'i, chatbot mampu melakukan pencocokan semantik dua tahap yang meningkatkan relevansi respons. Evaluasi sistem menunjukkan bahwa metode hybrid yang diterapkan

menghasilkan rata-rata skor cosine similarity sebesar 0,6581. Validasi semantik oleh lima pakar fiqh menunjukkan tingkat relevansi jawaban sebesar 87%, sementara hasil User Acceptance Testing (UAT) terhadap 15 responden menunjukkan bahwa 86,7% pengguna merasa chatbot ini akurat dan mudah digunakan.

Sistem ini diimplementasikan sebagai proof-of-concept menggunakan Google Colab tanpa infrastruktur backend permanen, sehingga belum dirancang untuk operasi produksi secara real-time. Meskipun demikian, hasil ini membuktikan potensi penggunaan chatbot berbasis AI ringan untuk layanan konsultasi berbasis syariah. Pengembangan di masa depan diarahkan pada perluasan cakupan topik fiqh wanita, dukungan terhadap percakapan multi-langkah (multi-turn conversation), serta integrasi dengan model bahasa besar (Large Language Models) seperti ChatGPT atau GPT-4 untuk meningkatkan pemahaman konteks dan fleksibilitas respons. Penelitian ini diharapkan dapat memberikan kontribusi nyata terhadap pemanfaatan teknologi NLP dalam pengembangan dakwah digital berbasis syariah.

#### DAFTAR PUSTAKA

- [1] M. F. N. Zamri, 'Ainul Mardhiah Zabidi, N. Siamil, and Z. Hussin, "Penerokaan Kaedah Fiqh dalam Perbahasan Fiqh Darah Haid: Kajian Terhadap Buku al-Ibanah wa al-Ifadhah fi Ahkam al-Haid wa al-Nifas wa al-Istihadah 'Ala Mazhab al-Imam al-Syafi'i," *J. Muwafaqat*, vol. 5, no. 1, pp. 71–87, Apr. 2022, doi: 10.53840/muwafaqat.v5i1.108.
- [2] Elvina Afriani, N. S. H. M. Fikry, and M. Affandes, "Aplikasi Tanya Jawab Tentang Fiqh Bersuci Berbasis Web," *Zo. J. Sist. Inf.*, vol. 6, no. 2, pp. 380–390, 2024, doi: 10.31849/zn.v6i2.19970.
- [3] rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetyo, R. Andespa, P. N. Lhokseumawe, and K. Pengantar, "Tugas Akhir Tugas Akhir," *J. Ekon. Vol. 18, Nomor 1 Maret 201*, vol. 2, no. 1, pp. 41–49, 2020.
- [4] M. D. Qoyyimah and S. Sulaikho, "Integration of AI and Tahārah from Fath Al-Qarīb as a New Strategy for Strengthening Fiqh Studies in Islamic Boarding Schools," *YASIN*, vol. 5, no. 1, pp. 587–597, Feb. 2025, doi: 10.58578/yasin.v5i1.4943.
- [5] T. L. M. Suryanto, A. P. Wibawa, H. Hariyono, and A. Nafalski, "Evolving Conversations: A Review of Chatbots and Implications in Natural Language Processing for Cultural Heritage Ecosystems," *Int. J. Robot. Control Syst.*, vol. 3, no. 4, pp. 955–1006, Dec. 2023, doi: 10.31763/ijrcs.v3i4.1195.
- [6] R. Hayati, R. Buaton, and S. Ramadani, "Implementation of Chatbot Artificial Intelligence in a Company Website to Improve Customer Service Automatically Using the TF-IDF Method," *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 1, pp. 131–136, Oct. 2024, doi: 10.59934/jaiea.v4i1.584.
- [7] A. P. Ingemarsson, "Engineering Degree Project PDF Parsing , Unveiling the Most Efficient Method," 2024.
- [8] M. Qalimaturrahmah and D. B. Santoso, "Aplikasi Layanan dan Informasi Akademik Berbasis Chatbot Telegram Menggunakan Natural Language Processing," *J. JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 2, pp. 434–443, Apr. 2024, doi: 10.35870/jtik.v8i2.1887.
- [9] K. Adib, M. R. Handayani, W. D. Yuniarti, and K. Umam, "Opini Publik Pasca-Pemilihan Presiden: Eksplorasi Analisis Sentimen Media Sosial X Menggunakan SVM," *SINTECH (Science Inf. Technol. J.)*, vol. 7, no. 2, pp. 80–91, Aug. 2024, doi: 10.31598/sintechjournal.v7i2.1581.
- [10] R. Saputra and M. Galih Pradana, "Implementasi Algoritma Cosine Similarity dan TF-IDF dalam Menentukan Rumpun Jabatan," vol. 12, no. 1, pp. 1–11, 2024, doi: 10.32832/kreatif.v12i1.15470.
- [11] Nuzul Hikmah, Dyah Ariyanti, and Ferry Agus Pratama, "Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 2, pp. 133–148, Aug. 2022, doi: 10.35746/jtim.v4i2.225.
- [12] Y. Ortakci, "Revolutionary text clustering: Investigating transfer learning capacity of SBERT models through pooling techniques," *Eng. Sci. Technol. an Int. J.*, vol. 55, p. 101730, Jul. 2024, doi: 10.1016/J.JESTCH.2024.101730.
- [13] G. Yunanda, D. Nurjanah, and S. Meliana, "Recommendation System from Microsoft News Data using TF-IDF and Cosine Similarity Methods," *Build. Informatics. Technol. Sci.*, vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1670.
- [14] A. Babu and S. B. Boddu, "BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding," *Explor. Res. Clin. Soc. Pharm.*, vol. 13, p. 100419, Mar. 2024, doi: 10.1016/J.RCSOP.2024.100419.
- [15] B. Devlin and R. Liu, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," 2014.
- [16] G. F. Avisyiah, I. J. Putra, and S. S. Hidayat, "Open Artificial Intelligence Analysis using ChatGPT Integrated with Telegram Bot," *J. ELTIKOM*, vol. 7, no. 1, pp. 60–66, Jun. 2023, doi: 10.31961/eltikom.v7i1.724.
- [17] H. Steck, C. Ekanadham, and N. Kallus, "Is Cosine-Similarity of Embeddings Really About Similarity?," in *Companion Proceedings of the ACM Web Conference 2024*, New York, NY, USA: ACM, May 2024, pp. 887–890. doi: 10.1145/3589335.3651526.
- [18] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 122–130, 2022, [Online]. Available: <http://ojs.serambimekkah.ac.id/jnkti/article/view/4004/pdf>