

## Pengembangan Chatbot Berbasis AI untuk Mendukung Pelayanan Perpustakaan

### *Development of AI-Based Chatbot to Support Library Services*

Muhammad Ahsanu Qaulan<sup>1\*</sup>, Wahyuni<sup>2</sup>, Pitrasacha Adytia<sup>3</sup>

<sup>1,2</sup>Teknik Informatika, STMIK Widya Cipta Dharma

<sup>3</sup>Sistem Informasi, STMIK Widya Cipta Dharma

<sup>1</sup>2143028@wicida.ac.id, <sup>2</sup>wahyuni@wicida.ac.id, <sup>3</sup>pitra@wicida.ac.id

#### Abstract

*The limitations of library staff in handling a large number of inquiries from students often result in delays in service and information delivery in academic libraries. This research addresses the problem by developing an artificial intelligence (AI)-based chatbot to support library information services at STMIK Widya Cipta Dharma. The chatbot development was conducted using CRISP-DM method and LLM technology (Llama3.2) integrated through Retrieval-Augmented Generation (RAG) approach, by applying semantic vector embedding through SentenceTransformer and efficient search with FAISS indexing. The dataset used consists of 11 question-answer pairs related to library services, which are then subjected to preprocessing, vector embedding, and document search using FAISS. Evaluation was conducted using the BERTScore metric to measure semantic similarity between chatbot answers and references, with an average precision of 0.6513, recall of 0.7924, and F1-Score of 0.7124. These values indicate that the chatbot has good semantic capabilities in answering general questions related to library services, although the limited amount of training data indicates that these results are preliminary and still require further development to improve accuracy on complex questions.*

*Keywords: BERTScore, Chatbot, CRISP-DM, Library Services, LLM*

#### Abstrak

Keterbatasan staf perpustakaan dalam menangani sejumlah besar pertanyaan dari mahasiswa sering mengakibatkan penundaan layanan dan penyampaian informasi di perpustakaan akademik. Penelitian ini mengatasi masalah tersebut dengan mengembangkan chatbot berbasis kecerdasan buatan (AI) untuk mendukung layanan informasi perpustakaan di STMIK Widya Cipta Dharma. Pengembangan chatbot dilakukan dengan metode CRISP-DM dan teknologi LLM (Llama3.2) yang diintegrasikan melalui pendekatan Retrieval-Augmented Generation (RAG), dengan menerapkan embedding vektor semantik melalui SentenceTransformer dan pencarian efisien dengan pengindeksan FAISS. Dataset yang digunakan terdiri dari 11 pasangan pertanyaan-jawaban terkait layanan perpustakaan, yang kemudian dilakukan proses preprocessing, embedding vektor, dan pencarian dokumen menggunakan FAISS. Evaluasi dilakukan menggunakan metrik BERTScore untuk mengukur kesamaan semantik antara jawaban chatbot dan referensi, dengan hasil rata-rata precision sebesar 0.6513, recall sebesar 0.7924, dan F1-Score sebesar 0.7124. Nilai tersebut menunjukkan bahwa chatbot memiliki kemampuan semantik yang baik dalam menjawab pertanyaan umum terkait layanan perpustakaan, meskipun keterbatasan jumlah data latihan menunjukkan bahwa hasil ini bersifat awal dan masih memerlukan pengembangan lebih lanjut untuk meningkatkan akurasi pada pertanyaan yang kompleks.

Kata kunci: BERTScore, Chatbot, CRISP-DM, Layanan Perpustakaan, LLM

#### 1. Pendahuluan

Perpustakaan merupakan salah satu unit penunjang dalam kegiatan pembelajaran di lingkungan perguruan tinggi. Secara umum keberadaan perpustakaan adalah memberikan layanan kepada pengguna, menumbuhkan minat baca, serta memperkaya wawasan dan pengetahuan guna mencerdaskan kehidupan bangsa [1]. Di STMIK Widya Cipta Dharma (Wicida), terdapat perpustakaan digital sebagai layanan untuk mengakses segala informasi terkait perpustakaan secara online.

Akan tetapi, untuk layanan menjawab pertanyaan dari mahasiswa masih secara manual, sehingga staf mengalami kesulitan dalam menjawab pertanyaan yang begitu banyak [2].

Untuk mengatasi permasalahan tersebut, dibutuhkan sistem yang dapat menjawab pertanyaan secara otomatis. Pada penelitian ini, dibangun sebuah chatbot berbasis kecerdasan buatan (AI) untuk menjawab pertanyaan mahasiswa terkait layanan yang ada di perpustakaan [3]. Chatbot adalah suatu program yang

dapat melakukan percakapan melalui media tulisan atau pesan yang dapat merespon suatu perintah yang diberikan [4].

Chatbot yang dibangun dalam penelitian ini menggunakan bahasa pemrograman Python karena penggunaannya yang terbilang mudah dan sesuai dalam penelitian ini [5]. Selain itu juga dengan memanfaatkan teknologi Large Language Model (LLM), yang memungkinkan chatbot untuk memahami konteks pertanyaan pengguna secara lebih akurat dan menyusun respons yang logis dan relevan dalam bentuk percakapan [6]. LLM dilengkapi dengan arsitektur jaringan syaraf tiruan khusus yang disebut Transformers, yang memungkinkan mereka untuk melaksanakan tugas-tugas bahasa yang kompleks. LLM dilatih pada dataset besar yang terdiri dari milyaran kata dari berbagai sumber seperti situs web dan buku [7].

Proses pengembangan chatbot dilakukan dengan menerapkan metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining), yaitu sebuah pendekatan yang bersifat fleksibel dan banyak diterapkan dalam pengembangan sistem berbasis data mining maupun kecerdasan buatan [8]. Pada penelitian ini, CRISP-DM terdiri atas enam tahap yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment [9].

Untuk mengukur kualitas respons yang dihasilkan oleh chatbot, digunakan metode evaluasi BERTScore. BERTScore merupakan metrik evaluasi berbasis model transformator seperti BERT yang digunakan untuk mengukur kesamaan semantik antara teks hasil keluaran chatbot dan jawaban referensi [10]. Tidak seperti metrik konvensional berbasis n-gram, BERTScore menghitung kesamaan konteks dengan membandingkan representasi vektor dari setiap token dalam kalimat, sehingga mampu menangkap nuansa makna yang lebih dalam dan fleksibel [11]. Evaluasi dilakukan dengan tiga metrik utama yaitu precision, recall, dan F1-score terhadap pasangan kalimat prediksi dan referensi. Hasil pengujian dalam berbagai studi menunjukkan bahwa BERTScore mampu memberikan penilaian yang lebih akurat terhadap koherensi semantik, menjadikannya metode yang relevan untuk menilai kecocokan jawaban dalam sistem chatbot berbasis bahasa alami [12].

Sejumlah penelitian sebelumnya telah membuktikan bahwa chatbot berbasis AI mampu meningkatkan kualitas layanan informasi di institusi pendidikan. Salah satunya adalah pengembangan WIDYA, chatbot untuk layanan perpustakaan berbasis AIML yang diterapkan di Institut Teknologi Kesehatan Malang. Hasilnya menunjukkan chatbot tersebut mampu memberikan informasi secara cepat dan akurat, dengan tingkat akurasi mencapai 90%, serta membantu mengurangi beban kerja pustakawan [13].

Selain itu, kajian lain menyoroti peran chatbot dalam membangun ekosistem smart library, yaitu layanan perpustakaan yang tidak hanya responsif secara teknis tetapi juga terpersonalisasi sesuai kebutuhan pengguna. Dalam konteks ini, chatbot menjadi bagian penting dari transformasi layanan menuju pendekatan yang lebih berpusat pada pengguna (*user-centered service*) [14].

Temuan serupa juga ditemukan di lingkungan sekolah, di mana chatbot berbasis AI dinilai efektif dalam menyampaikan informasi, memberikan respons instan, dan selalu siap sedia 24 jam. Meskipun konteksnya berbeda, namun tantangan dan kebutuhan informasi serupa juga terjadi di lingkungan kampus, terutama di perpustakaan [15].

Berdasarkan uraian beberapa penelitian sebelumnya, berikut disajikan perbandingan antara penelitian yang dilakukan dengan penelitian terkait lainnya untuk menunjukkan letak kebaruan yang diusulkan dalam penelitian ini.

Tabel 1. Kebaruan Penelitian

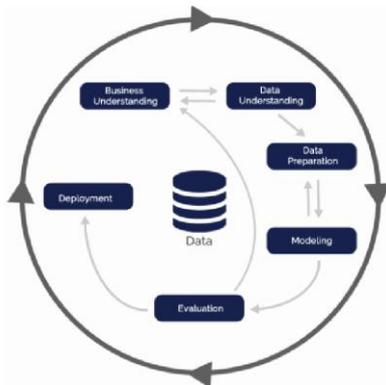
Aspek	Penelitian Ini	Penelitian [13]	Penelitian [14]	Penelitian [15]
Fokus Utama	Layanan informasi perpustakaan berbasis AI dengan kemampuan semantik tinggi	Chatbot AIML untuk layanan tanya jawab perpustakaan	Konsep smart library dengan pendekatan an user-centered	Chatbot sekolah untuk layanan umum dan akademik
Pendekatan	RAG berbasis LLM (Llama3.2) dan embedding vektor menggunakan SentenceTransformer	AIML (rule-based), tanpa pemrosesan semantik	Pendekatan umum tanpa integrasi NLP modern	Dialog statis dan rule-based
Metode Tambahan	Evaluasi performa semantik dengan BERTScore (Precision, Recall, F1)	Evaluasi berdasarkan akurasi manual	Evaluasi konseptual dan kualitatif	Tidak menyertakan evaluasi performa
Sumber Pengetahuan	Dataset pertanyaan-jawaban layanan perpustakaan (11 pasang), berbasis konteks kampus lokal	Dokumen statis berbasis layanan umum	Data layanan dan ekosistem smart library	Informasi sekolah dan akademik dasar
Kebaruan (Novelty)	Integrasi teknologi LLM dengan RAG; Evaluasi otomatis berbasis semantik	Tidak mendukung pemahaman konteks	Belum mengarah pada sistem cerdas	Tidak ada sentuhan teknologi NLP modern
Keunggulan Utama	Kontekstual, responsif, siap integrasi LMS	Cepat namun terbatas pada pola baku	Strategis tapi belum operasional	24 jam aktif, namun tidak relevan di kampus

Dari perbandingan pada Tabel 1 dapat disimpulkan bahwa kebaruan utama dalam penelitian ini terletak pada penggunaan model LLM berbasis RAG untuk menangani pertanyaan layanan perpustakaan secara kontekstual, serta penerapan evaluasi otomatis dengan BERTScore yang jarang digunakan dalam penelitian

sejenis. Pendekatan ini diharapkan dapat menjadi solusi inovatif dalam mengatasi beban kerja pustakawan serta meningkatkan kualitas layanan informasi di lingkungan perguruan tinggi.

## 2. Metode Penelitian

Penelitian ini bertujuan untuk merancang dan menilai kinerja chatbot berbasis kecerdasan buatan (AI) yang mampu menyampaikan informasi terkait layanan perpustakaan di STMIK Widya Cipta Dharma. Metode yang diterapkan dalam penelitian ini mengacu pada pendekatan CRISP-DM. Gambar 1 adalah gambaran alur proses tahapan CRISP-DM:



Gambar 1. Alur CRISP-DM

CRISP-DM terdiri dari enam tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment [16]. Berikut adalah penjelasan rinci dari masing-masing tahapan tersebut.

### 2.1 Business Understanding

Tahapan ini bertujuan untuk memahami kebutuhan bisnis dan tujuan utama dari pengembangan sistem [17]. Dalam penelitian ini, kebutuhan utama yang diidentifikasi adalah kebutuhan akan layanan informasi perpustakaan yang lebih cepat dan efisien, mengingat banyaknya pertanyaan yang diterima oleh staf perpustakaan setiap harinya.

Untuk mengatasi permasalahan tersebut, dibutuhkan sistem otomatis berupa chatbot untuk menangani pertanyaan tersebut secara real-time dan mengurangi beban kerja staf. Penggunaan chatbot diharapkan mampu meningkatkan kualitas layanan informasi kepada mahasiswa serta menciptakan pengalaman pengguna yang lebih responsif dan interaktif.

### 2.2 Data Understanding

Tahap ini melibatkan pengumpulan dan eksplorasi awal terhadap data yang akan digunakan. Dengan cara mengumpulkan data berupa pertanyaan yang sering diajukan oleh mahasiswa kepada staf perpustakaan.

Data diperoleh melalui observasi langsung dan wawancara dengan staf [18].

Proses ini bertujuan untuk memahami struktur, isi, serta konteks dari pertanyaan-pertanyaan yang ada. Selain itu, eksplorasi data juga membantu mengidentifikasi pola dan frekuensi pertanyaan tertentu, sehingga dapat menjadi dasar yang kuat dalam membentuk skenario interaksi chatbot secara lebih relevan dan natural.

### 2.3 Data Preparation

Setelah data berhasil dikumpulkan, langkah selanjutnya adalah mempersiapkan data agar dapat digunakan dalam proses pelatihan dan pemodelan. Tahapan ini mencakup pembersihan teks untuk menghilangkan karakter yang tidak diperlukan, tokenisasi untuk memecah kalimat menjadi bagian-bagian yang lebih kecil, dan penghapusan stopwords [19].

Kemudian, pasangan pertanyaan-jawaban disusun dalam format JSON dan diubah menjadi bentuk DataFrame menggunakan library *Pandas*. Selanjutnya, data ini diproses menggunakan model *SentenceTransformer* untuk menghasilkan representasi vektor yang akan digunakan dalam proses pencocokan dan penarikan informasi oleh chatbot.

### 2.4 Modeling

Pada tahap ini, pemodelan tidak dilakukan dengan melatih model dari awal, melainkan dengan memanfaatkan model *Large Language Model (LLM)* yang telah dilatih sebelumnya. Pendekatan yang digunakan adalah *Retrieval-Augmented Generation (RAG)*, yang memungkinkan sistem menghasilkan jawaban dengan menggabungkan hasil pencarian informasi dan kemampuan generatif dari LLM [20].

Komponen utama dalam sistem ini meliputi representasi embedding menggunakan *SentenceTransformer*, penyimpanan vektor menggunakan *FAISS* (Facebook AI Similarity Search), serta integrasi dengan pipeline LLM melalui API. Dengan pendekatan ini, chatbot dapat merespons pertanyaan pengguna secara lebih relevan dan kontekstual berdasarkan pengetahuan yang tersedia.

### 2.5 Evaluation

Evaluasi dilakukan dengan menggunakan metode evaluasi *BERTScore*. Metode ini digunakan untuk mengukur kualitas jawaban yang dihasilkan oleh chatbot dengan membandingkannya terhadap jawaban referensi. *BERTScore* menghitung kesamaan semantic antara dua kalimat menggunakan representasi vector dari model BERT [21].

Rumus *BERTScore* dapat dilihat pada Persamaan 1:

$$P = \frac{1}{|x|} \sum_{x_i \in x} \max_{y_j \in y} \text{sim}(x_i, y_j) \quad (1)$$

Precision mengukur seberapa baik setiap token dalam kalimat kandidat ( $x$ ) cocok dengan token terbaik dalam kalimat referensi ( $y$ ). Untuk setiap token dalam kalimat kandidat, BERTScore mencari token dalam referensi yang memiliki nilai kesamaan (*similarity*) tertinggi [22].

$$R = \frac{1}{|y|} \sum_{y_j \in y} \max_{x_i \in x} \text{sim}(y_j, x_i) \quad (2)$$

Recall menggambarkan seberapa banyak kelas positif yang diprediksi dengan benar. Recall menggunakan rumus pada Persamaan 2[23].

$$F1 = \frac{2PR}{P + R} \quad (3)$$

F1-Score seperti dapat dilihat pada Persamaan 3 merupakan nilai *harmonic mean* dari *precision* dan *recall*, yang menyeimbangkan kedua metrik tersebut untuk memberikan evaluasi yang lebih komprehensif tentang kecocokan keseluruhan antara kalimat kandidat dan referensi [24].

Nilai precision, recall, dan F1-score berada dalam rentang 0 hingga 1 (atau 0% hingga 100%), dengan nilai yang lebih tinggi menunjukkan performa yang lebih baik [25]. Berdasarkan kategorisasi yang umum digunakan dalam evaluasi model:

0.0-0.2 (0-20%): Sangat buruk. Model praktis tidak dapat membedakan kelas positif dan negatif dengan baik.

0.2-0.4 (20-40%): Buruk. Model memiliki kemampuan klasifikasi yang lemah.

0.4-0.6 (40-60%): Cukup. Model memiliki kemampuan klasifikasi moderat tetapi masih memerlukan perbaikan.

0.6-0.8 (60-80%): Baik. Model memiliki kemampuan klasifikasi yang substansial.

0.8-0.9 (80-90%): Sangat baik. Model memiliki kemampuan klasifikasi yang hampir optimal.

0.9-1.0 (90-100%): Sempurna. Model memiliki kemampuan klasifikasi yang sangat akurat.

Dengan kategori penilaian ini, dapat dengan mudah menginterpretasikan apakah skor yang dihasilkan menunjukkan kualitas jawaban yang baik atau tidak.

### 3. Hasil dan Pembahasan

Sesuai dengan metodologi CRISP-DM yang digunakan, pada bagian ini akan dijelaskan secara rinci tahapan-tahapan yang telah dilakukan dalam penelitian. Penelitian ini difokuskan hingga tahap evaluasi, di mana pada tahap tersebut diperoleh hasil akhir dari performa chatbot dalam menjawab pertanyaan layanan perpustakaan.

#### 3.1 Business Understanding

Dilakukan identifikasi kebutuhan dan permasalahan utama yang dihadapi oleh perpustakaan STMIK Widya Cipta Dharma. Berdasarkan hasil observasi dan wawancara dengan pihak pengelola perpustakaan, ditemukan bahwa meskipun perpustakaan telah memiliki sistem perpustakaan digital untuk akses informasi secara online, layanan untuk menjawab pertanyaan dari mahasiswa masih dilakukan secara manual. Hal ini menyebabkan staf perpustakaan mengalami kesulitan dalam menangani jumlah pertanyaan yang besar dari mahasiswa.

Berdasarkan identifikasi tersebut, ditetapkan bahwa solusi yang dibutuhkan adalah chatbot berbasis kecerdasan buatan yang mampu memahami pertanyaan dalam bahasa alami dan memberikan jawaban yang akurat mengenai layanan perpustakaan seperti jadwal operasional, prosedur peminjaman dan pengembalian buku, syarat bebas pustaka, hingga informasi tentang repository perpustakaan.

#### 3.2 Data Understanding

Tahap pemahaman data dilakukan dengan cara mengumpulkan dan menganalisis pertanyaan-pertanyaan yang sering diajukan oleh mahasiswa kepada staf perpustakaan. Data ini diperoleh melalui wawancara langsung dengan staf untuk mengidentifikasi jenis pertanyaan yang umum dan sering berulang.

Hasil dari proses ini kemudian disusun menjadi dataset yang berisi 11 pasangan pertanyaan dan jawaban terkait layanan perpustakaan di STMIK Widya Cipta Dharma. Meskipun jumlah data relatif sedikit, namun dataset tersebut telah mencakup berbagai topik utama yang sering ditanyakan mahasiswa, seperti pembuatan kartu anggota, prosedur peminjaman dan pengembalian buku, jadwal operasional, serta akses ke repository dan layanan digital lainnya. Keragaman isi dataset ini menjadi dasar penting dalam membentuk skenario interaksi chatbot agar mampu menjawab pertanyaan umum secara relevan dan kontekstual.

#### 3.3 Data Preparation

Setelah data pertanyaan dan jawaban dikumpulkan, dilakukan proses persiapan data untuk memastikan bahwa dataset siap digunakan dalam pelatihan chatbot. Tahapan persiapan data mencakup pembersihan, transformasi, dan konversi format data ke dalam struktur yang sesuai untuk pemrosesan lebih lanjut.

Pertama, dilakukan tahap pembersihan teks dengan cara mengubah seluruh huruf menjadi huruf kecil, menghapus tanda baca, dan melakukan tokenisasi menggunakan pustaka *Natural Language Toolkit (NLTK)*. Proses ini bertujuan agar teks lebih seragam dan mudah dianalisis. Setelah itu, dilakukan penghapusan stopwords, yaitu kata-kata umum dalam

bahasa Indonesia yang tidak memiliki makna penting secara semantik, seperti “yang”, “dan”, “di”, dan sebagainya. Tahap ini bertujuan untuk meningkatkan efisiensi dan fokus analisis terhadap kata-kata kunci yang lebih bermakna.

Setelah proses pembersihan selesai, data disusun dalam format JSON, lalu dikonversi ke bentuk DataFrame menggunakan pustaka *Pandas* untuk memudahkan proses manipulasi dan analisis. Data yang telah dibersihkan ini kemudian digunakan pada tahap pemodelan untuk diubah menjadi representasi vektor menggunakan *SentenceTransformer*, yang berperan dalam mengukur kesamaan semantik antara pertanyaan pengguna dan data yang tersedia.

Untuk memberikan gambaran umum tentang data yang digunakan, berikut ini ditampilkan dalam Tabel 2 contoh pasangan pertanyaan dan jawaban dari dataset yang telah disiapkan:

Tabel 2. Dataset Pertanyaan dan Jawaban

No	Pertanyaan	Jawaban
1	Bagaimana cara membuat kartu anggota perpustakaan?	Untuk membuat atau memperpanjang kartu anggota perpustakaan, silakan datang ke bagian layanan anggota dengan membawa kartu tanda mahasiswa (KTM) dan mengisi formulir pendaftaran.
2	Apa saja layanan yang tersedia di perpustakaan?	Perpustakaan menyediakan berbagai layanan seperti peminjaman dan pengembalian buku, ruang baca, layanan referensi, akses jurnal elektronik, ruang diskusi, dan layanan lainnya.
3	Bagaimana prosedur peminjaman dan pengembalian buku?	Peminjaman buku dapat dilakukan dengan membawa kartu anggota perpustakaan dan memilih buku yang ingin dipinjam, kemudian mendatangi staf layanan perpustakaan. Untuk pengembalian dapat dilakukan di tempat yang sama sebelum batas waktu yang ditentukan.
4	Apakah perpustakaan menyediakan layanan konsultasi dengan pustakawan?	Ya, perpustakaan menyediakan layanan konsultasi dengan pustakawan untuk membantu pencarian informasi, penggunaan database akademik, dan referensi penelitian. Layanan ini dapat dilakukan secara langsung.
5	Apakah ada layanan perpustakaan digital?	Ya, perpustakaan menyediakan akses ke koleksi digital seperti e-journal, e-book, dan database akademik yang dapat diakses melalui website perpustakaan melalui link berikut <a href="http://10.10.11.23/opac/">http://10.10.11.23/opac/</a>
6	Bagaimana cara menggunakan fasilitas ruang baca di perpustakaan?	Ruang baca dapat langsung digunakan selama fasilitas meja dan kursi masih tersedia.
7	Bagaimana jika saya terlambat mengembalikan buku?	Jika terlambat mengembalikan buku, akan dikenakan denda sesuai kebijakan perpustakaan.
8	Apa saja syarat Memo KKP atau PI?	Untuk informasi mengenai syarat Memo KKP atau PI bisa cek pada link

No	Pertanyaan	Jawaban
9	Apa saja syarat bebas pustaka?	berikut <a href="https://pustaka.wicida.ac.id/galeri/">https://pustaka.wicida.ac.id/galeri/</a> Untuk informasi mengenai syarat bebas pustaka bisa cek pada link berikut <a href="https://pustaka.wicida.ac.id/syarat-bebas-perpus/">https://pustaka.wicida.ac.id/syarat-bebas-perpus/</a>
10	Jam berapa dan hari apa saja jadwal layanan STMik Widya Cipta Dharma?	Jadwal layanan Perpustakaan STMik Widya Cipta Dharma: SENIN-JUM'AT (Pukul 08.00-21.00 WITA). Jam istirahat layanan: Siang (Pukul 12.00-13.00), Malam (Pukul 18.00-19.00). Atau bisa cek langsung di link berikut <a href="https://pustaka.wicida.ac.id/jadwal-layanan/">https://pustaka.wicida.ac.id/jadwal-layanan/</a>
11	Bagaimana cara upload di repository perpustakaan?	Untuk informasi mengenai upload di repository perpustakaan bisa cek pada link berikut <a href="https://repository.wicida.ac.id/">https://repository.wicida.ac.id/</a>

### 3.4 Modeling

Tahap modeling pada penelitian ini dilakukan dengan menerapkan pendekatan Retrieval-Augmented Generation (RAG) yang terintegrasi dengan teknologi Large Language Model (LLM), yaitu Llama 3.2. Pendekatan ini memungkinkan sistem chatbot menghasilkan jawaban yang tidak hanya akurat secara semantik, tetapi juga relevan dengan konteks pertanyaan yang diajukan oleh pengguna. Pada tahap ini, pemodelan tidak dilakukan dengan melatih model dari awal, melainkan dengan memanfaatkan LLM yang telah dilatih sebelumnya untuk menghemat sumber daya serta meningkatkan efisiensi dan akurasi.

Langkah awal pada tahap modeling adalah melakukan prapemrosesan terhadap data teks pertanyaan dan jawaban. Teks dibersihkan menggunakan pustaka *Natural Language Toolkit (NLTK)* melalui proses konversi huruf menjadi huruf kecil, penghapusan tanda baca, dan penghilangan *stopwords* atau kata-kata umum yang tidak memiliki nilai semantik penting. Setelah teks dibersihkan, kalimat-kalimat tersebut diubah ke dalam bentuk vektor numerik menggunakan model *paraphrase-MiniLM-L6-v2* dari pustaka *SentenceTransformer*. Model ini menghasilkan representasi vektor berdimensi 384 untuk setiap kalimat, di mana masing-masing dimensi menyimpan informasi semantik dari kalimat yang bersangkutan.

Vektor berdimensi 384 berarti bahwa setiap kalimat direpresentasikan sebagai titik dalam ruang vektor berdimensi tinggi yang terdiri dari 384 sumbu. Representasi ini bersifat numerik dan berskala real, memungkinkan sistem untuk mengukur kemiripan makna antara satu kalimat dengan kalimat lainnya. Vektor-vektor ini kemudian disimpan dan diindeks menggunakan pustaka *FAISS (Facebook AI Similarity Search)*, yang mendukung pencarian berdasarkan kemiripan semantik dengan efisiensi tinggi.

Ketika pengguna menginput pertanyaan, sistem akan mengubah pertanyaan tersebut menjadi vektor, kemudian melakukan pencarian terhadap tiga jawaban terdekat dari indeks berdasarkan nilai kemiripan vektor ( $k = 3$ ). Tiga jawaban terpilih tersebut digabungkan menjadi konteks tambahan yang akan dikirim bersama pertanyaan ke model LLM. Melalui integrasi API dengan Ollama, model Llama 3.2 memproses input tersebut dan menghasilkan jawaban secara kontekstual berdasarkan pemahaman semantik, bukan sekadar pencocokan kata kunci.

Hasil representasi vektor dari kalimat yang telah diproses berupa deretan angka real sebanyak 384 komponen. Mengingat jumlah dimensi yang cukup besar, maka hanya sebagian kecil dari nilai vektor yang ditampilkan sebagai ilustrasi. Tabel 3 merupakan hasil embedding berdasarkan dataset pertanyaan dan jawaban yang sesuai dengan Tabel 2 pada tahap Data Preparation sebelumnya:

Tabel 3. Hasil Representasi Vektor

Nilai Vektor					
0.719833	-0.07158813	0.66129065	-0.15781802	0.61082995	-0.34900543
-0.34722382	-0.03819844	0.08890308	-0.43072578	-0.12938522	0.4618232
0.2304309	-0.12442546	-0.08547947	-0.18029317	-0.17968895	-0.26630706
0.26680252	0.0614747	0.0232509	0.13840222	0.3512496	-0.00608033
0.03141745	-0.05354738	0.1940344	0.12437943	-0.31522715	0.24485292
...					
-0.05451731	0.12150189	0.3439875	-0.28079227	0.07371654	0.42481408

Nilai-nilai tersebut merupakan komponen dari vektor embedding yang digunakan dalam pencocokan semantik oleh sistem. Dengan membandingkan vektor pertanyaan pengguna terhadap basis vektor yang telah diindeks sebelumnya, sistem mampu menemukan jawaban yang paling relevan secara makna. Pendekatan ini menjadikan chatbot lebih responsif dan kontekstual dalam menjawab pertanyaan-pertanyaan terkait layanan perpustakaan.

Selain mengubah kalimat menjadi representasi vektor, proses embedding dalam penelitian ini juga dapat dilakukan terhadap satu kata tertentu. Sebagai contoh, kata “perpustakaan” dapat dikonversi menjadi vektor berdimensi 384 menggunakan model paraphrase-MiniLM-L6-v2 dari pustaka *SentenceTransformer*. Proses ini dilakukan dengan memberikan kata tersebut sebagai input ke dalam fungsi `encode()` yang disediakan oleh model. Model ini akan mengubah kata menjadi serangkaian angka real yang secara semantik merepresentasikan makna kata tersebut dalam ruang vektor berdimensi tinggi.

Langkah-langkahnya dimulai dengan memuat model paraphrase-MiniLM-L6-v2 ke dalam program. Setelah itu, kata “perpustakaan” disiapkan sebagai string dan dimasukkan ke dalam fungsi `model.encode()`. Hasil dari fungsi tersebut adalah array atau vektor berdimensi 384, yang masing-masing komponennya mencerminkan ciri semantik dari kata tersebut. Representasi ini memungkinkan sistem chatbot untuk memahami makna

dari satu kata secara kontekstual, dan membandingkannya dengan kata atau kalimat lain dalam pencarian semantik.

Sebagai contoh, pada Tabel 4 ditampilkan hasil 10 dimensi pertama dari embedding kata “perpustakaan”:

Tabel 4. Nilai Vektor Embedding untuk Kata “Perpustakaan”

Dimensi	Nilai Vektor
1	0.129844
2	-0.027140
3	0.223451
4	-0.104317
5	0.310729
6	-0.215802
7	0.064234
8	0.149732
9	-0.086295
10	0.201509

Meskipun hanya satu kata yang diproses, hasil embedding tetap memiliki jumlah dimensi yang sama seperti kalimat, yaitu 384 dimensi. Hal ini dikarenakan model yang digunakan memang dirancang untuk merepresentasikan unit teks apa pun baik satu kata, frasa, maupun kalimat dalam bentuk vektor berdimensi tetap. Proses ini menjadi bagian penting dalam pengembangan sistem berbasis pemahaman semantik, karena memungkinkan perbandingan makna antar kata atau kalimat dalam ruang vektor.

```

=== PERTANYAAN USER ===
cara bebas pustaka

=== JAWABAN DARI DATASET ===
Untuk informasi mengenai syarat bebas pustaka bisa cek p

=== JAWABAN DARI MODEL LLM (GENERATED) ===
Tentu, saya dapat membantu Anda dengan pertanyaan tersele

Untuk mendapatkan informasi tentang cara bebas pustaka,

Dalam situs web tersebut, Anda dapat menemukan syarat da

1. **Cari informasi tentang syarat bebas pustaka**: Di s
2. **Periksa kualifikasi**: Pastikan Anda memenuhi kual
3. **Lihat kebijakan peminjaman**: Cari informasi tentan
4. **Periksa syarat untuk penggunaan e-book**: Jika Anda

Dengan mengikuti langkah-langkah di atas, Anda dapat mer
    
```

Gambar 2. Jawaban Chatbot yang Sesuai dengan Dataset

```

=== PERTANYAAN USER ===
apakah perpustakaan menyediakan tempat untuk makan di tempat

=== JAWABAN DARI DATASET ===
Ya, perpustakaan menyediakan layanan konsultasi dengan pustakawan untuk memba

=== JAWABAN DARI MODEL LLM (GENERATED) ===
Ya, berdasarkan informasi yang diberikan, perpustakaan memang menyediakan beb

1. Layanan konsultasi dengan pustakawan: Memberikan bantuan dan saran dalam m
2. Penggunaan database akademik: Mengaktifkan akses untuk menggunakan databas
3. Referensi penelitian: Menyediakan sumber referensi yang akurat dan terkini

Namun, tidak disebutkan secara eksplisit bahwa perpustakaan menyediakan tempa
    
```

Gambar 3. Jawaban Chatbot yang Tidak Sesuai dengan Dataset

Pada Gambar 2 dan 3 ditampilkan hasil jawaban yang dihasilkan oleh chatbot ketika dimasukkan pertanyaan oleh pengguna. Dalam contoh ini, untuk mengetahui apakah chatbot mampu menjawab dengan baik, maka dimasukkan pertanyaan yang sesuai dan tidak sesuai dengan dataset.

### 3.5 Evaluation

Evaluasi performa chatbot dilakukan menggunakan metrik BERTScore, yang mengukur kesamaan semantik antara jawaban yang dihasilkan oleh chatbot dengan jawaban referensi. Skor BERTScore terdiri dari tiga komponen utama: Precision, Recall, dan F1-Score, yang masing-masing dihitung berdasarkan kemiripan vektor kata antar kalimat menggunakan representasi dari model BERT.

Berdasarkan hasil pengujian terhadap beberapa pertanyaan umum terkait layanan perpustakaan, chatbot menunjukkan performa dapat dilihat pada Tabel 5

Tabel 5. Hasil Evaluasi BERTScore

No	Precision	Recall	F1-Score
1	0.7177	0.8067	0.7596
2	0.7212	0.8494	0.7801
3	0.6550	0.7972	0.7191
4	0.6542	0.9128	0.7622
5	0.7365	0.8701	0.7978
6	0.5309	0.6753	0.5944
7	0.5747	0.8442	0.6838
8	0.5517	0.6131	0.5808
9	0.7122	0.8549	0.7771
10	0.7684	0.7253	0.7462
11	0.5413	0.7672	0.6348

Dari hasil yang diperoleh, evaluasi kinerja chatbot menggunakan BERTScore menunjukkan bahwa chatbot memiliki kemampuan yang baik dalam memahami dan merespons pertanyaan pengguna secara semantik. Nilai F1-Score tertinggi dicapai pada data uji ke-5 sebesar 0.7978, menandakan keseimbangan optimal antara precision dan recall. Sementara itu, beberapa data uji seperti ke-6 dan ke-8 menunjukkan performa yang lebih rendah dengan F1-Score masing-masing 0.5944 dan 0.5808, yang mengindikasikan perlunya peningkatan pemahaman konteks pada jenis pertanyaan tertentu.

Tabel 6. Hasil Rata-rata Evaluasi BERTScore

Precision	Recall	F1-Score
0.6513	0.7924	0.7124

Hasil rata-rata evaluasi BERTScore pada Tabel 6 menunjukkan bahwa chatbot memiliki nilai precision sebesar 0.6513, recall sebesar 0.7924, dan F1-Score sebesar 0.7124. Nilai precision yang moderat mengindikasikan bahwa tidak semua jawaban yang diberikan chatbot sepenuhnya relevan, namun nilai recall yang tinggi menunjukkan bahwa chatbot mampu menangkap sebagian besar informasi penting dari pertanyaan pengguna. F1-Score yang berada di atas angka 0.7 menandakan bahwa secara keseluruhan

chatbot memiliki performa semantik yang baik dalam merespons pertanyaan-pertanyaan yang berkaitan dengan layanan perpustakaan.

```
==== RATA-RATA BERTScore ====
Average Precision : 0.6513
Average Recall    : 0.7924
Average F1 Score  : 0.7124
```

Gambar 4. Hasil Rata-rata Evaluasi BERTScore

Skor F1 yang mencapai 0.7124 pada Gambar 4 menunjukkan performa yang baik namun masih memiliki ruang untuk peningkatan. Beberapa faktor yang dapat memengaruhi skor ini antara lain ukuran dataset yang digunakan relatif kecil (11 pasangan pertanyaan-jawaban) yang membatasi kemampuan model untuk menggeneralisasi berbagai pertanyaan. Selain itu, kompleksitas model Llama3.2 yang digunakan meskipun powerful, namun mungkin memerlukan fine-tuning khusus untuk domain perpustakaan. Faktor lainnya adalah representasi vektor dimana model embedding yang digunakan (paraphrase-MiniLM-L6-v2) mungkin kurang optimal untuk bahasa Indonesia dan terminologi perpustakaan.

Untuk meningkatkan skor, beberapa pendekatan yang dapat diterapkan. Pendekatan pertama adalah memperkaya dataset dengan menambah jumlah dan variasi pertanyaan-jawaban untuk meningkatkan pengetahuan domain model. Pendekatan kedua adalah fine-tuning model yaitu melakukan penyesuaian khusus pada model LLM dengan data perpustakaan untuk meningkatkan pemahaman domain. Selanjutnya dapat juga menggunakan embedding khusus bahasa Indonesia dengan mengadopsi model embedding yang dioptimalkan untuk bahasa Indonesia seperti IndoBERT. Pendekatan lainnya adalah meningkatkan konteks retrieval dengan menambah jumlah dokumen yang diambil ( $k > 3$ ) atau mengimplementasikan reranking untuk meningkatkan kualitas konteks. Terakhir, dapat menerapkan hybrid evaluation dimana selain BERTScore, metrik evaluasi lain seperti ROUGE (untuk mengukur overlap n-gram), BLEU (untuk keselarasan terjemahan), atau bahkan evaluasi manusia dapat memberikan penilaian yang lebih komprehensif tentang kualitas jawaban chatbot. Khususnya, evaluasi manusia dapat menilai aspek pragmatis dan kegunaan jawaban yang mungkin tidak tertangkap oleh metrik otomatis.

Meskipun BERTScore memberikan wawasan berharga tentang kesamaan semantik, penggunaan kombinasi metrik evaluasi bersama dengan penilaian manusia dapat memberikan gambaran yang lebih holistik tentang performa chatbot dan area spesifik yang memerlukan perbaikan.

#### 4. Kesimpulan

Penelitian ini berhasil membangun chatbot berbasis kecerdasan buatan (AI) untuk mendukung layanan informasi perpustakaan di STMIK Widya Cipta Dharma dengan menerapkan metode CRISP-DM. Chatbot dibangun menggunakan pendekatan Retrieval-Augmented Generation (RAG) dan model LLM Llama3.2, serta dievaluasi menggunakan metrik BERTScore. Hasil evaluasi menunjukkan bahwa chatbot mampu memberikan jawaban yang cukup baik terhadap pertanyaan pengguna dengan nilai precision sebesar 0.6513, recall sebesar 0.7924, dan F1-Score sebesar 0.7124. Nilai F1-Score ini menunjukkan kualitas semantik jawaban berada dalam kategori "Baik". Meskipun demikian, terdapat beberapa pertanyaan kompleks yang menghasilkan skor rendah, sehingga pada pengembangan selanjutnya disarankan untuk menambah variasi dataset, melakukan fine-tuning model, dan mengevaluasi sistem secara manual guna meningkatkan akurasi dan relevansi jawaban yang diberikan oleh chatbot.

#### Daftar Rujukan

- [1] W. Widyasari, A. Marini, R. R. Aliyyah, A. A. Gafar, and R. Ruhimat, "Fungsi Pendidikan dan Penelitian pada Perpustakaan Perguruan Tinggi Saat Pandemi Covid-19," *Jurnal Basicedu*, vol. 5, no. 6, pp. 6011–6016, Nov. 2021.
- [2] R. A. Prasetyo et al., "Membangun Sistem Informasi Perpustakaan Digital Menggunakan Metode FAST pada STMIK Widya Cipta Dharma," *Jurnal Sebatik*, vol. 29, no. 1, Jun. 2025.
- [3] L. Costaner, "Aplikasi Chatbot untuk Layanan Informasi dan Akademik Kampus Berbasis Artificial Intelligence Markup Language (AIML)," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 11, no. 2, Nov. 2020.
- [4] A. Hikmah, F. Azmi, and R. A. Nugraheni, "Implementasi Natural Language Processing pada Chatbot untuk Layanan Akademik," *e-Proceeding of Engineering*, vol. 10, no. 1, pp. 371–376, Feb. 2023.
- [5] S. Nova, N. Khotimah, and M. Y. Aryati Wahyuningrum, "Pemanfaatan Chatbot Menggunakan Natural Language Processing untuk Pembelajaran Dasar-Dasar GUI Tkinter pada Bahasa Pemrograman Python," *Jurnal Pendidikan Teknologi dan Kejuruan*, vol. 11, no. 1, Jan. 2024.
- [6] I. L. Alberts et al., "Large language models (LLM) and ChatGPT: What will the impact on nuclear medicine be?," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 50, no. 6, pp. 1549–1552, May 2023.
- [7] B. R. Anassai and P. Josaphat, "Pembangunan Chatbot Sistem Informasi KBLI dan KBJI Berbasis LLM," *Jurnal Teknologi Informasi*, 2024.
- [8] P. Adytia, W. Wahyuni, D. Asmaramany, and K. Sussolaikah, "Coffee Type Classification Using Backpropagation Artificial Neural Network," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 7, no. 1, p. 193, Apr. 2024.
- [9] T. E. Putri and G. Ramadhan, "Penerapan Chatbot sebagai Alat Pembelajaran untuk Pengembangan Pendidikan Karakter," *Jurnal Pendidikan Karakter*, vol. 15, no. 1, Jun. 2024.
- [10] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating Text Generation with BERT," *arXiv preprint*, arXiv:1904.09675, Feb. 2020.
- [11] Y. Yuniati et al., "Analisis Performa Ekstraksi Konten GPT-3 dengan Metrik BERTScore dan ROUGE," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 11, no. 6, Dec. 2024.
- [12] E. Durmus, H. He, and M. Diab, "FEQA: A Question Answering Evaluation Framework for Faithfulness Assessment in Abstractive Summarization," in *Proc. ACL*, May. 2020.
- [13] M. A. Pratiwi, D. Syahbaniar, and A. H. Robbani, "WIDYA (Web Information Dialog Your Assistant): AI-Powered Chatbot for Library Online Service Innovation," *Jurnal FPPTI*, vol. 2, no. 2, pp. 42–52, Aug. 2024.
- [14] S. Sugiono, "Peran Chatbot dalam Mendukung Smart Service pada Smart Library," *Visi Pustaka*, vol. 23, no. 3, Dec. 2021.
- [15] E. Erlina et al., "Penerapan Artificial Intelligence pada Aplikasi Chatbot sebagai Sistem Pelayanan dan Informasi Online pada Sekolah," *Journal of Information System and Technology*, vol. 4, no. 3, pp. 221–230, Nov. 2023.
- [16] Y. Suhanda, I. Kurniati, and S. Norma, "Penerapan Metode CRISP-DM dengan Algoritma K-Means Clustering untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *Jurnal Teknologi Informatika dan Komputer MH Thamrin*, vol. 6, no. 2, Sep. 2020.
- [17] C. Schröder, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021.
- [18] M. R. Amalsyah, D. Kurniawan, A. Rifai, and P. Sari, "Sentiment Analysis of Fintech Application User Reviews Using the CRISP-DM Framework for Product Development Prioritization," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 2, pp. 813–825, Feb. 2025.
- [19] Y. A. Singgalen, "Analyzing an Interest in GPT 4o through Sentiment Analysis using CRISP-DM," *Journal of Information Systems and Informatics*, vol. 6, no. 2, pp. 882–898, Jun. 2024.
- [20] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, "The Evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks," *SBC Reviews on Computer Science*, vol. 4, no. 1, pp. 28–43, Oct. 2024.
- [21] M. I. Syah et al., "Penerapan Retrieval Augmented Generation Menggunakan Langchain dalam Pengembangan Sistem Tanya Jawab Hadis Berbasis Web," *ZONAsi: Jurnal Sistem Informasi*, vol. 6, no. 2, pp. 370–373, May 2024.
- [22] F. R. Fatolah, D. S. Maylawati, and E. Nurlatifah, "Chatbot Edukasi Pra-Nikah berbasis Telegram Menggunakan Bidirectional Encoder Representations From Transformers (BERT)," *Jurnal Algoritma*, vol. 21, no. 2, pp. 29–40, Nov. 2024.
- [23] A. C. Saputra, A. S. Saragih, and D. Ronaldo, "Perbandingan Nilai Akurasi DistilBERT dan BERT pada Dataset Analisis Sentimen Lembaga Kursus," *Jurnal Teknologi Informasi*, vol. 18, no. 2, Aug. 2024.
- [24] J. S. Wibowo, H. Februariyanti, and H. Listiyono, "Model Penjawab Pertanyaan Otomatis Berdasarkan Peringkat Relevansi Kalimat Menggunakan Model BERT," *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 5, no. 3, pp. 1100–1108, Jul. 2024.
- [25] J. H. Cabot dan E. Gyang Ross, "Evaluating Prediction Model Performance," *Surgery*, vol. 174, no. 3, hal. 723–726, Sep. 2023.