

INTISARI

Pengembangan Chatbot Berbasis RAG untuk Optimalisasi Layanan Informasi Terpadu pada Website Statistika Universitas Islam Indonesia

(Studi Kasus : Bidang Statistika)

Faiz Mustafid Amna

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Penelitian ini bertujuan untuk mengembangkan chatbot berbasis *Retrieval-Augmented Generation* (RAG) guna mengoptimalkan layanan informasi terpadu pada website Program Studi Statistika Universitas Islam Indonesia (UII). Permasalahan utama yang dihadapi adalah kesulitan pengguna dalam menemukan informasi spesifik akibat penyebaran konten pada berbagai halaman *website* sehingga diperlukan sistem yang mampu memberikan jawaban secara cepat, akurat, dan kontekstual. Sumber data penelitian dibatasi pada informasi yang terdapat pada halaman visi misi, tujuan, dan sejarah *website* Statistika UII. Metode yang digunakan adalah pendekatan RAG yang menggabungkan proses *retrieval* dan *generation*. Data dikumpulkan melalui *scraping* pada *web* Statistika UII menggunakan Python kemudian diproses melalui tahapan *indexing* meliputi data *loading*, *chunking*, *embedding*, dan penyimpanan pada *vector database* Pinecone. Sistem dikembangkan menggunakan *platform* n8n dengan integrasi model *Large Language Model* (LLM) Mistral Medium. Evaluasi dilakukan menggunakan *semantic similarity*, *metrics*, persentase jawaban benar, uji McNemar, serta *User Acceptance Test* (UAT) dengan metode *System Usability Scale* (SUS). Hasil pengujian menunjukkan bahwa *chatbot* memiliki tingkat *semantic similarity* rata-rata sebesar 0,95 yang menunjukkan kesamaan makna yang sangat tinggi antara jawaban *chatbot* dan *ground truth*. Nilai *F1-score* sebesar 0,76 menunjukkan performa yang baik dalam kesesuaian kata. Selain itu, *chatbot* berbasis RAG memperoleh persentase jawaban benar sebesar 100%, sedangkan *chatbot* berbasis LLM memperoleh 0%. Hasil uji McNemar menunjukkan nilai $\chi^2_{hitung} = 20$ lebih besar dari $\chi^2_{tabel} = 18.307$ sehingga terdapat perbedaan yang signifikan antara performa *chatbot* berbasis RAG dan LLM tanpa RAG. Hasil UAT juga menunjukkan bahwa sistem berada pada kategori “*Good*” dan dapat diterima oleh pengguna.

Kata Kunci : RAG, n8n, *chatbot*, LLM, *website*.

ABSTRACT

Development of a RAG-Based Chatbot to Optimize Academic Information Services on the Statistics Website of the Islamic University of Indonesia (Case Study : Field of Statistics)

Faiz Mustafid Amna
Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

This study aims to develop a Retrieval-Augmented Generation (RAG)-based chatbot to optimize the integrated information services on the website of the Statistics Study Program, Universitas Islam Indonesia (UII). The primary challenge addressed in this study is the difficulty users experience in locating specific information due to content being distributed across multiple web pages. Therefore, a system capable of providing fast, accurate, and contextually relevant responses is required. The data source was limited to information available on the Vision and Mission, Objectives, and History pages of the Statistics UII website. The proposed method employs the RAG approach, which combines retrieval and generation processes. Data were collected through web scraping of the Statistics UII website using Python and subsequently processed through indexing stages, including data loading, chunking, embedding, and storage in the Pinecone vector database. The chatbot was developed using the n8n platform with the integration of the Mistral Medium Large Language Model (LLM). System evaluation was conducted using semantic similarity, text-based metrics, correct answer percentage, the McNemar test, and the User Acceptance Test (UAT) using the System Usability Scale (SUS). The experimental results indicate that the chatbot achieved an average semantic similarity score of 0.95, demonstrating a very high level of semantic agreement between the chatbot responses and the ground truth. An F1-score of 0.76 indicates good performance in terms of lexical similarity. Furthermore, the RAG-based chatbot achieved a 100% correct answer rate, while the LLM-only chatbot achieved 0%. The McNemar test yielded a calculated chi-square value of $\chi^2 = 20$, which exceeded the critical value of $\chi^2 = 18.307$, indicating a statistically significant difference between the performance of the RAG-based chatbot and the LLM without RAG. In addition, the UAT results showed that the system was classified as "Good" and was well accepted by users.

Keywords: RAG, n8n, chatbot, LLMs, website