

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Pada era globalisasi, kemajuan di bidang teknologi tentu tidak dapat dihindari. Era Revolusi Industri 4.0 secara fundamental telah mengubah cara manusia dalam berpikir, menjalani hidup, dan berhubungan satu sama lain [1]. Selain itu, era Revolusi Industri 4.0 juga mempermudah manusia dalam mendapatkan informasi. Saat ini, setiap individu membutuhkan informasi untuk menunjang kegiatan sehari-hari. Sehingga setiap individu akan berusaha untuk mendapatkan informasi yang dibutuhkan secepat mungkin [2].

Berdasarkan data yang diperoleh dari *wearesocial* [3], data pengguna internet di Indonesia pada tahun 2024 yaitu sebesar 5,35 milyar. Jumlah tersebut naik sebesar 97 juta pengguna atau 1,8% dari tahun 2023 dengan alasan utama penggunaan internet terbesar yaitu untuk mendapatkan informasi dengan persentase sebesar 83,1%. Banyaknya pengguna internet dengan tujuan untuk mendapatkan informasi dapat dimanfaatkan oleh institusi seperti universitas untuk dapat menyampaikan Informasi secara efektif melalui internet.

Universitas Islam Negeri (UIN) Sunan Gunung Djati Bandung memiliki beragam informasi yang dapat diakses oleh civitas akademika maupun masyarakat luas. Media penyebaran informasi yang digunakan oleh pihak universitas juga beragam, seperti surat edaran, banner, website, dan media sosial seperti Instagram.

Selain mendapatkan informasi melalui postingan Instagram, pengguna Instagram khususnya civitas akademika UIN Sunan Gunung Djati Bandung juga dapat mencari informasi tambahan dengan cara bertanya melalui kolom komentar. Namun, tidak jarang pertanyaan yang diajukan merupakan pertanyaan berulang yang sudah pernah ditanyakan oleh orang

lain. Pertanyaan berulang tersebut tentu kurang efisien apabila dijawab secara manual satu per satu oleh admin akun Instagram universitas.

Berdasarkan permasalahan di atas, terdapat solusi untuk menangani pertanyaan berulang, yaitu dengan melakukan otomatisasi penyajian informasi dengan mengembangkan pembangkit *Frequently Asked Questions* (FAQ) menggunakan komentar yang sering muncul pada postingan akun Instagram resmi Universitas. *Frequently Asked Questions* (FAQ) merupakan informasi yang berisi daftar pertanyaan yang paling sering ditanyakan oleh pengguna pada suatu layanan [4].

Proses penyusunan FAQ terdiri dari dua tahap. Tahap pertama yaitu membuat pertanyaan dari komentar yang sering muncul pada postingan akun Instagram. Tahap kedua yaitu mencari jawaban untuk pertanyaan yang dihasilkan pada tahap pertama. Oleh karena itu, proses pengembangan pembangkit FAQ akan dibagi menjadi dua tahap, yaitu tahap pengembangan model pembangkit pertanyaan dan tahap pengembangan model pembangkit jawaban.

Tahap pertama pengembangan pembangkit FAQ telah dilakukan oleh Naufal Rahman pada tahun 2023 dalam penelitian yang berjudul “Kesesuaian Pembangkit Pertanyaan Untuk Layanan Akademik Dengan Cosine Similarity” [5]. Penelitian tersebut berhasil mengembangkan model pembangkit pertanyaan yang mampu menghasilkan pertanyaan berdasarkan komentar yang sering muncul pada postingan akun Instagram menggunakan jenis algoritma *N-Gram* yaitu *Bigram* dan *Term Frequency*, lalu mengevaluasinya dengan algoritma *Cosine Similarity*.

Tahap kedua pengembangan pembangkit FAQ yaitu pengembangan model pembangkit jawaban yang menjadi topik utama pada penelitian ini. Perkembangan pesat pada bidang *question answering* telah memunculkan banyak inovasi dalam pembuatan model untuk menjawab pertanyaan dari pengguna. Salah satu perkembangan dari bidang *Question Answering* yaitu munculnya arsitektur *retriever-reader* yang disebut juga sebagai *Modern Open-Domain Question Answering* (OpenQA) [6].

Pada arsitektur *retriever-reader*, proses pencarian jawaban pertanyaan dilakukan dalam dua tahapan, yaitu tahap pencarian dokumen relevan dengan membandingkan secara sintaksis antara dokumen pada *database* dengan pertanyaan yang diajukan, dan tahap penentuan jawaban akhir dengan membandingkan secara semantik antara pertanyaan yang diajukan dengan dokumen yang telah didapatkan pada tahap sebelumnya. Karena kemampuan tersebut, arsitektur *retriever-reader* dipilih untuk mengembangkan model pembangkit jawaban pada penelitian ini.

Dalam mengimplementasikan arsitektur *retriever-reader*, penelitian ini menggunakan model BM25 sebagai *retriever*, dan model IndoBERT sebagai *reader*. Model BM25 dipilih karena memiliki ketepatan tinggi dalam mengurutkan hasil pencarian [7]. Sementara itu, model IndoBERT dipilih karena memiliki performa yang baik sebagai *reader* dibanding model lainnya seperti RoBERTa, distilbert, dan all-MiniLM [8].

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, adanya permasalahan mengenai pertanyaan berulang pada kolom komentar postingan Instagram universitas, serta kehadiran arsitektur *retriever-reader* yang dapat digunakan untuk mengembangkan model pembangkit atau generator jawaban pada pembangkit FAQ. Hal tersebut menjadi latar belakang untuk mengangkat sebuah penelitian tugas akhir yang berjudul **“Generator Jawaban Dalam Sistem Frequently Asked Question (FAQ) Otomatis”**.

1.2 Rumusan Masalah Penelitian

Berdasarkan latar belakang yang telah dipaparkan pada sub-bab sebelumnya, maka rumusan masalah pada penelitian ini yaitu:

1. Bagaimana cara model BM25 sebagai *retriever* dan model IndoBERT sebagai *reader* mencari jawaban yang relevan dengan pertanyaan yang telah dihasilkan oleh model pembangkit pertanyaan?
2. Bagaimana kinerja dari model BM25 sebagai *retriever* dan model IndoBERT sebagai *reader* dalam menjawab pertanyaan yang dihasilkan oleh model pembangkit pertanyaan?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan model BM25 sebagai *retriever* dan model IndoBERT sebagai *reader* pada pembangkit jawaban untuk mencari jawaban yang relevan dengan pertanyaan yang dihasilkan oleh model pembangkit pertanyaan.
2. Mengetahui kinerja dari model BM25 sebagai *retriever* dan model IndoBERT sebagai *reader* dalam menjawab pertanyaan yang dihasilkan oleh model pembangkit pertanyaan.

1.4 Batasan Masalah Penelitian

Batasan-batasan masalah pada penelitian ini yaitu:

1. Fokus utama penelitian ini mengembangkan model pembangkit jawaban untuk menjawab pertanyaan yang dihasilkan oleh model pembangkit pertanyaan yang telah dikembangkan pada penelitian sebelumnya.
2. Arsitektur yang digunakan dalam mengembangkan model pembangkit jawaban yaitu arsitektur *retriever-reader*.
3. Data komentar yang digunakan oleh model pembangkit pertanyaan dalam membuat pertanyaan yaitu data komentar dari akun Instagram *uinsgd.official* pada tanggal 1 Januari 2024 hingga 30 November 2024.
4. Data yang digunakan sebagai jawaban dari pembangkit pertanyaan bersumber dari media sosial dan website resmi UIN Sunan Gunung Djati Bandung.
5. Kombinasi model yang digunakan pada arsitektur *retriever-reader* yaitu BM25 dan IndoBERT dengan *Elasticsearch* sebagai *database*.
6. Data yang digunakan untuk melakukan *fine-tuning* model IndoBERT yaitu SQuAD versi 2.0 berbahasa Indonesia.
7. Framework yang digunakan dalam pengembangan model dengan arsitektur *retriever-reader* yaitu *Haystack*.
8. Hasil dari penelitian ini yaitu model pembangkit jawaban yang dapat menghasilkan jawaban relevan terhadap pertanyaan yang dihasilkan oleh model pembangkit pertanyaan berdasarkan informasi yang tersedia pada *database*.

1.5 Kerangka Pemikiran Penelitian



Gambar 1.1 Kerangka Pemikiran

Pada Gambar 1.1, kerangka pemikiran penelitian ini diawali dari adanya permasalahan mengenai pertanyaan berulang pada kolom komentar postingan akun Instagram resmi UIN Sunan Gunung Djati Bandung. Pertanyaan berulang tersebut tentu tidak efisien apabila dijawab satu per satu. Saat ini, terdapat arsitektur *question answering* yaitu *retriever-reader* yang dapat menghasilkan jawaban yang telah dianalisis secara sintaksis dan semantik terhadap pertanyaan berdasarkan informasi yang tersedia pada *database*.

Pendekatan yang digunakan dalam menyelesaikan masalah yaitu menggunakan model pembangkit pertanyaan yang telah dihasilkan pada penelitian sebelumnya oleh Naufal Rahman (2023) [5], kemudian membuat model pembangkit jawaban menggunakan arsitektur *retriever-reader* untuk menjawab pertanyaan yang dihasilkan model pembangkit pertanyaan. Metode pengembangan yang digunakan pada penelitian ini yaitu *Cross-industry Standard Process for Data Mining (CRISP-DM)*. Implementasi dari pengembangan model pembangkit jawaban yaitu menggunakan bahasa pemrograman *python*, *framework Haystack* untuk mengembangkan model dengan arsitektur *retriever-reader*, *database elasticsearch*, dan *flask* untuk *deployment*.

Hasil dari penelitian ini yaitu model pembangkit jawaban yang dapat menghasilkan jawaban relevan terhadap pertanyaan yang dihasilkan oleh model pembangkit pertanyaan berdasarkan informasi yang tersedia pada *database*.

1.6 Sistematika Penulisan

Pada penelitian ini, sistematika penulisan meliputi gambaran bagaimana isi pada tiap bab, urutan penulisan, serta keterkaitan antara satu bab dengan bab lainnya. Sistematika penulisan pada penelitian ini terbagi menjadi 5 bab, yang dapat dilihat sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang penelitian, rumusan masalah penelitian, tujuan penelitian, batasan masalah penelitian, kerangka penelitian, dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Bab ini menjelaskan tentang literatur dan penelitian terdahulu, serta menjelaskan juga konsep dan teori pendukung yang menjadi landasan dalam penelitian yang akan dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang bagaimana rangkaian metodologi yang digunakan atau implementasinya pada penelitian, hal tersebut mencakup langkah-langkah yang akan dilakukan selama penelitian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini memaparkan dua hal utama, pertama adalah pemaparan tentang temuan atau hasil penelitian berdasarkan tahapan penelitian yang dilakukan. Kedua adalah pembahasan hasil temuan penelitian, hal ini bertujuan untuk menjawab pertanyaan yang telah dirumuskan pada awal penelitian.

BAB V SIMPULAN DAN SARAN

Bab ini membahas kesimpulan dari hasil penelitian yang telah ditemukan dan pembahasan yang telah dipaparkan sebelumnya, selain itu diberikan saran-saran untuk peningkatan pada penelitian selanjutnya.