

BAB I

PENDAHULUAN

1.1 Latar Belakang

Chatbot dapat dideskripsikan sebagai program komputer yang berinteraksi dengan *user* menggunakan bahasa natural (alami). Teknologi ini mulai dikembangkan pada awal 1960 dengan tujuan awal untuk melihat apakah *chatbot* dapat menghasilkan respon yang dapat mengelabui *user* dalam mengidentifikasi sistem *chatbot* tersebut sebagai manusia asli [1]. *Chatbot* dapat memberikan respon menggunakan bahasa alami berdasarkan topik yang dibangun ke dalam pemahaman internal dari model *chatbot* tersebut — tergantung pada basis pengetahuan apa yang diprogram ke dalam *chatbot* oleh pembuat *chatbot* tersebut [2].

Dalam implementasinya di dunia nyata, *chatbot* dapat menjadi solusi dalam menangani pertanyaan yang diajukan oleh *user*, khususnya jika pertanyaan-pertanyaan tersebut berulang kali diajukan oleh *user* yang berbeda sehingga membentuk sebuah pola tertentu. Banyak cara yang dapat dilakukan untuk mengenali pola tersebut, namun secara garis besar pola yang dimaksud dapat diklasifikasikan ke dalam dua kategori, yaitu pertanyaan dengan domain terbuka dan domain tertutup [3]. Kedua kategori pertanyaan tersebut dapat menghadirkan perbedaan yang cukup signifikan. Baik dalam pemilihan metode yang hendak digunakan, jenis model yang cocok untuk kategori pertanyaan tersebut, sampai hasil akhir berupa respon yang paling relevan dengan pertanyaan yang diajukan oleh *user*. Terdapat banyak jenis algoritma dan infrastruktur yang digunakan untuk membangun sebuah *chatbot*. Salah satunya yaitu BERT (Bidirectional Encoding Representation from Transformers) yang merupakan salah satu model pembelajaran mesin yang dikembangkan oleh google pada 2018 [16].

Kasus hadirnya berbagai pertanyaan dengan pola berulang dari berbagai user salah satunya yaitu pada kegiatan penerimaan mahasiswa baru. Informasi mengenai penerimaan mahasiswa baru dapat diakses melalui halaman resmi dari setiap universitas. Di UIN Sunan Gunung Djati Bandung misalnya, informasi mengenai penerimaan mahasiswa baru dapat diakses melalui tautan pmb.uinsgd.ac.id. Disana pun terdapat pertanyaan yang sering ditanyakan oleh calon pendaftar seputar penerimaan mahasiswa baru. Akan tetapi, pada praktiknya di lapangan seringkali ada calon pendaftar yang masih menanyakan informasi-informasi yang sebenarnya sudah tercantum di halaman resmi tersebut.

Untuk mengantisipasi hal-hal seperti demikian, pihak universitas mencantumkan nomor kontak yang dapat dihubungi agar calon pendaftar dapat berdiskusi lebih lanjut. Untuk mendukung layanan tanya jawab seputar penerimaan mahasiswa baru, pada salah satu

penelitian oleh Pasya Alvan Gail Zulvan berjudul “Pendekatan model BERT dalam mendukung layanan tanya jawab penerimaan mahasiswa baru di perguruan tinggi” [32], dibuatlah sebuah *chatbot* yang dapat menjawab pertanyaan seputar penerimaan mahasiswa baru di UIN Sunan Gunung Djati Bandung.

Penelitian tersebut membahas tentang penggunaan model BERT untuk membangun sebuah *chatbot* yang terintegrasi dengan aplikasi WhatsApp untuk mendukung proses tanya jawab seputar penerimaan mahasiswa baru. Varian BERT yang digunakan yaitu IndoBERT yang merupakan sebuah model BERT yang sudah dilatih menggunakan kosakata bahasa indonesia. Hasilnya, ditemukan bahwa tingkat akurasi (*confidence score*) mencapai 99.8%. Hal ini merupakan hasil dari proses pengolahan kata dari setiap pertanyaan yang diajukan, dimana setiap kata tersebut akan dikomparasikan dengan dataset pertanyaan yang dikumpulkan pada penelitian tersebut. Kendati demikian, terdapat varian baru dari BERT yakni SBERT (*Sentence Bidirectional Encoding Representations from Transformer*) yang merupakan pengembangan berkelanjutan dari BERT. Sedikit berbeda dari BERT, SBERT ini memiliki proses pengolahan kalimat, dimana tahap pengolahannya dilakukan berdasarkan susunan kata yang membangun suatu kalimat pertanyaan.

Untuk memahami bagaimana varian baru dari BERT ini bekerja serta apa saja yang membedakan SBERT dan BERT pada studi kasus tanya jawab seputar penerimaan mahasiswa baru, maka peneliti melakukan penelitian ini yang berjudul **“Implementasi Bidirectional Encoding Representation From Transformer untuk Klasifikasi Teks pada Model Chatbot Tanya Jawab Penerimaan Mahasiswa Baru”**. Penelitian ini akan berfokus pada pemanfaatan teknologi *Sentence Bidirectional Encoding Representation from Transformer* (SBERT) untuk klasifikasi teks berdasarkan kalimat dalam membangun sebuah *chatbot* untuk menangani interaksi tanya-jawab antara *user* yang akan mendaftar ke perguruan tinggi atau memerlukan informasi terkait penerimaan mahasiswa baru.

Dengan hadirnya penelitian ini, diharapkan keilmuan tentang SBERT dapat lebih diperluas lagi serta dapat ditemukan bagaimana cara kerja SBERT jika dibandingkan dengan model BERT yang sudah ada.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan yang dipaparkan sebelumnya, terdapat berbagai masalah yang dirumuskan sebagai berikut:

1. Bagaimana proses pengembangan *chatbot* menggunakan model berbasis SBERT pada sistem tanya jawab penerimaan mahasiswa baru?

2. Bagaimana akurasi dan kecepatan respons model berbasis SBERT dalam pengimplementasiannya pada chatbot sistem tanya jawab penerimaan mahasiswa baru?
3. Apa saja perbedaan antara model SBERT dengan model BERT dari segi persiapan hingga hasil akhir?

1.3 Tujuan Penelitian

Berdasarkan penjabaran rumusan masalah yang telah dipaparkan sebelumnya, dibentuklah tujuan penelitian sebagai berikut:

1. Mengembangkan *chatbot* menggunakan SBERT untuk mendukung proses tanya jawab sekaligus mempermudah aksesibilitas informasi seputar penerimaan mahasiswa baru.
2. Mengevaluasi akurasi dan kecepatan respons model berbasis SBERT dalam pengimplementasiannya pada chatbot sistem tanya jawab penerimaan mahasiswa baru.
3. Memahami perbedaan antara model SBERT dengan model BERT dari segi persiapan hingga hasil akhir.

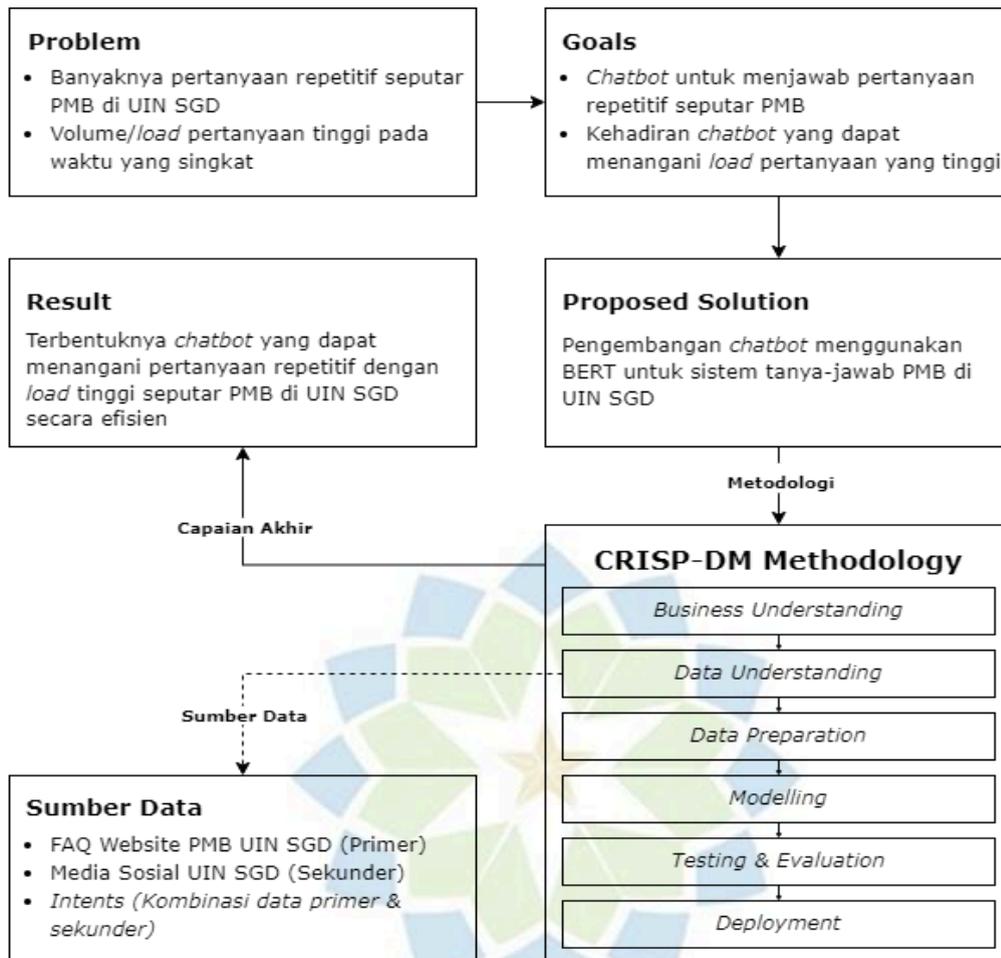
1.4 Batasan Masalah

Untuk memusatkan fokus masalah penelitian sehingga dapat dipahami dalam perspektif yang sama dan membatasi cakupan penelitian, terdapat batasan masalah yang bisa dirumuskan sebagai berikut:

1. *Chatbot* yang dikembangkan pada penelitian ini akan berfokus pada sistem tanya-jawab seputar penerimaan mahasiswa baru di UIN Sunan Gunung Djati Bandung bagi calon mahasiswa baru.
2. Penelitian ini berfokus pada pengembangan *chatbot* menggunakan SBERT;
3. Sumber data utama yang digunakan berisikan pertanyaan-pertanyaan berbasis teks dalam bahasa Indonesia yang tertera pada halaman FAQs di situs resmi penerimaan mahasiswa baru UIN Sunan Gunung Djati Bandung.
4. *Chatbot* akan dilatih menggunakan dataset yang merupakan hasil dari pengembangan sumber data utama, dimana pengembangan yang dimaksud berupa penyesuaian bentuk dan pola dari pertanyaan serta respon dari pertanyaan tersebut.

1.5 Kerangka Pemikiran Penelitian

Adapun kerangka pemikiran dari penelitian ini adalah sebagai berikut:



Gambar 1.1 Kerangka Pemikiran Penelitian

Pada gambar 1.1 dijelaskan bagaimana peneliti menyusun kerangka pemikiran guna memberikan patokan terhadap setiap tahapan penelitian dan mendokumentasikan setiap tahapan tersebut agar menjadi penelitian yang terstruktur. Adapun tahapan awal yang dilewati pada penelitian ini dimulai dari pengidentifikasian masalah dan *goals* yang ingin dicapai, pembentukan solusi yang diajukan, metodologi yang digunakan untuk mencapai solusi, dan gambaran capaian akhir yang diinginkan. Dalam mencapai solusi yang diinginkan, peneliti menggunakan metodologi CRISP-DM untuk pengembangan model dan juga *data analysis*.

1.6 Metode Penelitian

1.6.1 Metode Pengumpulan Data

Peneliti menggunakan metode pengumpulan data secara kuantitatif dengan mengumpulkan pertanyaan yang tertera di halaman FAQ PMB UIN Sunan Gunung Djati Bandung. Selain itu, terdapat sumber lain yang digunakan sebagai sumber data pada penelitian ini, yaitu media sosial instagram UIN Sunan Gunung Djati Bandung.

1.6.2 Metode Pengembangan Model

Peneliti menggunakan metode CRISP-DM. Metode CRISP-DM merupakan metode yang sering digunakan untuk eksplorasi data atau data mining, dimana metode ini mencakup tahapan-tahapan yang terstruktur dan sistematis sehingga memudahkan dapat mempermudah peneliti dalam memetakan tahapan-tahapan yang akan dilakukan pada penelitian ini. Adapun tahapan yang tercakup pada metode CRISP-DM yaitu: *Business Understanding; Data Understanding; Data Preparation; Modelling; Testing & Evaluation; Deployment*.

1.7 Sistematika Penulisan

Informasi dan hasil yang didapat dari penelitian ini disusun dalam bentuk tugas akhir. Adapun sistematika penulisannya dapat diuraikan sebagai berikut:

BAB I PENDAHULUAN

Bab I terdiri dari beberapa subbab, yakni Latar Belakang, Rumusan Masalah, Tujuan Penelitian, Batasan Masalah, Kerangka Pemikiran, dan Sistematika Penulisan.

BAB II KAJIAN LITERATUR

Bab II berisikan studi literatur sebagai dasar acuan untuk penelitian ini. Studi literatur pada penelitian terdahulu dimaksudkan untuk membentuk penelitian ini agar sesuai dengan objek dan arah penelitian yang sudah ditentukan.

BAB III METODOLOGI PENELITIAN

Bab III menjelaskan langkah-langkah dan metode penelitian. Selain itu, metodologi yang digunakan disesuaikan untuk memenuhi kebutuhan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini akan menjelaskan seluruh proses pembuatan sistem. Proses ini dimulai dengan tahap perencanaan kebutuhan sistem, tahap perancangan desain sistem, yang digambarkan dalam diagram UML, tahap pengkodean sistem, dan tahap pengujian sistem.

BAB V PENUTUP

Pada bab akhir, terdapat simpulan yang membahas jawaban dari pertanyaan pada rumusan masalah. Selain kesimpulan, bab ini pun memuat saran yang ditujukan untuk menunjang dan meningkatkan kualitas pada penelitian selanjutnya.

Daftar Pustaka

Bagian ini berisi daftar dari sumber referensi yang digunakan untuk penelitian.

Lampiran

Dokumen-dokumen tambahan yang digunakan dalam proses penyusunan dan perancangan penelitian dimuat pada bagian lampiran.