

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Indonesia telah memasuki fase endemi Covid-19 setelah pemerintah menyatakan berakhirnya masa pandemi Covid-19 pada tanggal 21 Juni 2023 [1]. Dampak dari pandemi Covid-19 adalah adanya kepanikan, rasa was-was dan kecurigaan terhadap orang asing yang dapat menyebabkan menurunnya kesehatan masyarakat fisik dan juga kesehatan mental [2]. Adanya masa karantina dan isolasi membuat masyarakat dituntut untuk bisa beradaptasi menggunakan teknologi digital khususnya pada bidang kesehatan [3]. Dalam bidang kesehatan, penggunaan layanan kesehatan digital seperti aplikasi *healthcare* meningkat secara signifikan [4] yang berujung pada timbulnya tantangan untuk menyediakan informasi terpercaya dan akurat secara cepat. Pertanyaan dasar yang cukup penting seperti “Apa gejala Covid-19?”, “Bagaimana pertolongan pertama pada pasien Covid-19?” sering ditanyakan oleh masyarakat. Untuk mencari informasi tentang covid-19, masyarakat menggunakan internet sebagai medianya. Hal ini dikarenakan keputusan dan pilihan yang diambil oleh masyarakat lebih banyak didasarkan pada informasi dari internet, terutama media sosial [5]. Berdasarkan kebiasaan tersebut, informasi yang diperoleh terkadang belum spesifik dengan yang diharapkan. Oleh karena itu untuk mengatasi permasalahan tersebut maka perlu Sistem Question Answering (QA) untuk mendapatkan informasi spesifik.

Sistem QA digunakan untuk memberikan informasi secara spesifik yang dibutuhkan oleh pengguna. Berbeda dengan search engine, sistem QA bertujuan untuk memberikan jawaban atas pertanyaan secara langsung daripada hanya memberikan tautan yang relevan, sehingga memberikan kemudahan dan efisiensi dalam mendapatkan informasi [6]. Pengguna hanya perlu memberikan pertanyaan, kemudian sistem akan otomatis menghasilkan jawaban yang sesuai. Sistem QA ini dapat digunakan untuk memudahkan masyarakat untuk mendapatkan jawaban seputar kesehatan dalam waktu yang singkat. Salah satu

framework yang cocok untuk sistem QA adalah BERT. BERT dianggap sebagai *state-of-the-art* dalam tugas NLP (*Natural Language Processing*) [7].

Pada penelitian ini, model BERT yang digunakan adalah IndoBERT, yaitu model BERT yang di-*train* menggunakan dataset berbahasa Indonesia [8]. Pada beberapa penelitian, IndoBERT memiliki hasil yang unggul dibandingkan model berbasis BERT yang lain. Penelitian [9] yang membandingkan IndoBERT dengan RoBERTa pada studi kasus sistem QA terhadap dataset Buku Panduan Akademik menghasilkan F1-score 91,32 dan Exact Match 81,17 untuk IndoBERT dan F1-score 90,18 dan Exact Match 79,53 untuk RoBERTa. Penelitian lain [10] melakukan uji coba terhadap data BPS (Badan Pusat Statistik) Indonesia menggunakan model IndoBERT, distilbert, all-MiniLM-L6-V2, dan RoBERTa. Hasil dari penelitian tersebut didapat F1-score masing-masing 46,03; 35,81; 32,67; 36,63, dan masing-masing Exact Match 1,33; 0,33; 1,0; 1,0.

Evaluation metric yang digunakan pada penelitian ini adalah Exact Match (EM) dan F1-score yang juga umum digunakan pada BERT untuk sistem QA [11].

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana merancang sistem QA menggunakan model pre-trained IndoBERT dengan dataset dari artikel pada *website* halodoc?
2. Bagaimana performa model yang dibangun?
3. Bagaimana hasil *answer* model yang dibangun?

1.3. Tujuan dan Manfaat

Berdasarkan latar belakang dan rumusan masalah yang telah dijelaskan, tujuan dan manfaat pada penelitian ini adalah sebagai berikut.

1. Membuat model IndoBERT QA dengan dataset dari artikel pada *website* halodoc

2. Mengetahui hasil performa model yang dibangun
3. Sistem dapat menghasilkan prediksi jawaban (*answer*) yang tepat beserta *score*-nya

1.4. Batasan Masalah

Batasan dalam penelitian Tugas Akhir ini adalah membuat model BERT QA dengan dataset kesehatan yang berasal dari artikel web halodoc untuk memprediksi jawaban (*answer*) dari pertanyaan (*question*) mengenai artikel tersebut. Model pre-trained BERT yang digunakan adalah IndoBERT. *Question* dan *Answer* hanya yang berkaitan dengan artikel kesehatan saja. *Output* dari penelitian ini adalah prediksi *answer* beserta *score*-nya.

1.5. Jadwal Pelaksanaan

Jadwal pelaksanaan pengerjaan Tugas Akhir ini dapat dilihat pada Tabel 1.1.

Tabel 1.1. Jadwal Pelaksanaan Tugas Akhir

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan Data						
3	Perancangan Sistem						
4	Pengujian Model						
5	Analisis dan Evaluasi						
6	Penyusunan Laporan Tugas Akhir						

1. Studi Literatur

Melakukan pengumpulan informasi dari sumber yang relevan seperti jurnal, artikel ilmiah mengenai BERT, IndoBERT dan Sistem *Question Answering*.

2. Pengumpulan Data

Melakukan pengumpulan data yang dibutuhkan dalam penelitian, seperti penyusunan dataset.

3. Perancangan Sistem

Merancang Sistem *Question Answering* dengan menggunakan dataset yang telah disusun.

4. Pengujian Model

Melakukan pengujian model dengan beberapa skenario percobaan untuk mengetahui output terbaik dari sistem yang telah dibangun.

5. Analisis dan Evaluasi

Melakukan analisis dan evaluasi terhadap hasil dari skenario percobaan yang telah dilakukan.

6. Penyusunan Laporan Tugas Akhir

Membuat laporan tugas akhir yang berisi latar belakang, rumusan masalah, tujuan penelitian, perancangan sistem, evaluasi dari proses pengujian, dan kesimpulan penelitian.