

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

*Chatbot* merupakan program berbasis komputer yang berfungsi untuk mensimulasikan percakapan dengan manusia melalui pendengaran dan tekstual. *Chatbot* dirancang untuk dapat mengenali masukan pengguna dengan menggunakan pencocokan pola dan mengakses informasi dari database untuk memberikan respon yang sesuai (Dahiya, 2017). *Chatbot* sering digunakan untuk membantu pekerjaan manusia dalam menyampaikan informasi seperti penggunaan chatbot dalam sistem pembelajaran virtual, penerapan dalam dunia bisnis seperti layanan *customer service*, dan pada bidang lainnya. Oleh karena itu, saat perancangan *chatbot* perlu diperhatikan *knowledge* yang digunakan pada sistem pemberian jawaban agar dapat memberikan jawaban yang sesuai dengan pertanyaan *user*.

Pada era ini, *web* menjadi sebuah *repository* pengetahuan besar yang menyediakan beragam informasi dalam berbagai bentuk seperti teks, audio, grafik, video dan multimedia (Mowla dkk., 2017). Informasi yang ada tidak terbatas dan dapat diakses, disimpan, serta diperbaharui oleh pengguna dari seluruh dunia. Pada saat melakukan pencarian informasi, akan ada sejumlah *link* berisikan informasi terkait yang muncul. Di dalam sejumlah *link* yang tersedia, akan ada banyak informasi berupa *unstructured text* yang tidak mudah digunakan oleh komputer untuk memproses data (Dang dan Ahmad, 2014). Maka dari itu, diperlukan peranan *text mining* untuk mengekstrak informasi penting yang dibutuhkan.

Instansi seperti Universitas Multimedia Nusantara menggunakan chatbot yang diperkenalkan pertama kali oleh Steven Wijaya (2019) dengan nama “*Jacob*” untuk membantu tugas dari *staf* marketing. *Chatbot* berbasis web ini menggunakan platform Wit.ai untuk menyediakan informasi program *Joint Degree* Informatika seperti biaya, kurikulum, mata kuliah, fasilitas, dan karir dengan menggunakan bahasa Inggris (Wijaya dan Wicaksana, 2019). Aplikasi *chatbot* Jacob telah dikembangkan oleh peneliti sebelumnya yang merupakan alumni mahasiswa Universitas Multimedia Nusantara yaitu Steven Wijaya (2019), Octavany (2019), Akino Archiles (2019), Felicia Priscilla Lovely (2020), dan Melvin Hendronoto (2020) dengan menambahkan modul baru pada *chatbot* Jacob. Jacob saat ini telah dirancang untuk dapat mengatasi kebutuhan yang dinamis di mana Jacob dapat melakukan *Text Mining* secara daring, tetapi *Text Mining* yang dilakukan hanya terbatas pada Bahasa Inggris. Selain itu, aplikasi *chatbot* Jacob saat ini hanya dapat menerima *input* dan menghasilkan *output* dalam Bahasa Inggris.

Penggunaan bahasa selain Bahasa Inggris diperlukan supaya *chatbot* Jacob dapat melayani pertanyaan *user* yang menggunakan bahasa selain Bahasa Inggris. Saat melakukan interaksi dengan *chatbot* Jacob, pengguna bisa saja ada yang tidak secara leluasa melakukan percakapan dengan menggunakan Bahasa Inggris. Oleh karena itu, diperlukan sebuah opsi berupa pemilihan bahasa yang akan digunakan oleh *user* saat melakukan interaksi dengan *chatbot* Jacob. Bahasa lain yang digunakan adalah Bahasa Indonesia sebagai bahasa mayoritas dari pengguna *chatbot* Jacob. Selain itu, diperlukan sistem *text mining* dalam Bahasa Indonesia supaya *chatbot* Jacob dapat mencari dan menerima informasi Berbahasa Indonesia, serta dapat memberikan jawaban yang sesuai kepada *user*. IndoBERT-lite dan

RoBERTa merupakan contoh arsitektur *neural network* yang dapat digunakan untuk melakukan *text mining*.

Pada penelitian yang dilakukan oleh Wilie dkk (2020), diciptakan sebuah *benchmark* diberi nama dengan “IndoNLU”. IndoNLU merupakan *benchmark* pertama untuk proses *training*, *evaluation*, dan *Indonesian natural language understanding* pada *machine learning* dengan 12 *task* yang berbeda-beda. *Benchmark* dirancang untuk memenuhi pemahaman mesin terhadap bahasa formal maupun tidak formal pada Bahasa Indonesia. IndoNLU juga memperkenalkan arsitektur *neural network* BERT Bahasa Indonesia yang diberi nama dengan IndoBERT, serta arsitektur ALBERT dengan nama IndoBERT-lite. Hasil penelitian menunjukkan *pre-trained model* IndoBERT dan IndoBERT-lite memiliki nilai *F-score* tertinggi dibandingkan dengan *pre-trained model* lainnya (Wilie dkk., 2020).

Selain IndoBERT-lite, terdapat arsitektur *neural network* lainnya yang merupakan perluasan dari arsitektur BERT dengan nama RoBERTa. RoBERTa dikemukakan oleh Liu dkk (2019) merupakan akronim dari “*Robustly Optimized BERT Pretraining Approach*” yang dapat lebih baik bahkan melebihi kinerja dari metode post-BERT. Pada penelitian yang dilakukan, didapatkan hasil yang menunjukkan arsitektur RoBERTa memiliki *F-score* tertinggi dengan nilai sebesar 94.6% pada *dataset* SQuAD v1.1 dan 89.4% pada SQuAD v2.0 dibandingkan dengan arsitektur BERT (Devlin dkk., 2019) dan XLNet (Yang dkk., 2019).

Pada penelitian ini, dirancang suatu modul yang dapat melayani pertanyaan *user* dalam Bahasa Indonesia. Selain itu, dirancang sebuah *web service* yang mengimplementasikan sistem *text mining* Bahasa Indonesia pada aplikasi *chatbot* Jacob. *Web service* dibangun dengan menggunakan bahasa pemrograman Python

3.6 dan *framework web serive* Flask Python 1.0.2. *Pre-trained model* yang dipilih adalah IndoBERT-lite dan RoBERTA yang telah di *fine-tuning* dengan menggunakan *dataset* SQuAD 2.0 Bahasa Indonesia dan TyDI QA untuk menghasilkan *output* berupa jawaban dari pertanyaan *user*. *Pre-trained model* IndoBERT tidak dipilih dikarenakan jumlah parameter yang digunakan sangat banyak, serta tidak memungkinkan untuk dilakukan *fine-tuning* dengan spesifikasi sistem yang ada. *Testing* dilakukan dengan pendekatan *white box testing* untuk melihat nilai akurasi dan *F-score* dari hasil implementasi IndoBERT-lite dan RoBERTa.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah ditelaah, dapat dirumuskan rumusan masalah sebagai berikut.

1. Bagaimana mengimplementasikan IndoBERT-lite dan RoBERTa untuk *text mining* pada aplikasi *chatbot* Jacob?
2. Berapa nilai akurasi dan *F-Score* yang dihasilkan IndoBERT-lite dan RoBERTa pada aplikasi *chatbot* Jacob?

## 1.3 Batasan Masalah

Berdasarkan rumusan masalah yang telah ditelaah, batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Penggunaan bahasa pada masukan yang diberikan dan jawaban yang disampaikan menggunakan Bahasa Indonesia.
2. *Text Mining* dilakukan dengan menggunakan Bahasa Indonesia.

3. *Input* dan *output* yang diberikan melalui suara dengan menggunakan mikrofon sebagai masukan dan pengeras suara sebagai keluaran.
4. *Text Mining* hanya dapat dilakukan jika *user* menyalakan tombol untuk menjelajah internet dan memberikan *input* yang diawali dengan kata tanya.

#### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah ditelaah, tujuan dilaksanakannya penelitian ini adalah sebagai berikut.

1. Mengimplementasikan IndoBERT-lite dan RoBERTa untuk *text mining* pada aplikasi *chatbot* Jacob.
2. Mengetahui nilai akurasi dan *F-Score* yang dihasilkan IndoBERT-lite dan RoBERTa pada aplikasi *chatbot* Jacob.

#### **1.5 Manfaat Penelitian**

Manfaat dilaksanakannya penelitian ini adalah sebagai berikut.

1. Memberikan tambahan kecerdasan kepada *voice chatbot* Jacob untuk dapat mengenali pertanyaan yang diberikan oleh *user* dalam Bahasa Indonesia.
2. Membantu *voice chatbot* Jacob untuk dapat melakukan *text mining* secara daring dan memberikan jawaban yang sesuai dengan pertanyaan *user* dalam Bahasa Indonesia.

#### **1.6 Sistematika Penulisan**

Sistematika penulisan yang digunakan dalam laporan Skripsi ini adalah sebagai berikut.

## BAB 1 PENDAHULUAN

Bab ini berisikan latar belakang masalah dari pemilihan judul Skripsi “Implementasi IndoBERT-lite dan RoBERTa untuk Text Mining pada Aplikasi Chatbot Jacob”, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

## BAB 2 LANDASAN TEORI

Bab ini berisikan beberapa literatur yang mendukung penelitian terkait dengan penelitian yang dilakukan. Beberapa literatur tersebut antara lain adalah literatur mengenai aplikasi *chatbot* Jacob, *text mining*, IndoBERT-lite, RoBERTa, SQuAD, TyDI QA, akurasi, dan *F-Score*.

## BAB 3 METODOLOGI PENELITIAN

Bab ini berisikan analisis kebutuhan, perancangan aplikasi, dan implementasi aplikasi dari *chatbot* Jacob Bahasa Indonesia dan *web service* yang digunakan. *Flowchart* dan gambaran mengenai skema kerja *chatbot* Jacob Bahasa Indonesia dan *web service* juga dijabarkan pada bab ini.

## BAB 4 HASIL DAN DISKUSI

Bab ini berisikan hasil implementasi dan hasil pengujian yang dilakukan dari perancangan *web service* dan aplikasi *chatbot* Jacob Bahasa Indonesia. Hasil penelitian dan hasil analisis untuk menjawab rumusan masalah penelitian ini juga dijabarkan pada bab ini.

## BAB 5 SIMPULAN DAN SARAN

Bab ini berisikan simpulan dan hasil penelitian dari tujuan yang ingin dicapai pada penelitian ini, beserta saran untuk pengembangan aplikasi *chatbot* Jacob lebih lanjut.