

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Informasi mengenai penerimaan mahasiswa baru di universitas biasanya sudah tersedia di web resmi universitas. Untuk informasi lebih lengkap biasanya kampus menyediakan kontak yang dapat dihubungi, untuk memperjelas pertanyaan dari calon mahasiswa ataupun orang tua dari mahasiswa. Namun jika banyak orang yang menelpon ke universitas, hanya untuk mencari info mengenai penerimaan mahasiswa baru maka petugas akan kebingungan. Banyak calon mahasiswa yang ingin mengetahui informasi mengenai universitas dan program studi yang diinginkan, oleh karena itu mereka cenderung akan menggali informasi mengenai universitas yang diinginkan.

Sekarang ini sudah banyak digunakan kecerdasan buatan salah satunya dalam bidang tanya jawab dengan menggunakan teknologi bernama *Chatbot*. Teknologi *Chatbot* merupakan salah satu bentuk aplikasi dari *Natural Language Processing* (NLP), NLP merupakan salah satu bidang ilmu Kecerdasan Buatan yang mempelajari komunikasi antara manusia dengan komputer melalui bahasa alami [1].

Chatbot adalah sebuah kecerdasan buatan yang tersimulasi sebagai percakapan seakan-akan seperti dengan orang sungguhan [2]. *Chatbot* juga kadang-kadang disebut *chat robot*, *bot*, *chatterbot*, *botchatting*, *chatterbox*, *V-Host*, *V-People*, *agent*, dan manusia virtual [3]. *Chatbot* dapat mengubah cara bagaimana suatu layanan disediakan, daripada kita harus menghubungi nomor telepon yang disediakan kita bisa langsung mengetikkan apa pertanyaan kita kedalam suatu antarmuka yang berbentuk seperti tampilan aplikasi *chatting* biasanya. Pertanyaan yang diajukan ke dalam antarmuka tadi akan di proses oleh sebuah kecerdasan buatan yang kemudian akan dijawab langsung oleh kecerdasan buatan tadi atau bisa

disebut *Chatbot*, sehingga petugas tidak harus siap setiap saat untuk menjawab pertanyaan dari pengguna dan pengguna merasa puas karena pertanyaannya dijawab.

Chatbot ini akan mengimplementasikan sebuah model *Natural Language Processing* (NLP). Model yang diimplementasikan adalah model *Retrieval Based*, model ini merupakan model yang akan mengeluarkan respon sesuai dengan apa yang di masukan kedalam sistem. Model ini akan mengecek masukan pertanyaan tadi kedalam *dataset*, lalu akan dicek kemiripan atau kesamaannya, saat sudah ditemukan maka akan dikeluarkan respon yang sesuai dengan yang ada di *dataset*. *Chatbot* yang menerapkan model ini mendapat keuntungan dari respon yang informatif, dikarenakan model ini memilih respon yang sesuai dari percakapan tergantung dari informasi yang dimiliki *Chatbot* [4]. Pada aplikasi *Chatbot* ini masukan berupa pertanyaan dan respon berupa jawaban. *Dataset* akan disusun dengan data dari pertanyaan yang sering ditanyakan mengenai penerimaan mahasiswa baru dan informasi mengenai universitas Atma Jaya Yogyakarta, *dataset* ini akan berisi dua buah atribut yaitu pertanyaan beserta jawabanya.

Model *Retrieval Based chatbot* ini lebih cocok digunakan pada kasus tanya – jawab (*Question and Answer*) atau bidang *Customer Service*, karena model ini akan langsung mengeluarkan respon sesuai yang ada di *dataset* sehingga jawaban yang keluar tidak akan asal karena jawaban akan langsung diambil sesuai yang ada pada *dataset*. Informasi yang diberikan kepada penanya haruslah terpercaya sehingga model ini sangat cocok digunakan pada *Chatbot* ini.

Berdasarkan hal tersebut maka perlu diterapkan aplikasi *Chatbot* di lingkup Universitas Atma Jaya Yogyakarta dalam kasus penerimaan mahasiswa baru. Model ini nantinya dapat menjawab pertanyaan mengenai informasi universitas , program studi, fasilitas, informasi mengenai penerimaan mahasiswa baru yang ada di universitas Atma Jaya Yogyakarta, dan dapat menjawab beberapa percakapan dasar. *Chatbot* nantinya juga dapat memberikan rekomendasi informasi yang ada jadi ketika pengguna mengakses *Chatbot*, *Chatbot* akan menyapa penggunanya sehingga pengguna tidak merasa bahwa dia harus memulai duluan untuk melakukan proses *chat* bersama *Chatbot*.

Dengan menerapkan *Chatbot* maka calon mahasiswa dan orang tua mahasiswa dapat mencari informasi lebih detail melalui aplikasi *chatting* jika informasi yang ada di *website* Atma Jaya Yogyakarta dirasa kurang lengkap atau sulit didapatkan. Sehingga calon dan orang tua mahasiswa merasa cocok dan ingin bergabung dalam Universitas Atma Jaya Yogyakarta

1.2. Rumusan Masalah

Berdasarkan apa yang sudah dijelaskan pada latar belakang di atas dapat dirumuskan sebagai berikut :

1. Bagaimana cara membangun *dataset* untuk kasus Informasi Penerimaan Mahasiswa Baru ?
2. Bagaimana cara membuat model *Chatbot* Retrieval-Based untuk pemberian informasi tentang Penerimaan Mahasiswa Baru dan informasi Universitas Atma Jaya Yogyakarta ?
3. Bagaimana cara mengimplementasikan *Chatbot* agar mudah di akses oleh pengguna ?

1.3. Tujuan Pembahasan

Berdasarkan rumusan masalah diatas, maka tujuan pembuatan tugas akhir ini adalah :

1. Membuat *dataset* yang digunakan sebagai sumber informasi dari *Chatbot* yang akan dibuat.
2. Membuat model *Chatbot Retrieval Based* yang dapat menjawab pertanyaan yang diberikan oleh pengguna sesuai dengan informasi yang dimiliki *Chatbot*.
3. Membuat web berupa tampilan chat yang dapat diakses oleh pengguna secara online.

Ruang Lingkup Kajian

Agar penulisan tugas akhir ini berfokus pada permasalahan yang ada, maka diperlukan batasan – batasan yaitu :

1.3.1. Batasan Perangkat Keras

Perangkat keras yang digunakan untuk melakukan pembangunan, pelatihan, dan pengujian untuk membuat web dan model *Chatbot* yaitu :

Laptop Asus dengan Windows 10 Home, processor Intel Core i7-7700HQ @2.80GHz, RAM 8 GB, dan GPU NVIDIA GeForce GTX 1060 6 GB.

1.3.2. Batasan Perangkat Lunak

Perangkat lunak yang dibutuhkan untuk mengakses atau menjalankan Model *Chatbot* ini adalah browser dan jupyter notebook yang sudah terinstal di PC. Jupyter Notebook digunakan untuk membuka file bertipe *.ipynb. Tampilan dari interaksi *Chatbot* juga ditampilkan melalui Jupyter Notebook namun juga tersedia secara online melalui web yang sudah dibuat. *Dataset* dibuat dengan format *.csv jadi dapat dibuka dengan aplikasi *spreadsheet* seperti Microsoft Excel.

1.3.3. Library Yang Digunakan

Dalam penelitian ini juga menggunakan beberapa *library* untuk bahasa pemrograman Python. *Library* yang digunakan dalam penelitian ini yaitu :

1. Numpy untuk pemrosesan array matriks
2. NLTK untuk pembersihan data dan tokenisasi
3. Pickle untuk menyimpan objek kedalam file (pada penelitian digunakan untuk menyimpan hasil training agar training hanya dilakukan sekali ketika *dataset* diperbarui saja.
4. *tfidfVectorizer* untuk ekstraksi fitur dari dokumen menjadi matriks *TFIDF*
5. *Cosine_similarity* mengukur kesamaan dari dokumen atau kata dari *dataset*.

6. Sastrawi untuk kosa kata dengan bahasa Indonesia, pada penelitian digunakan untuk menghilangkan *Stop word* dan melakukan *Stemming*

1.4.4. Batasan Masalah

Adapun batasan masalah agar tugas akhir ini dapat berfokus pada permasalahan yang ada, adalah sebagai berikut :

1. Kebutuhan data diperoleh dari dokumen yang diberikan Kantor Kerjasama dan Promosi (KKP) dan mengambil informasi dari situs Atma Jaya Yogyakarta.
2. *Chatbot* hanya dapat menjawab semua pertanyaan yang terdapat pada *dataset*.
3. *Chatbot* hanya dapat menjawab pertanyaan mengenai informasi universitas dan penerimaan mahasiswa baru program Sarjana atau S1.
4. Model yang digunakan untuk membuat *Chatbot* adalah *Retrieval Based* dengan penarikan respon menggunakan metode pembobotan *TFIDF* dan *cosine similarity*.
5. Pengaksesan data melalui web agak lama karena di *deploy* menggunakan *hosting* Heroku versi gratis.

1.4. Sumber Data

Untuk mendukung penelitian ini data yang digunakan pada penelitian ini didapatkan dari beberapa sumber yaitu :

1. Dokumen mengenai pertanyaan dan jawaban yang sering ditanyakan oleh calon mahasiswa yang diberikan oleh kantor kerjasama dan promosi (KKP) universitas Atma Jaya Yogyakarta, berupa file word.
2. Brosur Penerimaan Mahasiswa Baru Universitas Atma Jaya Yogyakarta
3. Buku Pedoman Penerimaan Mahasiswa Baru Universitas Atma Jaya Yogyakarta
4. Web Universitas Atma Jaya Yogyakarta (uajy.ac.id)

1.5. Metodologi Penelitian

Metode yang digunakan untuk pengumpulan data dan pembuatan model *Chatbot* informasi penerimaan mahasiswa baru adalah sebagai berikut :

1. Studi Pustaka

Studi pustaka dilakukan untuk mencari sumber pustaka yang berkaitan dengan aplikasi yang akan dibuat. Pada penelitian ini dilakukan pencarian sumber pustaka untuk model *Chatbot Retrieval Based*. Selain itu juga mencari pustaka tentang penggunaan *TFIDF* untuk pemberian bobot suatu kalimat. Studi pustaka membantu pekerjaan dalam hal pembuatan model dan memberikan informasi mengenai teori – teori yang sudah ada.

2. Pembuatan Dataset

Pembuatan *dataset* dilakukan secara manual dengan menginputkan data berupa Pertanyaan beserta jawabannya kedalam sebuah file bertipe .csv. Data yang diuji berupa pertanyaan yang sering ditanyakan oleh calon mahasiswa beserta jawabannya yang tepat untuk pertanyaan tersebut. Penulis mengambil data dari dokumen yang diberikan oleh Kantor Kerjasama dan Promosi Universitas Atma Jaya Yogyakarta dan juga mengambil informasi melalui situs web universitas dikarenakan calon mahasiswa cenderung akan mencari informasi dari web tersebut. *Dataset* ini hanya berisi dua buah atribut yaitu kolom *Question* dan kolom *Answer*. *Dataset* ini nantinya akan dibersihkan di dalam model yang dibuat.

3. Pembuatan Model

Model yang digunakan adalah *Retrieval Based* , model ini akan digunakan untuk menguji apakah pertanyaan yang diajukan sesuai dengan hasil yang diharapkan. Algoritma dari model ini adalah

1. Membersihkan *dataset* yang baru pertama kali dimasukan

2. Melakukan *preprocessing* data pertanyaan yang dimasukan oleh pengguna dan dari *dataset* yang akan digunakan jika *dataset* baru pertama kali dimasukan.
3. Memilih jawaban berdasarkan nilai *cosine similarity* dan jika tidak ada akan menampilkan informasi bahwa pertanyaan tersebut belum bisa dijawab oleh *Chatbot*.

4. Preprocessing

Dataset yang sudah dibuat akan di normalisasi, proses normalisasi ini akan melakukan beberapa hal yaitu. Menghilangkan karakter yang berulang di sebuah kata yang ada di kalimat. Menghilangkan karakter spesial seperti titik, koma, dan lain-lain. Menghilangkan kata-kata konjungsi seperti kata dan, atau, dan meskipun atau bisa disebut menghilangkan *stop word*. Melakukan tokenisasi dan menjadikannya ke *lowercase*, serta melakukan *stemming* untuk mencari kata dasar dari suatu kata. Lalu semua proses tadi digabungkan menjadi satu. Normalisasi tadi akan di lanjutkan dengan melakukan pembobotan menggunakan algoritma *TFIDF* yang akan menghasilkan nilai bobotnya yang akan digunakan untuk menghitung nilai *cosine similarity*. Proses *Preprocessing* akan ada di dalam model untuk memproses masukan pertanyaan dari pengguna yang tidak ada di *dataset*. Hasil dari *preprocessing* juga akan digunakan untuk pembuatan *Webservice*.

5. Evaluasi Model

Untuk mengetahui seberapa baik akurasi model, model akan dievaluasi dengan cara memasukan pertanyaan yang akan langsung dicek jawabannya sesuai atau tidak menurut pengguna. Perhitungan dari akurasinya akan dihitung dengan cara jumlah jawaban benar dibagi jumlah pertanyaan yang diajukan lalu dikalikan 100%.

6. Pembuatan *Webservice*

Hasil dari pelatihan data atau preprocessing data akan menghasilkan sebuah vektor *TFIDF* yang berupa token serta bobotnya. *Webservice* akan dibuat menggunakan hasil pelatihan data tersebut serta menggunakan *dataset* yang sudah dibuat. Pada *Webservice* akan mengirimkan data berupa jawaban dan nilai *cosine similarity* yang dihitung dari nilai *TFIDF* hasil pelatihan dengan nilai *TFIDF* dari pertanyaan yang diajukan, proses penghitungan *cosine similarity* dilakukan di dalam *web service*. *Webservice* dibuat untuk menghubungkan pengguna dengan *Chatbot* yang dibuat.

7. Pengintegrasian dengan Web

Web yang dibuat berupa sebuah tampilan antarmuka yang mirip aplikasi chatting dimana pengguna akan saling ngobrol dengan *Chatbot*. Tampilan dibuat mirip aplikasi chatting agar pengguna sudah merasa terbiasa dan langsung mengerti ketika menggunakannya. *Webservice* yang sudah dibuat tadi akan diintegrasikan dengan web yang sudah dibuat.

1.6. Sistematika Penyajian

Secara sistematis isi dari laporan ini disusun sebagai berikut:

BAB I : PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan pembahasan, ruang lingkup kajian, sumber data, metodologi penelitian dan sistematika penyajian.

BAB II : TINJAUAN PUSTAKA

Bab ini berisi penjelasan dari penelitian–penelitian sebelumnya yang telah dilakukan dan digunakan sebagai bahan referensi dalam pemecahan masalah.

BAB III : LANDASAN TEORI

Bab ini berisi dasar–dasar teori yang digunakan sebagai pedoman dan acuan dalam pemecahan masalah.

BAB IV : ANALISIS DAN PEMBAHASAN

Bab ini berisi penjelesan mengenai analisis dan perancangan model *Chatbot* yang dibuat beserta *datasetnya*.

BAB V : IMPLEMENTASI MODEL KEDALAM WEBSERVICE DAN WEB

Bab ini berisi penjelasan mengenai bagaimana model ini dapat di implementasikan kedalam *web service* dan *web* .

BAB VI : PENUTUP

Bab ini berisi kesimpulan mengenai model yang telah dibuat beserta saran–saran yang bermanfaat untuk pengembangan lebih lanjut.