

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Salah satu bentuk komunikasi atau interaksi manusia dengan komputer yaitu diterapkan pada Sistem Tanya Jawab. Sistem tanya jawab atau *Question Answering (QA) System* adalah suatu sistem yang membantu manusia untuk mendapatkan informasi dengan memberikan jawaban secara otomatis atas pertanyaan-pertanyaan yang diajukan oleh manusia dalam bahasa alami. Pemikiran mendasar di balik sistem QA adalah untuk membantu dan mempermudah interaksi antara manusia dan mesin (komputer) [1]. *Recurrent Neural Network (RNN)* adalah salah satu bagian dari keluarga *Neural Network* untuk memproses data yang bersambung (*sequential data*) dengan cara menyimpan informasi dari masa lalu atau pada proses sebelumnya, yaitu dengan melakukan perulangan atau *looping* di dalam arsitekturnya. Proses perulangan tersebut membuat setiap langkah proses dari RNN selalu dipengaruhi oleh keluaran yang dihasilkan oleh proses sebelumnya. RNN termasuk algoritma yang banyak digunakan saat ini untuk di bidang *Natural Language Processing* karena mampu mengingat hal-hal penting tentang masukan yang diterima, yang memungkinkan akan sangat tepat dalam memprediksi apa yang akan terjadi selanjutnya [2].

Beberapa penelitian telah dilakukan sebelumnya terkait sistem tanya jawab, salah satunya dilakukan oleh Ivan Karl Bindoff, BComp, yang menggunakan metode *Multiple Classification Ripple Down Rules (MCRDR)* untuk klasifikasi resep obat berdasarkan masukan berupa keluhan dari pasien. Sistem tanya jawab dari hasil penelitiannya mendapatkan nilai akurasi sebesar 60% [3]. Yudi Kawula Maharlika mencoba mengembangkan penelitian Ivan Karl Bindoff, BComp, dengan menggabungkan metode *rule-based MCRDR* dengan metode *keyword-matching* untuk sistem tanya jawab berbahasa Indonesia pada *Front Office* [4]. Sistem tanya jawab dalam penelitiannya

menghasilkan nilai akurasi yang lebih baik, yaitu 66,67%. Akan tetapi, sistem tanya jawab dengan *rule-based* sering kali bermasalah pada terbatasnya *rule* atau aturan, sehingga kurangnya aturan tersebut menyebabkan sistem tidak menemukan jawaban. Penelitian yang dilakukan oleh Denis Lukovnikov, Asja Fischer, Jens Lehmann, dan Sören Auer tidak lagi menggunakan *rule-based* pada sistem tanya jawabnya. Mereka melakukan penelitian tentang sistem tanya jawab melalui grafik pengetahuan pada tingkat kata dan karakter dengan *neural network-based*. Mereka menggunakan salah satu mekanisme dari algoritma *Recurrent Neural Network* (RNN), yaitu *Gated Recurrent Unit* (GRU) [5]. Dari nilai akurasi yang dihasilkan, sebesar 71,2%, sistem tanya jawab pada penelitian tersebut dapat dikatakan lebih baik dari dua penelitian sebelumnya yang menggunakan *rule-based*. Akan tetapi, penelitian-penelitian terkait sistem tanya jawab yang menggunakan *neural network-based*, khususnya RNN, masih menggunakan data berbahasa Inggris atau bahasa yang lain selain bahasa Indonesia.

Berdasarkan hal tersebut, untuk mengembangkan penelitian sebelumnya mengenai sistem tanya jawab berbahasa Indonesia pada *Front Office* maka dilakukan penelitian dengan mengimplementasikan *Recurrent Neural Network* pada kasus sistem tanya jawab berbahasa Indonesia.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang ada, diidentifikasi permasalahan, yaitu:

1. Sistem tanya jawab dengan *rule-based* sering bermasalah pada kurangnya aturan yang menyebabkan jawaban yang dihasilkan tidak sesuai, sehingga akurasi sistem tanya jawab dengan *rule-based* masih kurang baik.
2. Sistem tanya jawab yang menggunakan *neural network-based*, khususnya RNN, menghasilkan akurasi yang lebih baik

dibandingkan sistem tanya jawab dengan *rule-based*. Akan tetapi, masih menggunakan data berbahasa Inggris atau bahasa yang lain selain bahasa Indonesia.

3. Belum diketahuinya seberapa besar nilai akurasi dari sistem tanya jawab berbahasa Indonesia dengan penerapan *Recurrent Neural Network*.

### 1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk melakukan analisis dan mengimplementasikan *Recurrent Neural Network* pada kasus aplikasi tanya jawab berbahasa Indonesia. Adapun tujuan dari penelitian ini adalah untuk mengukur seberapa besar akurasi yang dihasilkan dari penerapan *Recurrent Neural Network* pada kasus aplikasi tanya jawab berbahasa Indonesia.

### 1.4 Batasan Masalah

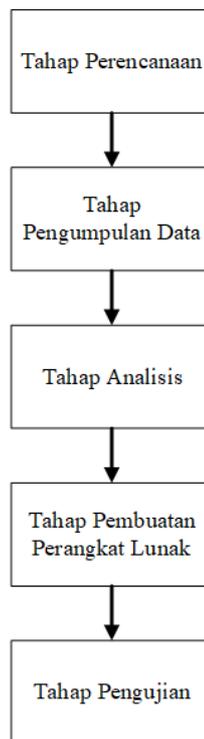
Adapun batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Data masukan berupa teks berisi seputar pertanyaan dalam bahasa Indonesia dengan format baku mengenai pendaftaran mahasiswa baru dan pembayaran kuliah di Universitas Komputer Indonesia yang biasa dilakukan di *Front Office* Universitas Komputer Indonesia.
2. *Data set* yang digunakan untuk pelatihan dan pengujian terdiri dari 100 data pertanyaan yang bersumber dari penelitian yang dilakukan sebelumnya oleh Yudi Kawula Maharlika [4] dan Zamaludin Abdulah [6].
3. Pada tahap *preprocessing*, metode-metode yang digunakan adalah metode *Case Folding*, *Filtering*, *Tokenization*, *Removing Stopword*, TF-IDF, dan *Normalization*.

4. Data keluaran berupa jawaban yang relevan dari pertanyaan yang diberikan dalam bentuk teks, dan nilai akurasi yang dihasilkan dari aplikasi tanya jawab.

### 1.5 Metode Penelitian

Berikut ini adalah alur dari penelitian yang dilakukan, yaitu pada Gambar 1.1.



Gambar 1.1 Alur Penelitian

#### 1.5.1 Tahap Perencanaan

Tahapan perencanaan dalam penelitian ini memiliki beberapa tahapan, yaitu:

1. Menetapkan tujuan penelitian

Penetapan tujuan penelitian dilakukan agar tujuan dalam penulisan lebih jelas dan dapat menjadi solusi dari permasalahan yang ada, yaitu mengukur seberapa besar

kinerja dari penerapan *Recurrent Neural Network* pada kasus aplikasi tanya jawab berbahasa Indonesia.

## 2. Studi pustaka

Dengan melakukan studi pustaka, informasi terkait penelitian tentang Implementasi *Recurrent Neural Network* pada Kasus Aplikasi Tanya Jawab semakin bertambah sehingga membantu penelitian ini berjalan dengan lancar. Studi pustaka yang dilakukan, yaitu membaca dan mempelajari hasil penelitian sebelumnya, buku, jurnal, dan artikel tentang *Question Answering System*, *Recurrent Neural Network*, dan lainnya yang terkait dengan penelitian ini.

## 3. Identifikasi masalah

Identifikasi masalah perlu dilakukan untuk menemukan permasalahan dalam penelitian yang dilakukan. Masalah yang telah diidentifikasi, yaitu belum diketahuinya seberapa besar nilai akurasi dari sistem tanya jawab berbahasa Indonesia dengan penerapan *Recurrent Neural Network*.

## 4. Menentukan data yang diperlukan

Menentukan data yang diperlukan, yaitu data set untuk pelatihan dan pengujian, dilakukan agar mempermudah saat melakukan analisis.

### 1.5.2 Tahap Pengumpulan Data

Metode pengumpulan data yang dilakukan untuk penelitian ini adalah metode Studi Pustaka dan Literatur. Metode pengumpulan data ini dilakukan dengan cara mengumpulkan dan mempelajari sumber kepustakaan, yang berupa penelitian sebelumnya, buku referensi, artikel dan jurnal yang terkait dengan judul penelitian untuk membantu penelitian, seperti penelitian Ivan Karl Bindoff [3], Yudi Kawula Maharlika [4], dan Denis Lukovnikov [5]. Pada tahap ini juga dilakukan pengumpulan *data set* yang bersumber dari penelitian sebelumnya yang dilakukan oleh Zamaludin Abdulah [6].

### 1.5.3 Tahap Analisis

Analisis yang dilakukan dalam penelitian ini meliputi:

1. Analisis Data Masukan

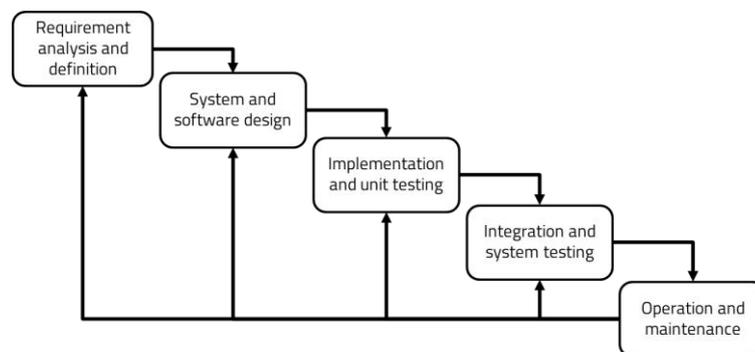
Analisis ini dilakukan agar dapat mengolah data yang digunakan menjadi informasi sehingga karakteristik data tersebut bisa dipahami dan bermanfaat untuk permasalahan yang berkaitan dengan penelitian.

2. Analisis Algoritma

Analisis ini dilakukan untuk mengetahui karakteristik dan performa dari algoritma yang digunakan dalam menyelesaikan masalah, dalam penelitian ini, yaitu algoritma *Recurrent Neural Network* (RNN).

### 1.5.4 Tahap Pembuatan Perangkat Lunak

Pembuatan perangkat lunak atau aplikasi menggunakan metode *waterfall*, yang meliputi beberapa tahapan, yaitu seperti pada Gambar 1.2 [7].



Gambar 1.2 Tahap-tahap Metode *Waterfall*

Berikut adalah penjelasan dari tahapan metode *waterfall* [7]:

1. *Requirement analysis and definition*

Tahap *requirement analysis and definition* adalah tahap di mana pengumpulan kebutuhan telah terdefinisi secara lengkap kemudian dianalisis dan didefinisikan kebutuhan yang harus

dipenuhi oleh program yang dibangun. Fase ini harus dikerjakan secara lengkap untuk bisa menghasilkan desain yang lengkap.

2. *System and software design*

Tahap *system and software design* merupakan tahap mendesain perangkat lunak yang dikerjakan setelah kebutuhan selesai dikumpulkan secara lengkap.

3. *Implementation and unit testing*

Tahap *requirement analysis and definition* merupakan tahap hasil desain program diterjemahkan ke dalam kode-kode dengan menggunakan bahasa pemrograman yang sudah ditentukan. Program yang dibangun langsung diuji berdasarkan unit-unitnya.

4. *Integration and system testing*

Tahap *integration and system testing* merupakan tahap penyatuan unit-unit program kemudian sistem diuji secara keseluruhan.

5. *Operation and maintenance*

Tahap *operation and maintenance* merupakan tahap mengoperasikan program dan melakukan pemeliharaan, seperti penyesuaian karena adaptasi dengan situasi yang sebenarnya.

### **1.5.5 Tahap Pengujian Tujuan Penelitian**

Tahap pengujian tujuan penelitian dilakukan dengan menghitung seberapa besar nilai akurasi atau ketepatan yang dihasilkan dari pengimplementasian *Recurrent Neural Network* pada kasus aplikasi tanya jawab. Pengujian ini dilakukan untuk mengetahui apakah aplikasi sudah sesuai atau belum dengan tujuan yang telah ditentukan dari penelitian. Metode Pengujian yang digunakan pada penelitian ini adalah metode *Confusion Matrix*.

## 1.6 Sistematika Penulisan

Sistematika penulisan ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan secara umum adalah sebagai berikut:

### **BAB 1 PENDAHULUAN**

Berisi penjelasan tentang latar belakang mengapa penelitian ini dilakukan, rumusan masalah berdasarkan latar belakang, maksud dan tujuan, batasan-batasan dari penelitian yang dilakukan, metodologi penelitian yang digunakan, serta sistematika penulisan untuk penelitian yang dilakukan.

### **BAB 2 LANDASAN TEORI**

Berisi teori dan konsep yang digunakan sebagai landasan atau acuan yang diperlukan, yaitu antara lain penjelasan terkait penelitian yang dilakukan seperti tentang *Artificial Intelligent*, *Recurrent Neural Network*, *Question Answering*, dan sebagainya.

### **BAB 3 ANALISIS DAN PERANCANGAN**

Berisi analisis dan penjelasan rancangan aplikasi, mulai dari tujuan perangkat lunak yang digunakan, pembuatan *prototype* aplikasi, dan analisis data dan algoritma pada aplikasi.

### **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Berisi implementasi dari hasil analisis dan perancangan yang telah dibuat dan disertai juga dengan hasil pengujian mengenai tingkat keakurasian dari implementasi aplikasi.

### **BAB 5 KESIMPULAN DAN SARAN**

Berisi kesimpulan dari hasil pengujian yang dilakukan, apakah tujuan awal dalam penelitian telah tercapai atau tidak, serta berisi saran untuk penelitian atau pengembangan selanjutnya.