BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Sistem Tanya Jawab *(Question Answering System)* merupakan sebuah sistem pada komputer yang dapat menjawab pertanyaan menggunakan bahasa alami yang biasa dipakai manusia secara otomatis. Jawaban yang diperoleh dari sistem berisi informasi yang berasal dari sebuah sumber basis data [1]. Sebuah sistem tanya jawab, diharapkan dapat membantu orang untuk mendapatkan informasi lebih mengenai hal tertentu tanpa harus bertanya kepada orang lain [2].

Pada penelitian sebelumnya tentang sistem tanya jawab dalam bahasa Indonesia sudah dilakukan, salah satunya dengan menggunakan *rule base* [2] [3]. Pada penelitian tersebut belum menghasilkan akurasi yang cukup baik yaitu 60% dan memiliki kekurangan yaitu hanya mendeteksi kata kunci yang terdapat pada *rule*, sehingga jika menemukan kata kunci selain pada *rule* maka menghasilkan jawaban tidak relevan.

Mengatasi kata kunci yang tidak terdapat pada *rule* dapat diterapkan metode pendekatan *Machine Learning* [4]. Pada penelitian tersebut menjelaskan dengan menggunakan pendekatan *Machine Learning* dapat menghasilkan sistem yang mencakup ribuan fitur pertanyaan dan mengklasifikasikan pertanyaan tersebut secara otomatis dan meningkatkan produktivitas sistem tanya jawab.

Adapun penelitian lain tentang sistem tanya jawab yang menggunakan 5 metode pendekatan *Machine Learning* yaitu *K-Nearest Neighbor*, *Naive Baiyes*, *Support Vector Machine* (SVM), *Sparse Network of Winnows* (SNoW) dan *Decision Tree* (DT) [5]. Pada penelitian tersebut metode SVM menghasilkan akurasi sebesar 87,4%. Penelitian tersebut menggunakan bahasa Inggris.

Adapun penelitian lainnya mengenai sistem tanya jawab dengan menggunakan metode SVM. Hasil yang didapat yakni nilai akurasi diatas 80-90% [6] [7] [8] [9]. Pada penelitian tersebut menggunakan bahasa Inggris-Arab.

Oleh karena itu, pada penelitian ini bermaksud untuk mengembangkan sistem tanya jawab pada *Front Office* UNIKOM dengan menerapkan metode

pendekatan *machine learning Support Vector Machine* (SVM) dalam bahasa indonesia. Penelitian ini diharapkan dapat menghasilkan dan akurasi yang cukup baik.

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan di atas maka dapat diidentifikasi permasalahan yaitu belum diketahuinya nilai akurasi dari sistem tanya jawab dengan menerapkan metode pendekatan *machine learning* SVM dalam Bahasa Indonesia.

1.3. Maksud dan Tujuan

1.3.1. Maksud

Maksud dari penelitian ini adalah untuk mengembangkan sistem tanya jawab pada penelitian sebelumnya dengan menerapkan metode pendekatan *machine learning* SVM untuk *Front Office* UNIKOM dalam bahasa indonesia.

1.3.2. Tujuan

Tujuan dari penelitian ini adalah untuk mengukur akurasi yang dihasilkan dari sistem tanya jawab dengan menerapkan metode SVM pada kasus *Front Office* dalam bahasa indonesia.

1.4. Batasan Masalah

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

Data Masukan:

1. Pertanyaan yang digunakan pada sistem tanya jawab menggunakan Bahasa Indonesia.
2. Format *file* masukan yaitu data berupa *text* yang berasal dari *database*.
3. Cakupan pertanyaan adalah informasi-informasi mengenai Autodebet, Pendaftaran Mahasiswa Baru (PMB), pertanyaan umum jurusan dan mengenai kampus UNIKOM.

4. Data pertanyaan tidak mengandung singkatan kata.
5. Data yang digunakan sebanyak 146 data yang terdiri dari 82 data latih, 59 data dari penelitian sebelumnya [2], 23 data dari hasil kuesioner *online* terhadap pengunjung dan 64 data uji dari hasil kuesioner *online* pengunjung *front-office*.
6. Data Autodebet dan PMB yang digunakan sebagai jawaban berdasarkan data semester ganjil tahun 2017.

Proses:

1. Tahap *preprocessing* yang dilakukan adalah *case folding*, *cleansing* (penghilangan simbol), *tokenizing*, *remove stopwords*, dan TF-IDF.
2. Klasifikasi pertanyaan dilakukan dengan menggunakan *Library Support Vector Machine* (LibSVM*)*.
3. Kernel SVM yang akan digunakan adalah RBF.
4. Metode *Multi Class Support Vector Machine* yang digunakan adalah *Multiclass One Against One.*
5. Kelas yang digunakan berdasarkan cakupan tema yang telah ditentukan yaitu Waktu Autodebet, Informasi PMB, Waktu PMB, Lokasi PMB, Informasi Autodebet, Biaya PMB, Biaya Autodebet
6. Sistem akan memberikan output atau hasil berupa 1 jawaban yang sesuai dengan pertanyaan yang diinputkan.

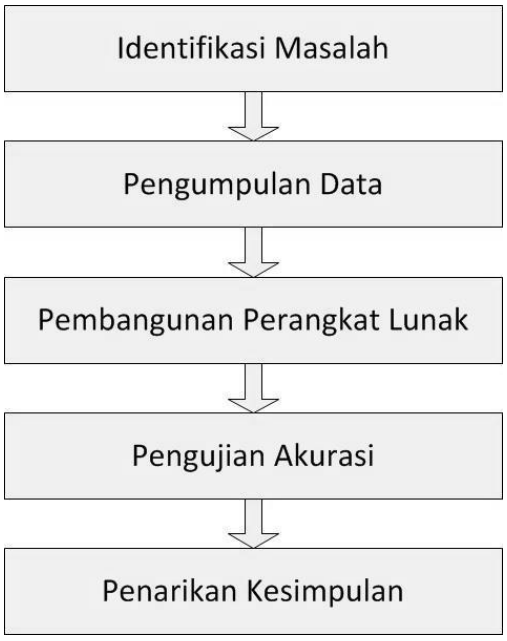
Output:

1. Proses dan jawaban dari pertanyaan yang diinputkan
2. Nilai akurasi yang dihasilkan.

1.5. Metode Penelitian

Metode penelitian yang digunakan adalah metode deskriptif. Metode ini dipilih karena sesuai dengan jangkauan penelitian yang hanya menjangkau populasi tertentu, selain itu juga sistem yang akan dibangun dapat menggambarkan fakta-fakta yang didapat dari tahap pengumpulan data [10]. Pada penelitian ini dimulai dari identifikasi masalah yang terdapat pada penelitian sebelumnya, pengumpulan data berupa studi literatur yang berhubungan dengan

penelitian sebelumnya dan pengumpulan data melalui wawancara dengan narasumber terhadap penelitian yang sedang berjalan, implementasi kedalam perangkat lunak hingga mendapatkan kesimpulan dari penelitian yang telah dilakukan. Alur penelitian yang akan digunakan dalam penelitian ini dapat dilihat pada Gambar 1.1.



Gambar 1.1 Metode Penelitian

1.5.1. Identifikasi Masalah

Identifikasi masalah dengan cara menganalisis penelitian sebelumnya yang sudah dilakukan dimana salah satu hasil analisis adalah belum diketahuinya akurasi dengan menerapkan metode pendekatan *machine learning* SVM dan apakah dapat meningkatkan akurasi dari jawaban yang dihasilkan dari sistem tanya jawab berbahasa Indonesia pada kasus *Front Office* atau tidak.

1.5.2. Pengumpulan data

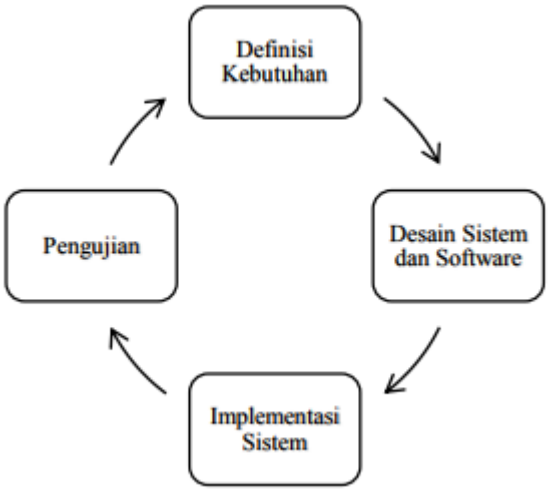
Metode pengumpulan data yang digunakan dalam penelitian ini adalah dengan menggunakan metode studi literatur dan metode wawancara. Studi literatur adalah cara pengumpulan data dengan menelusuri sumber-sumber tulisan yang pernah disusun sebelumnya. Tulisan tersebut berupa hasil penelitian yang

berhubungan dengan sistem tanya jawab dengan menggunakan metode SVM dan metode SVM.

Wawancara adalah suatu teknik pengumpulan data yang dilakukan dengan tanya jawab secara lisan, baik langsung atau tidak langsung dengan sumber data responden (terwawancara) [10]. Pada penelitian ini yang menjadi narasumber adalah mahasiswa unikom. Topik yang ditanyakan mengenai pertanyaan yang sering diajukan ketika sedang berada di *front office*.

1.5.3. Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan adalah metode *prototyping*. Metode ini dipilih karena hasil dari pembangunan dapat di evaluasi kembali jika terjadi kesalahan. Berikut adalah tahapan-tahapan pembangunan perangkat lunak menggunakan metode *prototyping.*



Gambar 1.2 Model *Prototyping* [11]

1. Definisi Kebutuhan

   Pada tahap ini, menetapkan seluruh kebutuhan pembangunan sistem tanya jawab. Kebutuhan tersebut meliputi analisis data pertanyaan sebagai data masukan hingga klasifikasi data pertanyaan, penggunaan metode *support vector machine* sebagai klasifikasi data pertanyaan, dan penggunaan metode *cosine similarity* untuk pemberian jawaban dari sistem tanya jawab.

2. Desain Sistem dan *Software*

   Dalam tahap ini membuat gambaran yang dikerjakan dan bagaimana tampilan dari sistem yang dibangun.

3. Implementasi Sistem

   Tahap ini merupakan implementasi dari metode dengan pemrograman (*coding*). Pembuatan *software* dipecah menjadi modul-modul kecil yang nantinya akan digabungkan dalam sistem.

4. Pengujian

   Tahap ini dilakukan pengujian sistem terhadap modul yang dibuat dengan *Black-box testing*, apakah sudah memenuhi fungsi yang diinginkan atau belum.

1.5.4. Pengujian Akurasi

Pada tahap ini, hasil dari sistem tanya jawab diukur dari akurasi kebenaran jawaban dengan cara validasi oleh petugas PMB Unikom. Akurasi tersebut dihitung menggunakan perhitungan akurasi.

1.5.5. Penarikan Kesimpulan

Pada tahap ini akan dilakukan analisis kesimpulan berdasarkan hasil dari pengujian sistem pada pembangunan perangkat lunak. Hasil yang diambil adalah dengan mencocokkan jawaban yang dihasilkan dari sistem tanya jawab dengan sumber data yang valid yang dijadikan sebagai tolak ukur dari akurasi sistem tanya jawab. kesimpulan dibuat menurut analisis yang sudah dilakukan.

1.6. Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum mengenai penelitian yang dikerjakan. Sistematika penulisan penelitian sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini berisi tentang penguraian latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian, serta sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini berisi tentang bahasan konsep dasar dan teori-teori yang berkaitan dengan topik penelitian yang dilakukan seperti Sistem Tanya Jawab (*question and answering*), teori *preprocessing* meliputi *case folding, cleansing, tokenizing, remove stopword,* metode pembobotan dengan TF-IDF, metode klasifikasi dengan SVM, metode *cosine similarity* dan metode yang digunakan dan hal-hal yang berguna dalam proses analisis dan perancangan sistem.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini berisi mengenai analisis data, analisis kebutuhan perangkat lunak, serta perancangan sistem yang meliputi perancangan struktur menu, alur data, basis data, dan jaringan semantik.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang hasil implementasi dari rancangan sistem yang telah dibangun sesuai dengan perancangan sistem yang telah dibuat, juga disertakan pengujian dari sistem tersebut. Pengujian akan dilakukan secara *offline* yakni membandingkan jawaban yang relevan terhadap pertanyaan yang diajukan oleh pengunjung atau pengguna.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dari hasil pembangunan sistem yang telah diimplementasikan disertai juga dengan saran yang diharapkan dapat berguna untuk tahap penelitian lanjutan dari sistem ini.