

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Keberhasilan dalam proses belajar mengajar biasanya ditentukan oleh salah satu bentuk evaluasi dari proses pembelajaran yang biasa disebut dengan ujian. Ujian dengan sistem esai merupakan bentuk evaluasi di mana pilihan jawaban tidak disediakan, dan siswa harus menjawab dengan kalimat[1]. Ujian dengan sistem esai dapat melatih dalam menyampaikan sesuatu informasi secara verbal.

Salah satu penerapan ujian dengan sistem esai adalah dalam proses *e-learning*. Perangkat komputer yang menjadi komponen penting dalam proses *e-learning* lebih mudah dalam melakukan penilaian ujian pilihan ganda dan ujian singkat secara akurat dibandingkan dengan melakukan penilaian jenis ujian esai. Hal ini yang melatar belakangi lahirnya berbagai macam bentuk penilaian yang didesain untuk membantu para guru, dosen, dan tenaga pengajar dalam menyelesaikan penilaian soal esai dengan komputerisasi.

Mengoreksi jawaban dari soal esai secara manual adalah hal yang membutuhkan waktu lama, selain itu terdapat banyak kesulitan jika harus mengoreksi jawaban esai secara manual. Untuk dapat mengoreksi jawaban dengan cepat dan akurat maka dibutuhkan sistem untuk melakukan otomatisasi penilaian

jawaban esai. Dengan adanya sistem ini, mengoreksi jawaban soal esai akan menjadi lebih mudah dan tidak memerlukan waktu yang lama.

Berbagai algoritma yang biasa digunakan untuk mendeteksi kemiripan teks adalah algoritma Rabin Karp, Smith Waterfall dan Cosine Similarity. Algoritma Cosine Similarity adalah suatu algoritma untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah *vector* dengan menggunakan *keywords* (kata kunci) dari sebuah dokumen sebagai ukuran[2]. Penulis memilih Algoritma Cosine Similarity karena memiliki tingkat keakurasian yang tinggi. Selain itu alasan lain penulis memilih cosine similaity adalah karena metode ini tidak terpengaruh terhadap panjang pendeknya suatu dokumen. Karena yang diperhitungkan hanya nilai term dari masing-masing dokumen. semakin besar jumlah kemunculan kata (term) dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar[3].

Berdasarkan uraian tersebut, penulis berharap dengan adanya otomatisasi dalam penilaian soal tipe esai ini dapat membantu mempercepat dan mempermudah tenaga pengajar dalam mengoreksi jawaban dari soal esai secara akurat dan efisien.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang diatas, penulis merumuskan masalah sebagai berikut:

1. Bagaimana proses penerapan algoritma *Cosine Similarity* dalam otomatisasi penilaian jawaban esai?

2. Bagaimana tingkat akurasi dari algoritma *Cosine Similarity*?

### 1.3 Batasan Penelitian

Berdasarkan rumusan masalah diatas, maka penulis membuat batasan-batasan masalah sebagai berikut:

1. Sistem dibuat untuk mengoreksi jawaban esai berbahasa indonesia yang merupakan jawaban dari pertanyaan pada sistem, bukan merupakan esai karangan yang panjang.
2. Data meliputi soal-soal beserta jawaban sistem yang telah dikumpulkan dari tenaga pengajar.
3. Sistem ini menggunakan algoritma *Cosine Similarity* dan pembobotan TF-IDF.
4. Sistem ini memiliki 3 user yaitu admin, guru, dan siswa.
5. Sistem ini berbasis website.
6. Penilaian dilakukan dengan cara mencocokkan jawaban ujian siswa dan kunci jawaban soal.

### 1.4 Maksud dan Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka ditetapkan maksud dan tujuan dari penelitian ini adalah:

1. Implementasi *text mining* dengan menggunakan algoritma *Cosine Similarity* untuk menganalisis jawaban esai.



2. Menganalisis tingkat keakuratan kesamaan pada jawaban siswa dan kunci jawaban.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Bagi penulis, menambah wawasan penulis mengenai *text mining* terutama algoritma *Cosine Similarity*.
2. Bagi Masyarakat, diharapkan dapat memudahkan dalam mengoreksi jawaban esai secara akurat, karena penelitian ini dilakukan untuk mempermudah dan mempersingkat proses pengoreksian jawaban esai melalui otomatisasi penganalisaan dan pengoreksian menggunakan *text mining*.
3. Bagi Universitas Amikom Yogyakarta, dapat dijadikan sebagai bahan pertimbangan atau dikembangkan lebih lanjut, serta referensi terhadap penelitian yang sejenis bagi mahasiswa yang sedang menyelesaikan skripsi.

### 1.6 Metode Penelitian

Penelitian ini menggunakan metode penelitian dan pengembangan (*Research and Development*), yakni metode penelitian yang digunakan untuk menghasilkan produk tertentu dan menguji efektivitas produk tersebut.

Dalam hal ini, penulis menghasilkan produk berupa sistem *text mining* dan untuk menguji efektivitas produk tersebut dengan melihat apakah produk yang dihasilkan dapat bekerja sesuai dengan kegunaannya.

### 1.6.1 Metode Pengumpulan Data

Berdasarkan sumber data yang digunakan pada penelitian ini, maka metode pengumpulan data yang penulis gunakan adalah sebagai berikut:

#### 1. Pengamatan Langsung (*Observasi*)

Dilakukan pengamatan secara langsung terhadap sistem yang sudah ada (manual), sehingga dapat dengan jelas terlihat alur data (informasi) yang sedang berjalan.

#### 2. Studi Pustaka (*Study Library*)

Studi pustaka adalah mempelajari karya ilmiah, buku ilmiah, dan sumber ilmiah lainnya yang sesuai dengan penelitian dan memiliki hubungan dengan masalah yang diteliti. Pada penelitian ini, penulis mempelajari dan mengumpulkan data-data teoritis dari buku, jurnal, internet, serta literatur yang ada dengan maksud untuk mendapatkan teori-teori dan bahan-bahan yang berkaitan dengan pembuatan dan masalah yang diambil.

#### 3. Tenaga Pengajar

Mengumpulkan data-data soal, kunci jawaban, dan jawaban soal dari para siswa yang bersumber dari tenaga pengajar yang nantinya digunakan untuk proses testing.

### 1.6.2 Metode Analisis

Pada tahap ini digunakan untuk mengolah data yang ada dan kemudian melakukan analisis terhadap hasil studi literatur yang diperoleh sehingga menjadi sebuah informasi. Dalam text mining terdapat istilah *preprocessing* data, yaitu proses pendahulu yang diterapkan terhadap data teks yang bertujuan untuk menghasilkan data *numerik*. Pada proses *preprocessing* merupakan tahap di mana deskripsi ditangani untuk dapat siap diproses memasuki tahap *text mining*[4]. *Text mining* penelitian ini terdiri dari *text preprocessing* yaitu *case folding*, *tokenizing*, *stopword removal/ filtering*, *stemming*, dan *analyzing*, tahap pembobotan term dengan metode *tf-idf*, tahap *Cosine Similarity* dan evaluasi menggunakan *confusion matrix*.

### 1.6.3 Metode Perancangan

Implementasi dari penelitian ini adalah sistem ujian esai berbasis komputer yang memiliki sistem penilaian otomatis menggunakan Algoritma *Cosine Similarity*. Penelitian ini menggunakan sebagai perancangan, yaitu *use case diagram* yang berguna untuk merepresentasikan kebutuhan fungsional dari sistem, *class diagram* yang berguna untuk merepresentasikan tabel pada database dan relasinya, *activity diagram* yang berguna untuk merepresentasikan alur kerja dari sistem yang akan dibuat, *sequence diagram* untuk menunjukkan rangkaian pesan yang dikirim dan juga interaksi antar object, Bootstrap sebagai pihak ketiga dalam perancangan antarmuka.

Dalam melakukan perancangan sistem ini terdapat beberapa bahasa pemrograman yang digunakan. Bahasa pemrograman adalah sekumpulan instruksi



standar untuk memerintah komputer. Bahasa pemrograman ini merupakan suatu himpunan dari aturan sintaks dan semantik yang dipakai untuk mendefinisikan program komputer. Bahasa pemrograman yang digunakan pada penelitian ini adalah bahasa pemrograman PHP.

#### 1.6.4 Metode Coding

Metode coding yang digunakan pada penelitian ini yaitu *waterfall*. Metode pengembangan *waterfall* adalah model pengembangan sistem yang bersifat sistematis, serta berurutan dalam pengembangan *software*, mulai dari kebutuhan analisis, perancangan, implementasi, pengujian dan penerapan serta pemeliharaan sistem. Jadi jika langkah pertama belum dikerjakan, maka langkah kedua tidak dapat dikerjakan.

#### 1.6.5 Metode Testing

Metode testing yang digunakan oleh peneliti untuk menguji sistem menggunakan dua macam pengujian, yaitu menggunakan metode *white-box testing* dan *black-box testing* sebagai pengukuran kualitas sistem pendukung keputusan yang akan dibangun, dengan mencari kemungkinan *error* (kesalahan) yang ada pada program yang selanjutnya dilakukan evaluasi dan memperbaiki kesalahan yang terjadi.

Lalu untuk menguji tingkat keakurasian dari hasil perhitungan yang sudah dilakukan akan digunakan metode evaluasi *Confusion Matrix*. *Confusion Matrix* adalah sebuah metode yang biasa digunakan untuk perhitungan akurasi. Dalam pengujian keakuratan hasil pencarian akan dievaluasi nilai *recall*, *precision*, *accuracy*, dan *error rate*.

## 1.7 Sistematika Penulisan

Sistematika yang digunakan dalam penulisan penelitian ini akan diuraikan dengan pokok-pokok bahasan yang telah disusun sebagai berikut:

### **BAB I PENDAHULUAN**

Pada bab ini, berisi latar belakang masalah, rumusan masalah, batasan penelitian, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan

### **BAB II LANDASAN TEORI**

Pada bab ini menjelaskan teori-teori yang digunakan dalam penelitian dan membahas pengertian program yang digunakan dalam mengimplementasikan perangkat lunak yang dibangun

### **BAB III ANALISIS DAN PERANCANGAN**

Pada bab ini, membahas tentang penganalisaan dan perancangan aplikasi yang akan dibangun.

### **BAB IV IMPLEMENTASI & PEMBAHASAN**

Pada bab ini, berisi penjelasan tentang pembuatan *database*, rancangan tampilan antarmuka, serta cara kerja program.

### **BAB V PENUTUP**

Pada bab ini, berisi kesimpulan dan saran dari apa yang sudah dilakukan dalam penelitian Tugas Akhir.



## **DAFTAR PUSTAKA**

Berisi keterangan referensi dan acuan proses pembuatan skripsi dari buku-buku, jurnal dan atau sumber yang lainnya

