

Bab I Pendahuluan

1.1 Latar Belakang

Teknologi tercipta untuk membantu pekerjaan manusia agar lebih cepat dan tepat. Dengan berkembangnya teknologi informasi maka pengolahan data akan lebih mudah untuk dilakukan. Banyaknya data di internet maka dapat diolah untuk dijadikan informasi yang bermanfaat bagi orang lain. Pekerjaan yang biasanya dilakukan secara manual oleh manusia, kini sudah dapat digantikan oleh teknologi informasi. Dengan memanfaatkan teknologi informasi, maka dapat mempercepat waktu dan mengurangi tenaga yang dibutuhkan dibanding dilakukan oleh manusia. Selain memakan waktu yang cukup lama untuk membaca semua dokumen, kelemahan dari pengecekan manual adalah terkadang nilai yang diberikan masih bersifat subjektif dan tidak berimbang.

Menurut Kamus Besar Bahasa Indonesia, kemiripan adalah hampir sama atau serupa. Dalam tugas akhir ini akan dibahas mengenai cara untuk mengetahui tingkat kemiripan suatu tugas akhir dengan tugas akhir lainnya dilihat dari judul dan abstraknya. Hal ini dilakukan karena abstrak adalah rangkuman dari seluruh bab yang dikerjakan di tugas akhir dan di dalam abstrak terdapat kata kunci yang bersifat unik dan mencakup seluruh pembahasan di tugas akhir terkait.

Text Mining adalah suatu bahasan yang menganalisis dan mengelola teks menjadi sebuah informasi yang dapat diolah untuk tujuan tertentu. Pada *Text Mining* dikenal istilah *case folding*, *tokenizing*, *filtering*, dan *stemming*. Hal ini yang dilakukan pertama kali terhadap data abstrak sebelum dilakukan pembobotan *tf-idf* dan perhitungan *cosine similarity*.

Berdasarkan latar belakang tersebut, maka judul tugas akhir ini adalah **“Perhitungan Kemiripan Dokumen Bahasa Indonesia Menggunakan Metode Cosine Similarity (Studi Kasus : Abstrak Tugas Akhir Fakultas Informatika Universitas Telkom) “**

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan di atas, adapun rumusan masalahnya adalah :

1. Bagaimana menerapkan *text mining* dalam menghitung kemiripan suatu tugas akhir dengan tugas akhir lainnya?
2. Melakukan perhitungan terhadap data yang diperoleh yaitu bagian abstrak tugas akhir mahasiswa yang sudah lulus

1.3 Batasan Masalah

Adapun batasan masalah dalam tugas akhir ini adalah :

1. Dokumen yang digunakan adalah data abstrak tugas akhir
2. Ruang lingkup yang digunakan adalah S1 Informatika Telkom University
3. Data tugas akhir yang akan dikelola adalah terbitan tahun 2010 – 2015

1.4 Tujuan

Tujuan dari tugas akhir ini adalah dengan menerapkan *Text Mining* dalam perhitungan kemiripan dokumen maka dapat mengurangi waktu yang dibutuhkan untuk membandingkan dua dokumen dan nilai yang dihasilkan tidak bersifat subjektif.

1.5. Metodologi Penyelesaian Masalah

Metodologi dari tugas akhir ini adalah :

1. Identifikasi Masalah

Masalah – masalah yang terdapat di beberapa perguruan tinggi berkaitan dengan kasus plagiat karya orang lain serta masalah – masalah yang didapat dari kumpulan jurnal yang sudah ada selanjutnya dirumuskan menjadi poin permasalahan tugas akhir ini.

2. Studi literatur

Pada tahapan ini yang dilakukan adalah pemahaman terkait metode yang akan digunakan serta memahami karya ilmiah yang sudah ada yang berkaitan dengan tugas akhir ini yang selanjutnya akan dijadikan referensi tugas akhir ini. Informasi yang didapat berasal dari karya ilmiah yang sudah ada, buku, maupun diskusi bersama orang – orang yang mempunyai pemahaman lebih di bidang *text mining* sehingga

lebih memudahkan dalam pemahaman. Studi literatur harus disiapkan dan dilakukan secara maksimal karena berkaitan dengan pengerjaan tugas akhir secara keseluruhan.

3. Pengambilan Data

Pengambilan data tugas akhir mahasiswa (berupa NIM, Nama mahasiswa, Tahun Ajaran, Judul Tugas Akhir, dan bagian Abstrak) di perpustakaan Telkom University.

4. Melakukan Tahapan *Preprocessing*

Pada tahapan ini yang dilakukan adalah *Case Folding*, *Tokenizing*, *Filtering*, dan *Stemming*. Hasil dari melakukan *preprocessing* ini adalah sebuah *term* yang selanjutnya akan digunakan pada implementasi.

5. Implementasi

Setelah tahap *preprocessing* selesai dan didapatkan *term* maka dilakukan implementasi yaitu proses perhitungan bobot tiap kata menggunakan *tf-idf* dan implementasi metode *cosine similarity*.

6. Analisis Hasil

Pada tahapan ini yang dilakukan adalah melihat keluaran yang ditampilkan oleh program yang sudah dibuat dan melakukan analisis terhadap hasil tersebut.

7. Pembuatan Laporan

Setelah proses analisis hasil dilakukan maka tahapan akhir yang dilakukan adalah pembuatan laporan. Hal ini dilakukan untuk menyimpulkan hasil analisis yang sudah dilakukan sebelumnya apakah keluaran dari sistem sudah sesuai dengan tujuan awal melakukan penelitian ini. Selain itu laporan ini juga dapat digunakan sebagai referensi untuk karya ilmiah lain yang dapat dikembangkan dari kelebihan dan kekurangan yang ada pada sistem yang sudah dibuat di tugas akhir ini.