

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam menempuh perkuliahan dibutuhkan syarat untuk kelulusan. Skripsi atau pun tugas akhir merupakan salah satu mata kuliah wajib yang harus ditempuh mahasiswa dalam jenjang sarjana sebagai syarat kelulusan. Pengambilan judul skripsi harus disesuaikan dengan kompetensi mahasiswa dan konsentrasi mahasiswa sehingga sesuai dengan tujuan awal mahasiswa dalam proses pengambilan konsentrasi supaya maksimal. Dalam perkembangan teknologi sudah mempermudah mahasiswa dalam proses pencarian tema atau judul skripsi. Hal ini bisa mempunyai kelebihan dan kekurangan masing - masing. Selain itu hal ini dapat dijadikan referensi yang baik dengan banyaknya sumber referensi, sisi lainnya dapat memicu tingkat kemiripan judul yang beragam yang memicu plagiat yang tinggi sehingga menyebabkan kurang beragamnya tema atau judul penelitian.

Proses pendeteksian kemiripan judul yang dilakukan oleh Universitas Teknologi Digital Indonesia masih sebatas membandingkan judul skripsi yang diajukan oleh mahasiswa oleh dosen pembimbing masing - masing. Hal ini akan berdampak kurang lengkapnya sumber informasi judul yang sudah di ajukan dalam basis data yang tersedia di Universitas Teknologi Digital Indonesia. Setiap dosen akan mempunyai preferensi penilaian yang berbeda - beda terhadap proposal judul yang diajukan, apakah layak diangkat sebagai judul skripsi atau

tidak tanpa adanya variabel pengecekan tentang kemiripan judul yang sudah pernah diajukan.

Proses pencarian kemiripan judul dapat dijadikan alasan dalam penolakan proposal judul skripsi. Penggunaan metode yang tepat dapat membantu dalam proses presentase kemiripan tema atau topik. Salah satu metode yang digunakan untuk proses perhitungan kemiripan teks khususnya untuk judul adalah metode Cosine Similarity. Cosine Similarity adalah perhitungan kesamaan antara dua vektor n dimensi dengan mencari kosinus dari sudut diantara keduanya dan sering digunakan untuk membandingkan dokumen dalam text mining (Zhiqiang, 2009). Penggunaan cosine similarity dapat dipadukan dengan pembobotan setiap kata dengan metode *Term Frequency Inverse Document Frequency* (TF-IDF). Metode TF-IDF merupakan metode untuk menghitung bobot suatu kata (term) terhadap dokumen. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat (Ma'rif dan Aziz, 2015). Frekuensi data kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata tersebut di dalam dokumen.

Penelitian ini dilakukan untuk menguji tingkat kemiripan judul proposal skripsi yang akan diajukan dengan dibandingkan dengan judul yang sudah ada di Universitas Teknologi Digital Indonesia. Berdasarkan uraian di atas maka peneliti mengambil judul "Pembobotan Tf-Idf Dan Metode Cosine Similarity Untuk Deteksi Kemiripan Dalam Pengajuan Topik Skripsi i UTDI".

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas dapat dirumuskan suatu permasalahan yaitu: “Bagaimana metode cosine similarity dan pembobotan TF-IDF dapat melakukan deteksi tingkat kemiripan judul skripsi beserta ruang lingkup untuk dapat dijadikan alat penyaringan judul yang diajukan.”

1.3 Ruang Lingkup

Dalam pembuatan skripsi ini ditentukan suatu lingkup atau batasan masalah yang bertujuan untuk memudahkan pengerjaan dan menghindari adanya bahasan di luar sasaran yang tidak diinginkan. Agar tidak menyimpang jauh dari permasalahan yang ada, maka diperlukan adanya batasan masalah, yaitu:

1. Aplikasi yang dibangun adalah aplikasi yang mampu deteksi kemiripan teks judul skripsi dan ruang lingkup skripsi yang sudah ada di Universitas Teknologi Digital Yogyakarta.
2. Data judul skripsi dan ruang lingkup diambil dari tahun 2021 dan 2022 yang dijadikan sebagai data pembandingan untuk deteksi kemiripan dengan jumlah data master masing masing 50 judul pada program studi Informatika di Universitas Teknologi Digital Yogyakarta.
3. Pengguna adalah dosen pembimbing sebagai admin dan Mahasiswa UTDI sebagai user yang akan mengajukan topik penelitian.
4. Hak akses dosen pembimbing atau admin dapat memasukkan, merubah, menghapus, dataset judul dan ruang lingkup skripsi dan melakukan pengecekan presentase kemiripan. Sedangkan hak akses mahasiswa hanya dapat melakukan pengecekan kemiripan dan hasil kemiripan dapat di

unduh dalam format pdf untuk dijadikan lampiran saat mengajukan proposal.

5. Dosen pembimbing atau admin dapat login untuk menjalankan sistem. Mahasiswa tidak perlu login untuk menjalankan sistem.
6. Metode yang digunakan untuk deteksi kemiripan menggunakan cosine similarity dengan terlebih dahulu dilakukan dengan pembobotan TF-IDF pada hasil kata yang sudah di proses.
7. Hasil dari kemiripan berupa nilai presentase dengan nilai kemiripan dari tertinggi sampai terendah.
8. Tahapan stemming menggunakan library tambahan yaitu library sastrawi.
9. Aplikasi yang dibangun berbasis web dengan bahasa program PHP dan basis data MySQL.
10. Tahapan pengujian metode dilakukan dengan perbandingan hasil kemiripan antara yang dilakukan dengan proses stemming dan tanpa stemming.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Membuat suatu sistem yang dapat mendeteksi tingkat kemiripan dari judul skripsi yang diajukan dibandingkan dengan basis data judul skripsi yang ada di Universitas Teknologi Digital Indonesia.
2. Menguji tingkat kemiripan judul skripsi dengan membandingkan text processing dengan tahapan stemming dibandingkan tanpa tahapan stemming.

1.5 Manfaat Penelitian

Manfaat yang dapat diambil dari penelitian ini adalah :

1. Membantu pengguna dalam proses deteksi tingkat kemiripan judul skripsi yang akan diajukan oleh mahasiswa UTDI.
2. Membantu pengguna dalam proses pencarian informasi perbandingan kemiripan judul skripsi dengan tahapan proses tanpa stemming dan dengan stemming.

1.6 Sistematika Penulisan

Tugas akhir ini nantinya disusun dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Berisi tentang Latar Belakang, Rumusan Masalah, Ruang lingkup, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penulisan.

BAB II LANDASAAAN TEORI

Landasan Teori berisikan tinjauan pustaka dengan menelaah penelitian terdahulu yang sejenis. Serta didukung teori-teori yang digunakan sebagai landasan dalam penyelesaian Tugas Akhir ini. Teori yang ada diantaranya berisikan metode yang digunakan, alat perancangan dan alat untuk pembuatan sistem

BAB III METODE PENELITIAN

Berisi mengenai bagaimana metode penelitian yang akan dilakukan. Dalam pembahasan ini akan dijabarkan metode pengumpulan data yang dilakukan

disertakan dengan perancangan yang dilakukan mulai dari perancangan basis data dan perancangan tampilan.

BAB IV HASIL DAN PEMBAHASAN

Berisi tentang proses pengujian hasil perancangan, kemudian menganalisis hasil uji coba yang telah didapat dan pembahasannya sehingga didapatkan hasil yang sesuai yang diharapkan.

BAB V PENUTUP

Berisi tentang kesimpulan dari pembahasan bab-bab sebelumnya dan saran-saran, serta kemungkinan pengembangan tugas akhir ini.