

BAB I PENDAHULUAN

1.1 Latar Belakang

Pada dunia pendidikan era digital yang berkembang sangat pesat ini, terdapat banyak platform-platform digital yang menyediakan referensi pembelajaran bagi mahasiswa. Seperti website *google scholar* yang menyediakan berbagai jurnal, buku, tesis, artikel, tugas akhir, dan skripsi. Dengan adanya platform digital tersebut dapat menjadi acuan oleh mahasiswa dalam memilih judul tugas akhir dan skripsi yang relevan dan menghindari *plagiarisme* [1]. Namun masih terdapat banyak mahasiswa memiliki judul yang hampir sama dengan mahasiswa lainnya. Oleh karena itu diperlukan adanya sebuah sistem yang dapat memberikan informasi mengenai tingkat kemiripan judul dan dapat menjadi referensi bagi mahasiswa dalam menentukan judul yang unik, relevan, dan tidak serupa dengan mahasiswa lainnya. Sistem tersebut perlu menerapkan sebuah algoritma seperti *cosine similarity* agar dapat menganalisis kemiripan judul melalui teks.

Algoritma *cosine similarity* merupakan sebuah metode untuk menghitung kemiripan antara dua buah objek yang dinyatakan dalam sebuah vektor dengan menggunakan kata kunci dari sebuah dokumen sebagai ukuran [2]. Sebelum dilakukan penghitungan menggunakan algoritma *cosine similarity*, terdapat pembobotan kata menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). Metode TF-IDF merupakan suatu mekanisme pencarian yang menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen. TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen yang akan menghasilkan pencarian dokumen sesuai tingkat pembobotan berdasarkan *query* yang dimasukkan oleh pengguna [3]. Algoritma dan metode tersebut dapat digunakan untuk mendeteksi kemiripan judul karena penghitungan dari dua buah vektor dan pembobotan kata yang akan mendapatkan tingkat presentase kemiripan yang akurat.

Judul tugas akhir dan skripsi yang digunakan sebagai dataset dalam sistem deteksi kemiripan ini diambil dari situs <https://eprints.amikom.ac.id/> yang merupakan *repository* Universitas Amikom Yogyakarta. Dalam *repository* tersebut terdapat banyak data judul tugas akhir, skripsi, tesis, artikel, buku, dan karya ilmiah. Data yang akan digunakan dalam pengembangan sistem ini hanya data judul tugas akhir dan skripsi. Pengambilan data tersebut menggunakan teknik *web scraping*. *Web scraping* merupakan teknik mendapatkan informasi dari website secara otomatis tanpa harus menyalinnya secara manual [4]. *Web scraping* dalam sistem ini berfokus dalam mendapatkan data dan informasi yang berupa judul, nama penulis, tautan, tahun publikasi, program studi, dan perguruan tinggi dengan cara pengambilan dan ekstraksi.

1.2 Perumusan masalah

Berdasarkan latar belakang masalah pada point 1.1, penulis merumuskan masalah sebagai berikut:

1. Bagaimana proses algoritma *cosine similarity* dalam sistem deteksi kemiripan judul?
2. Berapa tingkat akurasi yang didapat dengan menggunakan algoritma *cosine similarity*?

1.3 Tujuan Penelitian

Berdasarkan perumusan masalah pada point 1.2, tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui cara kerja algoritma *cosine similarity* dalam sistem deteksi kemiripan judul tugas akhir dan skripsi.
2. Untuk membantu mahasiswa dalam menentukan judul tugas akhir dan skripsi yang relevan dan tidak plagiat dengan tingkat akurasi yang tinggi.

1.4 Batasan Masalah

Berdasarkan perumusan masalah pada point 1.2, batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Sistem hanya dapat digunakan oleh mahasiswa yang ingin melakukan

pengecekan judul.

2. Sistem deteksi kemiripan hanya merujuk berdasarkan data *repository* Universitas Amikom Yogyakarta.
3. Sistem hanya dapat mendeteksi kemiripan judul tugas akhir dan skripsi.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mempermudah mahasiswa dalam menentukan judul tugas akhir dan skripsi yang relevan dan tidak serupa dengan mahasiswa lainnya.
2. Dapat menjadi referensi kepada mahasiswa dalam menentukan judul tugas akhir dan skripsi.
3. Dapat bermanfaat bagi dosen pembimbing dalam mempertimbangkan judul tugas akhir dan skripsi yang akan diajukan oleh mahasiswa agar tidak serupa dengan judul terdahulu.

