

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam jenjang pendidikan perguruan tinggi di seluruh Indonesia tuntutan akademiknya sangatlah berbeda-beda. Khusus pada perguruan tinggi, mahasiswa dituntut untuk mengerjakan tugas akhir/skripsi sebagai syarat kelulusan (Akbar and Anggraeni 2017). Skripsi sendiri adalah suatu karya ilmiah yang ditulis oleh mahasiswa untuk bukti mahasiswa dalam melakukan penelitian lalu dipertahankan dalam suatu ujian sidang (Akbar and Anggraeni 2017). Skripsi merupakan salah satu syarat yang harus dibuat oleh mahasiswa di seluruh perguruan tinggi Indonesia untuk mendapatkan gelar Sarjana Strata 1 (S1) (Nuraeni, Sudiarjo, and Rizal 2021). Begitupun juga Universitas Nahdlatul Ulama Sunan Giri Bojonegoro yang mengharuskan mahasiswa nya membuat skripsi sebagai persyaratan menempuh S1.

Dalam pemilihan topik skripsi atau judul skripsi yang tepat sangatlah berpengaruh terhadap mahasiswa dalam ketepatan waktu dalam menyelesaikan skripsi, selain itu juga akan mempengaruhi pada akreditasi program studi (Wibowo, Saifudin, and Darmawan 2022). Pada saat peneliti melakukan observasi di lapangan yaitu wawancara kepada salah satu mahasiswa UNUGIRI yaitu Milla mengungkapkan bahwa "Karena banyaknya topik skripsi sehingga banyak mahasiswa yang kesulitan memilih topik yang akan menjadi fokus kajian penelitian berdasarkan dari judul yang diajukan". Maka dari itu, sistem yang memiliki kemampuan dalam mendeteksi persoalan topik berdasarkan judul penelitian yang diajukan sangat diperlukan.

Dalam pengolahan data tersebut menggunakan proses *text mining* yang mana menurut (Wibowo, Saifudin, and Darmawan 2022) *text processing* adalah langkah yang dilakukan dalam membangun model *meachine learning* dengan *text mining*. Dalam *text processing* ada berbagai tahap yang perlu dilakukan, tahapan yang paling sering digunakan yaitu, *lower*

conversion, tokenzation, stemming dan stop-word removal. Text processing menurut (Atmaja and Mandala 2020) merupakan *model meachine learning* dengan *text mining, text mining* sendiri yaitu teori pengolahan pada koleksi dokumen yang berjumlah besar dari beberapa waktu yang menggunakan analisis dengan tujuan pengolahan teks agar mengetahui informasi dari sumber data sehingga dapat diidentifikasi dan ekplorsi pola yang menarik pada koleksi dokumen yag tidak terstruktur.

Metode-metode yang sering digunakan para peneliti dalam klasifikasi pelabelan topik skripsi yaitu *Support Vector Machine (VSM)* (Liani 2020), *K-Nearest Neighbor (KNN)* (Mahasiswa and Komputer, n.d.) dan juga *Naive Bayes* (Sihotang and Ghaniy 2019). KNN memiliki kelebihan merupakan algoritma *machine learning* yang paling sederhana walaupun data yang digunakan banyak (Wibowo, Saifudin, and Darmawan 2022), karena berdasarkan aturan yang memilih jarak minimum dari data uji ke sampel pelatihan untuk menentukan K tetangga terdekat, akan tetapi, memiliki kelemahan dalam mengukur kemiripan suatu teks atau dokumen.

VSM memiliki kelebihan mengukur kemiripan antara suatu dokumen dengan suatu query user dengan menggunakan cosinus dari sudut antar vektor yang dibentuk oleh dokumen dengan vektor dari kata kunci yang diinputkan oleh *user* (Sihotang and Ghaniy 2019). Salah satu kelemahan dari VSM adalah metode ini menganggap bahwa setiap term pada dokumen bersifat independen, yaitu metode ini tidak melihat hubungan makna dengan term lain (Sihotang and Ghaniy 2019). Sebagai contoh, apabila *user* melakukan pencarian dengan kata kunci “*programming*” maka hasil pencariannya adalah semua dokumen yang hanya memiliki kata “*programming*” saja, padahal masih banyak dokumen-dokumen yang masih berhubungan makna dengan kata “*programming*” seperti “*PHP*”, “*Java*”, dan lain-lain. Metode untuk mengetahui kategori pada sebuah judul tugas akhir secara otomatis tanpa harus dikategorikan dibaca satu persatu maka perlu dilakukan pengukuran kemiripan dokumen terkait dengan menggunakan metode *Naive Bayes* yaitu proses pengenalan teks dan dokumen.

Naive Bayes adalah klasifikasi statistik yang bisa memprediksi probabilitas sebuah kelas, dan kelebihan dari metode ini adalah tingkat akurasi yang tinggi, juga waktu komputasi yang lebih cepat (Sihotang and Ghaniy 2019). Hanya saja bagaimana perancangan sistem yang tepat untuk membuat sistem label topik skripsi. Maka dari itu metode pendekatan *Design Thinking* sangat diperlukan untuk perancangan sistem yang diperlukan karena Manfaat dari pendekatan *Design Thinking* sendiri yaitu menyelesaikan permasalahan dengan memfokuskan kebutuhan pengguna, dimana pendekatan ini dilakukan dengan memberi ide-ide yang berupa solusi dari permasalahan dan kebutuhan pengguna lalu diimplementasikan ke dalam aplikasi yang akan dibangun.

1.2 Rumusan Masalah

Karena pentingnya sistem pelabelan topik skripsi untuk mahasiswa khususnya mahasiswa UNUGIRI Bojonegoro, diperlukan algoritma dan juga metode yang cocok, maka rumusan masalah yang peneliti buat adalah sebagai berikut: Bagaimana cara mengimplementasikan algoritma Naïve Bayes dalam sistem pelabelan topik skripsi dengan menggunakan pendekatan *Design Thinking*?

1.3 Tujuan

Tujuan yang ingin dicapai oleh penulis dari penelitian Implementasi algoritma *Naïve Bayes* dalam sistem pelabelan topik skripsi dengan menggunakan pendekatan *Design Thinking* adalah untuk mengetahui topik yang akan diambil oleh mahasiswa jika mengajukan judul tersebut.

1.4 Manfaat

Sistem pelabelan topik skripsi dengan pendekatan *Design Thinking* menggunakan algoritma *Naive Bayes* ini diharapkan dapat membantu mahasiswa-mahasiswa dalam menentukan topik yang akan di gunakan dalam judul skripsi yang diajukan dengan tepat. Sehingga ketika pemilihan topik skripsi tepat akan mempermudah dan mempercepat dalam pengerjaan skripsi para mahasiswa.



UNUGIRI