

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tugas Akhir (TA) atau skripsi adalah karya ilmiah yang disusun berdasar atas kaidah keilmuan dan ditulis menurut kaidah bahasa Indonesia dengan bimbingan sekaligus pengarahan dosen pembimbing.[1]. Tugas akhir merupakan salah satu persyaratan yang harus dipenuhi oleh mahasiswa untuk mendapatkan gelar sarjana, untuk melakukan penelitian dan menghasilkan dokumen tugas akhir, mahasiswa memanfaatkan pengetahuan yang telah mereka pelajari[2]. Tugas akhir mahasiswa di jurusan teknik informatika dikelompokkan dalam empat bidang minat yaitu komputasi bahasa alami, kecerdasan komputasi, sistem terdistribusi dan jaringan, teknologi perangkat lunak.

Semakin banyaknya lulusan maka banyak juga dokumentasi tugas akhir yang dihasilkan. Dokumen yang dihasilkan akan sulit untuk menemukan banyak materi karya ilmiah atau tugas akhir jika tidak disusun berdasarkan topik. Tantangan menentukan cara mengklasifikasikan karya ilmiah atau tugas akhir merupakan tantangan bagi banyak mahasiswa. Karena selama ini dalam menentukan topik karya ilmiah atau tugas akhir hanya berdasarkan perkiraan isi yang diteliti oleh mahasiswa, sehingga kesesuaian antara topik dan isi terkadang sangat berbeda.

Dengan adanya permasalahan yang ada diperlukan sebuah sistem yang dapat secara otomatis mengkategorikan bidang minat yang ada, sistem yang dapat mengkategorikan bidang minat tugas akhir secara otomatis. Diharapkan sistem yang akan dikembangkan dapat membantu pihak jurusan dalam mengklasifikasikan tugas akhir yang diajukan mahasiswa secara lebih efektif dan efisien. Tugas akhir dalam penelitian ini akan dikategorikan ke dalam beberapa bidang minat jurusan teknik informatika.

Machine learning (ML) merupakan sebuah subbidang dari kecerdasan buatan untuk mempelajari data untuk memfokuskan pengembangan sistem. *Machine learning* dapat didefinisikan sebagai aplikasi komputer

dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang [3]. kecerdasan buatan (Artificial Intelligent) berkembang sangat pesat, masalah-masalah yang sebelumnya sangat sulit dipecahkan manusia, dengan adanya kecerdasan buatan masalah tersebut dapat diselesaikan dengan mudah, kecerdasan buatan (Artificial Intelligent) sangat menarik dan sangat penting untuk dipelajari dan diterapkan, karena kelebihanannya yang mampu belajar seperti jaringan saraf otak manusia [4]. Sebelum masuk pada metode *Mechine Learning* diperlukan data untuk melatih metode-metode yang ada pada *Machine Learning*. Salah satu cara untuk mendapatkan data dengan melakukan mining atau crawling data pada suatu platform. *Data Mining* didefinisikan sebagai menafsirkan atau menyelidiki data dari sejumlah besar data yang tidak teridentifikasi. Namun, hal tersebut dapat dimengerti dan memiliki manfaat dari *database* yang masif serta dimanfaatkan sebagai pembuat keputusan bisnis yang penting. *Data Mining* adalah metode penggalian informasi yang berguna dari database yang dapat digunakan dalam menganalisis data yang dapat membantu dalam mengambil beberapa keputusan berdasarkan data [5]. Setelah itu, data akan diolah dan diproses sesuai dengan subjek yang diperlukan, termasuk regresi, klasifikasi, *Clustering*, dan lain-lain.

Klasifikasi adalah salah satu aspek dari *data mining*. Prosedur klasifikasi melibatkan melihat dan menilai dokumen teks baru yang belum diberi kelas. Proses mengidentifikasi fungsi atau model yang mencirikan atau membedakan ide atau informasi. kelas sebagai kategorisasi. Tujuan dari klasifikasi adalah untuk mendapatkan kemampuan dalam memperkirakan kelas sebuah objek yang namanya tidak diketahui [6]. Penggunaan *Naïve Bayes Classifier* pada klasifikasi dokumen dapat dilihat dari prosesnya yang mengambil tindakan berdasarkan data-data yang telah ada[7]. Metode ini memungkinkan klasifikasi dokumen yang dipersonalisasi, yang mempertimbangkan kebutuhan dan preferensi setiap individu.

Metode *naïve bayes* sudah banyak digunakan pada penelitian tentang text mining, salah satu kelebihan *naïve bayes* yaitu algoritma sederhana tetapi mempunyai nilai akurasi yang cukup tinggi[8]. Algoritma *Naïve Bayes Classifier (NBC)* adalah algoritma klasifikasi sederhana, efisien, dan kinerja yang baik dalam set data, Algoritma *Naïve Bayes Classifier (NBC)* adalah algoritma yang paling populer serta cocok digunakan pada dataset besar dan dapat mengatasi data yang kosong (missing value), serta mampu mengelola atribut yang beragam dan gangguan dalam data [9].

Pada penelitian ini untuk mengidentifikasi jenis bidang minat pada tugas akhir menggunakan metode *Naive Bayes Classifier*. Berdasarkan permasalahan yang telah dijelaskan, adapun tujuan dilakukan penelitian ini adalah untuk menentukan kategori bidang minat yaitu komputasi bahasa alami, kecerdasan komputasi, sistem terdistribusi dan jaringan, teknologi perangkat lunak. Hasil akhir berupa, hasil klasifikasi pada bidang minat menggunakan metode *Naive Bayes Classifier*.

1.2 Perumusan Masalah

Perumusan masalah pada penelitian ini sebagai berikut:

1.2.1 Permasalahan

Meningkatnya dokumen pada portal tugas akhir membuat mahasiswa ragu ragu untuk memilih bidang minat yang ada pada pemilihan topik yang di inginkan, dengan adanya klasifikasi pada bidang minat seperti komputasi bahasa alami, kecerdasan komputasi, sistem terdistribusi dan jaringan, teknologi perangkat lunak menggunakan metode klasifikasi teks berbasis *Naive Bayes Classifier* adalah salah satu strategi untuk kategorisasi teks. Metode *Naive Bayes Classifier* dapat meningkatkan klasifikasi teks pada halaman web tugas akhir teknik informatika dengan melihat pola data yang kompleks.

1.2.2 Metode Usulan

Penelitian ini mengusulkan untuk melakukan klasifikasi bidang minat pada portal tugas akhir menggunakan metode *Naive Bayes Classifier* yang

memiliki waktu komputasi yang sangat cepat dalam tahap pelatihan dan klasifikasi.

1.2.3 Pertanyaan Penelitian

1. Berapa nilai akurasi menggunakan *Naive Bayes Classifier* ?
2. Berapakah nilai akurasi paling tinggi menggunakan *K-Fold*

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Setelah pertanyaan penelitian, tujuan dari penelitian ini mengetahui nilai akurasi dari metode *Naive Bayes Classifier*. Mengelompokkan abstrak berdasarkan bidang minat yang menciptakan klasifikasi sesuai bidang minat yang di inginkan.

1.3.2 Manfaat Penelitian

Menghasilkan nilai akurasi dari metode *Naive Bayes Classifier*. Dengan menggunakan metode *Naive Bayes Classifier* dalam melakukan klasifikasi teks pada artikel dapat mengetahui performa metode *Naive Bayes Classifier* dalam klasifikasi. Penelitian diharapkan dapat memberikan kontribusi yang bagi pengembangan pengetahuan lebih lanjut dalam bidang ini.

1.4 Batasan-Batasan

1. Dokumen yang di uji berjumlah 816 jurnal jurusan Teknik Informatika pada portal tugas akhir dengan proses *crawling* dari web <https://pta.trunojoyo.ac.id/byprod>
2. Data teks yang digunakan berupa data yang menggunakan Bahasa Indonesia.
3. Pembagian kategori atau label dibagi menjadi 4 bagian yaitu kategori komputasi bahasa alami, kecerdasan komputasi, sistem terdistribusi dan jaringan, teknologi perangkat lunak.
4. Portal tugas akhir yang di klasifikasikan di fokuskan pada kategori dalam memilih bidang minat yang di inginkan.

1.5 Jadwal Penelitian

Berikut merupakan perencanaan jadwal penelitian skripsi dijelaskan pada Tabel 1.1.

Tabel 1.1 Jadwal penelitian skripsi

Kegiatan	Februari	Maret	April	Mei	Juni
Studi Literatur & Penyusunan Bab 1-3					
Seminar Proposal					
Revisi Seminar Proposal					
Pengujian Skenario					
Evaluasi Model					
Penyusun Bab 4 & 5					
Sidang Skripsi					
Revisi Sidang Skripsi					

1.6 Sistematika Penulisan Proposal

Pada penelitian ini rencana susunan sistematika penulisan proposal sebagai berikut:

BAB I PENDAHULUAN

Bab ini memuat latar belakang penelitian, rumusan masalah penelitian, klasifikasi teks, batasan masalah penelitian, tujuan penelitian, keunggulan penelitian, dan sistem penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini dilakukan kajian terhadap literatur dan teori terkait penelitian yang dilakukan yang diperoleh dari crawling pada web portal

tugas akhir. Beberapa teori yang digunakan dalam penelitian ini antara lain abstrak portal tugas akhir, preprocessing, *Naïve Bayes Classifier*, dalam klasifikasi portal tugas akhir berdasarkan bidang minat.

BAB III METODE USULAN

Dalam bab ini dirujuk langkah-langkah penelitian agar penelitian yang dilakukan mencapai tujuan yang maksimal. Tahapan penelitian meliputi tahap penelitian, perancangan sistem, pengumpulan data, tahap preprocessing, tahap klasifikasi, tahap evaluasi dan tahap skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan secara rinci mengenai penerapan tahapan yang disusun pada bab sebelumnya dan hasil klasifikasi text pada portal tugas akhir berdasarkan bidang minat dengan menggunakan metode *Naïve Bayes Classifier*.

BAB V PENUTUP

Pada bab ini terdapat kesimpulan dari hasil penelitian serta sebagai acuan saran untuk penelitian selanjutnya dengan merujuk pada hasil penelitian in