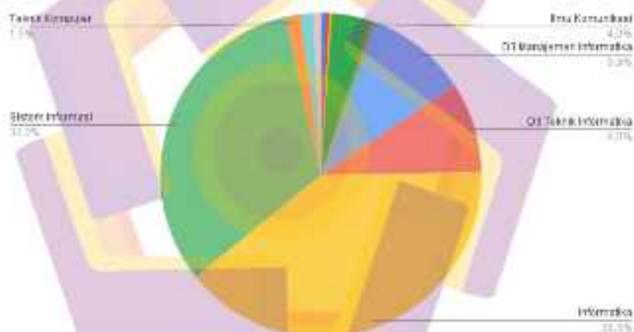


## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Data penelitian tugas akhir skripsi di perguruan tinggi selalu mengalami pertumbuhan dari waktu ke waktu, dikarenakan tugas akhir merupakan salah satu syarat yang harus dipenuhi mahasiswa agar dapat lulus dari perguruan tinggi tersebut. Setiap perguruan tinggi sudah menentukan batasan topik apa saja yang sebaiknya diambil oleh setiap mahasiswa, hal ini bertujuan supaya topik penelitian yang diambil mahasiswa sesuai dengan program studi atau konsentrasi yang diambil.



**Gambar 1. 1** Statistik Jumlah tugas akhir dan skripsi

Berdasarkan hasil olah statistik dari data web *digital library* Universitas Amikom Yogyakarta, total data penelitian skripsi dan tugas akhir tercatat pada bulan Agustus tahun 2022 sebanyak 5.533 penelitian. Topik skripsi yang diambil mahasiswa umumnya sesuai dengan mata kuliah bidang konsentrasi yang diajarkan, akan tetapi mata kuliah bidang konsentrasi selalu mengikuti perkembangan teknologi agar kualitas lulusannya dapat mengikuti persaingan industri. Hal ini tentunya berpengaruh terhadap topik skripsi yang diambil oleh mahasiswa dari waktu ke waktu. Semakin bervariasinya topik yang diambil mahasiswa membuat pengelompokan skripsi berdasarkan topik relatif sulit.

Dari banyaknya penelitian skripsi yang semakin beragam, membuat pengelompokan topik harus menyesuaikan dengan perkembangan tersebut. Untuk itu diperlukan metode *text mining* untuk mengubah data-data yang tidak terstruktur menjadi data-data yang terstruktur[1]. *Text mining* merupakan teknik penggalian data berupa teks, data yang digali umumnya berasal dari suatu dokumen. Tujuan dari pengaplikasian *text mining* adalah memperoleh pola-pola data, trend, dan ekstraksi dari pengetahuan-pengetahuan yang potensial dari data teks [2].

*Text classification* merupakan sub bidang dari *text mining* yang menentukan kategori dari sebuah dokumen berdasarkan kontennya [1]. Metode yang sering digunakan untuk melakukan *text classification* antara lain adalah metode *Naïve Bayes*, *Support Vector Machine (SVM)*, *Random Forest*, *Decision Tree*, dan yang lainnya. *Naïve Bayes* sangat luas dipakai dalam berbagai bidang, khususnya untuk *text classification*. Metode *Naïve Bayes* merupakan *Supervised Learning* yang membutuhkan data pelatihan yang telah diberikan topik terlebih dahulu[3].

Berdasarkan pada observasi penulis pada data-data penelitian skripsi Universitas Amikom Yogyakarta, skripsi memungkinkan untuk memuat lebih dari satu topik penelitian. Hal ini menunjukkan bahwa klasifikasi data skripsi berbeda dengan klasifikasi data teks pada umumnya. Klasifikasi yang sesuai untuk kasus ini adalah klasifikasi multi-label. Klasifikasi *multi-label* merupakan metode klasifikasi dimana satu dokumen dapat diberikan lebih dari satu kelas target[4].

Klasifikasi *multi-label* perlu dilakukan perubahan dari bentuk dasar formula *Naïve Bayes* agar bisa mengklasifikasikan suatu dokumen ke banyak kelas. *Multinomial Naïve Bayes* merupakan turunan dari *Naïve Bayes* yang denominatornya dianggap konstan. Berdasarkan laporan dari hasil penelitian sebelumnya, metode *Naïve Bayes* memberikan hasil yang bagus ketika diimplementasikan baik ke dalam dokumen berbahasa Inggris maupun dokumen berbahasa Indonesia[5].

Oleh karena itu, berdasarkan latar belakang di atas, maka perlu dilakukan penelitian tentang klasifikasi terhadap data abstrak skripsi Universitas Amikom Yogyakarta Program Studi Informatika dengan metode *Multinomial Naïve Bayes*. Tujuan dari penelitian ini adalah untuk mengimplementasikan dan melakukan pengujian kinerja algoritma *Multinomial Naïve Bayes* terhadap studi kasus penelitian.

### 1.2 Rumusan Masalah

Dari pemaparan latar belakang di atas, didapatkan rumusan masalah sebagai berikut :

1. Bagaimana mengimplementasikan *Multinomial Naïve Bayes Classifier* untuk mengklasifikasikan skripsi berdasarkan abstrak ?
2. Bagaimana kinerja metode *Multinomial Naïve Bayes Classifier* pada penelitian ini ?

### 1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Data penelitian diambil dari abstrak skripsi Universitas Amikom Yogyakarta Program Studi Informatika.
2. Topik-topik penelitian yang dijadikan sebagai target kelas disesuaikan dengan konsentrasi sesuai dengan tahun kurikulum pada data.
3. Data dibagi menjadi 80% data training dan 20% data testing, dataset yang diolah sebanyak 1366 data skripsi.
4. Metode yang digunakan untuk melakukan klasifikasi adalah *Multinomial Naïve Bayes* dengan TF-IDF sebagai metode pembobotan kata.
5. Bahasa pemrograman yang digunakan adalah Python sebagai pengolah sekaligus pemroses data dan *Framework* Vue Javascript sebagai *frontend*

untuk menyajikan informasi, teknologi *frontend* yang digunakan tidak menjadi pembahasan utama dalam penelitian ini, hanya untuk mempermudah dalam implementasi.

#### 1.4 Tujuan Penelitian

Berdasarkan pemaparan masalah diatas tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Mengimplementasikan *Multinomial Naïve Bayes Classifier* pada klasifikasi skripsi berdasarkan abstrak.
2. Mengukur kinerja yang diperoleh dari sistem *Multinomial Naïve Bayes Classifier* yang dibuat pada penelitian ini.

#### 1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

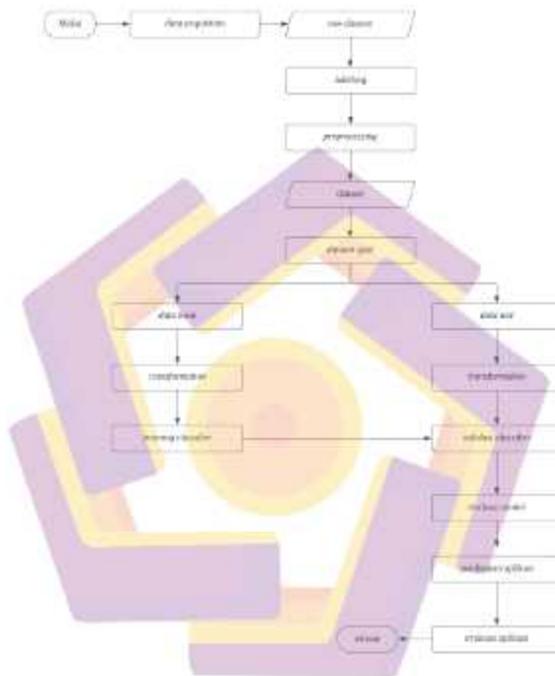
1. Memberikan pengetahuan/wawasan terhadap akademisi mengenai tingkat keberagaman topik-topik skripsi dari waktu ke waktu.
2. Memberikan pengetahuan kepada pembaca tentang langkah-langkah melakukan klasifikasi dan evaluasi pada data teks menggunakan *Multinomial Naïve Bayes*.
3. Mengetahui kinerja sistem *Multinomial Naïve Bayes Classifier* yang dibuat pada penelitian ini.
4. Sebagai rujukan pada penelitian lainnya dalam kasus klasifikasi pada data teks.

#### 1.6 Metodologi Penelitian

Penelitian ini diperlukan beberapa metode agar mencapai hasil yang diharapkan, namun perlu dilakukan studi pustaka terlebih dahulu. Penelitian Klasifikasi Multi-label menggunakan algoritma *Multinomial Naive Bayes* sudah ada sebelumnya, sehingga perlu mengumpulkan data penelitian sebelumnya dari

buku-buku, jurnal-jurnal baik nasional maupun internasional, dan juga dari internet yang berkaitan dengan permasalahan yang diangkat dalam penelitian ini.

Gambar 1.2 merupakan alur penelitian yang menjelaskan langkah-langkah mulai dari pengumpulan data hingga evaluasi.



**Gambar 1. 2 Alur metodologi penellitian**

Penjelasan Gambar 1.2 sebagai berikut:

### 1. Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari website arsip digital Resource Center Universitas Amikom Yogyakarta (<https://eprints.amikom.ac.id>). Data yang digunakan bersifat data terbuka yang dapat diakses oleh siapapun, meskipun sumber data bersifat terbuka penelitian ini sudah mendapat izin Resource Center untuk dilakukan pengolahan. Atribut utama yang digunakan untuk

pengolahan yaitu abstrak skripsi, sedangkan atribut pendukungnya yaitu judul penelitian, tahun, dan penulis.

## 2. Metode Pelabelan Data

Klasifikasi merupakan *supervised learning*, dimana perlu pengawasan manusia dalam proses pengelompokan datanya. klasifikasi membutuhkan data training yang sudah diberikan label atau kelas target terlebih dahulu. Pada tahapan ini dilakukan labeling secara semi otomatis dan dilakukan analisis terhadap data yang sudah diberi label supaya jumlah dokumen di setiap labelnya berimbang. proses pelabelan tentunya melibatkan persetujuan dosen-dosen informatika yang kompeten dibidangnya.

## 3. Metode Preprocessing

Metode preprocessing merupakan serangkaian proses yang digunakan untuk menjadikan data siap untuk diklasifikasikan, dalam tahap ini terdiri dari dua proses utama yaitu pembersihan kata dan pembobotan kata. Bentuk dari data abstrak skripsi adalah teks yang memuat kata serapan, angka, dan kata yang menggunakan aksen karakter tertentu. Data abstrak skripsi memiliki struktur yang lebih baku daripada data teks lainya oleh karena itu metode pembersihan kata yang digunakan adalah *cleansing*, *filtering*, *stemming*, dan *tokenizing*. Tahapan selanjutnya adalah merubah data teks menjadi angka menggunakan pembobotan kata, metode pembobotan kata yang digunakan TF-IDF.

## 4. Metode Klasifikasi

Pada tahapan ini dilakukan pengolahan *dataset* yang sudah dilakukan *preprocessing*. Dokumen abstrak skripsi dikelompokkan kedalam kelas-kelas tertentu sesuai target. Hasil dari tahapan ini berupa model klasifikasi yang nantinya digunakan untuk pengklasifikasian data test setelah dilakukan pengujian model. Metode yang digunakan pada tahapan ini adalah *Multinomial Naïve Bayes*.

## 5. Metode Perancangan dan Implementasi Interface

Tujuan dari aplikasi yang akan dibangun adalah untuk mempermudah penyajian dan pengaksesan informasi. Aplikasi yang akan dibuat berbasis website yang terdiri aplikasi *backend* dan *frontend*. Metode perancangan perangkat lunak yang digunakan adalah *Waterfall*. Metode ini dilakukan dengan pendekatan yang sistematis mulai dari tahap analisis kebutuhan sistem, desain, perancangan, dan pengujian[6].

## 6. Metode Evaluasi

Pengujian dilakukan terhadap model klasifikasi dan aplikasi yaitu sebagai berikut :

### a. Evaluasi Model Klasifikasi

Metode yang digunakan untuk melakukan evaluasi terhadap hasil klasifikasi multi-label adalah *Hamming Loss* kemudian digambarkan dalam bentuk grafik. Semakin kecil nilai *Hamming Loss* maka semakin baik kinerja dari model klasifikasi yang dibuat[7].

### b. Evaluasi Aplikasi Interface

Pengujian aplikasi *interface* fokus terhadap validasi kualitas aplikasi yang dibangun berdasarkan model klasifikasi. Pengujian aplikasi yang akan dilakukan pada penelitian ini adalah *black box testing/functional testing*.

## 1.7 Sistematika Penulisan

Guna memudahkan pembaca dalam mengikuti apa yang dipaparkan dalam laporan skripsi ini, sistematika penulisan tersusun dari 5 (lima) bab yaitu:

### **BAB I            PENDAHULUAN**

Bab ini berisikan gambaran umum penelitian yang dilakukan, meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian metode penelitian, dan sistematika penelitian.

**BAB II      TINJAUAN PUSTAKA**

Bab ini membahas tentang dasar-dasar teori yang mendukung dalam melakukan proses pengklasifikasian abstrak skripsi menggunakan algoritma *Multinomial Naïve Bayes* mulai dari tahap data *preprocessing*, pembobotan kata, klasifikasi dan evaluasi

**BAB III     METODOLOGI PENELITIAN**

Bab ini membahas tentang metodologi yang digunakan dalam penelitian yang mencakup metode klasifikasi teks menggunakan algoritma *Multinomial Naïve Bayes*, serta metode-metode penunjang lainnya seperti pengumpulan dataset, *Preprocessing*, *Feature Selection*, *Feature extraction*, *Classification Stage*, dan Evaluasi.

**BAB IV     HASIL DAN PEMBAHASAN**

Bab ini berisi penjelasan dari hasil dan pembahasan klasifikasi. Hasil klasifikasi yang didapatkan menggunakan algoritma *Multinomial Naïve Bayes* berdasarkan 4 tahap proses klasifikasi yaitu *Preprocessing*, *Feature Selection*, *Feature extraction*, *Classification Stage*. Selanjutnya pembahasan dilakukan dengan melihat pada teori sebelumnya dengan memperhatikan dan menimbang berdasarkan eksperimen yang dilakukan.

**BAB V      PENUTUP**

Bab ini merupakan bab terakhir yang berisi kesimpulan dari apa yang telah diuraikan pada bab sebelumnya serta saran untuk perkembangan penelitian yang akan data