

# BAB I

## PENDAHULUAN

### 1.1 LATAR BELAKANG MASALAH

*Data mining* yaitu kegiatan meliputi pengumpulan, pemakaian data *historis* dalam menemukan keteraturan, pola, atau hubungan dalam *dataset* berukuran besar. *Output* dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan[1]. *Data mining* merupakan proses yang mempekerjakan satu atau lebih teknik untuk menganalisis pengetahuan secara otomatis[2]. Teknik *data mining* dapat mengklasifikasikan data baru dengan memanipulasi data yang sudah ada yang telah diklasifikasi dan menggunakan hasilnya untuk memberikan sejumlah aturan[3]. *Data mining* adalah proses mengekstrak informasi penting dari sebuah dataset dengan tujuan untuk mengekstrak informasi berupa pola, tren, atau hubungan yang tidak terlihat secara langsung pada data.

Salah satu teknik *data mining* yang umum digunakan adalah klasifikasi. Klasifikasi adalah teknik melihat perilaku dan atribut kelompok yang telah ditentukan[3]. Teknik klasifikasi adalah teknik pembelajaran untuk prediksi suatu nilai dari target variabel kategori[4]. Teknik klasifikasi melibatkan pemeriksaan variabel yang dihasilkan dari data yang ada[5]. klasifikasi adalah proses menemukan variabel yang menggambarkan dan membedakan kelas data dengan tujuan untuk memprediksi kelas dari objek yang label kelasnya telah diketahui ataupun belum diketahui.

Pada klasifikasi terdapat beberapa algoritma yang sering digunakan diantaranya adalah Support Vector Machine, Decision Tree, dan Naive Bayes[6]. Naïve Bayes adalah metode yang sederhana dan efisien, namun dapat melampaui metode klasifikasi yang lebih kompleks[7]. *Naïve Bayes* merupakan prediksi berbasis probabilitas sederhana yang mengacu pada teorema Bayes yang memiliki asumsi independensi yang kuat[8]. *Naïve Bayes* menggunakan salah satu cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan-kemungkinan klasifikasi[9]. *Naïve Bayes* adalah teknik klasifikasi sederhana berbasis probabilitas untuk mendapatkan peluang terbesar pada klasifikasi.

Penerimaan siswa baru merupakan proses pendaftaran kepada siswa baru untuk masuk sekolah dan gerbang awal yang dilalui peserta didik didalam objek Pendidikan[10]. Dengan menggunakan algoritma Naïve Bayes, institut Pendidikan dapat mengklasifikasi data pendaftar seperti nilai, prestasi di luar akademik, dll untuk menerima siswa terbaik dan berpotensi.

Pendidikan tinggi di Indonesia telah mengalami peningkatan dari waktu ke waktu, saat ini sudah mencapai 3098 perguruan tinggi baik negeri maupun swasta[11]. Peningkatan juga terjadi pada jumlah pendaftar di perguruan tinggi[12]. Peningkatan ini harus diimbangi dengan strategi seleksi penerimaan mahasiswa yang baik agar mendapatkan calon mahasiswa yang berkualitas.[13]. Mahasiswa yang berkualitas bukan hanya yang memiliki kemampuan intelektual tinggi, tetapi juga harus dapat menyelesaikan pendidikannya dengan baik. Salah satu cara yang dapat digunakan untuk menyeleksi mahasiswa baru adalah dengan

menggunakan klasifikasi[14]. Meningkatnya jumlah pendaftar mengharuskan institut pendidikan melakukan seleksi untuk menerima mahasiswa terbaik bagi universitas.

Dalam melakukan klasifikasi penerimaan mahasiswa dibutuhkan sebuah *dataset*. Pada penelitian kali ini sumber *dataset* yang akan digunakan penulis adalah *dataset* yang didapat dari *Kaggle Repository* dengan nama *Data Mahasiswa*[15]. *dataset* ini memiliki 3587 data dengan 25 fitur dan 1 label. *Dataset* ini memiliki label pada setiap data, yang membuat *dataset* ini akan digunakan untuk *supervised learning*. Pada *supervised learning* sistem diberikan *training dataset* berupa informasi masukan dan keluaran yang diinginkan, sehingga sistem akan mempelajari berdasarkan data yang telah ada. Sistem akan mencari pola dari *dataset*, kemudian pola itu akan dijadikan sebagai acuan untuk kumpulan data berikutnya[16]. *Dataset* merupakan komponen penting pada sebuah penelitian. Sebuah *dataset* yang baik akan menghasilkan penelitian yang baik pula.

Berdasarkan penelitian sebelumnya oleh Reji Pikriyansah dkk, dalam klasifikasi daftar ulang calon mahasiswa baru menggunakan metode klasifikasi *Naïve Bayes*, algoritman *Naïve Bayes* berhasil mendapatkan tingkat akurasi sebesar 63.83% [17]. Dan pada penelitian oleh Aries Saifudin pada metode data mining untuk seleksi calon mahasiswa pada penerimaan mahasiswa baru di universitas Pamulang, algoritma *Naïve Bayes* berhasil mendapatkan tingkat akurasi tertinggi dengan nilai 60.85% [18]. Berdasarkan penelitian sebelumnya, penelitian ini dilakukan untuk mengklasifikasi penerimaan mahasiswa baru.

Dari permasalahan yang telah diuraikan dan berdasarkan penelitian terdahulu, penulis tertarik untuk membuat penelitian dengan judul “Penerapan Algoritma *Naïve Bayes* Dalam Klasifikasi Penerimaan Mahasiswa Baru”.

## **1.2 RUMUSAN MASALAH**

Berdasarkan dari latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana menerapkan algoritma *Naïve Bayes* dalam mengklasifikasi penerimaan mahasiswa baru?
2. Seberapa akurat algoritma *Naïve Bayes* dalam mengklasifikasi data penerimaan mahasiswa baru?

## **1.3 BATASAN MASALAH**

Untuk tidak memperluas pembahasan, perlu adanya Batasan-batasan untuk menyederhanakan permasalahan, yaitu:

1. *Dataset* yang akan digunakan pada penelitian diambil dari *KAGGLE repository*, dengan nama Dataset Mahasiswa Baru. *Dataset* ini memiliki 3587 data dengan 25 fitur dan 1 label.
2. Dalam melakukan klasifikasi Data Mahasiswa Baru menggunakan algoritma *Naïve Bayes*.
3. Menggunakan *tools WEKA* dalam menjalankan proses klasifikasi.
4. Untuk Mengukur performa akurasi algoritma berdasarkan nilai akurasi, presisi, *recall*, dan *confusion matrix*.

## **1.4 TUJUAN DAN MANFAAT PENELITIAN**

Berdasarkan latar belakang masalah yang telah dijelaskan, berikut adalah tujuan dan manfaat yang dapat diperoleh dari penelitian ini.

### **1.4.1 Tujuan Penelitian**

Tujuan penulis dalam melakukan penelitian ini adalah:

1. Menerapkan algoritma *Naïve Bayes* dalam mengolah *dataset* Data Mahasiswa Baru dalam proses klasifikasi penerimaan mahasiswa baru.
2. Mengukur tingkat akurasi, presisi, dan *recall* algoritma *Naïve Bayes* dalam mengklasifikasi *dataset* Data Mahasiswa Baru.

### **1.4.2 Manfaat Penelitian**

Manfaat penelitian ini adalah:

1. Untuk membantu institut pendidikan dalam mengklasifikasi mahasiswa yang telah mendaftar.
2. Memberikan kontribusi dalam pengembangan metode klasifikasi penerimaan mahasiswa baru.
3. Penelitian ini dapat digunakan untuk mengetahui performa algoritma *Naïve Bayes* dalam mengklasifikasi data penerimaan mahasiswa baru.
4. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi data mining khususnya dalam klasifikasi penerimaan mahasiswa baru.

## 1.5 SISTEMATIKA PENELITIAN

Sistematika dibuat untuk memberi suatu gambaran yang jelas terhadap penulisan ilmiah agar menghindari terjadinya pembahasan diluar dari tema dan judul penelitian. Secara garis besar penulisan ilmiah terdiri dari 5 bab, dapat dilihat dari sistematika penulisan sebagai berikut:

### – **BAB I: PENDAHULUAN**

Bab ini merupakan bab pendahuluan yang membahas tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

### – **BAB II: LANDASAN TEORI**

Bab ini memuat tentang landasan teori yang mendasari pembahasan laporan secara khusus dan berisikan definisi - definisi yang melandasi penelitian yang dilakukan dengan melakukan studi literature sebagai dasar dalam melakukan analisa dan perancangan.

### – **BAB III: METODOLOGI PENELITIAN**

Bab ini membahas secara rinci kerangka kerja penelitian, metode atau pendekatan yang digunakan, dan tools (alat bantu) yang digunakan dalam menjawab masalah penelitian untuk mencapai suatu tujuan.

### – **BAB IV: ANALISIS DAN PENGUJIAN**

Bab ini menyajikan pembahasan hasil yang diperoleh dan model deteksi menggunakan klasifikasi yang dibangun dalam penelitian

ini. Serta evaluasi dan pengujian sistem yang telah dihasilkan untuk melihat kesesuaian dengan tujuan yang diharapkan.

– **BAB V: KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dan saran-saran dari penelitian untuk membantu penelitian – penelitian selanjutnya

