

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Naïve Bayes Classifier merupakan salah satu metode yang populer digunakan untuk menyelesaikan masalah klasifikasi dan mudah diimplementasikan. Metode tersebut berpotensi baik untuk mengklasifikasikan data karena kesederhanaannya. Algoritma *Naïve Bayes* menghitung perbandingan peluang antara jumlah dari masing-masing kriteria nilai variabel terhadap nilai hasil prediksi sesungguhnya [1]. *Data mining* adalah sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan [2]. *Data mining* sering juga disebut dengan *Knowledge Discovery in Database* atau disingkat KDD, adalah kegiatan yang meliputi pengumpulan, pemakaian data histori untuk menelusuri data yang ada untuk membangun sebuah model agar dapat mengenali pola data yang lain berukuran besar.

Universitas Dinamika Bangsa (UNAMA) memiliki database mahasiswa selama menempuh proses kegiatan perkuliahan. Data alumni semakin bertambah setiap tahunnya dan database tersebut bisa dimanfaatkan. Sumber data dapat mencakup database, gudang data (*data warehouse*) [3]. Berdasarkan banyaknya data alumni dan data jumlah kelulusan mahasiswa, perlu dilakukan analisa untuk mengetahui informasi penting berupa

pengetahuan baru (*Knowledge Discovery*). Dalam kondisi normal, mahasiswa S1 dari UNAMA dapat menyelesaikan studinya selama 4 tahun atau 8 semester. Kelulusan mahasiswa dapat dikatakan salah satu item penilaian penting dalam proses akreditasi institusi suatu perguruan tinggi. BAN-PT memiliki point nilai tersendiri untuk lulusan mahasiswa bagi setiap perguruan tinggi, sehingga jika dalam suatu perguruan tinggi memiliki lulusan tepat waktu setiap semesternya maka dapat membantu suatu perguruan tinggi tersebut dalam proses penilaian akreditasi, pada penerepannya tingkat kelulusan mahasiswa tidak bisa kita pastikan mereka lulus dengan tepat waktu [4]. Penelitian ini dilakukan pada mahasiswa di Universitas Dinamika Bangsa (UNAMA) Jambi sebagai objek penelitian dan sampel, dengan mengambil data dari para alumni yang akan dibandingkan antara prediksi waktu kelulusan dengan kondisi kelulusan sebenarnya.

Dengan menggunakan metode *Naive Bayes* untuk meneliti prediksi lama masa studi mahasiswa dengan atribut yang digunakan diantaranya yaitu jenis kelamin, sekolah asal, jurusan sekolah asal, IPK, kelas lulus (tepat, terlambat). Untuk mengklasifikasikan data kelulusan. Dalam proses prediksi lamanya masa studi mahasiswa yang penulis buat, BAAK terlebih dahulu memberikan data-data yang diperlukan lalu memasukkannya ke dalam file dikomputer. Setelah itu, maka akan dilakukan proses data cleaning dari microsoft excel yang ada pada file di laptop, kemudian melakukan pengukuran tingkat akurasi metode *Naive Bayes*. Selesai melakukan pengukuran tingkat akurasi metode *Naive Bayes*, lalu hasilnya akan sebagai sebuah tabel. Dengan proses klasifikasi prediksi lamanya masa studi

mahasiswa yang penulis buat, diharapkan dapat memudahkan mahasiswa dalam meningkatkan kualitas di bidang pendidikan dari program studi maupun universitas. Algoritma Naive Bayes (NB) merupakan metode yang sederhana dalam klasifikasi berdasarkan teori probabilitas yang dikemukakan pertama kali oleh ilmuwan Inggris bernama Thomas Bayes. Disebut naif karena menyederhanakan masalah yang bergantung pada dua asumsi penting [5].

Terdapat penelitian sebelumnya yang memakai *Naive Bayes* untuk prediksi berbagai kondisi, diantaranya dalam penelitian Tias Mugi Rahayu [6] dibuktikan Hasil perhitungan *naïve bayes* data yang terklasifikasi dengan benar menunjukkan tingkat akurasi sebesar 69,33%, pada penelitian Ananda Fiqri Firdaus [7] dibuktikan Dari rasio data training dan data testing yang telah diujikan yaitu 60:40, 70:30, dan 80:20. Didapatkan hasil bahwa dengan menggunakan rasio 80:20 dapat menghasilkan tingkat akurasi, presisi dan recall yang paling optimal secara keseluruhan, Serta dalam penelitian Hozairi [8] Hasil penelitian ini menunjukkan bahwa setelah menggunakan model *K-Nearest Neighbor*, *Decision Tree* serta *Naive Bayes* untuk mengklasifikasi status kelulusan mahasiswa Teknik Informatika Universitas Islam Madura diperoleh hasil bahwa kinerja *Naive Bayes* lebih unggul dari *K-Nearest Neighbor* serta *Decision Tree*. Terbukti bahwa dari 35 data uji yang digunakan *Naive Bayes* memiliki nilai akurasi 89%, presisi 88% sedangkan *K-Nearest Neighbor* memiliki nilai akurasi 77% , presisi 76% dan *Decision Tree* memiliki nilai akurasi 74% dan presisi 84%.

Hal inilah yang melatar belakangi penulis untuk melakukan penelitian guna

memberi solusi terhadap masalah yang terjadi dengan mengangkat judul “**Algoritma Naive Bayes Untuk Mencari Lama Waktu Studi Mahasiswa (Studi Kasus : Universitas Dinamika Bangsa)**” yang nantinya dapat mengklasifikasikan lama masa studi / kategori kelulusan mahasiswa yang telah menempuh kuliah dengan menggunakan dasar data kelulusan sebelumnya.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah bagaimana mengklasifikasikan dan menganalisis mencari lama masa studi mahasiswa Universitas Dinamika Bangsa (UNAMA) menggunakan metode *naïve bayes*?

1.3 BATASAN MASALAH

Agar penelitian ini dapat berjalan dengan baik, terarah, dan tidak keluar dari topik pembahasan maka penulis menetapkan batasan-batasan masalah yaitu :

1. Penelitian ini hanya mengklasifikasikan dan menganalisis lama masa studi mahasiswa program studi Sistem Informasi Universitas Dinamika Bangsa (UNAMA) tahun masuk 2017 dan 2018.
2. Analisis menggunakan teknik *data mining* klasifikasi dengan metode *naïve bayes*.
3. Atribut yang digunakan yaitu jenis kelamin, sekolah asal, jurusan sekolah asal, IPK, kelas lulus (tepat, lambat).
4. Informasi yang dihasilkan berupa data pengklasifikasian pencarian lama masa studi mahasiswa dengan dua kategori yaitu tepat dan lambat.

5. Pengujian hasil analisis menggunakan *tools* WEKA.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan

Tujuan dari penelitian ini adalah :

Mengklasifikasikan dan menganalisis data-data mahasiswa Sistem Informasi dengan metode *naïve bayes* sehingga dapat mencari lama masa studinya. dan dapatkan akurasi yang baik untuk klasifikasi prediksi lama masa studi mahasiswa dengan menggunakan metode *naïve bayes*.

1.4.2 Manfaat

Manfaat dari penelitian ini yaitu :

1. Dapat menggunakan hasil analisis untuk menentukan strategi dalam meningkatkan kualitas dan kuantitas kelulusan mahasiswa pada tahun-tahun berikutnya.
2. Dapat menggunakan hasil analisis untuk menentukan strategi dalam meningkatkan kualitas dan kuantitas kelulusan mahasiswa pada tahun-tahun berikutnya.
3. Penulis dapat menambah ilmu dan wawasan baru mengenai pengklasifikasian dan analisis prediksi lama masa studi mahasiswa.
4. Dapat digunakan sebagai acuan untuk melakukan penelitian berikutnya.

1.5 SISTEMATIKA PENULISAN

Adapun sistematika penulisan tugas akhir ini, penulis menguraikan dalam beberapa bab yaitu:

BAB I : PENDAHULUAN

Pada bab ini akan diuraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab landasan teori ini membahas tentang teori-teori dan pendapat para ahli yang berhubungan dengan permasalahan yang dianalisis. Teori-teori yang digunakan antara lain mengenai data mining, klasifikasi, prediksi, lama masa studi, *naïve bayes*, dan WEKA.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang kerangka kerja penelitian, metode pengumpulan data, metode klasifikasi, serta alat bantu yang digunakan pada penelitian ini.

BAB IV : ANALISIS

Pada bab ini dilakukan perhitungan analisis menggunakan metode *naïve bayes* terhadap data-data mahasiswa yang tersedia, hasil dari analisis dan bentuk visualisasi analisis dari tools WEKA yang digunakan.

BAB V : PENUTUP

Bab ini yang berisikan kesimpulan-kesimpulan yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil penelitian.