

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis yang menemukan keteraturan, pola, dan hubungan dalam dataset berukuran besar (Rahmawali, Hansun, & Wiratama Widya, 2017). *Data mining* sendiri memiliki beberapa teknik salah satunya klasifikasi. Klasifikasi merupakan salah satu proses pada data mining yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar. *Decision tree* dan *naïve bayes* adalah bagian dari metode klasifikasi. Kemudian metode *decision tree* memiliki beberapa algoritma, algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokkan dan bersifat prediktif. Algoritma *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (*naif*) akan independensi dari masing-masing kondisi / kejadian.

Program sarjana (S1) STIKOM Dinamika Bangsa memiliki beban studi sekurang-kurangnya 144 sks yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester dan paling lama 14 semester.

Peneliti menemukan masih banyak mahasiswa yang menempuh lama studi tidak sesuai target yang dijadwalkan. Setiap tahun jumlah data mahasiswa pada STIKOM DB Jambi selalu bertambah dan tidak ada tindak lanjut manfaat dari data-data tersebut. Jika data yang melimpah ini hanya dibiarkan menumpuk, maka hanya akan menjadi beban database yang dimiliki. Padahal data tersebut bisa diolah menjadi sebuah informasi baru dengan data mining.

Prediksi masa studi mahasiswa dapat dimanfaatkan oleh pembimbing akademik untuk mengetahui status kelulusan mahasiswa. Hal ini dapat menjadi rekomendasi atau sebagai peringatan agar mahasiswa bisa lulus dengan tepat waktu sehingga dapat menekan angka keterlambatan kelulusan sekaligus meningkatkan kualitas jurusan itu sendiri. Penggunaan *data mining* dengan Algoritma C4.5 dan *Naïve Bayes* dalam memprediksi masa studi mahasiswa dapat menjadi alternatif yang tepat. Namun sampai saat ini belum diketahui algoritma yang paling akurat dalam memprediksi masa studi mahasiswa ini. Untuk itu maka dalam penelitian ini akan dilakukan komparasi metode Algoritma C4.5 dan *Naïve Bayes* untuk mengetahui algoritma yang memiliki akurasi lebih tinggi dalam memprediksi masa studi mahasiswa.

Hal inilah yang melatar belakangi penulis untuk melakukan penelitian guna memberi solusi terhadap masalah yang terjadi dengan mengangkat judul **“KLASIFIKASI PREDIKSI LAMA MASA STUDI MAHASISWA SISTEM INFORMASI MENGGUNAKAN PERBANDINGAN ALGORITMA C4.5 DAN NAÏVE BAYES (Studi Kasus : STIKOM DB JAMBI)”** yang nantinya dapat mengklasifikasikan lama masa studi / kategori kelulusan mahasiswa yang

telah menempuh kuliah dengan menggunakan dasar data kelulusan sebelumnya dan menentukan algoritma manakah yang paling akurat untuk memprediksi lama masa studi mahasiswa.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah :

1. Bagaimana mengklasifikasikan dan menganalisis prediksi masa studi mahasiswa Sistem Informasi STIKOM DB menggunakan Algoritma C4.5 dan *Naïve Bayes* ?

1.3 BATASAN MASALAH

Untuk mempermudah pembahasan dalam penelitian ini agar lebih terarah dan berjalan baik, dan untuk menghindari terjadinya pembahasan yang tidak berkenaan dari judul laporan. Maka batasan masalah yang akan dibahas adalah :

1. Penelitian ini hanya menggunakan data mahasiswa program studi sistem informasi tahun 2013 dan 2014.
2. Atribut yang digunakan yaitu jenis kelamin, asal sekolah, jurusan sekolah asal, IPK, status kuliah, nilai D, dan kelas lulus (tepat, lambat).
3. Data *training* menggunakan 70% data mahasiswa, sedangkan data *testing* menggunakan 30% data mahasiswa.

4. Informasi yang dihasilkan berupa data pengklasifikasian dengan 2 (dua) kategori yaitu tepat dan lambat, serta perbandingan dan akurasi yang efisien dan riil.
5. Pengujian hasil menggunakan *tools* Weka.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Mengklasifikasikan dan menganalisis data-data mahasiswa sistem informasi dengan algoritma C4.5 dan *Naïve Bayes* sehingga dapat memprediksi masa studinya.
2. Mendapatkan akurasi yang baik untuk klasifikasi prediksi lama masa studi mahasiswa dengan menggunakan algoritma C4.5 dan *Naïve Bayes*.
3. Mengetahui algoritma yang paling akurat untuk memprediksi lama masa studi mahasiswa sistem informasi.

1.4.2 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dalam penelitian ini adalah sebagai berikut :

1. Melalui penelitian ini dapat mengetahui prediksi lama masa studi masing-masing mahasiswa sistem informasi.

2. Penulis dapat menambah ilmu dan wawasan baru mengenai pengklasifikasian dan analisis prediksi lama masa studi mahasiswa.
3. Penelitian ini dapat digunakan sebagai referensi bagi pengembangan penelitian sejenis dimasa yang akan datang.

1.5 SISTEMATIKA PENULISAN

Sistematika penulisan menggambarkan secara umum tentang apa yang akan di bahas dalam setiap bab dari penulisan tugas akhir ini. Penulisan tugas akhir ini terdiri dari enam bab.

Adapun isi pokok bahasan masing-masing bab sebagai berikut :

BAB I : PENDAHULUAN

Bab ini merupakan bab pertama yang menjelaskan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika.

BAB II : LANDASAN TEORI

Bab ini merupakan bab kedua yang menjelaskan teori-teori yang mendasari pembahasan laporan secara khusus berisi definisi-definisi yang di dapat melalui studi pustaka sebagai dasar dalam penelitian.

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 algoritma C4.5 dan *naïve bayes* terhadap data-data mahasiswa yang tersedia.

BAB V : HASIL ANALISIS DAN VISUALISASI

Pada bab ini akan ditampilkan hasil dari analisis dan bentuk visualisasi analisis dari tools WEKA yang digunakan.

BAB VI : PENUTUP

Pada bab ini merupakan penutup dari penelitian ilmiah ini yang berisi kesimpulan dari pembahasan bab-bab sebelumnya dan juga saran-saran yang berguna bagi pihak-pihak yang berkaitan dengan penelitian ilmiah ini.