

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Seiring dengan berkembangnya teknologi penyimpanan data dan informasi, semakin berkembang pula kemampuan seseorang dalam mengumpulkan dan mengolah data. Data yang terkumpul dan berukuran besar tersebut merupakan aset yang dapat dimanfaatkan untuk dianalisis yang hasilnya berupa pengetahuan atau informasi berharga untuk masa mendatang. Tidak hanya dunia bisnis, namun instansi seperti perguruan tinggi lainnya juga mengalami penumpukan data.

Jurusan Teknik Informatika Universitas Muhammadiyah Jember adalah program pendidikan perguruan tinggi yang memiliki beban studi sekurang - kurangnya 144 SKS (satuan kredit semester) yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester. Dari data kelulusan yang diperoleh dari sekretariat Jurusan Teknik Informatika Program Sarjana (S1) angkatan 2010, 2011, 2012, dan 2013 hanya sedikit mahasiswa yang lulus dalam 8 semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) reguler di Jurusan Teknik Informatika yang menempuh lama studi lebih dari 8 semester.

Melihat kondisi tersebut diperlukan penelitian untuk menggali data yang dimiliki oleh Jurusan Teknik Informatika Universitas Muhammadiyah Jember. Data yang akan dimanfaatkan dalam penelitian ini adalah data mata kuliah umum, yaitu Pengantar Teknologi Informasi, Programan Terstruktur, Struktur Data, Sistem Informasi, Analisis Desain Berorientasi Objek, Programan Berorientasi Objek, Jaringan Komputer, Praktikum Jaringan Komputer, Programan Jaringan, Praktikum Programan Jaringan Komputer, dan nilai IPK semester 1-6, dan data nilai akademik mahasiswa baik yang sudah lulus, belum lulus atau sedang menempuh yang digunakan sebagai data training dan data testing untuk memprediksi masa studi mahasiswa. Penelitian ini dirasa perlu karena jika masa studi mahasiswa dapat diketahui lebih dini, maka pihak jurusan dapat melakukan tindakan-tindakan yang

dirasa perlu supaya mahasiswa dapat lulus tepat waktu sekaligus meningkatkan kualitas jurusan itu sendiri.

Dari data masalah dan pengalaman historis yang sudah terjadi dibutuhkan suatu teknik dan metode klasifikasi yang merupakan salah satu teknik dari klasifikasi untuk menganalisis data di Jurusan Teknik Informatika. Untuk menganalisa kasus tersebut. Algoritma C4.5 dan Naive Bayes adalah salah satu metode klasifikasi yang dapat di uji untuk memproses dan memprediksi kemungkinan mahasiswa yang lulus lebih dari 8 semester dan lulus tepat waktu.

*Algoritman C4.5* merupakan algoritma pengembangan dari algoritma ID3 yang digunakan untuk membentuk pohon keputusan. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan (Sofiana, 2012).

*Naive Bayes* merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidak tergantungan yang kuat. Selain itu naive bayes juga dapat menganalisa variabel-variabel yang paling mempengaruhinya dalam bentuk peluang (Prasetyo, 2012).

Menurut Wu, et al. (2007), terdapat sejumlah algoritma fungsi data mining yang dipilih menjadi Top 10 algorithm data mining. Hasilnya, diperoleh bahwa algoritma klasifikasi menjadi mayoritas dalam Top 10 algorithm. Adapun algoritma klasifikasi yang terbaik dalam Top 10 algorithm data mining adalah salah satunya pohon keputusan C4.5, K-Nearest Neighbor (K-NN), Naive Baiyes, dan Classification and Regression Trees (CART).

Galathiya, et al. (2012), melakukan penelitian dengan melakukan pengembangan pada algoritma pohon keputusan C4.5 dan melakukan perbandingan terhadap algoritma *Naive Bayes*, *Sequential Minimal Optimization (SMO)*, dan multilayer perceptron dengan menggunakan sejumlah data sets dan pengujian akurasi berdasarkan kepada Correctly Classified Instances dan Relative Absolute Error. Hasilnya, algoritma pohon keputusan C4.5 yang dikembangkan memiliki tingkat akurasi sekitar 1-3% lebih baik pada 7 data sets dari total 9 data sets.

*Othman & Yau (2007)* melakukan penelitian berjudul “Comparison of Different Classification Techniques using WEKA for Breast Cancer”. Metode pengujian yang digunakan berupa percentage split sebesar 75%, artinya 75% data digunakan sebagai data training dan 25% sisanya sebagai data testing. Pengujian akurasi berdasarkan kepada Correctly Classified Instances, Incorrectly Classified Instances, dan Time Taken. Hasilnya, algoritma Naive Bayes memiliki tingkat akurasi tertinggi dan waktu pembangunan model tercepat dibandingkan dengan Radial Basis Function, Decision Tree and Pruning, Single Conjunctive Rule Learner, dan Nearest Neighbors Algorithm.

Berdasarkan masalah tersebut di atas, penulis mencoba melakukan penelitian untuk menganalisa dan membandingkan akurasi dua metode algoritma yang merupakan bagian dari Top 10 algorithm data mining, yaitu algoritma C4.5 dan Naive Bayes pada sejumlah data sets. Adapun sumber data sets adalah data mahasiswa jurusan teknik informatika (Universitas Muhammadiyah Jember) dimana setiap data sets memiliki jumlah data (*instances*) dan jumlah atribut (*attributes*) yang berbeda. Perbandingan pengukuran akurasi yang di uji akan menghasilkan informasi yang efisien, akurat dan rill, Atas dasar inilah, penulis tertarik untuk mengambil judul tugas akhir : **“ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN ALGORITMA NAIVE BAYES UNTUK PREDIKSI KELULUSAN MAHASISWA ( Studi Kasus : Prodi Teknik Informatika, Universitas Muhammadiyah Jember)”**.

## 1.2. Perumusan Masalah

Permasalahan yang dijadikan objek penelitian pada tugas akhir ini berdasarkan latar belakang yang telah dipaparkan adalah :

1. Bagaimana Memproses Algoritma C4.5 dan Algoritma Naive Bayes berdasarkan *Implementasi* algoritma.
2. Bagaimana Membandingkan *Performance* Algoritma C4.5 dan Algoritma Naive Bayes berdasarkan *Implementasi* Algoritma ?
3. Berapa *Persentase Akurasi* yang didapat dalam *Implementasi* Algoritma C4.5 dan Algoritma Naive Bayes ?

## 1.3. Batasan Masalah

Untuk lebih memfokuskan pengerjaan penelitian ditetapkan pembahasan hanya dibatasi pada :

1. Parameter yang digunakan 11 parameter, yaitu mata kuliah seperti Pengenalan Teknologi Informasi, Pemrograman Terstruktur, Struktur Data, Sistem Informasi, Analisis & Desain Berorientasi Objek, Pemrograman Berorientasi Objek, Jaringan Komputer, Praktikum Jaringan Komputer, Programan Jaringan, Praktikum Programan Jaringan Komputer, dan IPK semester 1 – 6.
2. Data yang di gunakan sebagai pengujian yaitu data mahasiswa angkatan 2010, 2011, 2012, dan 2013 jurusan teknik Informatika Universitas Muhammadiyah Jember, training data angkatan 2010, 2011 2012, dan 100 data angkatan 2013 sebagai testing atau data uji.
3. Metode yang di gunakan Algoritma C4.5 dan Algoritma Naive Bayes.
4. Hasil akhir pengujian adalah perbandingan dan akurasi berupa informasi kelulusan mahasiswa angkatan 2013.

## 1.4. Tujuan

Tujuan penelitian pada tugas akhir ini adalah

1. Menganalisa *Performance* Algoritma c4.5 dan Algoritma Naive Bayes
2. Menganalisa *Implementasi* Algoritma c4.5 dan Algoritma Naive Bayes
3. Melakukan *Perbandingan* Akurasi Algoritma c4.5 dan Algoritma Naive Bayes

## 1.5. Manfaat

Manfaat penelitian dibedakan menjadi dua jenis yaitu manfaat secara teoritis dan manfaat secara praktis.

### 1.5.1 Manfaat Teoritis

Penelitian ini dapat digunakan sebagai referensi bagi pengembangan penelitian sejenis dimasa yang akan datang.

### 1.5.2 Manfaat secara Praktis

#### a. Manfaat bagi akademik

Bagi akademik, manfaat yang dapat diperoleh dari proses pembuatan proposal skripsi ini adalah berupa sumbangsih laporan buku Tugas Akhir, yang di harapkan mampu digunakan dengan sebaik-baiknya, dan sebagai tambahan buku di perpustakaan, dan juga bisa lebih di kembangkan untuk masa yang akan dantang.

#### b. Bagi Mahasiswa

Bagi mahasiswa, Tugas Akhir ini dapat dijadikan sebagai sumber referensi maupun pedoman bagi mahasiswa lain yang mengambil penelitian sejenis.

#### c. Bagi Universitas

Mampu memprediksi masa studi mahasiswa pada program sarjana (S1) Jurusan Teknik Informatikan (atau lainnya).