

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi telah menyebabkan banyak orang dapat memperoleh data dengan mudah bahkan cenderung berlebihan. Data tersebut semakin lama semakin banyak dan terakumulasi, akibatnya pemanfaatan data yang terakumulasi tersebut menjadi tidak optimal. Salah satu cara mengatasi masalah ini adalah dengan menggunakan *Data Mining* (DM) dengan teknik klasifikasi. Kegiatan pengklasifikasian yang dilakukan oleh manusia masih memiliki keterbatasan, terutama pada kemampuan manusia dalam menampung jumlah data yang ingin diklasifikasikan (Yusuf, 2014).

Mahasiswa merupakan salah satu aspek penting dalam evaluasi keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi. Pemantauan mahasiswa yang masuk, peningkatan kemampuan mahasiswa, prestasi yang dicapai mahasiswa, rasio kelulusan terhadap jumlah total mahasiswa, dan kompetensi lulusan seyogyanya mendapatkan perhatian yang serius untuk memperoleh kepercayaan *stakeholder* dalam menilai dan menetapkan penggunaan lulusannya (Ridwan, Suyono, & M.Sarosa, 2013).

Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (*knowledge Discovery*), misalnya informasi mengenai pengklasifikasian data mahasiswa berdasarkan profil dan data akademik. Pengetahuan baru tersebut dapat membantu pihak universitas untuk melakukan klasifikasi mengenai tingkat kelulusan mahasiswa guna menentukan strategi untuk meningkatkan kelulusan pada tahun – tahun berikutnya (Yuda, 2014).

Machine Learning merupakan suatu teknologi baru yang dikembangkan untuk persoalan pembelajaran data dalam jumlah sedikit. *Machine Learning* atau pembelajaran mesin merupakan pendekatan dalam *Artificial Intelligence* yang banyak digunakan untuk menggantikan atau

menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Informasi ini dapat dijadikan alternatif solusi dalam proses pengambilan keputusan.

Sementara itu, klasifikasi merupakan proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Proses klasifikasi didasarkan pada empat komponen, yaitu kelas, *predictor*, *training dataset* dan *testing dataset*.

Dalam suatu angkatan, terdapat mahasiswa yang lulus tepat waktu namun ada juga yang tidak lulus tepat waktu. Beberapa hal yang dapat mempengaruhi salah satunya nilai Indeks Prestasi Semester, banyaknya Mahasiswa yang lulus tepat waktu dapat mempengaruhi akreditasi pada Program Studi tersebut. Oleh sebab itu, dibutuhkan sebuah metode untuk peringatan dini prediksi kelulusan Mahasiswa, sehingga pihak Fakultas atau Universitas mampu mengambil keputusan sejak dini pada kelulusan tepat waktu.

Penelitian yang dilakukan (Zul, 2015) mengajukan penggunaan algoritma *K-Nearest Neighbor* (K-NN) dan *Naïve Bayes* untuk prediksi nilai akhir semester yang sedang berjalan. Penggunaan algoritma *K-Nearest Neighbor* dan *Naive Bayes* untuk proses mencari *rule* dari klasifikasi nilai akhir semester yang sedang berjalan, sehingga dapat dimanfaatkan untuk rekomendasi pertimbangan untuk menentukan nilai akhir mahasiswa. Pada penelitian yang dilakukan, menyimpulkan bahwa algoritma terbaik yang didapatkan adalah K-NN dengan k sebanyak 15. Akurasi pengujian dengan menggunakan algoritma ini mencapai 95%.

Penelitian yang dilakukan (Selvia, Wendi, & Astrid, 2014) Mengajukan penggunaan algoritma *K-Nearest Neighbor* untuk prediksi masa studi mahasiswa. Penggunaan algoritma *K-Nearest Neighbor* bekerja berdasarkan jarak terdekat dari objek baru ke objek lama dengan menentukan nilai k. Dengan menggunakan teknik data mining tersebut maka di perguruan tinggi dapat memanfaatkan data akademik mahasiswa yaitu

indeks prestasi (IP) untuk memprediksi masa studi mahasiswa berdasarkan kategori kelulusan yaitu tepat waktu (4-5 Tahun) dan tidak tepat waktu (5 tahun lebih). Dengan *data training* berjumlah 30 dan *data testing* berjumlah 30 untuk enam semester yaitu nilai k sebanyak 10 didapat tingkat akurasi sebesar 81,66%, sedangkan pada *data training* berjumlah 60 dan *data testing* berjumlah 61 diperoleh nilai k sebanyak 20 dan k sebanyak 30 maka didapatkan tingkat akurasi sebesar 85%. Berdasarkan latar belakang tersebut, penelitian Tugas Akhir ini akan membangun sebuah penelitian pengklasifikasian dengan judul “**Klasifikasi Mahasiswa Program Studi Teknik Informatika Menggunakan Algoritma K-Nearest Neighbor**” untuk membantu Program Studi dalam memantau dan mengevaluasi perkembangan studi mahasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang maka rumusan masalah yang dihadapi adalah sebagai berikut :

1. Berapa tingkat presisi untuk Klasifikasi Kelulusan Mahasiswa Teknik Informatika di Universitas Muhammadiyah Jember Menggunakan Metode K-Nearest Neighbor ?
2. Berapa tingkat akurasi untuk Klasifikasi Kelulusan Mahasiswa Teknik Informatika di Universitas Muhammadiyah Jember Menggunakan Metode K-Nearest Neighbor ?

1.3 Tujuan

Berdasarkan perumusan masalah di atas, maka tujuan penelitian yang akan dilakukan adalah :

1. Mengetahui tingkat presisi untuk Klasifikasi Kelulusan Mahasiswa Teknik Informatika di Universitas Muhammadiyah Jember Menggunakan Metode K-Nearest Neighbor.

2. Mengetahui tingkat akurasi untuk Klasifikasi Kelulusan Mahasiswa Teknik Informatika di Universitas Muhammadiyah Jember Menggunakan Metode K-Nearest Neighbor.

1.4 Manfaat

Manfaat dari penelitian ini berdasarkan latar belakang adalah :

1. Sebagai alternatif untuk mendapatkan informasi yang berguna tentang tingkat kelulusan Mahasiswa di Universitas Muhammadiyah Jember.
2. Sebagai bahan pertimbangan yang strategis dalam upaya pencapaian tujuan universitas sehingga universitas mampu memenuhi tuntutan masyarakat untuk menghasilkan output yang kompeten dan berkualitas.

1.5 Batasan Masalah

Batasan yang digunakan agar tujuan penelitian ini dapat tercapai antara lain :

1. Dataset berupa data mahasiswa Jurusan Teknik Informatika dengan tahun angkatan 2011, 2012, dan 2013
2. Kriteria data yang digunakan yaitu:
 - a. NIM
 - b. Nama
 - c. IPS semester 1-6
 - d. Tanggal masuk Akademik
 - e. Tahun lulus Akademik
3. Menggunakan Algoritma K-Nearest Neighbor dan perhitungan jarak *Euclidean*
4. Validasi data uji dan data latih menggunakan *K-fold cross validation* dengan 5 skenario percobaan.
5. Bahasa pemrograman yang digunakan adalah PHP dengan *database* MySQL.