

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pertumbuhan data akademik mahasiswa pada perguruan tinggi setiap tahun terus mengalami peningkatan. Data ini meningkat dengan cepat tanpa banyak diketahui manfaatnya kepada manajemen. Pertambahan data yang pesat ini menciptakan kondisi dimana suatu perguruan tinggi memiliki tumpukan data. Pada saat ini, tumpukan data tersebut belum dimanfaatkan secara maksimal (Mutiara Ayu, *et al.* 2015). Jika data tersebut diolah dengan tepat akan sangat berguna bagi perguruan tinggi untuk terus meningkatkan mutu pendidikan dan akreditasi (Slamet dan Taufiq, 2018).

Pertambahan data akademik mahasiswa menyebabkan timbulnya pola yang bervariasi pada atribut data mahasiswa. Hal ini mempunyai potensi untuk melakukan penggalian data (*data mining*) yang akan memberikan sebuah pengetahuan baru. Salah satu teknik dalam data mining adalah memprediksi sebuah keputusan dari pola yang diciptakan oleh tumpukan data yang ada. Karena itu teknik data mining ini akan sangat berguna bagi perguruan tinggi yang akan melakukan pencarian nilai dari ketepatan kelulusan mahasiswa. Salah satu teknik untuk bisa memprediksi kelulusan mahasiswa salah satunya adalah dengan menggunakan Algoritma *K-Nearest Neighbor* (K-NN).

K-NN merupakan metode pengklasifikasian data yang memiliki konsisten yang kuat, dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot. Metode K-NN ini lebih efektif dalam melakukan training data yang besar dan dapat menghasilkan data yang lebih akurat (Sumarlin, 2015). Pendekatan dalam Algoritma K-Nearest Neighbor adalah dengan cara menghitung kedekatan antara kasus baru dengan kasus yang lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah atribut yang ada.

Misalkan akan dicari waktu kelulusan seorang mahasiswa yang masih semester 6 dengan menggunakan data dari mahasiswa yang sudah lulus. Untuk mencari waktu kelulusan mahasiswa tersebut, dihitung kedekatan kasus mahasiswa yang akan lulus dengan semua kasus mahasiswa yang sudah lulus dengan kedekatan terbesar lah yang akan diambil sebagai solusi pada kasus mahasiswa yang akan lulus tersebut.

Pada penelitian ini data yang digunakan adalah data mahasiswa Fakultas Teknologi Informasi Universitas Andalas Program Studi Sistem Informasi. Dan beberapa atributnya seperti data indeks prestasi mahasiswa semester 1-6, jumlah sks sampai semester 6, dan keterangan lulus dari mahasiswa angkatan 2011. Menurut Sumarlin (2015), terdapat banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama diantaranya *euclidean* dan *manhattan*. Dalam pencarian nilai *K-Nearest Neighbor* rumus yang digunakan berdasarkan dengan perhitungan jarak (*euclidean*) dan data lama (*data training*) yang digunakan adalah seluruh mahasiswa angkatan 2011 yang sudah mempunyai keterangan lulus dan *drop out*, sedangkan data baru (*data testing*) menggunakan mahasiswa angkatan 2016 yang diambil secara acak.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, permasalahan penelitian ini dirumuskan sebagai berikut :

1. Bagaimana proses metode *k-nearest neighbor* digunakan untuk memprediksi kelulusan mahasiswa ?
2. Berapa tingkat akurasi hasil prediksi kelulusan mahasiswa dari metode *k-nearest neighbor* ?

## 1.3 Batasan Masalah

Untuk mencapai tujuan penelitian ini, peneliti memfokuskan batasan penelitian sebagai berikut:

1. Data yang digunakan merupakan data mahasiswa Fakultas Teknologi Informasi, Program Studi Sistem Informasi angkatan 2011.
2. Atribut yang digunakan yaitu NIM, Nama, Indeks Prestasi Semester 1-6, jumlah SKS sampai semester 6, dan keterangan lulus.
3. Metode yang digunakan adalah *K-Nearest Neighbor*.

## 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Menerapkan metode *K-Nearest Neighbor* dengan menggunakan rumus perhitungan jarak euclidean.
2. Untuk memprediksi kelulusan mahasiswa fakultas Teknologi Informasi, program studi Sistem Informasi, dengan akurasi yang cukup tinggi.

## 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Dapat memberikan informasi prediksi kelulusan mahasiswa yang lulus tepat waktu maupun yang tidak tepat waktu.
2. Dapat membantu mengevaluasi kelulusan mahasiswa pada sistem akademik yang berjalan di perguruan tinggi.

## 1.6 Sistematika Penulisan

Adapun sistematika penulisan tugas akhir ini dibagi ke dalam enam bab yang dapat dijelaskan sebagai berikut :

### **Bab I: Pendahuluan**

Berisi Latar Belakang, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penelitian.

### **Bab II : Landasan Teori**

Pada bab ini dijelaskan teori-teori mengenai teknik untuk menunjang penulisan tesis. Adapun dibahas dalam bab ini adalah teori yang berkaitan tentang data mining dan metode *K-Nearest Neighbor*.

### **Bab III : Metodologi Penelitian**

Bab ini membahas tentang analisa dan penggunaan secara matematis metode *K-Nearest Neighbor*.

### **Bab IV : Analisa dan Perancangan**

Bab ini membahas hasil implementasi metode *K-Nearest Neighbor* pada data kelulusan mahasiswa Fakultas Teknologi Informasi Universitas Andalas.

**Bab V : Implementasi dan Hasil**

Bab ini membahas hasil implementasi metode *K-Nearest Neighbor* pada data kelulusan mahasiswa Fakultas Teknologi Informasi Universitas Andalas.

**Bab VI : Kesimpulan dan Saran**

Bab ini membuat kesimpulan dan hasil penelitian keunggulan metode *K-Nearest Neighbor* dalam memprediksi kelulusan mahasiswa pada fakultas Teknologi Informasi, Universitas Andalas.