

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Universitas Teknologi Digital Indonesia adalah perguruan tinggi swasta di Daerah Istimewa Yogyakarta. Di perguruan tinggi ini terdapat salah satu program pemerintah untuk membantu siswa memperoleh beasiswa ke jenjang pendidikan yang lebih tinggi yaitu perguruan tinggi yang disebut program Kartu Indonesia Pintar Kuliah (KIP-Kuliah) menjadi KIP Kuliah Merdeka yang memberikan bantuan belajar, perluasan akses, jaminan biaya pendidikan dan bantuan biaya hidup untuk meningkatkan modalitas ekonomi dan mobilitas sosial bagi mahasiswa. Berdasarkan panduan Pedoman Pendaftaran Kartu Indonesia Pintar Kuliah– KIP Kuliah Merdeka 2022 jangka waktu pemberian bantuan KIP Kuliah Merdeka pada Program Sarjana maksimal 8 (delapan) semester dan Diploma Tiga maksimal 6(enam) semester.

Dengan adanya pembatasan waktu dalam pemberian bantuan KIP Kuliah Merdeka maka perguruan tinggi harus melakukan proses monitoring prestasi akademik para penerima beasiswa KIP Kuliah Merdeka. Adapun angkatan penerima beasiswa yang digunakan yaitu tahun 2020/2021 dan 2021/2022 pada program Diploma Tiga, alasan menggunakan tahun 2020/2021 karena pada semester ini merupakan semester penentuan apakah akan lulus tepat waktu atau tidak sedangkan pada tahun 2021/2022 di semester tersebut masih memiliki beberapa semester untuk memperbaiki indeks prestasi ketika prediksi dilaksanakan

sebagai bentuk pencegahan agar lulus tepat waktu . Adanya suatu sistem prediksi ini diharapkan dapat membantu perguruan tinggi dalam hal tersebut.

Beberapa penelitian telah dilakukan untuk memprediksi ketepatan waktu kelulusan menggunakan algoritma C5.0, diantaranya penelitian (Nusari, Purbasari, & Puspaningrum, 2020) menggunakan algoritma C5.0 untuk membuat sistem prediksi kelulusan mahasiswa dengan 5 atribut memperoleh nilai *recall* sebesar 72,22%, nilai *precision* sebesar 92,85%, nilai *accuracy* sebesar 83,78%, dan nilai *error* sebesar 16,21%. Hal ini menunjukkan bahwa C5.0 memiliki tingkat akurasi yang tinggi.

Dengan demikian penerapan data mining dapat digunakan untuk melakukan prediksi terhadap kelulusan mahasiswa penerima KIP-Kuliah Merdeka. Prediksi ini hampir sama dengan algoritma klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang (Kusrini & Luthfi, 2009). Data Mining adalah kegiatan yang meliputi pengumpulan, pemakaian dan historis untuk menemukan keteraturan pola dan hubungan dalam set data berukuran besar (Santosa, 2007).

Salah satu algoritma yang dapat digunakan untuk memprediksi dan menerapkan metode pohon keputusan adalah algoritma C5.0. Di mana algoritma ini merupakan algoritma penyempurnaan dari algoritma C4.5 yang mana pada algoritma C5.0 tingkat akurasi yang dimiliki lebih tinggi, lebih cepat dalam pengambilan keputusan, dan dapat menangani missing value serta data dalam jumlah besar. Algoritma ini juga dapat melakukan *training* data dalam waktu yang cepat untuk digunakan dalam *testing* data (Benediktus & Oetama, 2020).

Berdasarkan uraian yang ada dan penelitian yang telah dilakukan sebelumnya, maka akan diusulkan penelitian mengenai Implementasi Algoritma C5.0 untuk Prediksi Ketepatan Kelulusan Mahasiswa Penerima KIP-Kuliah Merdeka (Studi Kasus Universitas Teknologi Digital Indonesia). Kemudian menggunakan *confusion matrix* untuk mengukur hasil akurasi dari model prediksi yang digunakan.

### **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dijelaskan di atas maka diperoleh rumusan masalah yaitu:

1. Bagaimana menerapkan algoritma C5.0 untuk prediksi ketepatan kelulusan mahasiswa penerima KIP-Kuliah Merdeka
2. Bagaimana akurasi dari prediksi menggunakan algoritma C5.0.

### **1.3 Ruang Lingkup**

Penelitian yang dilakukan memiliki batasan dan ruang lingkup sebagai berikut:

1. Metode prediksi yang digunakan adalah algoritma C5.0.
2. Dataset yang digunakan adalah data penerima KIP Kuliah Merdeka di Universitas Teknologi Digital Indonesia sejumlah 195 *record* data, di mana dari data tersebut diambil sebanyak 80% untuk data latih dan 20% untuk data uji.
3. Ada 2 tahun angkatan yang digunakan pada data, yaitu tahun 2020/2021 dan 2021/2022.

4. Menggunakan atribut SKS (Satuan Kredit Semester), IPS1 (Indeks Prestasi Semester 1), IPS2 (Indeks Prestasi Semester 2), IPS3 (Indeks Prestasi Semester 3), IPS4 (Indeks Prestasi Semester 4), dan LABEL. Dimana LABEL memiliki 2 makna yaitu 0=lulus tepat waktu dan 1=tidak lulus tepat waktu. Serta menggunakan IPS2, IPS1 dan IPS4 untuk penerapan pohon keputusan berdasarkan nilai *gain ratio* tertinggi.
5. Proses *preprocessing* dilakukan di luar sistem menggunakan k-means untuk melakukan pengelompokan pada LABEL berdasarkan atribut yang ada.
6. Proses *training* dilakukan di luar sistem prediksi.
7. Kualitas hasil prediksi dihitung dengan menggunakan *confusion matrix*.
8. Aplikasi yang dibuat berbasis web yang menampilkan hasil prediksi.

#### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk menguji metode algoritma C5.0 pada ketepatan kelulusan mahasiswa penerima KIP-Kuliah Merdeka dan mengukur akurasi untuk melihat seberapa baik model yang dibuat.

#### **1.5 Manfaat Penelitian**

Penelitian ini diharapkan mampu memberikan kontribusi pada perguruan tinggi dalam memonitoring mahasiswa penerima KIP-Kuliah Merdeka untuk memaksimalkan semester yang tersisa agar lulus tepat waktu.

## 1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini terdiri dari lima bab yang akan diuraikan pada penjelasan berikut:

### a. Bab I Pendahuluan

Pada bab I ini berisi pengantar sekaligus memberikan hipotesa awal mengenai penelitian yang sedang dilakukan. Bab I berisi latar Belakang, rumusan masalah, ruang lingkup, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

### b. Bab II Tinjauan Pustaka dan Dasar Teori

Pada bab II ini berisi tinjauan Pustaka dan dasar teori. Dimana tinjauan pustaka digunakan sebagai acuan primer yang berisi tentang hasil penelitian terdahulu berkaitan dengan penelitian yang dilakukan, sedangkan teori berisi definisi dan uraian dari berbagai sumber yang telah dipublikasikan sehingga dapat digunakan untuk memahami pengertian dasar dan istilah dalam penelitian ini. Terdiri dari definisi *data mining*, prediksi, algoritma C5.0, dan *confusion matrix*.

### c. Bab III Metode Penelitian

Pada bab III ini berisi langkah-langkah metode penelitian yang dilakukan penulis meliputi bahan/data yang digunakan, prosedur dan pengumpulan data, serta analisis dan perancangan sistem.

### d. Bab IV Implementasi dan Pembahasan

Pada bab IV ini berisi laporan hasil penelitian berupa Implementasi sistem, Implementasi antarmuka, uji coba sistem, hasil uji coba sistem, dan pembahasan sistem.

e. Bab V Penutup

Pada bab V ini berisi kesimpulan dari hasil penelitian yang telah dilakukan dan saran untuk pengembangan penelitian selanjutnya.