

# **BAB I. PENDAHULUAN**

## **1.1. Latar Belakang**

Mahasiswa adalah orang yang belajar pada jenjang perguruan tinggi [1]. Menurut Hartaji (2012), mahasiswa adalah seseorang yang tengah menimba ilmu atau belajar dan terdaftar pada salah satu bentuk perguruan tinggi, yang terdiri dari akademi, politeknik, sekolah tinggi, institut, hingga universitas. Mahasiswa memiliki peranan yang sangat penting dalam kehidupan suatu negara. Mahasiswa berperan untuk menjamin kelangsungan hidup negara dan bangsa serta mewujudkan cita – cita pembangunan nasional. Pada umumnya mahasiswa akan menjalani proses pendidikan yang diselenggarakan oleh perguruan tinggi, yaitu mengikuti proses belajar mengajar, melakukan penelitian dan melakukan pengabdian pada masyarakat. Mahasiswa dapat dinyatakan lulus dari sebuah perguruan tinggi apabila sudah memenuhi syarat – syarat kelulusan yang ada di perguruan tinggi tersebut.

Universitas Atma Jaya Yogyakarta (UAJY) adalah salah satu perguruan tinggi swasta (PTS) yang ada di Indonesia. Universitas ini didirikan pada tanggal 27 September 1965 oleh kaum awam dan dikelola oleh Yayasan Slamet Rijadi – Yogyakarta, dibawah naungan Santo Albertus Magnus. Universitas ini memiliki empat kampus dengan satu kampus terletak di daerah Mrican, Yogyakarta dan tiga kampus terletak di daerah Babarsari, Yogyakarta. Universitas Atma Jaya Yogyakarta termasuk ke dalam 10 universitas swasta terbaik Indonesia versi QS AUR 2023 [2]. Dengan status tersebut maka tidak heran jika banyak siswa yang telah lulus dari jenjang Sekolah Menengah Atas (SMA), berminat untuk mendaftar dan menempuh pendidikan yang lebih tinggi di Universitas Atma Jaya Yogyakarta.

Untuk mempertahankan statusnya sebagai 10 universitas terbaik atau bahkan lebih baik lagi, maka Universitas Atma Jaya Yogyakarta harus memiliki penilaian akreditasi yang unggul. Akreditasi merupakan kegiatan yang dilakukan oleh organisasi atau badan mandiri di luar perguruan tinggi, dalam melakukan penilaian kelayakan dan mutu dari suatu perguruan tinggi atau program studi [3]. Akreditasi bertujuan untuk menentukan kelayakan dan mutu eksternal dari program studi atau

perguruan tinggi, berdasarkan kriteria yang mengacu pada Standard Nasional Perguruan Tinggi. Salah satu dari banyak faktor yang mempengaruhi penilaian akreditasi adalah kelulusan mahasiswa. Jika mahasiswa lulus tepat waktu maka akan sangat membantu dalam penilaian akreditasi perguruan tinggi tersebut. Sebaliknya, jika mahasiswa lulus tidak tepat waktu maka dapat memberikan pengaruh negatif terhadap penilaian akreditasi dari perguruan tinggi tersebut. Kriteria dari mahasiswa yang lulus tepat waktu adalah mahasiswa yang dapat menyelesaikan syarat-syarat kelulusan dalam kurun waktu empat tahun atau kurang.

Universitas Atma Jaya Yogyakarta membuka beberapa program studi salah satunya adalah Program Studi Informatika. Setiap tahunnya Program Studi Informatika menerima mahasiswa baru. Selain itu, Program Studi Informatika juga akan meluluskan mahasiswa yang sudah memenuhi syarat – syarat kelulusan di Universitas Atma Jaya Yogyakarta. Dengan adanya hal tersebut maka program studi ini juga akan menghasilkan data mahasiswa dalam jumlah yang besar.

Data mahasiswa yang dihasilkan dapat diolah menggunakan beberapa metode, agar bisa menghasilkan suatu informasi yang dapat dimanfaatkan oleh Program Studi Informatika. Salah satu proses yang bisa digunakan untuk mengolah data adalah dilakukannya proses *data mining*. *Data mining* adalah proses penggalian dan analisis sekumpulan data yang bertujuan untuk menemukan informasi yang belum diketahui sebelumnya [4]. Data yang ada akan diolah sedemikian rupa, sehingga mendapatkan informasi ataupun *insight* yang dapat membantu pembentukan sistem atau kebijakan. Kebijakan yang didapat dari pengolahan data tersebut nantinya dapat menjadi solusi dari permasalahan yang ingin diselesaikan.

Berdasarkan data mahasiswa angkatan 2012 hingga 2016 pada Program Studi Informatika di Universitas Atma Jaya Yogyakarta. Dari 717 mahasiswa, terdapat 291 mahasiswa yang lulus tepat waktu dan 426 mahasiswa yang lulus tidak tepat waktu, sehingga diperoleh persentase kelulusan sebesar 40.6%. Masalah tersebut tentunya akan berdampak negatif pada penilaian akreditasi program studi apabila dibiarkan terus menerus. Selain itu, masalah tersebut juga akan berdampak pada minat calon mahasiswa baru untuk mengambil Program Studi Informatika di

Universitas Atma jaya Yogyakarta.

Guna menyelesaikan permasalahan sesuai dengan uraian diatas maka digunakan proses *data mining* untuk mengolah data mahasiswa yang jumlahnya besar. Penelitian ini bertujuan untuk mendapatkan informasi tersembunyi yang terdapat pada data mahasiswa tersebut. Teknik *data mining* yang akan digunakan adalah klasifikasi. Klasifikasi adalah proses analisis data yang menghasilkan model yang menggambarkan dan membedakan kelas data atau konsep. Model hasil klasifikasi nantinya bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Teknik ini akan membagi mahasiswa menjadi dua yaitu mahasiswa yang lulus tepat waktu dan mahasiswa yang lulus tidak tepat waktu. Untuk algoritma, dalam penelitian ini algoritma yang digunakan adalah algoritma *decision tree* C4.5. Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk pohon keputusan. Algoritma ini memiliki kelebihan seperti dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, dan efisien dalam menangani atribut bertipe kontinyu dan diskrit [5]. Dalam proses pengklasifikasian, akan dilakukan dua buah proses yaitu proses pelatihan dan proses pengujian.

Penelitian ini juga akan menggunakan metode tambahan yaitu metode *ensemble* yang bernama *AdaBoost*. Algoritma *AdaBoost* atau *Adaptive Boosting* merupakan algoritma *machine learning* yang pertama kali diperkenalkan oleh Freund dan Schapire tahun 1995. Algoritma *AdaBoost* merupakan algoritma yang bersifat adaptif, karena hasil pembelajaran yang salah akan diklasifikasikan dan digunakan kembali pada proses klasifikasi selanjutnya. Algoritma *AdaBoost* akan membentuk *strong classification* dengan mengkombinasikan beberapa *weak classification* secara linier [6]. Tujuan dari penggunaan metode tambahan ini adalah diharapkan hasil yang diperoleh dalam penelitian ini, akan semakin baik dan akurat dalam memprediksi kelulusan mahasiswa.

Hasil penelitian diharapkan dapat membantu menemukan informasi ataupun *insight* tersembunyi dari data mahasiswa Program Studi Informatika, Universitas Atma Jaya Yogyakarta. *Insight* yang didapatkan dari penelitian ini diharapkan dapat membantu pihak program studi, dalam membuat sebuah kebijakan atau strategi

untuk mengatasi masalah mahasiswa yang lulus tidak tepat waktu.

## 1.2. Rumusan Masalah

Rumusan masalah yang ingin diselesaikan pada penelitian ini antara lain:

- a. Bagaimana menerapkan teknik *data mining* untuk menggali informasi ataupun *insight* terhadap data mahasiswa khususnya untuk memprediksi kelulusan mahasiswa?
- b. Bagaimana perbandingan dari performa algoritma *desicion tree* C4.5 dengan metode *AdaBoost* yang menggunakan algoritma C4.5 sebagai dasar pemodelan data?
- c. Bagaimana mengekstrak hasil klasifikasi agar dapat menjadi pengetahuan yang berguna bagi program studi untuk membantu mengambil keputusan?

## 1.3. Batasan Masalah

Untuk mengatasi penelitian yang tidak terarah, maka batasan penelitian pada penelitian ini adalah sebagai berikut:

- a. Ruang lingkup penelitian dibatasi pada data mahasiswa Program Studi Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
- b. Data mahasiswa yang digunakan pada penelitian ini adalah data mahasiswa angkatan 2012 dan 2016.

## 1.4. Tujuan Penelitian

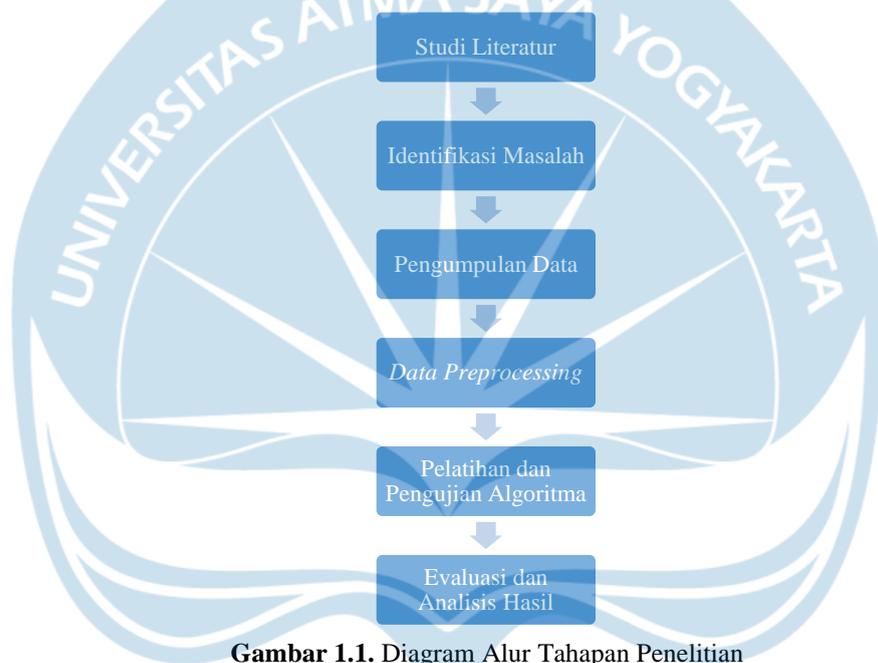
Adapun tujuan dari penelitian ini adalah:

- a. Mendapatkan model klasifikasi untuk memprediksi kelulusan mahasiswa.
- b. Untuk mengetahui performa dari algoritma *decision tree* C4.5 dan metode *AdaBoost* yang menggunakan algoritma C4.5 sebagai dasar pemodelan data dalam memprediksi kelulusan mahasiswa.
- c. Untuk mendapatkan informasi atau *insight* dari mengolah *dataset*

mahasiswa, yang dapat membantu program studi dalam mengambil keputusan dan memprediksi kelulusan mahasiswa.

### 1.5. Metode Penelitian

Metode penelitian adalah tahapan dan langkah – langkah penelitian yang disusun secara sistematis. Metode penelitian ini dilakukan dengan tujuan supaya hasil dari proses penelitian dapat sejalan dengan target dari penelitian. Diagram alur dari tahapan penelitian yang dilakukan pada penelitian ini dapat dilihat pada gambar 1.1.



Gambar 1.1. Diagram Alur Tahapan Penelitian

#### 1.5.1. Studi Literatur

Pada tahap ini, dilakukan proses studi dengan membaca beberapa buku dan penelitian terdahulu yang berhubungan dengan topik penelitian yang dibuat. Tujuan dilakukannya studi literatur ini adalah untuk mengumpulkan referensi dan memahami pengetahuan yang telah ada dari buku dan penelitian terdahulu. Referensi yang digunakan untuk penelitian ini berasal dari jurnal, buku, *e-book*, dan tulisan ilmiah lainnya.

### **1.5.2. Identifikasi Masalah**

Pada tahap ini, dilakukan proses pengidentifikasian masalah yang terjadi dan ingin diselesaikan pada penelitian ini. Tujuan dilakukannya identifikasi masalah adalah untuk menetapkan tujuan penelitian.

### **1.5.3. Pengumpulan Data**

Dalam tahap ini, dilakukan proses pengumpulan data yang diperlukan untuk melakukan penelitian. Pengumpulan data dilakukan dengan cara meminta data secara langsung kepada pihak Kantor Sistem Informasi, Universitas Atma Jaya Yogyakarta. Data yang dikumpulkan berupa Data Lengkap Mahasiswa dan Data IPS Mahasiswa dari Program Studi Informatika angkatan 2009 hingga 2019.

### **1.5.4. Data Preprocessing**

Pada tahap ini, data-data yang telah berhasil diperoleh akan langsung diidentifikasi dan di *preprocessing*. *Data Preprocessing* merupakan proses mengolah data untuk menghapus data yang tidak diperlukan pada penelitian, data yang tidak konsisten, *outlier*, dan *missing value*. Tujuan dilakukannya *Data Preprocessing* adalah untuk mentransformasikan data mentah menjadi data yang lebih rapi, mudah dibaca, dan lebih berkualitas. *Data Preprocessing* yang dilakukan pada penelitian ini meliputi pembersihan data, integrasi data, transformasi data, reduksi data, dan diskritisasi data.

### **1.5.5. Pelatihan dan Pengujian Algoritma**

Pada tahap ini, akan dilakukan proses klasifikasi dengan dua tahap berbeda yaitu pelatihan dan pengujian. Data latih yang sudah diproses dan diolah pada tahap sebelumnya akan diolah menggunakan algoritma *Decision Tree C4.5* dan metode *AdaBoost*. Penerapan algoritma *Decision Tree C4.5* dan metode *AdaBoost* pada data latih dilakukan

dengan tujuan, untuk menemukan model yang akan digunakan pada proses pengujian. Sedangkan data uji yang sudah diproses dan diolah pada tahap sebelumnya, akan diolah menggunakan model yang didapatkan pada proses pelatihan. Penerapan model pada data uji bertujuan untuk memprediksi kelulusan mahasiswa. Selain itu, hasil dari penerapan algoritma ini digunakan untuk mendapatkan informasi dari *dataset* yang digunakan dalam penelitian ini.

#### **1.5.6. Evaluasi dan Analisis Hasil**

Pada tahap ini, akan diuraikan hasil dan kesimpulan yang didapat dari proses dilakukannya penelitian. Hasil yang didapat adalah pengukuran performa dari algoritma dengan melihat nilai akurasi dan nilai *Area Under Curve* (AUC). Selain itu, hasil dari penelitian ini berupa informasi yang didapatkan dari hasil menganalisis *dataset* yang digunakan pada penelitian ini. Informasi yang didapatkan nantinya bisa berguna bagi program studi untuk membantu pengambilan keputusan.

### **1.6. Sistematika Penulisan**

#### **1.6.1. BAB I PENDAHULUAN**

Pada bab satu yaitu pendahuluan membahas tentang latar belakang mengapa penelitian ini dilakukan, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

#### **1.6.2. BAB II TINJAUAN PUSTAKA**

Pada bab dua yaitu tinjauan pustaka membahas tentang penelitian terdahulu dan perbandingan penelitian yang terkait dengan topik pada penelitian ini.

#### **1.6.3. BAB III LANDASAN TEORI**

Pada bab tiga yaitu landasan teori membahas tentang teori-teori yang digunakan pada penelitian ini.

#### **1.6.4. BAB IV DATASET DAN DATA PREPROCESSING**

Pada bab empat yaitu *dataset* dan *data preprocessing* yang menjelaskan tentang *dataset* yang digunakan dalam penelitian. Selain itu, bab ini juga membahas tentang pengolahan data dengan cara *Data Preprocessing* yang meliputi pembersihan data, integrasi data, transformasi data, reduksi data, dan diskritisasi data.

#### **1.6.5. BAB V PENERAPAN ALGORITMA DAN PENGUJIAN MODEL**

Pada bab lima akan membahas tentang penerapan algoritma, proses pelatihan, proses pengujian, dan analisis hasil penelitian terhadap *dataset* yang digunakan. Hasilnya berupa performa dan *rule* dari model algoritma serta informasi yang dapat digunakan untuk membantu mengambil keputusan.

#### **1.6.6. BAB VI PENUTUP**

Pada bab enam yaitu penutup membahas tentang kesimpulan dan saran dari hasil penelitian. Kesimpulan berisikan tentang ringkasan hasil penelitian dan jawaban dari rumusan masalah pada penelitian ini. Saran berisikan tentang saran-saran yang berguna untuk memaksimalkan penelitian selanjutnya.