

# **BAB I**

## **PENDAHULUAN**

Pada bab pendahuluan akan dijelaskan mengenai latar belakang, identifikasi dan perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan tesis. Latar belakang membahas tentang latar belakang yang mendasari penelitian ini di lakukan.

### **1.1 Latar Belakang**

Kemajuan teknologi dalam berbagai bidang sangatlah pesat (APJII, 2020), salah satunya teknologi di bidang Pendidikan yaitu E-Learning (McCue, 2018). Saat ini E-Learning merupakan sesuatu yang wajib untuk menunjang proses pembelajaran (Hartanto, 2016), salah satunya Akademi Anti Korupsi merupakan platform belajar dengan basis digital yang mudah diakses oleh semua masyarakat sehingga dapat belajar mengenai korupsi dan cara-cara memerangnya. Akademi Anti Korupsi sudah berjalan sejak tahun 2018 (Indonesia Corruption Watch, 2019). Terdapat dua puluh mata kuliah yang dapat diakses secara virtual. Pada dasarnya tidak ada persyaratan khusus untuk menjadi peserta dalam program ini.

Platform Akademi Anti Korupsi menggunakan website sebagai penunjang utama pelaksanaannya. Langkah pertama yang dilakukan peserta untuk mengakses Akademi Anti Korupsi adalah dengan melengkapi form registrasi peserta pada halaman [akademi.antikorupsi.org](http://akademi.antikorupsi.org) berupa jenis kelamin, pekerjaan, pengalaman pendidikan anti korupsi, umur, institusi, domisili, pengalaman *e-learning* dan alasan mengikuti *e-learning* anti korupsi. Setiap modul pembelajaran atau *course* berisikan konten video pembelajaran yang menjelaskan materi perkuliahan, kuis, tugas, dan forum yang harus diselesaikan peserta setelah menerima materi pembelajaran hingga dapat mengunduh sertifikat pada aktifitas terakhir. Kurun waktu pembelajaran dapat diselesaikan dalam waktu tiga bulan.

Akademi Anti Korupsi memiliki jumlah peserta pada kurun waktu 2018-2021 sebanyak 12.385 (Indonesian Corruption Watch, 2022). Pada tahun 2021 mengalami penurunan jumlah pendaftar secara drastis dari tahun sebelumnya sebesar 52% (Indonesian Corruption Watch, 2021), (Indonesian Corruption Watch, 2022). Tetapi tidak diimbangi juga dengan jumlah lulusan yang dihasilkan oleh akademi Anti Korupsi sebesar 71%. Akademi Anti Korupsi memiliki tujuan strategi bisnis untuk meningkatkan jumlah pendaftaran peserta yang dapat lulus (Indonesian Corruption Watch, 2022). Jumlah pendaftar mengalami penurunan dari tahun sebelumnya sehingga hal ini akan mengakibatkan tidak tercapainya tujuan strategi Akademi Anti Korupsi dari segi peserta yaitu tidak relevannya persentase jumlah peserta yang mendaftar dengan jumlah peserta yang lulus setiap tahunnya.

Konsep e-learning dapat memberikan banyak informasi melalui data yang dihasilkan dalam proses yang ada di dalam sistem. Beberapa informasi yang bisa didapatkan adalah sasaran atau target adopter dari program Akademi Anti Korupsi. Untuk mengolah data tersebut dilakukan dengan memantau prediksi kelulusan jumlah mata kuliah (*course*) yang dapat diambil melalui form registrasi peserta e-learning. Sehingga merujuk pada data tersebut, dapat diketahui sasaran atau target dalam Akademi Anti Korupsi untuk mengatasi permasalahan menurunnya jumlah pendaftar pada tahun 2021 dan meningkatkan kelulusan peserta.

Masalah ini dapat diselesaikan dengan menggunakan data mining. Data mining dapat dibangun menggunakan teknik perangkungan yang pada dasarnya menggunakan teknik klasifikasi (Suyanto, 2019), (Fabriane et al., 2017). *Decision Tree* adalah salah satu yang paling terkenal metode klasifikasi karena proses formulasi nya yang relatif cepat dan mudah dan mudah dipahami karena kesederhanaan (Witten et al., 2016). *Decision Tree* sebagai metode klasifikasi yang menggunakan representasi struktur pohon dimana node mewakili atribut tertentu dan daun mewakili kelas dengan node paling atas menjadi root (Witten et al., 2016). Metode ini dapat digunakan pada masalah seperti kinerja siswa (D. Funcion, 2017), (Hemantkumar & Adholiya, 2019) dan mengklasifikasikan

mahasiswa putus sekolah (Wiyono et al., 2020), dan memprediksi tingkat kelulusan mahasiswa (Putri et al., 2019). Kemudian terdapat metode yang memiliki konsep mirip decision tree yaitu random forest. Metode Random Forest ini merupakan kombinasi masing-masing tree dari decision tree yang kemudian digabungkan menjadi satu model (Petkovic et al., 2016). Studi kasus dengan menggunakan Akademi Antikorupsi dipilih sebagai objek untuk menerapkan pendekatan *Decision Tree* dan *Random Forest* dengan fokus pada penggunaan klasifikasi untuk mengetahui faktor kelulusan course dan memprediksi kelulusan course.

Terdapat beberapa penelitian sebelumnya yang menerapkan model klasifikasi dibidang pendidikan. Penelitian yang dilakukan oleh (Utari et al., 2020) melakukan prediksi potensi siswa putus sekolah dan mengidentifikasi faktor-faktor potensial berkaitan dengan siswa putus sekolah menggunakan *Random Forest*. Utari dkk menghilangkan data *missing value* pada data preparation dengan pendistribusian dataset adalah cross-validation. Hasil penelitian diperoleh metode *Random Forest* mendapatkan akurasi sebesar 93,43%. Kemudian, peneliti (Budiman et al., 2018) melakukan prediksi kelulusan mahasiswa menggunakan metode *Decision Tree* dengan algoritma C4.5. Budiman dkk menghilangkan data *missing value* pada data preparation dengan pendistribusian dataset adalah *split validation*. Hasil penelitian menunjukkan bahwa Algoritma C4.5 memiliki akurasi (AC) sebesar 78,57% dengan true positive rate (TP) sebesar 76,72% dengan menggunakan data *testing* 90% memiliki nilai akurasi kinerja terbaik. Selanjutnya, peneliti (Benediktus & Oetama, 2020) melakukan prediksi prestasi siswa menggunakan metode *Decision Tree* dengan algoritma C5.0 merupakan penyempurnaan algoritma C4.5. Benediktus dan Oetama tidak menghilangkan data *missing value* pada data preparation dengan pendistribusian dataset secara random. Hasil diperoleh bahwa aalgoritma C.50 dibagi menjadi data latih sebesar 75% dan data uji sebesar 25% diperoleh akurasi sebesar 71,667%.

Berdasarkan penelitian-penelitian yang dilakukan sebelumnya dalam melakukan prediksi kelulusan menggunakan model klasifikasi, terdapat

perbedaan dalam melakukan persiapan data. Beberapa penelitian ada yang menghilangkan data *missing value* (Utari et al., 2020), (Budiman et al., 2018). Tetapi ada juga yang tidak menghilangkan data *missing value* (Benediktus & Oetama, 2020). Kemudian terdapat perbedaan dalam melakukan pendistribusian dataset. Beberapa peneliti ada yang menggunakan cross validation (Utari et al., 2020), secara random (Benediktus & Oetama, 2020), dan menggunakan *split validation* (Baswardono et al., 2019). Hal tersebut menjadi salah satu fokus pada tujuan penelitian ini, yaitu untuk menentukan bagaimana pengaruh *missing value* pada data dalam melakukan prediksi kelulusan dan mengetahui konsistensi hasil perhitungan data *training* masing-masing metode klasifikasi menggunakan *split validation*. Selain itu, belum terdapat penelitian terkait perbandingan metode *Decision Tree* dan *Random Forest* untuk melakukan prediksi kelulusan E-Learning. Sehingga peneliti juga akan membandingkan keakurasian dari model klasifikasi menggunakan metode *Decision Tree* menggunakan algoritma C5.0 dan metode *Random Forest*. Banyak *Tools* atau *software* yang ditawarkan untuk melakukan proses klasifikasi salah satunya adalah aplikasi Orange (Ishak et al., 2020). Orange merupakan suatu aplikasi yang dapat digunakan untuk menganalisa data yang mengedepankan *Visual Programming (Free Coding)* pada data mining Orange dikembangkan menggunakan Bahasa pemrograman Python (Ram & Vishwakarma, 2021).

Dalam penelitian ini akan membandingkan akurasi dari metode *Decision Tree* dengan algoritma C5.0 dan metode *Random Forest* dalam memprediksi kelulusan *course*. Dengan adanya penelitian ini diharapkan dapat diketahui faktor yang mempengaruhi kelulusan *course* untuk dijadikan sasaran atau target bisnis Akademi Anti Korupsi.

## 1.2 Identifikasi dan Rumusan Masalah

### 1.2.1 Identifikasi Masalah

Identifikasi Masalah berdasarkan latar belakang yang telah disampaikan sebagai berikut:

1. Terdapat beberapa faktor yang dapat mempengaruhi kelulusan *course* peserta Akademi Anti Korupsi.
2. Belum ada penelitian yang melakukan perbandingan model klasifikasi menggunakan Metode *Decision Tree* dan *Random Forest* untuk prediksi kelulusan *course* pada *E-Learning*.
3. Dalam melakukan prediksi kelulusan dengan model klasifikasi, terdapat beberapa perbedaan pada penelitian saat melakukan data *preparation* yaitu ada yang menghilangkan *missing value* dan ada yang tidak menghilangkan *missing value*.
4. Belum terdapat data konsistensi data *training* dan data *testing* pada perhitungan masing-masing model algoritma klasifikasi.

### 1.2.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Faktor-faktor apa yang mempengaruhi kecenderungan kelulusan *course* peserta Akademi Anti Korupsi.
2. Bagaimana perbandingan hasil akurasi metode *Decision Tree* dan metode *Random Forest* untuk memprediksi kelulusan *course*.
3. Bagaimana pengaruh *missing value* dalam melakukan prediksi kelulusan *course* menggunakan model klasifikasi.
4. Bagaimana konsistensi hasil perhitungan masing-masing model algoritma klasifikasi metode *Decision Tree* dan *Random Forest* pada berbagai tingkatan jumlah data *training*.

### **1.3 Tujuan Penelitian**

Sesuai dengan perumusan masalah yang ada maka tujuan penelitian sebagai berikut:

1. Untuk mengetahui faktor – faktor yang mempengaruhi kelulusan *course* peserta Akademi Anti Korupsi lulus.
2. Untuk mengetahui perbandingan hasil akurasi antara metode Decision Tree dan Random Forest dalam memprediksi kelulusan *course*.
3. Untuk mengetahui pengaruh *missing value* dalam melakukan prediksi kelulusan *course* dengan model klasifikasi.
4. Untuk mengetahui konsistensi perhitungan masing-masing metode *Decision Tree* dan *Random Forest* pada tingkatan jumlah data *training*.

### **1.4 Manfaat Penelitian**

Sesuai dengan rumusan masalah dan tujuan yang telah disebutkan, maka penelitian ini diharapkan berguna bagi Operasional, bagi pengembang ilmu pengetahuan, dan bagi penulis sendiri.

#### **1.4.1 Manfaat Operasional**

Kegunaan penelitian ini bagi operasional sebagai berikut:

1. Dapat dijadikan sebuah alat prediksi dengan memasukan faktor kelulusan yang dimiliki.
2. Sebagai pedoman dalam pemanfaatan metode *Decision Tree* dan *Random Forest* dalam menentukan kemungkinan lulus seseorang.

### **1.4.2 Manfaat Penggunaan Ilmu**

Kegunaan penelitian ini bagi pengembangan ilmu sebagai berikut:

1. Dengan menggunakan model algoritma yang memiliki akurasi tinggi, dapat dijadikan rujukan untuk menentukan apakah seseorang dapat dinyatakan lulus.
2. Mengetahui faktor-faktor yang menyebabkan kelulusan seseorang yang dijadikan sebagai dasar untuk menentukan tingkat kelulusan seseorang melalui data *training* dan *testing*.

### **1.5 Batasan Masalah**

Membuat model yang memberikan klasifikasi tentang jumlah *course* yang dapat diambil peserta pada awal masa kuliah e-learning dan tidak menangani pemberian solusi manajerial untuk mengatasi permasalahan tidak lulus peserta tersebut.

### **1.6 Sistematika Penulisan**

Sistematika yang digunakan dalam penulisan tesis ini adalah sebagai berikut:

#### **1. BAB I PENDAHULUAN**

Berisi Latar Belakang yang mendorong dilakukannya penelitian, identifikasi masalah, perumusan masalah serta tujuan dan manfaat penelitian.

#### **2. BAB II KAJIAN PUSTAKA**

Berisi tentang penjelasan dasar ilmu yang mendukung pembahasan penelitian ini meliputi: Model Klasifikasi, *Decision Tree*, Algoritma C5.0, Metode CRISP-DM dan Penelitian terdahulu.

#### **3. BAB III METODOLOGI PENELITIAN**

Metodologi penelitian menguraikan secara rinci pendekatan yang digunakan dalam penelitian yaitu metode *Decision Tree* dan metode

*Random Forest* dengan menggunakan metode CRISP-DM serta bantuan aplikasi *Excel* dan *Orange*.

#### 4. BAB IV ANALISA DAN PEMBAHASAN

Berisi tentang Analisa dan Pembahasan penelitian mengenai model klasifikasi dalam memprediksi jumlah *course* yang dapat diambil oleh peserta menggunakan algoritma *Decision Tree* dan metode *Random Forest*.

#### 5. BAB V KESIMPULAN DAN SARAN

Berisi tentang kesimpulan dari hasil penelitian dan saran untuk mengembangkan penelitian yang telah dilakukan.