

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dalam era perkembangan teknologi dan informasi seperti saat ini, penerapan *database* sudah banyak digunakan khususnya di dunia pendidikan. Penerapan *database* tidak dapat dipisahkan dari dunia pendidikan contohnya Universitas Negeri Jakarta sebagai salah satu perguruan tinggi yang menggunakan *database* untuk menyimpan data yang berupa data mahasiswa, data dosen, data fakultas, dan data lainnya yang berkaitan dengan universitas.

*Database* yang ada dapat digunakan universitas untuk menghasilkan informasi tertentu. Data mahasiswa yang telah lulus merupakan salah satu *database* yang menghasilkan informasi penting. Data mahasiswa yang lulus dapat dimanfaatkan dengan maksimal untuk memberikan informasi yang bermanfaat bagi universitas. Data kelulusan mahasiswa tepat waktu menjadi salah satu indikator penilaian kelayakan program studi sebagai unit pelaksana pendidikan pada perguruan tinggi (BAN-PT, 2019), di mana menurut Buku Pedoman Akademik (UNJ, 2020, p. 142) tepat waktu adalah apabila program studi diselesaikan selama masa studi 8 (delapan) semester dan batas maksimal 14 (empat belas) semester dengan minimal jumlah satuan kredit semester 144 sks.

Berdasarkan data yang diperoleh dari Program Studi Pendidikan Teknik Informatika dan Komputer yang merupakan salah satu program studi di Universitas Negeri Jakarta, jumlah mahasiswa yang menempuh pendidikan selama 3,5 tahun sebanyak 23 orang dari angkatan 2010 sampai angkatan 2018. Sedangkan jumlah mahasiswa Program Studi Pendidikan Teknik Informatika dan Komputer yang lulus tepat waktu sebanyak 595 lulusan. Pada tahun 2010 terdapat 98 mahasiswa, sedangkan mahasiswa yang dapat lulus tepat waktu sebanyak 18 sehingga diperoleh presentasi kelulusan sebesar 18,36% pada periode tersebut. Pada tahun 2011 terdapat 131 mahasiswa, sedangkan mahasiswa yang dapat lulus tepat waktu sebanyak 20 sehingga diperoleh presentasi kelulusan sebesar 15,26%. Pada tahun 2012 terdapat 133 mahasiswa, sedangkan yang dapat lulus tepat waktu sebanyak 72 sehingga diperoleh presentasi kelulusan sebesar 54,13%. Pada tahun 2013 terdapat

90 mahasiswa, sedangkan mahasiswa yang dapat lulus tepat waktu sebanyak 33 sehingga diperoleh presentasi kelulusan sebesar 36,67%. Pada tahun 2014 terdapat 74 mahasiswa, sedangkan mahasiswa yang dapat lulus tepat waktu sebanyak 29, sehingga diperoleh presentasi kelulusan sebesar 39,18%. Berdasarkan persentase di atas dapat disimpulkan selama 5 tahun persentase rata-rata kelulusan mahasiswa Pendidikan Teknik Informatika dan Komputer sebesar 32,72%.

Data yang diperoleh dari Program Studi Teknik Informatika dan Komputer sebelum dijadikan sebuah prediksi harus melalui proses *data mining*. *Data mining* merupakan suatu proses penemuan sebuah informasi baru dalam sekumpulan data yang sangat besar (Iriadi & Nuraeni, 2016, p. 133). Penemuan data dengan proses *data mining* ini diharapkan dapat menghasilkan suatu pola atau aturan yang dapat digunakan untuk menghasilkan suatu informasi seperti prediksi kelulusan mahasiswa.

Prediksi kelulusan mahasiswa dapat digunakan lebih lanjut untuk membantu program studi dalam mengevaluasi dan memperbaiki sistem pembelajaran sehingga program studi dapat menghasilkan lulusan yang berkualitas. Penelitian ini akan dilaksanakan di Program Studi Teknik Informatika dan Komputer. Program Studi Teknik Informatika dan Komputer telah berdiri dari tahun 2009 dan memiliki sasaran untuk menjadi program studi unggulan. Oleh karena itu, hasil prediksi kelulusan mahasiswa dapat membantu Program Studi Teknik Informatika dan Komputer dalam mengambil langkah strategis sehingga dapat meningkatkan jumlah mahasiswa lulus tepat waktu.

Berdasarkan penelitian sebelumnya tentang prediksi kelulusan mahasiswa menggunakan Teorema Bayes (Studi Kasus: Universitas Pendidikan Ganesha) didapatkan nilai akurasi tertinggi yang dihasilkan Teorema Bayes dalam prediksi lama studi mahasiswa adalah 80,96% yang melibatkan seluruh fitur klasifikasi (full testing). Nilai rata-rata *recall* adalah 80,3% (Bendesa Subawa, 2019, p. 234).

Hasil penelitian lain menyatakan bahwa prediksi tingkat kelulusan mahasiswa menggunakan algoritma C4.5 dengan pembagian data sebanyak 70:30 menghasilkan nilai akurasi sebesar 76,10%. Atribut yang mempengaruhi kelulusan mahasiswa adalah IPK semester 1 (Budiman et al., 2019, p. 92).

Hasil penelitian lain yang dilakukan oleh (Permana et al., 2022, p. 168) tentang Perbandingan Algoritma C4.5 dan Naïve Bayes Untuk Prediksi Ketepatan Waktu Studi Mahasiswa (Studi Kasus: Program Studi Statistika Universitas Mulawarman) menghasilkan tingkat akurasi pada algoritma Naïve Bayes sebesar 69,97% sedangkan tingkat akurasi algoritma C4.5 sebesar 61,99%

Penelitian yang dilakukan oleh (Ariadi, 2020, p. 7) mengenai Analisa Perbandingan Algoritma DT C.45 dan Naïve Bayes Dalam Prediksi Penerimaan Kredit Motor menghasilkan tingkat akurasi pada algoritma C4.5 sebesar 91,43% sedangkan tingkat akurasi pada algoritma Naïve Bayes sebesar 89,52%.

Berdasarkan penelitian sebelumnya, maka penulis ingin menggabungkan perbandingan kedua algoritma yaitu algoritma Naïve Bayes dan algoritma C4.5 dalam prediksi kelulusan mahasiswa.

Algoritma Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes dengan asumsi independensi (ketidaktergantungan) yang kuat menurut (Prasetyo, 2012) yang diacu dalam (Wijaya & Dwiasnati, 2020, p. 3). Sedangkan Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal (Hairul Umam et al., 2017, p. 3).

Algoritma Naïve Bayes dan algoritma C4.5 merupakan algoritma yang bersifat *supervised learning* yang artinya dalam pengklasifikasian kelas memerlukan pelatihan atau pembelajaran terlebih dahulu (Bendesa Subawa, 2019, p. 229). Sehingga diperlukan *data training* dan *data testing*. *Data training* adalah data yang digunakan untuk melakukan pembentukan model. Sedangkan *data testing* adalah data yang digunakan untuk pengujian model. Pada penelitian ini digunakan data lulusan tahun 2017-2020 dengan perbandingan *data training* dan *data testing* sebesar 70:30.

Klasifikasi algoritma untuk prediksi dalam data mining dapat dibagi menjadi dua, yakni *rule-based* dan *non-rule based*. *Rule-based* merupakan klasifikasi yang menggunakan aturan 'jika-maka'. Algoritma yang termasuk kategori *rule-based* adalah C4.5 dan Classification and Regression Trees (CART). Sedangkan *non-rule based* tidak menggunakan aturan 'jika-maka'. Algoritma yang termasuk kategori

*non-rule based* adalah Support Vector Machine (SVM), Hidden Markov Model (HMM), Naïve Bayes. (Hadiwandura, 2019, p. 77) melakukan perbandingan kinerja model klasifikasi pada 4 dataset berbeda. Hasil penelitian menunjukkan tingkat akurasi kelas model Bayesian lebih stabil dibanding kelas model lainnya apabila diberi dataset dengan jumlah atribut yang lebih banyak.

Pengelompokkan klasifikasi akan menggunakan atribut berdasarkan nama, NIM (Nomor Induk Mahasiswa), jenis kelamin, nilai IPK dari semester 1 – 5, asal sekolah, asal daerah dan jalur masuk perguruan tinggi. Dengan kelas data yaitu tepat waktu dan tidak tepat waktu. Setelah menentukan atribut dan kelas data, langkah selanjutnya adalah mengimplementasikan algoritma Naïve Bayes dan algoritma C4.5 untuk memprediksi kelulusan mahasiswa.

Berdasarkan masalah dan peluang yang ada maka penulis tertarik untuk mencari mekanisme dalam memprediksi kelulusan mahasiswa dengan menggunakan perbandingan algoritma yang dipilih adalah algoritma Naïve Bayes dan algoritma C.45. Perbandingan algoritma Naïve Bayes dan algoritma C.45 untuk mencari tahu dari kedua algoritma tersebut mana yang memberikan kinerja terbaik. Hasil dari perbandingan ini akan dijadikan sebagai prediksi kemungkinan mahasiswa yang lulus apakah tepat waktu atau tidak tepat waktu. Diharapkan hasil dari perbandingan ini dapat memprediksi kelulusan mahasiswa Pendidikan Teknik Informatika dan Komputer serta dapat digunakan sebagai sarana dalam menentukan langkah dan kebijakan bagi mahasiswa yang menargetkan lulus tepat waktu.

## 1.2 Identifikasi Masalah

Identifikasi masalah yang dapat digunakan sebagai bahan untuk penelitian yaitu:

1. Membandingkan kedua algoritma yakni algoritma Naïve Bayes dan algoritma C4.5 yang memberikan kinerja terbaik dalam prediksi kelulusan mahasiswa.
2. Mencari tahu perbandingan akurasi dari algoritma Naïve Bayes dan algoritma C4.5 dalam melakukan prediksi kelulusan mahasiswa.

3. Perlu adanya suatu teknik untuk pencarian informasi prediksi ketepatan waktu kelulusan mahasiswa di Program Studi Pendidikan Teknik Informatika dan Komputer.

### **1.3 Batasan Masalah**

Agar dapat memfokuskan pengerjaan penelitian, maka ditetapkan batasan permasalahan yaitu:

1. Data yang digunakan yaitu data mahasiswa Program Studi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta yang sudah lulus dalam kurun waktu 4 tahun terakhir yaitu lulusan 2017 – 2020
2. Metode prediksi yang digunakan untuk memprediksi waktu kelulusan mahasiswa menggunakan algoritma Naïve Bayes dan algoritma C4.5.

### **1.4 Rumusan Masalah**

Berdasarkan latar belakang permasalahan yang telah diuraikan sebelumnya, maka perlu merumuskan beberapa pokok permasalahan yang diuraikan sebagai berikut:

“Bagaimana tingkat akurasi dan hasil perbandingan performa dari Algoritma Naïve Bayes dan Algoritma C.45 dalam memprediksi kelulusan mahasiswa Program Studi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta?”

### **1.5 Tujuan Penelitian**

Tujuan penelitian tugas akhir ini adalah membandingkan tingkat akurasi dan kinerja algoritma Naïve Bayes dan algoritma C4.5 dalam memprediksi kelulusan mahasiswa Program Studi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta.

## 1.6 Kegunaan Penelitian

Dalam penelitian ini di harapkan akan memberikan kegunaan baik secara teoritis maupun praktis, sebagai berikut:

### 1.6.1 Kegunaan Teoritis

Secara teoritis, hasil penelitian ini diharapkan dapat memberikan kegunaan yaitu:

1. Memberikan pemikiran dan karya ilmiah baru dalam ilmu Pendidikan Teknik Informatika dan Komputer dalam hal prediksi kelulusan mahasiswa.
2. Sebagai referensi untuk penelitian-penelitian selanjutnya yang berhubungan dengan perbandingan algoritma dalam menentukan prediksi kelulusan mahasiswa serta menjadi bahan kajian lebih lanjut.

### 1.6.1 Kegunaan Praktis

Secara akademis, diharapkan penelitian ini dapat memberikan kegunaan diantaranya:

1. Bagi Akademik

Kegunaan bagi akademik yaitu dapat memberikan karya baru yang diharapkan dapat memberikan kegunaan dan dapat dikembangkan untuk masa yang akan datang

2. Bagi Mahasiswa

Bagi mahasiswa, tugas akhir ini dapat dijadikan sebagai referensi ataupun pedoman bagi mahasiswa yang lain yang hendak mengambil penelitian sejenis.

3. Bagi Universitas

Kegunaan bagi universitas yaitu dapat memprediksi masa studi mahasiswa Program Studi Pendidikan Teknik Informatika dan Komputer atau yang lainnya.