

BAB 3

METODE PENELITIAN

Penelitian ini menggunakan metode *Decision Tree* dengan algoritma *CART* (*Classification and Regression Tree*) untuk menganalisis dan memprediksi tingkat kelulusan mahasiswa tugas akhir di FTTI Unjaya. *CART* merupakan algoritma pohon keputusan yang membentuk struktur pohon biner, di mana setiap node hanya memiliki dua cabang, dan proses pemisahan data dilakukan berdasarkan kriteria *Gini Index* untuk memperoleh kelompok data yang paling homogen. Tahapan penelitian dimulai dengan identifikasi masalah dan pengumpulan data dari mahasiswa tugas akhir FTTI Unjaya, yang mencakup IPK, IPS, nama, jenis kelamin, program studi, jumlah SKS yang diambil, jumlah mata kuliah yang diikuti, serta mata kuliah yang harus diulang. Data yang telah dikumpulkan kemudian melalui tahap *preprocessing*, termasuk pembersihan, transformasi, dan normalisasi, sebelum dibagi menjadi data latih (70%) dan data uji (30%). Model *Decision Tree* dengan algoritma *CART* dibangun dengan menentukan atribut yang paling berpengaruh secara hierarkis hingga terbentuk struktur pohon keputusan yang optimal. Evaluasi model dilakukan dengan mengukur akurasi, *Confusion Matrix*, serta nilai *Precision*, *Recall*, dan *F1-Score* untuk menilai performa model. Hasil penelitian ini digunakan untuk memprediksi tingkat kelulusan mahasiswa, mengidentifikasi faktor utama yang memengaruhi kelulusan, serta memberikan rekomendasi bagi pihak akademik dalam meningkatkan keberhasilan studi mahasiswa.

3.1 BAHAN DAN ALAT PENELITIAN

Dalam penelitian ini, perangkat yang dimanfaatkan adalah komputer dengan spesifikasi yang cukup untuk memastikan sistem dan perangkat lunak dapat berjalan secara optimal, serta didukung oleh koneksi internet yang stabil guna menunjang kelancaran seluruh tahapan penelitian.

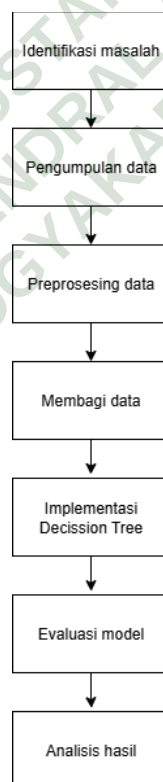
Sistem Operasi dan program aplikasi yang dipergunakan dalam pengembangan aplikasi ini adalah:

1. Sistem Operasi: Windows 11
2. Perangkat Keras : Processor intel Core i5, RAM 8GB.
3. Perangkat Lunak : Google Collab, Streamlit

Bahan penelitian yang digunakan adalah data mahasiswa dari angkatan tahun 2018 sampai 2021, yang berisi nama, angkatan, Indeks Prestasi Kumulatif (IPK), Satuan Kredit Semester Kumulatif (SKSK), Indeks Prestasi Semester (IPS), jumlah Satuan Kredit Semester (SKS), dan data tersebut diperoleh dari bagian akademik FTTI UNJAYA.

3.2 JALAN PENELITIAN

Jalan penelitian ini meliputi identifikasi masalah, pengumpulan data, Preprosesing data, membagi data, implemanetasi decision tree, evaluasi model, dan analisis hasil. Jalan penelitian dapat dilihat pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

Berikut ini uraian dari masing-masing tahap:

1. Pengumpulan Data

Data dikumpulkan melalui sumber yang relevan, seperti database akademik atau rekam jejak mahasiswa, mencakup variabel seperti IPK, jumlah SKS yang ditempuh, dan kehadiran.

2. Preprocessing Data

Data yang telah terkumpul kemudian diproses melalui tahap pembersihan untuk menangani normalisasi data, mengidentifikasi dan menangani outlier, serta melakukan transformasi agar data siap digunakan dalam proses pemodelan.

3. Pembagian Data

Data dikelompokkan ke dalam dua bagian, yaitu data latih dan data uji, dengan pembagian yang umumnya menggunakan rasio seperti 70:30. Data latih berfungsi untuk membentuk atau melatih model, sedangkan data uji digunakan untuk mengevaluasi kinerja model yang telah dibangun.

4. Implementasi Decision Tree

Decision Tree diterapkan pada data training untuk membangun model prediksi. Model ini akan menghasilkan pohon keputusan yang dapat diterapkan untuk menentukan mahasiswa yang berpotensi lulus sesuai masa studi atau tidak.

5. Evaluasi Model

Model dievaluasi menggunakan data testing untuk menilai akurasi, presisi, dan performanya secara keseluruhan. Dalam proses evaluasi ini, digunakan metrik seperti confusion matrix, precision, recall, dan F1-score. Confusion matrix adalah sebuah tabel yang menyajikan jumlah prediksi benar dan salah untuk masing-masing kelas, yang berguna untuk menganalisis kesalahan dalam proses klasifikasi. Precision menunjukkan seberapa besar proporsi prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dihasilkan oleh model. Recall menunjukkan sejauh mana model mampu mengidentifikasi seluruh kasus positif yang sebenarnya ada dalam data. Adapun F1-score merupakan rata-rata harmonis

antara precision dan recall, yang berguna untuk memberikan penilaian seimbang, terutama saat distribusi kelas dalam data tidak merata.

6. Analisis Hasil

Pada tahap ini, hasil model dianalisis untuk mengidentifikasi faktor utama yang mempengaruhi kelulusan mahasiswa.

PERPUSTAKAAN
UNIVERSITAS JENDRAL ACHMAD YANI
YOGYAKARTA