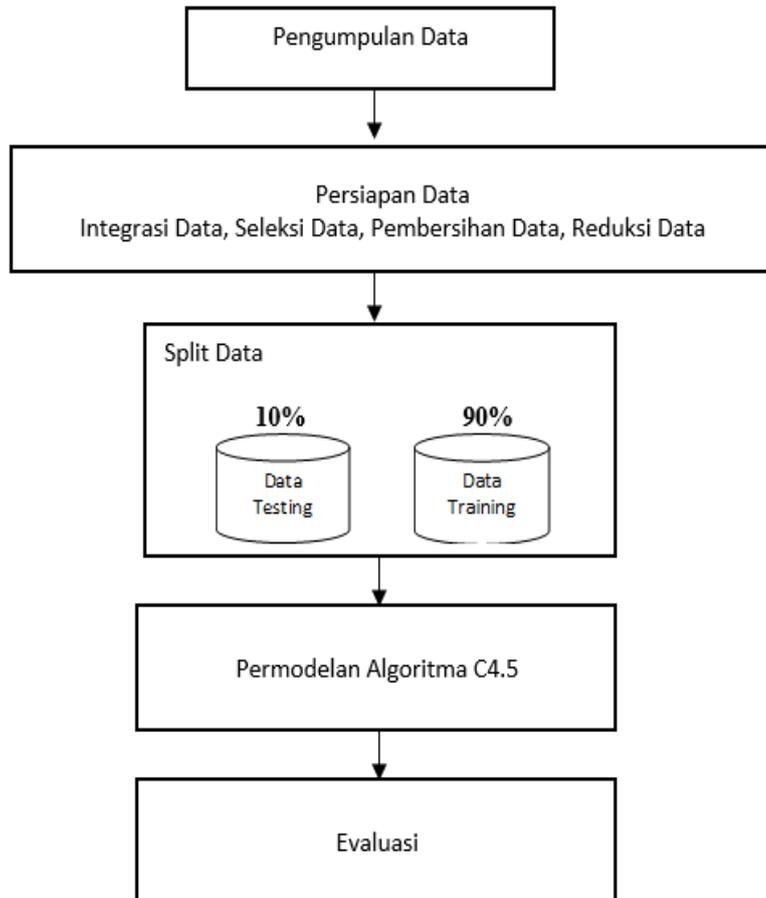


BAB 3

METODE PENELITIAN

Pada penelitian ini berfokus pada prediksi keterlambatan mahasiswa dalam membayar biaya menggunakan algoritma C4.5 dengan tahapan penelitian sebagai berikut.



Gambar 3.1 Tahapan Penelitian

3.1. Pengumpulan Data

Data yang diambil adalah data mahasiswa yang membayar biaya kuliah dari tahun 2019 s/d 2021. Data didapatkan pada bagian Biro Administrasi Keuangan dengan jumlah 3 atribut seperti nama, nim, dan keterangan. Data juga didapatkan pada bagian Biro Administrasi Akademik dengan jumlah 10 atribut seperti nama, nim, jenis kelamin, fakultas, prodi, angkatan, penghasilan ayah, pendidikan ayah, penghasilan ibu, dan pendidikan ibu.

Tabel 3.1 Atribut Data Mahasiswa dari Biro Administrasi Keuangan

Atribut	Penjelasan
Nama	Nama mahasiswa
NIM	Nomor Induk Mahasiswa
Keterangan	Keterangan mahasiswa terlambat atau tepat waktu dalam membayar biaya kuliah

Tabel 3.2 Atribut Data Mahasiswa dari Biro Administrasi Akademik

Atribut	Penjelasan
Nama	Nama Mahasiswa
NIM	Nomor Induk Mahasiswa
Jenis Kelamin	Jenis kelamin mahasiswa
Fakultas	Fakultas mahasiswa
Prodi	Program Studi Mahasiswa
Angkatan	Angkatan mahasiswa
Pendidikan Ayah	Pendidikan ayah mahasiswa
Penghasilan Ayah	Penghasilan ayah mahasiswa
Pendidikan Ibu	Pendidikan ibu mahasiswa
Penghasilan Ibu	Penghasilan ibu mahasiswa

3.2. Persiapan Data

Persiapan data dilakukan untuk memperbaiki masalah yang terdapat pada data dengan melakukan integrasi data, seleski data, pembersihan data, dan reduksi data.

1. Integrasi Data

Tahap integrasi dilakukan untuk data yang bersumber dari bagian Biro Administrasi Keuangan dan Biro Administrasi Akademik akan dilakukan penggabungan data sehingga menghasilkan sebuah data yang baru. Penggabungan data dari Biro Administrasi Keuangan menggunakan atribut nim untuk mengambil data mahasiswa yang berasal dari Biro Administrasi Akademik seperti jenis kelamin, fakultas, prodi, angkatan, penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu.

2. Seleksi Data

Tahap seleksi data dilakukan dengan membuang variabel yang tidak digunakan atau tidak relevan terhadap model yang akan dibangun untuk menghindari hasil yang tidak akurat.

3. Pembersihan Data

Tahap pembersihan data dilakukan untuk menghapus data yang tidak lengkap dan data yang error karena sebelum dianalisis harus bersih, lengkap (tidak mengandung *null*), dan tidak memiliki duplikasi agar dapat memberikan hasil yang akurat.

4. Reduksi Data

Tahap reduksi data dilakukan untuk menghadapi masalah pada himpunan data yang sangat besar karena memiliki kelas target yang tidak seimbang dilakukan dengan membuang sebagian kelas target sehingga dataset menjadi lebih kecil dengan tetap menjaga integritas data asli.

3.3. Split Data

Split data dilakukan dengan membagi *dataset* menjadi dua bagian untuk data *training* sebesar 90% digunakan dalam membangun model, kemudian untuk data *testing* sebesar 10% digunakan sebagai evaluasi model yang telah dibangun.

3.4. Permodelan Algoritma C4.5

Permodelan dilakukan dengan memilih model yang sesuai agar mendapatkan hasil yang optimal menggunakan algoritma C4.5 dengan membangun pohon keputusan sebagai berikut.

1. Mempersiapkan data *training* dengan mengambil data masa lalu yang pernah terjadi sebelumnya.
2. Menghitung nilai *entropy* menggunakan persamaan 2.2.
3. Menghitung nilai *gain* di setiap atribut menggunakan persamaan 2.1.
4. Ulangi langkah ke-2 hingga semua *record* terpartisi
5. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam *record* yang dipartisi lagi.

- c. Tidak ada *record* didalam cabang yang kosong.

3.5. Evaluasi

Evaluasi dilakukan untuk mendapatkan hasil yang sesuai terhadap model yang telah dibangun dengan teknik *confusion matrix* digunakan untuk mengukur keakuratan hasil serta kualitas data *training*. Pengukuran *confusion matrix* yang digunakan adalah akurasi menggunakan persamaan 2.3.