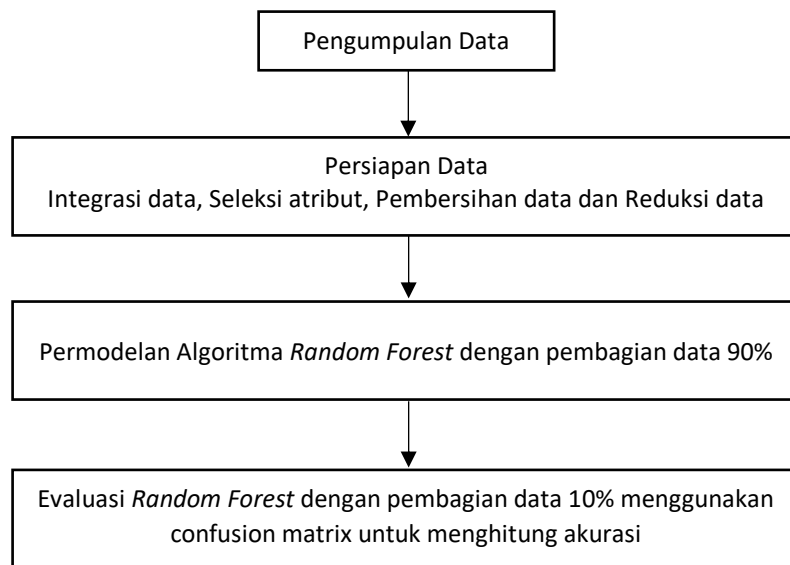


## BAB 3

### METODE PENELITIAN

#### 3.1 Tahapan Penelitian

Berikut adalah tahapan penelitian yang akan dilakukan pada penelitian ini yang terdiri dari 4 tahapan yang ditunjukkan pada gambar 3.1.



**Gambar 3.1 Tahapan Penelitian**

#### 3.2 Pengumpulan Data

Tahap awal dalam penelitian ini adalah pengumpulan data. Data yang dikumpulkan diperoleh dari dari Biro Administrasi Keuangan dan Bagian Administrasi Akademik pada tahun 2019-2021. Data dari Biro Administrasi Keuangan terdiri dari atribut nama, nim dan keterangan. Sedangkan data dari Biro Administrasi Akademik terdiri dari atribut nama, nim, gender, program studi, fakultas, angkatan, penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu.

**Tabel 3.1 Atribut Data Mahasiswa dari Biro Administrasi Keuangan**

Atribut	Keterangan
Nama	Nama Mahasiswa
NIM	Nomor Induk Mahasiswa

**Tabel 3.1 Atribut Data Mahasiswa dari Biro Administrasi Keuangan (Lanjutan)**

Atribut	Keterangan
Keterangan	Keterangan mahasiswa terlambat atau tepat waktu dalam membayar biaya kuliah

**Tabel 3.2 Atribut Data Mahasiswa dari Biro Administrasi Akademik**

Atribut	Keterangan
Nama	Nama Mahasiswa
NIM	Nomor Induk Mahasiswa
Gender	Jenis kelamin mahasiswa
Fakultas	Fakultas mahasiswa
Prodi	Prodi Mahasiswa
Angkatan	Angkatan mahasiswa
Penghasilan Ayah	Penghasilan ayah mahasiswa
Penghasilan Ibu	Penghasilan ibu mahasiswa
Pendidikan Ayah	Pendidikan Ayah
Pendidikan Ibu	Pendidikan Ibu

### **3.3 Persiapan Data**

Pada tahap ini akan dilakukan proses integrasi data, seleksi data, pembersihan dan reduksi data yang nantinya akan diperlukan dalam tahap permodelan dan evaluasi. Berikut ini adalah proses persiapan data yang akan dilakukan yaitu:

#### **1) Integrasi data**

Pada tahap ini integrasi data dilakukan untuk menggabungkan 2 buah data dari Biro Administrasi Keuangan dan Administrasi Akademik. Data digabungkan dengan cara menyamakan nim dari data Biro Administrasi Keuangan dan Administrasi Akademik.

## 2) Seleksi atribut

Pada tahap ini seleksi data dilakukan untuk melakukan pemilihan atribut relevan yang akan digunakan dalam penelitian, atribut yang tidak relevan akan dihapus dan tidak akan digunakan. Pada penelitian ini seleksi atribut menggunakan *information gain ratio*.

## 3) Pembersihan data

Pada tahap ini pembersihan data dilakukan dengan cara menghapus data yang tidak lengkap atau kosong, menghilangkan noise dan data yang duplikat, sehingga data tersebut dapat diolah untuk melakukan proses data mining. Setelah semua data yang dibutuhkan telah melalui tahap pembersihan data maka data akan disimpan dalam dataset baru.

## 4) Reduksi data

Pada tahap ini reduksi data dilakukan dengan cara pengurangan data terhadap label kelas yang datanya tidak seimbang, dimana kelas tepat lebih banyak daripada kelas terlambat, sehingga perlu dilakukan reduksi dimensi data agar tidak terjadi klasifikasi lebih condong ke kelas mayoritas dibandingkan dengan kelas minoritas menggunakan teknik *random under sampling*.

### 3.4 Permodelan

Adapun tahapan-tahapan permodelan yang dilakukan pada penelitian ini sebagai berikut:

- 1) Menyiapkan dataset.
- 2) Menghitung nilai *entropy* menggunakan rumus seperti pada persamaan 2.1.
- 3) Menentukan akar dari pohon dengan cara menghitung nilai *gain* tertinggi dari masing-masing atribut menggunakan rumus seperti pada persamaan 2.2.
- 4) Ulangi langkah 2 dan 3 sebanyak k kali, sehingga terbentuknya sebuah hutan yang terdiri atas k pohon.
- 5) Melakukan perhitungan suara terbanyak untuk setiap target yang akan diprediksi pada setiap *tree*. Hasil kelas yang paling banyak akan menjadi hasil

akhir dari algoritma *random forest*.

### **3.5 Evaluasi**

Setelah fase modeling selesai, model tersebut perlu dilakukannya evaluasi sebelum menggunakannya. Untuk mengukur evaluasi dari model yang dihasilkan menggunakan akurasi. Rumus untuk menghitung akurasi seperti pada persamaan 2.6.