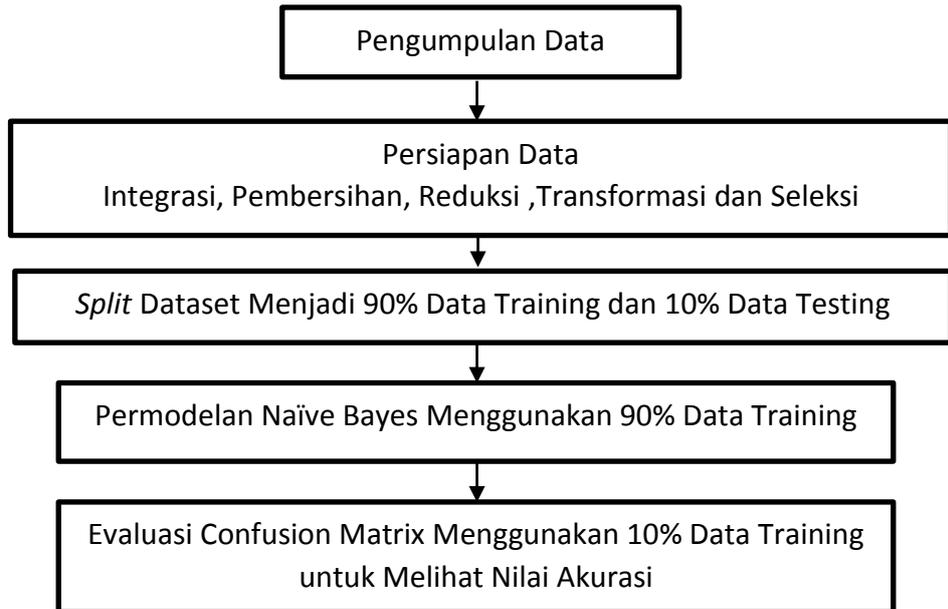


BAB 3

METODE PENELITIAN

3.1. Tahapan Penelitian

Penelitian ini akan menggunakan 5 tahapan, adapun tahapannya sebagai berikut :



Gambar 3.1 Tahapan Penelitian

3.2. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data keterlambatan pembayaran biaya kuliah yang bersumber dari Biro Administrasi Keuangan dan data Bagian Administrasi Akademik pada tahun 2019 sampai 2021. Data tersebut memiliki beberapa atribut seperti nama, NIM, jenis kelamin, program studi, fakultas, angkatan, penghasilan ayah, pendidikan ayah, pendidikan ibu, penghasilan ibu dan Label.

Tabel 3. 1 Atribut Data Mahasiswa Dari Biro Administrasi Keuangan

Atribut	Keterangan
Nama	Nama mahasiswa
NIM	Nomor Induk Mahasiswa
Label	Label class target tepat atau terlambat

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
Program Studi	Program Studi Mahasiswa
Angkatan	Angkatan
Penghasilan Ayah	Penghasilan Ayah Mahasiswa
Penghasilan Ibu	Penghasilan Ibu Mahasiswa
Pendidikan Ayah	Pendidikan Terakhir Ayah Mahasiswa
Pendidikan Ibu	Pendidikan Terakhir Ibu Mahasiswa

3.3. Persiapan Data

Pada tahapan ini terdapat beberapa persiapan seperti Integrasi data, proses pembersihan data, reduksi data, seleksi data serta transformasi data untuk mendapatkan data yang berkualitas agar dapat mempermudah dalam proses *data mining* dan dijadikan masukan dalam tahap pemodelan (*modeling*). Proses *preparation* ini mencakup lima hal utama yaitu :

1) Integrasi data

Tahap integrasi data dilakukan dengan menggabungkan data dari Bagian Administrasi keuangan yang memiliki 3 atribut seperti nama, NIM, Label. kemudian digabungkan dengan data Bagian Administrasi Akademik sebagai penunjang data, adapun data yang diambil seperti fakultas, prodi, angkatan, gender, penghasilan ayah, penghasilan ibu, pendidikan ayah, pendidikan ibu. Proses penggabungan disesuaikan dengan NIM dari data Bagian administrasi akademik dan Bagian Administrasi Keuangan.

2) Pembersihan data

Proses pembersihan data dilakukan dengan membersihkan data yang tidak memiliki nilai atribut lengkap (*missing value*), terdapat data yang ganjil (*noise*), dan data yang terduplikat agar tidak mengganggu proses kerja algoritma sehingga memberikan hasil akurasi yang baik.

3) Reduksi data

Pada tahap ini dilakukan reduksi data agar dataset yang mulanya memiliki label target tepat sebanyak 23.341 lebih banyak ketimbang kelas target terlambat sebanyak 6.204, karena akan menyebabkan permodelan menjadi tidak seimbang. Sehingga dilakukan pengambilan terhadap dataset yang memiliki label target tepat secara random menjadi 6.204 agar seimbang dengan label kelas target terlambat sebanyak 6.204.

4) Seleksi data

Pada tahap ini akan dilakukan pembuangan atribut-atribut yang tidak digunakan seperti atribut nama dan NIM, sehingga atribut yang akan digunakan lebih informatif dan dapat meningkatkan akurasi serta efisien terhadap algoritma *Naïve Bayes*.

5) Transformasi Data

Pada tahap ini, data-data yang bertipe kategorial akan dilakukan transformasi menjadi tipe data numerik agar dapat digunakan pada proses perhitungan yang dilakukan algoritma *Gaussian Naïve Bayes*. Adapun metode yang digunakan untuk transformasi adalah metode *One-Hot Encoding*.

Tabel 3. 3 Atribut Setelah di seleksi

Atribut	Keterangan
Fakultas	Fakultas
Program Studi	Program Studi Mahasiswa
Angkatan	Angkatan
Gender	Jenis Kelamin
Penghasilan Ayah	Penghasilan Ayah Mahasiswa
Penghasilan Ibu	Penghasilan Ibu Mahasiswa
Pendidikan Ayah	Pendidikan Terakhir Ayah Mahasiswa
Pendidikan Ibu	Pendidikan Terakhir Ibu Mahasiswa
Label	Tepat atau terlambat

3.4. Split Data

Penelitian ini menggunakan Split data agar mendapatkan hasil estimasi akurasi yang baik dari proses dataset. Dataset yang digunakan kemudian dibagi menjadi 90% data *training* dan 10% data *testing*.

3.5. Permodelan

Pada tahap ini dilakukan pemilihan model yang sesuai agar dapat mengoptimalkan hasil, adapun algoritma yang digunakan dalam pemodelan ini adalah algoritma *Naïve Bayes*. langkah – langkah permodelan sebagai berikut:

1. Membaca *data training*
2. Menghitung jumlah kelas target pada data training
3. perhitungan menggunakan data numerik

langkah awal dalam perhitungan data numerik yaitu mencari nilai *mean* dan nilai standar deviasi dari masing-masing atribut yang menggambarkan data angka. Adapun rumus yang digunakan untuk menghitung nilai *mean* seperti pada persamaan 2.4. Langkah berikutnya menghitung nilai standar deviasi, adapun rumusnya dapat dilihat pada persamaan 2.5

4. Nilai Distribusi *Gaussian*

Selanjutnya menghitung nilai probabilitas untuk fitur *Data Testing* yang memiliki data numerik menggunakan perhitungan seperti pada persamaan 2.6.

5. Menentukan Nilai Akhir

Setelah mendapatkan nilai distribusi Gaussian, langkah selanjutnya yaitu mengkalikan semua nilai *Gaussian* yang memiliki label tepat dan terlambat.

$$P(\text{Kelas} | X) = P(\text{Kelas}) \times P(X) \quad (3.1)$$

Kemudian setelah mendapatkan nilai akhir, langkah selanjutnya adalah membandingkan nilai antara Probabilitas keterangan “Tepat” dan Probabilitas keterangan “Terlambat”. Maka dihasilkan nilai prediksi dari perbandingan tersebut.

3.6. Evaluasi

Evaluasi merupakan tahapan yang digunakan untuk membantu pengukuran pada algoritma. Penulis menggunakan *Confusion Matrix* untuk melakukan pengukuran evaluasi pada model. Adapun rumus menghitung *accuracy* ditunjukkan pada persamaan 2.7.