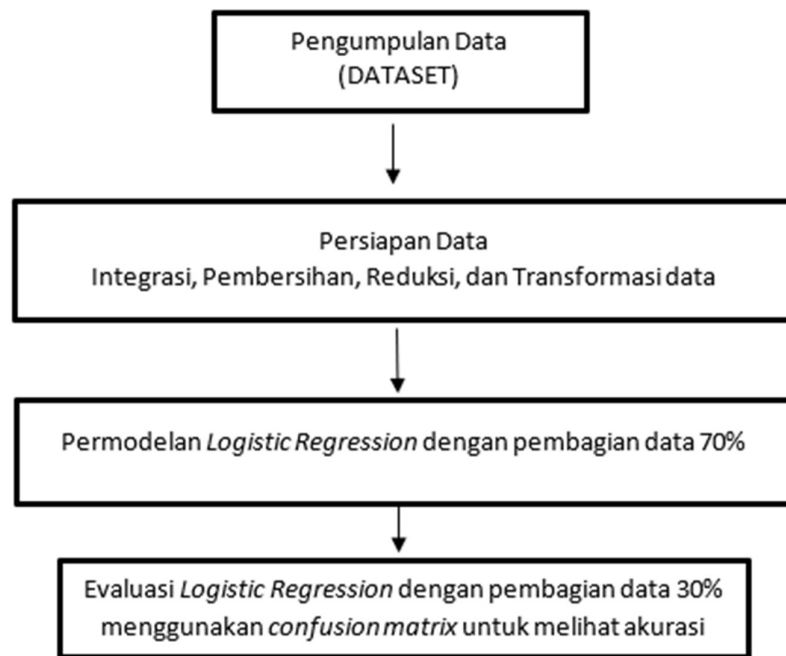


BAB 3

METODE PENELITIAN

3.1 Tahapan Penelitian

Dalam menyelesaikan penelitian dibutuhkan sebuah tahapan yang akan digunakan untuk penelitian sehingga dapat dilakukan secara konsisten. Pada penelitian ini metode yang akan digunakan yaitu *Logistic Regression*. Tahapannya bisa dilihat pada gambar 3.1.



Gambar 3.1 Tahapan Analisis Data

3.2 Pengumpulan Data (Dataset)

Pada tahap ini melakukan pengumpulan data. Data yang akan digunakan dalam penelitian ini adalah data dari Biro Administrasi Akademik dan Biro Administrasi Keuangan pada tahun 2019-2021. Data yang didapat tersebut terdapat beberapa atribut seperti nama, nim, angkatan, gender, program studi, fakultas, penghasilan ayah, pendidikan ayah, penghasilan ibu, pendidikan ibu, dan keterangan (tepat atau terlambat). Dapat dilihat pada tabel 3.1 dan tabel 3.2.

Tabel 3.1 Atribut Data Mahasiswa Dari Biro Administrasi Keuangan

Atribut	Keterangan
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	Keterangan
Nama	Nama Mahasiswa
Nim	Nomor Induk Mahasiswa
Program Studi	Program Studi Mahasiswa
Gender	Jenis Kelamin Mahasiswa
Fakultas	Fakultas Mahasiswa
Angkatan	Angkatan Mahasiswa
Pendidikan Ayah	Pendidikan Ayah Mahasiswa
Penghasilan Ayah	Penghasilan Ayah Mahasiswa
Pendidikan Ibu	Pendidikan Ibu Mahasiswa
Penghasilan Ibu	Penghasilan Ibu Mahasiswa

3.3 Persiapan Data

Tahap ini terdiri dari Integrasi data, proses pembersihan data, reduksi data, transformasi data serta seleksi data agar mendapatkan data yang berkualitas untuk dilanjutkan ke dalam tahap Permodelan. Proses ini mempunyai 5 tahapan yaitu :

1) Integrasi Data

Tahap ini dilakukan penggabungan atribut data yang telah dikumpulkan dengan cara mencocokkan atribut seperti nama dan nim dari Bagian Administrasi Keuangan dan Bagian Administrasi Akademik.

2) Pembersihan Data

Pada tahap ini dilakukannya suatu proses analisa mengenai kualitas dari data keterlambatan. Maka dilakukan penghapusan pada data yang tidak memiliki nilai atribut lengkap (*missing value*), terdapat kesalahan dalam pengisian data (*noise*), dan data yang terduplikat agar tidak mengganggu proses kerja pada algoritma.

3) Reduksi Data

Tahap selanjutnya dilakukan reduksi data yang mana pada dataset yang memiliki label kelas target tepat lebih banyak daripada kelas target terlambat, dikarenakan dapat menyebabkan permodelan menjadi tidak seimbang. Maka diperlukan reduksi data dengan menggunakan teknik *undersampling*.

4) Transformasi Data

Tahap ini dilakukan perubahan data pada atribut yang bertipe kategori menjadi numerik agar memudahkan algoritma *Logistic Regression* pada saat perhitungan. Contoh dari perubahan data tersebut yaitu mengubah setiap nilai dalam kolom menjadi angka yang berurutan seperti 1,2,3,4, dan seterusnya sesuai pada abjad data tersebut.

3.4 Pembagian Data

Sebelum dilakukannya permodelan, data yang telah di persiapkan dibagi menjadi 70% data training dan 30% data testing agar mendapatkan hasil perkiraan akurasi yang baik ketika dilakukan permodelan.

3.5 Modeling

Tahapan ini memilih model yang sesuai untuk dibangun agar mendapatkan hasil yang optimal dengan menggunakan algoritma *Logistic Regression* untuk menentukan model yang akan dibangun. Berikut adalah tahapan dalam pembuatan model untuk prediksi dengan rumus :

$$p(Y = 1) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (3.1)$$

Atau

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \quad (3.2)$$

Rumus di atas adalah rumus yang dapat digunakan dalam modeling untuk mencari nilai probabilitas dimana P adalah probabilitas kemudian e^x atau ditulis "e" adalah fungsi eksponen.

- 1) Membuat model dengan mengimpor paket yang diperlukan, yaitu *Logistic Regression* dalam `sklearn.linear_model`.
- 2) Kemudian untuk melatih model, fungsi `fit()` di panggil dengan memasukkan training dataset yang telah disiapkan.
- 3) Setelah itu algoritma machine learning sudah menemukan koefisien untuk masing-masing fitur pada training dataset.

3.6 Evaluasi

Pada tahap ini dilakukan pengukuran akurasi untuk mendapatkan hasil yang sesuai kepada model yang telah dibangun berdasarkan algoritma *Logistic Regression* dengan teknik *confusion matrix* guna mengukur seberapa akurat hasil serta kualitas dari *data training*. Berikut tabel 3.3 dan rumus dalam mengukur tingkat akurasi menggunakan *confusion matrix*:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.3)$$

Nilai akurasi dalam evaluasi pada *confusion matrix* menunjukkan seberapa akurat model dalam melakukan klasifikasi.