

## BAB 3

### METODOLOGI PENELITIAN

#### 3.1. Tahapan Penelitian

Tahapan ini merupakan acuan dalam melaksanakan proses penelitian agar lebih tertata dan konsisten dimana alurnya tertera pada gambar 3.1.



**Gambar 3. 1 Tahapan Penelitian**

#### 3.2. Identifikasi Masalah

Tahap ini merupakan identifikasi permasalahan tentang jumlah mahasiswa yang terlambat membayar SPP di Universitas Muhammadiyah Kalimantan Timur dan telah melebihi jumlah mahasiswa tepat waktu dalam membayar SPP. Sehingga diperlukan data analisis untuk prediksi agar dapat mengatasi terjadinya keterlambatan pembayaran biaya kuliah di Universitas Muhammadiyah Kalimantan Timur.

### 3.3. Pengumpulan Data

Tahap ini adalah memahami data mulai dari pengumpulan dilanjutkan dengan proses untuk memperoleh pemahaman data secara mendalam yang akan digunakan untuk melakukan prediksi dengan tahapannya sebagai berikut :

a) Pengumpulan data awal

Data yang diperoleh merupakan data dari Bagian Admisnistrasi Akademik dan data keterlambatan pembayaran SPP pada periode tahun 2019 - 2021.

b) Mendeskripsikan data

Berikut deskripsi dari sampel atribut data yang akan di gunakan dalam membangun model *data mining* yang tertera pada tabel 3.1.

**Tabel 3. 1 Atribut Data**

No	Atribut	Proses	Keterangan
1.	Nama	Data Cleaning	Nama Mahasiswa
2.	NIM	Data Cleaning	NIM Mahasiswa
3.	Jenis Kelamin	Data Cleaning	Jenis Kelamin Mahasiswa
4.	Fakultas	Data Cleaning	Fakultas
5.	Program Studi	Data Cleaning	Program Studi Mahasiswa
6.	Angkatan	Data Cleaning	Angkatan
7.	Pendidikan Ayah	Digunakan Sebagai Atribut	Pendidikan Ayah
8.	Penghasilan Ayah	Digunakan Sebagai Atribut	Penghasilan Ayah
9.	Pendidikan Ibu	Digunakan Sebagai Atribut	Pendidikan Ibu
10.	Penghasilan Ibu	Digunakan Sebagai Atribut	Penghasilan Ibu
11	Label	Digunakan Sebagai Label Taerget	Ketentuan Yang Telah Ditetapkan Kampus

### 3.4. Persiapan Data

Tahap ini mencakup integrasi, pembersihan, seleksi, reduksi, dan transformasi agar dapat dilanjutkan ke dalam tahap pemodelan (*modeling*). Data yang digunakan merupakan data dari Bagian Administrasi Akademik dan Bagian Biro Keuangan UMKT periode 2019 – 2021 dengan proses persiapannya sebagai berikut :

a) Integrasi data

Tahap integrasi data dilakukan untuk menggabungkan data dari Bagian keuangan dan Bagian Administrasi Akademik. Data yang didapatkan dari bagian keuangan memiliki atribut yaitu nim, nama, label kemudian data yang didapatkan dari Bagian Administrasi Akademik memiliki atribut yaitu nama, nim, fakultas, program studi, angkatan, pendidikan ayah, penghasilan ayah, pendidikan ibu, penghasilan ibu. Data digabungkan dengan cara melihat kecocokan nim sehingga atribut keseluruhan menjadi nama, nim, fakultas, program studi, angkatan, pendidikan ayah, penghasilan ayah, pendidikan ibu, penghasilan ibu dan label.

b) Pembersihan data

Tahap pembersihan data dilakukan agar menghilangkan data ganjil (*noise*), tidak konsisten, dan data yang kosong (*missing value*). Tahap ini dilakukan agar data yang digunakan pada tahap *modeling* benar - benar akurat.

c) Seleksi data dan reduksi data

Tahap seleksi data dilakukan terhadap atribut dan data yang telah dikumpulkan agar dapat meningkatkan akurasi dan efisiensi dari algoritma. Atribut data yang di pilih agar mendapatkan kesesuaian dengan algoritma yang akan di gunakan. Kemudian mereduksi data dimana dataset yang memiliki kelas target paling banyak (mayoritas) akan menyebabkan pemodelan menjadi bias terhadap kelas target minoritas. Untuk menghindari hal tersebut maka dilakukan pengurangan dataset terhadap jumlah kelas target mayoritas dengan teknik undersampling.

**Tabel 3. 2 Atribut Data Yang Dipilih**

No	Atribut	Keterangan
1.	Pendidikan Ayah	Pendidikan Ayah
2.	Penghasilan Ayah	Penghasilan Ayah
3.	Pendidikan Ibu	Pendidikan Ibu
4.	Penghasilan Ibu	Penghasilan Ibu
5.	Label	Ketentuan Yang Telah Ditetapkan Kampus

#### d) Transformasi data

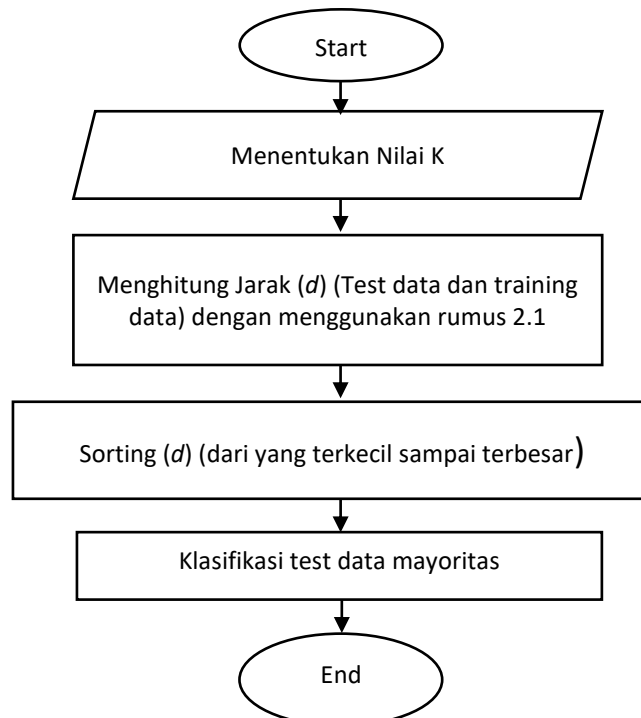
Tahap ini dilakukan untuk merubah semua tipe data menjadi numerik agar algoritma *K-Nearest Neighbor* dapat melakukan proses perhitungan. Teknik yang digunakan adalah *ordinal encoding* dengan cara data ordinal yang memiliki skala paling kecil diubah ke numerik dengan bilangan 0 hingga N sebanyak kategori per fitur.

### 3.5. Membagi Data

Tahap membagi data yaitu membagi dataset yang telah diolah pada proses persiapan data dengan pembagian 80:20. Pembagian tersebut terbagi menjadi 80% *data training* dan menjadi 20% *data testing*.

### 3.6. Pemodelan Dan Evaluasi

Tahap pemodelan merupakan bagian dari menemukan informasi atau pola menarik dari suatu data yang dipilih berdasarkan proses dari persiapan data. Teknik yang digunakan untuk *modeling* prediksi adalah menggunakan algoritma *K-Nearest Neighbor* dimana teknik ini mempunyai atribut yang di inialisasi menjadi nilai  $k$ , yaitu jumlah tetangga terdekat yang dijadikan acuan pada algoritma dengan prosesnya tertera pada gambar 3.2



**Gambar 3. 2 Diagram Alur kerja *K-Nearest Neighbor***

Setelah dilakukan tahap pemodelan kemudian melakukan evaluasi untuk melihat tingkat performa dari pola yang di hasilkan algoritma. Parameter yang digunakan untuk evaluasi algoritma adalah *confusion matrix* dengan melihat nilai akurasi, nilai tersebut dapat diperoleh melalui perhitungan pada rumus 2.2.