

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Jenis dan Sumber Data**

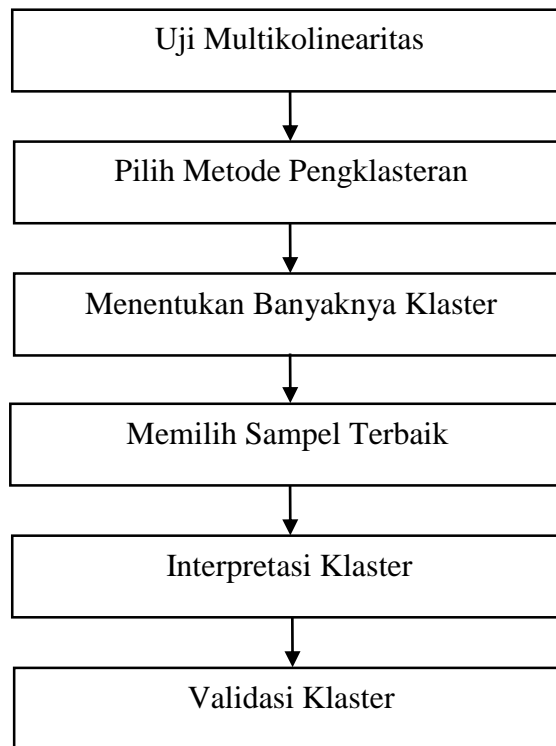
Jenis data yang digunakan dalam penelitian ini adalah data multivariat dengan masing-masing data memiliki tiga variabel atau biasa disebut sebagai data trivariat. Data yang dipakai adalah data simulasi yang mana semua data yang akan diuji dalam penelitian ini berasal dari kebijakan atau subjektivitas penulis sendiri.

#### **3.2 Analisis Data**

Pada studi kasus ini, pengolahan data dilakukan dengan menggunakan bantuan *software R (3.4.2)* yang sudah dilengkapi dengan *packages cluster*. *Software* ini adalah salah satu *software* statistik terbaik dalam mengklasterkan suatu data atau objek.

#### **3.3 Langkah-Langkah Penelitian**

Pada proses analisis kluster ini, terlebih dahulu dilakukan pemilihan metode pengklasteran yaitu metode CLARA (*Clustering Large Applications*) karena menggunakan data atau objek yang besar lebih dari seratus buah objek dan metode pengklasterannya dilakukan secara sampling acak sehingga hasil pengklasterannya akan diperoleh secara cepat dan tepat. Hal terpenting dalam analisis kluster adalah menentukan jumlah kluster. Dalam menentukan banyaknya kluster yang terbentuk dari masing-masing metode bergantung pada subjektivitas peneliti. Jadi penentuan banyaknya kluster pada data simulasi trivariat ini akan diambil sebanyak tiga kluster oleh peneliti. Karena pada metode CLARA tahan terhadap pencilan maka uji pencilan tidak dilakukan. Dalam penelitian ini, tahapan pengolahan data dapat dilihat pada gambar di bawah ini:



**Gambar 4.1 Tahap Pengolahan Data**

### 3.3.1 Uji Multikolinearitas

Sebelum mengolah data ke dalam metode CLARA (*Clustering Large Application*) akan dicari terlebih dahulu nilai korelasi antar variabel tersebut. Ketika diperoleh nilai korelasi yang tinggi antar variabel ini menandakan adanya multikolinearitas pada data tersebut maka untuk menanggulangnya akan ditransformasi data tersebut ke dalam nilai *z-score* menggunakan analisis komponen utama (AKU).

### 3.3.2 Pilih Metode Pengklasteran

Karena himpunan data yang digunakan pada penelitian ini adalah himpunan data yang besar, maka akan diolah dengan metode CLARA (*Clustering Large Application*) dalam proses pengelompokannya karena metode ini berbasis sampling atau sampelnya akan diolah secara acak. Penghitungan untuk mencari medoidnya menggunakan metode PAM

(*Partitioning Around Medoids*) di mana objek-objek yang dipilih menjadi pusat kluster (medoid) cenderung mirip dengan seluruh himpunan data.

### 3.3.3 Menentukan Banyaknya Kluster

Berdasarkan pemaparan pada metode pengklasteran, peneliti akan mengelompokkan himpunan data yang besar tersebut ke dalam 3 buah kelompok (kluster).

### 3.3.4 Memilih Sampel Terbaik

Sampel terbaik yang terpilih berdasarkan nilai jarak (*dissimilarities*) rata-rata dari seluruh data objek yang paling kecil. Karena semakin kecil jarak yang terpilih pada sampel tersebut maka semakin tinggi kemiripan karakteristik antar objek dalam satu kluster dan semakin tinggi juga perbedaan karakteristik objek antar kluster.

### 3.3.5 Interpretasi Kluster

Pada tahap ini, hasil dari sampel terbaik tersebut akan diinterpretasikan, di sini akan dipaparkan ukuran dari tiap-tiap kluster, nilai medoidnya, jarak antar klasternya, jarak objek dalam satu kluster, dan penempatan himpunan data besar ke dalam tiga kluster tersebut.

### 3.3.6 Validasi Kluster

Proses validasi ini bertujuan untuk mengukur seberapa akurat pengklasteran yang telah dilakukan dari tahap sebelumnya. Pada tahap ini akan diperoleh nilai *silhouette coefficient*nya (nilai rata-rata *silhouette width* tiap-tiap kluster) di mana semakin tinggi nilai *silhouette coefficient*nya maka semakin baik atau akurat pengelompokan data yang telah dibuat tersebut begitu juga sebaliknya. Yang artinya, tingkat kemiripan antar objek dalam satu kluster sangat tinggi atau sangat dekat dan tingkat kemiripan antar kluster sangat kecil atau sangat jauh. Nilai *silhouette coefficient* berada antara 0 sampai 1.