

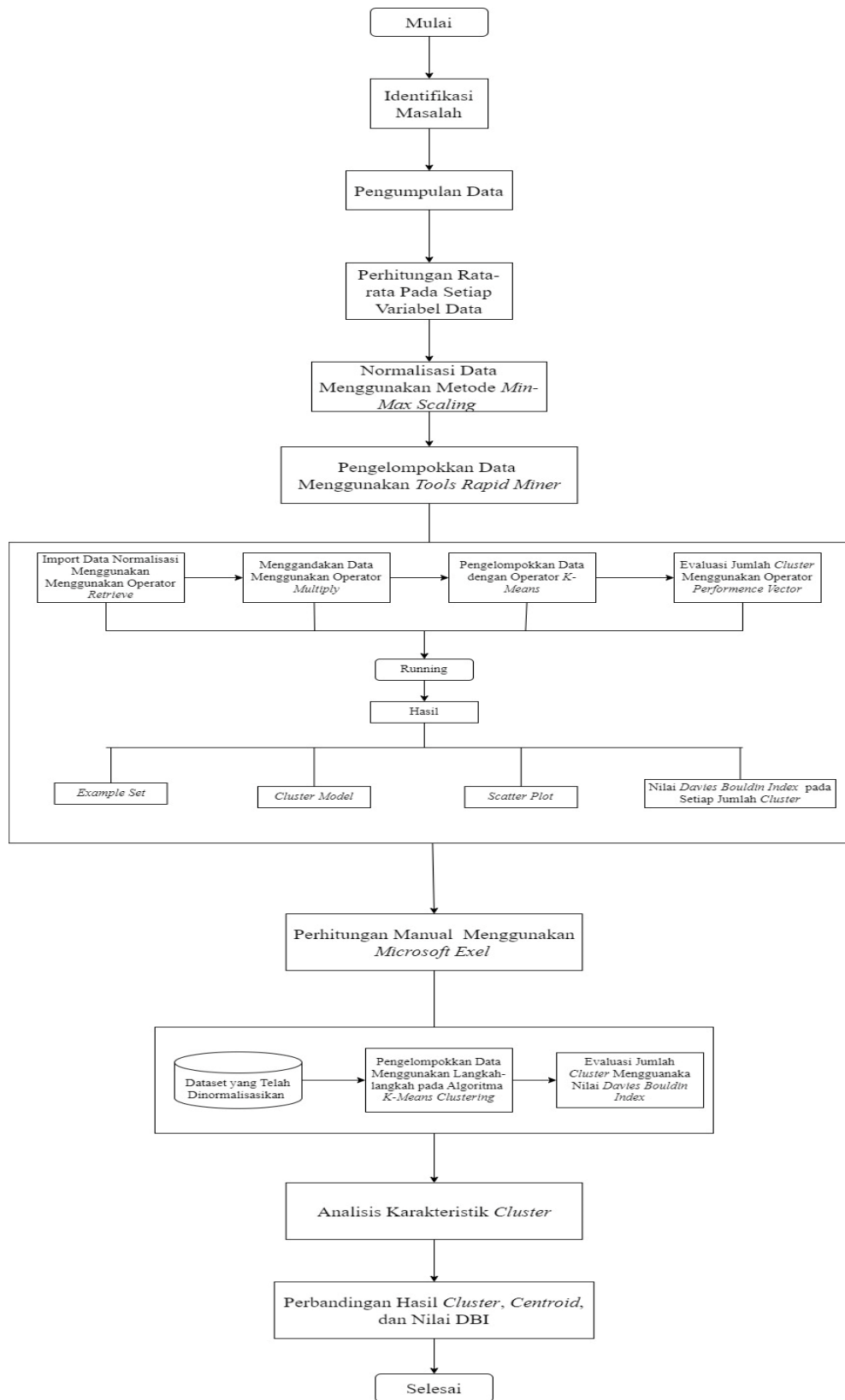
BAB III

METODE PENELITIAN

3.1 Tahapan Metode Penelitian

Manfaat dari metode penelitian adalah memberikan arahan dan struktur untuk mengumpulkan, menganalisis, dan menyajikan data dengan tujuan memperoleh pemahaman yang lebih mendalam tentang suatu permasalahan yang diteliti. Pada penelitian yang dilakukan, peneliti menemukan masalah yang terjadi pada kabupaten/kota di Provinsi Nusa Tenggara Timur. Masalah tersebut mencakup pembangunan sosial-ekonomi yang tidak merata sehingga mengakibatkan kemiskinan dan kesenjangan sosial-ekonomi antara kabupaten/kota. Dalam upaya mengatasi masalah tersebut, peneliti menggunakan algoritma *k-means*, untuk mengelompokkan kabupaten/kota ke dalam kelompok yang memiliki karakteristik pembangunan sosial-ekonomi yang sama atau serupa. Untuk mendapatkan pengelompokan tersebut, peneliti menggunakan 4 variabel data yang merepresentasikan pembangunan sosial-ekonomi dalam suatu kabupaten/kota. Keempat variabel data tersebut akan dianalisis dan dikelompokkan menggunakan *tools Rapid Miner* dan perhitungan manual menggunakan *microsoft excel*.

Metode penelitian yang lebih rinci akan dipaparkan sebagai solusi terhadap permasalahan yang menjadi fokus penelitian. Metode penelitian dapat dilihat pada Gambar 3.1 (Solichin & Khairunnisa, 2020).



Gambar 3.1 Alur Penelitian

3.1.1 Identifikasi Masalah

Dalam penelitian yang dilakukan, masalah difokuskan pada pembangunan sosial-ekonomi di kabupaten/kota Provinsi NTT yang belum merata sehingga mengakibatkan kemiskinan dan terjadinya kesenjangan sosial-ekonomi antara kabupaten/kota. Maka diperlukan pengelompokan kabupaten/kota berdasarkan data kemiskinan di Provinsi NTT menggunakan algoritma *k-means* agar alokasi sumber daya pembangunan sosial-ekonomi lebih merata ke kabupaten/kota yang membutuhkan.

3.1.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data yang diperoleh dari Kantor Badan Pusat Statistik Provinsi NTT. Untuk mendapatkan hasil pengelompokan data berdasarkan karakteristik pembangunan sosial-ekonomi, penelitian ini menggunakan 4 variabel data dari tahun 2017 sampai tahun 2023 (Riadi, 2023). Berikut ini dipaparkan 4 variabel data

1. Data jumlah penduduk miskin.
2. Data rata-rata lamanya sekolah.
3. Data rata-rata pengeluaran perkapita sebulan makanan dan *non* makanan.
4. Data persentase penduduk yang sulit mengakses layanan kesehatan.

3.1.3 Perhitungan Rata-rata pada Setiap Variabel Data

Pada tahap ini, data yang telah terkumpul akan diolah terlebih dahulu menggunakan *microsoft excel* dengan menghitung rata-ratanya pada keempat variabel data dari tahun 2017 sampai tahun 2023. Hal ini bertujuan untuk menganalisis karakteristik dari hasil *cluster*.

3.1.4 Normalisasi Data

K-Means bekerja dengan mengelompokkan data ke dalam *cluster* berdasarkan jarak antara titik data. Normalisasi membantu memastikan bahwa setiap data memiliki nilai yang seimbang terhadap pengukuran jarak. Pada penelitian ini, ada 4 variabel data yang digunakan dengan bentuk nilai yang berbeda. Oleh karena itu, Sebelum melakukan pengelompokkan data menggunakan algoritma *k-means*, diperlukan normalisasi data menggunakan *microsoft excel* untuk mencegah variabel data dengan rentang nilai yang besar mendominasi proses *clustering*. Normalisasi data menggunakan metode *min-max scaling* menggunakan Persamaan 3.1.

$$NB = \frac{NL - N_{Min \text{ dalam data}}}{N_{Max \text{ dalam data}} - \text{Nilai}_{Min \text{ dalam data}}} \dots\dots\dots(3.1)$$

Keterangan:

NB = Nilai Baru

NL = Nilai Lama

N_{Min} = Nilai minimum dalam data

N_{Max} = Nilai maksimal dalam data

3.1.5 Pengelompokkan Data Menggunakan *Tools Rapid Miner*

Pengelompokkan data dilakukan menggunakan algoritma *k-means* di dalam *tools Rapid Miner*. Berikut ini adalah tahapan-tahapan yang terlibat dalam proses pengelompokkan data pada *tools Rapid Miner*:

1. *Import* data yang telah dinormalisasikan di dalam *microsoft excel* menggunakan metode *min-max scaling* ke dalam *tools Rapid Miner*.
Import data dilakukan dengan menggunakan operator *retrieve*.
2. Menyambungkan *retrieve* dengan operator *multiply* untuk menggandakan *dataset*.
3. Proses pengelompokkan data menggunakan operator *k-means* pada *tools Rapid Miner*.
4. Evaluasi jumlah *cluster* menggunakan operator *performance vector* untuk menentukan jumlah *cluster* paling optimal yang dilihat dari nilai DBI paling rendah pada setiap jumlah *cluster*.
5. Setelah melakukan tahapan dengan menyambungkan berbagai operator, langkah berikutnya yang dilakukan adalah melakukan proses *running*, setelah itu *tools Rapid Miner* akan menyajikan informasi hasil *cluster* dalam bentuk *example set*, *cluster model*, grafik *scatter plot*. Setelah itu *Rapid Miner* akan menampilkan nilai *Davies Bouldin Index* (DBI) dari hasil penggunaan operator *performance vector*. Jumlah *cluster* dengan nilai DBI yang kecil merupakan jumlah *cluster* yang paling optimal dalam pengelompokkan data.

3.1.6 Perhitungan Manual Menggunakan *Microsoft Excel*

Pada tahap ini akan dilakukan perhitungan manual menggunakan *microsoft excel* dengan rumus *ecludience distance* pada algoritma *k-means*. Berikut ini adalah tahapan-tahapan yang terlibat dalam proses pengolahan data dan pengelompokkan data pada perhitungan manual:

1. Data yang telah dihitung rata-ratanya akan dinormalisasi menggunakan metode *min-max scaling*.
2. Data yang telah dinormalisasikan akan dikelompokkan menggunakan langkah-langkah pada algoritma *k-means*. Langkah-langkahnya sebagai berikut:
 - a) Tentukan jumlah *cluster* (*k*) dalam *dataset*.
 - b) Selanjutnya, nilai *centroid* ditentukan secara acak pada tahap awal, sedangkan pada tahap iterasi, digunakan rumus yang tercantum dalam Persamaan 3.2 (Margolang dkk., 2019).

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_k \dots\dots\dots(3.2)$$

Keterangan:

V_{ij} = rata-rata pusat (*cluster* ke-i) untuk variabel ke-j

N_i = jumlah anggota dalam kelompok pertama (*cluster* ke-i)

k = indeks yang merujuk ke *cluster*

j = indeks yang merujuk ke variabel

X_{kj} = nilai variabel ke-j pada data ke-k dalam kelompok tersebut

c) Tentukan Jarak Setiap Data ke Pusat *Cluster*.

Mengukur jarak antara setiap data dengan *centroid* menggunakan rumus *euclidean distance* (D) sebagaimana dijelaskan dalam Persamaan 3.3 (Solichin & Khairunnisa, 2020).

$$D_{(i,j)} = \sqrt{\sum_{k=1}^n (X_{ik} - C_{jk})^2} \dots\dots\dots(3.3)$$

Keterangan:

D = jarak *cluster*

X_{ik} = nilai data (i,k)

C_{jk} = nilai *centroid* (j,k)

n = jumlah *cluster*

d) Pengelompokan data berdasarkan *cluster* terdekat atau kelompokkan data berdasarkan jarak terdekat ke *centroid*. Mengamati *cluster* yang memiliki jarak terdekat dengan *centroid*, lalu masukkan data ke dalam *cluster* yang sesuai.

e) Bandingkan *cluster* baru dengan *cluster* sebelumnya. Apabila ada perubahan dalam *cluster* baru atau pengelompokkan data yang berbeda dengan *cluster* sebelumnya, maka proses akan diulang kembali ke tahap 2. Jika *cluster* atau pengelompokkan data telah sama dengan *cluster* sebelumnya, proses dapat dihentikan dan hasil pengelompokan akhir telah diperoleh.

3. Evaluasi Jumlah *Cluster* Menggunakan Nilai DBI

Davies Bouldin Index (DBI) adalah metrik evaluasi, untuk mengukur seberapa optimal pengelompokan (*clustering*) yang dibuat. Semakin rendah nilai indeks *davies-bouldin* atau mendekati nol positif, maka semakin baik pengelompokan (*clustering*) tersebut (Kaligis & Yulianto, 2022). Perhitungan manual evaluasi jumlah *cluster* dilakukan menggunakan *microsoft excel*. Perhitungan manual dapat dilakukan dengan langkah-langkah berikut:

1. Menghitung *Sum of Squares Within-Cluster* (SSW)

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, C_i) \dots \dots \dots (3.4)$$

Keterangan:

m_i = Jumlah data *cluster* ke-i

$d(X_j, C_j)$ = Jarak data ke *centroid*

X_j = data pada setiap *cluster* tersebut

C_j = *centroid cluster* ke-i

2. Menghitung *Sum of Squares Between-Cluster* (SSB)

$$SSB_{i,j} = d(C_i, C_j) \dots \dots \dots (3.5)$$

Keterangan:

C_i = *Cluster* Satu

C_j = *Cluster* lainnya

$d(C_i, C_j)$ = jarak antara *centroid* satu dengan lainnya

3. Menghitung Rasio

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \dots \dots \dots (3.6)$$

Keterangan:

R_{ij} = Rasio antar *cluster*

SSW_i = *cluster* 1

SSW_j = *cluster* 2 atau *cluster* lainnya

$SSB_{i,j}$ = pembagian dari *cluster* 1 dan *cluster* 2 atau *cluster* lainnya

4. Menghitung *Davies Bouldin Index*

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \dots \dots \dots (3.7)$$

Keterangan:

K = *Cluster* yang ada

$R_{i,j}$ = Rasio antara *cluster* i dan j

Max = rasio maksimum

3.1.7 Analisis Karakteristik *Cluster*

Pada tahap ini akan dilakukan analisis *cluster* dengan menggunakan analisis deskriptif. Analisis deskriptif adalah proses merangkum, meringkas, dan mengorganisir data menjadi bentuk yang lebih mudah dipahami dan diinterpretasikan. Tujuan dari analisis deskriptif adalah untuk memberikan gambaran umum tentang karakteristik dasar dari suatu *dataset*.

Pada tahap ini analisis deskriptif dilakukan dengan menganalisis karakteristik dari masing-masing *cluster* yang terbentuk. Untuk

membantu proses analisis dibutuhkan nilai dari 4 variabel data yang sebelumnya telah dihitung rata-ratanya. Kemudian menentukan *range* data (batas minimum dan batas maksimum) pada setiap variabel data yang telah dihitung rata-ratanya di masing-masing *cluster* yang terbentuk. Setelah itu bandingkan *range* pada setiap variabel data yang dimiliki oleh masing-masing *cluster*.

Setelah perbandingan *range* pada setiap variabel data di masing-masing *cluster* dilakukan, maka akan ditemukan karakteristik setiap *cluster* yang terbentuk, sehingga dapat diketahui bahwa *cluster* apa saja yang merepresentasikan pembangunan sosial-ekonomi yang rendah, sedang dan tinggi (Utari, 2021).

3.1.8 Perbandingan Hasil *Cluster* dan Nilai DBI

Pada tahap ini, dilakukan perbandingan hasil pengelompokan data menggunakan algoritma *k-means* antara *tools Rapid Miner* dan *microsoft excel*. Tujuannya adalah untuk menentukan kesamaan atau perbedaan hasil klaster pada setiap kabupaten/kota. Selanjutnya, akan dilakukan perbandingan perhitungan nilai DBI (*Davies Bouldin Index*) antara kedua *tools Rapid Miner* dan *microsoft excel* tersebut. Langkah ini penting guna mengetahui konsistensi dari klasterisasi yang dihasilkan.