

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksploratif. Tujuan dari penelitian ini adalah untuk mengidentifikasi dan pengelompokan wilayah berdasarkan data produksi susu sapi di Jawa Barat menggunakan metode *Hierarchical Clustering*. Penelitian ini bersifat non-eksperimental dan menggunakan teknik analisis data sekunder dari sumber resmi.

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi resmi <https://opendata.jabarprov.go.id/id> yaitu data produksi susu sapi per kabupaten/kota di Provinsi Jawa Barat.

Teknik pengumpulan data dilakukan melalui studi dokumentasi, yaitu dengan mengakses dan mengunduh dataset dari situs resmi <https://opendata.jabarprov.go.id/id> serta sumber pendukung lain yang relevan. Data yang diambil adalah data terbaru yang tersedia dan lengkap untuk seluruh wilayah kabupaten/kota di Jawa Barat.

3.2 Penggunaan Google Colaboratory (Colab)

Dalam penelitian ini, proses analisis data dan penerapan algoritma *Hierarchical Clustering* dilakukan menggunakan Google Colaboratory (Colab), sebuah platform komputasi awan yang dikembangkan oleh google untuk menjalankan kode python secara interaktif.

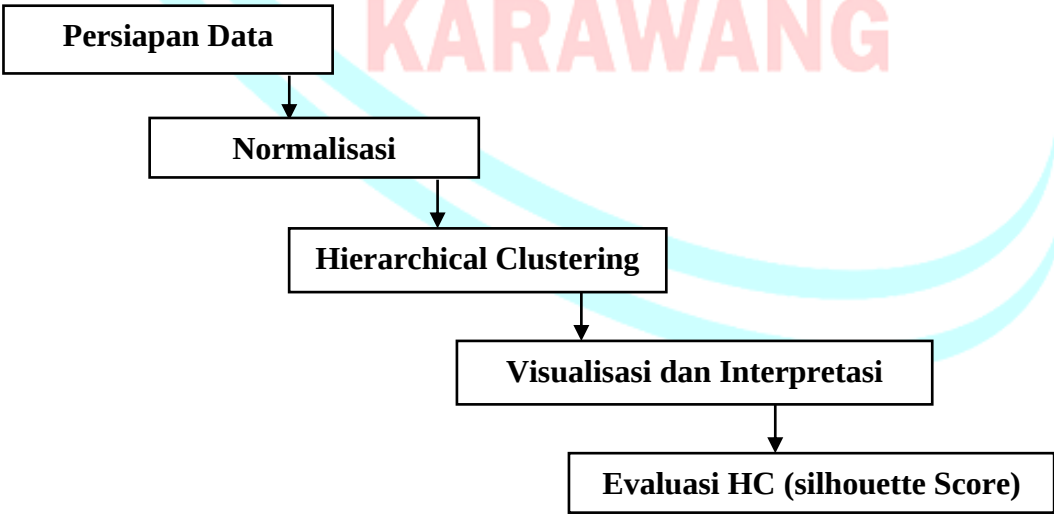
Google Colab dipilih karena memiliki beberapa keunggulan berikut :

- 1) Gratis dan berbasis cloud, sehingga tidak membutuhkan instalasi local maupun perangkat dengan spesifikasi tinggi.
- 2) Terintegrasi dengan Google Drive, memungkinkan penyimpanan dan pengelolaan data secara langsung di cloud
- 3) Mendukung Pustaka Python populer seperti *pandas*, *numpy*, *matplotlib*, *scikit-learn*, dan *scipy*, yang di gunakan dalam proses pembersihan data, normalisasi, clustering, evaluasi, serta visualisasi dendogram.
- 4) Mudah digunakan, bahkan oleh pengguna pemula, karena tampilannya menyerupai buku catatan interaktif (*notebook*)

3.3 Prosedur Penelitian

Dalam Upaya mengelompokan wilayah kabupaten/kota di Jawa Barat berdasarkan pola produksi susu sapi, diperlukan rangkaian tahapan analisis yang sistematis dan berbasis data. Proses analisis ini dimulai dari pengumpulan dan persiapan data, dilanjutkan dengan normalisasi agar data memiliki skala yang seimbang, kemudian dilakukan pengelompokan menggunakan metode *Hierarchical Clustering*, Infante (2002). Seluruh proses ini bertujuan untuk membentuk kelompok wilayah yang memiliki karakteristik produksi serupa, sehingga dapat menjadi dasar dalam penyusunan kebijakan distribusi dan pengembangan peternakan secara lebih efisien dan terarah.

Untuk memberikan gambaran yang jelas mengenai alur kerja analisis yang digunakan dalam penelitian ini, berikut disajikan diagram tahapan proses clustering. Diagram ini menunjukkan lima langkah utama yang dimulai dari persiapan data hingga evaluasi hasil pengelompokan menggunakan *Silhouette Score*. Setiap tahap saling berkesinambungan dan memainkan peran penting dalam menghasilkan analisis yang valid dan dapat diinterpretasikan dengan baik:



Gambar 1 Langkah langkah Penelitian

Diagram alur di atas menggambarkan urutan langkah-langkah yang ditempuh dalam penelitian ini untuk menganalisis pola distribusi produksi susu sapi menggunakan metode *Hierarchical Clustering*. Setiap tahapan memiliki fungsi yang saling berkaitan dan membentuk satu kesatuan proses analisis data yang utuh.

Penjelasan rinci mengenai masing-masing tahap akan dijabarkan pada bagian berikut ini, untuk memperlihatkan bagaimana proses tersebut dijalankan secara teknis dan logis dalam konteks penelitian.

1. Persiapan Data

Tahap pertama dalam proses analisis adalah persiapan data, di mana peneliti mengumpulkan, memverifikasi, dan membersihkan data produksi susu sapi dari kabupaten/kota di Jawa Barat. Data diperoleh dari sumber resmi seperti portal Open Data Provinsi Jawa Barat, yang menyajikan informasi produksi dari tahun 2013 hingga 2024. Data mentah ini mencakup variable-variabel penting seperti nama wilayah, tahun, dan jumlah produksi susu dan satuan ton.

Dalam proses ini juga dilakukan pembersihan data (*data cleaning*) untuk menghapus nilai-nilai kosong, duplikasi, atau entri yang tidak relevan. Menurut Bhattacharjee (2021) Tujuan dari tahap ini adalah memastikan bahwa data yang digunakan bersih, konsisten, dan layak untuk dianalisis lebih lanjut. Kualitas data sangat memengaruhi hasil akhir analisis clustering, sehingga tahap ini menjadi fondasi penting bagi keseluruhan proses.

2. Normalisasi

Setelah data disiapkan, Langkah selanjutnya adalah normalisasi. Proses ini bertujuan untuk menyertakan skala antar variable agar tidak ada variable yang mendominasi perhitungan jarak dalam proses clustering, Raymaekers and Zamar (2020). Dalam penelitian ini digunakan Min-Max Scaler, yang mengubah rentang nilai ke dalam skala $[0,1]$.

Normalisasi sangat penting terutama saat menangani data numerik dengan rentan nilai yang berbeda-beda secara drastis, seperti jumlah produksi dari wilayah dengan skala industri besar dibandingkan wilayah dengan produksi sangat kecil, Limpert and Stahel (2020). Dengan menyamakan skala, perhitungan jarak antar data menjadi lebih adil dan akurat, sehingga hasil clustering mencerminkan kedekatan karakteristik sebenarnya antar wilayah.

3. Hierarchical Cluster

Pada tahap ketiga, dilakukan proses pengelompokan menggunakan metode *Hierarchical Cluster* (HC). Penelitian ini menerapkan metode agglomerative dengan Teknik *linkage Ward*, yang menggabungkan data berdasarkan minimisasi variansi hingga seluruh data bergabung dalam satu hirarki.

Hierarchical Clustering dipilih karena mampu menghasilkan dendrogram, yaitu diagram pohon yang menggambarkan proses penggabungan cluster secara visual. Menurut Campello et al. (2020) Kelebihan metode ini adalah tidak perlu menentukan jumlah cluster di awal, sehingga lebih fleksibel dalam eksplorasi data. Hasil dari tahap ini akan menjadi dasar dalam membentuk cluster-cluster wilayah berdasarkan kemiripan produksi susu sapi.

4. Visualisasi dan Interpretasi

Setelah proses clustering selesai, dilakukan visualisasi hasil pengelompokan dalam bentuk dendrogram. Visualisasi ini menampilkan struktur hierarki antar wilayah berdasarkan jarak atau Tingkat kemiripan. Sumbu vertical menunjukkan jarak antar kelompok, sedangkan sumbu horizontal menggambarkan entitas yang dianalisis, yaitu kabupaten/kota di Jawa Barat.

Interpretasi dendrogram memungkinkan peneliti menentukan jumlah cluster optimal dengan cara “memotong” pohon pada level tertentu. Selain itu, visualisasi ini memberikan gambaran hubungan antarwilayah dan tingkat kedekatannya dalam konteks produksi susu, Pereira et al. (2021). Dengan begitu, hasil analisis menjadi lebih komunikatif dan mudah dipahami oleh pihak non-teknis seperti pengambil kebijakan.

5. Evaluasi HC (Silhouette Score)

Tahap terakhir adalah evaluasi hasil clustering menggunakan Silhouette Score, yang mengukur seberapa baik suatu data berada dalam clusternya, dibandingkan dengan cluster lainnya, Batool and Hennig (2021). Nilai silhouette berkisar antara -1 hingga 1, Dimana nilai mendekati 1 menunjukkan pemisah cluster yang sangat baik.

Evaluasi ini penting untuk menilai apakah hasil clustering mencerminkan struktur data yang optimal. Jika nilai silhouette rendah, maka perlu dipertimbangkan metode clustering lain atau dilakukan penyesuaian.