

BAB III

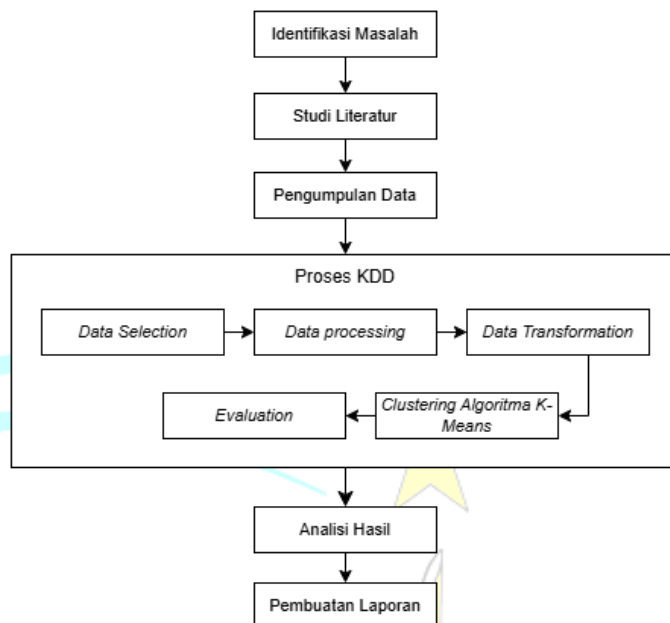
METODE PENELITIAN

3.1 Objek Penelitian

Data yang digunakan dalam penelitian ini berasal dari dataset yang dikeluarkan oleh Dinas Koperasi dan UMKM Provinsi Jawa Barat, yang mencakup berbagai atribut terkait UMKM di Jawa Barat. Dataset tersebut berisi data kategori usaha dan proyeksi jumlah UMKM di tingkat kabupaten/kota dari tahun 2017 sampai dengan 2023. Beberapa atribut penting yang akan dianalisis meliputi jumlah unit usaha, jenis usaha, dan lokasi geografis. Informasi ini menjadi dasar untuk membentuk klaster yang merepresentasikan pola persebaran dan potensi ekonomi UMKM di berbagai wilayah. Data yang dikumpulkan akan diproses dan dianalisis sesuai dengan kebutuhan penelitian untuk memastikan hasil yang akurat dan relevan (Rozaq et al., 2018).

3.2 Prosedur Penelitian

Penelitian ini menggunakan teknik klasterisasi, untuk memahami pola persebaran UMKM di Jawa Barat. Metode algoritma *K-Means* dipilih karena mampu mengelompokkan data berdasarkan kesamaan karakteristik sehingga setiap kelompok (klaster) yang terbentuk dapat merepresentasikan pola dan tren yang ada di data UMKM (Han et al., 2021). Pendekatan ini juga memungkinkan analisis yang lebih dalam terhadap kelompok-kelompok yang terbentuk, yang pada akhirnya dapat memberikan wawasan untuk pengembangan kebijakan atau program khusus di setiap klaster (Tan et al., 2020). Dalam hal ini, algoritma *K-Means* digunakan sebagai metode klasterisasi karena terkenal efektif dalam memisahkan data ke dalam kelompok yang homogen (Rashidi et al., 2022). Adapun tahap penelitian yang dilakukan adalah sebagai berikut.



Gambar 3. 1 Tahapan Penelitian

Metodologi dalam penelitian ini disusun melalui sejumlah tahapan utama yang dijelaskan sebagai berikut:

1. Identifikasi Masalah

Tahapan ini berfokus pada menentukan permasalahan utama yang akan diteliti, memahami konteks, latar belakang, dan ruang lingkup masalah yang relevan dengan topik penelitian.

2. Studi Literatur

Pada tahap ini, dilakukan penelusuran dan kajian literatur terkait dengan topik penelitian. Hal ini mencakup membaca jurnal, artikel, buku, atau sumber terpercaya lainnya untuk memahami teori dan metode yang relevan.

3. Pengumpulan Data

Data dikumpulkan menggunakan metode pengambilan data yang telah ada sebelumnya, sesuai dengan kebutuhan penelitian. Data yang dikumpulkan harus sesuai dengan tujuan penelitian dan dirancang untuk mendukung analisis yang direncanakan.

4. proses Knowledge Discovery in Databases (KDD)

Proses *Knowledge Discovery in Databases* (KDD) merupakan rangkaian tahapan sistematis yang digunakan untuk menggali pengetahuan tersembunyi dari sekumpulan data mentah. KDD tidak hanya mencakup proses analisis data, tetapi juga serangkaian langkah awal hingga akhir, mulai dari seleksi data hingga interpretasi hasil. Dalam konteks penelitian ini, proses KDD digunakan untuk melakukan analisis kluster terhadap data UMKM di Jawa Barat menggunakan algoritma *K-Means*. Berikut adalah uraian tiap tahapannya:

a. Data Selection

Pada tahap ini, dilakukan pemilihan atribut-atribut yang relevan, seperti kategori usaha dan jumlah UMKM per kabupaten/kota. Hanya data yang sesuai dengan tujuan analisis yang disertakan untuk memastikan fokus dan akurasi hasil klusterisasi.

b. Data Preprocessing

Data yang telah dipilih kemudian melalui tahap pembersihan (cleansing), normalisasi, dan standarisasi. Tujuan utamanya adalah untuk mengatasi permasalahan seperti nilai kosong (missing values), duplikasi data, serta inkonsistensi format agar data siap digunakan dalam proses analisis lebih lanjut.

c. Data Transformation

Pada tahap ini, data diubah ke dalam format atau struktur yang sesuai untuk proses data mining. Salah satu proses penting dalam tahap ini adalah **normalisasi** data, terutama karena algoritma *K-Means* menggunakan jarak *Euclidean* dalam pengelompokan. Fitur-fitur dengan skala yang berbeda perlu diseragamkan agar tidak mempengaruhi hasil kluster. Normalisasi dilakukan dengan metode **Min-Max Scaling**

d. Clustering Algoritma K-Means

Ini merupakan inti dari proses KDD, di mana algoritma statistik atau pembelajaran mesin diterapkan untuk menemukan pola atau struktur tersembunyi dalam data. Dalam penelitian ini, digunakan **algoritma K-Means**, yaitu metode klustering yang membagi data ke dalam k kelompok berdasarkan

kemiripan atributnya. Proses dimulai dengan inisialisasi *centroid*, dilanjutkan dengan penugasan klaster berdasarkan jarak terdekat, dan pembaruan posisi *centroid* hingga proses konvergen. Tujuannya adalah menemukan pola distribusi UMKM berdasarkan karakteristik usaha dan wilayahnya.

e. Evaluation Model Klasterisasi

Setelah proses klastering dilakukan, hasilnya perlu dievaluasi untuk menilai sejauh mana klaster yang terbentuk memiliki kualitas yang baik. Dalam penelitian ini, digunakan beberapa metrik evaluasi seperti, *Sum of Squared Errors* (SSE) untuk melihat kompaksi klaster, *Davies-Bouldin Index* (DBI) untuk mengukur pemisahan antar klaster, *Calinski-Harabasz Index* (CHI) untuk menilai rasio antara variansi antar dan dalam klaster, serta Silhouette Score untuk mengukur kesesuaian anggota klaster terhadap *centroid*-nya. Hasil evaluasi ini digunakan untuk menentukan jumlah klaster optimal dan menilai kualitas pengelompokan.

5. Analisis Hasil

Pada tahap ini, hasil dari analisis (hasil *clustering*) diinterpretasikan. Temuan tersebut dihubungkan dengan tujuan penelitian dan dijadikan dasar untuk menjawab pertanyaan penelitian.

6. Pembuatan Laporan

Semua hasil penelitian disusun dalam bentuk laporan yang mencakup latar belakang, metode, hasil, diskusi, dan kesimpulan. Laporan ini bertujuan untuk menyampaikan temuan penelitian kepada pembaca secara sistematis dan jelas.

3.3 Teknik Analisis Data

Teknik analisa data untuk penelitian menggunakan proses *Knowledge Discovery In Database* (KDD) dan menggunakan algoritma *K-Means* dalam pengelompokkannya. Penggunaan proses KDD berfungsi agar tahapan yang dilakukan berjalan secara sistematis. Proses ini untuk mencari secara sistematis suatu ikatan baru di dalam hubungan antara setiap variabel. Data mining serta KDD sering digunakan secara bergantian berfungsi untuk menggambarkan proses menggali data tersembunyi dari dasar informasi yang besar. Salah satu tahapan

dalam proses KDD merupakan data mining. Kedua hal itu memiliki konsep yang berbeda, tetapi memiliki kaitan satu sama lain. Tahapan *Knowledge Discovery in Databases* (KDD) terdiri dari *Selection*, *Preprocessing*, *Transformation*, *Data mining*, *Evaluation*, dan *Knowledge*. Dalam penelitian ini, proses analisis data dilakukan dengan memanfaatkan platform *Google Colab* untuk menjalankan pemrograman dan visualisasi hasil klasterisasi secara efisien dan fleksibel.

