

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

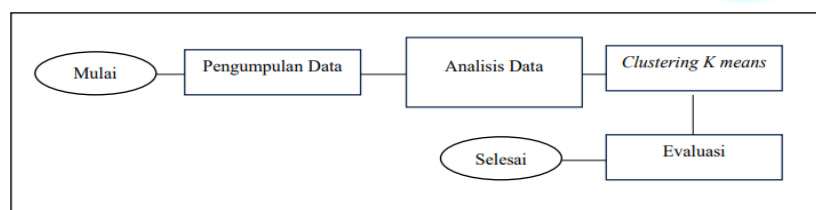
Penelitian ini menggunakan data *history* data penjualan di Toko Azko Technomart, data penjuala terdiri dari 300 produk dengan *history* data penjualan 15000 data dari bulan mei 2022 hingga juni 2025. Sampel data yang diambil di tujunjukan pada Tabel 3.1.

Tabel 3.1 Sampel Data

Nama Produk	Jumlah Produk
Kitchen Appliances	2.202
Regular Sports & Games	164
Electrical Supplies	1.277
Ladder	547
Protection	8.319
Residential	1.018
Auto Gadget & Comfort	195
Racking/Organizing	93
Electrical Light	3.049
Safes	33

3.2 Prosedur Penelitian

Tahap pertama pada prosedur penelitian pengumpulan data dilanjut dengan pembersihan data dengan menggunakan *Text Preproccesing* tahapan selanjutnya *Clustering* menggunakan *K-means* tahapan yang terakhir evaluasi hasil pengelompokan menggunakan metode *silhouette score* tahapan proseder penelitian digambarkan pada Gambar 3.1.



Gambar 3.1 Prosedur Penelitian

3.3 Pengumpulan Data

Pengambilan data penjualan dalam penelitian pada Toko Azko Technomart, database transaksi penjualan perusahaan melalui admin purchasing. Data yang terkumpul mencakup rincian transaksi penjualan selama periode Mei 2022 hingga Juni 2025. Data penjualan yang dianalisis mencakup variabel seperti nama produk, jumlah yang terjual, dan total harga. Proses pengumpulan data dilakukan untuk menjamin tersedianya informasi yang tepat dan sesuai dengan kebutuhan analisis.

3.4 Analisis Data

Pada proses analisis data terdapat tahap mentransformasi, membersihkan, memeriksa, dan memodelkan data dengan tujuan untuk menemukan informasi yang berguna, lalu membuat kesimpulan, dan mendukung pengambilan keputusan. Alur dari analisis data digambarkan pada tabel 3.2.

Tabel 3.2 Alur Analisis Data

Analisis Data
Input : Data Penjualan
Proses :
1. Baca kalimat
2. Menghapus Data yang tidak akurat
Output : Data menjadi akurat

3.5 Clustering Algoritma K-Means

K-Means Clustering merupakan metode pengelompokan berbasis partisi yang berfungsi untuk memecah sekumpulan data dalam sejumlah kluster (k) yang telah ditetapkan sebelumnya. Prinsip kerja utama algoritma ini adalah dengan meminimalkan jarak antara setiap titik data terhadap pusat kluster (*centroid*) terdekatnya. Adapun tahapan-tahapan dalam menjalankan algoritma *K-Means* yaitu sebagai berikut:

1. Menentukan jumlah kluster (k). Langkah pertama adalah menetapkan target jumlah kelompok atau kluster yang ingin dibentuk dari kumpulan data.
2. Mengelompokkan data ke dalam k kluster yang dibentuk berdasarkan lokasi pusat (*centroid*) yang telah dipilih atau ditentukan sebelumnya.
3. Menghitung ulang lokasi pusat (*centroid*) pada masing-masing kelompok. Nilai baru *centroid* diperoleh dengan mencari nilai rata-rata (*mean*) seluruh data yang ada di dalam kluster tersebut pada setiap fitur.

4. Mengalokasikan data ke *centroid* terdekat. Menempatkan kembali masing-masing data ke klaster yang memiliki *centroid* paling dekat. Pengukuran jarak pada titik data dengan *centroid* ini dilakukan memakai rumus *Euclidean (Euclidean Distance)*.

3.6 Evaluasi

Tahap evaluasi penelitian ini dilakukan dengan metode *Silhouette Score*. Tujuan utama dari pengujian ini adalah untuk mengukur tingkat akurasi dan kualitas dari model klasterisasi yang telah dibentuk. Terdapat tiga langkah utama dalam proses perhitungan *Silhouette Score*, yaitu:

1. Menghitung rerata jarak antara suatu objek dengan seluruh objek lain yang berada di dalam klaster yang sama menggunakan persamaan terkait.
2. Menghitung jarak objek tersebut terhadap seluruh objek pada klaster lain, lalu mengambil nilai jarak minimumnya menggunakan persamaan terkait.
3. Menghitung nilai akhir *Silhouette Coefficient* menggunakan persamaan.

