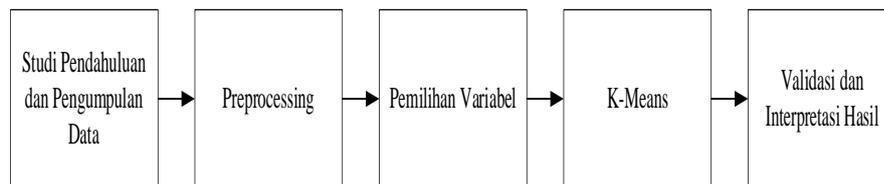


BAB III

METODOLOGI PENELITIAN

Metodologi penelitian mengacu pada perencanaan penelitian yang sistematis dan ilmiah. Metodologi penelitian komparatif diterapkan dalam penelitian ini. Data akan dianalisis dengan menggunakan pendekatan algoritma k-means, yang memisahkan data ke dalam kelompok-kelompok (cluster) berdasarkan kemiripan atribut. Analisis penelitian ini menggunakan perangkat lunak Orange untuk teknik algoritma k-means. Seperti yang ditunjukkan pada Gambar 3.1, prosedur yang dilakukan meliputi tinjauan literatur yang relevan, pengumpulan data, pra-pemrosesan, penerapan algoritma k-means, dan analisis temuan.



Gambar 3. 1 Tahapan Penelitian

3.1 Studi Pendahuluan

Peneliti meninjau literatur tentang segmentasi pasar, industri kopi, dan pendekatan *K-Means Clustering*. meninjau penelitian sebelumnya yang selaras dengan tujuan penelitian ini.

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari data penjualan bubuk kopi dari Januari 2018 hingga Maret 2023, yang meliputi penjualan harian dan stok di KSU Fa Masa. Data yang dikumpulkan berasal dari KSU Fa Masa.

3.3 Preprocessing

File excel yang berisi semua data penjualan disiapkan untuk pemeriksaan. Jika diperlukan, lakukan standarisasi data. Pisahkan data ke dalam variabel-variabel yang akan digunakan dalam penelitian. Sebagai contoh, KSU Fa Masa menyediakan data statistik penjualan bubuk kopi dari tahun 2018 hingga 2023, termasuk data tahun, nama produk, harga, dan jumlah produk (Alghifari & Juardi, 2021).

3.4 Pemilihan Variabel

Pilih variabel yang relevan untuk analisis segmentasi, seperti nama produk, harga, dan volume penjualan. Karena data tahun tidak termasuk dalam data yang dapat dianalisis menggunakan metode *k-means*, maka data tersebut tidak digunakan dalam analisis penghitungan.

3.5 Penerapan Algoritma *K-Means*

K-Means adalah salah satu algoritma pengelompokan yang paling banyak digunakan karena kecepatan dan kesederhanaannya. Algoritme ini membagi menjadi beberapa kluster dengan menetapkan setiap item ke pusat kluster terdekat - nilai rata-rata variabel untuk semua item dalam kluster tertentu - menggunakan

ukuran jarak yang telah ditentukan sebelumnya. Salah satu algoritma yang paling penting dalam penggalian data. (Aziz et al., 2022).

Menurut (Aziz et al., 2022), langkah-langkah algoritma k-means adalah:

1. Tentukan centroid dari k titik data, di mana k mewakili jumlah cluster yang digunakan.
2. Setiap titik (data) dicari centroid terdekatnya.
3. Sebuah cluster mengacu pada kumpulan titik data yang membentuk centroid.
4. Hitung ulang centroid untuk setiap cluster.
5. Ulangi langkah 1-4 sampai centroid tidak bergeser.

Pendekatan pengelompokannya adalah k-means, dan jarak euclidean digunakan untuk menghitung kedekatan data. Metode k-means berusaha untuk meminimalkan total jarak euclidean antara setiap titik dan cluster terdekat, secara spesifik. Sedangkan menurut Agusta, jarak euclidean dihitung dengan menggunakan persamaan sebagai berikut

$$d_{ij} = \sqrt{\sum_{k=1}^n \{x_{ik} - x_{jk}\}^2} \quad (1)$$

Keterangan:

d_{ij} = Jarak antara data ke-I dan ke-j

n = dimensi data

X_{ik} = koordinat data ke- i pada dimensi k

X_{jk} = koordinat data ke- j pada dimensi

3.6 Validasi Hasil Dan Interpretasi Hasil

Aplikasi Data Mining oranye akan digunakan untuk menganalisis hasil dari metode k-means clustering. Bergantung pada hasil analisis metode k-means clustering, data akan ditampilkan dalam bentuk tabel dan divisualisasikan dalam bentuk scatter plot.