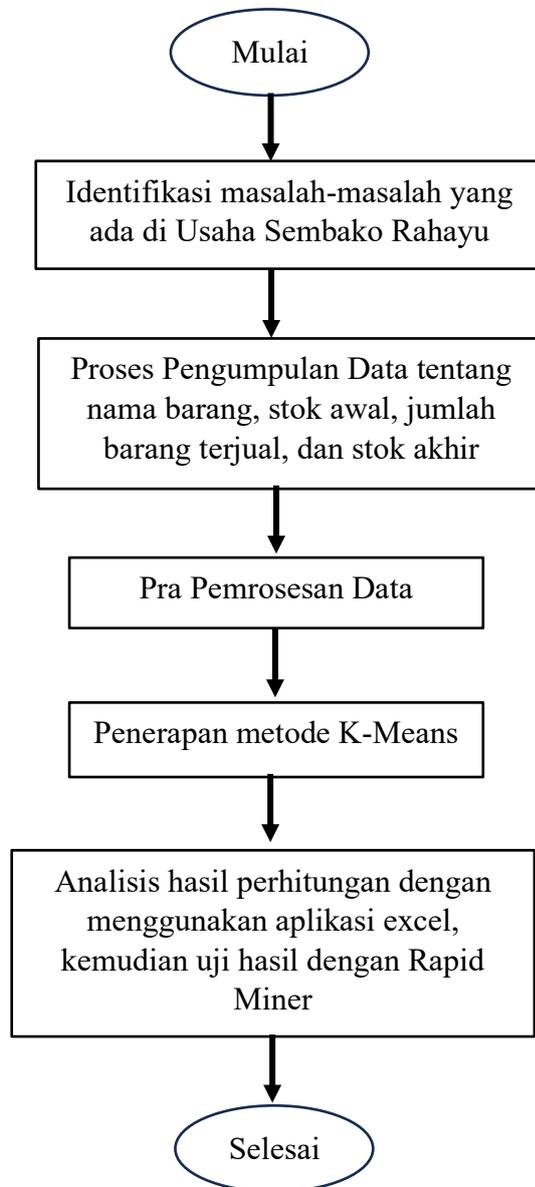


## BAB III

### METODE PENELITIAN

#### 3.1 Kerangka Kerja Penelitian



**Gambar 3.1** Kerangka Kerja Penelitian

Adapun desain penelitian ini dijelaskan sebagai berikut :

1. Penelitian dimulai dengan menetapkan fokus utama, yaitu mengidentifikasi jenis barang apa saja yang laku terjual
2. Pada tahap identifikasi masalah, fokusnya adalah memahami jenis barang, dan mengetahui stok awal
3. Pengumpulan data dilakukan dengan melakukan wawancara dan observasi langsung, dimana data tersebut diambil dari data transaksi selama 1 tahun yaitu tahun 2024 di Toko Sembako Rahayu
4. Data yang telah dikumpulkan lalu diproses untuk mengetahui berapa jumlah stok awal, jumlah barang yang terjual, dan stok akhir
5. Kemudian, setelah data diproses selanjutnya data tersebut akan diterapkan dengan menggunakan metode K-Means Clustering
6. Dilakukan analisis lebih lanjut untuk mengetahui hasil akhir yang sesuai dengan tujuan dari algoritma

## 3.2 Objek Penelitian

### 3.2.1 Lokasi Penelitian



**Gambar 3.2** Lokasi Penelitian

Objek penelitian yang digunakan dalam penelitian ini adalah Toko Usaha Dagang Rahayu yang beralamat di Dusun Wonodadi, Desa Mandalasena, Kecamatan Silangkitang

### **3.2.2 Waktu Penelitian**

Penelitian ini dilaksanakan terhitung dari perencanaan penelitian, pelaksanaan penelitian, sampai pembuatan laporan penelitian. Penelitian dilakukan di bulan Desember sampai dengan bulan Januari 2024

## **3.3 Teknik Pengumpulan Data**

Metode pengumpulan data dalam penelitian ini dilakukan dengan cara :

### **3.3.1 Observasi**

Observasi adalah salah satu teknik pengumpulan data yang melibatkan pengamatan langsung terhadap objek atau fenomena yang sedang diteliti. Dalam konteks penelitian, observasi digunakan untuk memperoleh data empiris dengan melihat dan mencatat perilaku, kejadian, atau proses yang terjadi secara alami di lingkungan yang relevan. Observasi pada penelitian ini dengan langsung melakukan pengamatan ke Toko Sembako Rahayu

### **3.3.2 Wawancara**

Wawancara atau interview adalah sebuah proses memperoleh keterangan untuk tujuan peneliti dengan cara tanya jawab sambil bertatap muka antara pewawancara dengan orang yang diwawancarai. Peneliti melakukan wawancara dengan pemilik toko dan beberapa konsumen untuk mendapatkan informasi mengenai Toko Sembako Rahayu.

### **3.3.3 Dokumentasi**

Dokumentasi adalah metode untuk mencari data mengenai hal-hal atau variabel yang berupa catatan-catatan, buku, bukti transaksi, dan sebagainya. Metode ini digunakan sebagai pelengkap guna memperoleh data sebagai bahan informasi yang penting

## **3.4 Model Penelitian**

Pengolahan data mining yang dilakukan pada penelitian ini mengikuti tahapan dalam Knowledge Discovery in Database (KDD) untuk menghasilkan informasi sesuai dengan urutan yang sudah ditentukan.

### **3.4.1 Data Selection (Pemilihan Data)**

Data set yang digunakan adalah data transaksi penjualan sembako pada Toko Sembako Rahayu, mencakup informasi tentang nama barang, stok awal, jumlah barang yang terjual, stok akhir, satuan, harga beli, dan jumlah harga. Data ini dikelompokkan berdasarkan nilai satuannya dan juga penyeleksian dalam variabelnya. Data ini dipilih untuk mengidentifikasi data penjualan sembako pada toko ini.

### **3.4.2 Preprocessing**

Data Cleaning (Pembersihan Data)

Mengatasi Missing Values : jika terdapat nilai kosong atau nilai yang hilang dengan metode imputasi (mean, median dan mode) atau menghapus entri yang tidak lengkap

### 3.4.3 Data Transformation (Transformasi Data)

Normalisasi : mengubah data kedalam skala tertentu, agar fitur dengan skala besar tidak mendominasi model clustering

### 3.4.4 Data Mining

Pada tahap data mining ini algoritma data mining digunakan untuk menemukan pola dalam data. Hubungan antara data mining dalam dataset antara lain :

1. Tentukan jumlah cluster (k)
2. Inisialisasi centroid secara acak.
3. Iterasi proses penugasan titik data ke centroid terdekat dan pembaruan posisi centroid hingga konvergensi tercapai.

### 3.4.5 Model evaluasi

1. Menentukan Centroid Awal

Memilih  $k$  centroid awal secara acak dari dataset. Centroid adalah pusat dari cluster

2. Menentukan jarak antara titik data dan centroid

Hitung jarak data dengan centroid menggunakan rumus jarak menggunakan rumus Eucliden yang dirumuskan sebagai berikut:

$$D_{(i,j)} = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (XKi - XKj)^2}$$

Dimana  $D(i, j)$  adalah jarak data ke  $i$  ke pusat cluster  $j$ ,  $X_{ki}$  adalah data ke  $i$  pada atribut data ke  $k$ ,  $X_{kj}$  adalah titik pusat ke  $j$  pada atribut ke  $k$ .

### 3. Mengelompokkan Titik Data ke Cluster Terdekat

Kelompokkan data berdasarkan kedekatan dengan centroid atau nilai yang paling minimum diantara jarak dari masing-masing centroidnya.

4. Setelah menyimpulkan bahwa data tersebut masuk kedalam centroid terdekatnya, kemudian jarak yang paling minimum di pangkatkan 2

### 5. Fungsi Objektif (Minimizing Inertia atau SSE)

Tujuan dari K-Means adalah untuk meminimalkan **Sum of Squared Errors (SSE)**, yang merupakan jumlah dari jarak kuadrat antara setiap titik data dan centroid cluster-nya:

$$SSE = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j\|^2$$

**Dimana :**

**$k$  : Jumlah Cluster**

**$x_i$  : Titik data dalam cluster  $C_j$**

**$c_j$  : Centroid dari cluster  $C_j$**

Untuk menghitung nilai SSE (Sum of Square Error) atau nilai yang eror, bisa juga melakukan penjumlahan nilai jarak minimum yang sudah di pangkatkan.

#### 6. Menghitung Ulang Centroid

Perbaharui nilai centroid baru dengan lokasi dari pusat cluster menggunakan persamaan:

$$C_j^{(t+1)} = \frac{1}{N_j} \sum_{x_i \in C_j} x_i$$

Dimana :

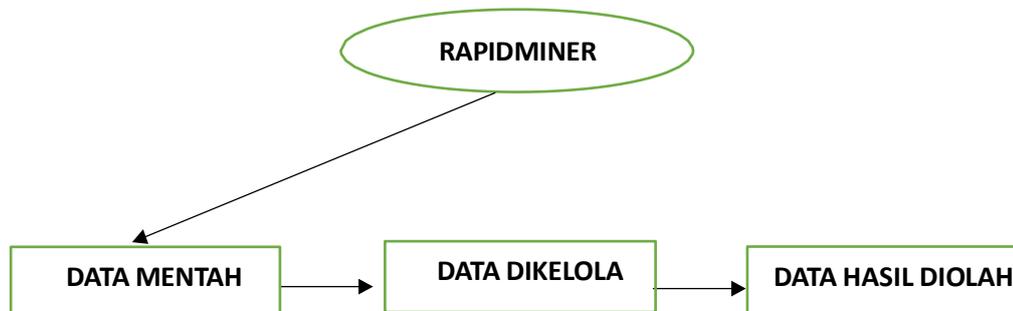
- a.  $C_j^{(t+1)}$  → Centroid baru untuk klaster ke-j pada iterasi berikutnya  $(t + 1)$
- b.  $N_j$  → Jumlah titik data yang termasuk dalam klaster  $C_j$
- c.  $x_i \in C_j$  → Semua titik data  $x_i$  yang menjadi anggota klaster  $C_j$
- d.  $\sum_{x_i \in C_j} x_i$  → Penjumlahan semua titik data dalam klaster  $C_j$
- e.  $\frac{1}{N_j}$  → Rata-rata koordinat dari semua titik dalam klaster  $C_j$

#### 7. Mengevaluasi Konvergensi

- a. Mengulangi langkah tersebut hingga salah satu kondisi berhenti tercapai
- b. Centroid tidak lagi berubah secara signifikan
- c. Tidak ada perubahan dalam keanggotaan cluster
- d. Jumlah iterasi maksimum tercapai

### 3.5 Rancangan Penganalisis Sistem Data

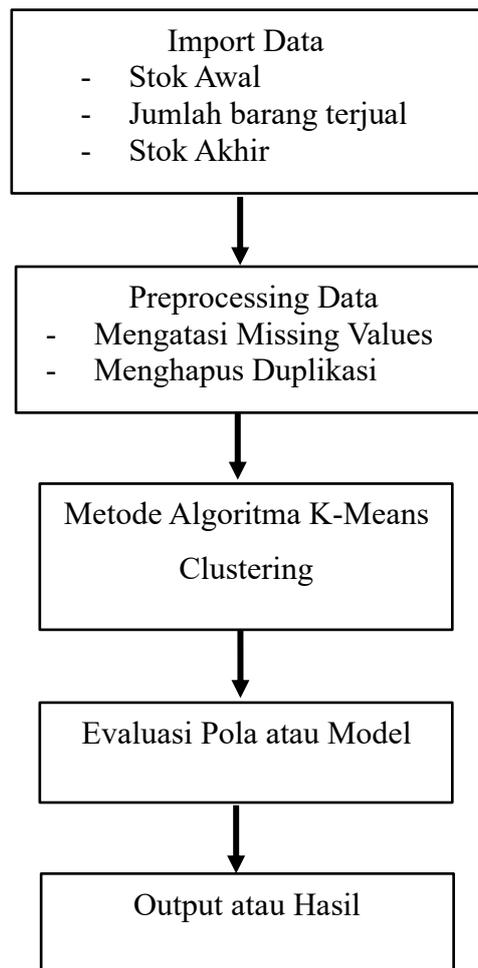
Pada bagian ini peneliti akan dapat menjelaskan dari sistem pendukung dan analisa pada data stok awal barang, jumlah barang yang terjual, dan stok akhir penjualan dari Toko Usaha Sembako Rahayu yang terdiri dari bentuk sistem dari rancangan yang diketahui dalam bentuk komponen yang digunakan oleh peneliti. Sehingga penelitian ini dapat menjelaskan berbagai perancangan sistem yang dikelola dan dirancang dengan menggunakan algoritma K-Means Clustering dengan menggunakan proses data mining, untuk lebih terarah dalam melakukan analisis sistem yang sedang digunakan. Untuk memperoleh gambaran mengenai keadaan sistem yang dibutuhkan dapat dilihat pada gambar dibawah 3.3 ini :



**Gambar 3.3** Rancangan Penganalisis Sistem Data

Dari gambar diatas, dijelaskan bahwa sebuah bentuk proses rancangan yang dibentuk oleh peneliti sendiri sehingga proses data identifikasi data penjualan pada Toko Sembako Rahayu pada rancangan ini terdapat data mentah yang kemudian dikelola sehingga menghasilkan suatu data dengan menggunakan algoritma K-Means Clustering dan dibantu dengan platform Rapid Miner.

### 3.6 Rapid Miner



**Gambar 3.4** Alur Rapid Miner Algoritma K-Means Clustering

Adapun desain alur Rapid Miner Algoritma K-Means Clustering ini dijelaskan sebagai berikut :

1. Import Data

Mengimport data set penjualan yang berisi nama barang, satuan, stok awal, jumlah barang terjual, dan stok akhir.

2. Preprocessing

Melakukan pembersihan data untuk menghindari missing values dan data yang duplikasi

3. Metode Algoritma K-Means Clustering

Membuat metode atau model klasifikasi dengan menggunakan algoritma K-Means Clustering

4. Evaluasi pola atau model

Mengevaluasi kinerja model untuk memastikan keakuratan perhitungannya

5. Output atau hasil

Menghasilkan data yang sudah diolah atau diproses untuk mendukung kesimpulan