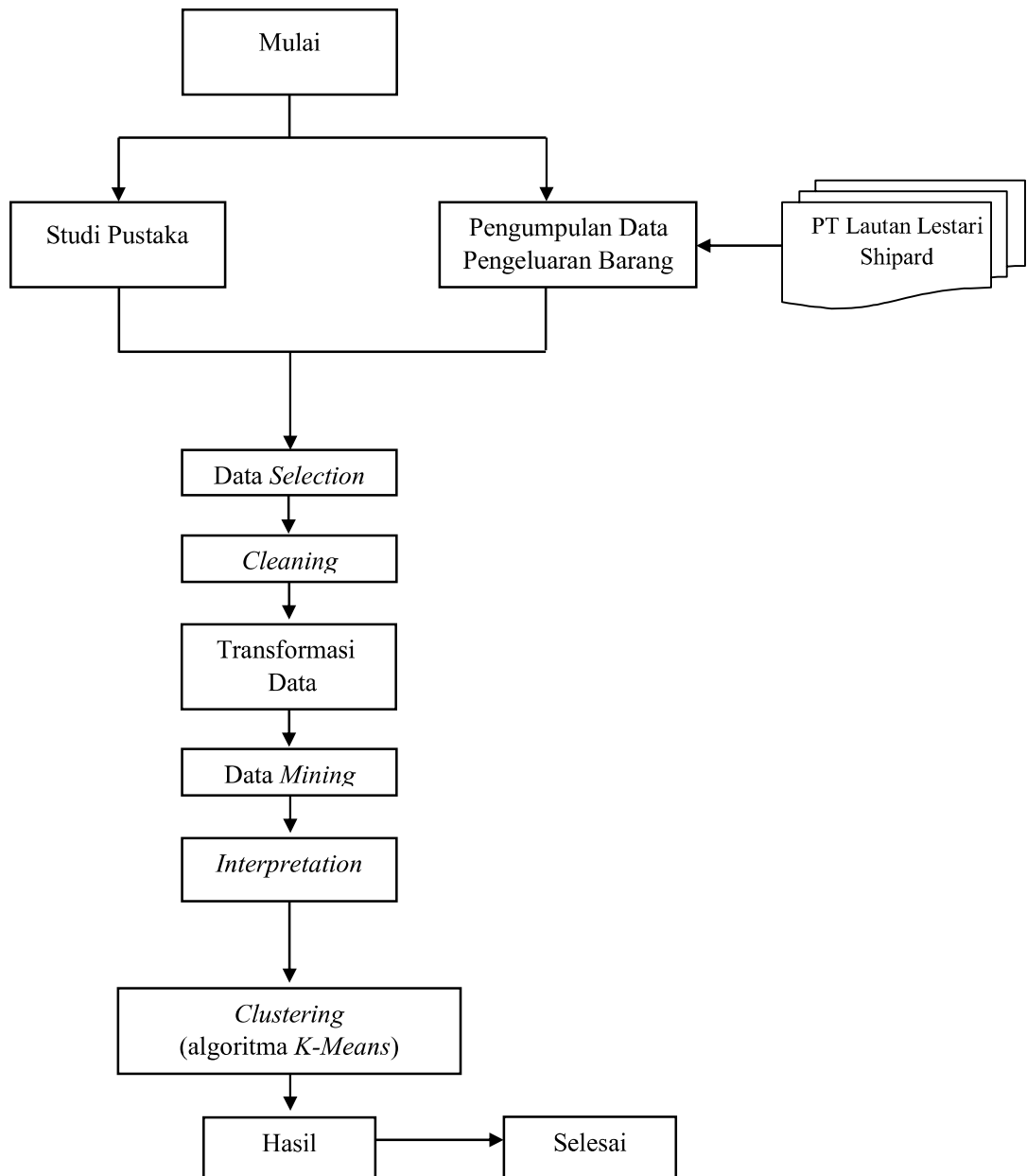


BAB III
METODOLOGI PENELITIAN

3.1 Desain Penelitian



Gambar 3.1 Tahapan Penelitian

Sumber: (Peneliti, 2022)

1. Studi Pustaka

Tinjau literatur terkait konsep data *mining* dengan menggunakan metode *clustering* dan algoritma *K-Means*. Literatur berasal dari data pengeluaran komoditas selama periode penelitian.

2. Pengumpulan Data

Untuk memahami data yang diperlukan, peneliti melakukan pengumpulan salinan data arsip pengeluaran barang dengan periode Januari-Desember 2020 dan Januari-November 2021.

3. Praproses data meliputi:

- a. Data Selection

Untuk memilih kumpulan data (*dataset*) yang akan digunakan dalam penelitian ini dalam bentuk data pengeluaran barang, yang mencakup informasi tentang jenis barang dan nilainya.

- b. *Cleaning*

Untuk merapikan data dengan cara yaitu menyempurnakan data, membuang data yang sama (duplikat).

- c. Transformasi data

Guna membangun data supaya bisa di *cluster* atau dapat juga dengan cara menambah atribut baru.

- d. Data *mining*

Menerapkan metode pengelompokan dan algoritma *K-Means* untuk memperoleh data pengeluaran produk, dan proses penambangan data akan dijabarkan menggunakan *software RapidMiner*.

3.2 Teknik Pengumpulan Data

Guna memperoleh data dari masalah penelitian perlu dilakukan pengumpulan data. Data yang telah dikumpulkan akan dilakukan pengolahan dengan data *mining* dan akan menghasilkan informasi yang bernilai guna. Ada beberapa metode pengumpulan data yang akan dikenakan, meliputi:

1. Observasi, Kegiatan yang dilakukan saat memasuki lokasi penelitian dengan mengamati langsung di lokasi penelitian. Peneliti dapat mengumpulkan data pengeluaran barang dan mengamati bagaimana barang ditempatkan di gudang PT Lautan Lestari Shipyard. Dalam penelitian ini yang akan diamati adalah penempatan barang digudang, proses pengambilan barang hingga proses rekapan data pengambilan barang.
2. Dokumentasi, diperlukan sebagai alat bukti bahwa peneliti mendapatkan data langsung dari lokasi penelitian di PT Lautan Lestari Shipyard. Dokumentasi tersebut dapat dalam bentuk foto kegiatan penelitian dan juga foto rak-rak barang.

3. Studi Pustaka

Dalam studi ini, peneliti mencari informasi pengembangan teoritis yang relevan dengan subjek dengan melakukan tinjauan literatur yang terkait dengan topik penelitian dan prosedur penelitian, terutama jurnal dan buku yang relevan.

3.3 Operasional Variabel

Dalam sebuah penelitian, variabel operasional digunakan untuk memastikan atribut dan nilai yang terkait. Pembersihan dan penggalian variabel dari data yang dikumpulkan selama proses penelitian. Karakteristik berikut dimasukkan dalam analisis sebagai variabel independen:

1. Jenis item atau kategori digunakan untuk menentukan hal-hal yang memiliki pola frekuensi tinggi dan kombinasi item,
2. Tahun transaksi digunakan untuk mengetahui total keseluruhan pengambilan barang berdasarkan item yang akan diolah dengan metode *clustering*.

3.4 Algoritma *K-Means Clustering*

Awalnya, algoritma *K-Means* memanfaatkan beberapa komponen populasi sebagai pusat dari *cluster* awal. Pusat *cluster* dipilih dengan cara random dari satu set data kependudukan di tahap ini. Sebagai langkah selanjutnya, *K-Means* menjalankan tes pada data *swarm* untuk menentukan jarak minimum antara setiap komponen dan masing-masing pusat *cluster* yang telah ditentukan. Setiap kali komponen data baru perlu ditugaskan ke pusat *cluster* tertentu, posisi pusat *cluster* dihitung ulang (Darmi & Setiawan, 2017: 150).

Dalam bukunya (Wahyudi et al. 2020: 6) menegaskan Dengan menggunakan algoritma pengelompokan *K-means*, data dapat dibagi menjadi beberapa kelompok berdasarkan seberapa jauh data dalam kelompok-kelompok itu satu sama lain. Algoritma menggunakan fungsi untuk membandingkan set data

yang memiliki banyak kesamaan. Fungsi *Euclidean* digunakan untuk mengetahui seberapa jauh sesuatu. Selanjutnya, data diurutkan berdasarkan kedekatan dan disimpan.

Berikut merupakan beberapa langkah dalam pengelompokan data (Witanto et al., 2019: 704)

1. Tentukan banyaknya cluster yang hendak dibentuk.
2. Inisialisasi awal dan pusat cluster dilakukan dengan cara random
3. Menurut jarak antara objek, setiap titik data diposisikan di tengah cluster terdekat. Pada titik ini, metode *Euclidean* digunakan untuk menghitung jarak dengan menentukan kesamaan atau perbedaan data. Berikut rumusnya:

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2}$$

dimana:

“ $d(x, y)$ = ukuran ketidakmiripan

$x_i = (x_1, x_2, x_3, \dots, x_n)$ yaitu ukuran ketidakmiripan

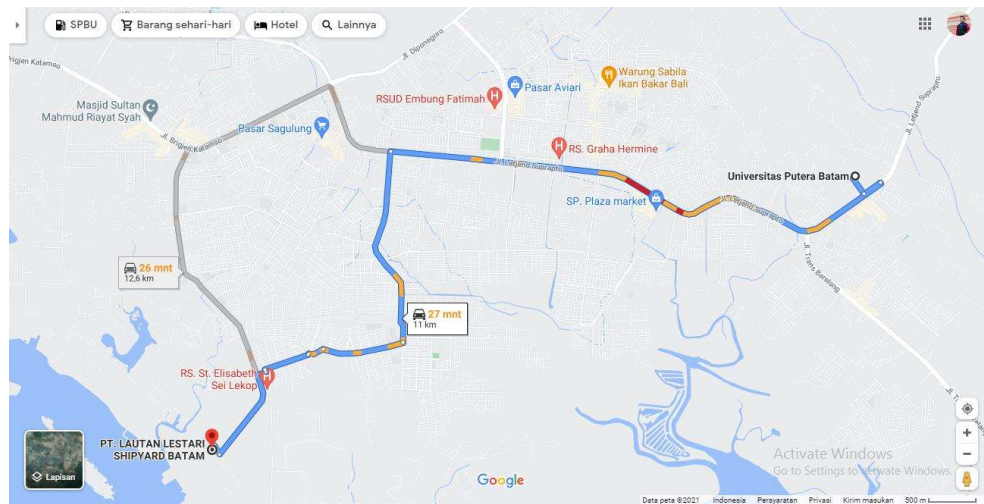
$y_i = (y_1, y_2, y_3, \dots, y_n)$ yaitu variabel pada titik pusat”

4. Rata-rata objek dalam cluster digunakan untuk menghitung pusat cluster. Selain itu, median juga dapat digunakan untuk membuat perhitungan.
5. Untuk menyelesaikan proses pengklasteran, harus menghitung ulang jarak antara setiap objek dan pusat cluster sampai cluster tetap tidak berubah.

3.5 Lokasi dan Jadwal Penelitian

3.5.1 Lokasi Penelitian

Penelitian ini dilakukan di PT Lautan Lestari Shipyards berlokasi di Sei Lekop Kecamatan Sagulung, dengan jarak tempuh sekitar 27 menit dari kampus UPB Tembesi.



Gambar 3.2. Peta Lokasi Penelitian

Sumber: (Peneliti, 2022)

3.5.2 Jadwal Penelitian

Penelitian ini membutuhkan waktu kurang lebih selama 3 bulan yaitu dilakukan pada bulan November 2021 hingga bulan Januari 2022. Berikut adalah jadwal yang peneliti lakukan.

Tabel 3.2 Jadwal Penelitian

Kegiatan	Nov. 2021				Des. 2021				Jan. 2022			
	1	2	3	4	1	2	3	4	1	2	3	4
Observasi												
Pengumpulan Data												
Pengumpulan Referensi												
Pengolahan dan Analisis Data												
Pengumpulan Laporan												

Sumber: (Peneliti, 2022)