

BAB III

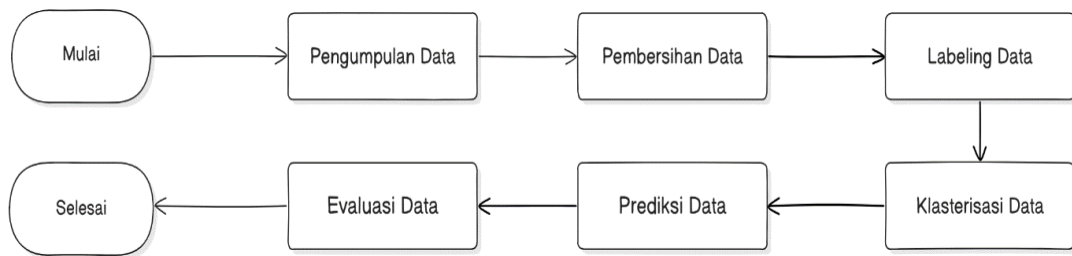
METODE PENELITIAN

3.1 Objek Penelitian

Penelitian dilakukan pada salah satu toko *vapestore* di Karawang. Data penelitian diambil langsung dari aplikasi pos toko yang meliputi informasi produk, jumlah penjualan, dan tanggal transaksi. Penelitian ini menggunakan algoritma *K-Means* dan *Linear Regression* untuk mengklasterisasi dan memprediksi data. Pengolahan data dilakukan menggunakan Phyton sebagai alat analisis utama untuk melakukan *preprocessing data*, eksplorasi, serta implementasi metode klasterisasi dan prediksi.

3.2 Prosedur Penelitian

Penelitian diawali dengan pengumpulan data penjualan dari *vapestore* XYZ, data yang telah diperoleh melalui laporan toko kemudian melalui tahap *preprocessing data*. Seperti pembersihan data, normalisasi, dan penghapusan nilai yang tidak relevan. Data yang telah di olah selanjutnya memasuki tahap berikutnya yaitu analisis dan klasterisasi data. Metode utama yang digunakan adalah algoritma *K-Means* dan *Linear Regression*, yang dibandingkan untuk mengevaluasi kinerja masing-masing dalam klasterisasi data penjualan dan prediksi tren dengan menggunakan metrik seperti akurasi, *MAE*, dan *RMSE*.(Adikara et al. 2017)



Gambar 3.1 Diagram Alur Penelitian

2.1.1

3.2.1. Pengumpulan Data

Peneliti datang ke toko *vapestore* XYZ, pengumpulan data dilakukan dengan memperoleh data langsung dari *vapestore*, yang mencakup laporan penjualan bulanan. Data tersebut akan dikumpulkan dalam format file CSV atau Excell untuk memudahkan analisis selanjutnya.

2.1.2

3.2.2. Pembersihan Data

Sebelum dilakukannya klasterisasi data, dataset penjualan toko perlu melalui tahapan *preprocessing data* untuk memastikan kualitas dataset yang lebih baik dan meningkatkan akurasi hasil analisis. Tahapan *preprocessing* yang akan dilaksanakan dalam penelitian meliputi beberapa proses utama seperti pembersihan data, cek duplikasi data, dan penghapusan data yang tidak relevan.

2.1.3

3.2.3. Labeling Data

Setelah tahap pembersihan data, peneliti melakukan tahap pelabelan data (*Labeling*). Pada tahap ini, setiap entri dalam dataset penjualan diberikan identifikasi yang jelas dan terstruktur, seperti penentuan nama produk, kategori produk, serta periode waktu penjualan. Proses pelabelan ini sangat penting untuk memastikan data memiliki atribut yang lengkap dan konsisten, sehingga memudahkan dalam proses analisis selanjutnya. Dengan data yang telah terlabel dengan baik, peneliti dapat melakukan segmentasi dan pengelompokan data secara lebih efektif, serta meningkatkan akurasi dalam penerapan model analisis.

Selain itu, pelabelan data juga berfungsi sebagai dasar dalam membangun model prediksi dan klasterisasi yang handal. Informasi yang terperinci mengenai karakteristik produk dan waktu penjualan memungkinkan peneliti untuk mengidentifikasi pola dan tren yang relevan. Tahap ini juga membantu dalam mengeliminasi ambiguitas data, sehingga hasil analisis yang diperoleh dapat lebih valid dan dapat dipertanggungjawabkan. Dengan demikian, pelabelan data merupakan langkah krusial yang menjembatani proses pembersihan data dan penerapan teknik analisis lanjutan dalam penelitian ini.

2.1.4

3.2.4. Klasterisasi Data

Pada tahap ini peneliti akan melakukan klasterisasi data yang sebelumnya telah dilabeli. Peneliti menggunakan algoritma *K-Means* sebagai algoritma klasterisasi data, proses dimulai dengan pemrograman menggunakan *Python* dan memanfaatkan pustaka seperti *Scikit-learn*. Dalam langkah ini, peneliti menyiapkan data yang telah diproses sebelumnya dan menentukan parameter optimal, termasuk jumlah klaster (k) yang akan digunakan. Pemilihan nilai k dilakukan dengan metode *Elbow* untuk mengidentifikasi titik optimal di mana penambahan klaster tidak lagi memberikan penurunan signifikan dalam *within-cluster sum of squares* (WCSS). Selain itu, peneliti melakukan inisialisasi centroid awal secara acak dan iterasi hingga tercapai konvergensi, di mana posisi *centroid* tidak lagi mengalami perubahan signifikan.

2.1.5

3.2.5. Prediksi Data

Dataset yang telah diklasterisasi akan melalui tahap prediksi dengan menggunakan algoritma regresi linier. Peneliti menganalisis hubungan antara variabel independen dan variabel dependen untuk memprediksi penjualan di masa depan. Proses ini melibatkan pembuatan model matematis yang menggambarkan hubungan linear antara variabel-variabel tersebut. Data historis yang telah dikumpulkan dan diproses digunakan untuk melatih model regresi linier, dengan memastikan asumsi dasar seperti linearitas, homoskedastisitas, dan normalitas residual terpenuhi.

2.1.6

3.2.6. Evaluasi Data

Setelah model diimplementasikan, tahap pengujian model dilakukan untuk mengevaluasi kinerja model yang telah dibuat. Pada tahap ini, metrik evaluasi seperti *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE) digunakan untuk mengevaluasi algoritma regresi linier.