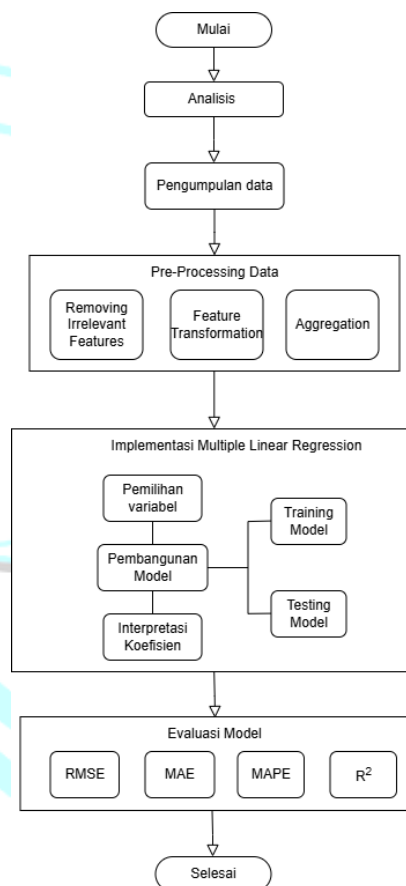


BAB III

METODE PENELITIAN

Penelitian ini dilakukan melalui tahapan yang dirancang secara sistematis untuk menghasilkan model klasifikasi pengeluaran perusahaan yang akurat menggunakan algoritma *Multiple Linear Regression*. Tahapan penelitian dapat dilihat pada Gambar 1. (Gandhi et al., 2021)



Gambar 1. Tahapan Penelitian

3.1 Analisis

Tahapan Analisis bertujuan untuk memahami permasalahan pengelolaan pengeluaran harian, mengidentifikasi pola pengeluaran, alokasi dana, serta kendala-kendala yang menyebabkan pengeluaran tidak terstruktur. Peninjauan literatur dilakukan untuk memastikan pendekatan yang digunakan sesuai dengan permasalahan dan kebutuhan perusahaan. Hasil dari proses ini menjadi landasan dalam merancang langkah-langkah penelitian selanjutnya.

3.2 Pengumpulan Data

Penelitian ini dimulai dengan mengumpulkan data pengeluaran harian PT. Harum Maju Mapan dari Januari hingga November 2024. Data meliputi tanggal transaksi, nama transaksi, kuantiti, jumlah pengeluaran, dan kode transaksi yang bersumber dari laporan digital perusahaan dalam format excel. Sebanyak 7.478 data transaksi terkumpul dan menjadi dasar analisis lebih lanjut, dengan penekanan pada kualitas yang kelengkapan data.

3.3 Pre-Processing Data

Setelah data terkumpul, tahap berikutnya adalah pra-pemrosesan data (*preprocessing*), yang bertujuan untuk meningkatkan kualitas data sebelum digunakan dalam algoritma *Linear Regression*. Pada tahap ini, terdapat beberapa metode *preprocessing* yang diterapkan, yaitu:

a. *Removing Irrelevant Features*

Fitur-fitur yang tidak relevan seperti nama dan kuantiti transaksi dihapus karena tidak memiliki kontribusi terhadap analisis yang dilakukan serta memiliki format yang tidak konsisten.

b. *Feature Transformation*

Pada langkah ini, tanggal transaksi diubah menjadi format numerik. Proses ini dilakukan agar algoritma yang dapat membaca dan memproses data dengan lebih baik.

c. *Aggregation*

Data transaksi diringkas dengan mengelompokkan nominal pengeluaran berdasarkan kategori untuk setiap tanggal. Proses ini membantu menyederhanakan data dengan tetap mempertahankan informasi penting yang diperlukan dalam analisis.

Setelah *preprocessing*, jumlah data yang awalnya sebanyak 7.478 baris berkurang menjadi 1.424 baris. Tahap ini bertujuan memastikan bahwa data yang digunakan sudah dalam kondisi optimal dan siap untuk membangun model *Multiple Linear Regression* yang akurat dalam memprediksi pengeluaran.

3.4 Implementasi *Multiple Linear Regression*

Pada tahapan ini, penelitian difokuskan pada penerapan algoritma regresi linier berganda untuk menganalisis hubungan antara variabel independen dan variabel dependen dalam dataset yang digunakan.

3.4.1 Pemilihan Variabel

Dataset yang digunakan dalam penelitian ini terdiri dari sejumlah variabel, di mana variabel-variabel tersebut dibagi menjadi dua kategori utama:

a. Variabel Dependen (*Target*)

Variabel yang ingin di prediksi, seperti tanggal transaksi.

b. Variabel Independen (*Predictor*)

Variabel yang digunakan untuk memprediksi nilai target, misalnya bulan, kategori transaksi, dan nominal transaksi. Pemilihan variabel dilakukan berdasarkan analisis awal terhadap data serta literatur terkait untuk memastikan relevansi dan signifikansi variabel terhadap target.

3.4.2 Pembangunan Model

Pembangunan model dimulai dari mempersiapkan dataset yang telah diproses. Selanjutnya dataset dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian.

a. Training Model

Pada tahap ini, dataset dibagi menjadi dua subset: data pelatihan (*training data*) dan data pengujian (*testing data*). Model dilatih menggunakan data pelatihan untuk menghitung koefisien regresi melalui kuadrat terkecil (*least squares method*). Proses ini bertujuan untuk meminimalkan *error residual* antara nilai prediksi dan nilai aktual.

b. Testing Model

Model yang telah dilatih diuji menggunakan data pengujian untuk mengevaluasi performa prediksi. Nilai prediksi yang dihasilkan dibandingkan dengan nilai aktual dalam data pengujian untuk mengukur tingkat akurasi dan keandalan model

3.4.3 Interpretasi Koefisien

Koefisien regresi digunakan untuk memahami pengaruh variabel independen terhadap target. Koefisien positif menunjukkan hubungan langsung, sementara koefisien negative menunjukkan hubungan berlawanan, membantu mengidentifikasi variabel yang paling signifikan.

3.5 Evaluasi Model

Tahap selanjutnya adalah evaluasi model untuk mengukur performa prediksi yang dihasilkan oleh algoritma *Multiple Linear Regression*. Evaluasi dilakukan dengan menggunakan metrik-metrik standar, antara lain:

a. **Mean Squared Error (MSE)**

Mengukur rata-rata kuadrat dari *error* residual, memberikan bobot lebih besar pada *error* yang lebih besar.

b. **Root Mean Squared Error (RMSE)**

Akar dari MSE, memberikan interpretasi *error* dalam satuan yang sama dengan variabel target.

c. **Mean Absolute Error (MAE) :**

Mengukur rata-rata nilai absolut dari *error* residual, memberikan penilaian yang lebih seimbang terhadap *error*.

d. **R-squared (R^2)**

Mengukur proporsi variansi dalam data yang dapat dijelaskan oleh model.

Hasil evaluasi model diinterpretasikan untuk memahami pola pengeluaran harian membandingkan nilai prediksi dan nilai aktual, serta identifikasi tren pengeluaran. Penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam kepada PT. Harum Maju Mapan serta mendukung perencanaan anggaran yang lebih tepat, efisien, dan strategis melalui implementasi algoritma *Multiple Linear Regression*.