BAB I
PENDAHULUAN

1.1 Latar Belakang

Industri ritel kosmetik merupakan sektor dengan pertumbuhan yang sangat dinamis, dipengaruhi oleh tiga faktor utama: pergeseran tren kecantikan yang cepat, perubahan preferensi dan perilaku konsumen, serta intensitas kompetisi promosi antar merek. Permintaan terhadap produk kosmetik menunjukkan pola yang sangat fluktuatif, khususnya pada kategori produk seperti *skincare*, *parfum*, dan *bodycare* yang sensitif terhadap musim dan kampanye pemasaran (Adriani, Ni Putu Erica Puspita, 2025). Menurut Agustina et al., (2025) Transaksi harian variatif karena segmentasi pelanggan dan sensitivitas diskon dan *loyalitas*, menjadikan data non-linear memerlukan analisis kompleks.

Dalam praktiknya, sebagian besar toko ritel kosmetik masih menggunakan pendekatan konvensional berbasis data *historis* sederhana atau intuisi, yang terbukti kurang mampu menangkap pola penjualan nonlinier dan interaksi kompleks antar variabel (Perin et al., 2025). Kondisi ini berdampak langsung pada pengelolaan persediaan, di mana fluktuasi permintaan yang tinggi dapat memicu terjadinya *overstock* maupun *stockout*. Kedua kondisi tersebut berpotensi menimbulkan kerugian operasional dan menurunkan tingkat kepuasan pelanggan. Oleh karena itu, diperlukan prediksi penjualan yang akurat sebagai dasar pengambilan keputusan dalam manajemen stok yang efisien.

Penjualan pada toko ritel kosmetik dipengaruhi oleh berbagai indikator yang berasal dari aktivitas transaksi dan perilaku konsumen. Indikator tersebut meliputi harga produk, diskon atau promosi, jumlah transaksi pelanggan, jumlah produk terjual, kategori produk, merek, serta waktu penjualan seperti hari libur, dan akhir pekan. Perubahan pada indikator-indikator tersebut menyebabkan nilai penjualan bersifat fluktuatif dan sulit diprediksi menggunakan cara konvensional. Oleh karena itu, diperlukan pendekatan berbasis data mining dan machine learning untuk menganalisis pola penjualan secara lebih akurat.

Sejumlah penelitian terdahulu menunjukkan bahwa pendekatan *machine learning*, khususnya algoritma Random Forest, mampu memberikan tingkat akurasi

prediksi yang lebih baik dibandingkan metode manual atau statistik konvensional dalam memodelkan data penjualan ritel yang kompleks dan bervariasi (Zulfia et al., 2025). Penelitian Rahmadewi et al., 2025; Falah et al, 2025) pada ritel kosmetik dan parfum menunjukkan bahwa algoritma Random Forest memiliki kinerja prediksi yang lebih unggul dibandingkan Regresi Linier dalam memodelkan data penjualan ritel berbasis *point of sales.* Berdasarkan evaluasi menggunakan metrik MAE, RMSE, MAPE, dan R^2, Random Forest menghasilkan nilai kesalahan yang lebih rendah dan kemampuan penjelasan variansi data yang lebih tinggi. Secara rata-rata, Random Forest memperoleh MAE sebesar 4,121, RMSE sebesar 5,246, MAPE sebesar 0,040, dan R^2 sebesar 0,519, sedangkan Regresi Linier menghasilkan MAE 7,269, RMSE 9,196, MAPE 0,071, dan R^2 0,251. Temuan ini menegaskan bahwa Random Forest lebih efektif dalam menangkap pola penjualan yang kompleks. Penelitian sebelumnya juga telah mengimplementasikan Random Forest untuk prediksi penjualan dan manajemen stok pada berbagai jenis ritel umum, seperti *minimarket*, usaha *frozen food*, dan ritel minuman berbasis *point of sales*. Hasil penelitian tersebut menunjukkan bahwa Random Forest mampu menghasilkan nilai kesalahan prediksi yang lebih rendah dan stabil, serta efektif dalam mendukung perencanaan stok dan pengambilan keputusan *restock* (Kurniawan et al., 2024; Hany et al., 2025; Rigorastio & Nusantara, 2025;)

Meskipun berbagai penelitian telah mengimplementasikan Random Forest untuk prediksi penjualan pada ritel umum seperti *minimarket*, *frozen food*, dan ritel minuman, penerapannya pada konteks ritel kosmetik masih relatif terbatas, khususnya yang menganalisis kontribusi masing-masing variabel terhadap penjualan. Oleh karena itu, penelitian ini mengintegrasikan variabel promosi, kategori, dan waktu dalam model Random Forest, evaluasi kinerja menggunakan MAE, RMSE, dan MAPE, serta menganalisis *feature importance* untuk mengidentifikasi faktor utama yang memengaruhi penjualan. Sebagai kontribusi praktis, model yang telah divalidasi di implementasikan dalam aplikasi *streamlit* dan mendukung keputusan bisnis pada toko ritel kosmetik.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana menerapkan model prediksi penjualan harian pada toko ritel kosmetik menggunakan algoritma Random Forest?
2. Bagaimana evaluasi kinerja model Random Forest dalam memprediksi penjualan harian berdasarkan nilai MAE, RMSE, dan MAPE?
3. Variabel apa saja yang memiliki pengaruh paling signifikan terhadap penjualan ritel kosmetik berdasarkan hasil analisis *feature importance* pada model Random Forest?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun model prediksi total penjualan harian menggunakan Random Forest.
2. Mengevaluasi performa model Random Forest menggunakan metrik regresi seperti nilai MAE, RMSE, dan MAPE untuk menentukan akurasi dan ketepatan hasil prediksi.
3. Menganalisis variabel yang paling berpengaruh terhadap penjualan ritel kosmetik berdasarkan hasil analisis *feature importance*.

1.4 Manfaat

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Memberikan kontribusi pada pengembangan literatur mengenai penggunaan algoritma Random Forest dalam memprediksi penjualan pada sektor ritel kosmetik yang memiliki karakteristik data non-linear.
2. Menyediakan model prediksi penjualan harian yang dapat membantu pengelola toko ritel kosmetik. Selain itu penelitian ini juga menghadirkan aplikasi prediksi penjualan berbasis Streamlit yang dapat digunakan secara langsung oleh pelaku usaha untuk mendukung pengambilan keputusan secara cepat, akurat, dan berbasis data