

Saya mendefinisikan churn sebagai kondisi ketika pelanggan berhenti bertransaksi dalam periode waktu tertentu setelah sebelumnya aktif. Untuk memprediksi kemungkinan churn, saya akan menggunakan pendekatan klasifikasi biner, karena target yang ingin diprediksi bersifat kategorikal — pelanggan akan diklasifikasikan sebagai churned atau active. Model seperti Logistic Regression atau Random Forest akan menjadi titik awal karena mudah diinterpretasi dan efektif untuk baseline prediksi.

Sebelum membangun model, saya akan membuat sejumlah fitur turunan dari data transaksi untuk merepresentasikan perilaku pelanggan, seperti: $\text{revenue} = \text{quantity} * \text{price}$, avg order_value per bulan, unique_customers, revenue_per_customer, serta metrik pertumbuhan seperti month-over-month (MoM) growth dan cumulative revenue (YTD). Saya juga akan menambahkan fitur berbasis waktu (weekday, hour) dan perilaku (is_returning customer) untuk menangkap pola frekuensi dan loyalitas pelanggan. Sebagai eksplorasi tambahan, saya akan melakukan clustering (misal dengan PCA + KMeans) untuk menemukan segmen perilaku pelanggan dan memahami pola churn sebelum membangun model klasifikasi utama.