

# **Analisis Data E-Commerce untuk Pertumbuhan Bisnis**

Wahyu



## Produk Terlaris

- 5 produk terlaris, yakni : Laptop Pro 15, Smartphone X, Sepatu Lari, Mechanical Keyboard, Tablet 10 inch
- Pada Laptop Pro 15 merupakan produk terlaris dengan total pendapatan mencapai Rp. 390.000.000

	product_id	product_name	total_revenue
0	PROD001	Laptop Pro 15	390000000
1	PROD002	Smartphone X	238000000
2	PROD008	Sepatu Lari	18750000
3	PROD004	Mechanical Keyboard	14400000
4	PROD041	Tablet 10 inch	13500000



## Pelanggan Paling Berharga

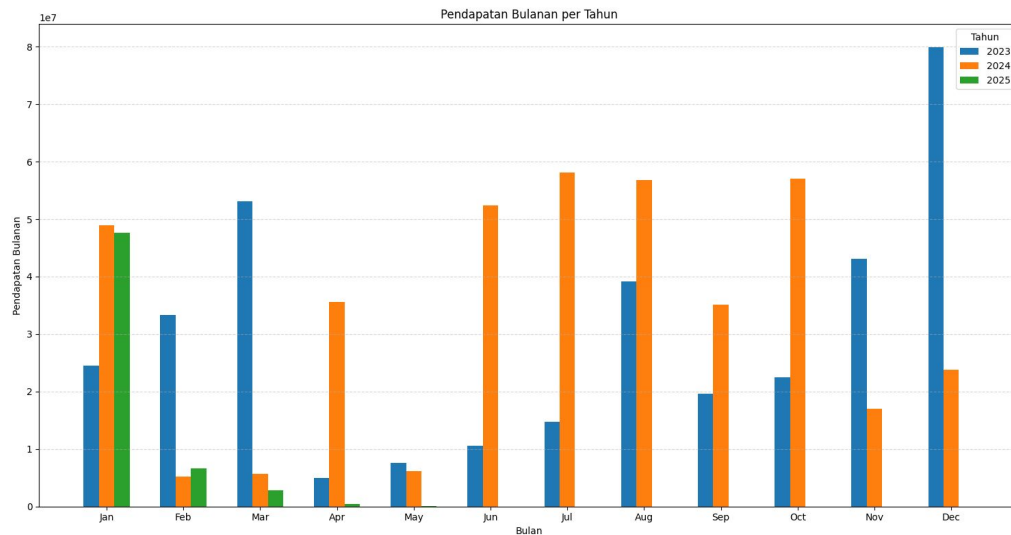
- Customer dengan ID **CUST005** merupakan customer paling royal yang telah membelanjakan di Toko RetailKita dengan total **Rp. 111.600.000**
- Lalu diikuti dengan customer dengan ID **CUST033** menghabiskan **Rp. 90.600.000**, dan dengan ID **CUST021** menghabiskan **Rp. 62.500.000**

	customer_id	total_spent
0	CUST005	111600000
1	CUST033	90600000
2	CUST021	62500000
3	CUST001	56590000
4	CUST029	52200000
5	CUST041	51800000
6	CUST018	50200000
7	CUST037	31500000
8	CUST025	31300000
9	CUST071	30000000

## Tren Penjualan Bulanan

- Total penjualan tertinggi terdapat pada tahun 2024
- Penjualan bulanan **tertinggi** terdapat pada bulan **Desember 2023**, sedangkan total penjualan bulanan **terendah** terdapat pada bulan **Mei 2025**

	sale_year	average_monthly_revenue
0	2023	29.43 Juta
1	2024	33.48 Juta
2	2025	11.51 Juta





# Membangun Model Identifikasi Customer Churn

Definisi **churn** adalah pelanggan yang **tidak bertransaksi/berlangganan** dalam **3 bulan** terakhir. Namun terdapat asumsi yang dapat dijadikan kemungkinan terjadinya customer churn:

- Nilai recency yang tinggi
- Jumlah transaksi historis rendah
- Pelanggan dengan revenue rendah
- Pelanggan dengan jeda transaksi yang panjang
- Variasi pembelian produk yang rendah
- Kuantitas kecil
- Pelanggan baru (masa bergabung singkat)
- Kategori produk
- Lokasi



# Membangun Model Identifikasi Customer Churn

Prediksi customer churn merupakan masalah klasifikasi biner karena hanya memiliki dua kelas, yaitu churn dan tidak churn. Model **Random Forest Classifier** dipilih sebagai titik awal karena mampu menangkap hubungan non-linier antar variabel, cukup tahan terhadap overfitting, serta dapat bekerja baik pada data tabular yang berisi kombinasi fitur numerik dan kategorikal. Selain itu, Random Forest juga menyediakan informasi mengenai **feature importance**, sehingga memudahkan dalam mengidentifikasi faktor-faktor utama yang memengaruhi churn pelanggan.



# Membangun Model Identifikasi Customer Churn

Fitur yang digunakan seperti:

- recency\_days
- transaction\_count
- total\_revenue
- avg\_time\_between\_transactions\_days
- unique\_products\_count
- total\_quantity
- duration\_since\_joining\_days
- category
- location



## Membangun Model Identifikasi Customer Churn

Setelah mendefinsikan fitur, model dilatih dengan Random Forest dan didapatkan skor akurasi mencapai 100%

```
Accuracy: 1.0000  
Precision: 1.0000  
Recall: 1.0000  
F1-score: 1.0000  
ROC AUC Score: 1.0000
```