

# E-COMMERCE CHURN

PRESENTED BY MUHAMAD WILDAN TRISIANLY

# Business Understanding

Perusahaan e-commerce mengalami masalah customer churn (pelanggan berhenti menggunakan layanan). Jika churn tinggi, perusahaan bisa kehilangan pendapatan dan biaya akuisisi pelanggan baru jadi lebih tinggi.

**Stake Holder :**

- Tim Marketing & Retention
- Manajemen Perusahaan:

**Target :**

- Memprediksi pelanggan yang akan churn agar bisa diberi promo untuk mencegah churn.



# Data Understanding

Dataset= data\_ecommerce\_customer\_churn.csv

Terdiri dari = 3941 baris

Tiap Baris Mewakili 1 customer

```
] 1 df.shape
    ✓ 0.0s

(3941, 11)
```

## Features

- **Tenure:** Tenure of a customer in the company..
- **WarehouseToHome:** Distance between the warehouse to the customer's home.
- **NumberOfDeviceRegistered:** Total number of deceives is registered on a particular customer.
- **PreferedOrderCat:** Preferred order category of a customer in the last month.
- **SatisfactionScore:** Satisfactory score of a customer on service.
- **MaritalStatus:** Marital status of a customer.
- **NumberOfAddress:** Total number of added on a particular customer.
- **Complaint:** Any complaint has been raised in the last month.
- **DaySinceLastOrder:** Day since last order by customer.
- **CashbackAmount:** Average cashback in last month
- **Churn:** Churn flag.

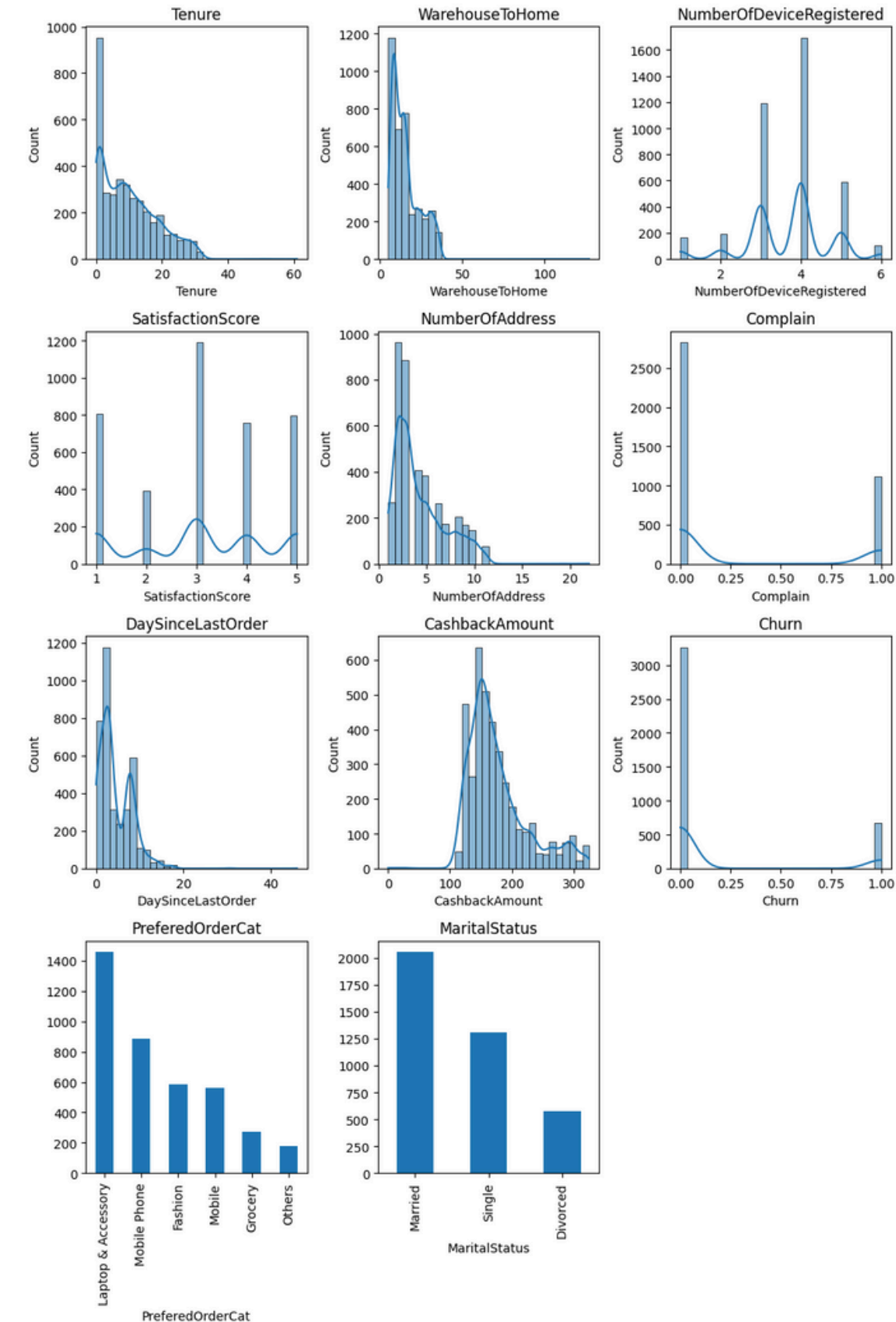
# Data Exploration

```
1 df.isna().sum() / len(df) * 100
```

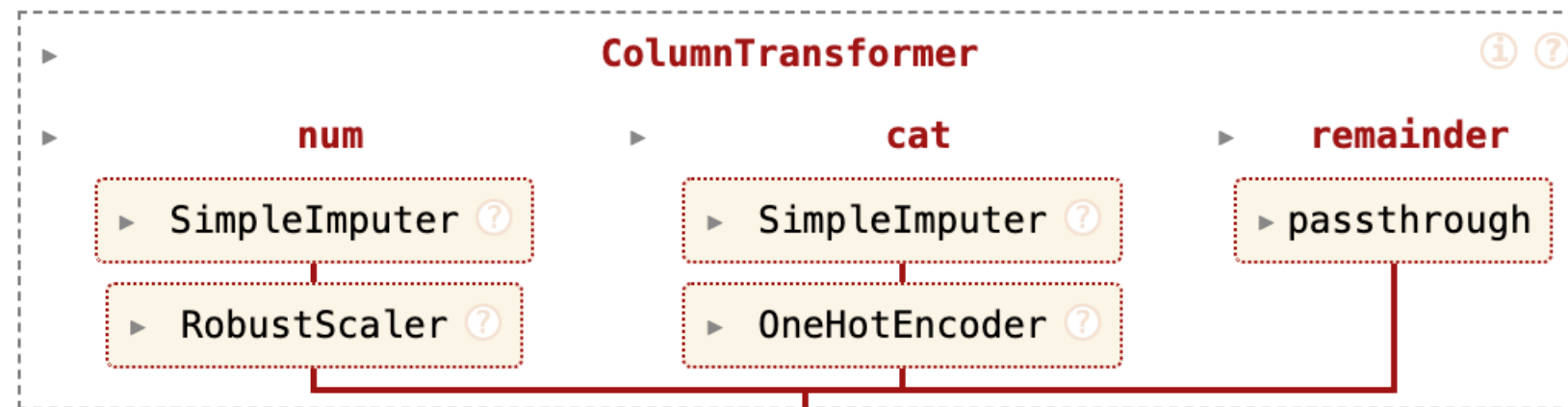
```
2
```

```
✓ 0.0s
```

Tenure	4.922608
WarehouseToHome	4.288252
NumberOfDeviceRegistered	0.000000
PreferredOrderCat	0.000000
SatisfactionScore	0.000000
MaritalStatus	0.000000
NumberOfAddress	0.000000
Complain	0.000000
DaySinceLastOrder	5.404720
CashbackAmount	0.000000
Churn	0.000000
dtype:	float64



# Pre-processing



## Preprocessing

- **Numerik:**
  - Imputasi nilai kosong dengan **median** untuk mengurangi pengaruh outlier ekstrem.
  - Scaling menggunakan **RobustScaler** agar skala data tidak terdistorsi oleh outlier.
- **Kategorikal:**
  - Imputasi nilai kosong dengan **modus**.
  - Encoding menggunakan **OneHotEncoder** agar bisa diproses oleh model machine learning.

# Cross Validation

	roc_auc	pr_auc	precision	recall	f1
model					
RandomForest	0.9558	0.8405	0.8430	0.6235	0.7153
XGBoost	0.9473	0.8309	0.7530	0.8016	0.7751
LogisticRegression	0.8801	0.6781	0.4499	0.8145	0.5796
DecisionTree	0.8526	0.6247	0.5912	0.7644	0.6655



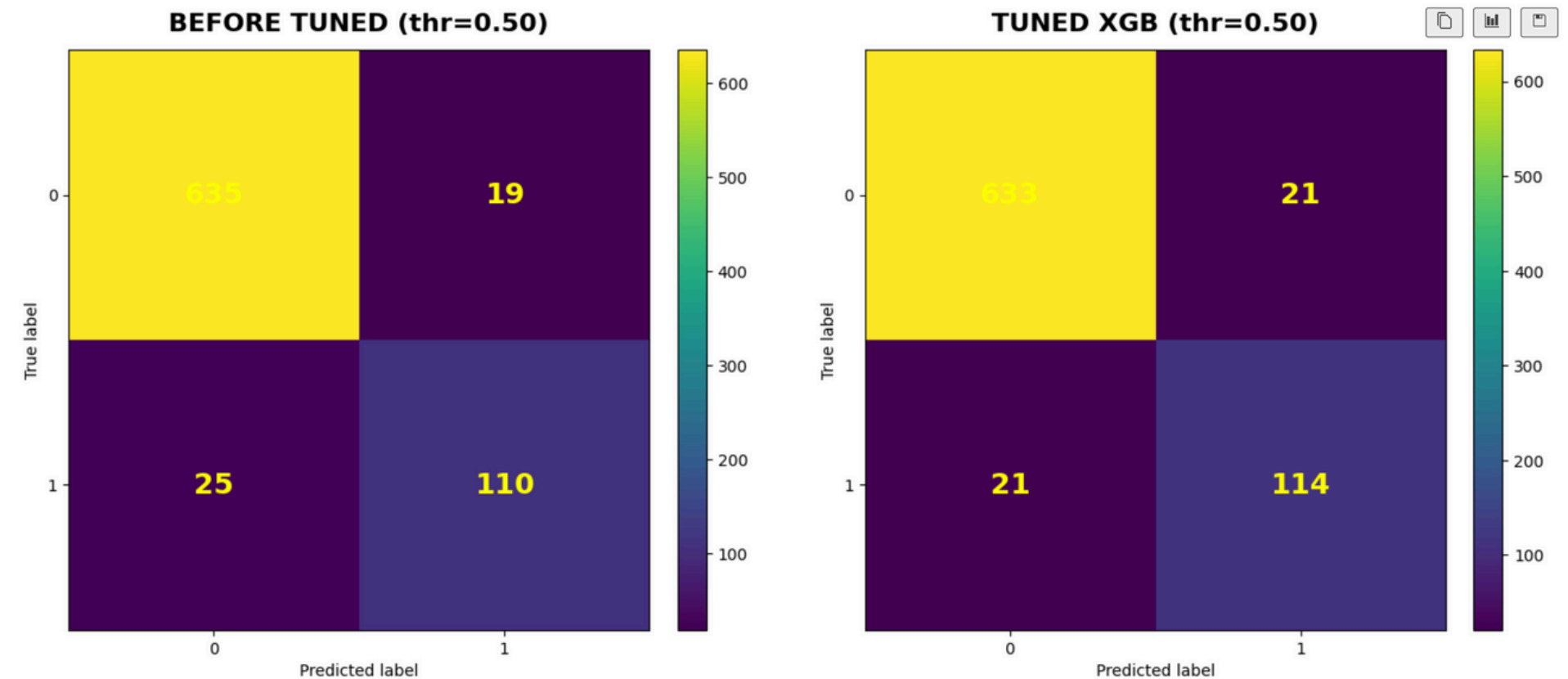
# Before dan After Tuning

=== Befor Tuned XGB @0.50 ===

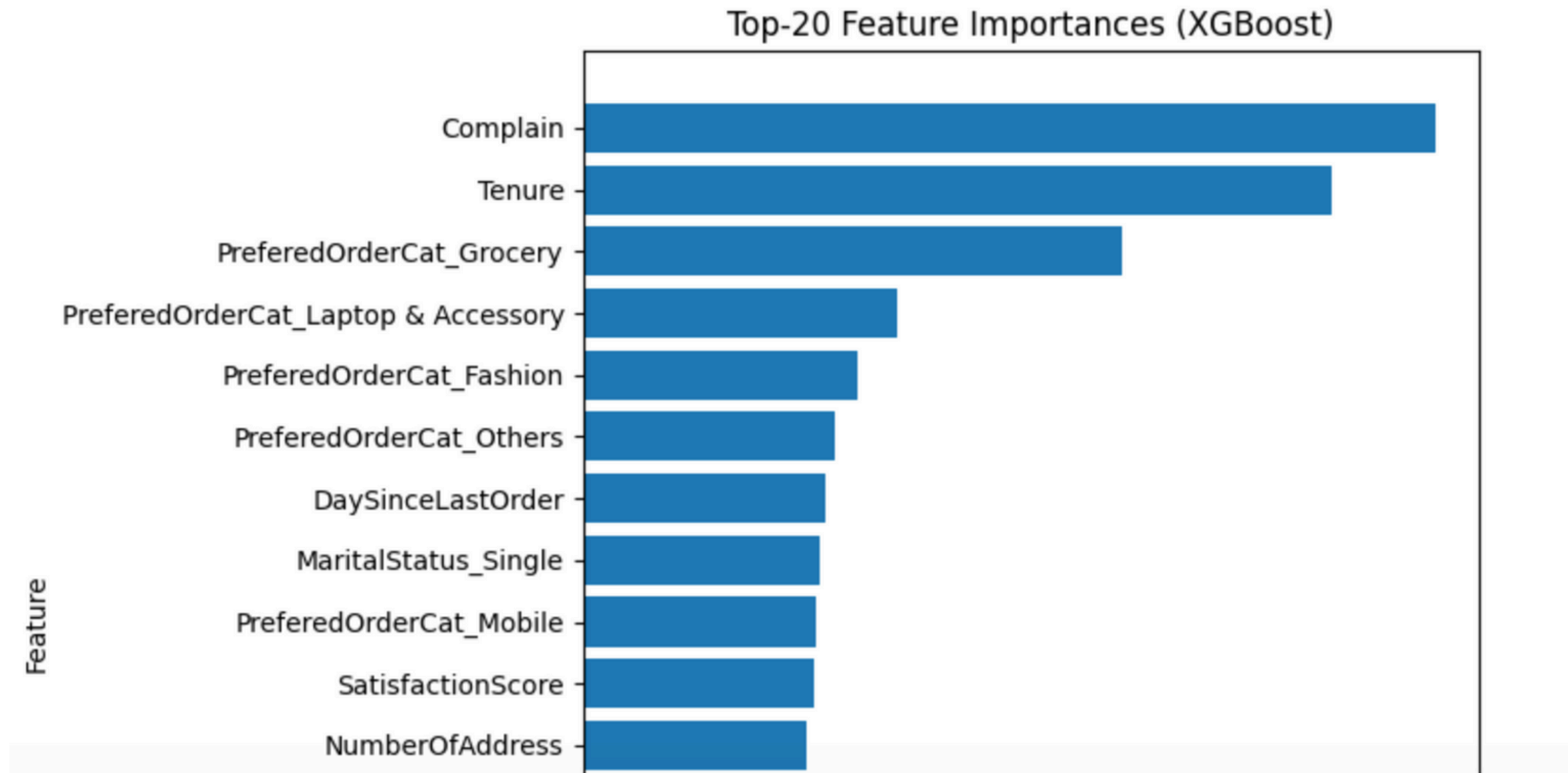
	precision	recall	f1-score	support
0	0.9621	0.9709	0.9665	654
1	0.8527	0.8148	0.8333	135
accuracy			0.9442	789
macro avg	0.9074	0.8929	0.8999	789
weighted avg	0.9434	0.9442	0.9437	789

=== TUNED XGB @0.50 ===

	precision	recall	f1-score	support
0	0.9679	0.9679	0.9679	654
1	0.8444	0.8444	0.8444	135
accuracy			0.9468	789
macro avg	0.9062	0.9062	0.9062	789
weighted avg	0.9468	0.9468	0.9468	789



# Feature Importance





# Kesimpulan

## 📌 Dari Sisi Model

- Model XGBoost memberikan performa terbaik dibanding model lain.
- Setelah tuning:
- Recall naik 0.8148 → 0.8444
- F1 naik 0.8333 → 0.8444
- Akurasi naik 0.9442 → 0.9468
- FN turun (25 → 21) → lebih banyak pelanggan churn berhasil terdeteksi
- TP naik (110 → 114) → prediksi churn semakin akurat

## 📌 Dari Sisi Bisnis

- Model mampu mendeteksi pelanggan yang berisiko churn secara dini
- Fitur paling berpengaruh: Complain, Tenure, PreferredOrderCat, DaySinceLastOrder, SatisfactionScore
- Pelanggan yang baru bergabung, pernah komplain, atau lama tidak belanja adalah kelompok paling berisiko churn

# Rekomendasi

## 📌 Untuk Bisnis

- Gunakan model secara bulanan untuk mendeteksi pelanggan berisiko churn
- Fokuskan tindakan pada:
  - Pelanggan baru (tenure rendah)
  - Pelanggan yang pernah komplain
  - Pelanggan lama tidak belanja
- Berikan promo, loyalty point, atau voucher untuk mempertahankan pelanggan ini
- Bangun dashboard internal agar tim marketing bisa langsung melihat daftar pelanggan berisiko churn

## 📌 Untuk Model

- Retrain model setiap 3–6 bulan agar menyesuaikan pola perilaku pelanggan terbaru
- Pantau metrik recall & FN sebagai indikator utama efektivitas model
- Sesuaikan threshold jika distribusi churn berubah, untuk menjaga keseimbangan recall dan precision
- Tambahkan fitur baru (misal: frekuensi login, klik, waktu respon komplain) untuk meningkatkan akurasi ke depan