

Predictive Analytics for Customer Churn

SwiftBuy

Leveraging Machine Learning to Improve Customer Retention

Patrick Jonathan

Data Analyst
SwiftBuy





Table of Content

- Executive Summary
- Data Understanding
- Exploratory Data Analysis
- Data Cleaning & Feature Engineering
- Modeling & Evaluation
- Business Impact
- Conclusion & Recommendation

Background

SwiftBuy adalah platform e-commerce yang menjual berbagai produk melalui marketplace online. Pendapatan diperoleh dari komisi penjualan, layanan premium untuk penjual, dan iklan di platform. Dengan persaingan yang ketat dan biaya akuisisi pelanggan baru yang tinggi, fokus bisnis kini bergeser pada retensi pelanggan untuk menjaga pertumbuhan pendapatan.





What's the Problem?



Goals

Membangun model prediktif berbasis data historis untuk mengidentifikasi pelanggan berisiko churn, memahami faktor penyebabnya, dan memberikan insight bagi strategi retensi yang tepat sasaran.

Analytical Approach

Proyek ini merupakan klasifikasi biner churn/tidak churn dengan tahapan data preparation, pengujian algoritma, tuning model terbaik, dan menghasilkan model prediksi churn beserta feature importance untuk insight bisnis.

Metric Selection

F2-score dipilih karena lebih menekankan recall untuk meminimalkan kerugian besar akibat false negative (kehilangan pelanggan) yang biayanya 4 kali lebih tinggi dibanding false positive (intervensi salah sasaran)

Data Understanding

FEATURE	DESCRIPTION
Tenure	Lama Pelanggan bergabung
WarehouseToHome	Jarak rumah pelanggan ke gudang (Km)
NumberOfDeviceRegistered	Jumlah perangkat yang terdaftar
PreferedOrderCat	Kategori produk yang sering dipesan
SatisfactionScore	Skor kepuasan pelanggan
MaritalStatus	Status pernikahan pelanggan
NumberOfAddress	Jumlah alamat yang pernah digunakan
Complain	Apakah pelanggan pernah mengajukan komplain (0 = tidak, 1 = ya)
DaySinceLastOrder	Hari sejak pesanan terakhir
CashbackAmount	Total cashback yang diterima
Churn	Apakah pelanggan churn atau tidak (0 = tidak churn, 1 = churn)



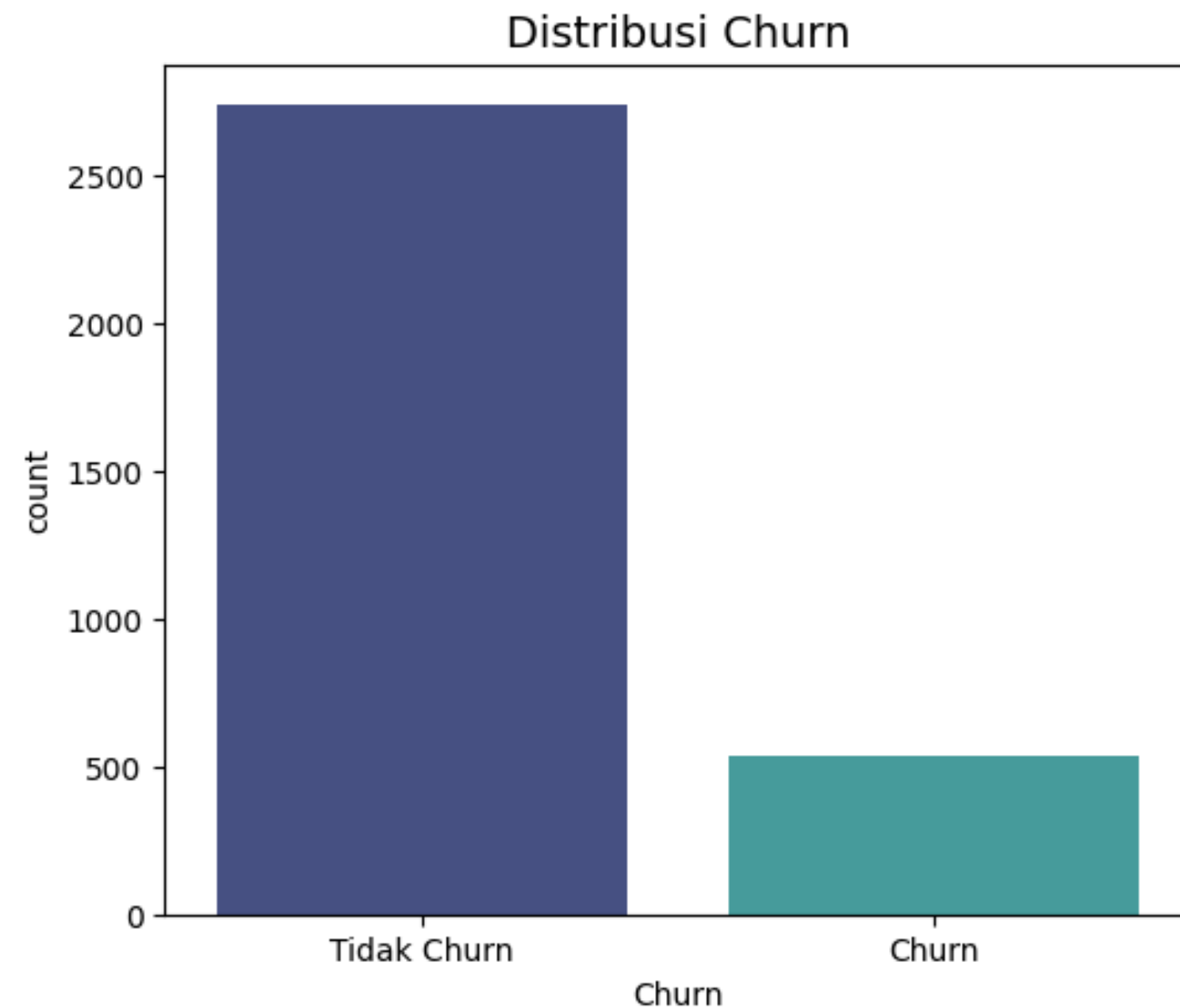
Observation: Setiap baris mewakili satu pelanggan SwiftBuy.



Data Source: Data transaksi & profil pelanggan internal SwiftBuy.

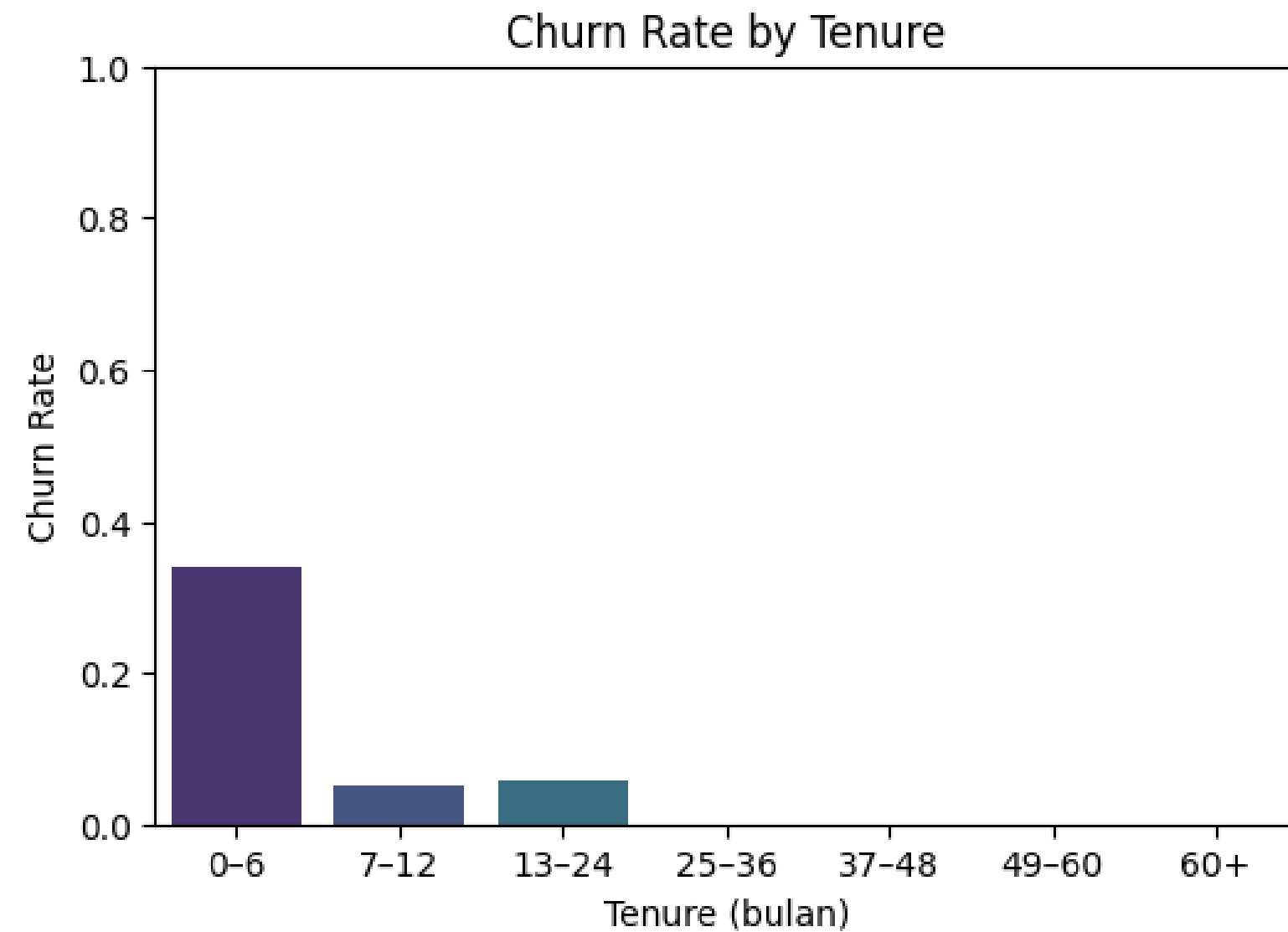
Exploratory Data Analysis

Data Tidak Seimbang: Churn Lebih Sedikit, Dampaknya Tetap Besar



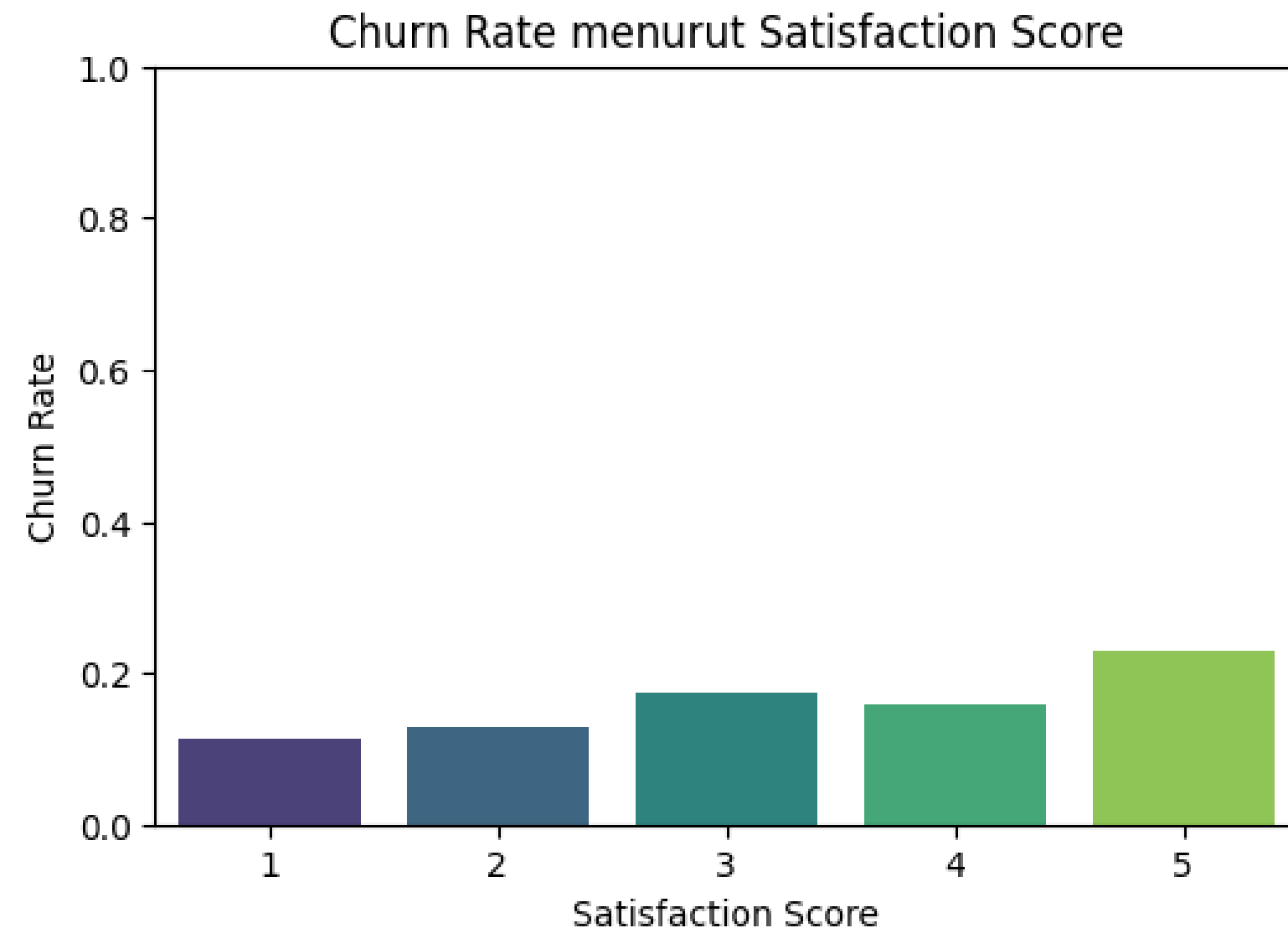
Walaupun jumlah pelanggan churn lebih sedikit, kehilangan mereka berdampak signifikan pada pendapatan dan biaya akuisisi

Pelanggan Baru Lebih Berisiko Churn



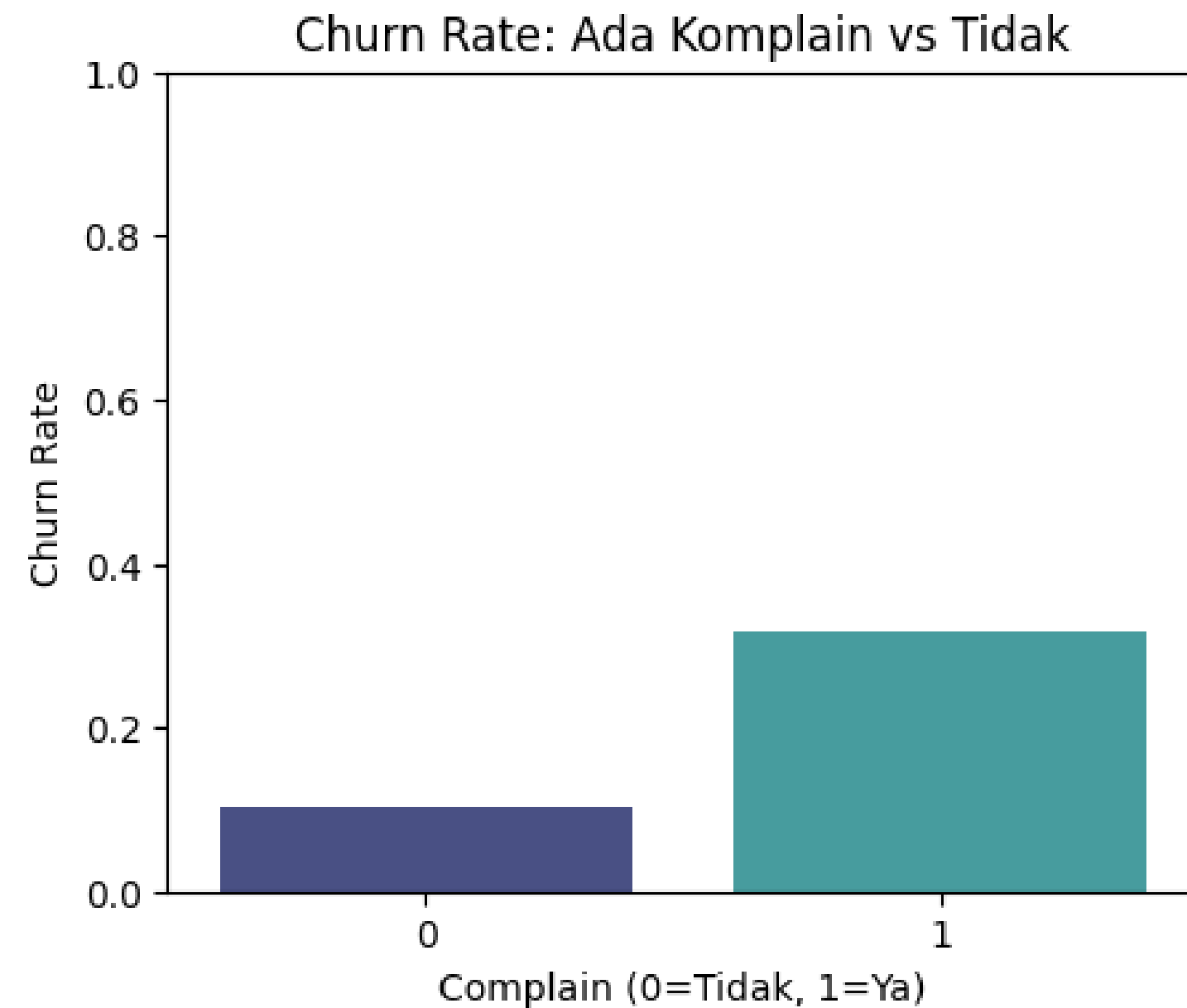
Pelanggan dengan masa bergabung 0-6 bulan memiliki churn rate tertinggi, menunjukkan bahwa pelanggan baru adalah kelompok paling rentan untuk berhenti menggunakan layanan. Setelah melewati 12 bulan, churn rate menurun signifikan.

Kepuasan Rendah Berkorelasi dengan Churn Tinggi



Churn rate menurun seiring meningkatnya skor kepuasan. Pelanggan dengan Satisfaction Score rendah (1-2) memiliki risiko churn jauh lebih tinggi dibandingkan mereka yang memberi skor 4-5.

Pelanggan yang Sering Mengeluh Cenderung Churn



Pelanggan yang pernah mengajukan keluhan (Complain = 1) memiliki churn rate yang sangat tinggi dibandingkan pelanggan tanpa keluhan. Hal ini menunjukkan bahwa pengalaman buruk yang tidak segera ditangani meningkatkan risiko churn.

Data Cleaning & Feature Engineering



Preprocessing Flow

- Data Splitting - train_test_split dengan stratify label → distribusi kelas seimbang.
- Missing Values - Imputasi median untuk numerik → tahan outlier.
- Feature Scaling - RobustScaler → stabil terhadap outlier.
- Encoding - OneHotEncoder (ignore unknown) → kategori aman di data baru.
- Imbalanced Data - SMOTE di train → seimbangkan kelas.
- Pipeline - ImbPipeline gabung semua tahap → konsisten & siap cross-validation.

Modeling

Cross-Validation

MODEL	RESAMPLING	MEAN SCORE	STANDARD DEVIATION
Logistic Regression	No Handling	0.686	0.034
Logistic Regression	SMOTE	0.682	0.035
K-Nearest Neighbors	No Handling	0.352	0.034
K-Nearest Neighbors	SMOTE	0.652	0.024
Decision Tree	No Handling	0.606	0.062
Decision Tree	SMOTE	0.667	0.032
Random Forest	No Handling	0.592	0.049
Random Forest	SMOTE	0.702	0.021
AdaBoost	No Handling	0.602	0.069
AdaBoost	SMOTE	0.664	0.038
Gradient Boosting	No Handling	0.604	0.036
Gradient Boosting	SMOTE	0.675	0.010
XGBoost	No Handling	0.689	0.036
XGBoost	SMOTE	0.710	0.035

- **Best Model: XGBoost using SMOTE with F2-score = 0.7107**

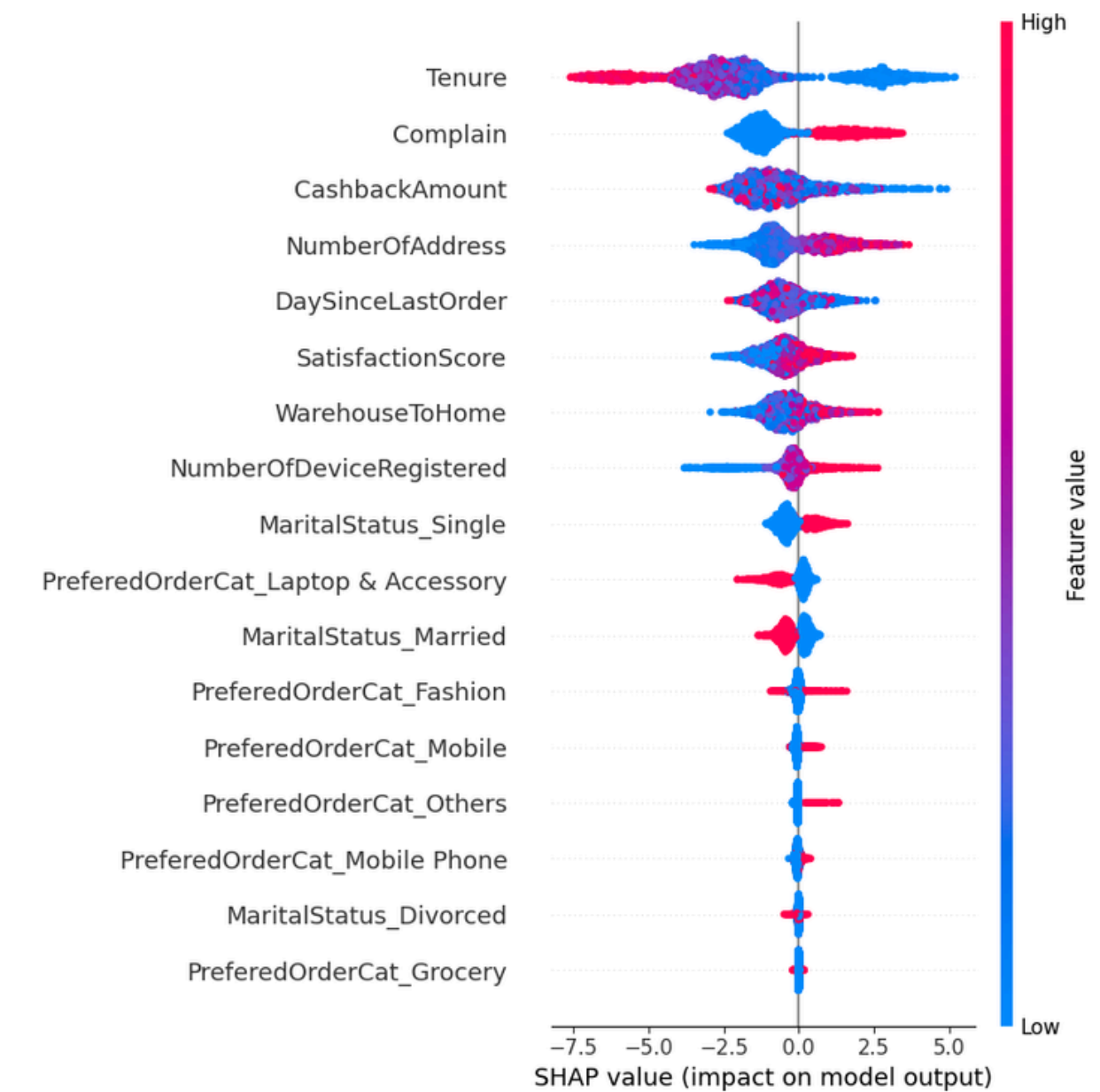
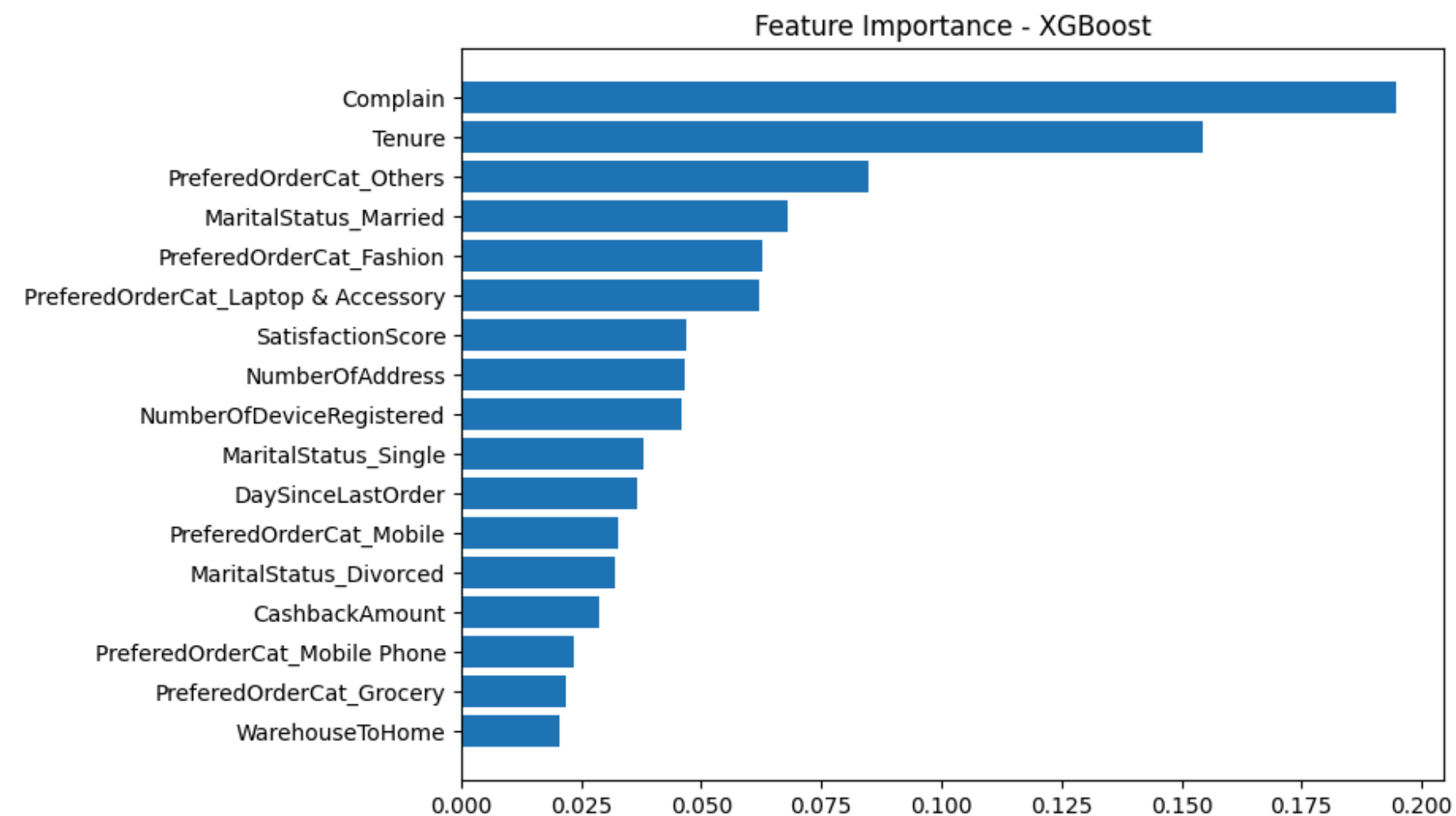
Hyperparameter Tuning Results (Top 5)

RANK	MEAN TEST SCORE	COL SAMPLE BYTREE	GAMMA	LEARNING RATE	MAX DEPTH	N ESTIMATOR	SUBSAMPLE
1	0.7344	0.8	0.0	0.2	7	300	1.0
2	0.7306	1.0	0.0	0.2	7	100	1.0
3	0.7299	1.0	0.0	0.2	7	300	1.0
4	0.7270	0.8	0.0	0.2	7	200	1.0
5	0.7270	0.8	0.0	0.1	7	300	1.0

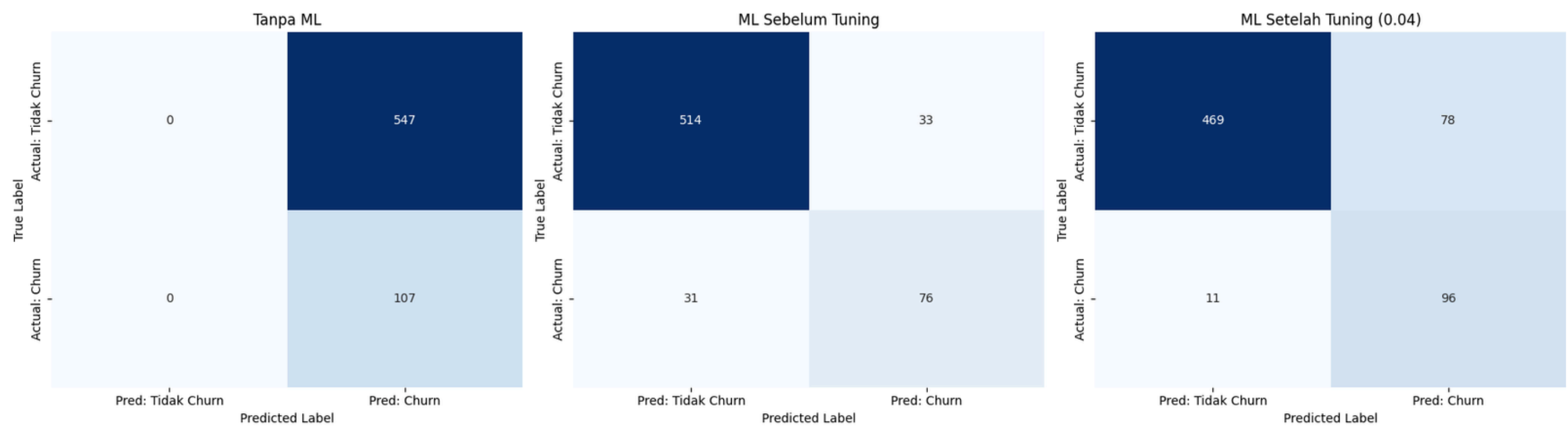
- **Model terbaik (Rank 1) pakai: colsample_bytree=0.8, learning_rate=0.2, max_depth=7, n_estimators=300.**
- **Skor tertinggi: 0.7344.**
- **Pola umum: max_depth=7 & gamma=0.0 muncul di semua top model.**
- **Perbedaan skor tipis → model stabil terhadap variasi hyperparameter.**

Modeling

Feature Importance & Explainability



Business Impact



- Keterangan**
- **FP (False Positive):**
Intervensi salah sasaran → biaya promosi.
 - **FN (False Negative):**
Kehilangan customer → biaya kehilangan lebih besar.
 - **Tuning threshold ke 0.04** menekan FN, sehingga kerugian total turun.

SKENARIO	COST FP (\$)	COST FN (\$)	TOTAL KERUGIAN (\$)	PENGHEMATAN (%)
Tanpa Machine Learning	136750	0	136750	0%
ML Sebelum Tuning	8250	31000	39250	71.29%
ML Setelah Tuning	19500	11000	30500	77.69%

Conclusion & Recommendation

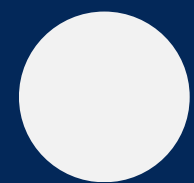
Conclusion

- Model XGBoost + SMOTE dengan threshold 0.04 terbukti efektif meminimalkan kerugian akibat churn, menghasilkan penghematan hingga 77.69% dibandingkan tanpa ML.
- Faktor paling berpengaruh terhadap churn adalah Satisfaction Score, Tenure, dan Complain, yang menguatkan temuan EDA sebelumnya.
- Pelanggan baru (0–6 bulan) dengan kepuasan rendah dan pernah mengeluh memiliki risiko churn paling tinggi.
- Model memiliki stabilitas performa yang baik terhadap variasi hyperparameter, sehingga dapat diandalkan untuk prediksi operasional.

Recommendation

- Manajemen → Integrasikan model churn ke strategi retensi, fokus pada pelanggan baru bernilai tinggi untuk memaksimalkan ROI.
- Marketing & Retention → Gunakan daftar pelanggan berisiko tinggi untuk kampanye promosi dan program loyalitas yang lebih tepat sasaran.
- Customer Service → Terapkan pelayanan proaktif pada pelanggan dengan kepuasan rendah atau riwayat komplain untuk menurunkan churn.
- Data Team → Pantau performa model secara berkala, lakukan retraining, dan kembangkan fitur tambahan berbasis perilaku pelanggan untuk meningkatkan akurasi.

Thank you!



cs@swiftbuy.com



www.swiftbuy.com



Green Office Park 9, 50th Floor

