



Purwadhika Digital
Technology School

Prediksi Customer Churn di Perusahaan Shopme.id



Sulthanur Iman Fatahillah
JCDS-2702



Outline



01

MASALAH BISNIS

02

DATA
UNDERSTANDING

03

DATA
PREPROCESSING

04

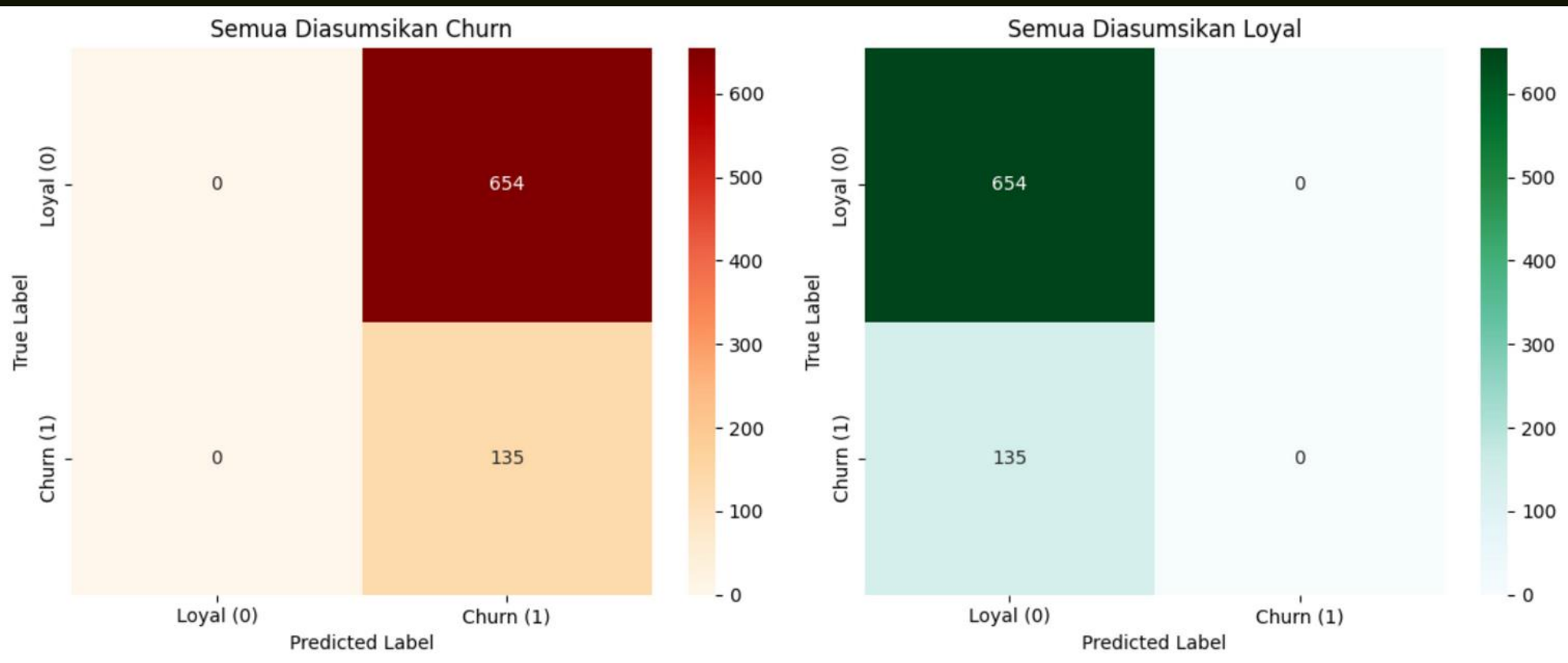
MODELING

05

KESIMPULAN DAN
REKOMENDASI



Latar Belakang



- Acquisition cost (A) = Rp 150.000 / pelanggan baru
- Retention cost (R) = Rp 20.000 / pelanggan yang diberi treatment

Skenario 1 — Treat All

- Total biaya kerugian perusahaan akibat menganggap semua churn adalah = Rp 13.080.000

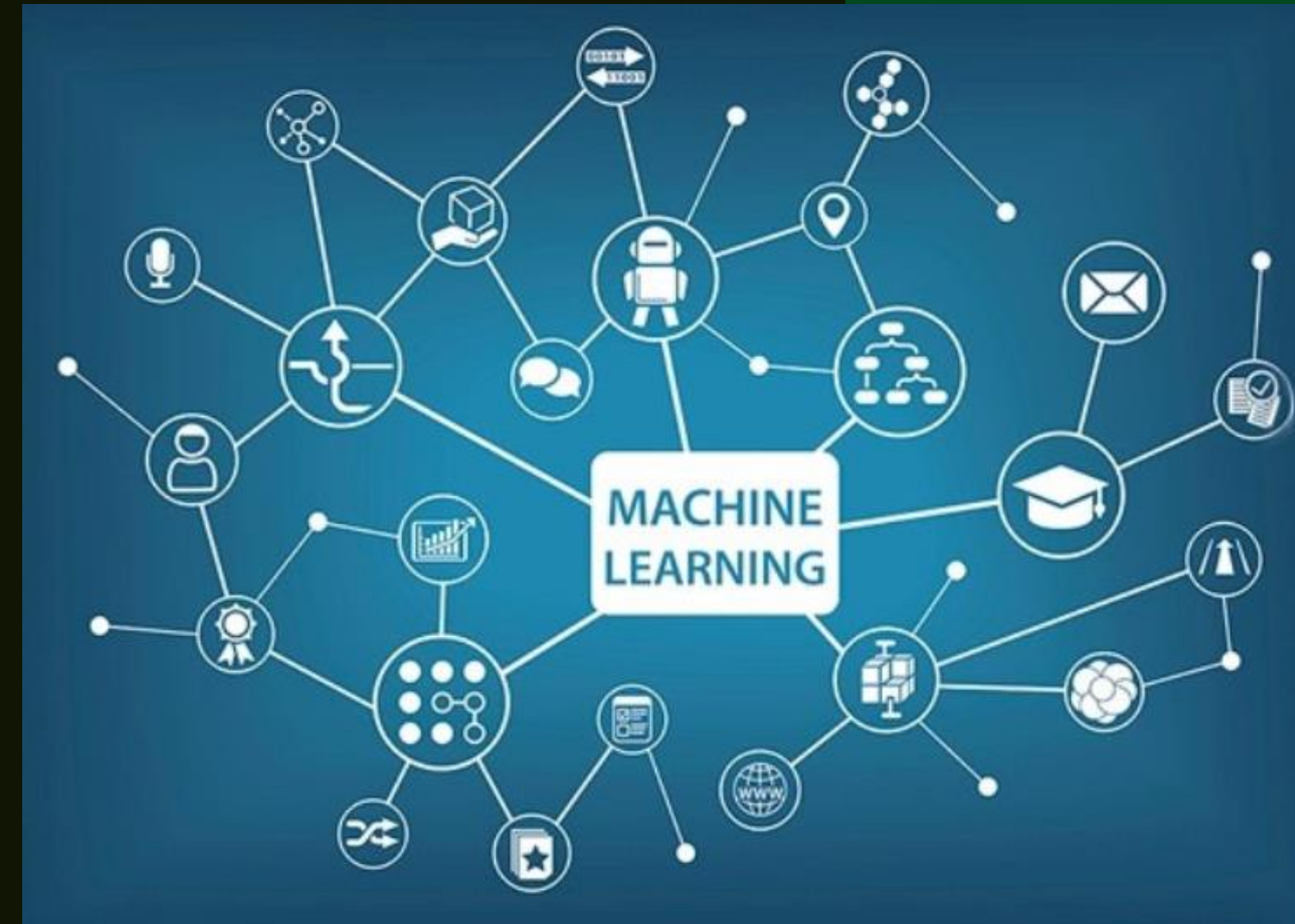
Skenario 2 — Treat None

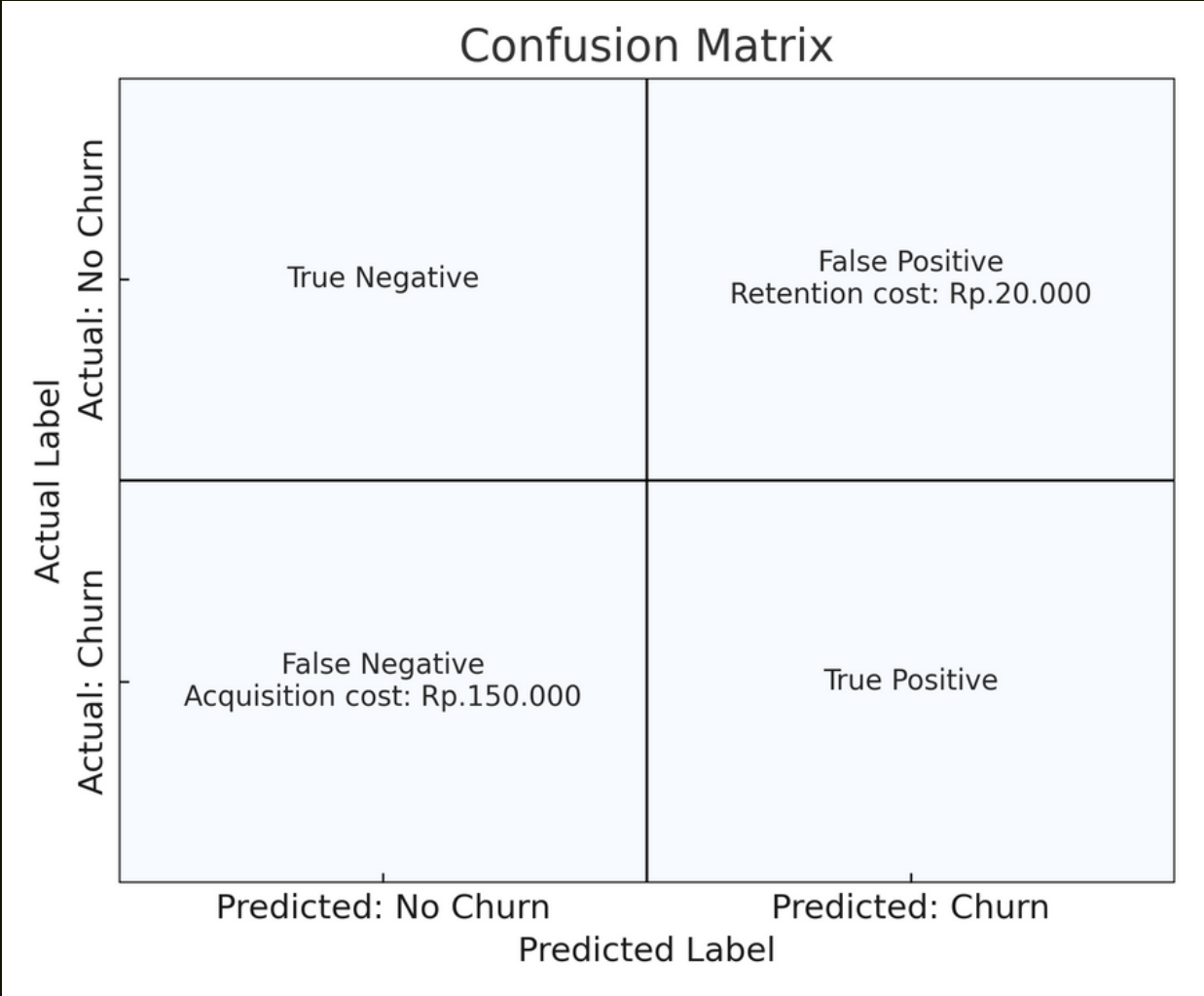
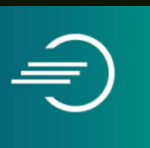
- Total biaya akibat kehilangan customer = Rp 20.250.000

Tanpa model ML, kedua strategi tidak efisien dan berpotensi membuang anggaran atau kehilangan potensi pendapatan.

Tujuan

- Menargetkan upaya retention hanya pada pelanggan yang benar-benar berisiko tinggi.
- Mengoptimalkan penggunaan anggaran pemasaran.
- Mengurangi total biaya gabungan (retention + akuisisi).
- Membantu Tim Marketing untuk prediksi customer churn untuk diberikan retention cost





Tolak Ukur Model

Penggunaan F-3 Score mengutamakan untuk menurunkan jumlah acqusition cost serendah mungkin.

Data Understanding

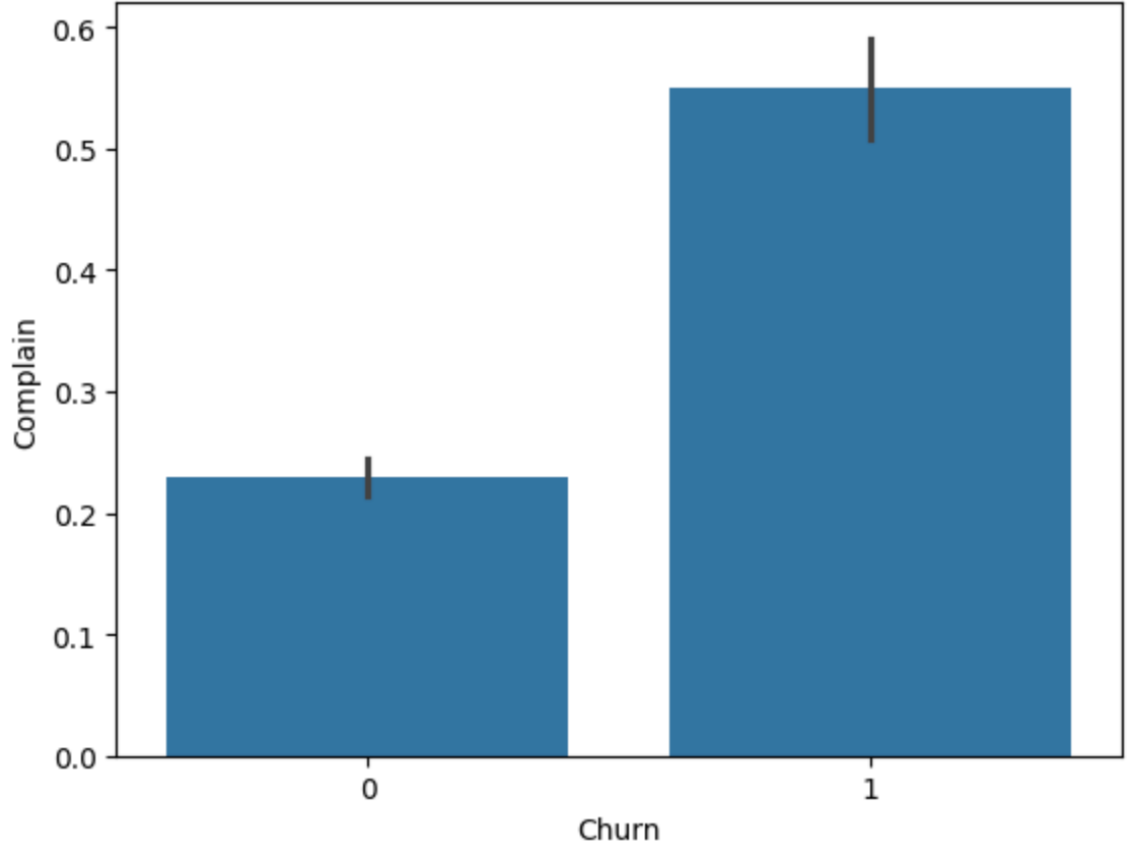
Fitur	Deskripsi
Tenure	Lama waktu pelanggan bergabung di perusahaan.
WarehouseToHome	Jarak antara gudang dan rumah pelanggan.
NumberOfDeviceRegistered	Jumlah perangkat yang terdaftar pada pelanggan tertentu.
PreferedOrderCat	Kategori pesanan yang paling sering dipilih pelanggan dalam sebulan terakhir.
SatisfactionScore	Skor kepuasan pelanggan terhadap layanan.
MaritalStatus	Status pernikahan pelanggan.
NumberOfAddress	Jumlah alamat yang terdaftar pada pelanggan tertentu.
Complaint	Apakah ada keluhan yang diajukan dalam sebulan terakhir.
DaySinceLastOrder	Jumlah hari sejak pesanan terakhir oleh pelanggan.
CashbackAmount	Rata-rata cashback dalam sebulan terakhir.

Data Historis ini akan digunakan model machine learning untuk memprediksi seorang customer adalah churn atau loyal. Dimana satu baris data mewakili satu pelanggan.

Target	Deskripsi
Churn	Penanda apakah pelanggan churn atau loyal.

Explanatory Data Analysis

Complain rate tinggi mengindikasikan pelanggan akan churn

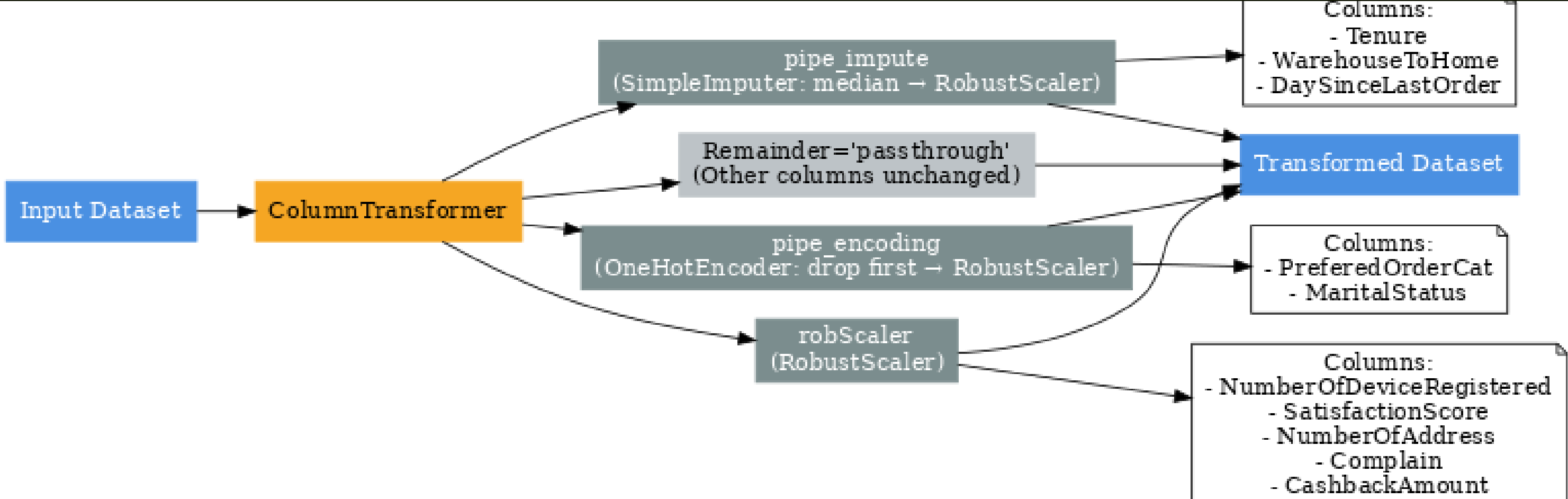


Pelanggan yang churn memiliki complain rate yang lebih tinggi dibandingkan dengan yang tidak churn.



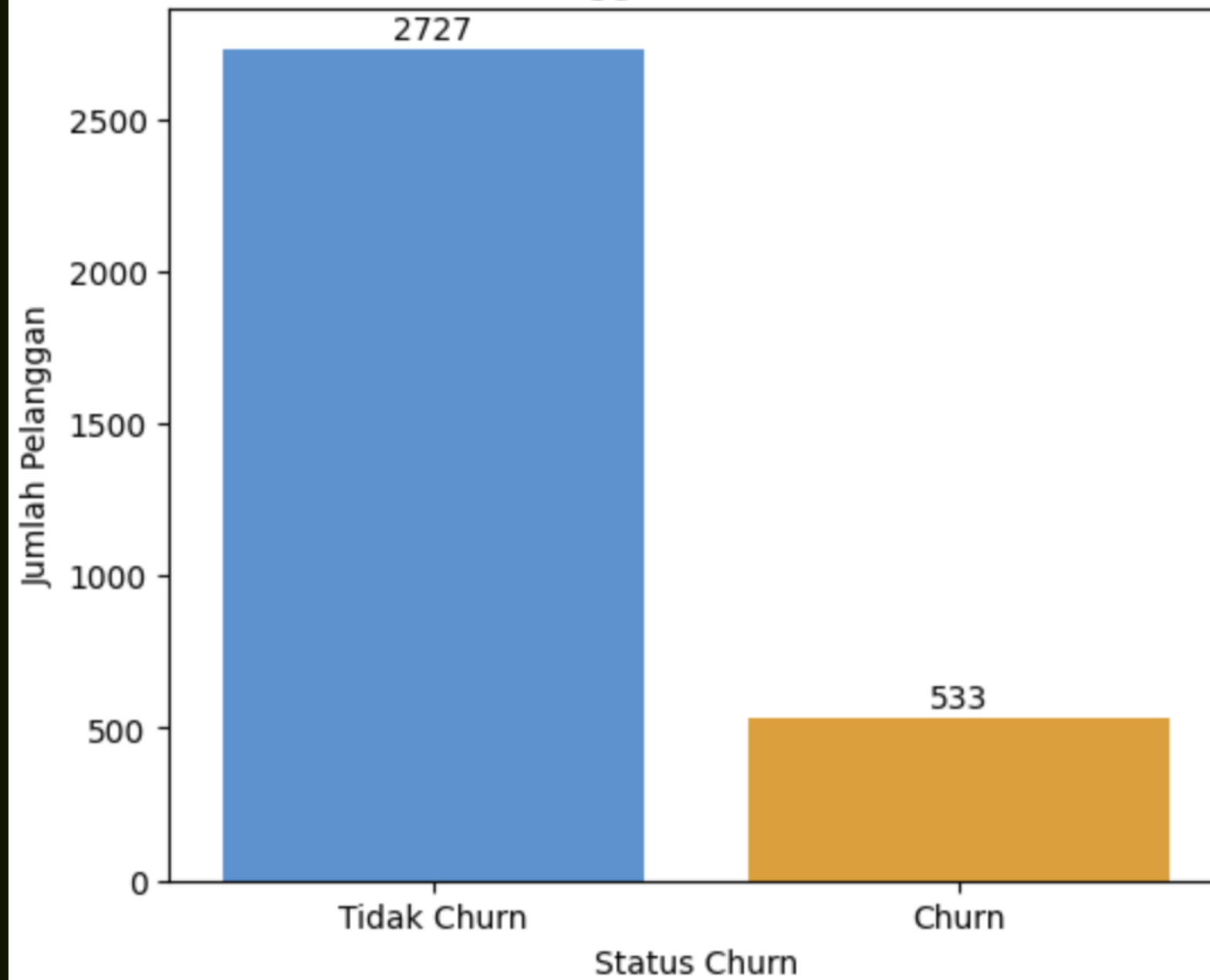
Data Preprocessing

- Imputing
- Encoding
- Scaling





Distribusi Pelanggan Churn vs Tidak Churn



Resampling

- None
- Random Over Sampling
- SMOTE Over Sampling



Modeling

- Logistic Regression
- KNN
- Decision Tree
- Random Forest
- Adaptive Boosting
- Gradient Boosting
- Extreme Gradient Boosting

Cross-Validation

	Algoritma	Resampling	Score	Mean	Std
0	LogisticRegression	None	[0.7675675675675675, 0.6427015250544662, 0.738...	0.722757	0.056468
1	LogisticRegression	SMOTE	[0.7700650759219091, 0.6413043478260868, 0.729...	0.720299	0.054321
2	LogisticRegression	RandomOverSampler	[0.7526881720430106, 0.6202393906420021, 0.719...	0.713086	0.062800
3	KNeighborsClassifier	None	[0.33707865168539325, 0.3117206982543641, 0.39...	0.363640	0.034900
4	KNeighborsClassifier	SMOTE	[0.7290533188248095, 0.6762749445676275, 0.647...	0.696263	0.031552
5	KNeighborsClassifier	RandomOverSampler	[0.7314410480349345, 0.6896551724137931, 0.652...	0.703378	0.038473
6	DecisionTreeClassifier	None	[0.6168446026097272, 0.5721096543504172, 0.584...	0.599257	0.017838
7	DecisionTreeClassifier	SMOTE	[0.6580493537015276, 0.5484247374562428, 0.652...	0.642154	0.079537
8	DecisionTreeClassifier	RandomOverSampler	[0.590318772136954, 0.5778301886792453, 0.5342...	0.575029	0.026839
9	RandomForestClassifier	None	[0.48327137546468396, 0.4551045510455105, 0.48...	0.493760	0.038428
10	RandomForestClassifier	SMOTE	[0.6738868832731648, 0.5819477434679334, 0.633...	0.657039	0.046424
11	RandomForestClassifier	RandomOverSampler	[0.6287787182587666, 0.5762304921968787, 0.644...	0.627576	0.055642
12	AdaBoostClassifier	None	[0.6264775413711584, 0.5931198102016607, 0.571...	0.611101	0.025210
13	AdaBoostClassifier	SMOTE	[0.6272189349112426, 0.580720092915215, 0.6321...	0.638950	0.045396
14	AdaBoostClassifier	RandomOverSampler	[0.6035502958579881, 0.5575326215895612, 0.581...	0.590876	0.023063
15	GradientBoostingClassifier	None	[0.5596107055961071, 0.5301204819277109, 0.597...	0.575094	0.047351
16	GradientBoostingClassifier	SMOTE	[0.6933019976498236, 0.5827505827505829, 0.702...	0.662111	0.064794
17	GradientBoostingClassifier	RandomOverSampler	[0.7648401826484017, 0.7361268403171007, 0.741...	0.755288	0.034362
18	XGBClassifier	None	[0.665859564164649, 0.6190476190476192, 0.5344...	0.630052	0.051141
19	XGBClassifier	SMOTE	[0.709987966305656, 0.6352941176470589, 0.5944...	0.661658	0.047761
20	XGBClassifier	RandomOverSampler	[0.6818181818181818, 0.6807511737089201, 0.648...	0.679939	0.017872

Model terbaik adalah GradientBoosting dan menggunakan RandomOverSampler yang mendapatkan mean F-3 Score tertinggi (75,5%) dan cukup stabil dengan Standar Deviasi yang masih relatif rendah (0,034).

Hyper-Parameter Tuning

Gradient Boosting

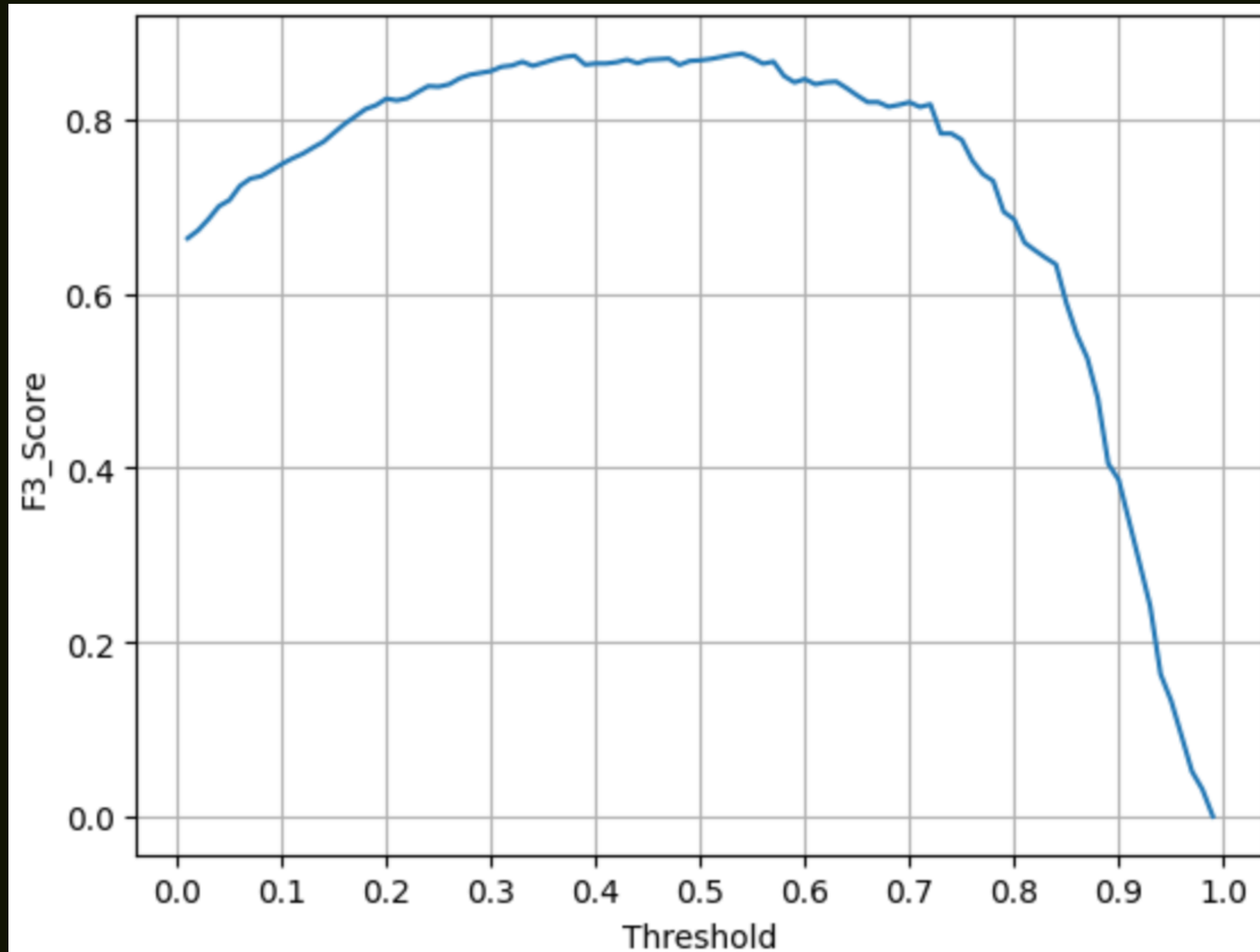
Hyperparameter	Nilai yang Diuji
n_estimators	100, 200
learning_rate	0.05, 0.1
max_depth	3, 5
min_samples_split	2, 5
min_samples_leaf	1, 2
subsample	0.8, 1.0

Best Score yang didapat saat melakukan parameter tuning adalah 85,08%

Pada saat melakukan prediksi ke test set, terjadi penurunan saat menggunakan tuning yang awal nya 86,7% menjadi 85,08%. Maka di sini saya akan menggunakan base model saja.



Tuning Threshold



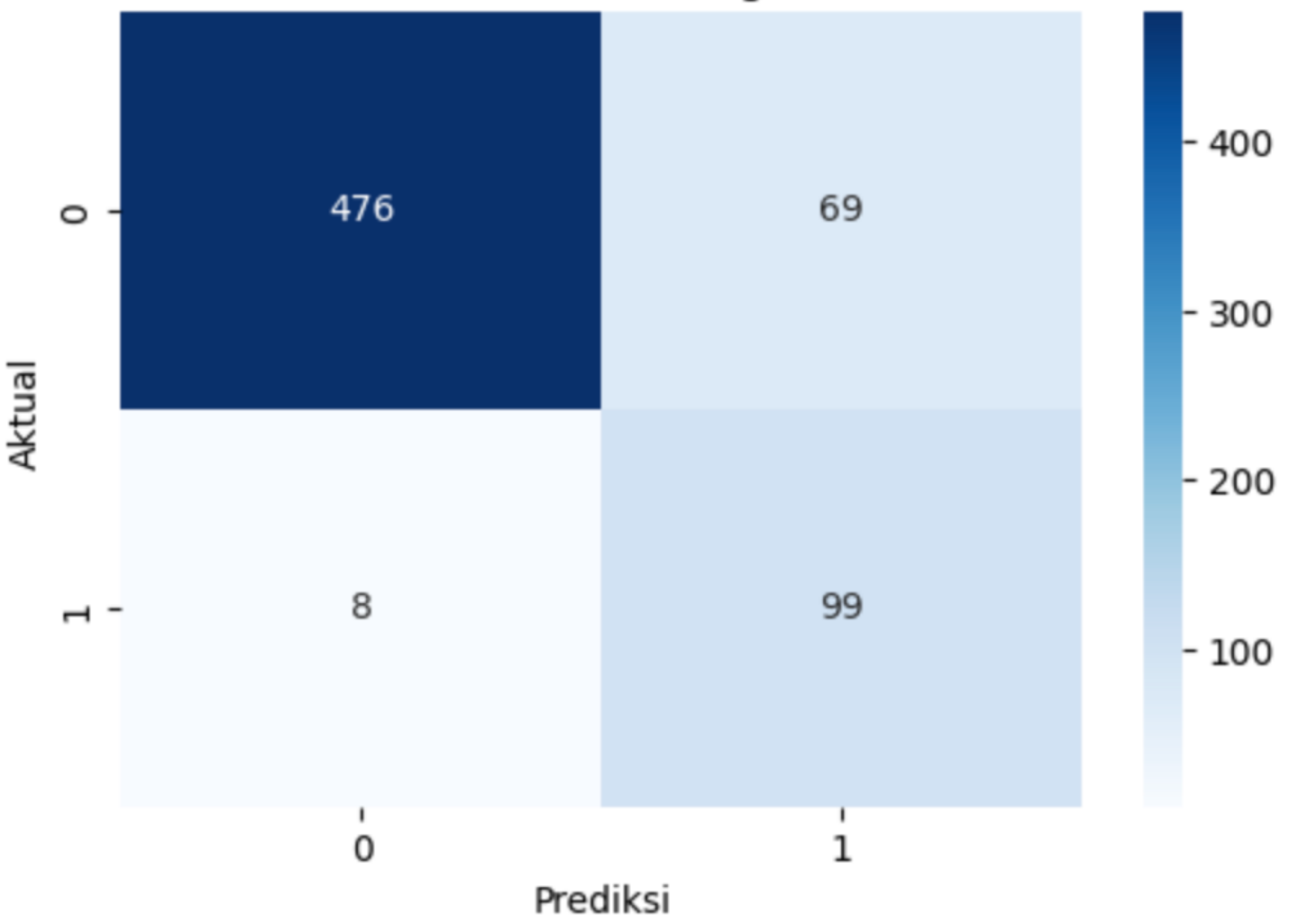
Model Gradient Boosting mendapatkan F-3 Score terbaik saat threshold berada di 0,54.



Kesimpulan dan Saran



Confusion Matrix GradientBossting (Threshold = 0.54)



Dampak Bisnis

Retention Cost yang tidak diperlukan(FP)
 $69 \times \text{Rp}20.000 = \text{Rp}1.380.000$

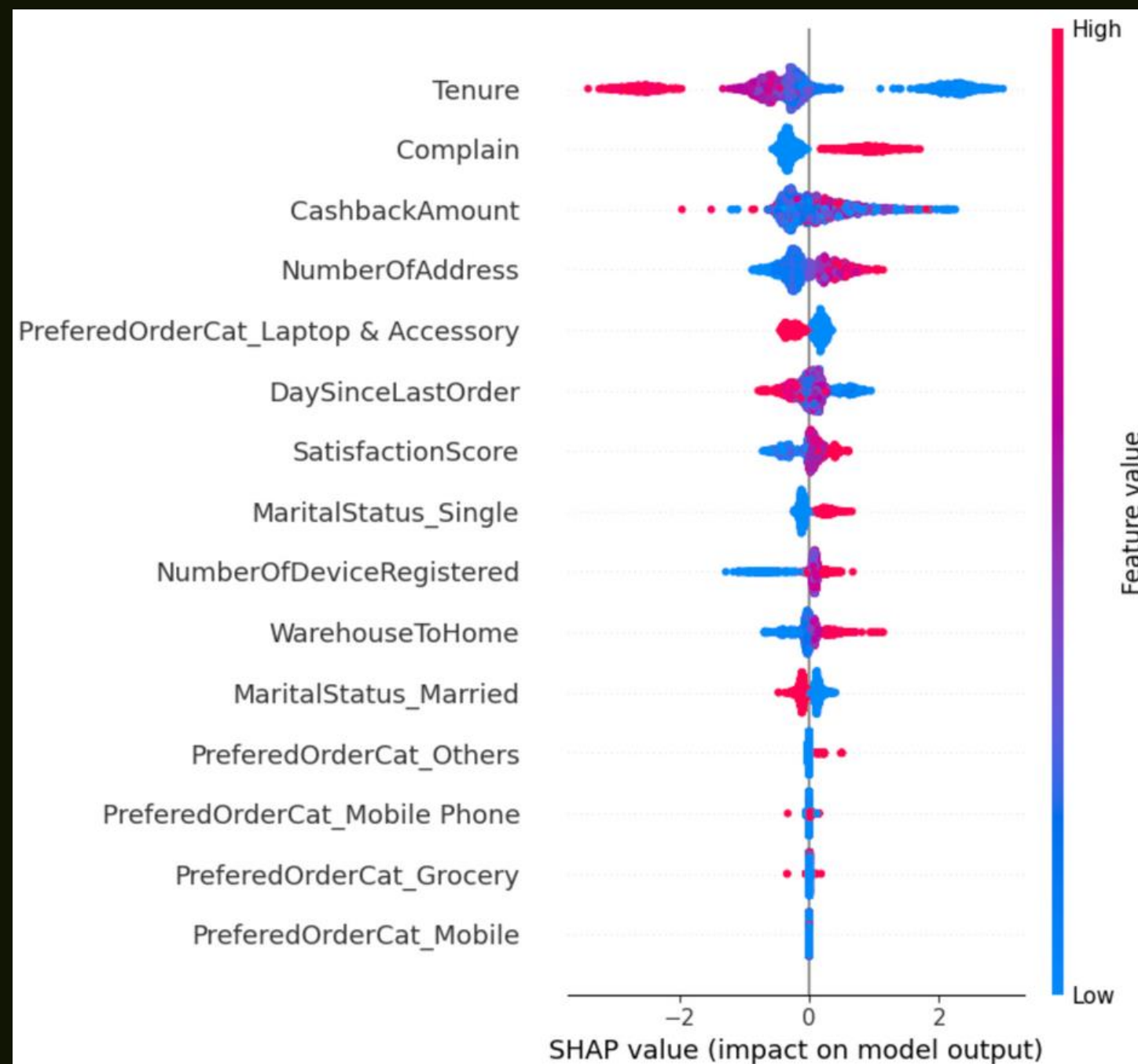
Acquisition diberikan ke pelanggan churn yang tidak di-treatment (FN)
 $8 \times \text{Rp}150.000 = \text{Rp}1.200.000$

Total Biaya :
 $\text{Rp}1.380.000 + \text{Rp}1.200.000 = \text{Rp}2.580.000$

Skenario	Total Biaya	Selisih Hemat dibanding ML
Treat All	Rp 13.080.000	Rp 10.500.000
Treat None	Rp 20.250.000	Rp 17.670.000
Machine Learning (0.54)	Rp 2.580.000	—



Rekomendasi Bisnis



1. Tenure rendah → risiko churn tinggi.
 - Buat program onboarding yang lebih engaging (email welcome series, edukasi fitur, promo khusus bulan pertama).
 - Berikan loyalty bonus setelah melewati 3–6 bulan pertama untuk mengurangi risiko churn di awal.
2. Pelanggan yang pernah complain → risiko churn tinggi.
 - Implementasikan Customer Support Priority untuk pelanggan yang pernah complain.
 - Setelah complain terselesaikan, kirim voucher atau cashback kecil sebagai bentuk apresiasi.
3. Cashback rendah → risiko churn tinggi.
 - Naikkan persentase cashback untuk segmen yang teridentifikasi memiliki risiko tinggi.
 - Buat cashback progresif: semakin lama pelanggan bertahan, semakin tinggi cashback-nya.
4. WarehouseToHome tinggi → risiko churn tinggi.
 - Optimalkan rute logistik atau buat warehouse mini di lokasi strategis.
 - Tawarkan gratis ongkir untuk segmen pelanggan dengan jarak pengiriman jauh.



Thank You



<https://github.com/sulthanfatahillah12>



<https://www.linkedin.com/in/sulthanur-iman-fatahillah>

Sulthanur Iman Fatahillah

