

E-Commerce Customers Churn Analysis and Prediction

Final Project Rakamin Academy Data Science Batch 25



Our Client



E-Commerce #1 Indonesia menurut
Majalah Bobo

Our Team



Garry



Fadla



Taufan



Diah



Ajeng



Rifky



Izza



Iqbal

Outline



Step 1

Business Understanding

Background, Problem Statement, Objective, Goals, Metrics

Step 2

Data Understanding

Data Overview

Step 3

Preprocessing

Duplicate & Missing Values, Encoding, Outliers, Split Data

Step 4

Modelling & Evaluation

Menentukan model yang "Best Fit"

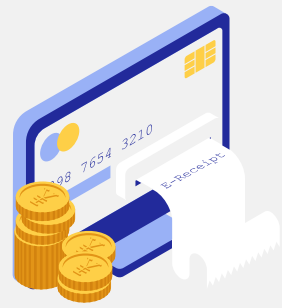
Step 5

Business Recommendations

Recommendations, Impact



Business Understanding



Business Understanding

Background Client

Tokopaedi adalah E-Commerce dengan perkembangan Nomor 1 di Indonesia (Berdasarkan Majalah Bobo) yang menjual berbagai macam jenis produk mulai dari Grocery, Electronic, Fashion, Laptop & Accessory dan lain-lain.

Terlepas dari perkembangan yang terjadi Tokopaedi mengalami masalah serius yaitu

Customer Churn

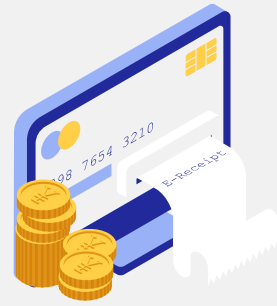


Apa itu Churn?

Customer Churn didefinisikan sebagai kecenderungan pelanggan untuk berhenti melakukan bisnis dengan sebuah perusahaan (Yu dkk. dalam Arifin, 2014).

Menurut Khakabi dkk., (dalam Arifin, 2014) untuk memperoleh pelanggan baru itu memerlukan biaya hingga 10 kali lipat lebih mahal dari biaya untuk mempertahankan pelanggan yang ada.

Source : <https://jurnal.umk.ac.id/index.php/SNA/article/viewFile/156/158>



Business Understanding

Problem Statement

Churn Rate

16.8% Churn

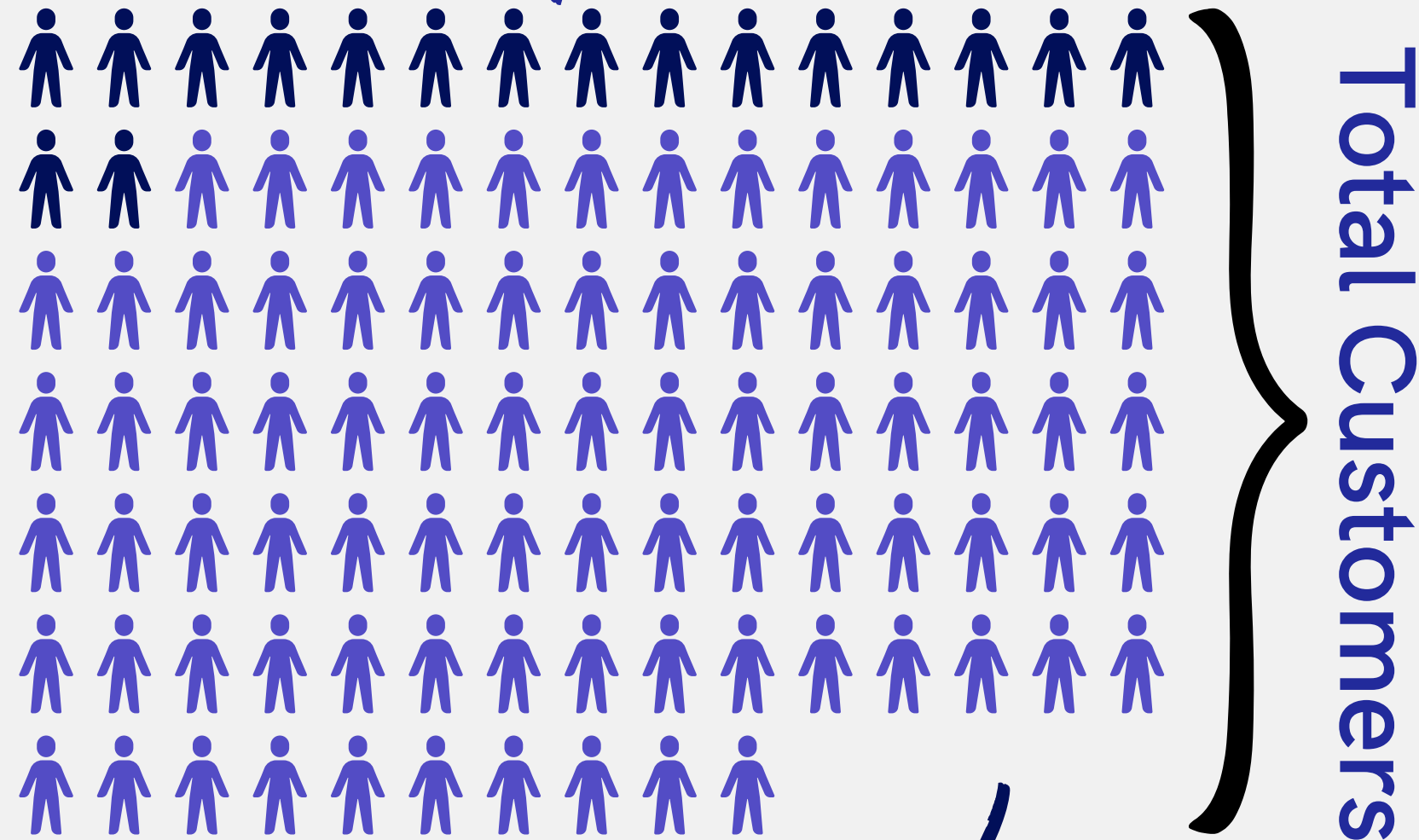
Lost Opportunity.



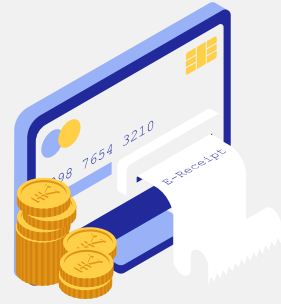
Rp473jt/Bulan

Lost Opportunity = Total Churn Customers x Average Monthly Spending User*

Source : <https://dataindonesia.id/digital/detail/berapa-pengeluaran-masyarakat-indonesia-untuk-belanja-online>



83.2% Not Churn



Business Understanding



Goals

Mengurangi tingkat churn customer dengan target churn rate dari 16,8% menjadi 5%



Objective

Membuat machine learning model yang dapat memprediksi churn dari perilaku atau ciri-ciri customer dan rekomendasi untuk dapat mengantisipasi perilaku churn customer berdasarkan insight.



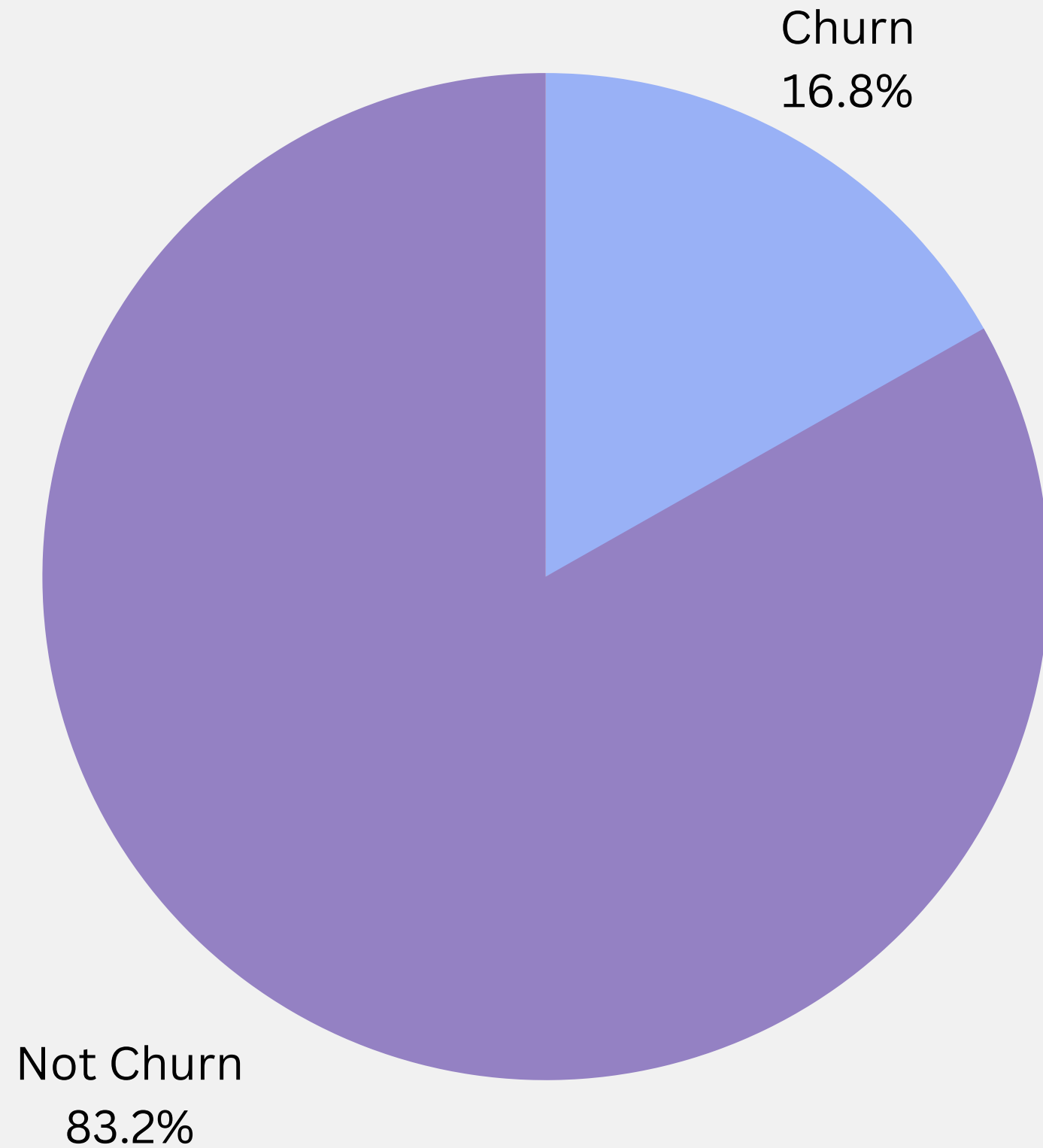
Metrics

$$\text{Churn Rate} = \frac{\text{Total Customers Churn}}{\text{Total Customers}}$$

Data Understanding



Data Understanding



About Dataset

Dataset berisi data penjualan Tokopaedi selama satu tahun

Shape

5630 Rows, 20 Columns

Data Overview

Variable	Discription
Churn	Churn Flag
CityTier	City tier
SatisfactionScore	Satisfactory score of customer on service
Complain	Any complaint has been raised in last month
PreferredLoginDevice	Preferred login device of customer
PreferredPaymentMode	Preferred payment method of customer
Gender	Gender of customer
PreferedOrderCat	Preferred order category of customer in last month
MaritalStatus	Marital status of customer
CustomerID	Unique customer ID
Tenure	Tenure of customer in organization
WarehouseToHome	Distance in between warehouse to home of customer
HourSpendOnApp	Number of hours spend on mobile application or website
NumberOfDeviceRegistered	Total number of deceives is registered on particular customer
NumberOfAddress	Total number of added added on particular customer
OrderAmountLikeFromlastYear	Percentage increases in order from last year
CouponUsed	Total number of coupon has been used in last month
OrderCount	Total number of orders has been places in last month
DaySinceLastOrder	Day Since last order by customer
CashbackAmount	Average cashback in last month

Target

Nominal

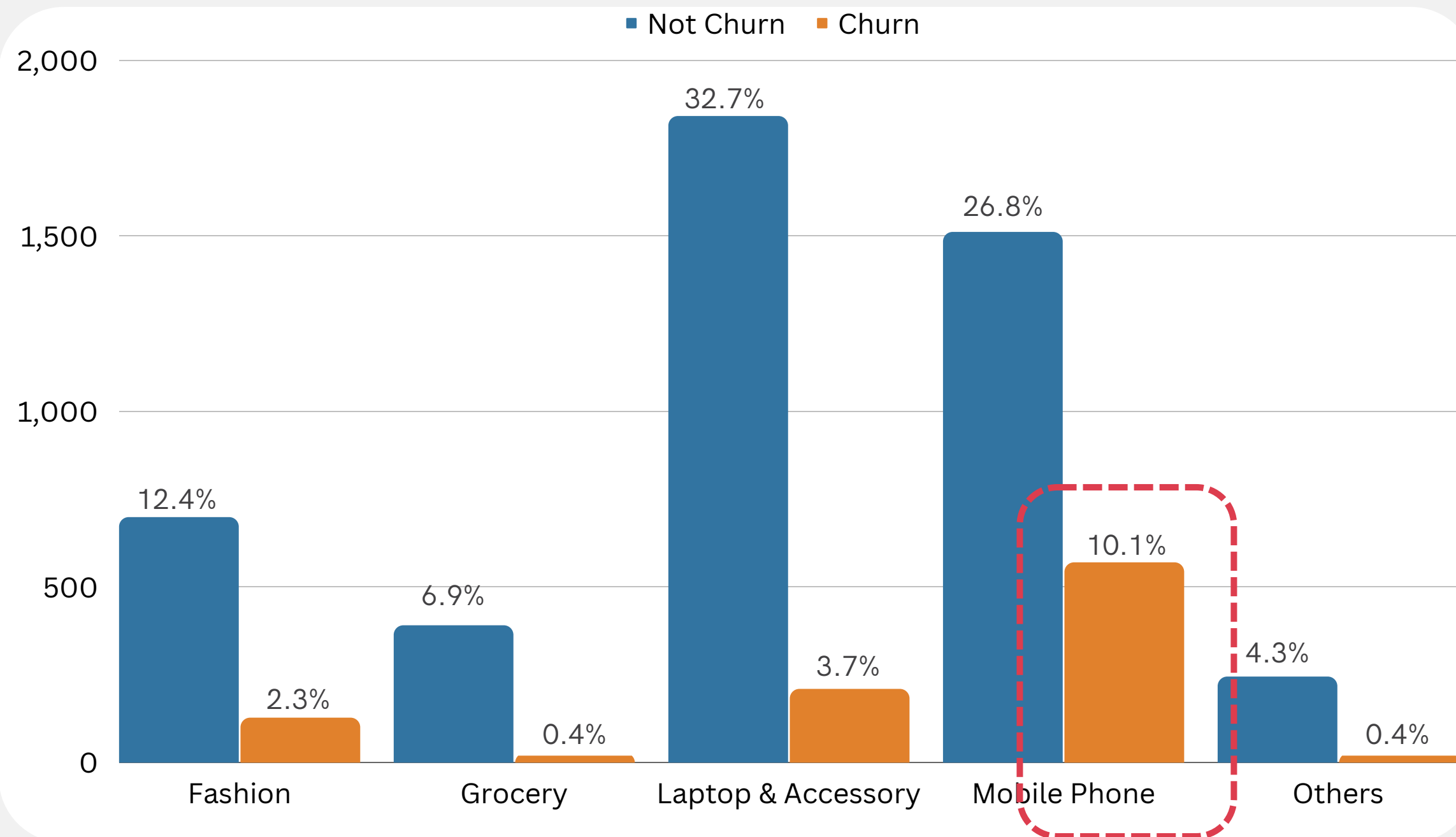
Ordinal

Interval

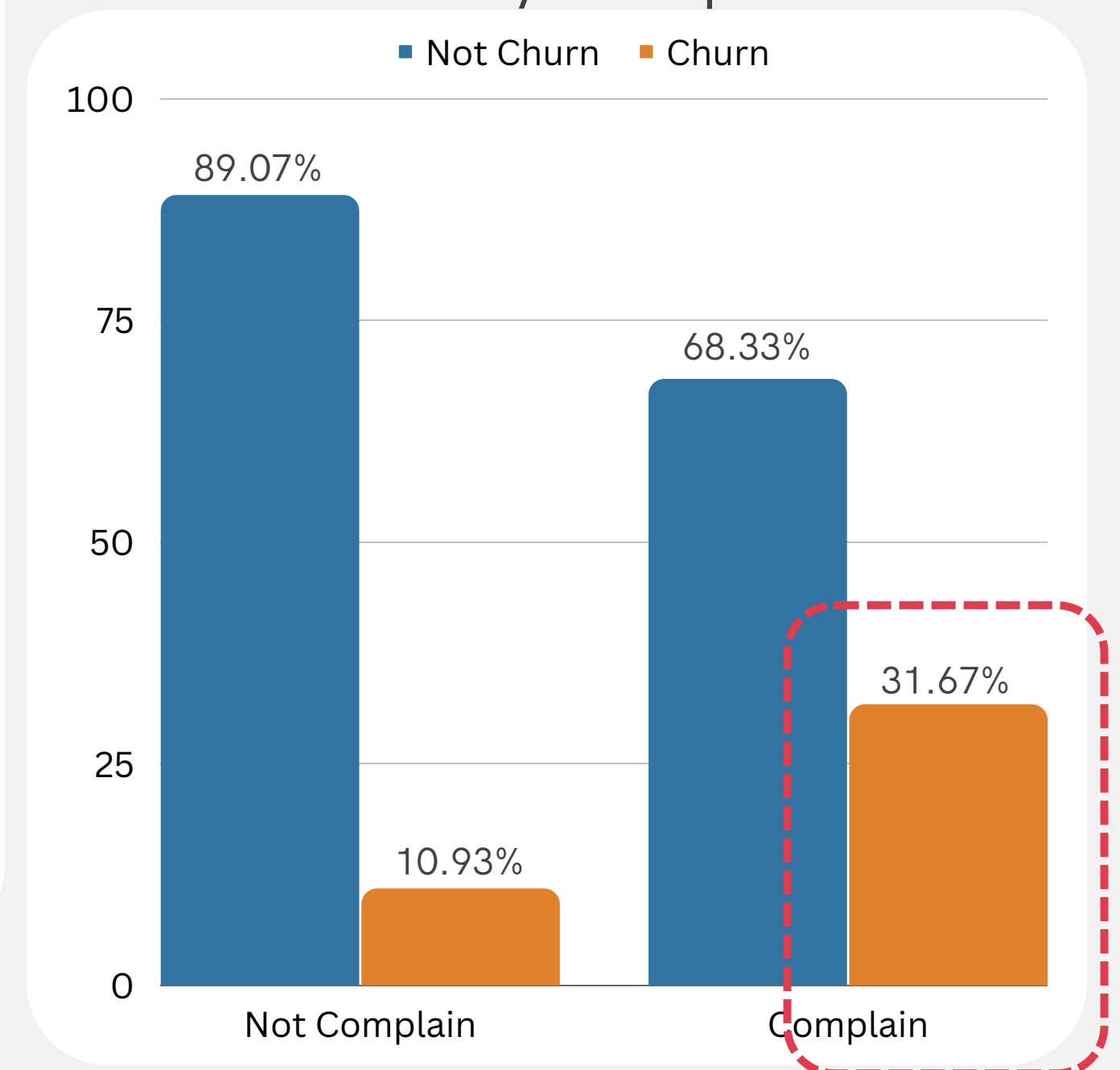
Data Understanding



Churn By Order Category

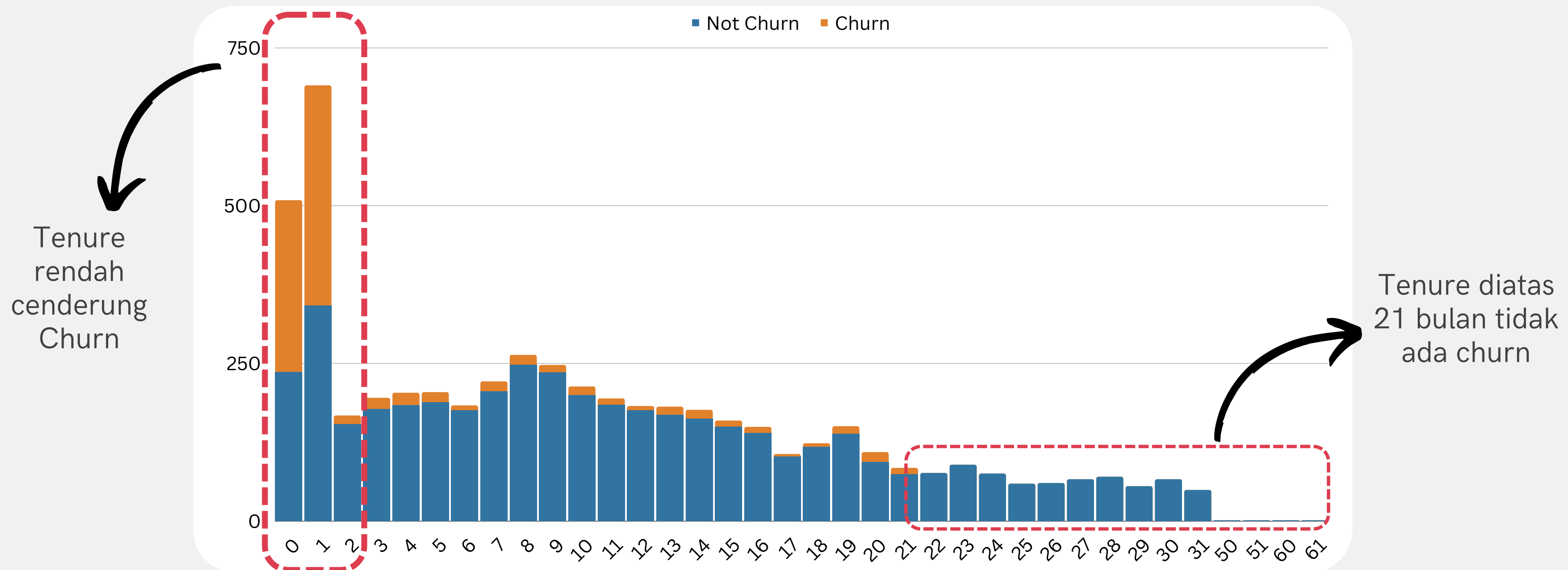


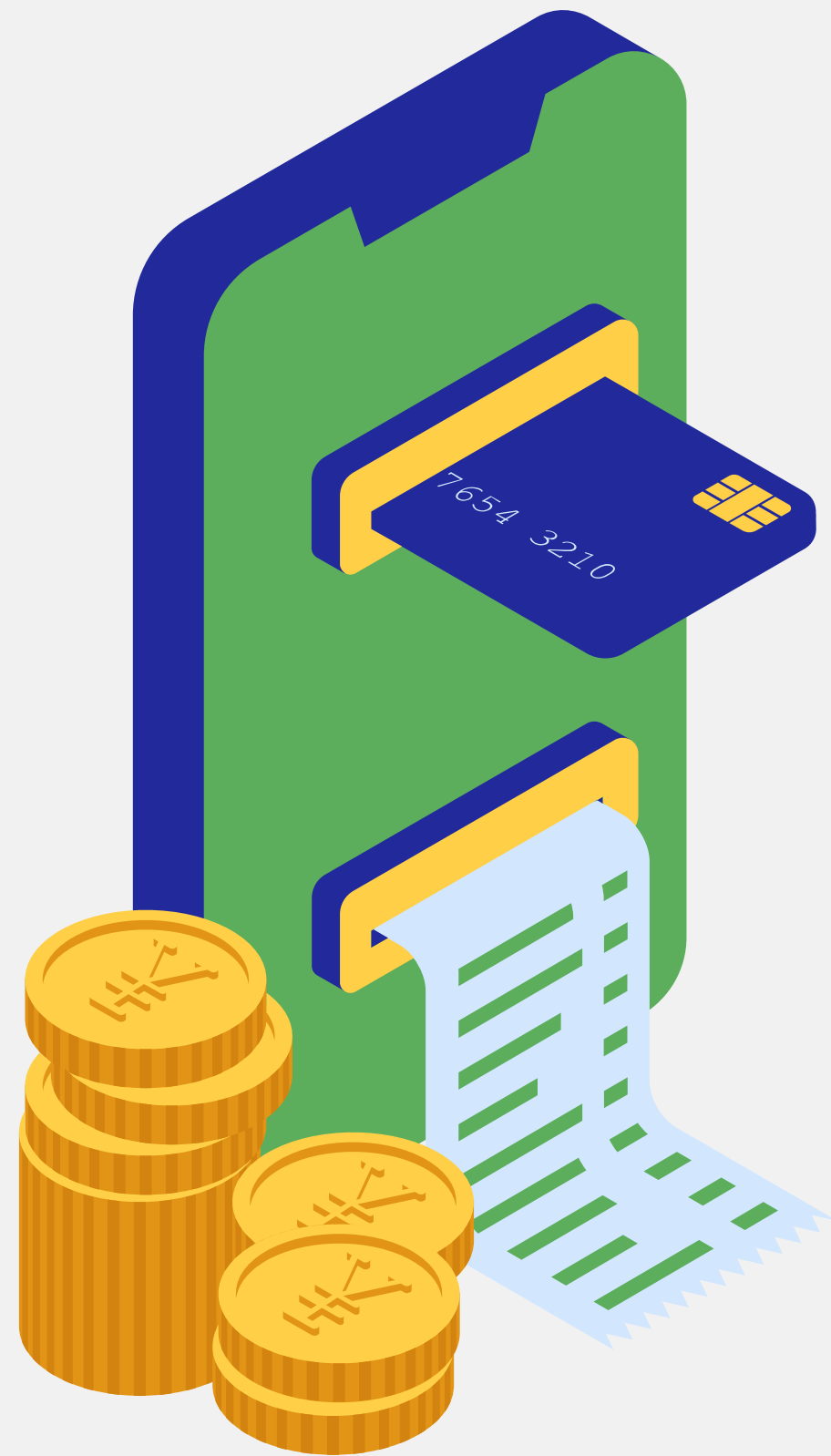
Churn By Complains



Data Understanding

Churn By Tenure





Data Preprocessing



Encoding

One-Hot Encoding:

- PreferredPaymentMode
- PreferredOrderCat
- MaritalStatus

Label Encoding

- Gender
- PreferredLoginDevice

Split Data

TRAIN 80%

TEST 20%

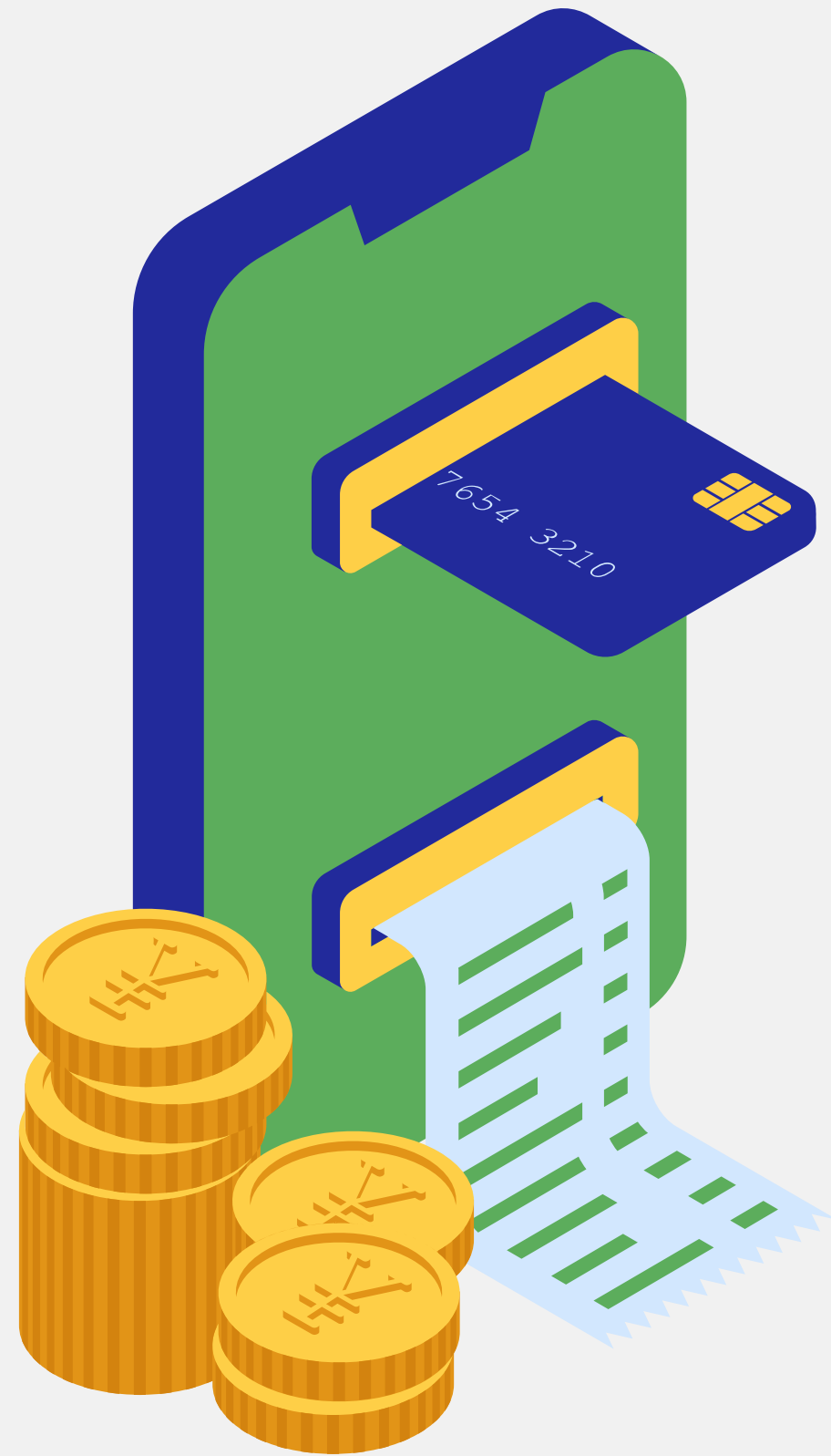
Missing Values



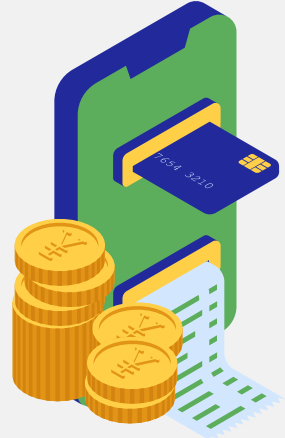
Treatment: Imputasi Median

Outliers

Outliers di keep karena masih menjadi bagian dari populasi



Modelling & Evaluation



Modelling & Evaluation

Train

Test

model_name	recall	precision	roc_auc	f1	accuracy	model_name	recall	precision	roc_auc	f1	accuracy
logistic regression	0.501966	0.750980	0.734009	0.601728	0.734009	logistic regression	0.524324	0.815126	0.750472	0.638158	0.750472
random forest	1.000000	1.000000	1.000000	1.000000	1.000000	random forest	0.859459	1.000000	0.929730	0.924419	0.929730
kNN	0.553080	0.791745	0.761704	0.651235	0.761704	kNN	0.421622	0.634146	0.686900	0.506494	0.686900
decision tree	1.000000	1.000000	1.000000	1.000000	1.000000	decision tree	0.924324	0.890625	0.951004	0.907162	0.951004
adaboost	0.579292	0.772727	0.772271	0.662172	0.772271	adaboost	0.567568	0.833333	0.772625	0.675241	0.772625
xgboost	1.000000	1.000000	1.000000	1.000000	1.000000	xgboost	0.886486	0.976190	0.941118	0.929178	0.941118
catboost	0.956750	0.997268	0.978108	0.976589	0.978108	catboost	0.821622	0.987013	0.909748	0.896755	0.909748

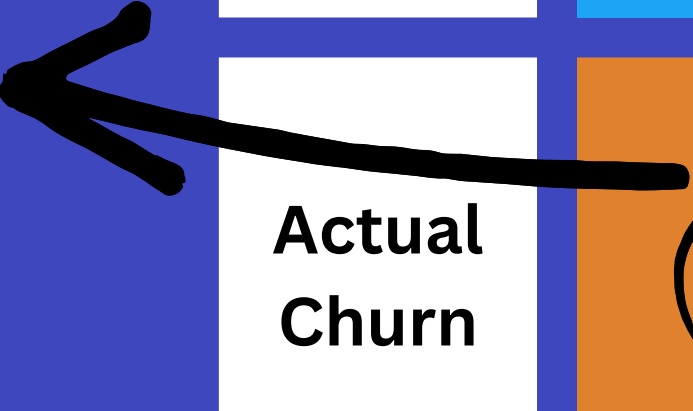
Model Decision Tree dengan recall 92%

Confusion Matrix

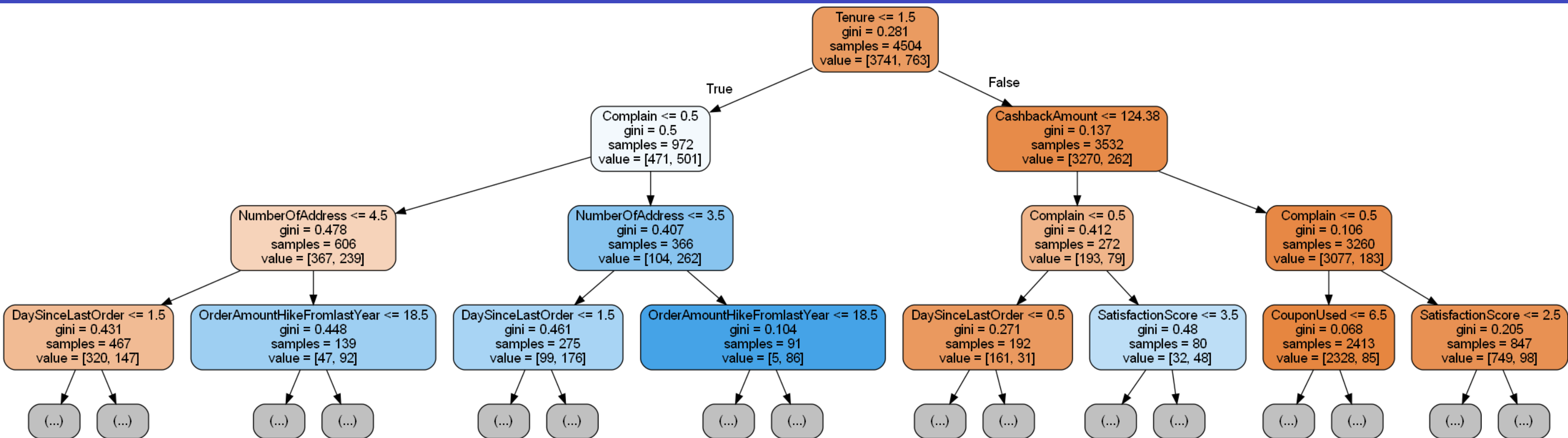
Mengapa Recall?

Tujuannya untuk mengetahui
sebanyak mungkin customer yang
actual Churn

	Predicted Not Churn	Predicted Churn
Actual Not Churn	881	17
Actual Churn	14	158



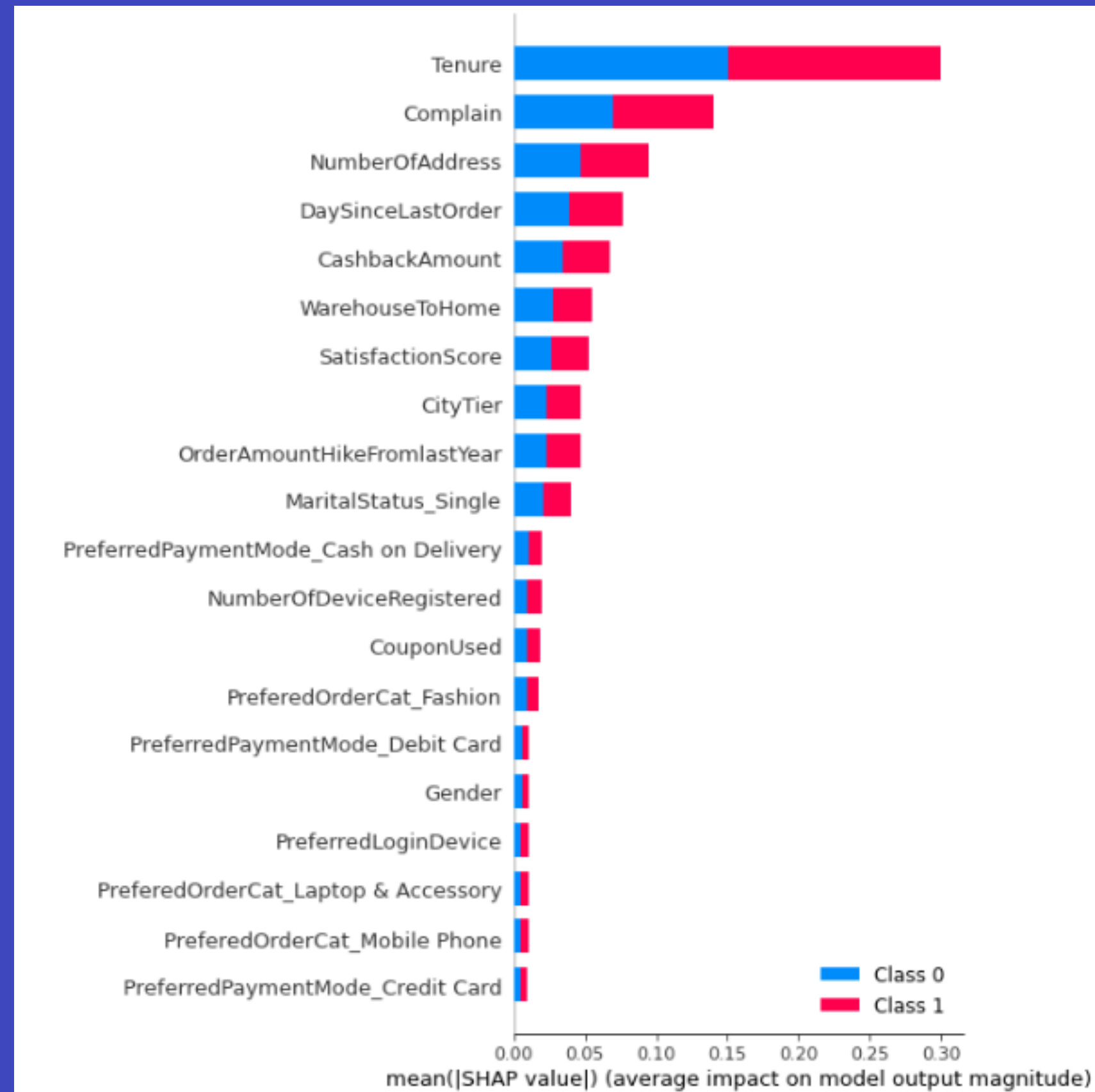
Decision Tree Visualization

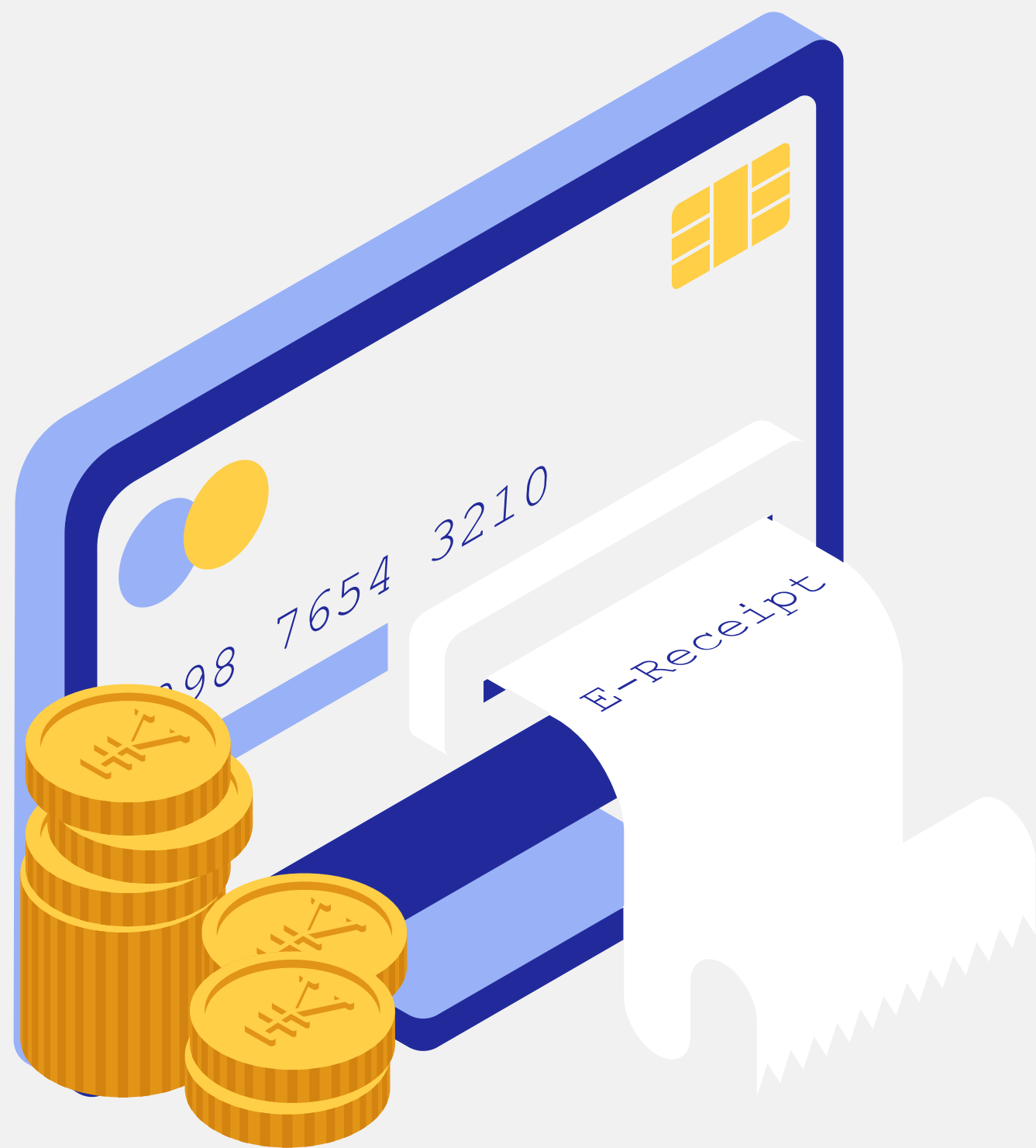


Ditampilkan potongan decision tree sampai 3 level percabangan, dimana pada tiap node menampilkan 4 informasi :

1. Nama kolom beserta treshold decision-nya
2. Nilai gini impurity fitur
3. Jumlah total sample yang diproses
4. Value = proporsi split data dalam node

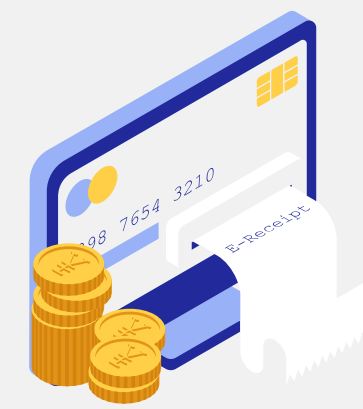
Features Importance



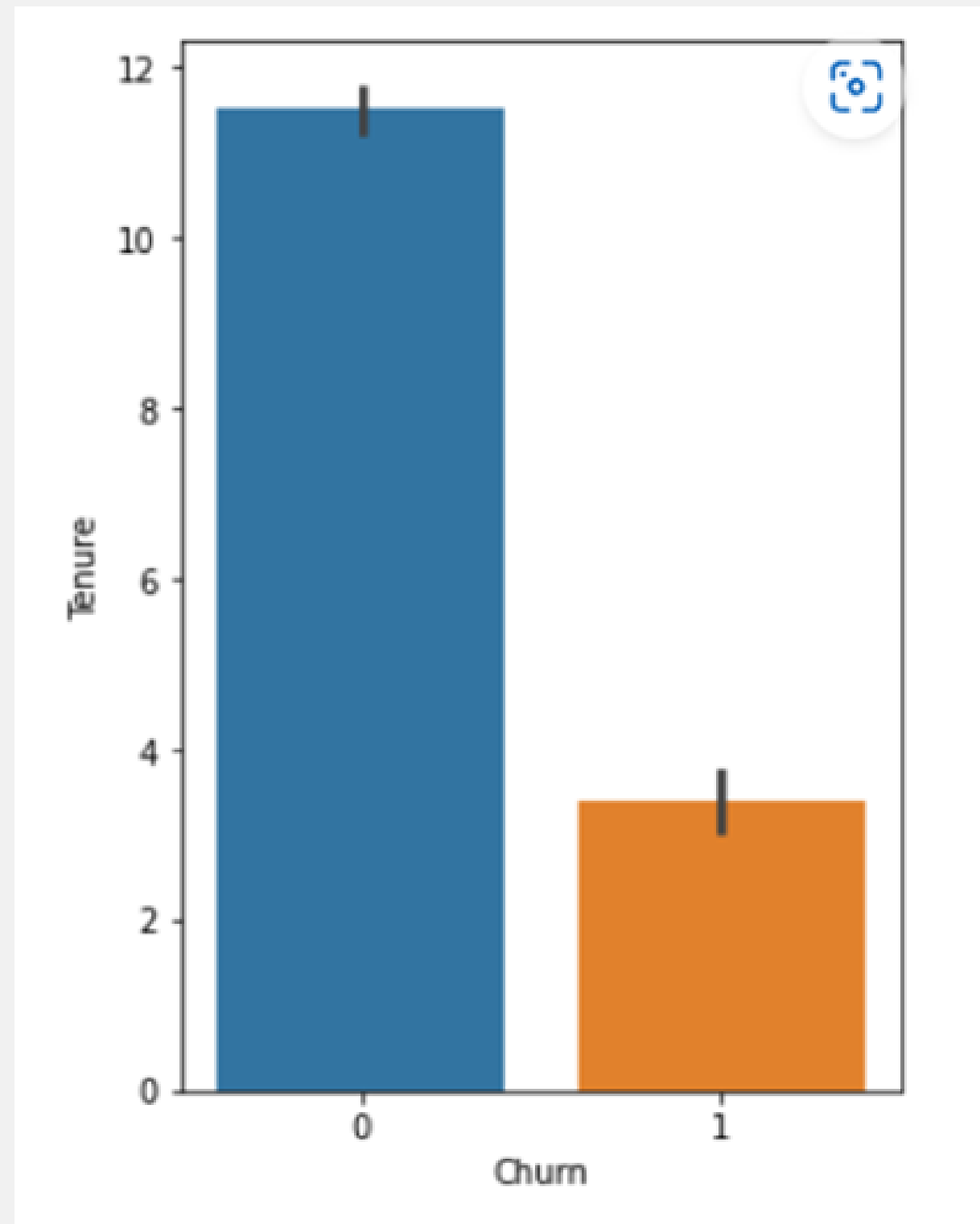


Business Recommendations

Benefit Recommendation



Churn By Tenure



Saat ini pelanggan yang churn terdistribusi pada tenor pembayaran di bawah 4 bulan

Mitigasi:

Tingkatkan program marketing (discount/coupon) pada pelanggan terduga churn yang memiliki tenor pembayaran kurang dari 4 bulan. Berikan penawaran benefit yang lebih untuk perpanjangan tenor pembayaran

Pencegahan:

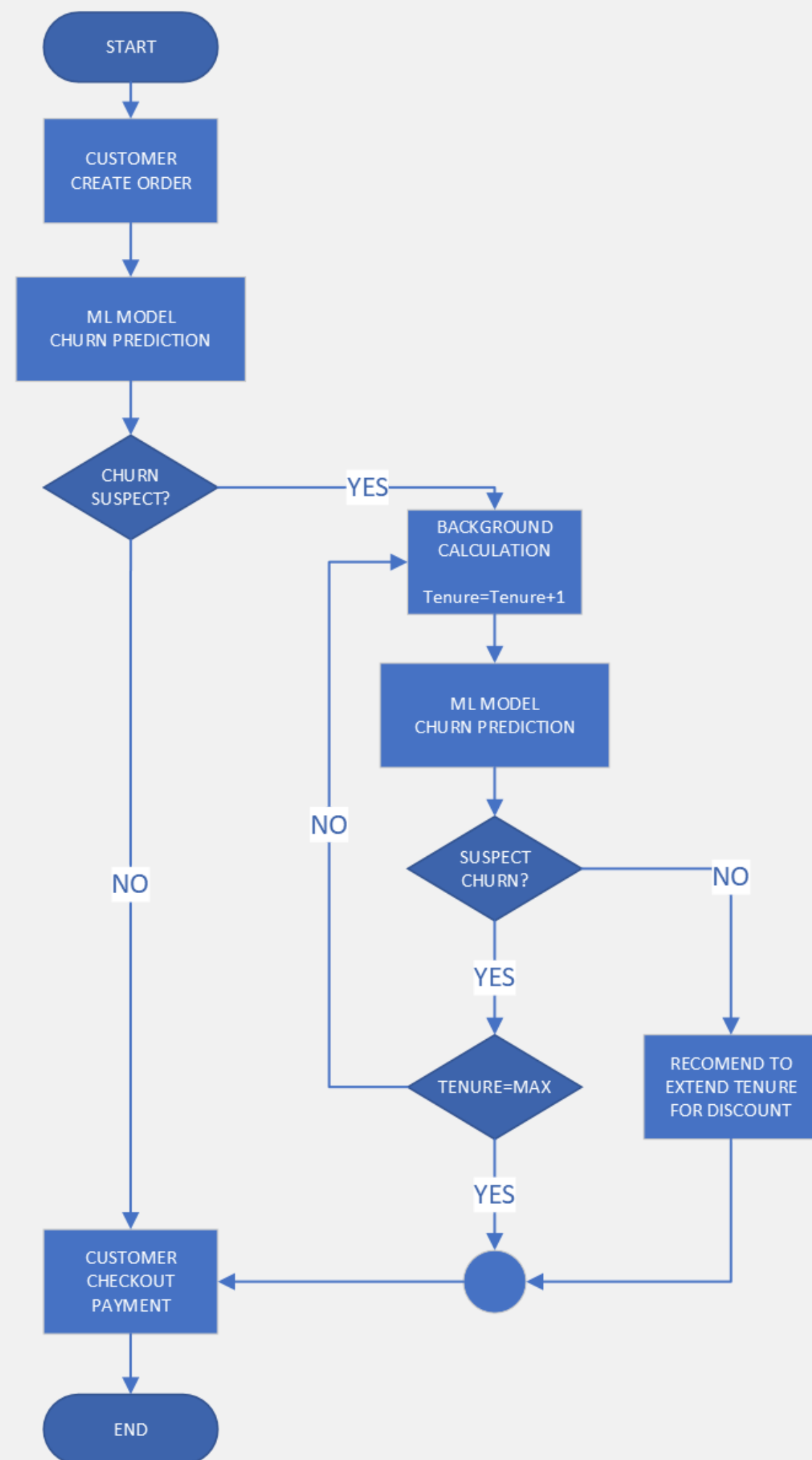
Perbanyak kerjasama dengan penyedia layanan kredit dengan tenor panjang yang memiliki bunga rendah

Benefit Recommendation Business Process

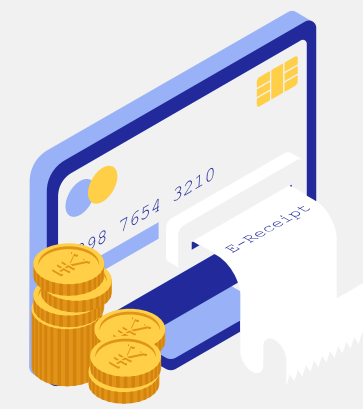
Controlled Variable: Tenure

Pada saat proses transaksi platform e-commerce, customer akan diberi penawaran diskon/cashback apabila:

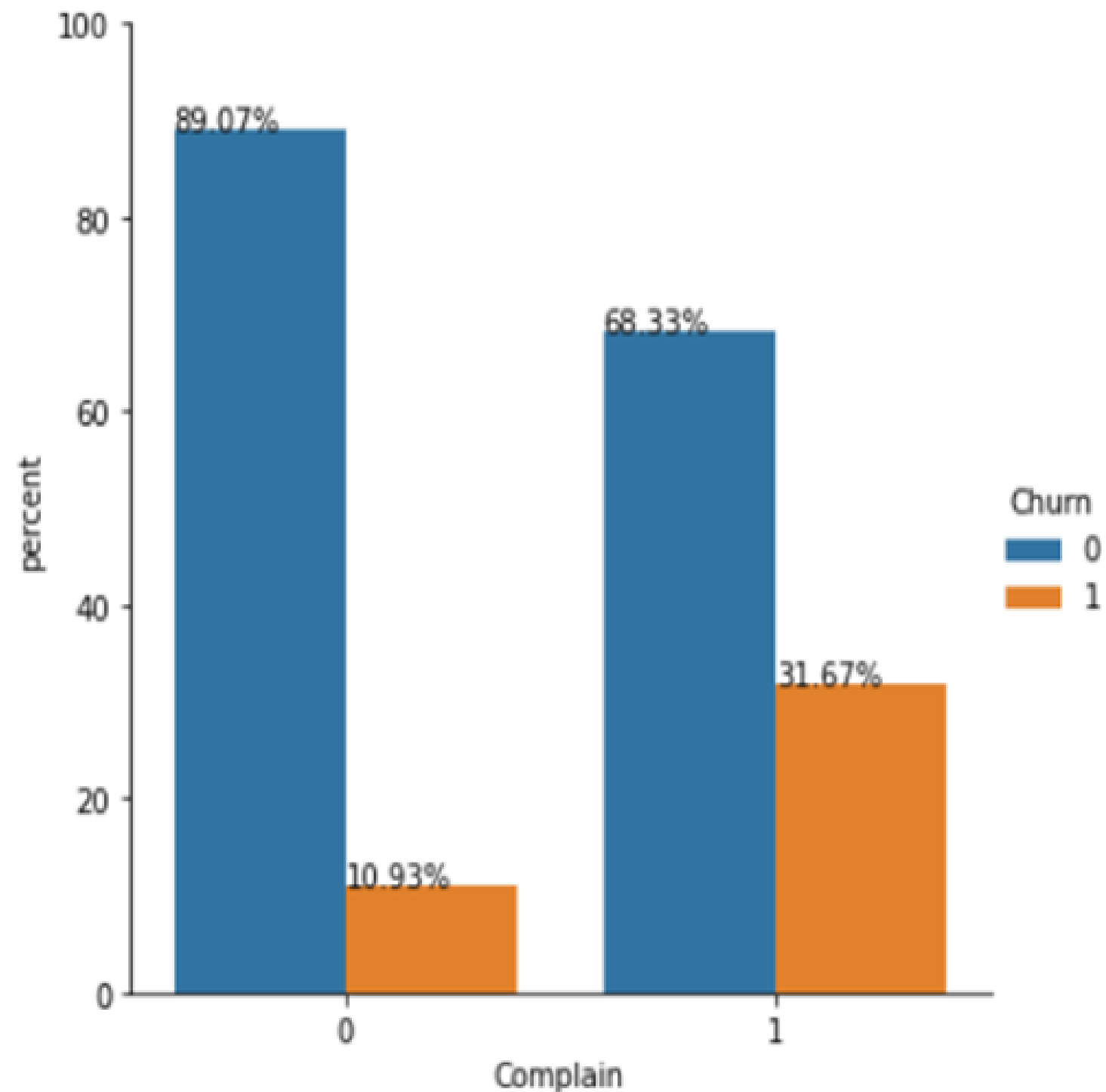
1. Di awal transaksi diprediksi churn
2. Tenure belum mencapai batas maksimal yang diperkenankan
3. Penambahan tenure dapat membuat hasil prediksi menjadi non-churn



Complaint Handling



Churn By Complain



31% customer yang pernah memiliki komplain, melakukan churn

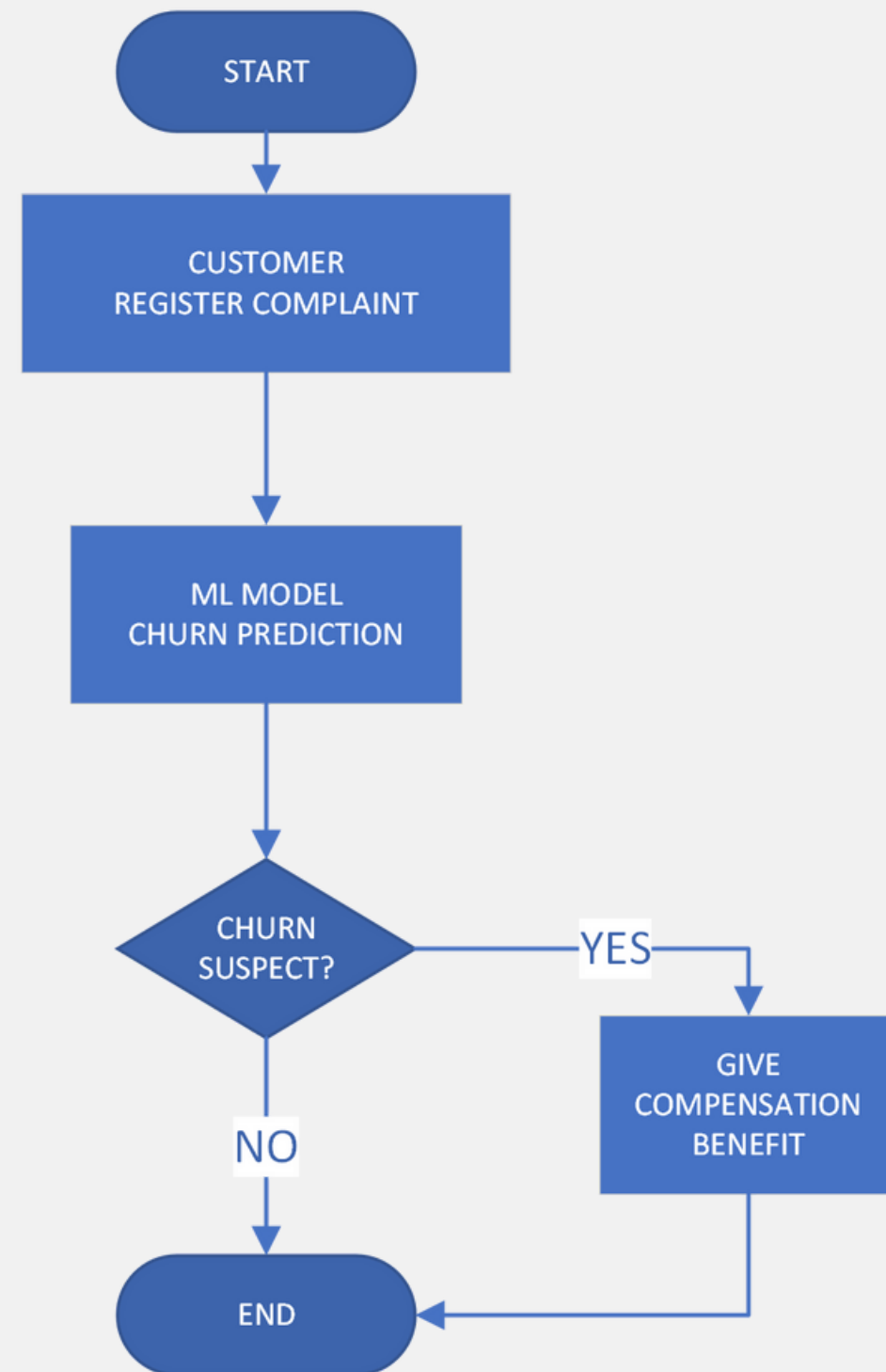
Rencana Mitigasi:

Penanganan komplain yang lebih baik pada customer terduga churn. Non-Aktifkan chat bot dan berikan kompensasi

Rencana Pencegahan

Tingkatkan standar layanan e-commerce untuk menghindari komplain

Complaint Handling Business Process



Pada saat customer melakukan complaint dan diprediksi akan churn pada sebuah transaksi maka perlakukan customer dengan value lebih:

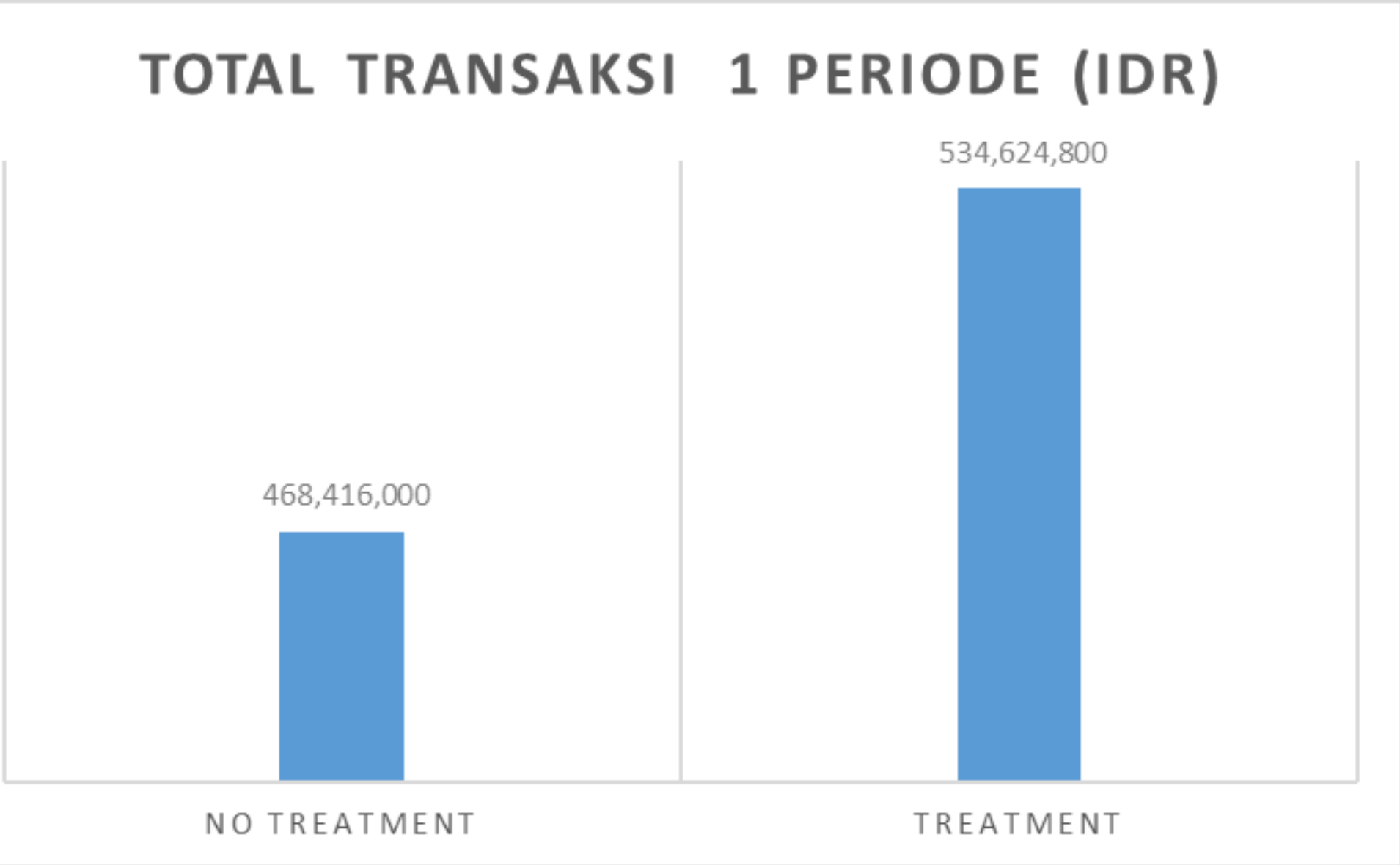
1. Mematikan chatbot, handle dengan customer service
2. Memberikan kompensasi berupa voucher discount

Impact Analysis

Asumsi Dasar:

- 1. Nilai transaksi per user rata-rata IDR 100.000
- 2. Success Ratio churn treatment 70%

Increase Revenue by 14,3%



Parameter	No Treatment	Treatment
Jumlah user	5,630	5,630
Jumlah user churn	946	284
Nilai transaksi @ user	100,000	100,000
Total Transaksi dalam 1 periode	468,416,000	534,624,800
Churn rate	16.8%	5.0%

Terima Kasih

