

# Final Project DS29

Numero Uno



# E-Commerce Churn Prediction

**Tim Numero Uno**

Amala Fahditia

Fawwaz El Gifari

Ismail Ashari

Stephen James

Azam Fathurahman

Marcelius Steven Susanto

Helmy Naufal Aziz

Mentor : Kevin Wu



# Table of Contents

1

...

## Background

Apa latar belakang masalahnya?

2

...

## EDA & Insight

Bagaimana analisis data dan *insight*nya?

3

...

## Pre-processing

Bagaimana mempersiapkan data sebelum Modelling?

4

...

## Modeling

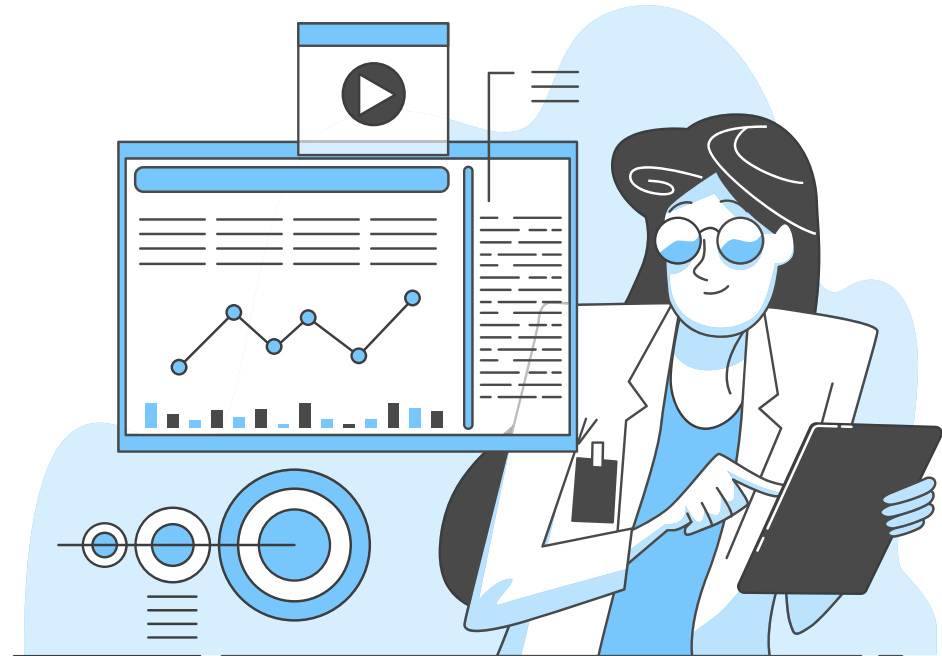
Bagaimana modeling datanya?

5

...

## Recommendation

Apa rekomendasi bisnis yang dapat diberikan?






# 01

# Background

Apa latar belakang masalahnya?



## Transaksi *E-Commerce* Turun, Masyarakat Mulai Banyak Transaksi Offline?

Nilai transaksi e-commerce tercapai Rp 476,3 triliun dengan volume 3.486 juta.

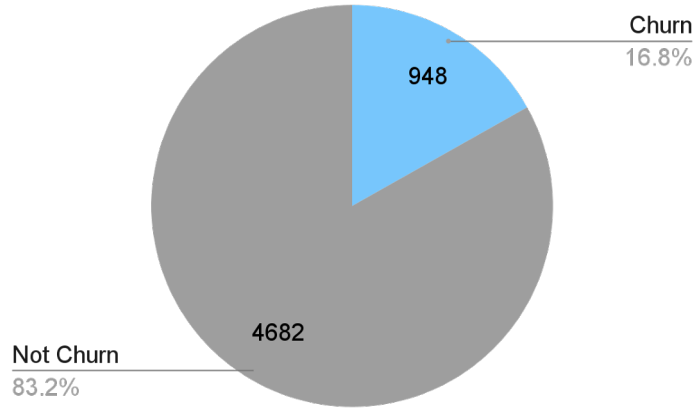
Rep: Rahayu Subekti/ Red: Lida Puspaningtyas

REPUBLIKA.CO.ID, JAKARTA -- Berdasarkan data sementara Bank Indonesia (BI), transaksi *e-commerce* pada 2022 menurun.

Padaahal, BI menargetkan transaksi *e-commerce* pada 2022 bisa mencapai Rp 489 triliun.

"Sampai hari ini ya data sementara kita lihat *e-commerce* tercapai Rp 476,3 triliun dengan volume 3.486 juta artinya sedikit di bawah target," kata Deputy Gubernur **BI** Doni P Joewono dalam konferensi pers Rapat Dewan Gubernur BI Januari 2023, Kamis (19/1/2023).





**1**  
NUMERO  
UNO

## PT. Numero Uno

Tim *data scientist* di perusahaan *e-commerce* Numero Uno bertugas untuk memberikan rekomendasi bisnis berdasarkan data untuk mengurangi persentase *customer* yang *churn*.

...

## What is a "good" churn rate?

This question is a bit harder to answer and is subjective to your industry and company age.

For SaaS companies, average churn rates are anywhere from 2% - 8% of monthly recurring revenue, per churn studies.

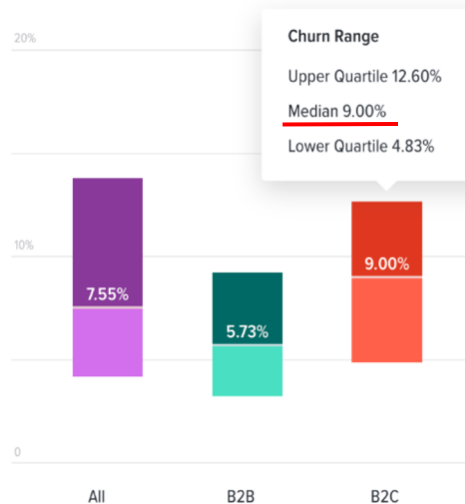
### Chargebee Blog

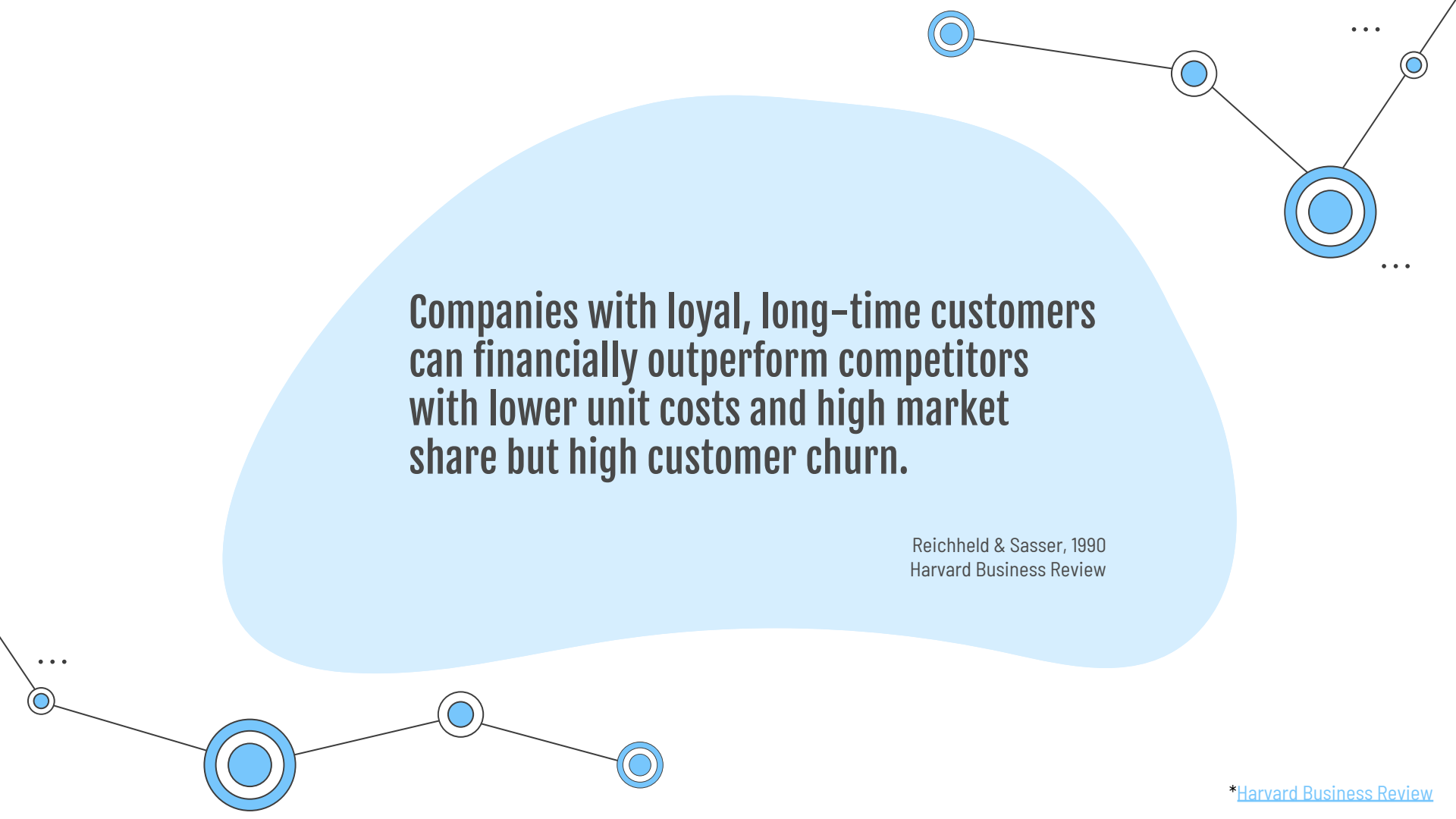
## What is a Good eCommerce Churn Rate?

It's widely believed that anything under 5% is a good churn rate for eCommerce companies. However, that's not always the case. Some enterprise companies might see churn rates of up to 10% and consider them a success.

## Recurly Research

### Consumer Goods



A decorative network diagram consisting of blue circles of varying sizes connected by thin black lines. The circles are arranged in a non-linear fashion, with some having concentric rings. Ellipses (...) are placed near some of the circles, suggesting a larger, interconnected network.

**Companies with loyal, long-time customers  
can financially outperform competitors  
with lower unit costs and high market  
share but high customer churn.**

Reichheld & Sasser, 1990  
Harvard Business Review



# Understanding the Problem



## Goal

Mengurangi *churn rate*  
hingga di bawah 10%

...



## Objective

Mengidentifikasi  
penyebab *churn*,  
membuat model  
prediksi, memberi  
rekomendasi bisnis

...



## Business Metric

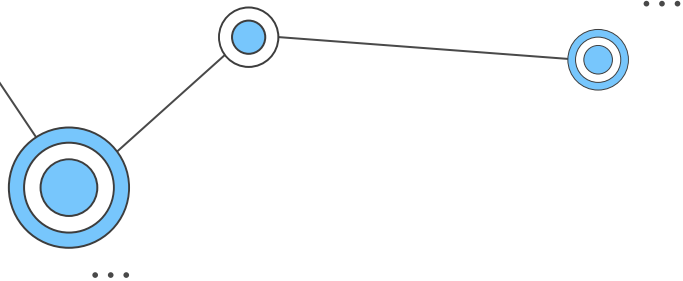
*Customer churn rate*

...

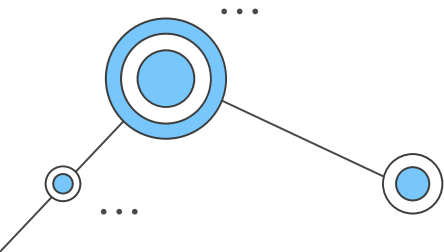
# 02

## EDA & Insight

Bagaimana analisis data dan  
*insightnya?*



- 20 kolom, 5630 baris
- Beberapa kolom memiliki nilai kosong
- Beberapa tipe data kurang sesuai
- Tidak ada data duplikat
- Kolom *Churn* yang akan menjadi target



### Incorrect Data Type

CustomerID  
int64 → object

CityTier  
int64 → object

CouponUsed  
float64 → int64

OrderCount  
float64 → int64

DaySinceLastOrder  
float64 → int64

### Missing Values

Tenure  
264 rows

WarehouseToHome  
251 rows

HourSpendOnApp  
255 rows

OrderAmountHikeFromLastYear  
265 rows

CouponUsed  
256 rows

OrderCount  
258 rows

DaySinceLastOrder  
307 rows

# Exploratory Data Analysis

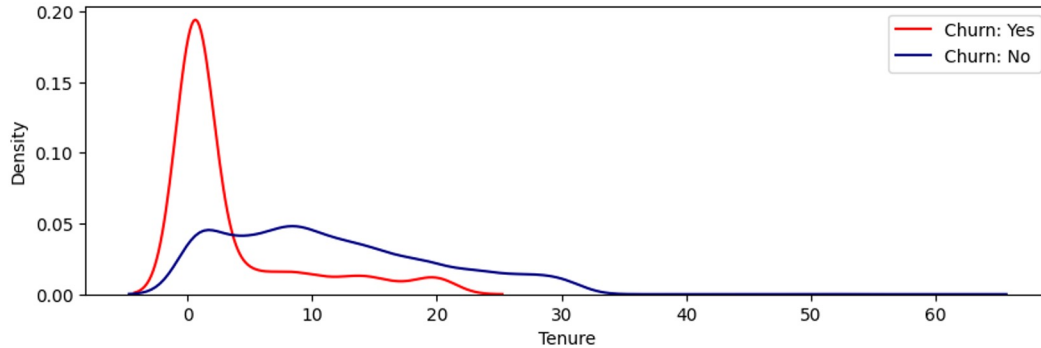
**Statistical** Descriptive

**Univariate** Analysis

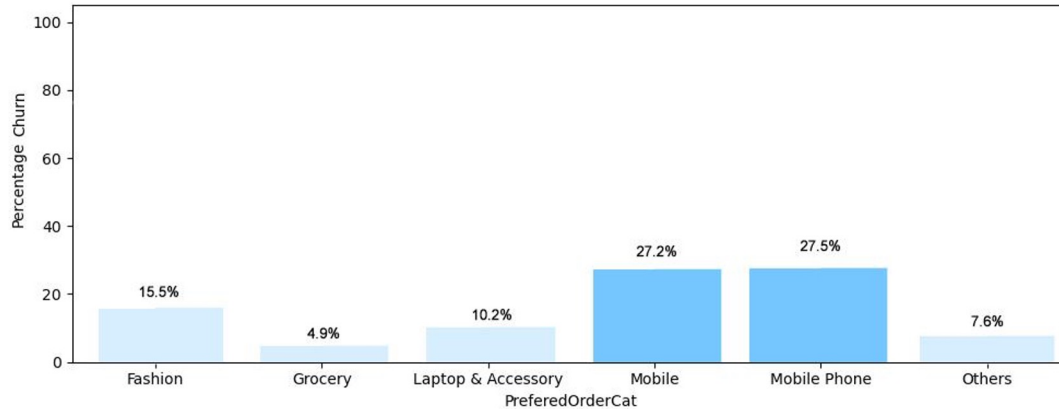
**Multivariate** Analysis



# Insight



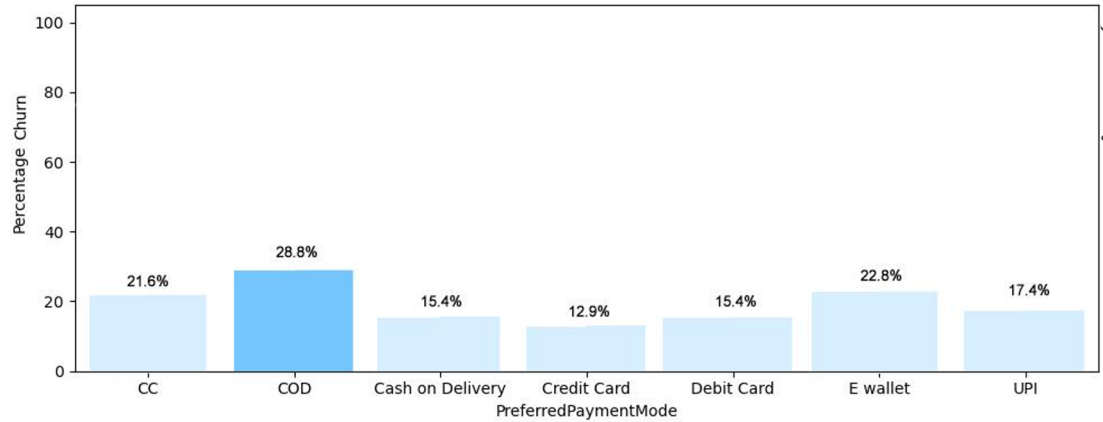
Customer berpotensi churn pada **bulan-bulan awal**.



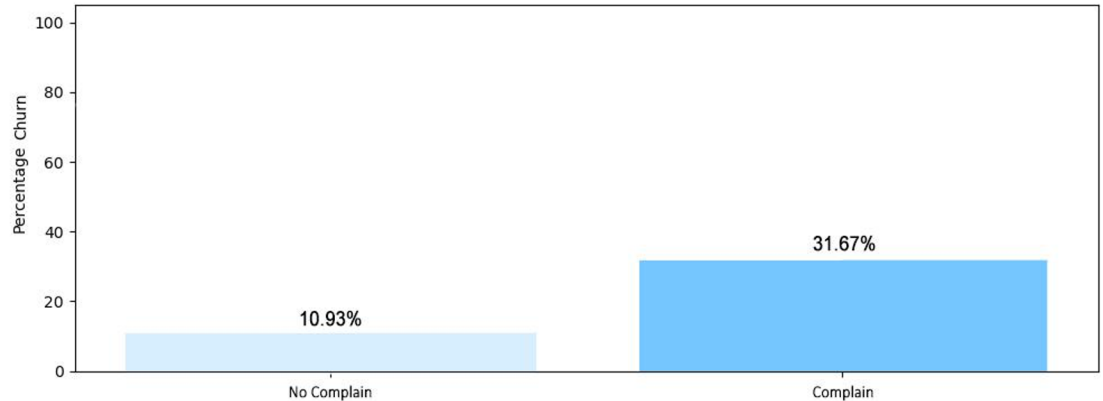
Customer yang melakukan transaksi pada kategori **Mobile** dan **Mobile Phone** memiliki kemungkinan churn tertinggi.

# Insight

Customer yang melakukan transaksi menggunakan metode **COD** memiliki kemungkinan churn tertinggi.



Customer yang **complain** memiliki kemungkinan yang tinggi untuk churn.



# 03

## Pre-Processing

Bagaimana mempersiapkan data  
sebelum Modelling?



# Data Pre-Processing

01

**Mengisi data yang kosong**

Right Skewed: Median  
Normal: Modus

02

**Membuang data Outliers**

Z-Score, 280 Outliers  
Loss

03

**Menambah Fitur baru**

AvgCashback

04

**Mentransformasi Fitur**

Log Transformation  
dan Standardization





# Data Pre-Processing



05

## Mengubah Fitur Kategorikal ke Fitur Numerikal

Label Encoding & One Hot Encoding

06

## Memilih Fitur

Satu target (Output) dan  
sebelas fitur (Input)

Churn

CityTier

NumberOfDeviceRegistered

SatisfactionScore

MaritalStatus

Tenure

CashbackAmount

DaySinceLastOrder

Complain

PreferredOrderCat  
Grocery

PreferredOrderCat  
Handphone

PreferredOrderCat  
Laptop

# 04

## Modeling

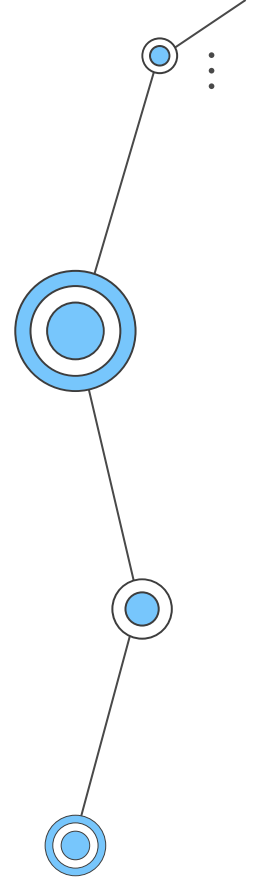
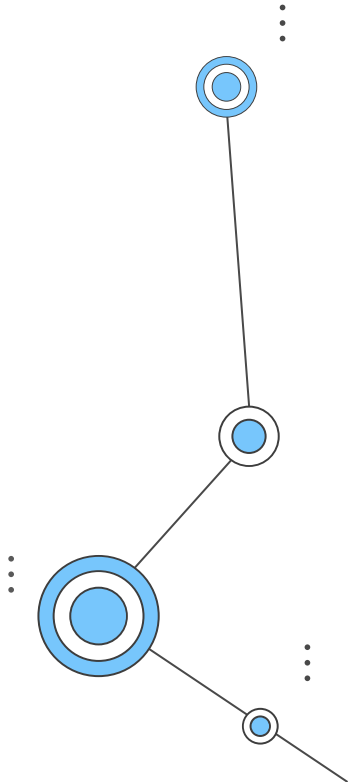
Bagaimana modeling datanya?

# Modeling

Split Data Set

Implementasi Algoritma

Modeling Impact



# Split Data Train & Data test

## Train

X\_train: 3745 row & 11 columns  
y\_train: 3745 row  
(0 = 3108, 1 = 637 )

## Test

X\_test: 1605 rows & 11 columns  
y\_test: 1605 rows

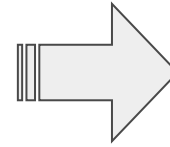
Hasil Data SMOTE → untuk Pembelajaran Model dan Memprediksi Target

X\_smote: 6216 rows dan 11 columns

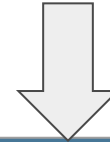
y\_smote: 6216 rows (0 → 3108, 1 → 3108)

# Implementasi Algoritma

Model	F2 Test	F2 Train
Logistic Regression	0,807	0,794
Decision Tree	0,943	1,0
KNeighbors	0,859	0,903
Random Forest	0,959	1,0
Ada Boost	0,853	0,862
Gradient Boosting	0,885	0,893
XGBoost	0,887	0,895



***Hyperparameter Tuning***

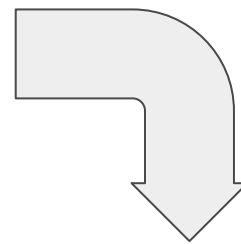


XGBOOST	
F2 Test	0,959
F2 Train	0,998

# Confusion Matrix

On Original Data

ACTUAL	Not Churn	Churn
	Not Churn	Churn
Not Churn	<b>True Negative</b> 4,373 81.74%	<b>False Positive</b> 65 1.21%
Churn	<b>False Negative</b> 48 0.90%	<b>True Positive</b> 864 16.15%



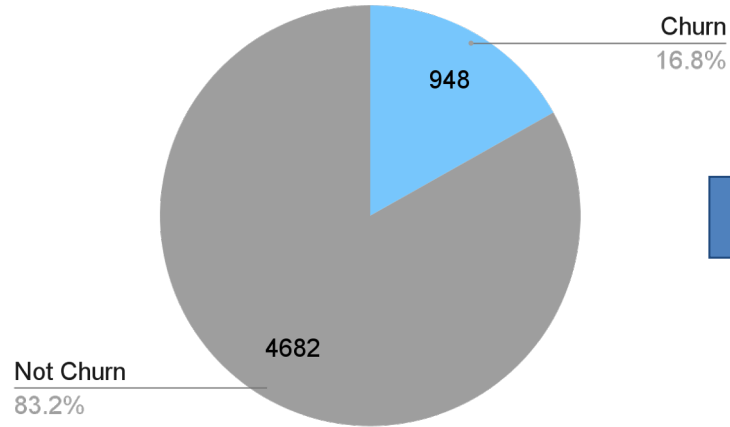
After Modeling

- (True Positive) 864 customers
- (False Positive) 65 customers
- Total 929 customers

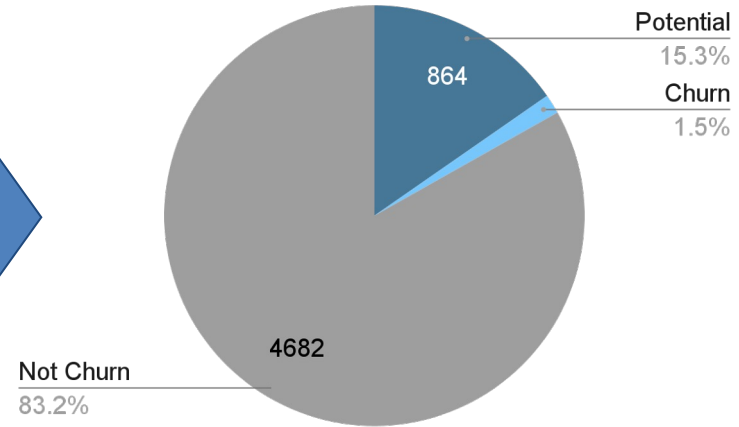
93% Prec

# Modeling Impact

## Before Modeling



## After Modeling



%Churn Predicted: **93%** → **864** customer

Jika semua customer yang terdeteksi churn dapat diantisipasi maka customer churn rate berpotensi berkurang hingga **1.5%**

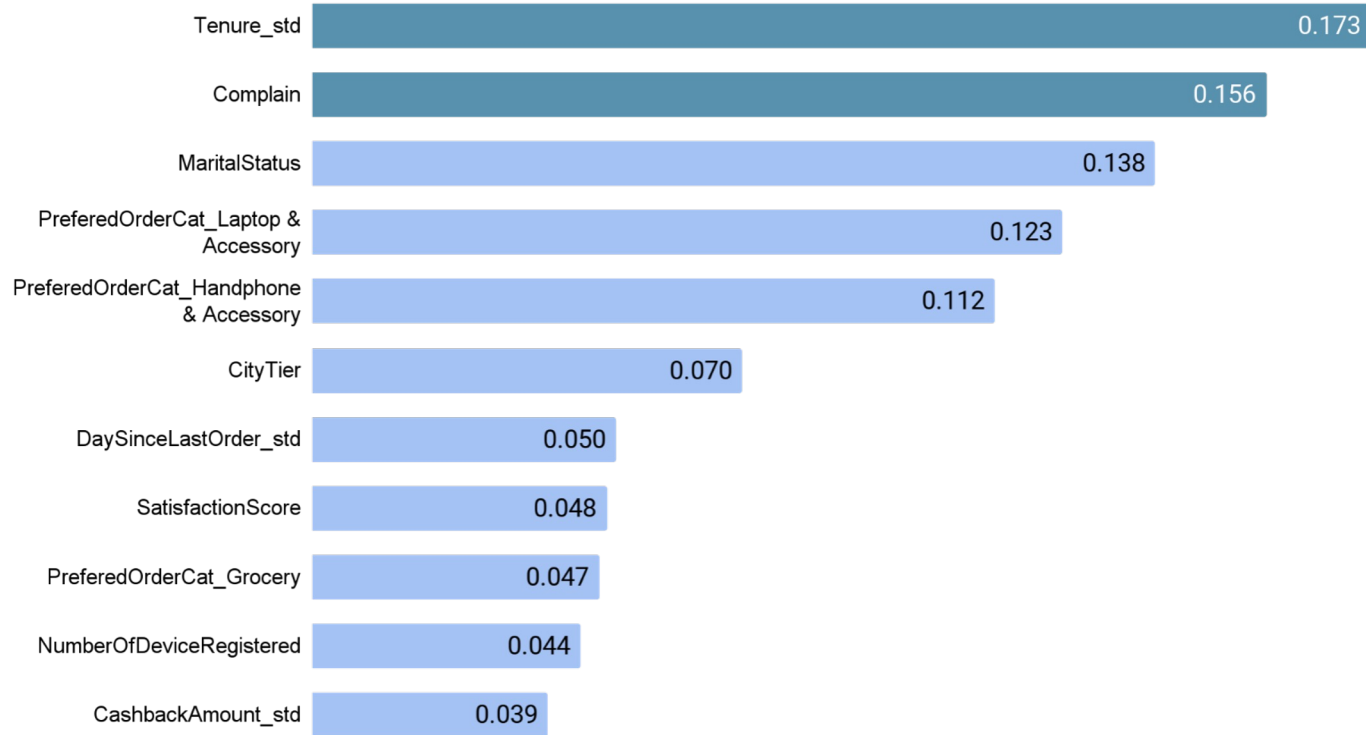
# 05

## Recommendation

Apa rekomendasi bisnis yang dapat  
diberikan?



# Feature Importance



...

...

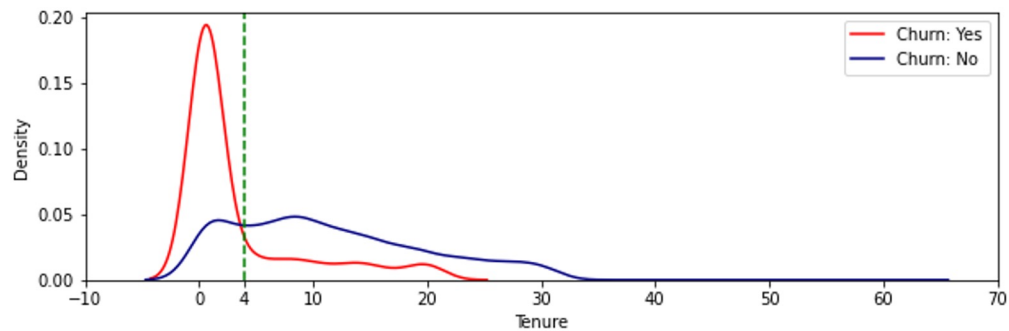
...

Untuk customer dengan  
tenure di bawah 5 bulan  
diberikan **promo** lebih  
banyak

...

Promo apa ya?

Mengapa tenure di bawah 5 bulan?



Terdapat 71% customer yang churn memiliki tenure di bawah 5 bulan

...

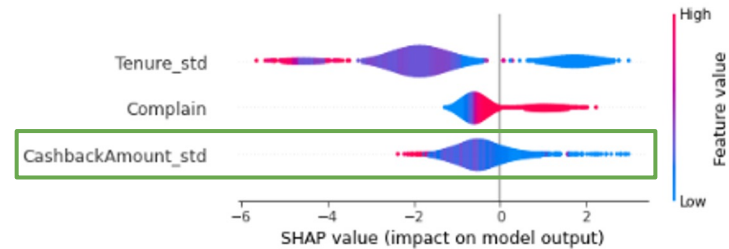
...

Untuk customer dengan  
tenure di bawah 5 bulan  
diberikan **cashback** lebih  
banyak

...

...

### Mengapa promo cashback?



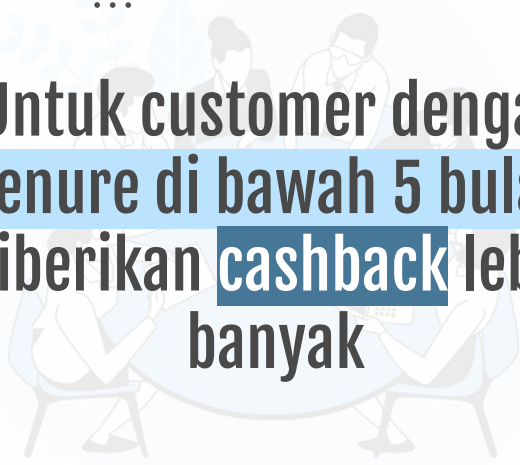
Cashback yang rendah dapat meningkatkan kemungkinan customer untuk churn

...



...

Untuk customer dengan  
tenure di bawah 5 bulan  
diberikan **cashback** lebih  
banyak

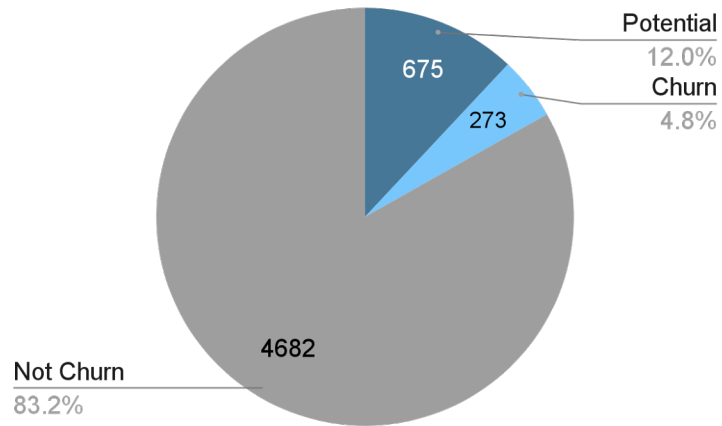


...



...

Seberapa besar pengaruhnya terhadap churn rate?

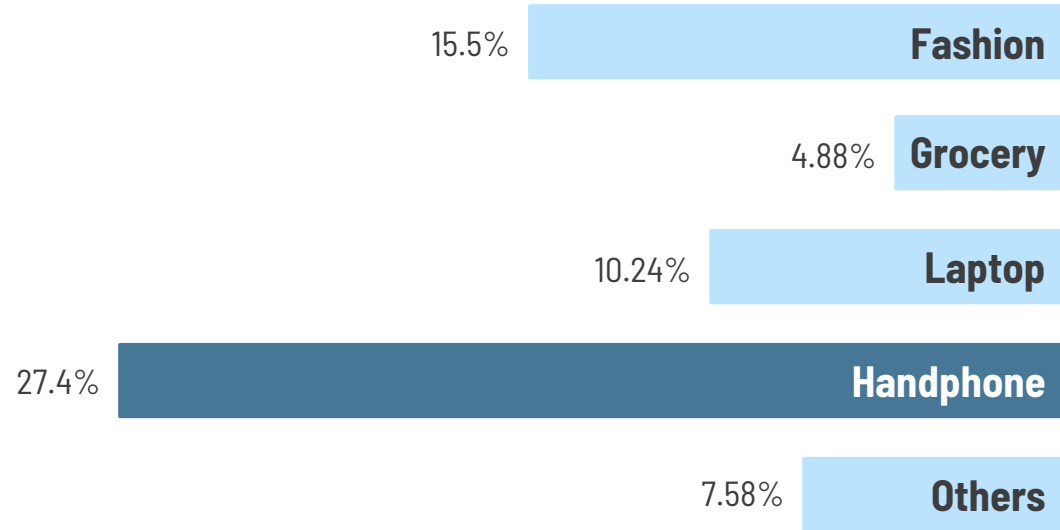


Jika customer yang churn dengan tenure di bawah 5 bulan dapat diantisipasi, terdapat potensi penurunan churn rate sebesar 12%



Untuk mengoptimalkan budget, cashback dapat difokuskan pada kategori Handphone

### Churn rate by preferred order category



Kategori Handphone memiliki churn rate tertinggi dengan 27.4%

...



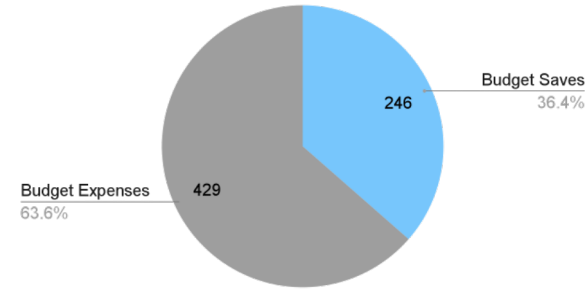
...

Untuk mengoptimalkan budget, cashback dapat difokuskan pada kategori Handphone

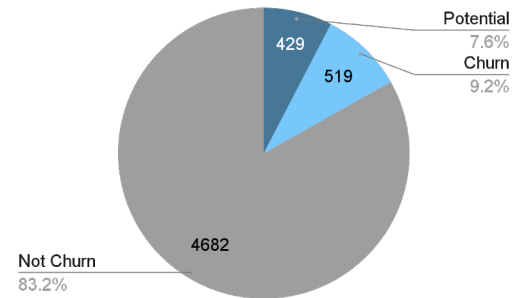
...



### Seberapa besar budget dioptimalkan?



### Seberapa besar pengaruhnya terhadap churn rate?



Jika kedua rekomendasi diaplikasikan, terdapat potensi penurunan churn rate sebesar 7.6% dengan optimalisasi budget sebesar 36.4%



## Meningkatkan kualitas customer service



“96% of consumers say  
customer service is key to their  
brand loyalty.”

Microsoft, 2017  
State of Global Customer Service

“91% of consumers are more  
likely to make another purchase  
after a positive customer  
service experience.”

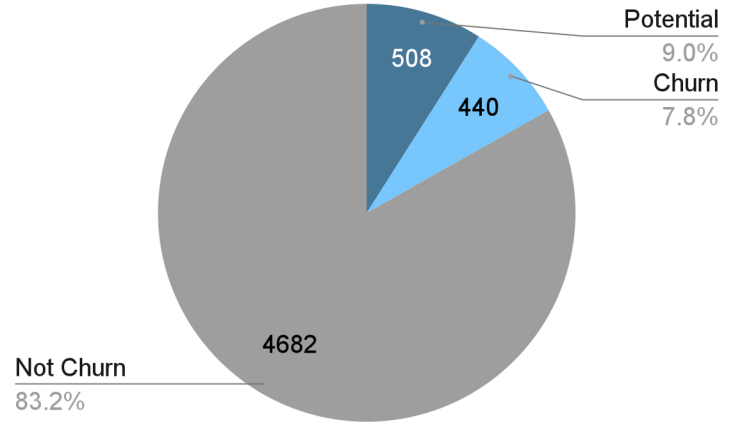
Salesforce, 2020  
State of The Connected Customer



## Meningkatkan kualitas customer service



### Seberapa besar pengaruhnya terhadap churn rate?



Jika customer yang complain dapat teratasi maka potensi penurunan churn rate sebesar 9%





## Further Improvements

# 01

### Categorize each feedback

Dapat memprioritaskan masalah yang harus diselesaikan

# 02

### Create a Chatbot

Mempercepat penanganan complain

# 03

### Start a reward and referral program

Potensi untuk meningkatkan loyalitas customer dan mendapatkan customer baru

Terima Kasih

