



# **E-COMMERCE** **CUSTOMERS CHURN** **ANALYSIS** **AND PREDICTION**

**FINAL PROJECT - BATCH 22**

**BUKAPEDIA**

# OUR TEAM



**FILDZAH**



**NABILAH**



**WAHYUNI**



**ACHEL**



**ALIFIAN**



**SAKTI**



**RYAN**



**BUKAPEDIA**

# OUTLINE



**01 BUSINESS UNDERSTANDING**

**02 DATA OVERVIEW**

**03 DATA CLEANING & PREPARATION**

**04 MODEL & EVALUATION**

**05 BUSINESS IMPACT & RECOMMENDATION**

# OUTLINE



**01 BUSINESS UNDERSTANDING**

**02 DATA OVERVIEW**

**03 DATA CLEANING & PREPARATION**

**04 MODEL & EVALUATION**

**05 BUSINESS IMPACT & RECOMMENDATION**



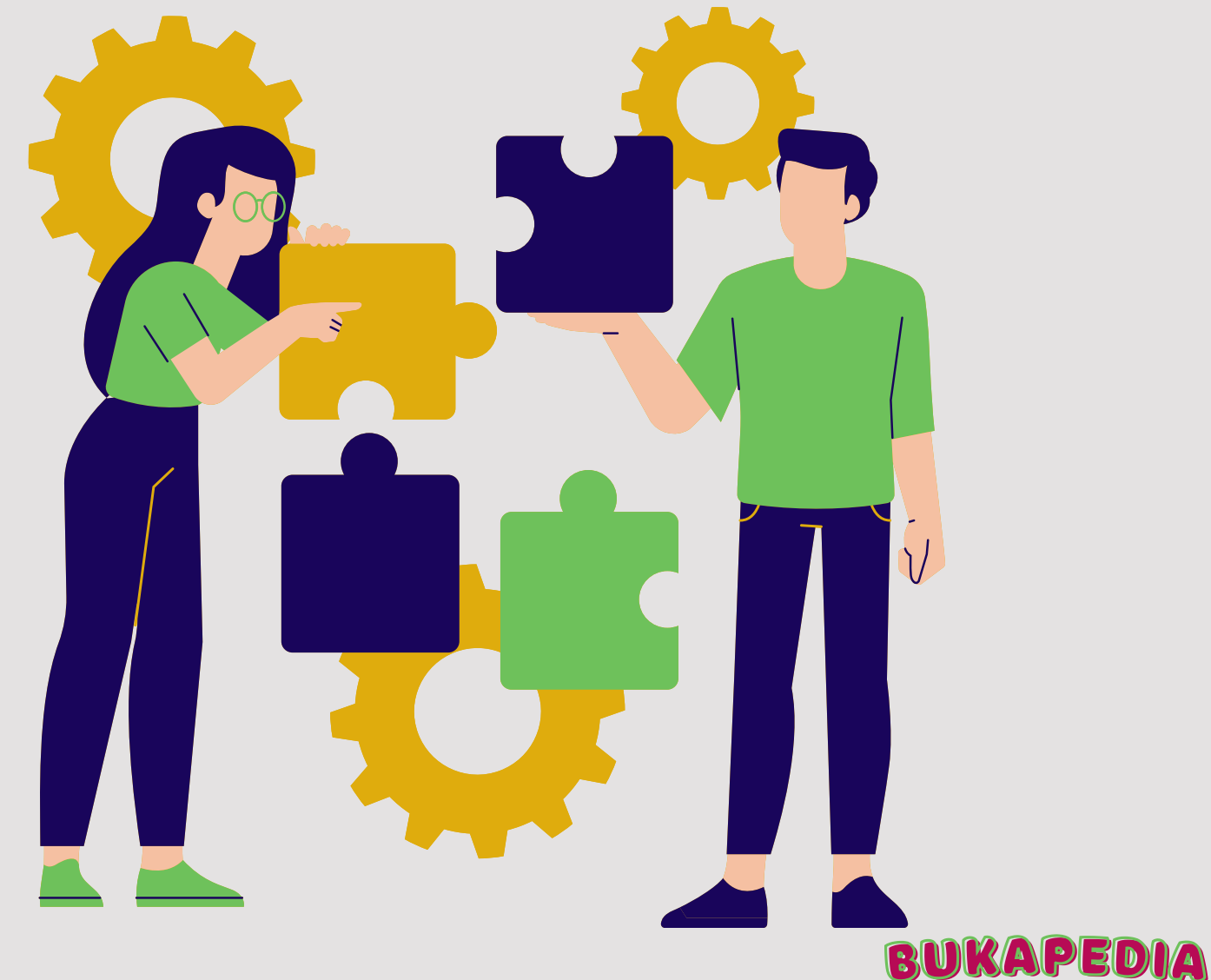


# WHO WE ARE?

Sebagai tim **data scientist** di perusahaan Bukapedia yang mampu memberikan **business recommendation** untuk menyelesaikan masalah yang ada di perusahaan.

# WHAT CASE?

Bukapedia adalah e-commerce di Indonesia. Bukapedia menjual berbagai jenis produk, mulai dari Grocery, Electronic, Fashion, Laptop & Accessory dan lain-lain. Namun, Bukapedia memiliki masalah terhadap **customer churn**.



BUKAPEDIA



# CHURN

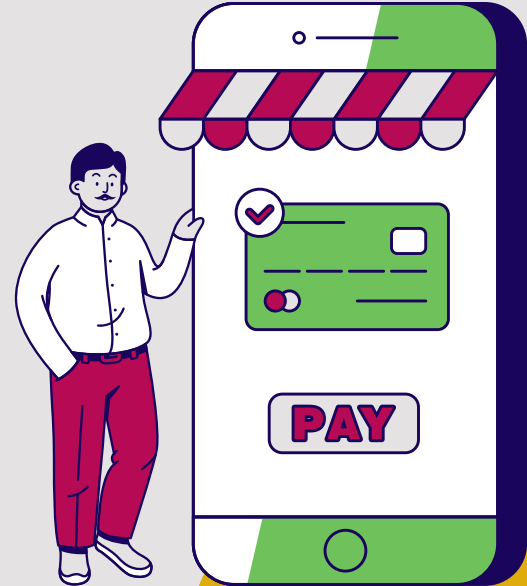
## WHAT IS CHURN?

**Churn** adalah kondisi dimana customers **berhenti berlangganan** atau **meninggalkan layanan**.

## HOW TYPES OF CHURN RATE?

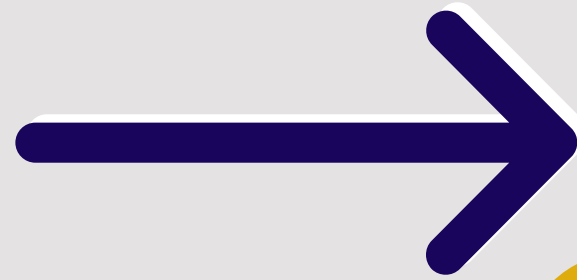
1. Customer Churn Rate
2. Employee Churn Rate

# PROBLEM STATEMENT



**948**

**CUSTOMERS CHURN**



**16,8%**

**CHURN RATE**



**\$ 216.144 /YEAR**

$$\text{CHURN RATE} = \frac{\text{CUSTOMER CHURN}}{\text{TOTAL CUSTOMERS}}$$

**LOST OPPORTUNITY**

$$\text{LOST OPPORTUNITY} = \text{TOTAL CUSTOMERS} \times \text{AVERAGE MONTHLY SPENDING USER}$$

## OBJECTIVE

MEMPREDIKSI CUSTOMER YANG BERPOTENSI **CHURN** DENGAN MEMBUAT MACHINE LEARNING MODEL. SERTA MEMBERIKAN **INSIGHT & REKOMENDASI** UNTUK MENGURANGI **CHURN RATE**

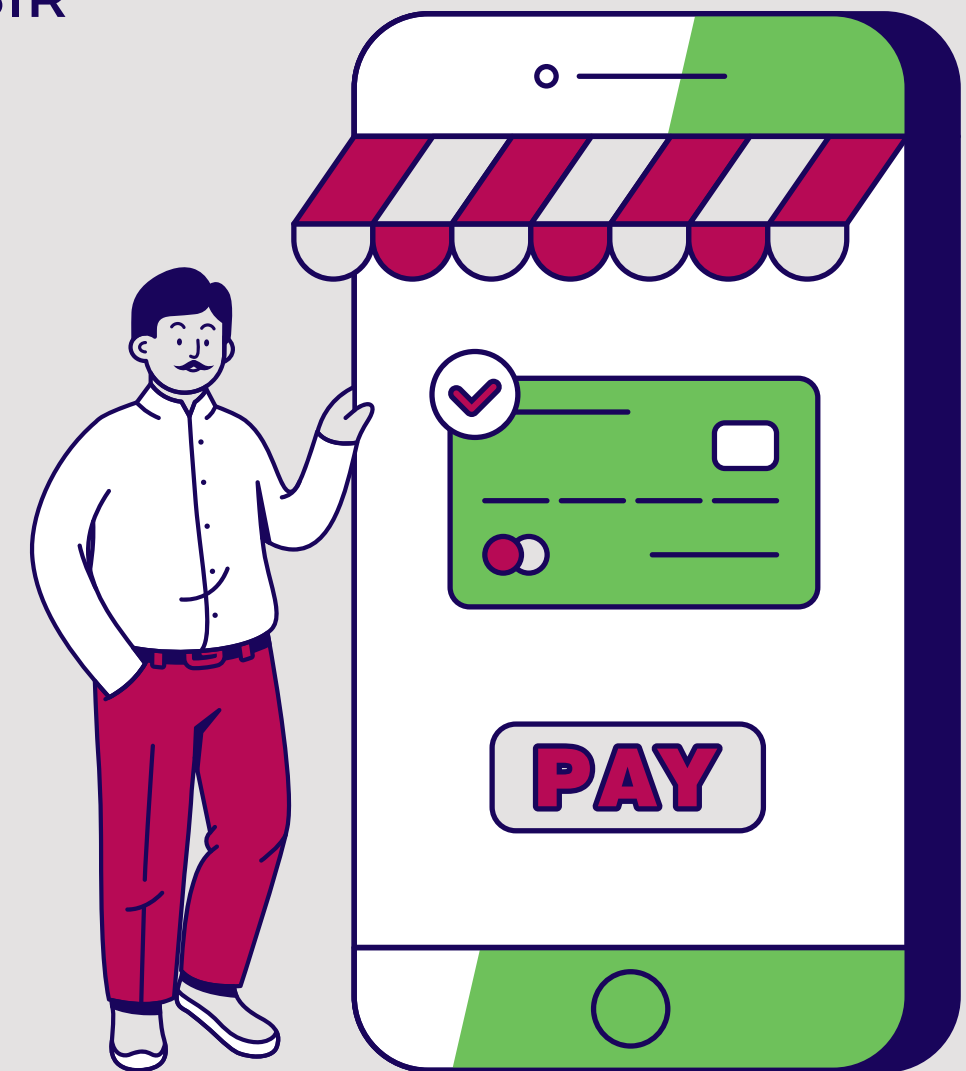
## GOALS

MENURUNKAN **CHURN RATE** DAN MEMINIMALISIR POTENTIAL LOSS AKIBAT CHURN

## BUSINESS METRICS

$$\text{CHURN RATE} = \frac{\text{CUSTOMER CHURN}}{\text{TOTAL CUSTOMERS}}$$

$$\text{LOST OPPORTUNITY} = \text{TOTAL CUSTOMERS COMPLAIN \& BERPOTENSI CHURN} \times \text{AVERAGE MONTHLY SPENDING USER}$$





# OUTLINE



01 BUSINESS UNDERSTANDING

02 DATA OVERVIEW

03 DATA CLEANING & PREPARATION

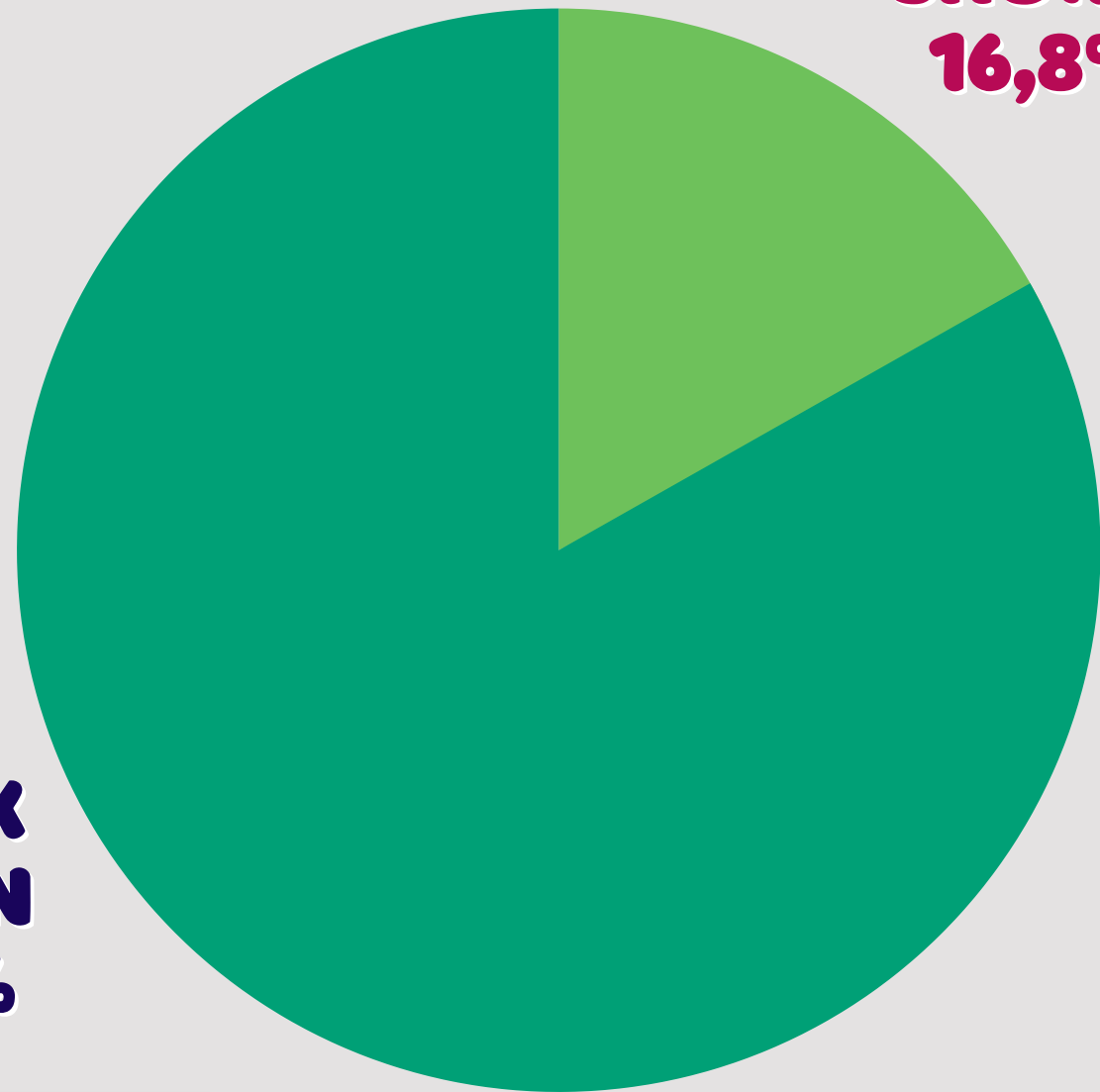
04 MODEL & EVALUATION

05 BUSINESS IMPACT & RECOMMENDATION

# DATA OVERVIEW

**CHURN**  
**16,8%**

**TIDAK  
CHURN**  
**83,2%**



## ABOUT DATASET

Dataset berisi data dari **satu tahun** penjualan di Ecommerce Bukapedia

**5.630 BARIS**

**20 KOLOM**

**BUKAPEDIA**



# DATA OVERVIEW

| Variable                    | Description  |
|-----------------------------|--|
| Churn                       | Churn Flag   |
| CityTier                    | Tingkatan kota   |
| SatisfactionScore           | Skor kepuasan pelanggan terhadap pelayanan                 |
| Complain                    | Keluhan yang diajukan pada bulan lalu                      |
| PreferredLoginDevice        | Perangkat login yang digunakan                             |
| PreferedOrderCat            | Kategori pesanan pilihan customer pada bulan lalu          |
| Gender                      | Jenis kelamin customer                                     |
| PreferredPaymentMode        | Metode pembayaran yang digunakan customer                  |
| MaritalStatus               | Status perkawinan  |
| CustomerID                  | ID Pelanggan   |
| Tenure                      | Masa penggunaan e-commerce                                 |
| WarehouseToHome             | Jarak antara Gudang ke rumah pelanggan                     |
| HourSpendOnApp              | Waktu yang dihabiskan di aplikasi seluler maupun situs web |
| NumberOfDeviceRegistered    | Total perangkat yang terdaftar                             |
| NumberOfAddress             | Total alamat yang dimasukkan user dalam setiap akun        |
| OrderAmountHikeFromlastYear | Presentase peningkatan secara berurutan dari tahun lalu    |
| CouponUsed                  | Kupon yang telah digunakan pada bulan lalu                 |
| OrderCount                  | Total pesanan yang telah dilakukan pada bulan lalu         |
| DaySinceLastOrder           | Hari sejak pesanan terakhir                                |
| CashbackAmount              | Cashback rata-rata di bulan lalu                           |



TARGET



3 FEATURES NOMINAL



5 FEATURES ORDINAL



10 FEATURES INTERVAL

# OUTLINE



01 BUSINESS UNDERSTANDING

02 DATA OVERVIEW

**03 DATA CLEANING & PREPARATION**

04 MODEL & EVALUATION

05 BUSINESS IMPACT & RECOMMENDATION

# DATA CLEANING & PREPARATION

One Hot Encoding : Category feature  
Label Encoding : Gender

**DATASET  
CHURN**

**SPLIT DATA**

| Data  | Percent | Rows |
|-------|---------|------|
| TRAIN | 80%     | 4504 |
| TEST  | 20%     | 1126 |

**ENCODING  
CATEGORICAL  
FEATURES**

**HANDLE  
MISSING  
VALUE**

Total missing value : 1,856  
Categorical : Imputasi Modus  
Numerikal : Imputasi Median



# OUTLINE



**01 BUSINESS UNDERSTANDING**

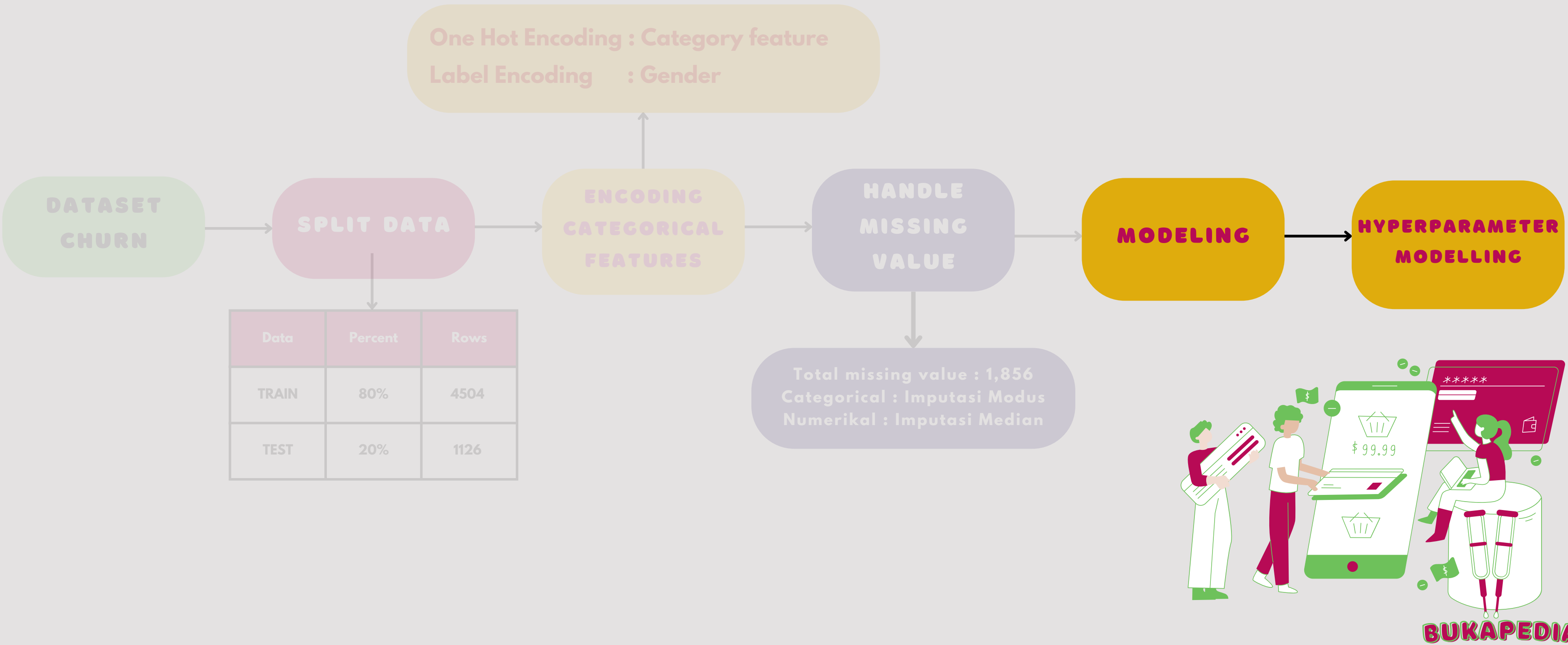
**02 DATA OVERVIEW**

**03 DATA CLEANING & PREPARATION**

**04 MODEL & EVALUATION**

**05 BUSINESS IMPACT & RECOMMENDATION**

# MODEL & EVALUATION



# RESULT MODELING

| MODEL               | RECALL |      | PRECISION |      | ROC_AUC |      | F1    |      | ACCURACY |      |
|---------------------|--------|------|-----------|------|---------|------|-------|------|----------|------|
|                     | TRAIN  | TEST | TRAIN     | TEST | TRAIN   | TEST | TRAIN | TEST | TRAIN    | TEST |
| Decision Tree       | 1      | 0.87 | 1         | 0.92 | 1       | 0.92 | 1     | 0.88 | 1        | 0.92 |
| Xgboost             | 0.66   | 0.62 | 0.88      | 0.85 | 0.82    | 0.80 | 0.76  | 0.72 | 0.82     | 0.80 |
| Catboost            | 0.95   | 0.81 | 0.99      | 0.97 | 0.97    | 0.90 | 0.97  | 0.88 | 0.97     | 0.90 |
| Adaboost            | 0.5    | 0.61 | 0.75      | 0.80 | 0.77    | 0.79 | 0.66  | 0.69 | 0.77     | 0.79 |
| Random Forest       | 1      | 0.82 | 1         | 1    | 1       | 0.91 | 1     | 0.90 | 1        | 0.91 |
| Logistic Regression | 0.51   | 0.55 | 0.72      | 0.79 | 0.77    | 0.76 | 0.66  | 0.65 | 0.77     | 0.76 |

Berdasarkan hasil modelling kami, decision tree dapat menentukan prediksi **Churn** terbaik dengan recall sebesar **87%**.



# RESULT HYPERPARAMETER MODELING

| MODEL               | RECALL |      | PRECISION |      | ROC_AUC |      | F1    |      | ACCURACY |      |
|---------------------|--------|------|-----------|------|---------|------|-------|------|----------|------|
|                     | TRAIN  | TEST | TRAIN     | TEST | TRAIN   | TEST | TRAIN | TEST | TRAIN    | TEST |
| Decision Tree       | 0.85   | 0.72 | 0.87      | 0.74 | 0.91    | 0.84 | 0.86  | 0.73 | 0.91     | 0.84 |
| Xgboost             | 0.99   | 0.89 | 1         | 0.9  | 0.99    | 0.95 | 0.99  | 0.94 | 0.99     | 0.95 |
| Catboost            | 1      | 0.93 | 1         | 0.99 | 1       | 0.96 | 1     | 0.96 | 1        | 0.97 |
| Adaboost            | 0.5    | 0.62 | 0.76      | 0.82 | 0.78    | 0.79 | 0.67  | 0.71 | 0.78     | 0.79 |
| Random Forest       | 0.99   | 0.79 | 1         | 1    | 0.99    | 0.89 | 0.99  | 0.88 | 0.99     | 0.91 |
| Logistic Regression | 0.26   | 0.24 | 0.79      | 0.83 | 0.62    | 0.62 | 0.40  | 0.38 | 0.62     | 0.61 |

## WHY USING CATBOOST ?

Setelah hyperparameter tuning. Hasil modeling catboost menunjukan hasil yang lebih baik untuk prediksi **Churn** terbaik dengan recall **93%**



# CONFUSION MATRIX

|                  | PREDICTED NOT CHURN       | PREDICTED CHURN           |
|------------------|---------------------------|---------------------------|
| ACTUAL NOT CHURN | 940<br>TRUE NEGATIVE (TN) | 1<br>FALSE POSITIVE (FP)  |
| ACTUAL CHURN     | 12<br>FALSE NEGATIVE (FN) | 173<br>TRUE POSITIVE (TP) |



$$\text{RECALL} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

## WHY USING RECALL ?

Tujuan untuk mengcapture sebanyak mungkin customer yang actual **Churn**

## SECONDAY MATRIX : PRECISION

Untuk mendapatkan hasil prediksi customer **Churn** terbaik



# OUTLINE



01 BUSINESS UNDERSTANDING

02 DATA OVERVIEW

03 DATA CLEANING & PREPARATION

04 MODEL & EVALUATION

05 BUSINESS IMPACT & RECOMMENDATION

# BUSINESS RECOMMENDATION (PRIORITY)

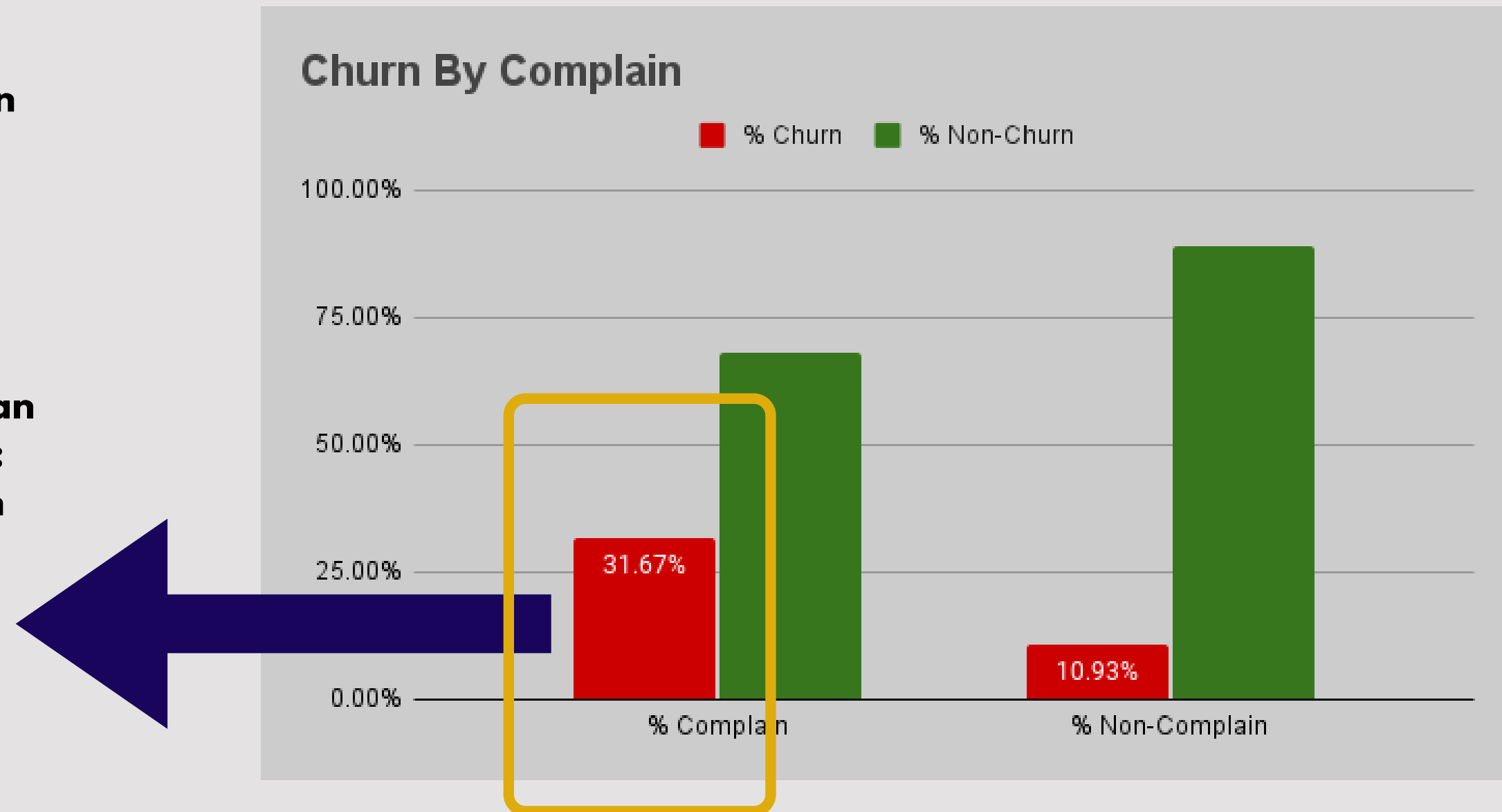
## INSIGHT

Persentase churn akibat customer complain (31.67%) hampir 3 kali lebih besar dari persentase churn total customer yang tidak complain (10.93%).

## BUSINESS RECOMMENDATION

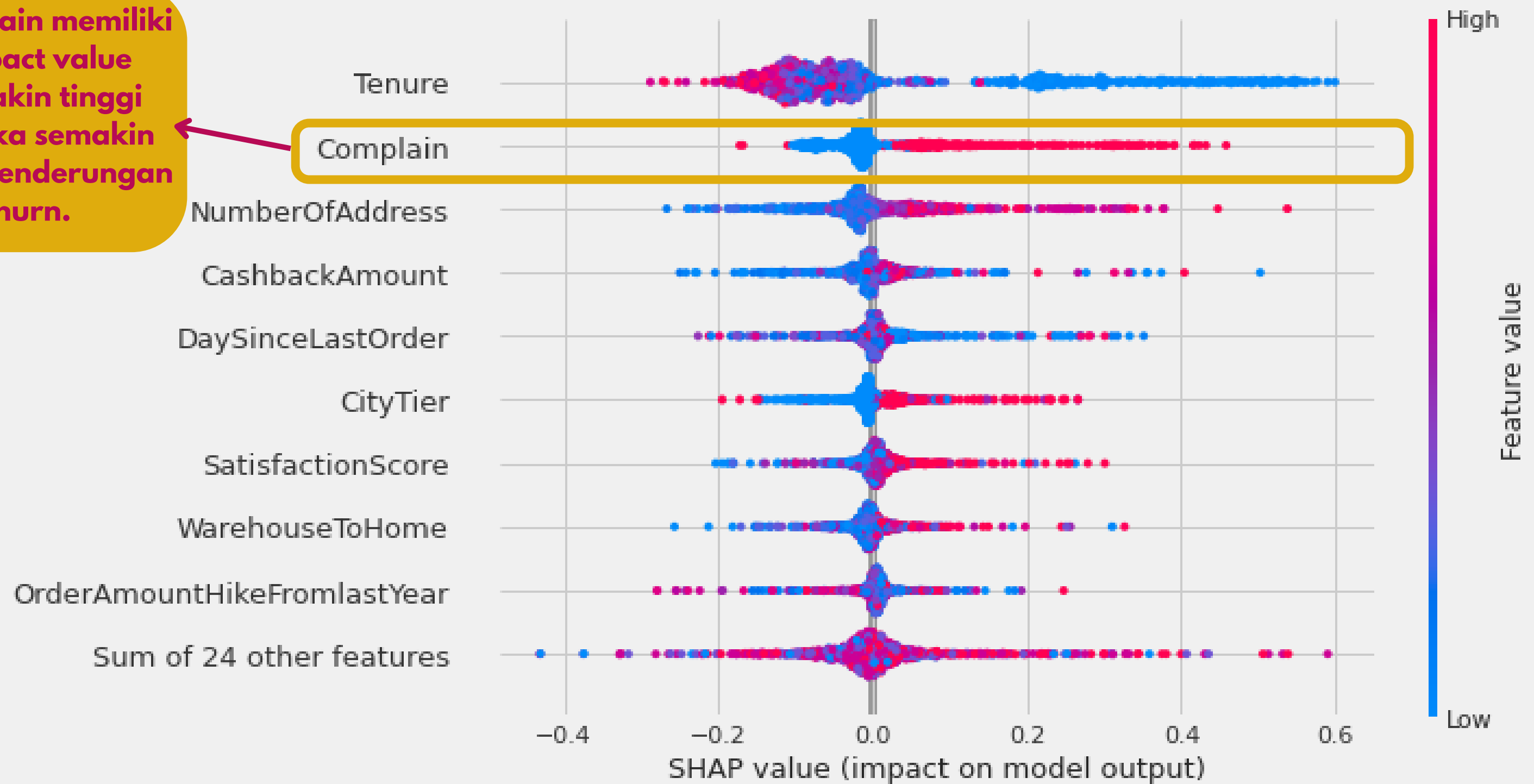
Mengutamakan customer satisfaction dengan cara membagi customer menjadi 2, yaitu : Priority Customer (customer complain dan berpotensi churn) dan Regular Customers.

**Priority Customer** mendapat penyelesaian atas complain nya lebih cepat sehingga customer tidak jadi berpotensi untuk churn.



# SHAP VALUE

Feature Complain memiliki positive impact value artinya semakin tinggi complain maka semakin tinggi juga kecenderungan untuk churn.



# BUSINESS RECOMMENDATION (PRIORITY)

67% of customer churn could be avoided if **resolved the customer's issue** during their first interaction.

Source : [https://www.getfeedback.com/resources/cx/40-stats-churn-customer-satisfaction/#::~text=67%25%20of%20customer%20churn%20could,\(Kolsky\)](https://www.getfeedback.com/resources/cx/40-stats-churn-customer-satisfaction/#::~text=67%25%20of%20customer%20churn%20could,(Kolsky))

82% of customer have stopped doing business with a ecommerce because of **bad customer service**.

Source : <https://www.zendesk.com/blog/why-companies-should-invest-in-the-customer-experience/>

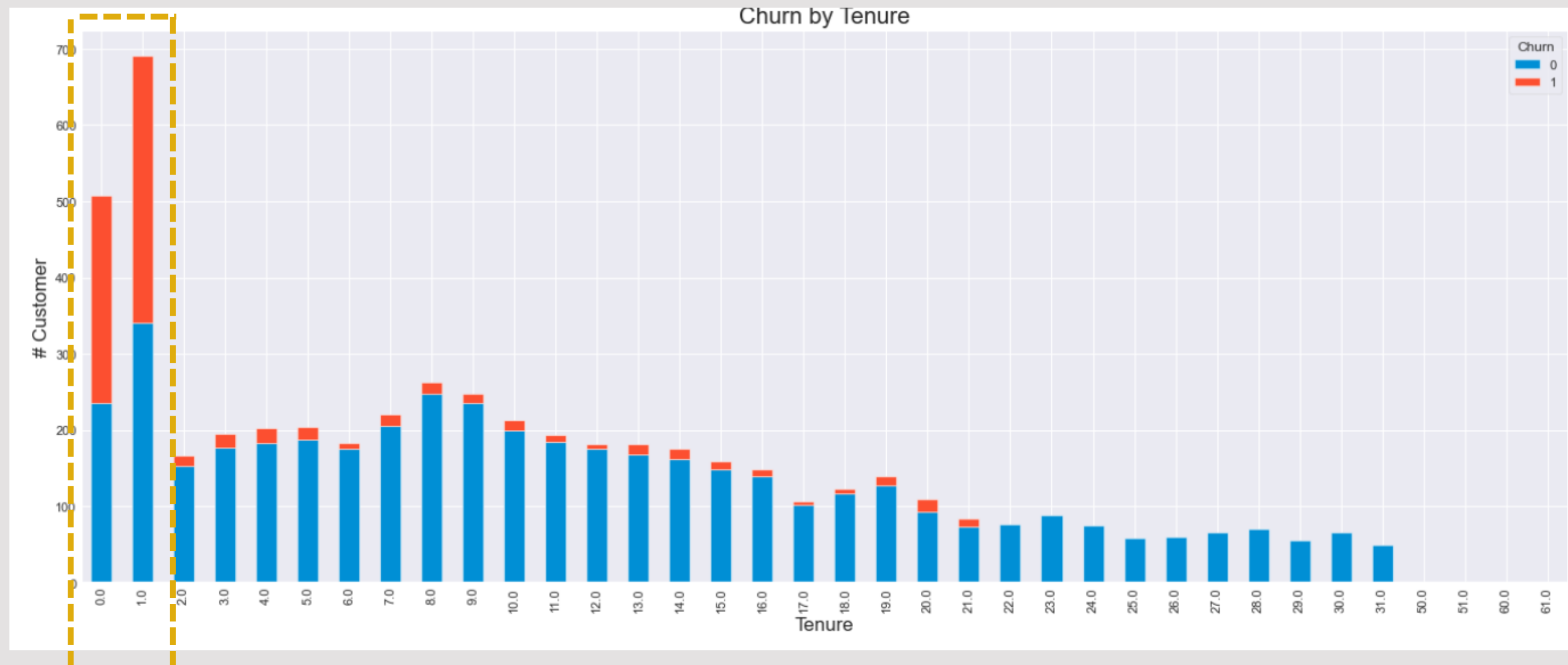
85% of customer churn because of **poor service** that could have been prevented.

Source : <https://www.slideshare.net/ekolsky/cx-for-executives>



## WHY COMPLAIN?

# BUSINESS RECOMMENDATION



## INSIGHT :

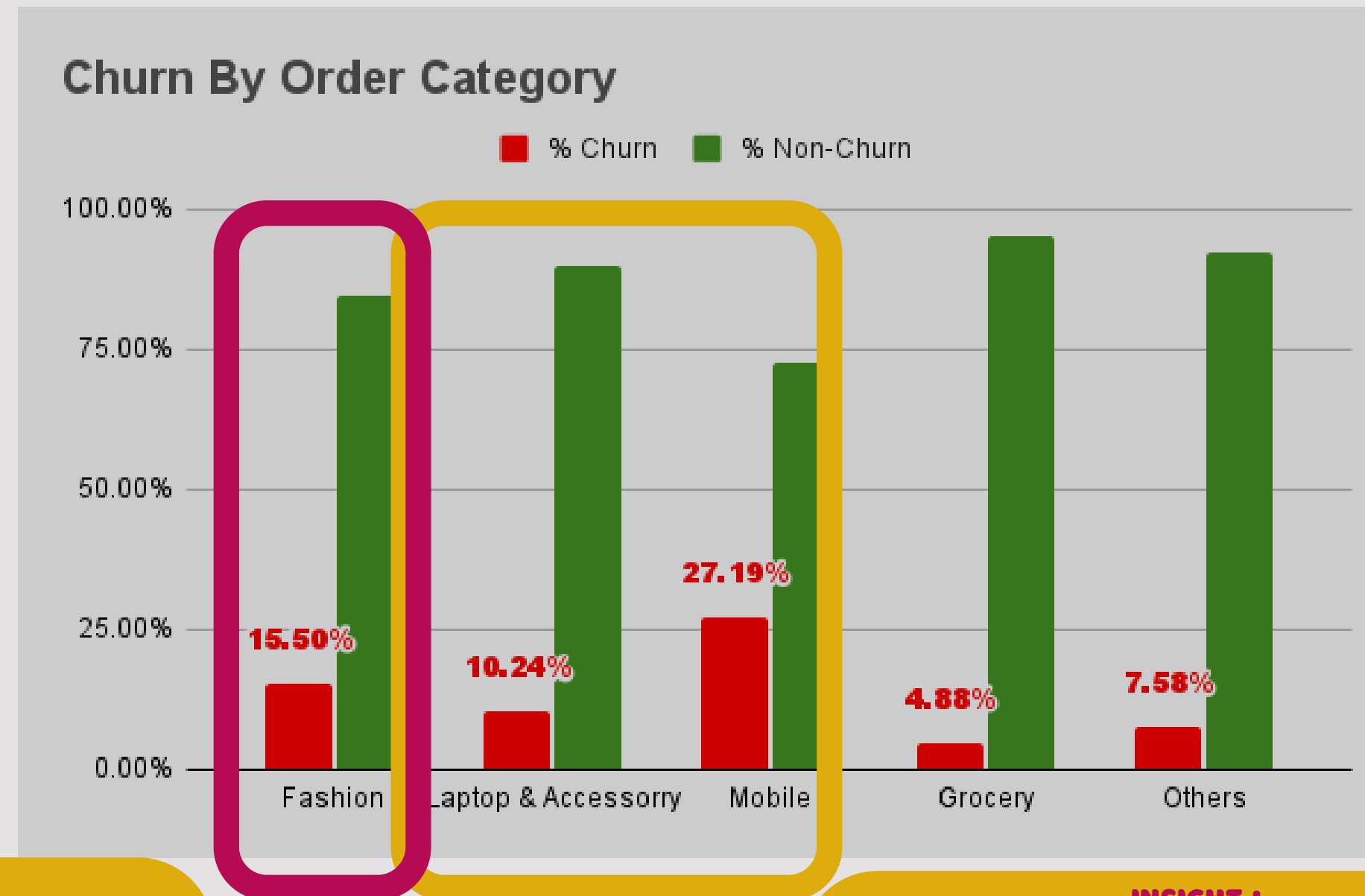
Customer dengan Tenure rendah (2 bulan pertama) memiliki potensi churn jauh lebih tinggi daripada tenure menengah keatas.

## BUSINESS RECOMMENDATION

Memberikan promo kepada customer dengan tenure < 2 bulan. Promo tersebut bisa berupa **Coupon** agar customer bisa terus bertahan dan kembali berbelanja di e-commerce. Selain itu, handle complain juga dapat membantu mendapatkan pelanggan royal



# BUSINESS RECOMMENDATION



## INSIGHT :

Fashion termasuk barang yang bisa dibeli setiap bulan seharusnya tingkat churn tidak tinggi

## BUSINESS RECOMMENDATION :

Memberikan penawaran produk yang berbeda dengan produk yang sudah dibeli oleh customer.

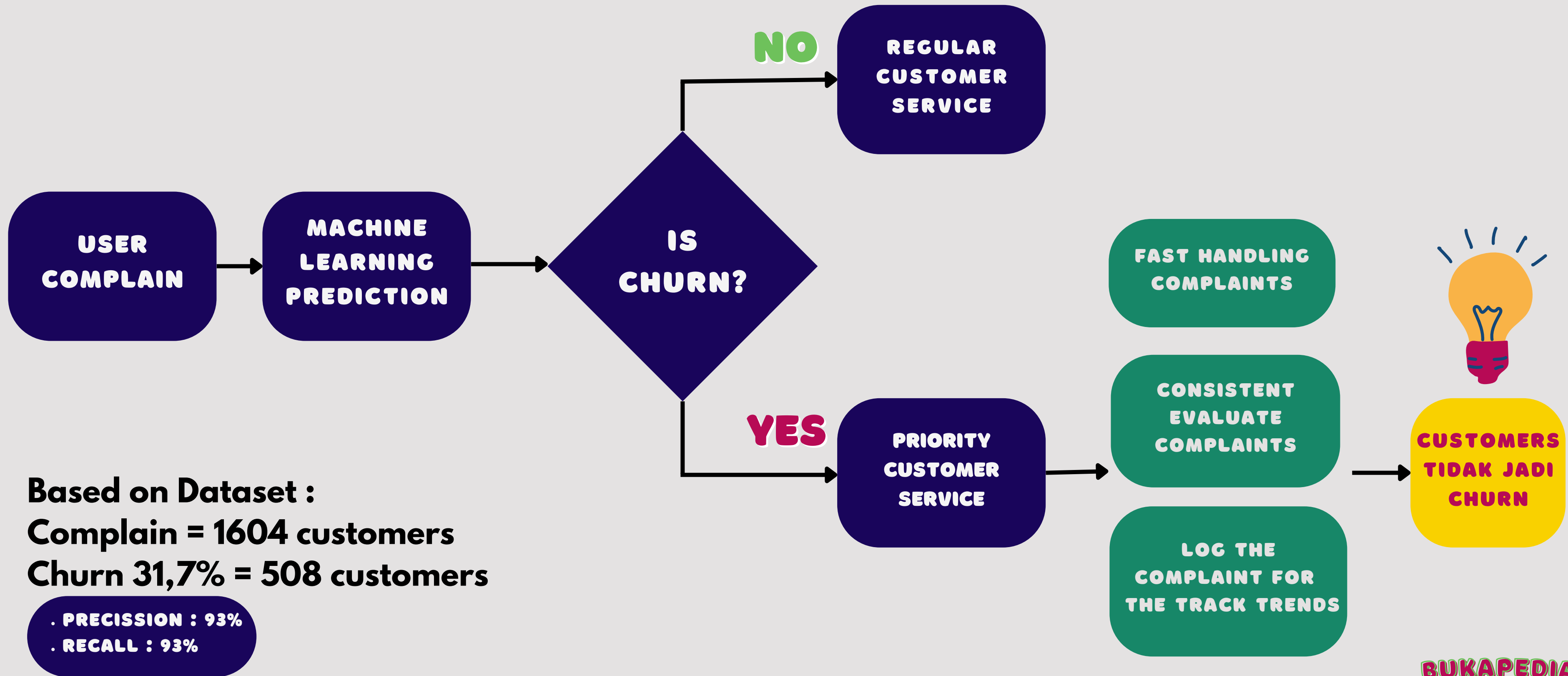
## INSIGHT :

Laptop & Accessorry dan Mobile bukan termasuk barang yang dibeli setiap bulan

## BUSINESS RECOMMENDATION :

Untuk customer yang baru membeli barang elektronik bisa dilakukan cross-selling dengan penawaran aksesoris elektronik, seperti keyboard, mouse, dll.

# BUSINESS IMPACT



# BUSINESS IMPACT

**SEBELUM  
MACHINE  
LEARNING  
PREDICTION**

**508 CUSTOMERS X \$ 228  
= \$ 115.824**

**LOST  
OPPORTUNITY**

**SETELAH MACHINE LEARNING PREDICTION**

**POTENTIAL SAVING  
OPPORTUNITY**

**= 20% X 508 CUSTOMERS X \$ 228  
= \$ 23.165 / YEAR**

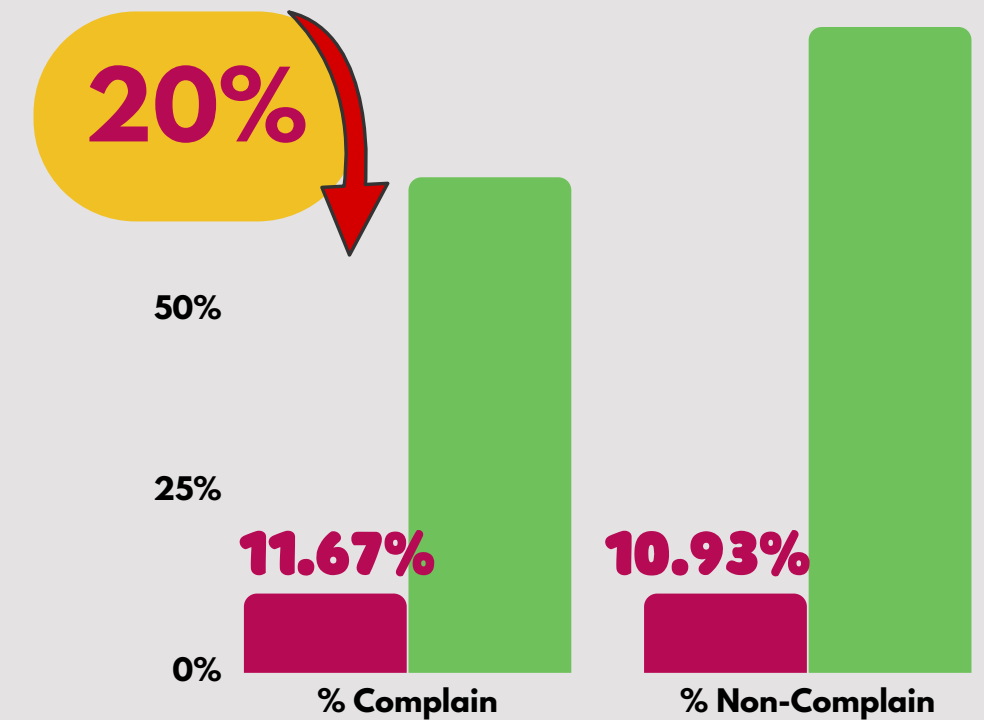
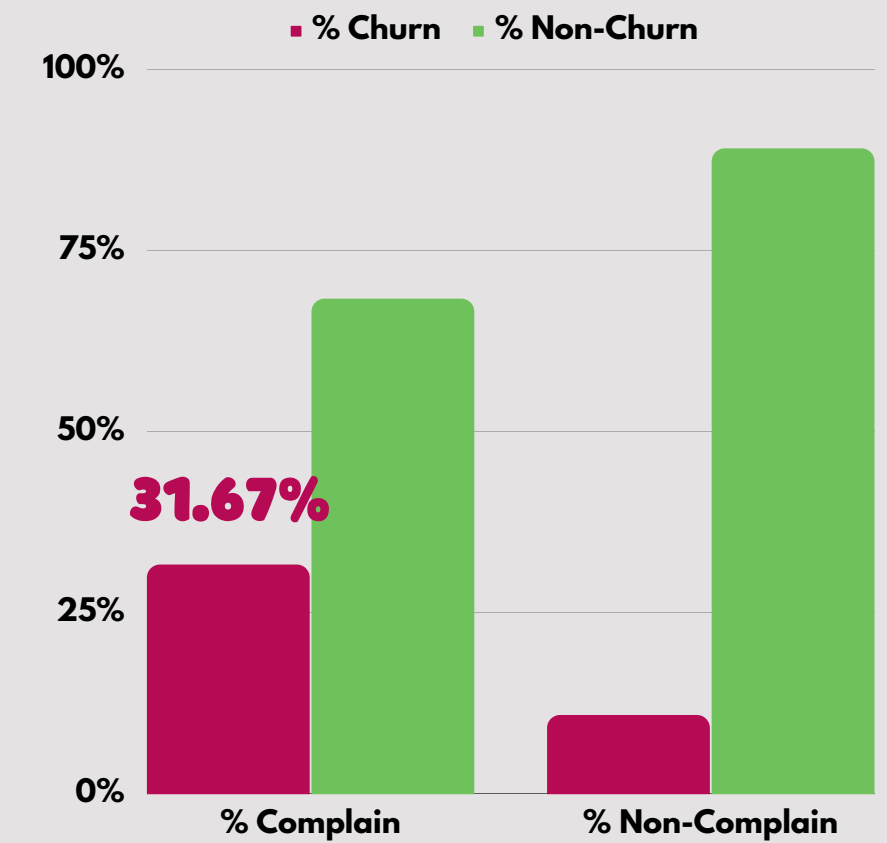
**406  
Customers**

**Fact:**

**Average Yearly Spending User = \$ 228 / Year\***

**\*Source :**

**<https://www.pcmag.com/news/how-much-do-your-devices-cost-you-per-year>**



**BUKAPEDIA**

**BUKAPEDIA**

**THANK YOU**

**“Without data, you are just another person with an  
opinion”**

**-W.Edwards Deming (Statistician. Professor, Author)**



# BUKAPEDIA

