E-COMMERCE CHURN ANALYSIS & PREDICTION

Hermes (Data Scientist)

Final Presentation Kelompok 5 - Batch 17

MEET OUR TEAM

Andre
Yudha Priyadi

Dwi
Susanto

Mhd **Fahmi**
Aziz

Nur Ayu
Asyifa

R.R **Intan**
Dwi Nuraini

OUTLINE

1. Business Understanding
2. Data Overview, EDA, & Insight
3. Data Cleaning & Preparation
4. Model & Evaluation
5. Business Insight & Recommendation

BUSINESS UNDERSTANDING

Who Are We?

Hermes adalah tim data scientist yang siap memberikan rekomendasi aksi untuk menyelesaikan masalah yang dihadapi perusahaan e-commerce

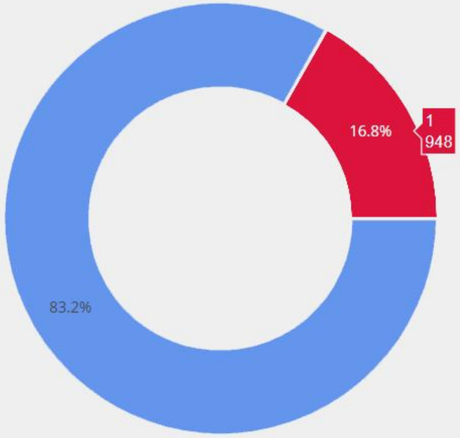
Who & What Case?

Amajoko adalah e-commerce apps yang berada di Indonesia dan sudah 5 tahun lebih berdiri.
Platform ini menjual produk fashion, handphone, laptop & aksesoris, groceries dan lain-lain.

Sayangnya, mereka memiliki masalah dengan banyaknya customer yang churn

Let’s Check The Data

Amajoko E-commerce Churn Rate



- ❏ Di tahun 2020, Amajoko memiliki total pelanggan sebanyak 5630.
- ❏ Namun, 16.8% dari total pelanggannya *churn*.

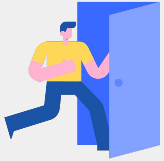
What is churn?

- **Churn customer** adalah kondisi dimana customer tidak lagi berlangganan ataupun berbelanja di suatu platform/e-commerce.
- Churn customer bisa disebabkan beberapa hal diantaranya:
  1. Avoidable Churn[1]**: Pindah** ke platform/e-commerce lain
  2. Unavoidable Churn[1]: Memang sudah **tidak belanja lagi** dimanapun (meninggal, tidak punya pendapatan, dll)
  3. Penyebab churn terbesar adalah karena **poor service**, sebesar 70%[2]

Reference: [1] www.sixteenventures.com/churn-classificatio [2] www.superoffice.com/blog/reduce-customer-churn

Problem Statement

Customer Churn



948

Customer

16.8%

Churn rate

Potential Loss

$4.63M

GMV* generated by churn customer last month

*asumsi:
Cashback = 10% dari GMV
GMV = Average cashback x #Order bulan lalu x 10

Jika tren churn rate masih **16.8% setiap bulan**, maka di bulan **ke-6** semua customer saat ini akan churn. Churn rate harus **dikurangi** agar bisnis bisa **sustain**.

Goals, Objective, Business Metrics

GOALS

1. **Menurunkan** tingkat **churn rate <5%**[1]
2. **Meminimalisasi potential loss** akibat churn

OBJECTIVE

1. Membuat **Machine Learning Model** untuk **memprediksi customer churn**
2. Memberikan **insight** dan **rekomendasi aksi** yang dapat membantu tim bisnis dan product untuk **mengurangi churn rate**

BUSINESS METRICS

Churn Rate

$$\frac{\#\ customer\ churn}{\#\ total\ customer}$$

Potential Loss

$$\#customer\ churn\ x\ potential\ GMV$$

Reference: [1]https://bloomintelligence.com/blog/is-your-retail-churn-rate-above-average/

DATA OVERVIEW, EDA & INSIGHT

Data Overview

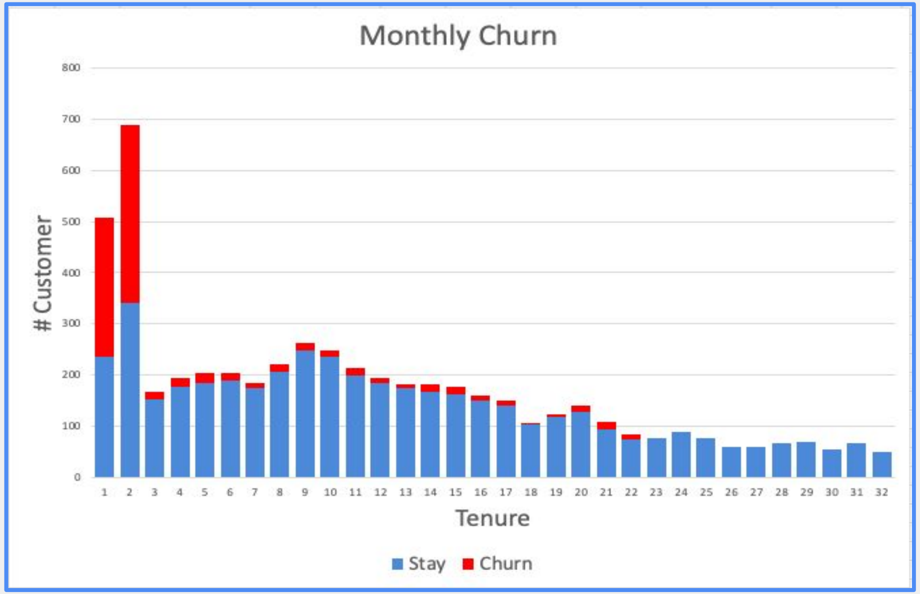
No(#)	Variable	Description
#1	CustomerID	Unique customer ID
#2	Churn	Churn Flag
#3	Tenure	Tenure of customer in organization
#4	PreferredLoginDevice	Preferred login device of customer
#5	CityTier	City tier
#6	WarehouseToHome	Distance in between warehouse to home of customer
#7	PreferredPaymentMode	Preferred payment method of customer
#8	Gender	Gender of customer
#9	HourSpendOnApp	Number of hours spend on mobile application or website
#10	NumberOfDeviceRegistered	Total number of deceives is registered on particular customer
#11	PreferedOrderCat	Preferred order category of customer in last month
#12	SatisfactionScore	Satisfactory score of customer on service
#13	MaritalStatus	Marital status of customer
#14	NumberOfAddress	Total number of added added on particular customer
#15	Complain	Any complaint has been raised in last month
#16	OrderAmountHikeFromlastYear	Percentage increases in order from last year
#17	CouponUsed	Total number of coupon has been used in last month
#18	OrderCount	Total number of orders has been places in last month
#19	DaySinceLastOrder	Day Since last order by customer
#20	CashbackAmount	Average cashback in last month

Load Data

- Target Label: Churn
- 18 feature

  Categorical feature:
  - #3 Tenure
  - #5 CityTier
  - #6 WarehouseToHome
  - #8 Gender
  - #9 HourSpendOnApp
  - #11 PreferedOrderCat
  - #13 MaritalStatus

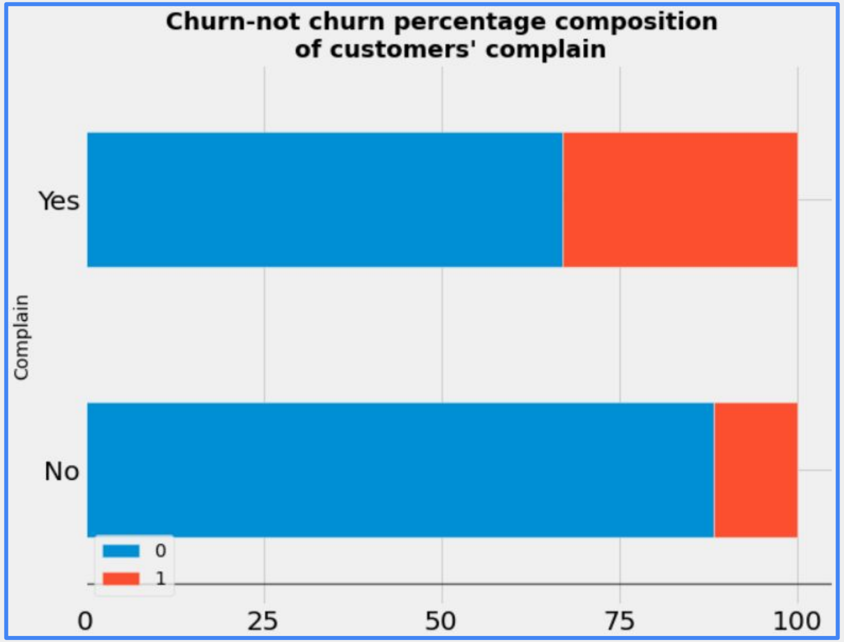
Exploratory Data Analysis (EDA)



Observation & Insight:

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

Exploratory Data Analysis (EDA)

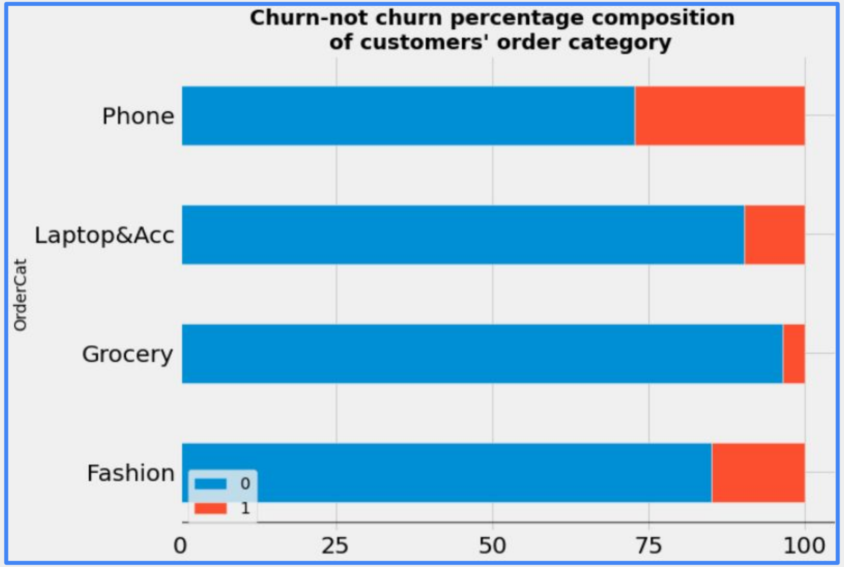


	Complain	
	Yes	No
Churn	32.93%	11.85%
Not Churn	67.07%	88.15%

Observation & Insight:

Persentase **churn akibat customer complain** (32.9%) **hampir 3 kali lebih besar** dari persentase churn dari total customer yang tidak complain (11.85%).

Exploratory Data Analysis (EDA)



	Order Category			
	Phone	Laptop&Acc	Grocery	Fashion
Churn	27.20%	9.70%	3.65%	15.02%
Not Churn	72.80%	90.30%	96.35%	84.98%

Observation & Insight:

Customer yang mengorder kategori **phone cenderung untuk churn** dibandingkan dengan customer yang mengorder kategori barang lainnya.

DATA CLEANING & PREPARATION

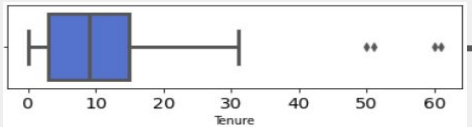
Data Cleaning

Why is it important?

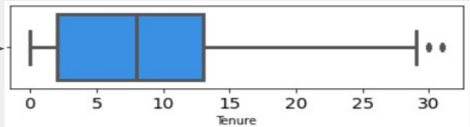
"Sparse quality data can not only harm the growth of an organization but can also signal many false **data insights**, leading to poor **decision-making**" [1]

How about our dataset?

- Handling Null Values & Imputing → 7 nums feature imputed with median
- Check & Remove Duplicates → 0 duplicate
- Handling Outliers (IQR method)



Before Handling Outliers



After Handling Outliers

Data before and after cleaning (rows): 5630 → 4040

Reference:
[1]www.simplilearn.com/data-cleaning-why-and-how-to-get-started-article

Data Preprocessing

Split Data
into train & test set

Train : Test = 70 : 30

Encoding
categorical feature

- Label Encoding
- One-Hot Encoding

Scaling & Transformation

- Log Transformation
- Standardization

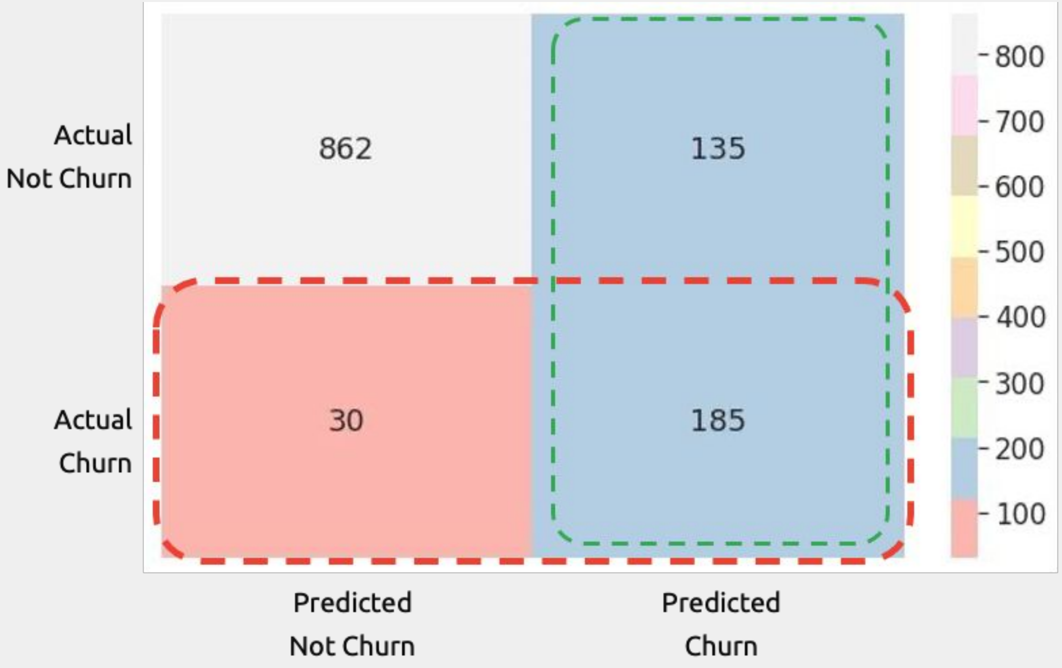
Balancing Data

Oversampling SMOTE
dengan rasio 0.33

Modelling

MODEL & EVALUATION

Confusion Matrix



Primary matrix: **Recall**

Secondary matrix: **Precision**

karena tidak memperbolehkan nilai **false negative** yang besar (customer yang sebenarnya churn namun dianggap tidak churn).

$$\text{Recall} = \frac{\text{True Positive}}{\text{True positive} + \text{False Negative}}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True positive} + \text{False Positive}}$$

Model After Hyperparameter Tuning

Model	Accuracy		Precision		Recall		F1-score		AUC	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
LOGISTIC REGRESSION	0.89	0.86	0.81	0.82	0.7	0.28	0.75	0.42	0.82	0.63
DECISION TREE	1	0.86	1	0.6	1	0.68	1	0.64	1	0.79
KNN	1	0.86	1	0.6	1	0.68	1	0.64	1	0.79
RANDOM FOREST	1	0.94	1	0.9	0.98	0.74	0.99	0.82	0.99	0.88
XGBOOST	0.95	0.86	0.94	0.58	0.86	0.86	0.9	0.69	0.92	0.86

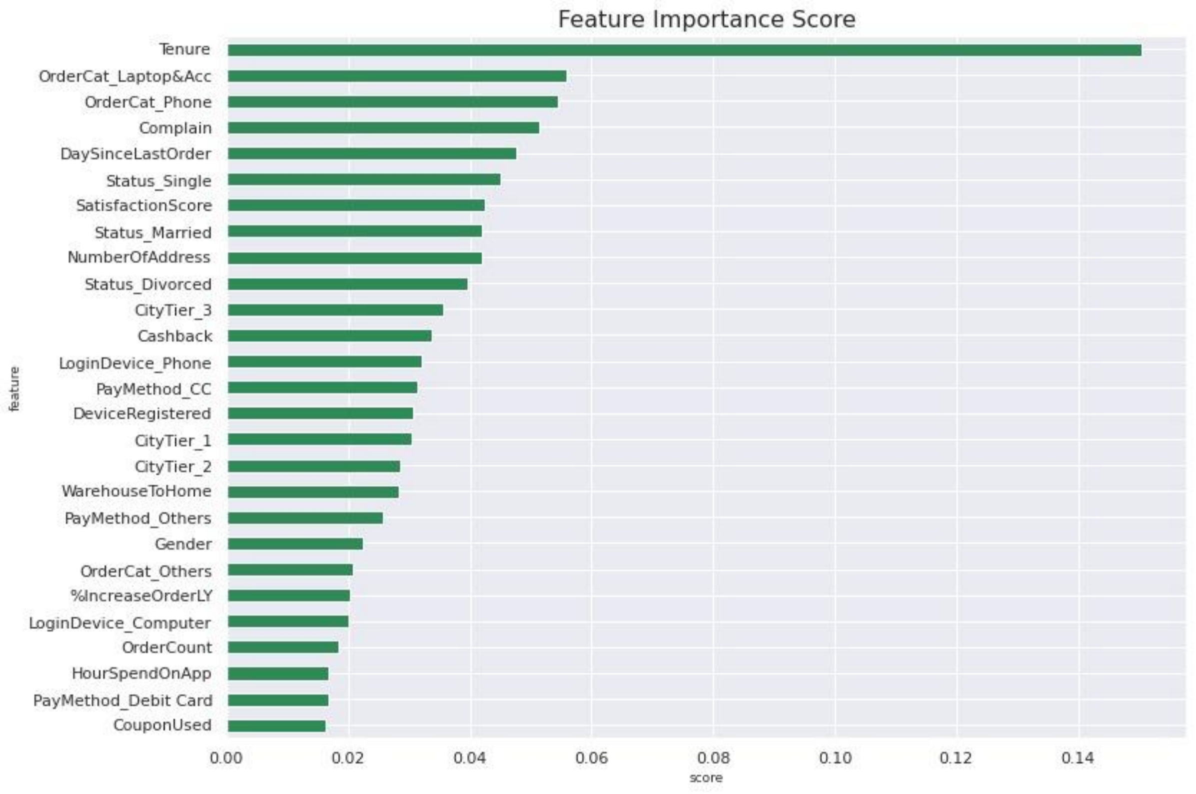
XGBOOST menjadi model yang dipilih karena:

- Model XGBoost adalah yang paling fit. Dapat dilihat dari train dan test **Recall-**nya memiliki nilai **delta terkecil.**
- Nilai **Precision test score di atas 0.5**

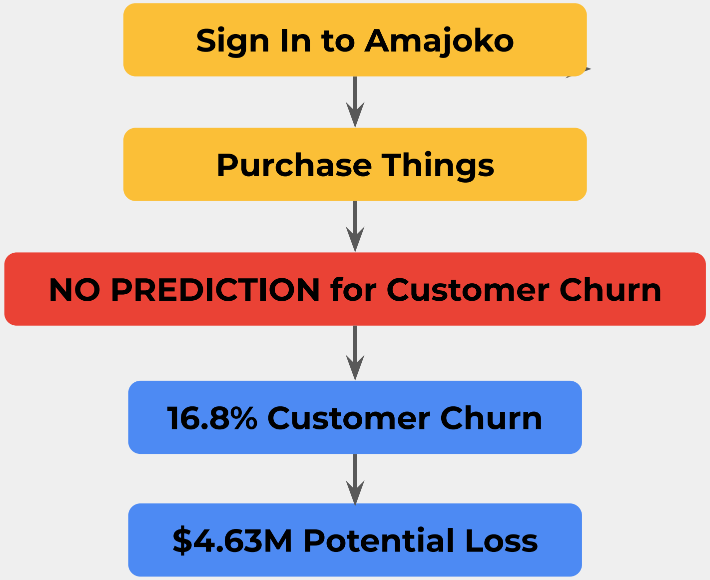
Untuk mendapatkan model XGBOOST yang best fit, beberapa trial yang dilakukan:

- Class imbalance
- Cross validation
- Hyperparameter tuning (max_depth, min_child_weight, gamma, eta, tree_method, colsample_bytree, num_round)

Feature Importance



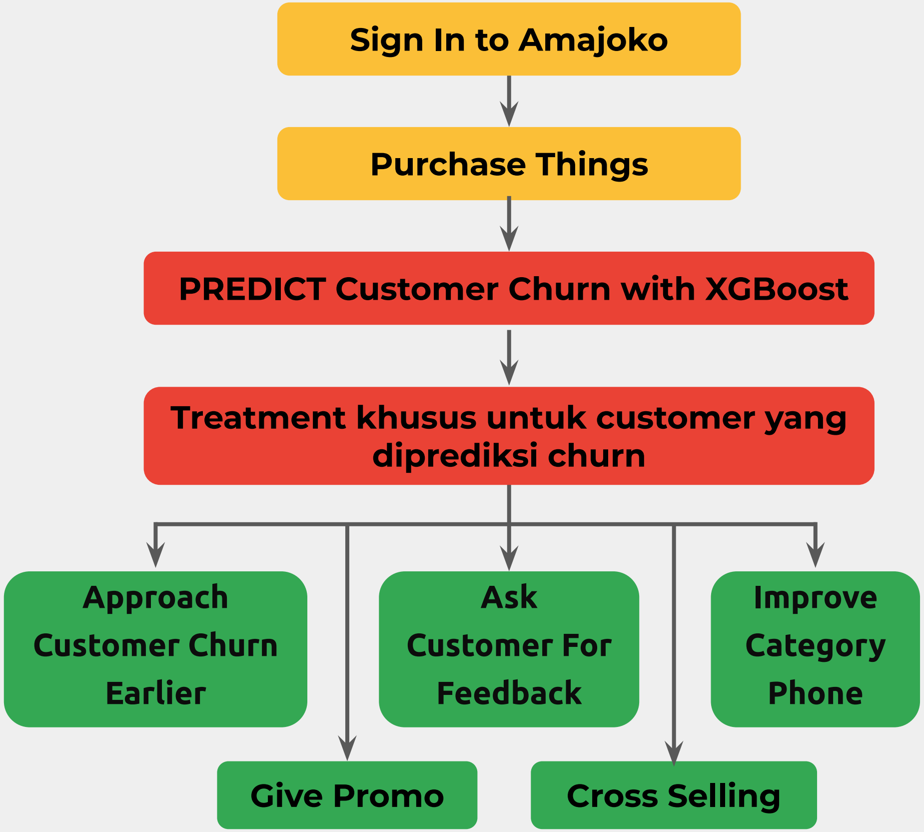
BEFORE Machine Learning Model



GMV* generated by churn customer last month

*asumsi:
Cashback = 10% dari GMV
GMV = Average cashback x #Order bulan lalu x 10

AFTER Machine Learning Model



BUSINESS INSIGHT & RECOMMENDATION

The Decrease of Potential Loss

Before using Model	After using Model (Recall: 0.86)
Total customer: 5630 (948 churn)	Remaining customer: 4682 (sisa customer yang tidak churn)
	Total customer based on prediction: Prediction x remaining customer = 86% x 4682 = 4026 (656 customer churn)
	Setelah **customer treatment** oleh tim bisnis, asumsi customer churn berkurang menjadi: 80% x 656 = 131 customer yang churn
(initial) Churn rate: **16,8%**	Jika diasumsikan 131 customer churn di bulan berikutnya, maka Churn rate: 131/4682 = **2,79%** (turun **14%**)
Potential loss akibat churn: $ 4.884 x 948 = **$ 4.630.000**	Predicted potential loss akibat churn: $ 4.884 x 131 = **$ 639.804** Potential loss yang terselamatkan: $ 4.884 x 525 = **$ 2.564.100**

Recommendation for Product Team

Improve order category phone karena banyak yang complain dan kemudian churn.

Meminta feedback user, seperti survey singkat mengenai apa yang perlu diubah untuk mendapatkan pengalaman yang lebih baik saat melakukan transaksi.

Reference:
www.mattsenkumar.com/e-commerce-customer-churn-best-practice

Recommendation for Business Team

Untuk **2 bulan pertama**, dimana kondisi customer dengan tenure rendah cenderung untuk churn, perlu **diberikan treatment khusus** seperti **memberikan promo** berupa free delivery dan **cashback** yang cukup tinggi.

Penyumbang churn terbesar adalah dari category product phone (60%), perlu adanya **cross selling untuk pembeli product phone** agar bisa menarik lagi customer yang sudah churn.

Melakukan **approach lebih awal** ke customer yang diprediksi akan churn untuk mendapatkan pain point mereka.

Reference: www.mattsenkumar.com/e-commerce-customer-churn-best-practice

DONE, THANK YOU

“Data telling the truth,
people tell stories and hopes”