

# Final Presentation E- Commerce Churn Prediction

MANDALIKA - GROUP 7



# Meet Our Team

**Lise Listianti**



**Ardilla Safitri**



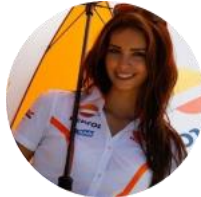
**Jonse Kennedy  
Oppusunggu**



**Fakhry Husnul**



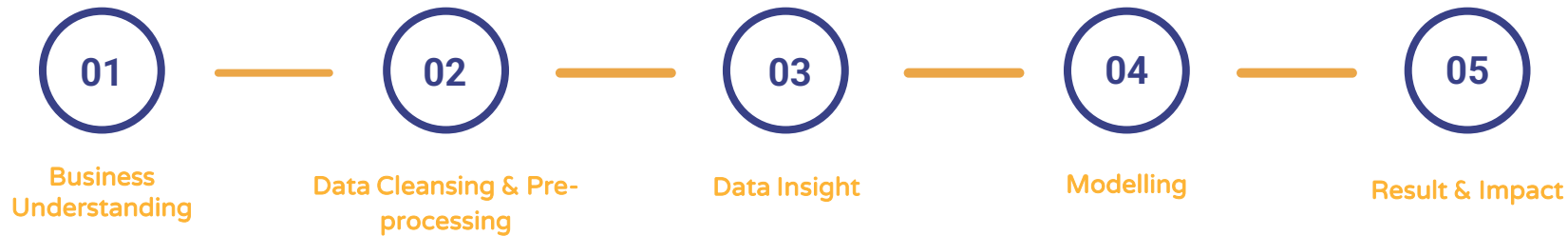
**Riska Diah N**



**Arni Cici Suryani**



# Outline





01

# Business Understanding

# About us

Perusahaan **Mandalika** bergerak di bidang e-Commerce yang menjual berbagai produk seperti *mobile phone*, laptop dan aksesoris, *fashion*, *groceries* yang sejauh ini memiliki total customers sebanyak 5.630.

Kami adalah **Tim Data Scientist** di perusahaan Mandalika.





# Background

## Apa itu *Churn Rate*?

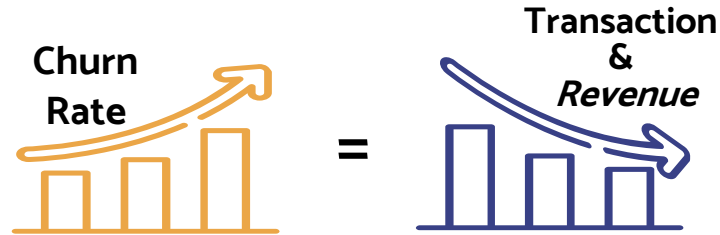
Menurut Corporate Finance Institute, **Churn Rate** dapat menunjukkan tingkat **persentase** di mana perusahaan **kehilangan pelanggan** atau **pendapatannya** karena berbagai keadaan.

## Customer Retention vs. Customer Acquisition

sumber: churnbuster.io

- Rata-rata, pelanggan tetap akan menghasilkan revenue **67% lebih banyak** dibanding pelanggan baru.
- Membawa pelanggan baru akan **25 kali lebih mahal** dibanding mempertahankan pelanggan lama.
- Meningkatkan tingkat retensi sebanyak 5% dapat meningkatkan keuntungan sebesar **15% hingga 95%**.

## Bagaimana dampak *Churn Rate*?



Dalam satu bulan terakhir ***churn rate*** perusahaan mencapai **16,83%** > 5%\*  
(standar Churn Rate perusahaan e-Commerce)

\*sumber: Open StartUp Survey, 2015

Pada Mandalika, dampak yang dihasilkan akibat Churn Rate **16,83%** adalah adanya potensi kehilangan *Revenue* sebesar **\$2.400.106.**

\*asumsi rata-rata transaksi memberikan cashback 7%



## GOALS

- **Menurunkan** Churn Rate dengan membuat Model Customer Churn Prediction
- **Menyelamatkan** potensi kehilangan *revenue*.  
\* asumsi rata-rata transaksi memberikan cashback 7%



## OBJECTIVE

Membuat **model** untuk mendeteksi pelanggan yang berpotensi berhenti berlangganan (churn) dengan target churn rate <5%.



## BUSINESS METRICS

- Main: ***Churn Rate***
- Supporting: ***Revenue***





02

## Data Cleansing & Pre-processing

# Data Overview



Data set terdiri dari data customer selama satu bulan terakhir.

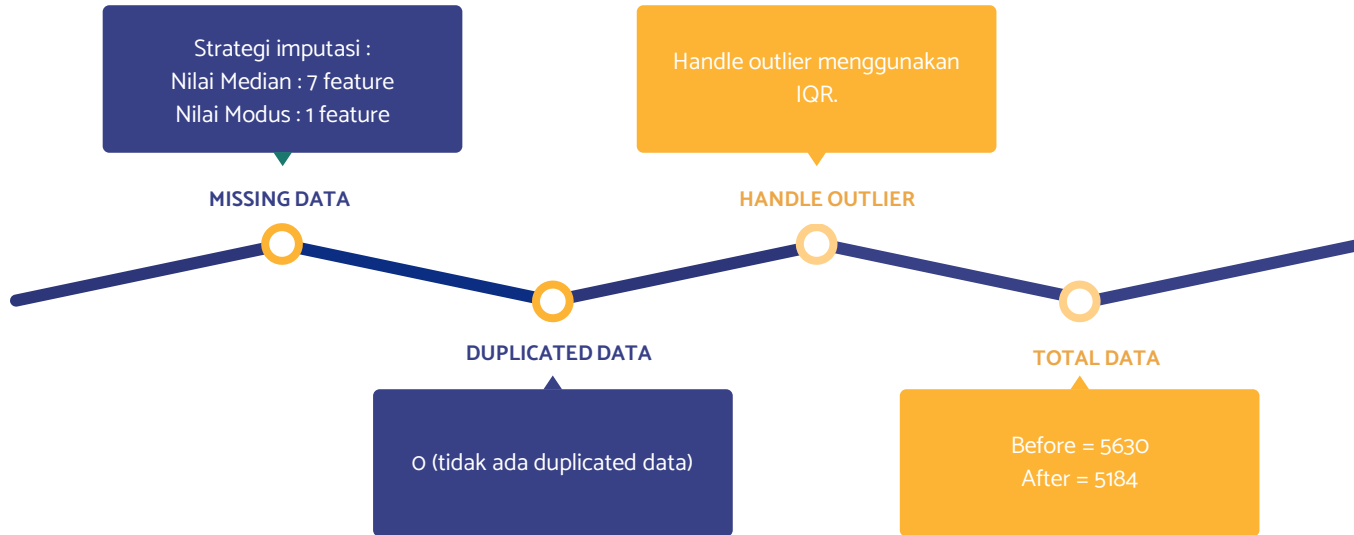


Data terdiri dari 19 features dan 5.630 baris.



Target label : Churn

# Data Cleansing



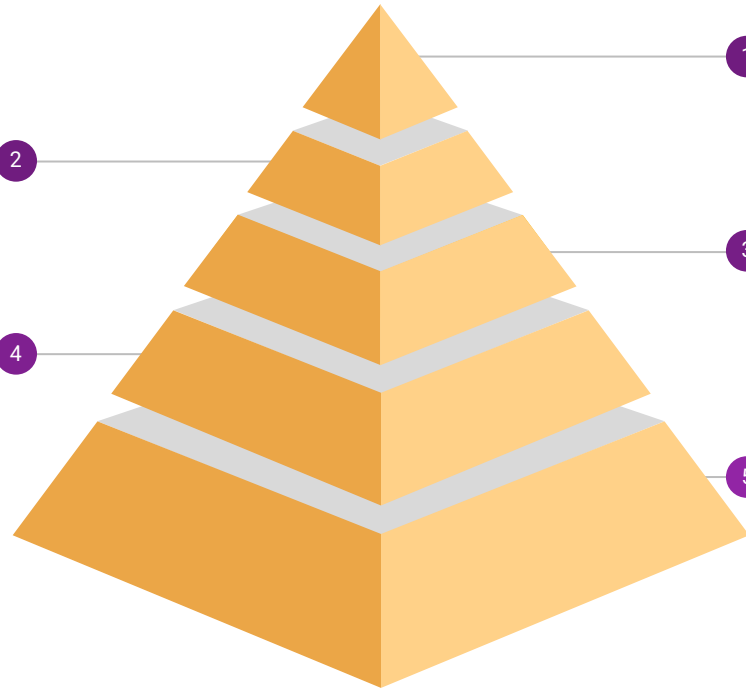
# Data Pre-processing

## FEATURE ENCODING

Rename jenis kategori yang sama.  
label encoding.  
one hot encoding.

## DUPLICATED DATA & SPLIT DATA

520 duplicated data.  
Data Train : Data Test = 80 : 20



## FEATURE TRANSFORMATION

Log Transformation pada 7 feature dan  
Standardization pada 4 feature.

## FEATURE SELECTION & EXTRACTION

5 feature tidak relevan di drop. Tersisa 26  
feature.  
Menambahkan kolom baru yaitu  
'OtherRevenue' \*

## CLASS IMBALANCE

Class imbalance menggunakan SMOTE.

\* 'OtherRevenue' = 'CashbackAmount' x cashback, asumsi cashback 7%.



03

# Data Insight

# Complain and Satisfaction Score

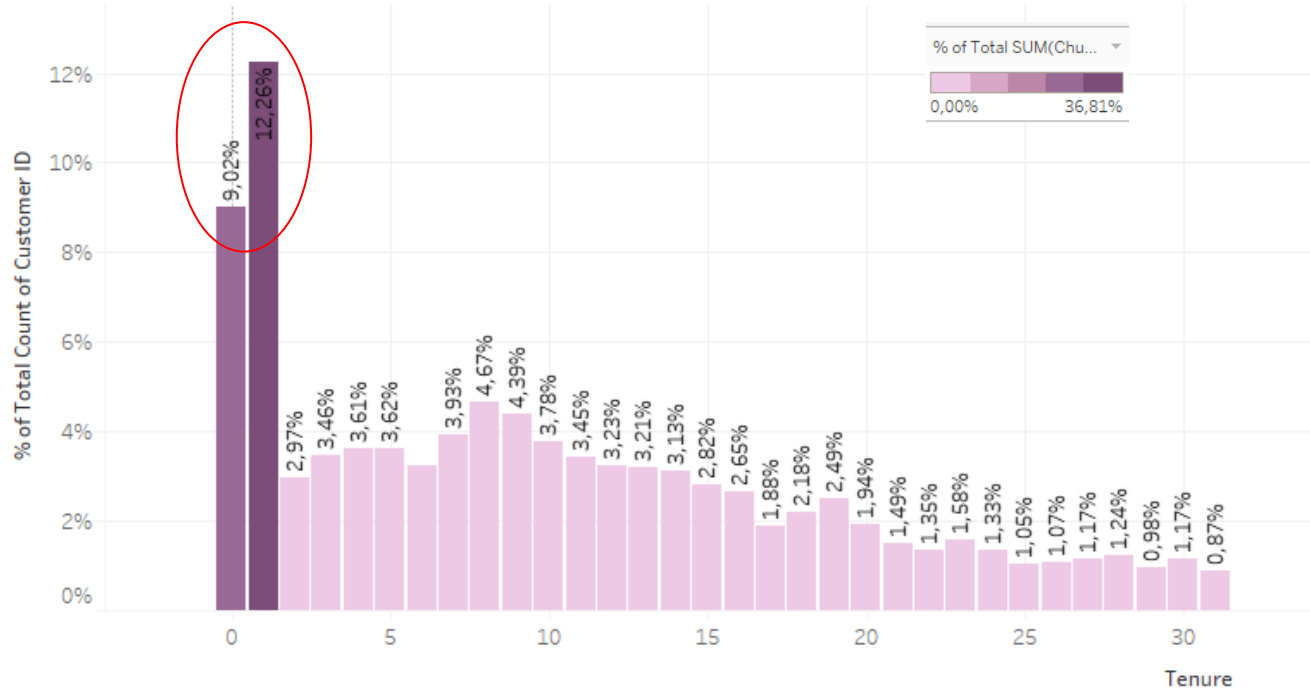


Complain rate 28%



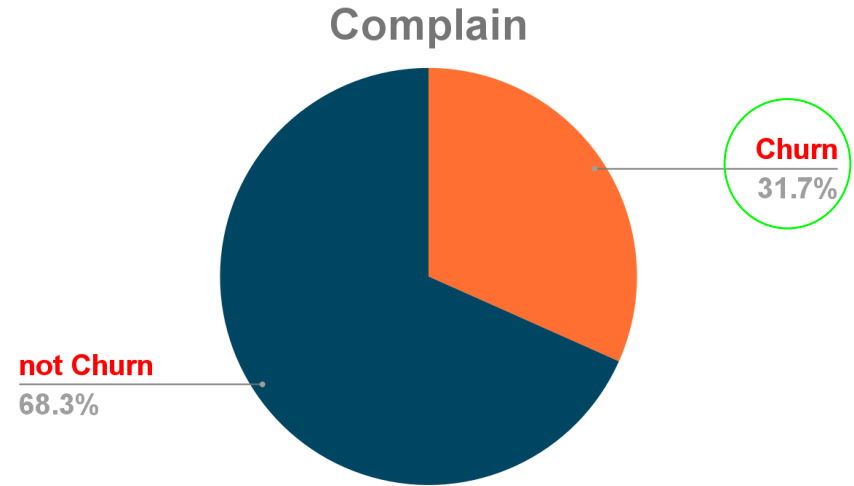
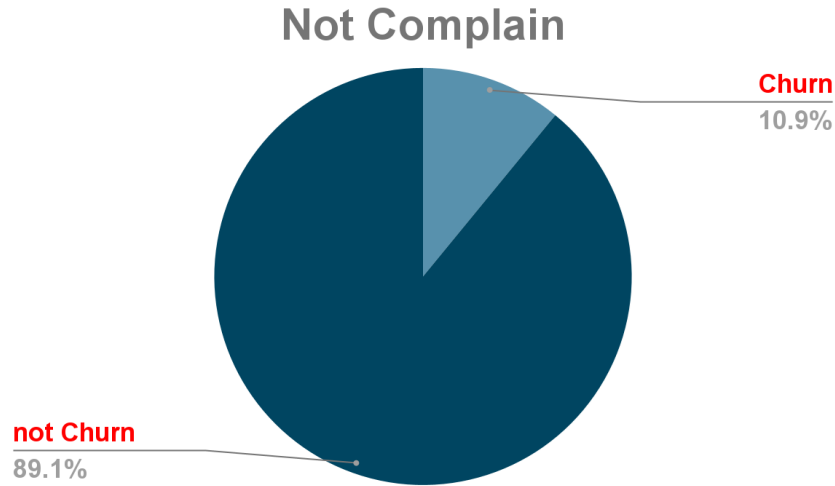
Satisfaction Score 3 out of 5

# Customer Churn Based On Tenure



Sebanyak 21,28% Customer dengan **Tenure Rendah** (kurang dari 1 minggu) memiliki potensi churn yang paling tinggi.

# Customer Churn Based on Complain

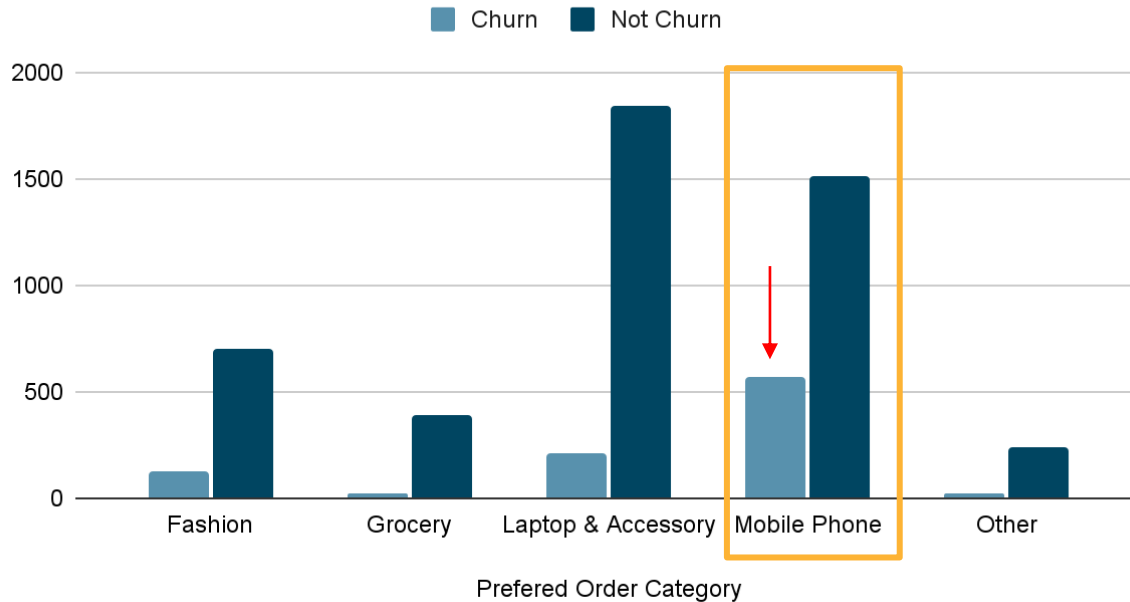


Pelanggan yang complain lebih **cenderung** untuk churn dari pada pelanggan yang tidak complain.



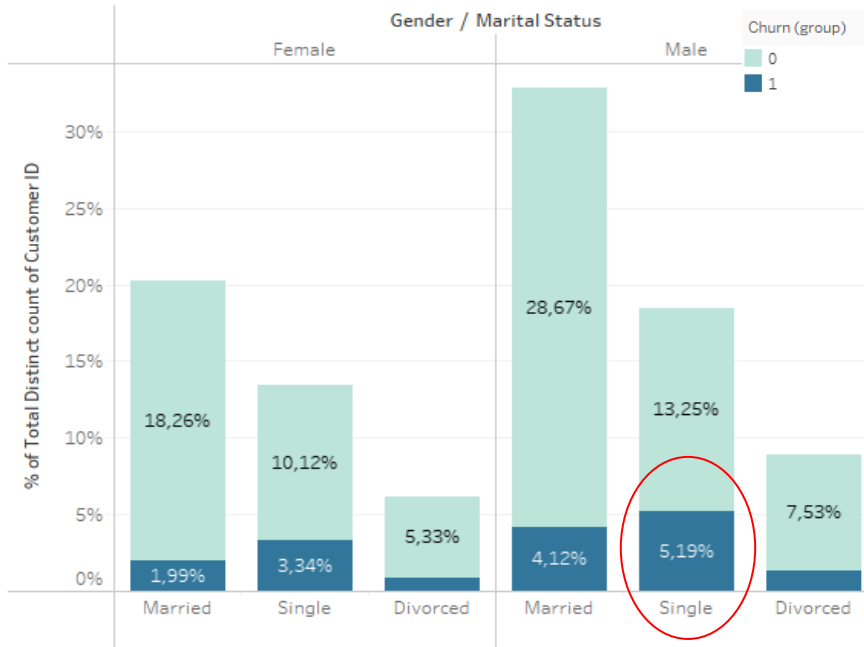
# Customer Churn Based on Order Category

Distribution of Preferred Order Category for Churn Customer



Sebanyak 570 Customer dengan Order Category **Mobile Phone** memiliki jumlah Churn paling tinggi.

# Customer Churn Based on Gender & Marital Status



**Laki-laki Single** memiliki jumlah Churn paling tinggi yaitu 5,19%.

# Why Customer Churn??



High Complain Rate

Low Score of Satisfaction

Tenure <1 week

Problem of Order Category

Male and Single Status



04

# Modelling

# Modelling

Metric yang digunakan ialah **RECALL** (primary) dan **PRECISION** (secondary).

MODEL	ACCURACY		PRECISION		RECALL		F1		AUC-ROC	
	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST
LOGISTIC REGRESSION	0.87	0.86	0.88	0.58	0.86	0.70	0.87	0.64	0.95	0.88
kNN	0.95	0.88	<b>0.92</b>	<b>0.61</b>	<b>1.00</b>	<b>0.90</b>	0.96	0.73	1.00	0.95
DECISION TREE	1.00	0.92	1.00	0.76	1.00	0.82	1.00	0.79	1.00	0.88
RANDOM FOREST	1.00	0.94	1.00	0.88	1.00	0.79	1.00	0.83	1.00	0.98
ADABOOST	0.90	0.86	0.90	0.58	0.90	0.70	0.90	0.63	0.96	0.89
XGBOOST	1.00	0.95	<b>1.00</b>	<b>0.90</b>	<b>1.00</b>	<b>0.84</b>	1.00	0.87	1.00	0.99

- Model **kNN** dan **XGBOOST** memiliki Recall tertinggi.
- Berdasarkan nilai Recall dan Precision tertinggi maka modelling yang optimal ialah **XGBOOST**.
- Model terpilih masih belum best-fit (overfit) sehingga perlu dilakukan **Hyperparameter Tuning**.

# Confusion Matrix

Metric yang digunakan ialah **RECALL** (primary) dan **PRECISION** (secondary).

LOGISTIC REGRESSION		Predictive	
		No Churn	Churn
Actual	No Churn	688	81
	Churn	50	114

kNN		Predictive	
		No Churn	Churn
Actual	No Churn	674	95
	Churn	16	148

DECISION TREE		Predictive	
		No Churn	Churn
Actual	No Churn	727	42
	Churn	30	134

RANDOM FOREST		Predictive	
		No Churn	Churn
Actual	No Churn	751	18
	Churn	35	129

ADABOOST		Predictive	
		No Churn	Churn
Actual	No Churn	687	82
	Churn	50	114

XGBOOST		Predictive	
		No Churn	Churn
Actual	No Churn	754	15
	Churn	27	137

- Model **kNN** dan **XGBOOST** memiliki False Negatif terendah.
- Model **kNN** memiliki False Positif paling tinggi → potensi penambahan cost.
- **XGBOOST** model yang paling optimal dilihat dari confusion matrix.

# XGBoost (Hyperparameter Tuning)

## BEFORE

METRICS	TRAIN	TEST
Accuracy	1.00	0.95
Precision	1.00	0.90
Recall	1.00	<b>0.84</b>
F1	1.00	0.87
AUC-ROC	1.00	0.99

Confusion Matrix		Predictive	
		No Churn	Churn
Actual	No Churn	754	15
	Churn	<b>27</b>	137

## AFTER

METRICS	TRAIN	TEST
Accuracy	1.00	0.96
Precision	1.00	0.91
Recall	1.00	<b>0.88</b>
F1	1.00	0.89
AUC-ROC	1.00	0.99

Confusion Matrix		Predictive	
		No Churn	Churn
Actual	No Churn	754	15
	Churn	<b>20</b>	144

### ❑ RECALL

Mengalami peningkatan sebesar **0.04%**.

### ❑ PRECISION

Mengalami peningkatan sebesar **0.01%**.

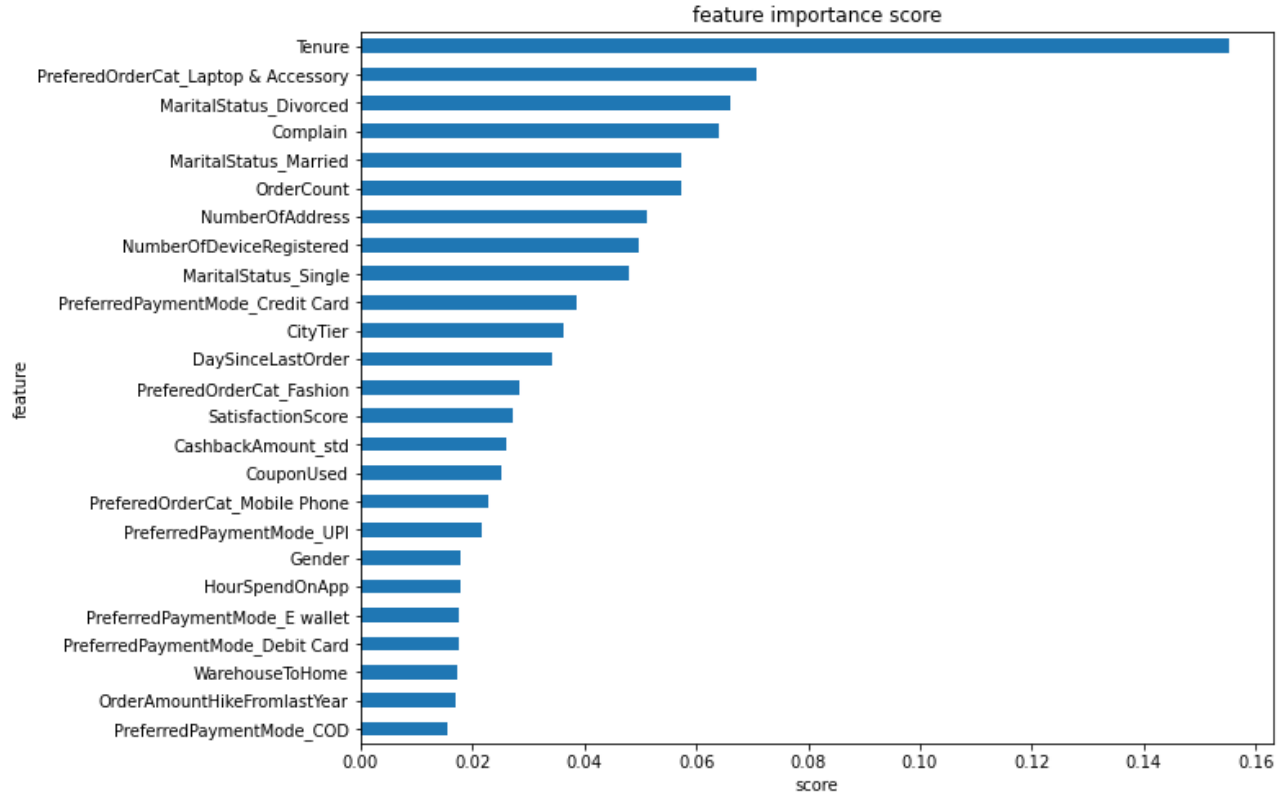
### ❑ FALSE NEGATIF

Mengalami penurunan sebesar **7**.

### ❑ Hyperparameter Tuning

'Max\_depth', 'lambda', 'alpha'.

# Feature Importance







05

# Result and Impact

## BEFORE MODELLING

MANDALIKA'S CUSTOMER



NO PREDICTION  
FOR CHURN CUSTOMER



**16.83%**  
Churn Rate



**\$2.400.106**  
Potential Losing Other Revenue

**\*\* Assumpstion**

Cashback = 7% from other revenue

Other revenue = Avg cashback/7%

Avg revenue = \$2.531/customer

## AFTER MODELLING

MANDALIKA'S CUSTOMER



PREDICT CUSTOMER CHURN  
WITH XGBOOST MODELLING



TREATMENT CUSTOMER  
CHURN BY PREDICT MODEL



GIVE LOYALTY REWARD  
(potongan ongkir, cashback, diskon, promo)



Improve our  
product



Ask Customer  
For Feedback



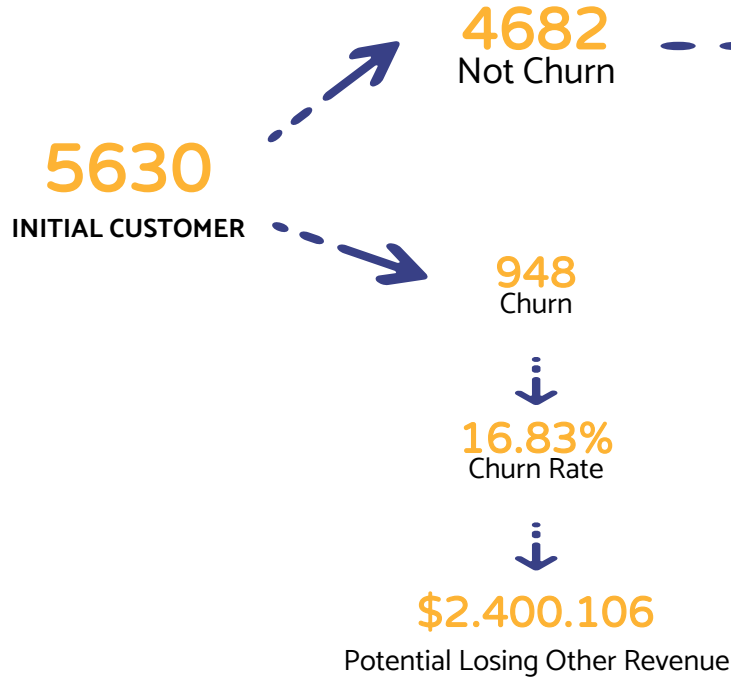
Improve our  
service



Approach Customer  
Churn Earlier

## BEFORE MODELLING

(LAST MONTH)



\*\*Asumption

Cashback = 7% from other revenue

Other revenue = Avg cashback/7%

Avg revenue = \$2.531/customer

## AFTER MODELLING

(NEXT MONTH)

PREDICT CUSTOMER CHURN  
WITH XGBOOST MODELLING  
Recall 88%

4120  
Not Churn

562  
Churn

TREATMENT CUSTOMER  
CHURN BY PREDICT MODEL

GIVE LOYALTY REWARD  
(potongan ongkir, cashback, diskon, promo)

TARGET TREATMENT SUCCESS : 80%

450  
Not Churn

112  
Churn

2.39%  
Churn Rate

# PREDICT CUSTOMER CHURN WITH XGBOOST MODELLING

Recall 88%

4120  
Not Churn

562  
Churn

TREATMENT CUSTOMER  
CHURN BY PREDICT MODEL

GIVE LOYALTY REWARD  
(potongan ongkir, cashback, diskon, promo)

COST LOYALTY  
REWARD = 7% revenue  
(cost →  $562 \times \$2531 \times 7\% = \$99.569$ )

TARGET  
TREATMENT  
SUCCESS : 80%

\*\*Assumpstion  
Cashback = 7% from other revenue  
Other revenue = Avg cashback/7%  
Avg revenue = \$2.531/customer



Impacts yang akan didapatkan ialah:

1. Churn Rate **2.39%**
2. Revenue terselamatkan sebesar **\$1.039.381**  
atau **8,77%** dari total revenue

112  
Churn

2.39%  
Churn Rate

450  
Not Churn

**\$1.138.950**  
Potential Saving  
Revenue  
(Total customer churn x  
avg revenue)

=  $450 \times \$2.531$

**\$1.039.381**

**Saving Revenue**  
(Potential Saving  
Other Revenue - Cost Loyalty)  
=  $\$1.138.950 - \$99.569$

# Business Insight & Recommendation



*Business insight* → Pengguna dengan tenure < 1 minggu beresiko untuk churn.

*Business recommendation:*

1. Memberikan **welcoming gift** untuk pengguna baru (gratis ongkir, cashback, promo).
2. Penawaran **membership** dengan keuntungan yang lebih menarik.
3. Memberikan **rekomendasi** produk sesuai dengan kategori gender, umur, marital status.
4. Pengembangan aplikasi agar lebih **user friendly**.

*Business insight* → Complain rate tergolong besar (28%)

*Business Recommendation :*

1. Meningkatkan **kualitas produk** yang dijual.
2. Meningkatkan **standar layanan** yang diberikan.
3. Membuat **layanan pengaduan** yang cepat, tanggap, dan solutif.
4. **Proactive** dalam menanggapi *Feedback*.



# Business Insight & Recommendation



*Business insight* → Order untuk kategori Mobile Phone memiliki resiko Churn paling tinggi.

*Business recommendation:*

1. Menyediakan **fitur verifikasi** barang original.
2. Menyediakan **jasa asuransi** produk.
3. Meningkatkan **kualitas produk** yang dijual.
4. Menawarkan **harga yang bersaing**.

*Business insight* → Nilai rata-rata satisfaction score tergolong biasa saja (3,06 out of 5).

*Business recommendation:*

1. Perbaiki **user interface**.
2. Menindak lanjuti **ulasan negatif** dan meminta **feedback** agar evaluasi bisa tepat sasaran.
3. Menyediakan **kolom kritik dan saran**.
4. Membuat **survey** mengenai kepuasan customer.





# Thanks!

Do you have any questions?

[business@mandalika.com](mailto:business@mandalika.com)

+62 2145893

[mandalika.com](http://mandalika.com)



# Learning Curves

