Grow-Ject
E-commerce Shipping Data

Team Members

Fakhri Nurrahmadi
(Leader)

Muhammad Iqbal
Tawakkal

Ryan Rizky F

Dyah Phitaloka

Dwi M Nurafli

Content

Our Job Role

Data Scientist
E-Commerce Grow-Ject

Masalah Bisnis

↓

Solusi Bisnis

↓

Peningkatan Performa Bisnis

Problem Statement

Research say...

- **17% of respondents will stop shopping** after receiving **a late delivery one time.**

- **55% of respondents will stop shopping** after receiving a **late delivery two to three times.**

Source : Hollingsworth

You're **5 times** more likely to sell to an **existing customer** than a new one.

Goals ➡️ Meningkatkan *customer retention rate*

Objectives :

- Model Machine Learning ➡️ Predict Late or Not

- From Data ➡️ Why its late?

- Late Delivery ➡️ Give Treatment

Business Metrics : *Customer Retention Rate*

Business Metrics :

$$\text{Customer Retention Rate} = \frac{\text{Jumlah customer di akhir periode} - \text{Jumlah customer baru di periode berjalan}}{\text{Jumlah customer di awal periode}} \times 100\%$$

Data Overview
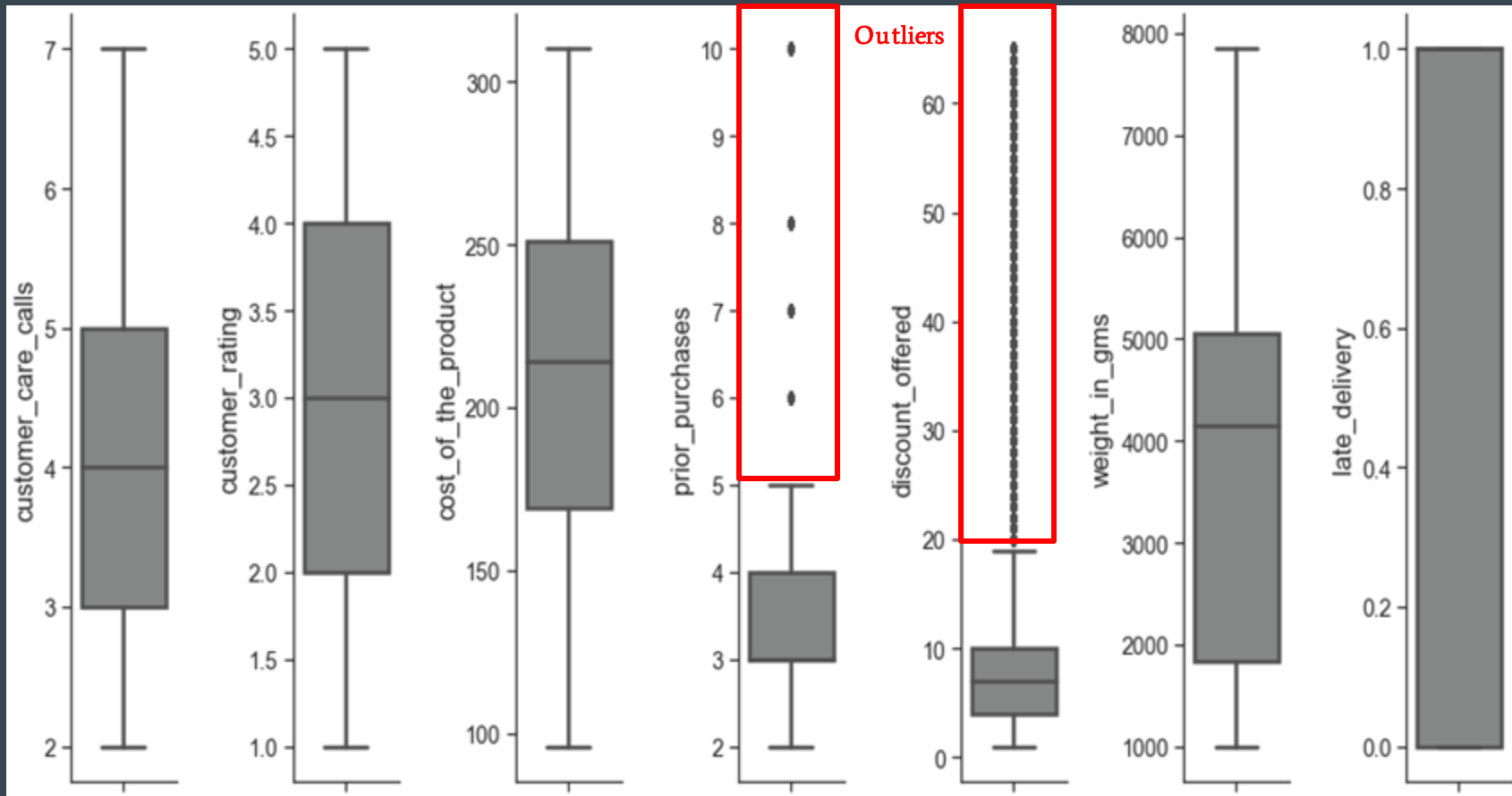
Data Overview

10999 Data of Shipping

COLUMNS

1. ID
2. Warehouse_block
3. Mode_of_shipment
4. Customer_care_calls
5. Customer_rating
6. Cost_of_the_product
7. Prior_purchases
8. Product_importance
9. Gender
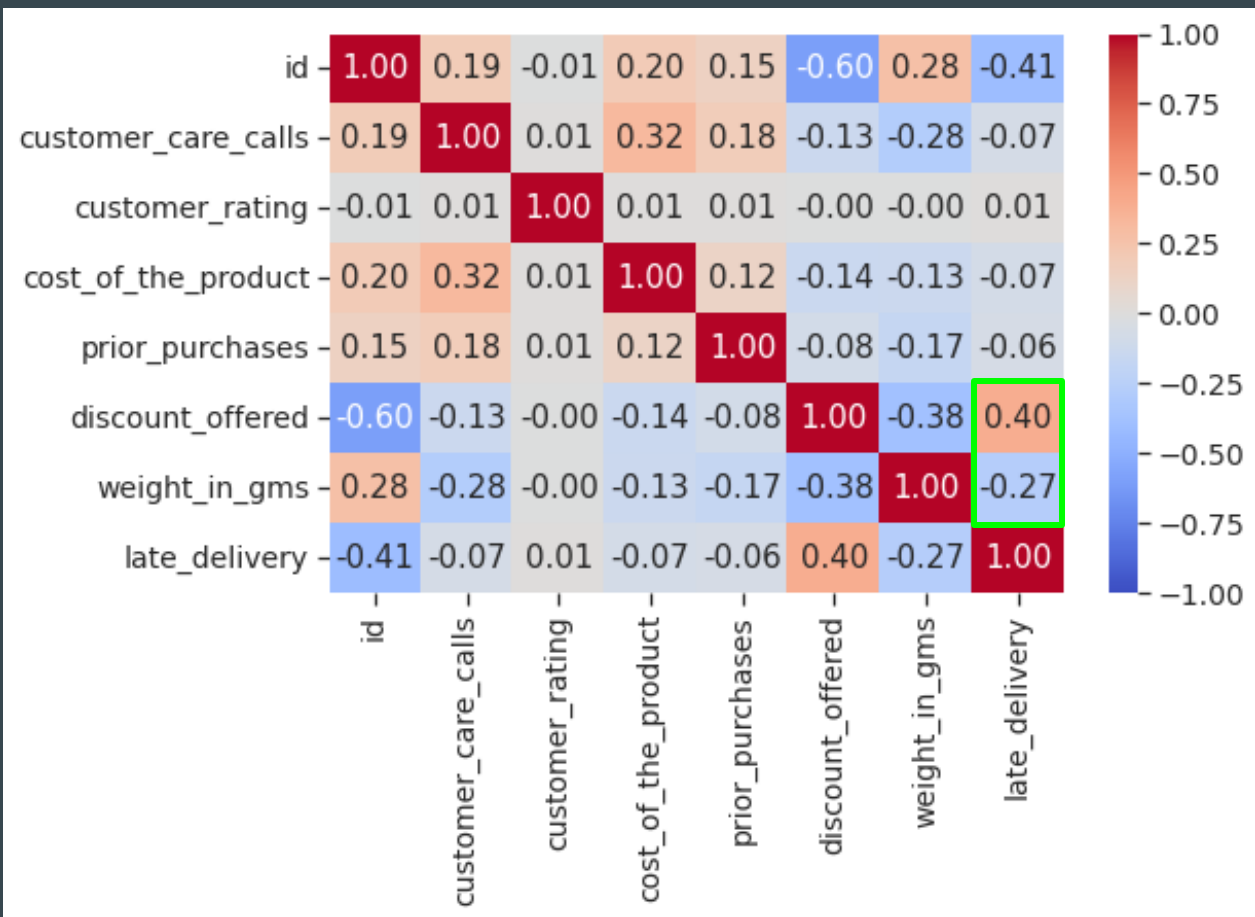10. Discount_offered
11. Weight_in_gms
12. Late_delivery

Column Target: Late_delivery

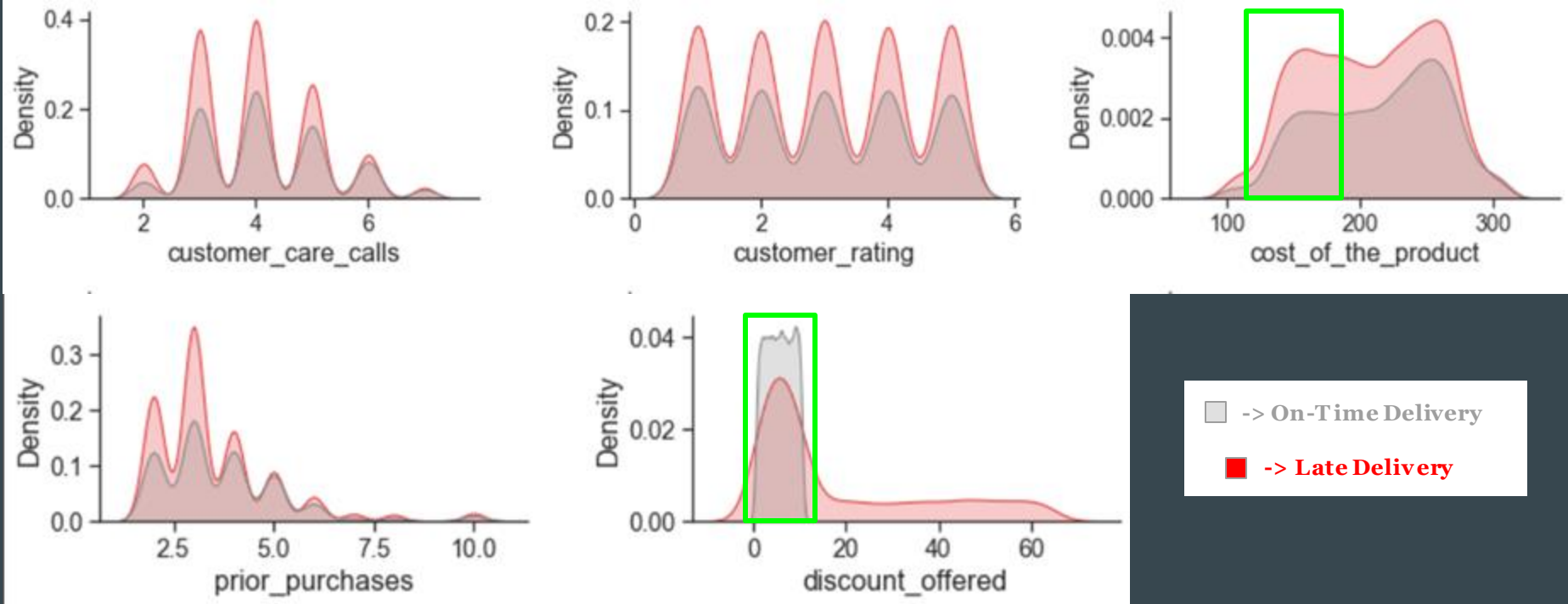# Exploratory Data Analysis

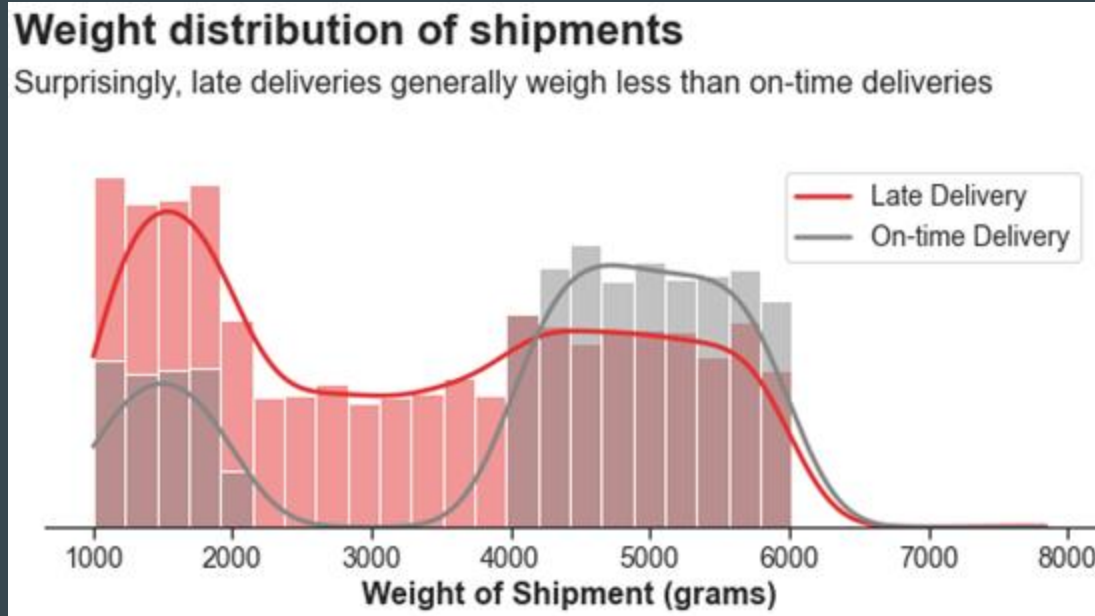# EDA Numerical

Multivariate Analysis

Multivariate Analysis

Distribusi data berdasarkan **late delivery** dan **on-time delivery**

Our Data say...

Weight distributions pada on time delivery dan late delivery

Our Data say...
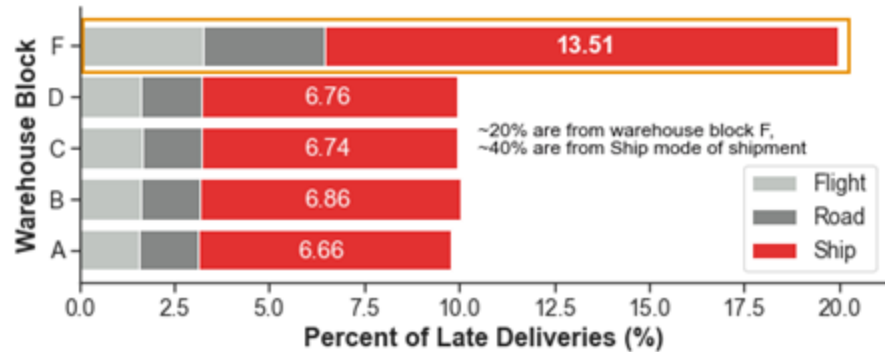
Warehouse block dan mode of shipments pada on time delivery dan late delivery

Late deliveries pada Product importance group

Business Insights

Our Data say...

More Late Deliveries than On-time Deliveries?

There are 20% more late deliveries than on-time deliveries.
Our late delivery rate is ~60%

6563

4436

On-time Delivery
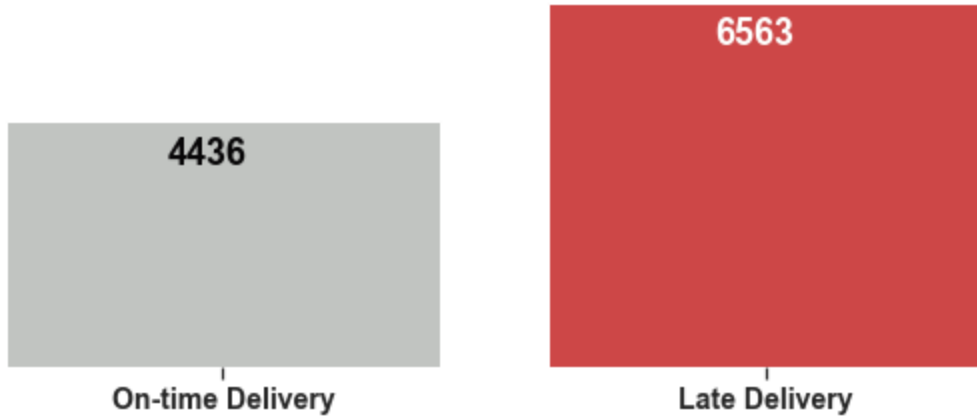
Late Delivery

Jumlah pembelian pada on time delivery dan late delivery

Our Data say...

- " **17% of respondents will stop shopping with a retailer after receiving a late delivery one time.** "

- " **55% of respondents will stop shopping** with a retailer after receiving a **late delivery two to three times.** "

Dari 6563 late deliveries → 17% Late 1x → 55% Late 2-3x → 27% akan churn dari total deliveries

Our Data say...

Average Discount Offered (%)

On-time Delivery: 5.55%
Late Delivery: 18.66%

Average discount pada on time delivery dan late delivery

Data Preprocessing

Data
cleaning

Handling
outlier

Class
Balancing

Splitting
Data

Data
Transformation

Feature
Normalizin
g

Feature
Encoding

Dilakukan untuk train dan test data
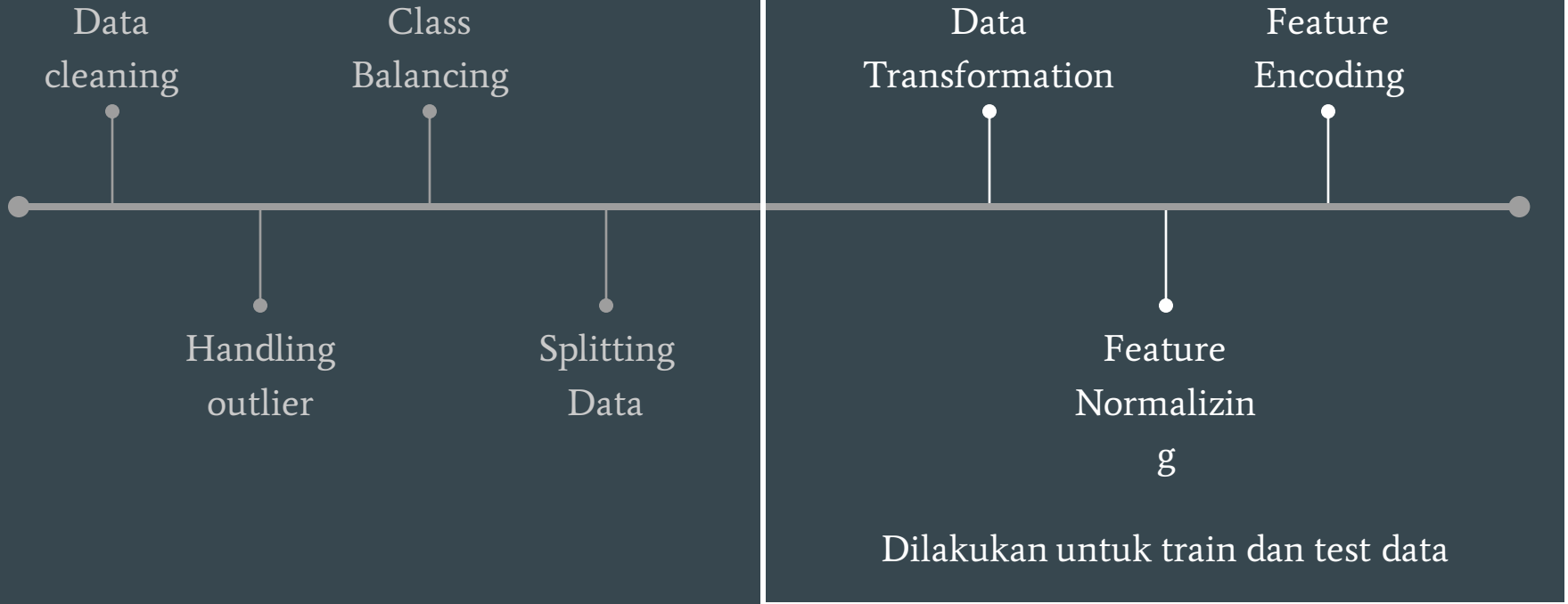
Feature Selection

Melalui beberapa metode* feature selection:

- *trf_weight_in_gms***
- *discount_offered*
- *trf_cost_of_the_product*
- *product_importance_high*

*chi-squared, RF feature importances, dsb

**trf = transformed
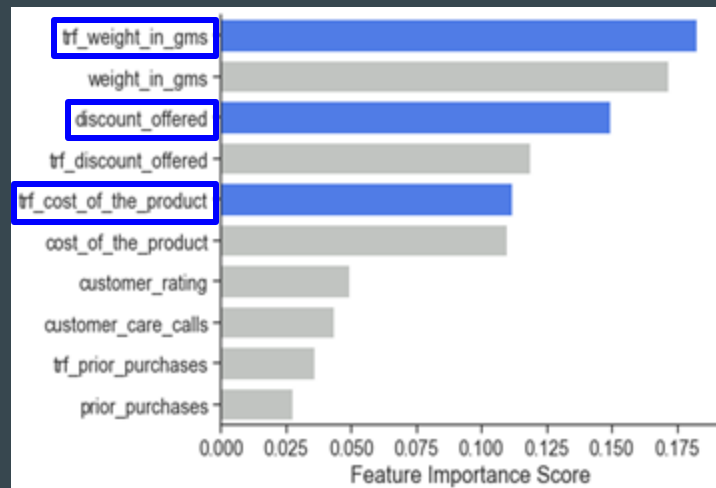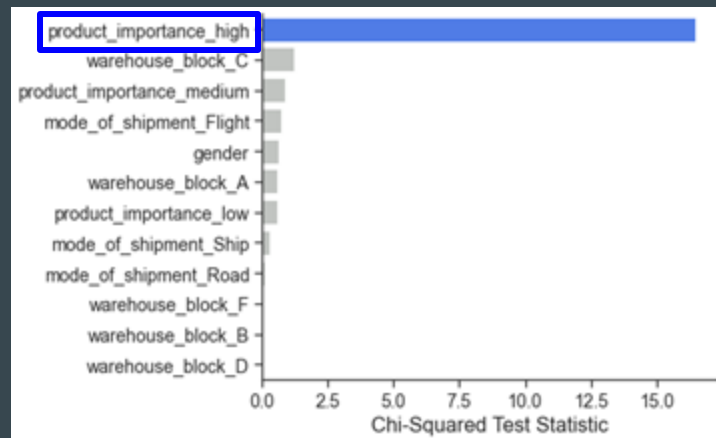
Machine Learning Modelling

Meminimalisir false negatives → Recall Score

Melalui beberapa tahapan model selection*:

- XGBoost RF Classifier

Milestone:

- 0.75 Recall untuk late delivery
- 0.50 Recall untuk on-time delivery

*lazypredict, cross validation, hyperparameter tuning, dsb

	precision	recall	f1-score	support
Late	0.68	0.75	0.71	1257
Ontime	0.61	0.53	0.57	942
accuracy			0.65	2199
macro avg	0.65	0.64	0.64	2199
weighted avg	0.65	0.65	0.65	2199
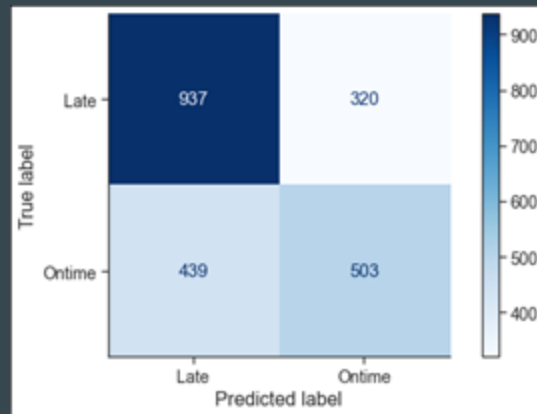
Feature Importances

Dari final model:

- *discount_offered* adalah fitur yang paling berperan dalam prediksi
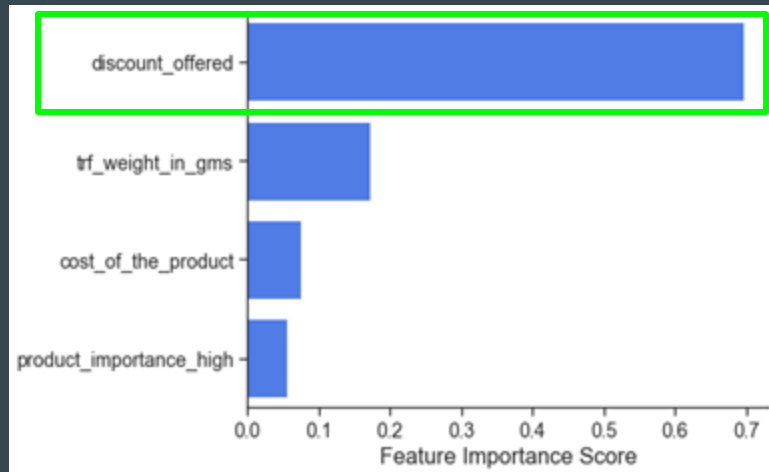
Business Recommendation

1. Flowchart Diagram
2. Future Works

Business Recommendation

1. Melakukan perbaikan manajemen distribusi barang:
   - *Mode of shipment*: jangan terlalu banyak menggunakan moda kapal (*ship*) → buat rekomendasi *mode of shipment* di *checkout page*.
   - *Warehouse*: warehouse block F meng-*handle* terlalu banyak barang, oleh karena itu perlu dilakukan pengaturan agar distribusi barang antar warehouse menjadi lebih merata.
2. Melakukan beberapa *improvement* untuk meningkatkan *customer retention rate*:
   a. Melakukan dual-late redefinition
   b. Memberikan kupon *partial refund* ongkos kirim atau diskon untuk pembelian selanjutnya

Dual-Late Redefinition

Membuat standar atau batas keterlambatan yang berbeda antara pihak logistik dan pihak pelanggan
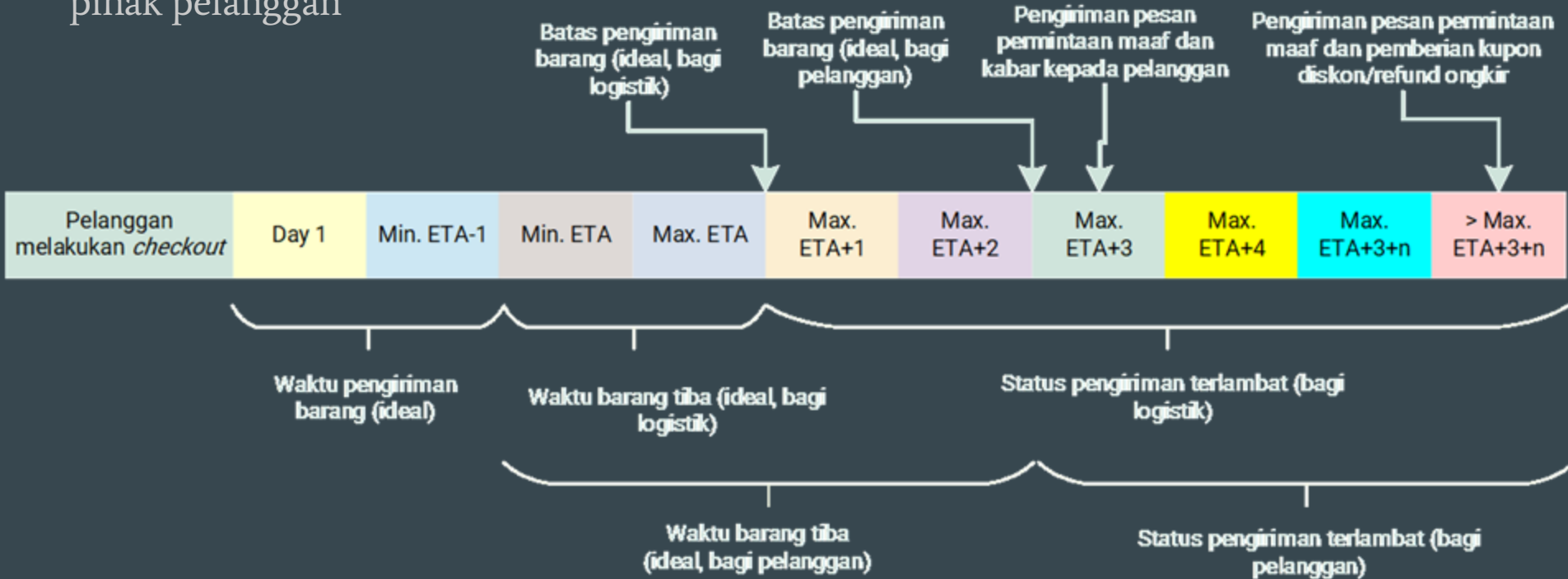
General Shipment Flow Process Diagram: ML and compensation applied
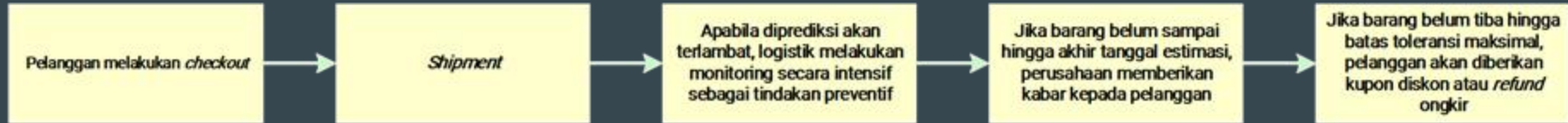
Detailed Shipment Flow Process Diagram: ML, Dual-Late Redefinition and Compensation Applied
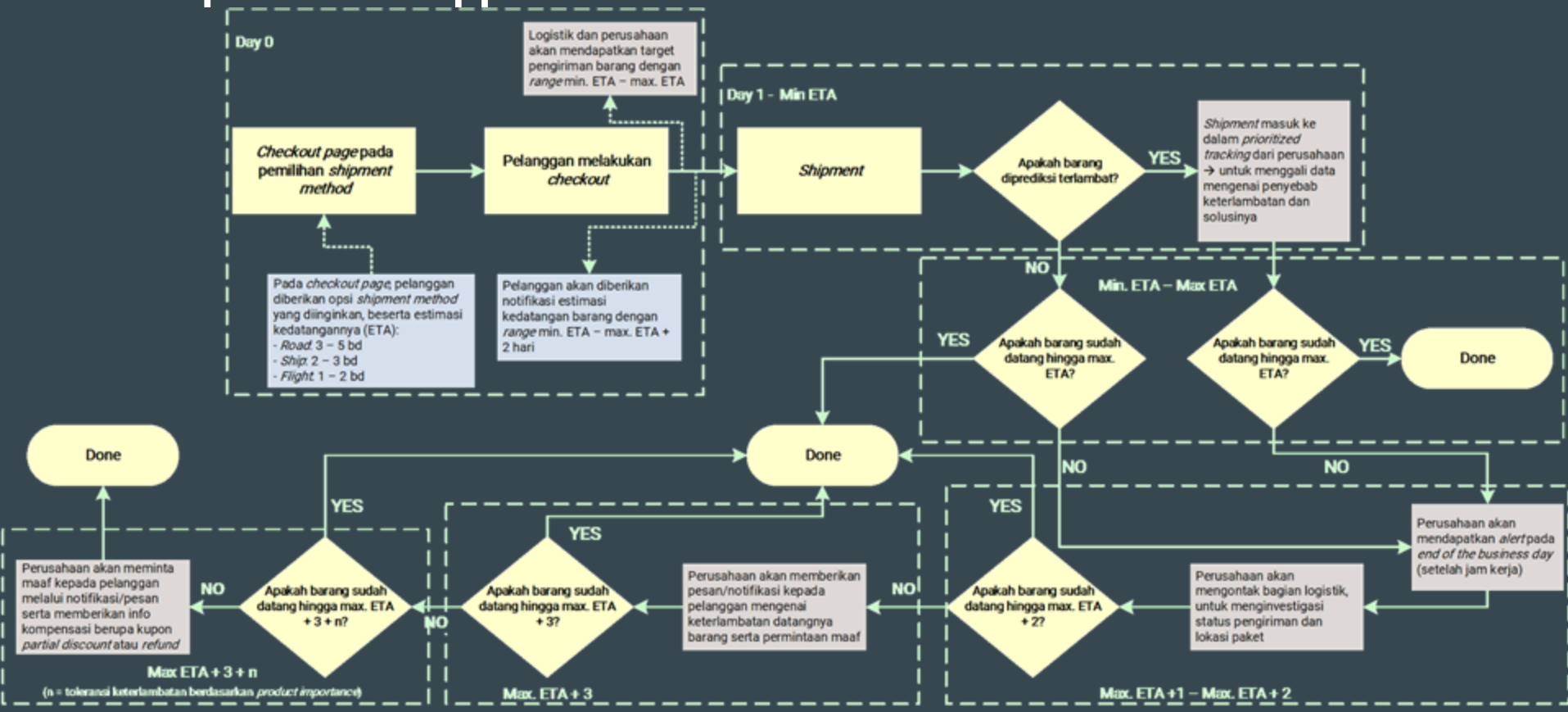
Customer Message Draft #1

Isi surat:
1. Permintaan maaf

1. Penyebab keterlambatan paket

1. Link untuk live tracking pengiriman paket

1. Customer akan dihubungi kembali setelah n hari jika paket belum juga sampai

Dear Mr. XXXX,

Thank you for being patient in waiting for your package. We apologize to you for the delay in your package. After we investigated, we found that your package was delayed due to bad weather during delivery. You can monitor and track your package delivery process by using the following link:

www.grow-ject.com/tracking

Thank you for your understanding. We will get back to you in 3 days, if your package hasn't arrived yet.

Best regards,

PT Grow-Ject Indonesia

Customer Message Draft #2

Isi surat:
1. Permintaan maaf

1. Pemberian kupon diskon atau refund

Dear Mr. XXXX,

Thank you for being patient in waiting for your package. We apologize to you for the delay in your package. As a compensation, we want to give you a 10% discount coupon. This coupon can be used on your next transaction for up to 6 months. Here's a coupon link that you can access:

www.grow-ject.com/coupon/rendeem

Thank you for entrusting the delivery of your goods to us. We are committed to continuously improving the company's performance and systems.

Best regards,

PT Grow-Ject Indonesia

Simulation: Independence and Controlled Variables

Asumsi yang digunakan (variabel terkontrol):

Description		Constant	Remark
Delivery Cost	Flight	8%	of total
	Road	6%	cost of the
	Ship	3%	product
Marketing cost/retention cost ratio		5	
%Churn prob./%Churn prob if refunded or discount ratio		75%	
%delivery problem solved by prioritized tracking		80%	
Compensation	Discount	5%	
	Refund	30%	
Arrived on	ETA+1 to ETA+2	50%	of late1
	ETA+3	60%	of late2
	ETA+4 to ETA+3+n	70%	of late3

Skenario yang digunakan (variabel terikat):

Scenario				
Description				
#	Type	Dual Late Redefinition		Compensation
A1	Conservative	No	No	No compensation
A2		Yes	No	No compensation
A3A		No	Yes	Refund Delivery Cost 30%
A3B				Discount Coupon 5%
A3C				Mixed (50:50)
A4A		Yes	Yes	Refund Delivery Cost 30%
A4B				Discount Coupon 5%
A4C				Mixed (50:50)
B1	ML Implementation	No	No	No compensation
B2		Yes	No	No compensation
B3A		No	Yes	Refund Delivery Cost 30%
B3B				Discount Coupon 5%
B3C				Mixed (50:50)
B4A		Yes	Yes	Refund Delivery Cost 30%
B4B				Discount Coupon 5%
B4C				Mixed (50:50)

Simulation Result

#	Type	Dual Late Redefinition		Compensation	Margin of net revenue (%)	Retention Rate (%)	Remarks
				Scenario			
		Description					
A1		No	No	No compensation	0.00%	72.76%	Base Case
A2		Yes	No	No compensation	2.05%	86.38%	
A3A		No	Yes	Refund Delivery Cost 30%	0.13%		
A3B	Conservative			Discount Coupon 5%	-0.47%	76.17%	
A3C				Mixed (50:50)	-0.17%		
A4A		Yes	Yes	Refund Delivery Cost 30%	2.07%		
A4B				Discount Coupon 5%	2.00%	86.79%	
A4C				Mixed (50:50)	2.03%		
B1		No	No	No compensation	2.52%	89.47%	Base Case
B2		Yes	No	No compensation	3.31%	94.72%	
B3A		No	Yes	Refund Delivery Cost 30%	2.54%		
B3B	ML Implementation			Discount Coupon 5%	2.45%	90.00%	
B3C				Mixed (50:50)	2.49%		
B4A		Yes	Yes	Refund Delivery Cost 30%	3.32%		
B4B				Discount Coupon 5%	3.29%	94.90%	
B4C				Mixed (50:50)	3.30%		

+22%

customer retention rate +

72.76% → 94.90%

Conclusion

1. Model *Machine Learning* yang dihasilkan memiliki *recall score* 0,75 untuk *late delivery* dan 0,53 untuk *on-time delivery*.

2. Penggunaan *machine learning*, *dual-late redefinition*, dan kompensasi berupa *partial refund delivery cost* akan memberikan *margin of net revenue* serta *customer retention rate* maksimal, dengan nilai masing-masing 3,32% dan 94,9%

3. Pemberian *discount coupon* dan *refund delivery cost* maksimal yang dapat diberikan masing-masing maksimal yang dapat diberikan 29,66% dan 56,51%.

Future Works

- Gali data *prioritized tracking* (dengan asumsi data lebih relevan dan lebih *targeted*), untuk mengetahui lebih mengenai penyebab *late delivery*
- Buat kembali *machine learning modelling* dengan data yang baru untuk memperbaiki kekuatan prediksi
- Melakukan eksplorasi data lapangan dan data riset agar konstanta dan asumsi yang digunakan dalam simulasi dapat lebih *reliable*.

```
            T
      A R I G A T O U G O Z A I M A S U                    M
   T     C       H       R                            H    M A
   A  K     K    A     C A M O N B A N                H V A L A  T U
      G     A    N         C                  T O D A         U R
 T E R I M A K A S I H           O   V            T E       L A N
   A     S       Y       A       B   I            E S              D
 A Z     A       O     S Y U K R A N   E          E K              Y A
 S I     H       U               B     A    F     K U  K O P K H U N K H A
 A E     A             D         I     K    H     U R              N U U
 N       M             A         G          H     R E      M         Y
 T       N             N         K    S A L A M A T U      E         U
 W E L A L I N         K I I T O S          R    R D      R C
         D             E       X            I    E  I     C I
       S P A S I B O           X        M U L T U M E S C  I
                               I            O       D R    I
             D Z I E K U J E                        R I
                                                    I
                                                    M
```