Grow-Ject: Tracking E-commerce Shipping

FINAL PROJECT

DATA SCIENCE BOOTCAMP BATCH 17

RAKAMIN ACADEMY

Team Members

Fakhri Nurrahmadi (Leader)

Muhammad Iqbal Tawakkal

Ryan Rizky Fathinanto

Dyah Phitaloka

Dwi Muhammad Nurafli

Outline

Introduction

Introduction

Syntax:
https://drive.google.com/file/d/1Ma7MWyoMjl1cFv4VXmwPK6AgvgMDwo5l/view?usp=sharing

Dataset:
https://www.kaggle.com/prachi13/customer-analytics

Our Job Role:

Masalah Bisnis

↓

Solusi Bisnis

↓

Peningkatan Performa Bisnis

Berdasarkan data dari Hollingsworth:

1. Sebanyak 17% pelanggan akan berhenti berbelanja jika paket mereka datang terlambat satu kali.

2. Sebanyak 55% pelanggan akan berhenti berbelanja jika paket mereka datang terlambat 2 – 3 kali.

3. Di sisi lain, usaha untuk mendapatkan satu pelanggan baru sama dengan usaha untuk mempertahankan lima pelanggan lama.

Introduction

Goal:
Meningkatkan *customer retention rate*

Objectives:
- Membuat model *machine learning* yang bertujuan untuk memprediksi *late delivery*
- Menggali data untuk mengetahui penyebab *late delivery*
- Memberikan treatment yang tepat kepada customer yang pengirimannya diprediksi sebagai *late delivery* untuk me-*retain customer* (sebagai bagian dari *risk control* dalam *risk management*)

Business Metrics:
customer retention rate

$$\boxed{CRR = \frac{CE - CN}{CS} . 100\%}$$

Dimana:

CRR = *customer retention rate*

CE = jumlah pelanggan di akhir periode

CN = jumlah pelanggan baru di periode berjalan

CS = jumlah pelanggan di awal periode

Exploratory Data Analysis (EDA)

Data Overview

Jumlah baris: 10999 data

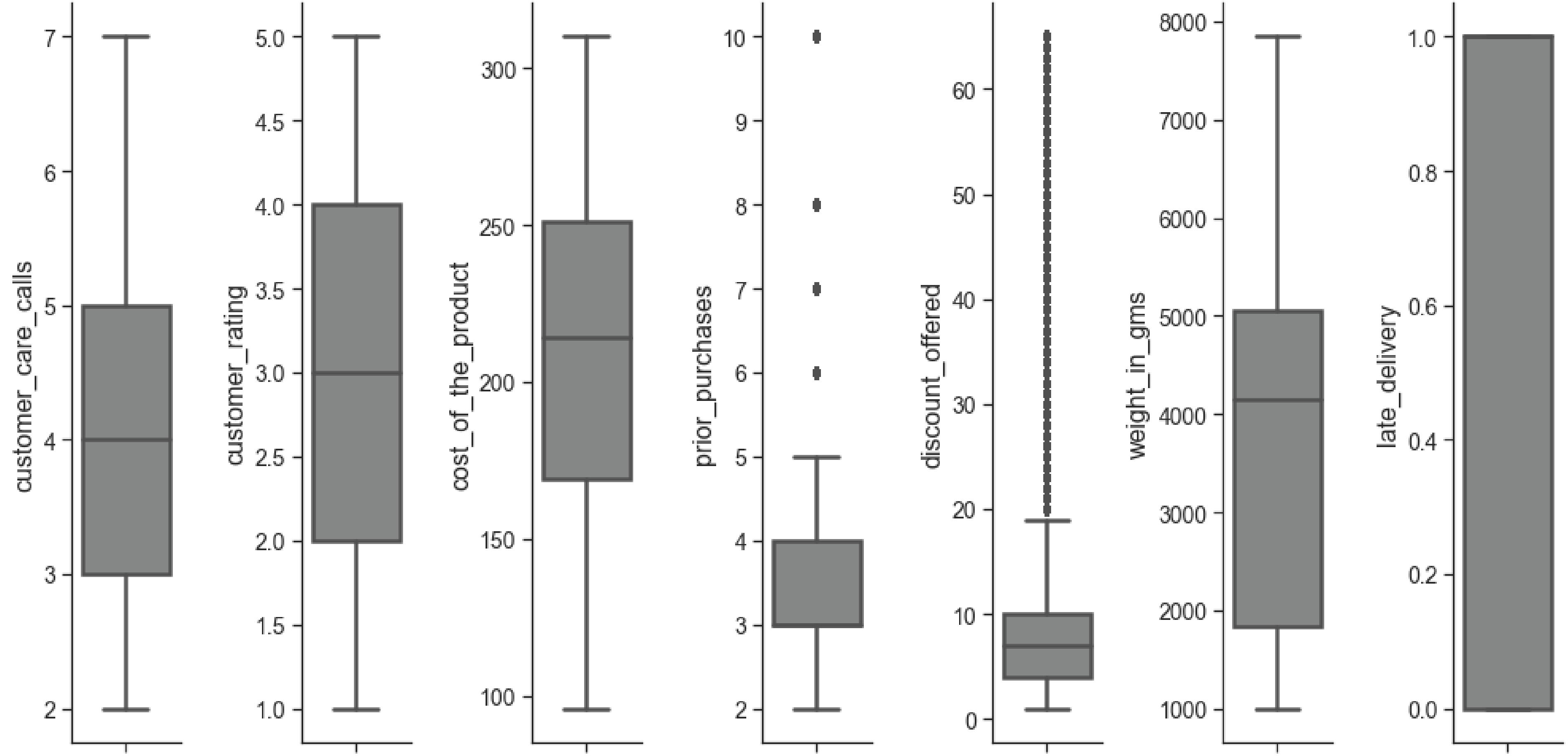
Kolom yang tersedia:

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 → Kolom target

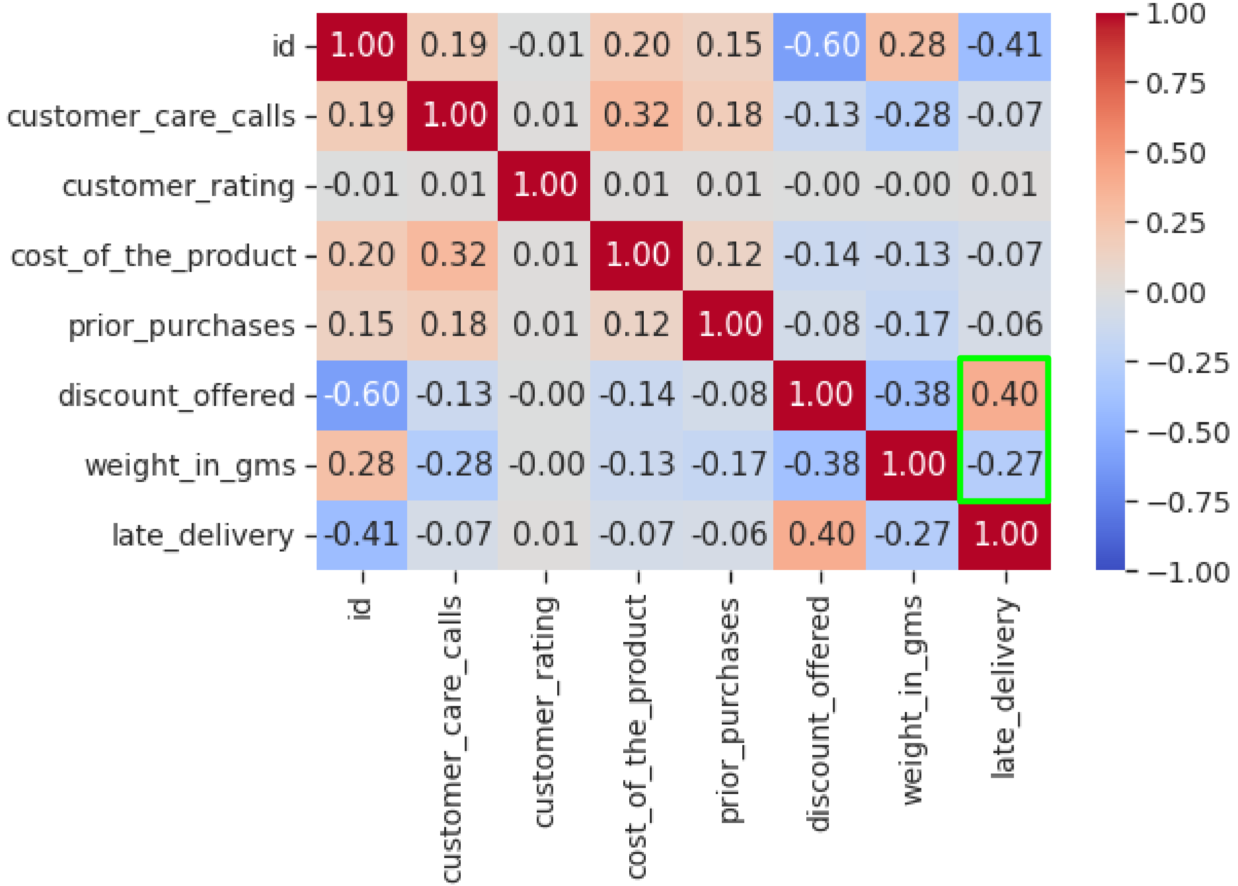
Hasil pengamatan:

1. semua tipe data sudah benar
2. tidak ada *null values* maupun *duplicated values*

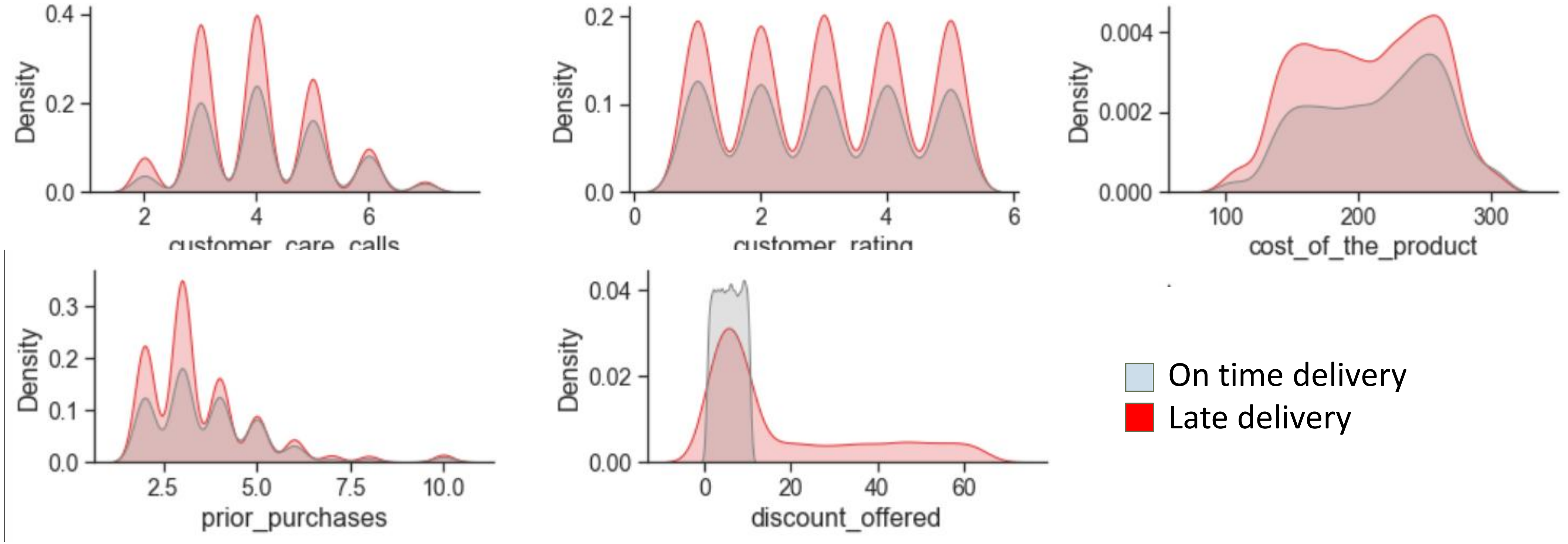
EDA: Numerical



Multivariate Analysis



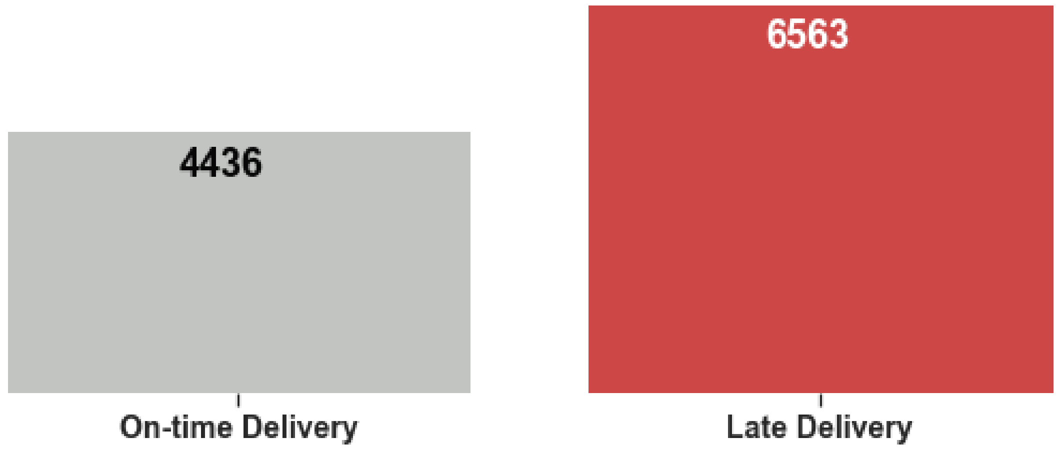
Multivariate Analysis:
Distribusi Data Berdasarkan Kolom late_delivery

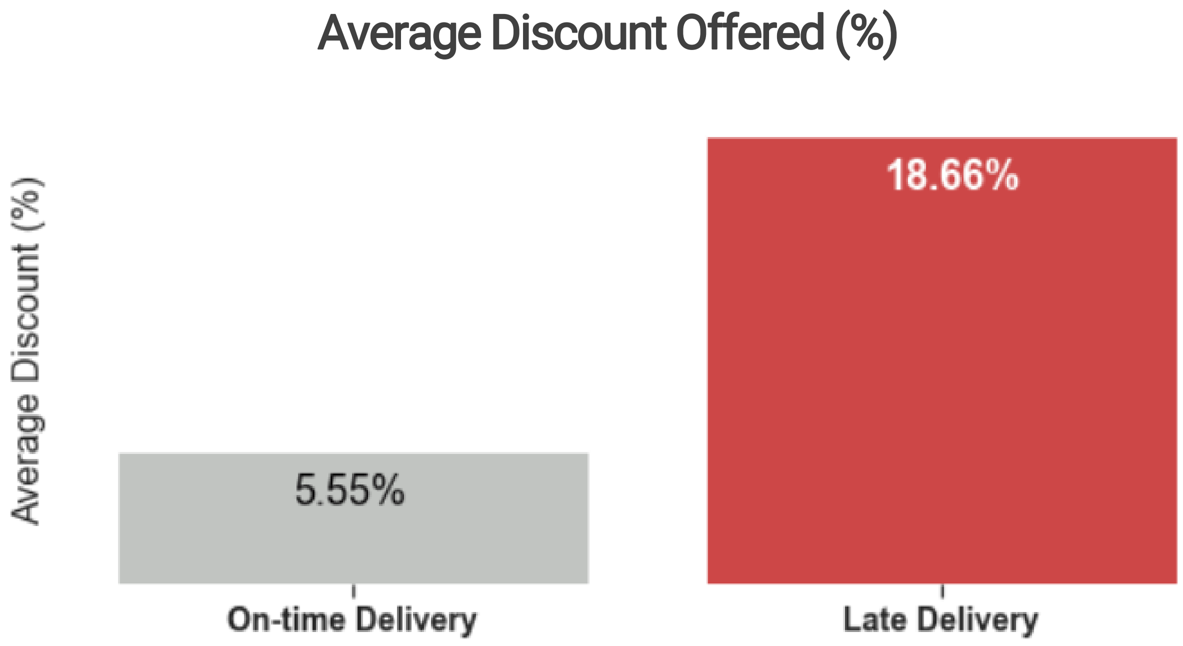


Business Insights

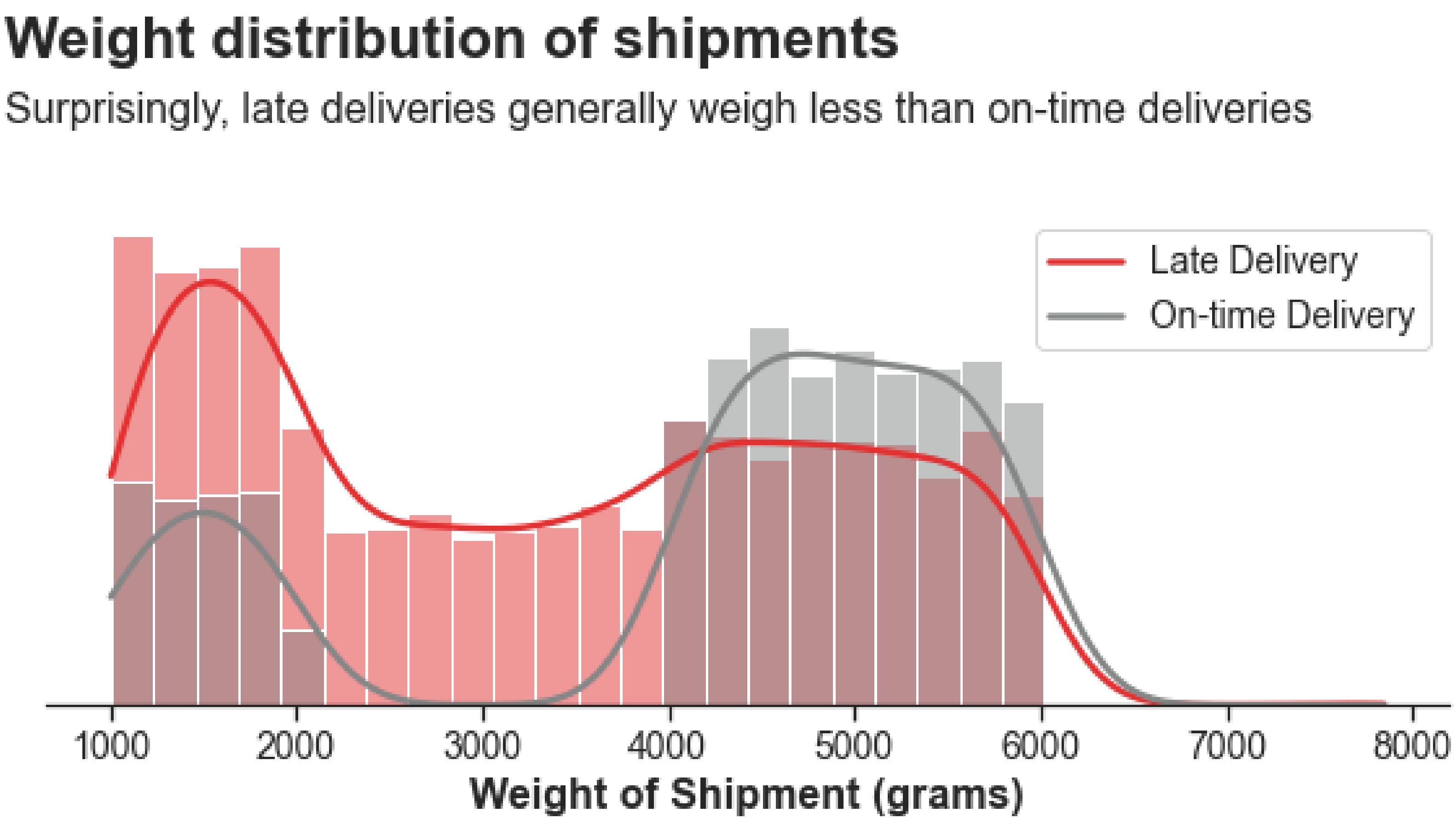
More Late Deliveries than On-time Deliveries?

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





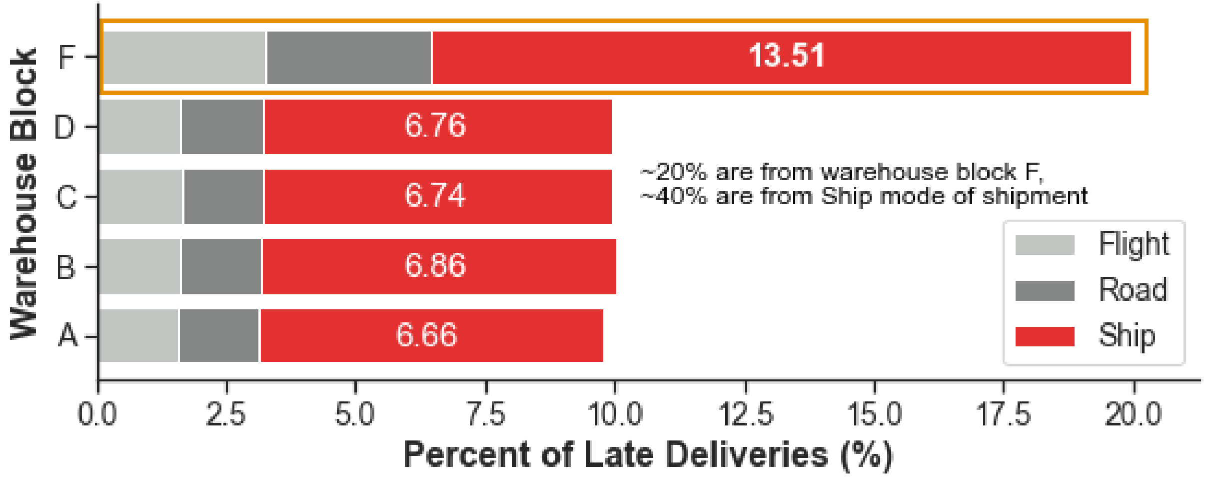
Business Insights



Business Insights

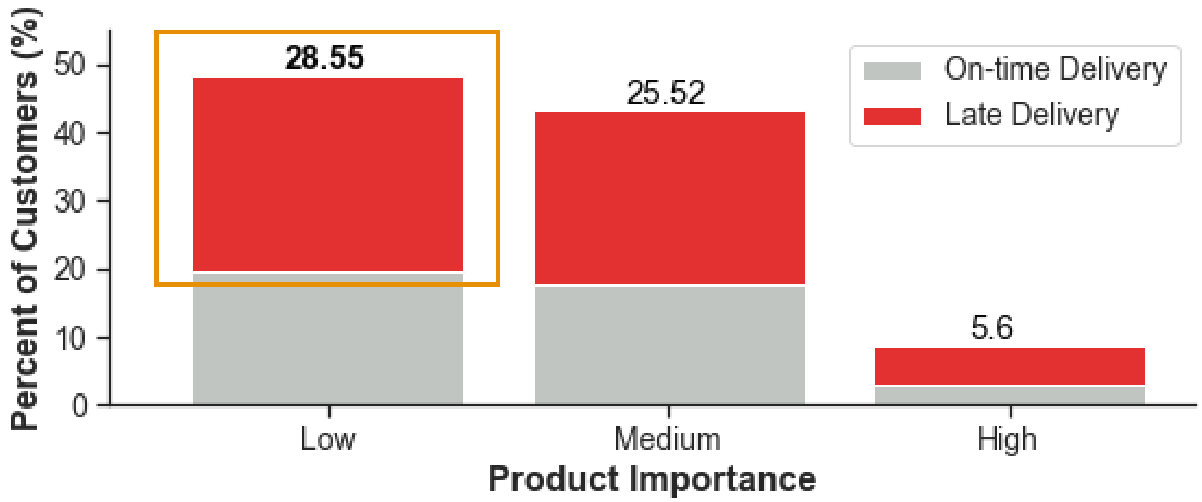
Distribution: warehouse blocks and shipment modes

Most of the late arrivals are shipments from warehouse block F and shipments using Ship mode of shipment. Redistribution is needed!



Late deliveries from each product importance group

"Low" has the most late deliveries. However, based on relative ratio, "High" has the highest relative ratio. The grouping of product importance is ambiguous.



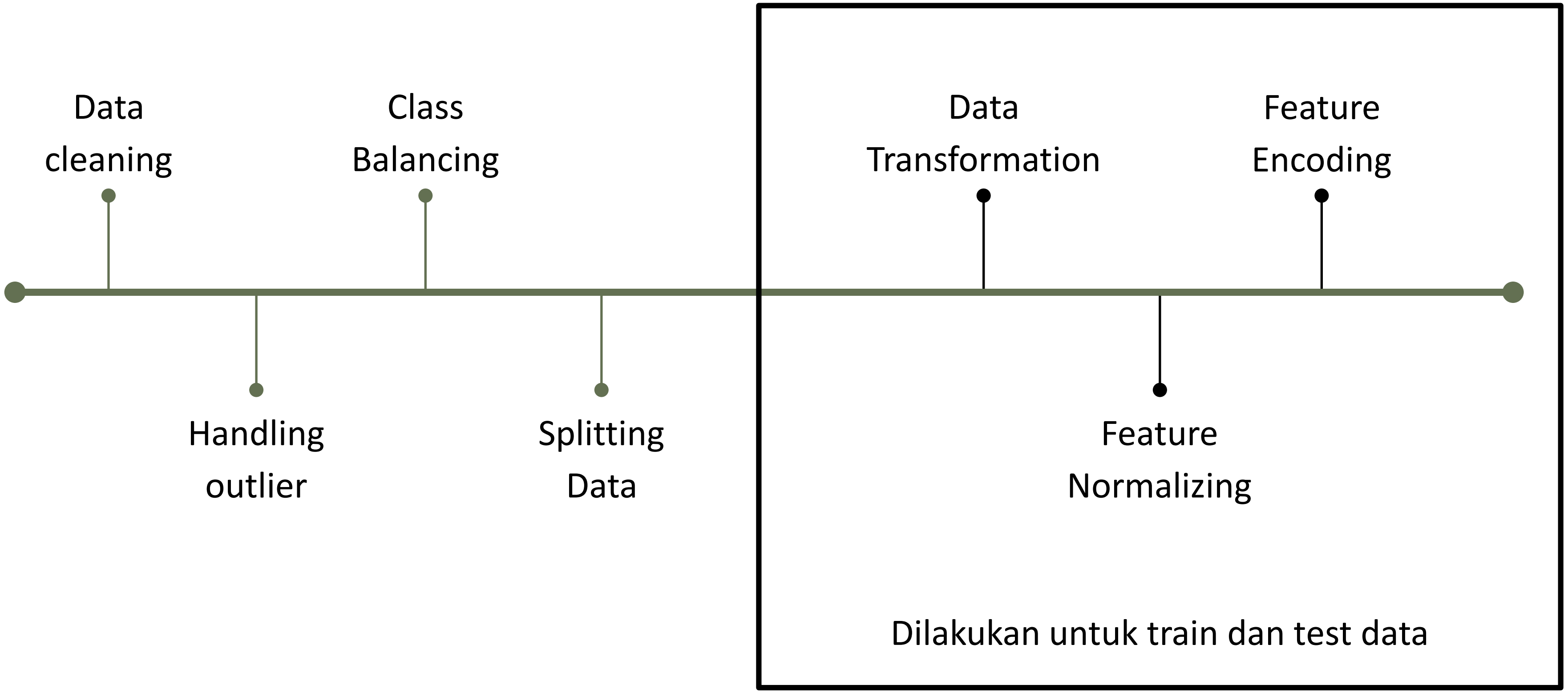
Business Insights

Berdasarkan data dari [Hollingsworth] yang telah dipaparkan sebelumnya:

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

Data Preprocessing & Machine Learning Modelling

Data Preprocessing



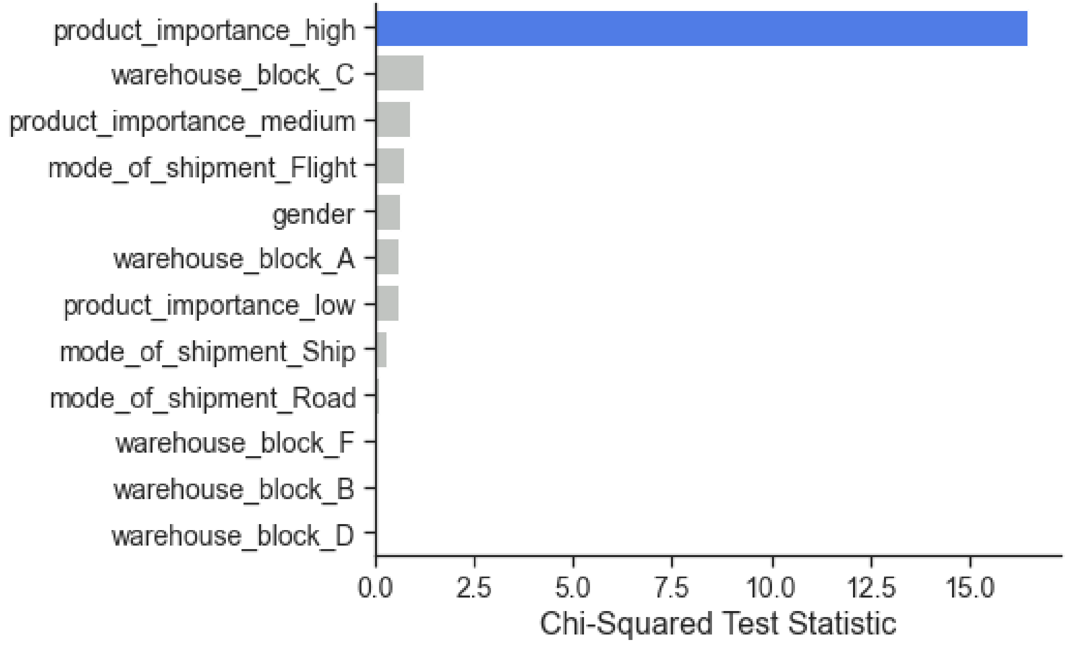
Feature Selection

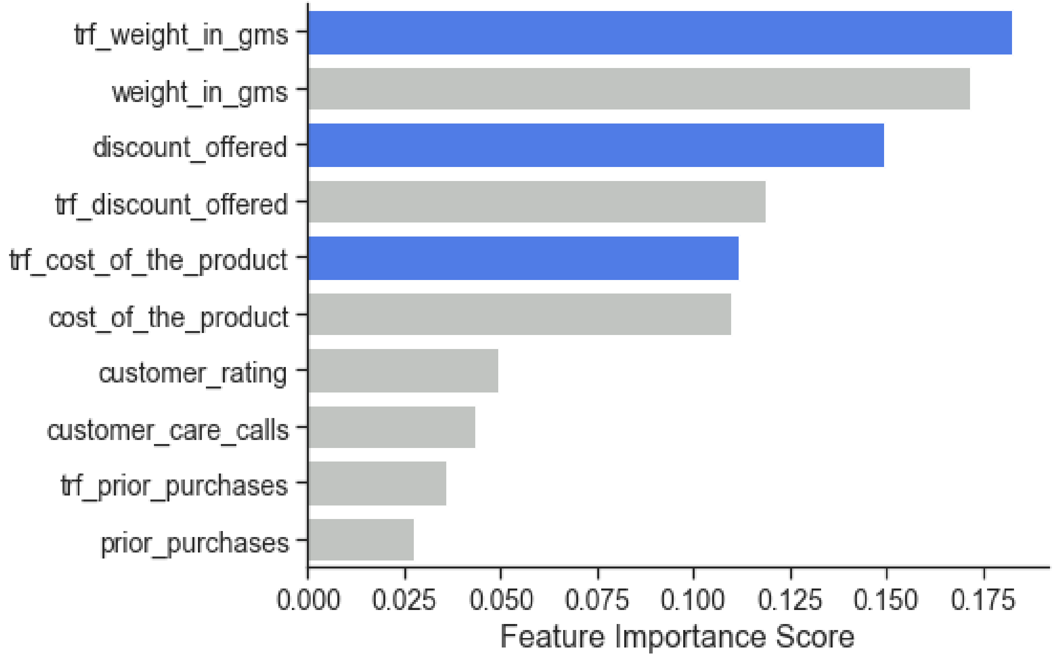
Melalui beberapa metode* feature selection:

- *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

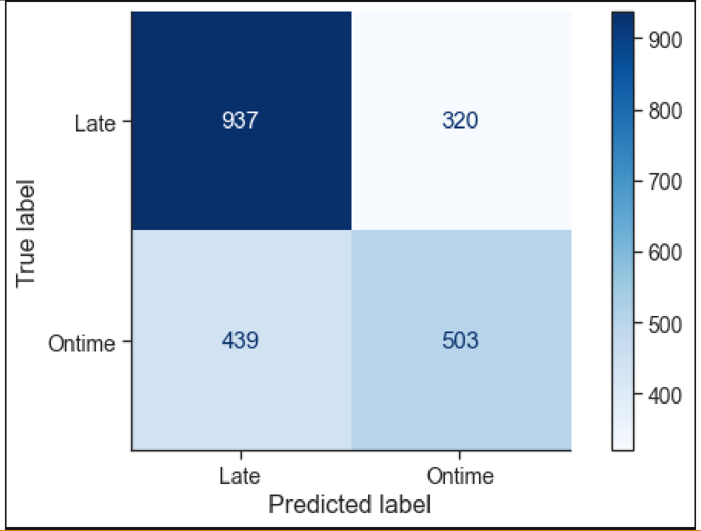
Melalui beberapa tahapan model selection*:

- XGBoost RF Classifier

Milestone:

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

	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



*lazypredict, cross validation, hyperparameter tuning, dsb

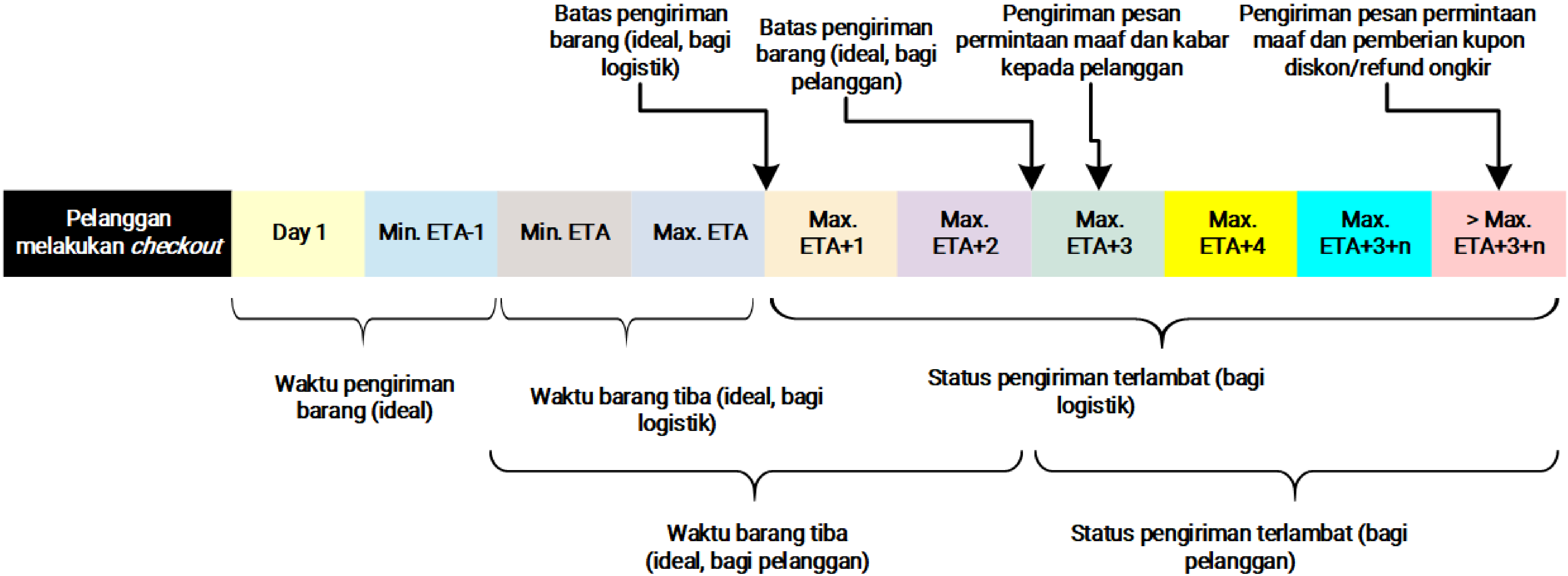
Business Recommendation

Business Recommendation

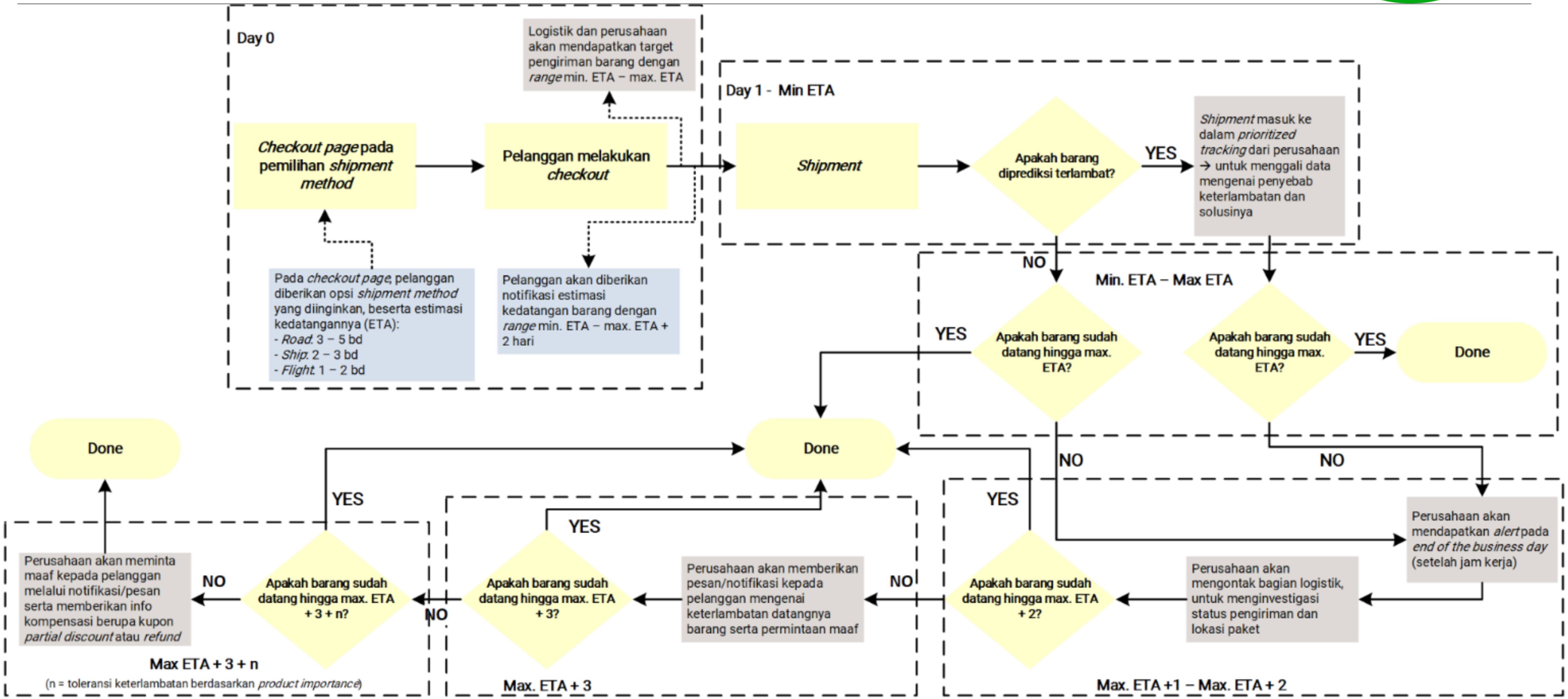
- **Melakukan perbaikan manajemen distribusi barang:**
    - *Mode of shipment*: jangan terlalu banyak menggunakan moda kapal (*ship*) → buat rekomendasi *mode of shipment* di *checkout page*.
    - *Warehouse*: warehouse block F meng-*handle* terlalu banyak barang, oleh karena itu perlu dilakukan pengaturan agar distribusi barang antar warehouse menjadi lebih merata.
- **Melakukan beberapa *improvement* untuk meningkatkan *customer retention rate*, seperti:**
    - Melakukan dual-late redefinition
    - 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, Dual Late Redefiniton, and Compensation Applied



Customer Message Draft #1

Isi surat:

1. Permintaan maaf
2. Penyebab keterlambatan paket
3. Link untuk live tracking pengiriman paket
4. 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

2. Pemberian kupon diskon atau refund ongkos kirim

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 cost of the product
	Road	6%	
	Ship	3%	
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)

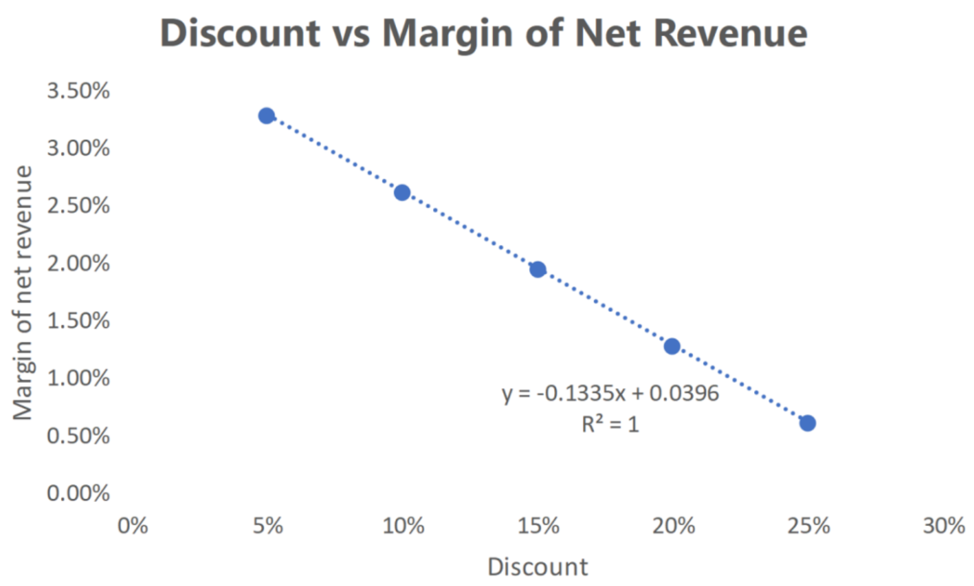
Simulation: Results

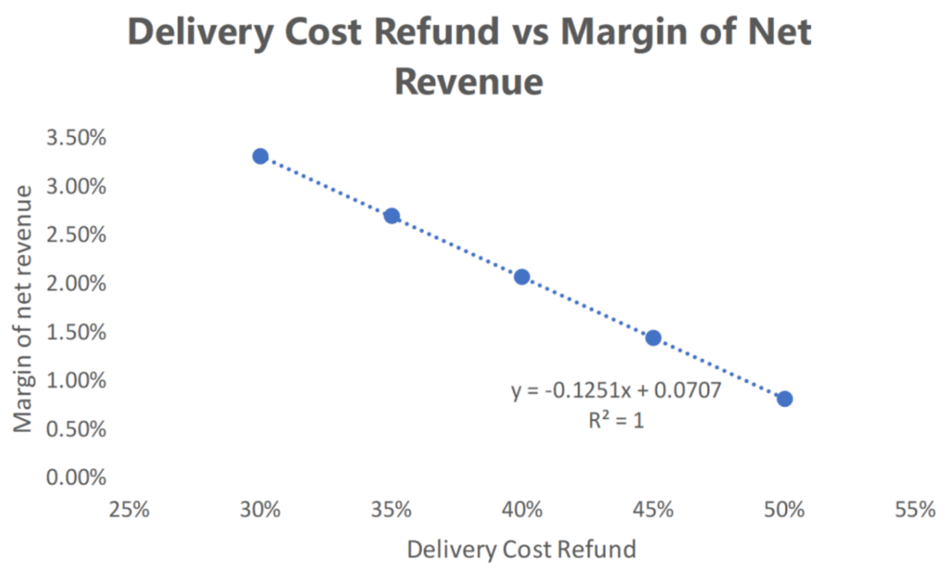
#	Scenario Description: Type	Dual Late Redefinition	Compensation		Margin of net revenue (%)	Retention Rate (%)	Remarks
A1	Conservative	No	No	No compensation	0.00%	72.76%	Base Case
A2		Yes	No	No compensation	2.05%	86.38%	
A3A		No	Yes	Refund Delivery Cost 30%	0.13%	76.17%	
A3B				Discount Coupon 5%	-0.47%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

Skenario A3B dan A3C menghasilkan *margin of net revenue* < 0, sehingga sangat tidak disarankan untuk diaplikasikan

Skenario B4A menghasilkan *margin of net revenue* dan *retention rate* paling tinggi → direkomendasikan untuk diaplikasikan

Sensitivity Analysis: Scenario B4 with Discount Rate & Refund Delivery Cost as Independent Variables





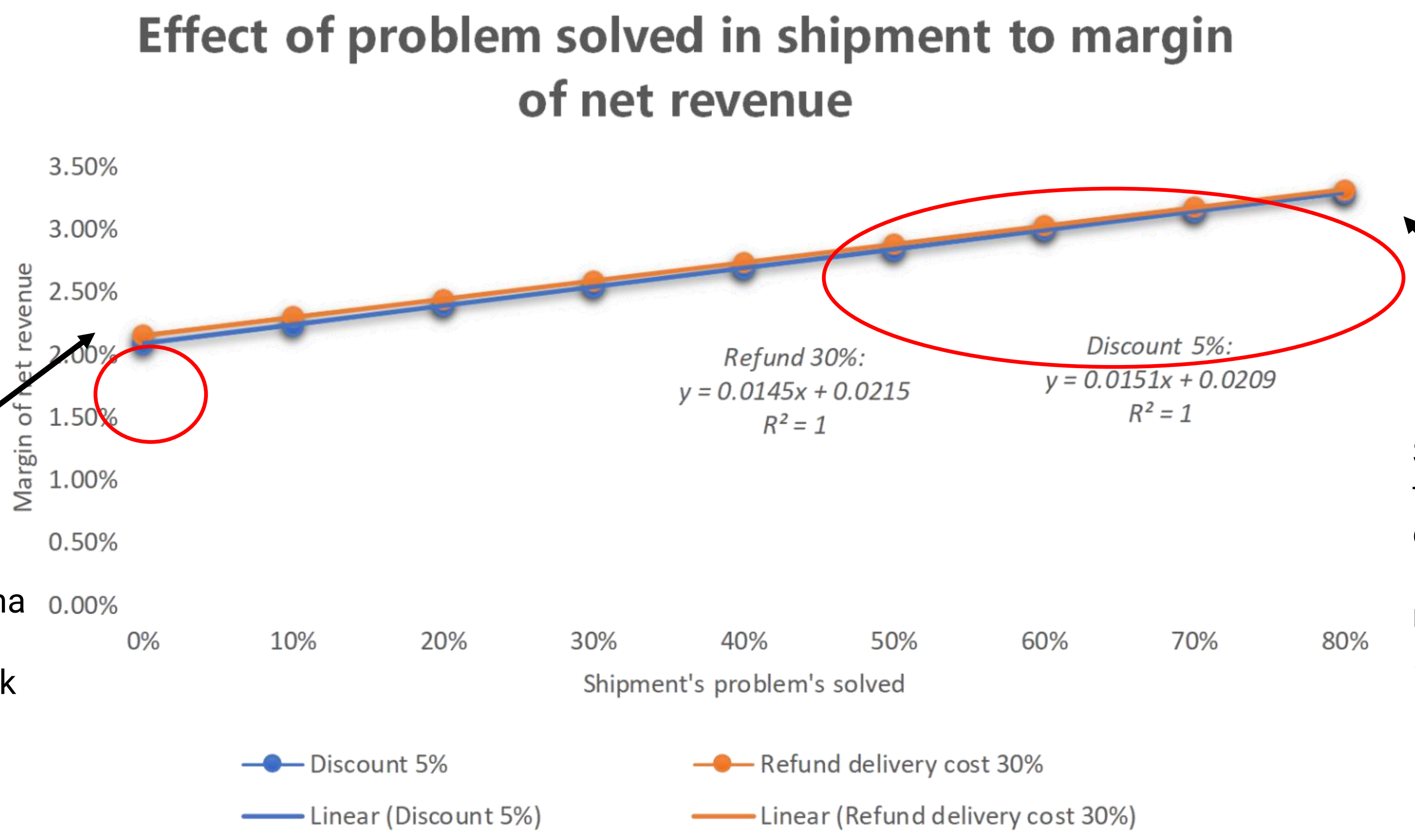
Variabel diskon lebih sensitif terhadap *margin of net revenue* dibandingkan variabel *delivery cost refund*

Kompensasi maksimal yang dapat yang diberikan:

Discount rate < 29.66%

Refund delivery cost < 56.51%

Sensitivity Analysis: Scenario B4 with Shipment's Problem Solved as Independent Variable

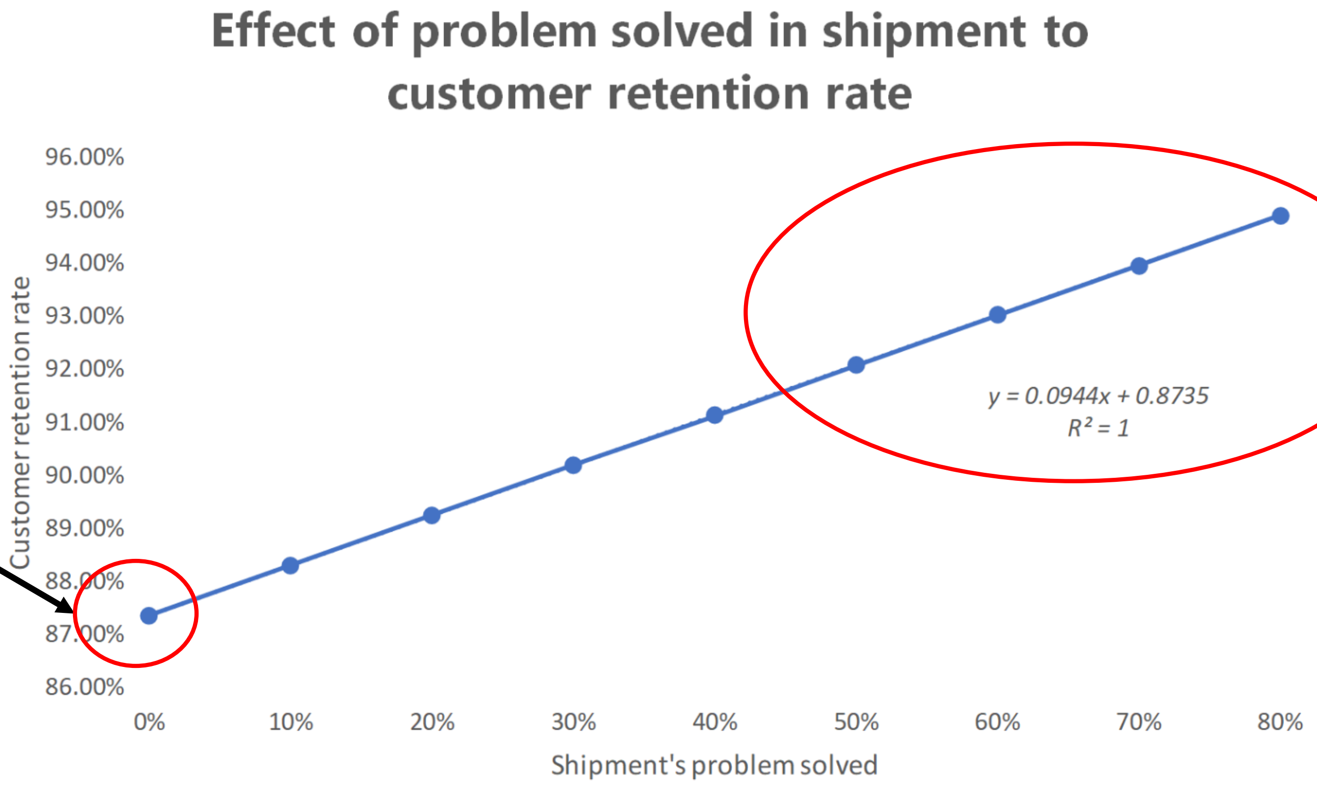


Nilai *margin of net revenue* sama dengan yang terjadi pada skenario A4 → ML menjadi tidak berguna jika tidak ada tindakan preventif dari pihak logistik

Semakin efektif tindakan preventif yang dilakukan pihak logistik, semakin tinggi nilai *margin of net revenue*

Sensitivity Analysis: Scenario B4 with Shipment's Problem Solved as Independent Variable

Nilai *customer retention rate* sama dengan yang terjadi pada skenario A4 → ML menjadi tidak berguna jika tidak ada tindakan preventif dari pihak logistik



Semakin efektif tindakan preventif yang dilakukan pihak logistik, semakin tinggi nilai *customer retention rate*

[Detailed simulation]

Kesimpulan

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

growject

- 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
  ARIGATOUGOZAIMASU       M
  C    H  R             H A
  K  K A CAMONBAN       V T
   G A N  C          TODA U
 TERIMAKASIH  O V    E  L R
   A S Y  A   B I    S  A N D
 A Z A O  SYUKRAN E  E    U Y
 S I H U      I A F  K    W A
 A E A    D   G K H  KOPKHUNKHA
 N   M    A  SALAMAT U    N U
 T   N    N X D   R  R M    Y
WELALIN   KIITOS  I  E E    U
     D    E E     S  D R
   SPASIBO  X  MULTUMESC
            I     O  R I
         DZIEKUJE
                     M
```

growject