



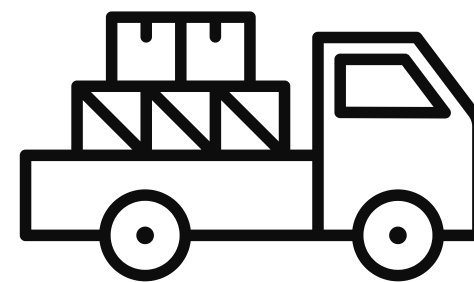
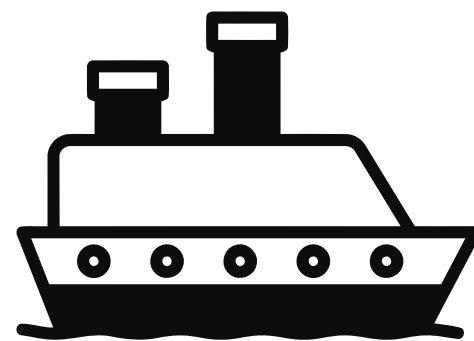
Final Project

by

Team Cobra

Bootcamp Data Science
Rakamin Academy – Batch 31

Product Shipment Delivery Prediction



Meet the Team



Ariella Arima
Aniendra



Dicky Maulana
Rozi



Fariyatul
Ainiyah



Mezky Matthew
Yandito



Mufid
Habiburrahman



Muhamad
Fariduddin Syafiq



Muhammad Raffi
Yudhistira

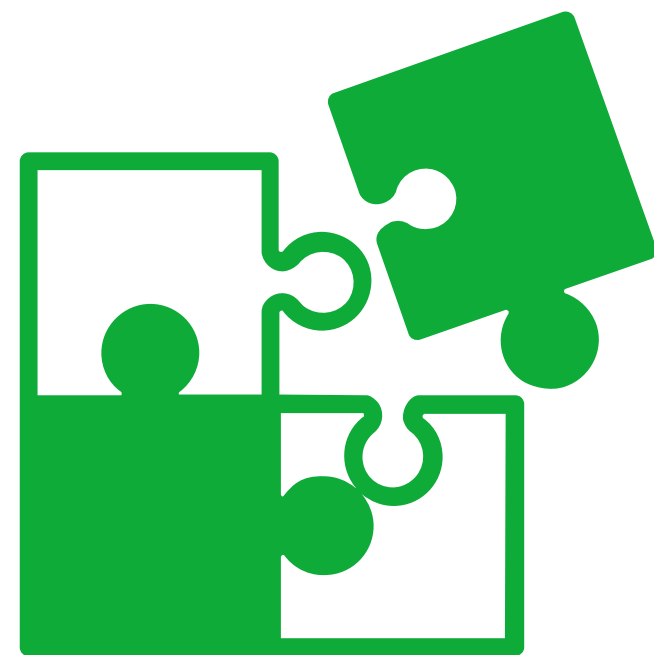
Data

Source

E-Commerce Shipping Data ([Kaggle](#))

Context

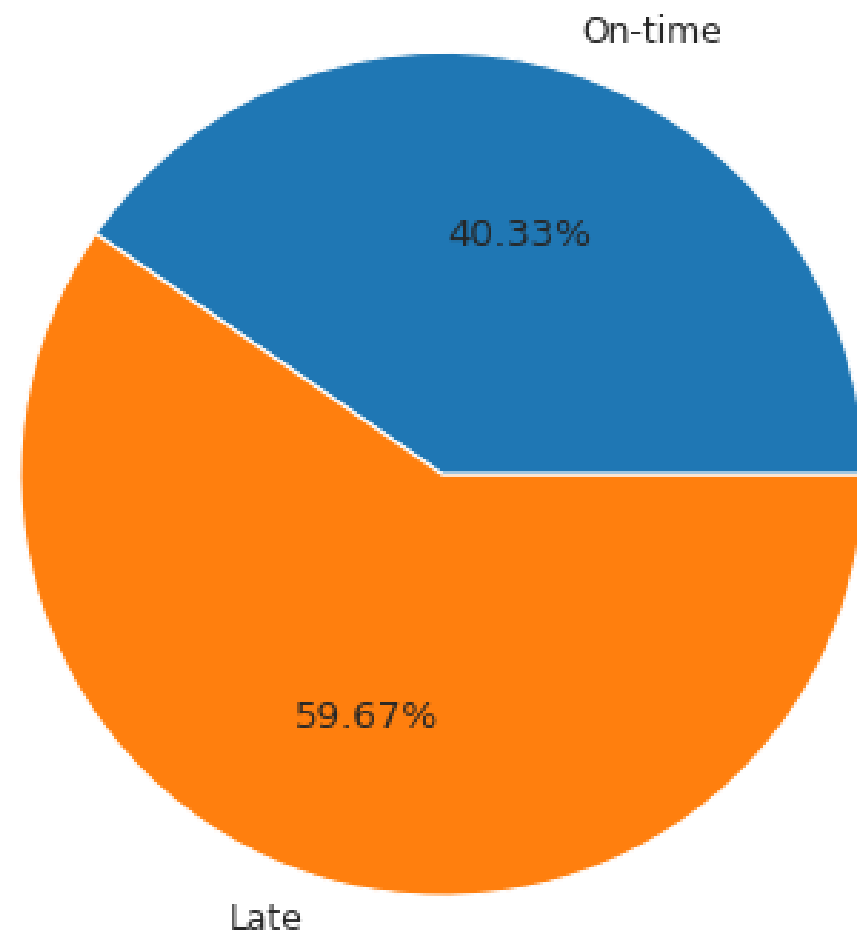
An international e-commerce company based wants to discover key insights from their customer database. They want to use some of the most advanced machine learning techniques to study their customers. The company sells electronic products.



The Case

What is the Problem?

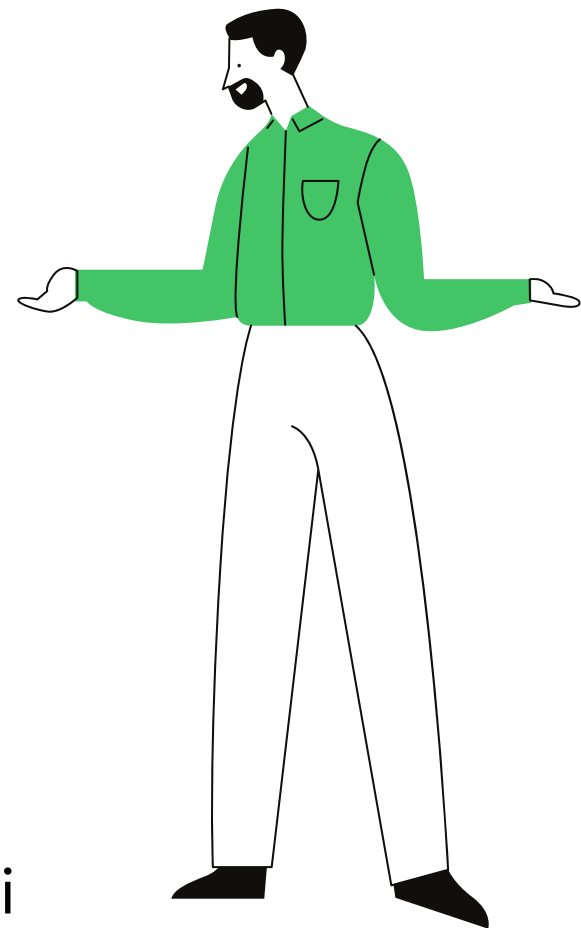
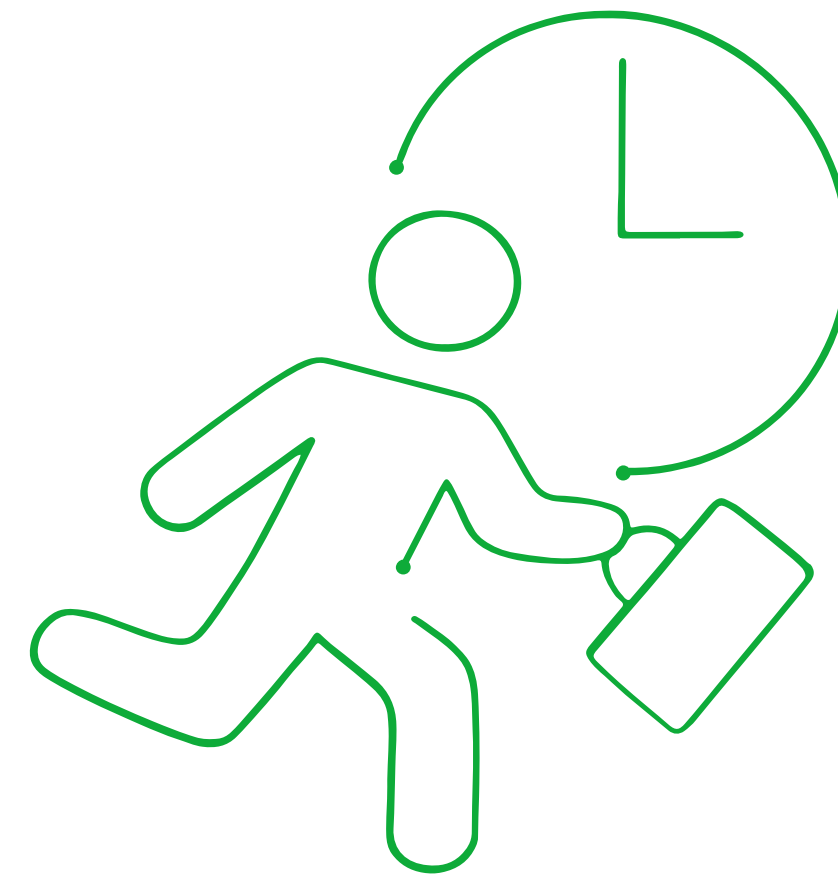
On-time vs Late Deliveries Ratio



Pada suatu perusahaan e-commerce sering terjadi **keterlambatan pengiriman barang**, lebih banyak dibandingkan pengiriman yang tepat waktu

Impacts of Late Deliveries*:

- Decrease in Customer **Retention Rate** & Customer **Lifetime Value**
- Additional shipping **costs**
- Tarnished company **reputation**



*source: softec.ai

What Can We Do?

Objective

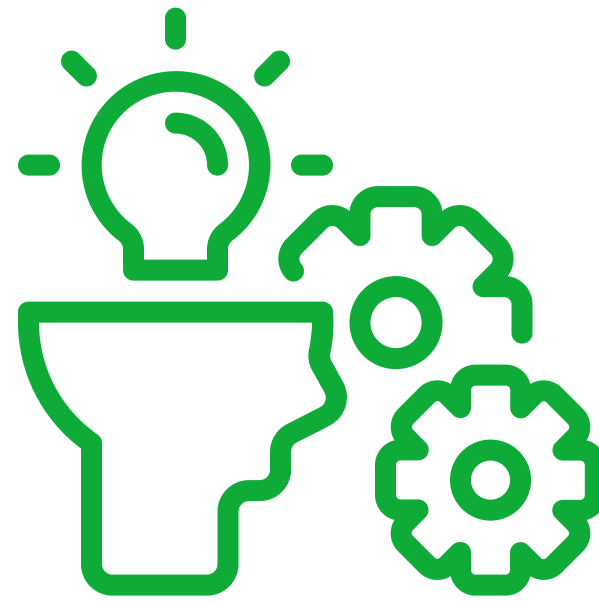
- Mengurangi rasio barang yang terlambat menjadi dibawah 30% dari total pengiriman

Goals

- Mencari faktor yang dapat mempengaruhi keterlambatan pengiriman
- Membentuk model Machine Learning (ML) untuk memprediksi keterlambatan pengiriman
- Memberikan rekomendasi yang terkait pengiriman barang

Business Metrics

- Main metric: Delivery on time ratio
- Supporting metric: Customer satisfaction



Process

Flow Modeling



Describe the Dataset

Features and Target

1

2

3

Data Exploration

ID

Prior_purchases

Warehouse_block

Product_importance

Mode_of_shipment

Gender

Customer_care_calls

Discount_offered

Customer_rating

Weight_in_gms

Cost_of_the_product

Reached.on.time_Y.N

- Jumlah baris data : 10999 baris
- 8 kolom tipe numerik (termasuk target) & 4 kolom tipe object/string
- Nama kolom **Reached.on.Time_Y.N** diubah menjadi **Late_delivery** untuk membuat penamaan lebih intuitif
- Kolom **Late_delivery** dijadikan sebagai kolom **target**
- **Tidak ada null values, setiap ID unik**

Exploratory Data Analysis (EDA)

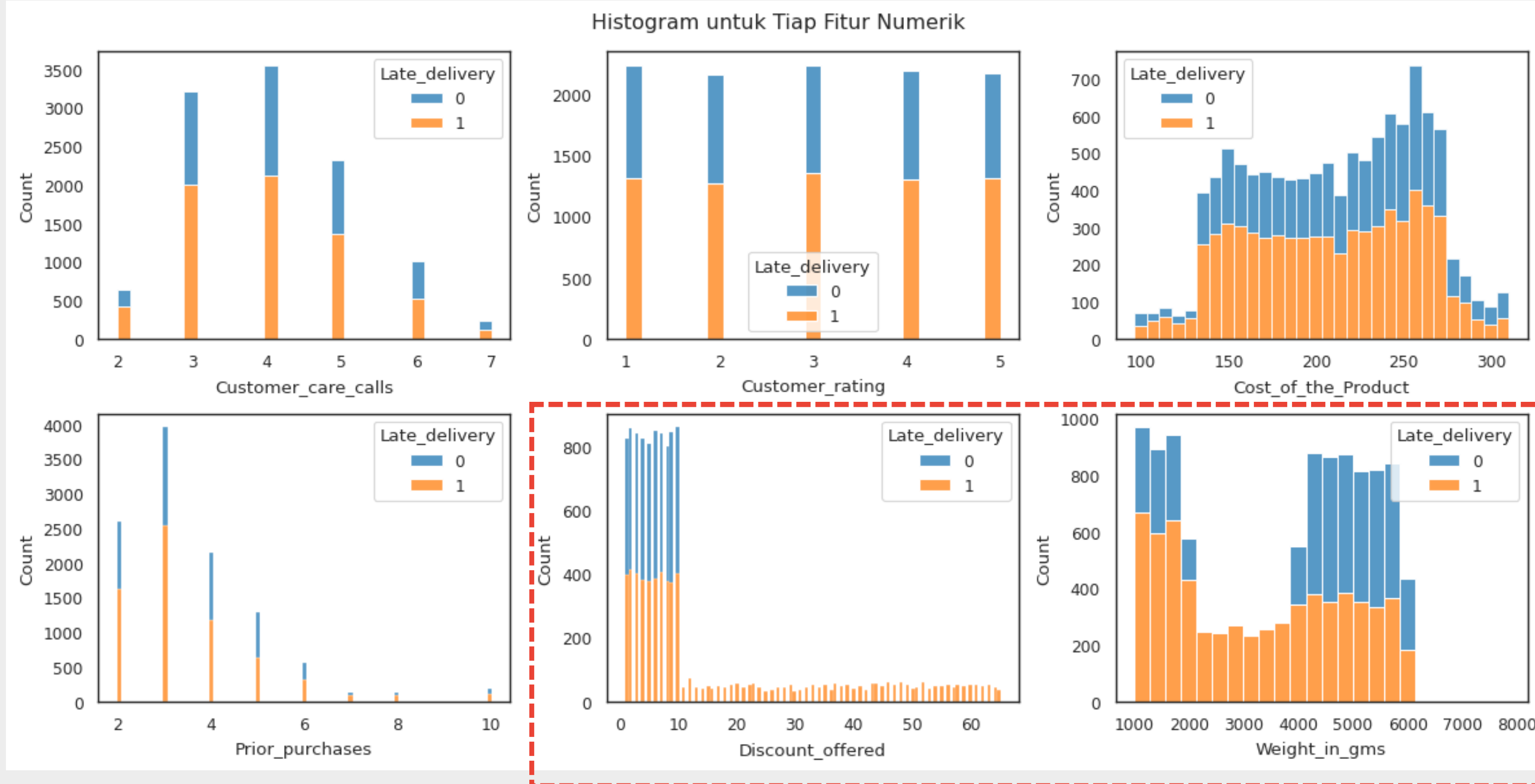
Distributions with Target as Hue

1

2

3

Data Exploration



- Barang yang diberikan diskon lebih dari 10% tidak ada yang pengirimannya tepat waktu
- Barang yang pengirimannya tepat waktu hanya terjadi pada barang yang beratnya ada di bawah 2000 gram atau di antara 4000–6000 gram.

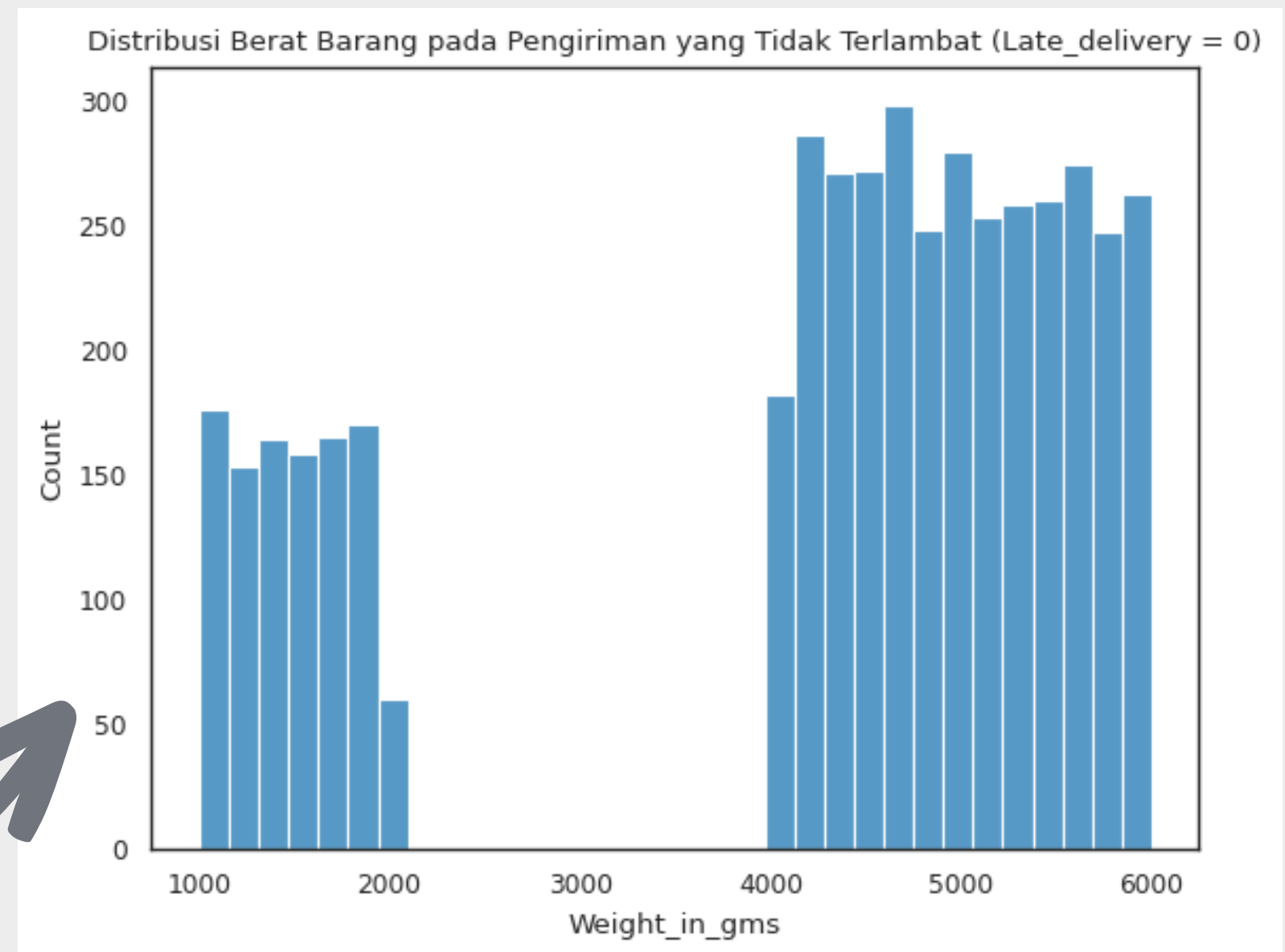
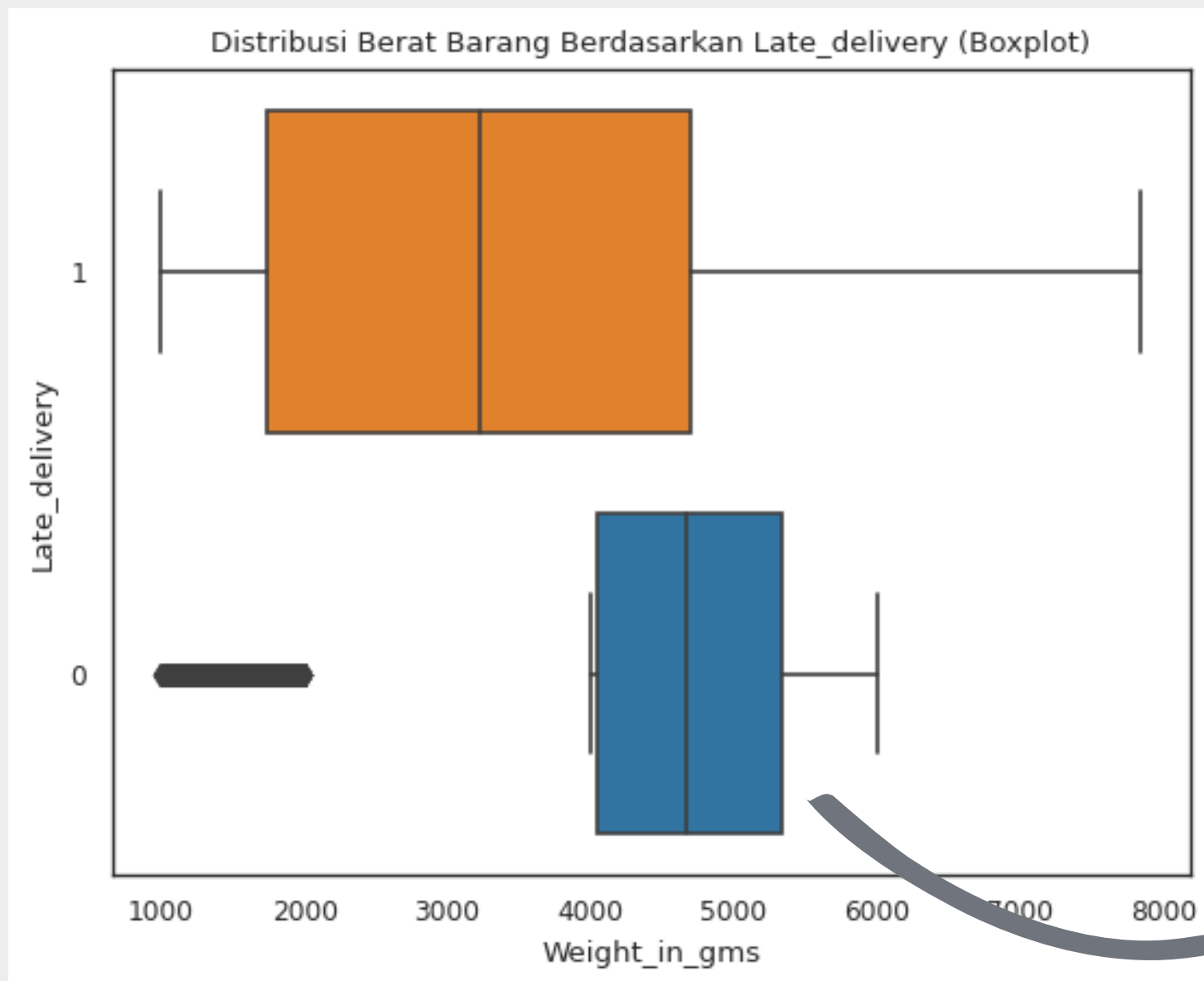
Product Weight on Items Delivered Late

1

2

3

Data Exploration



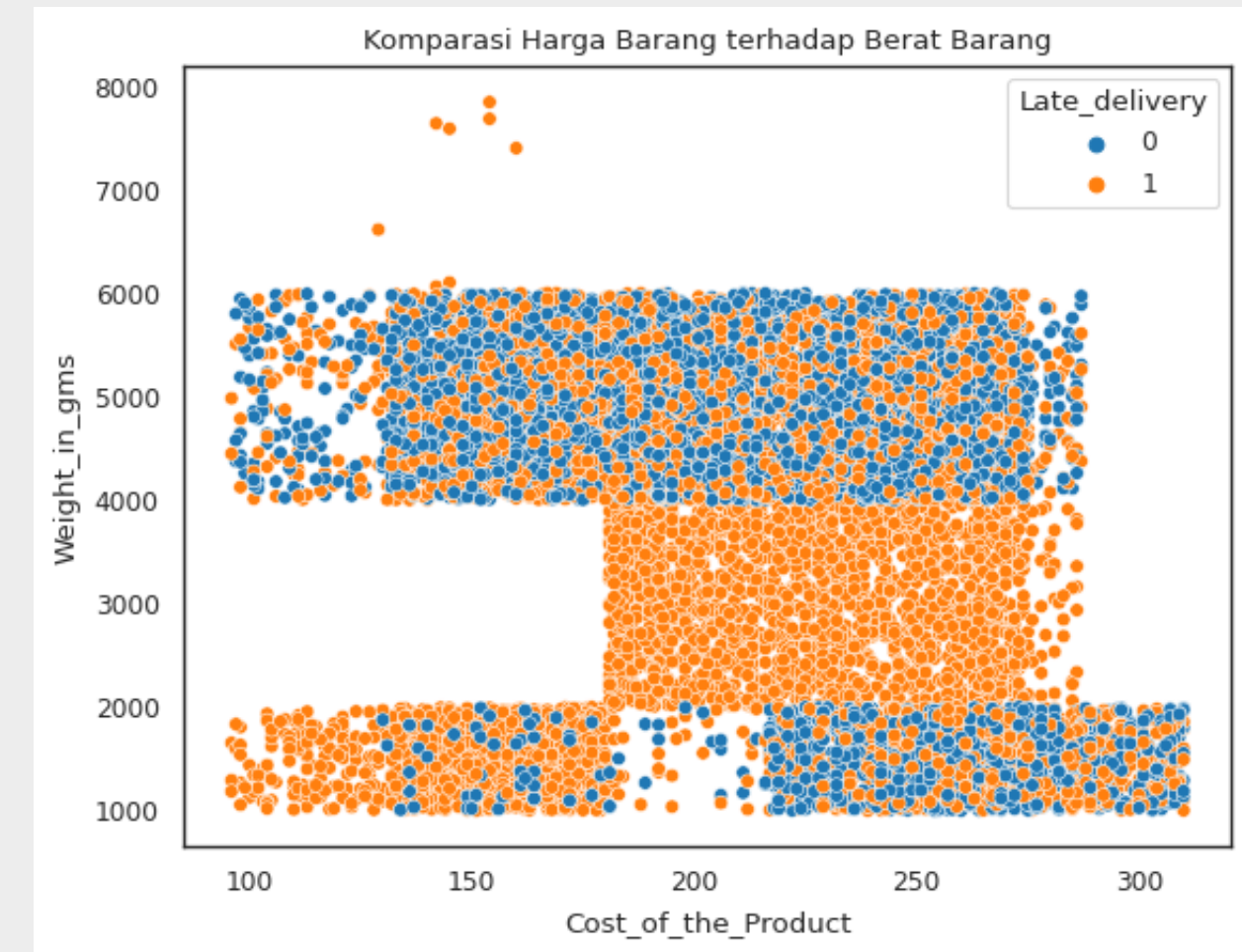
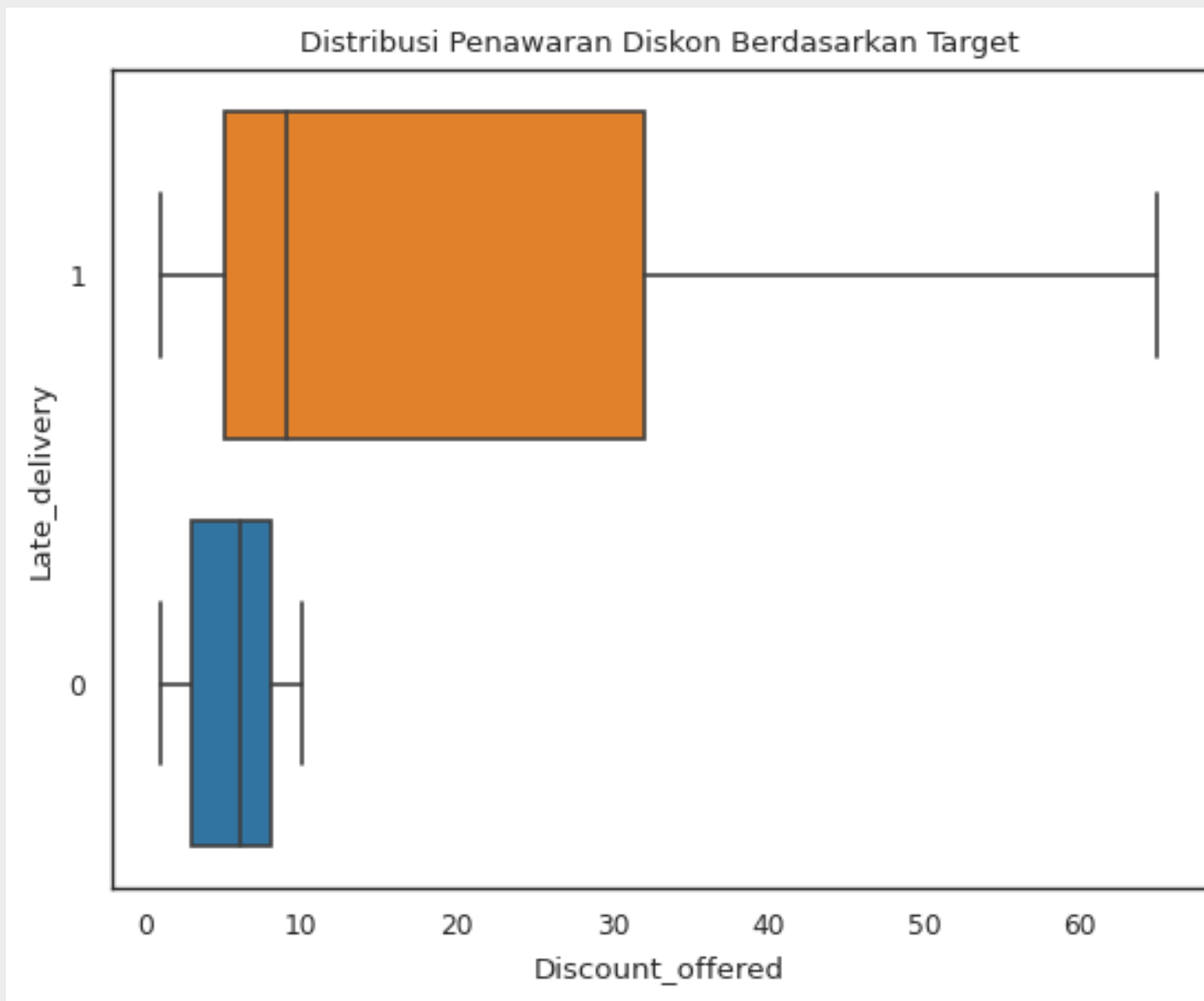
Discount Offered & Cost of Product

1

2

3

Data Exploration

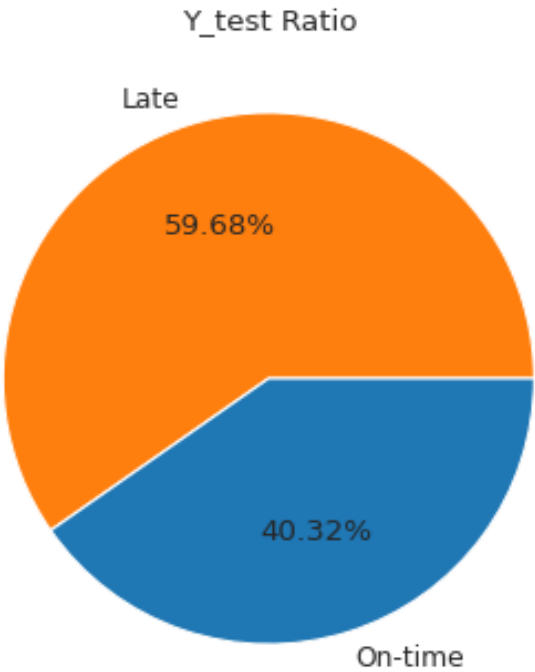
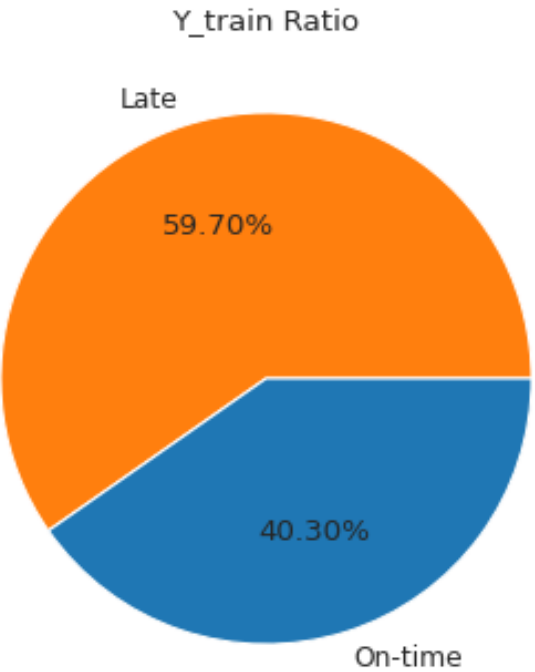
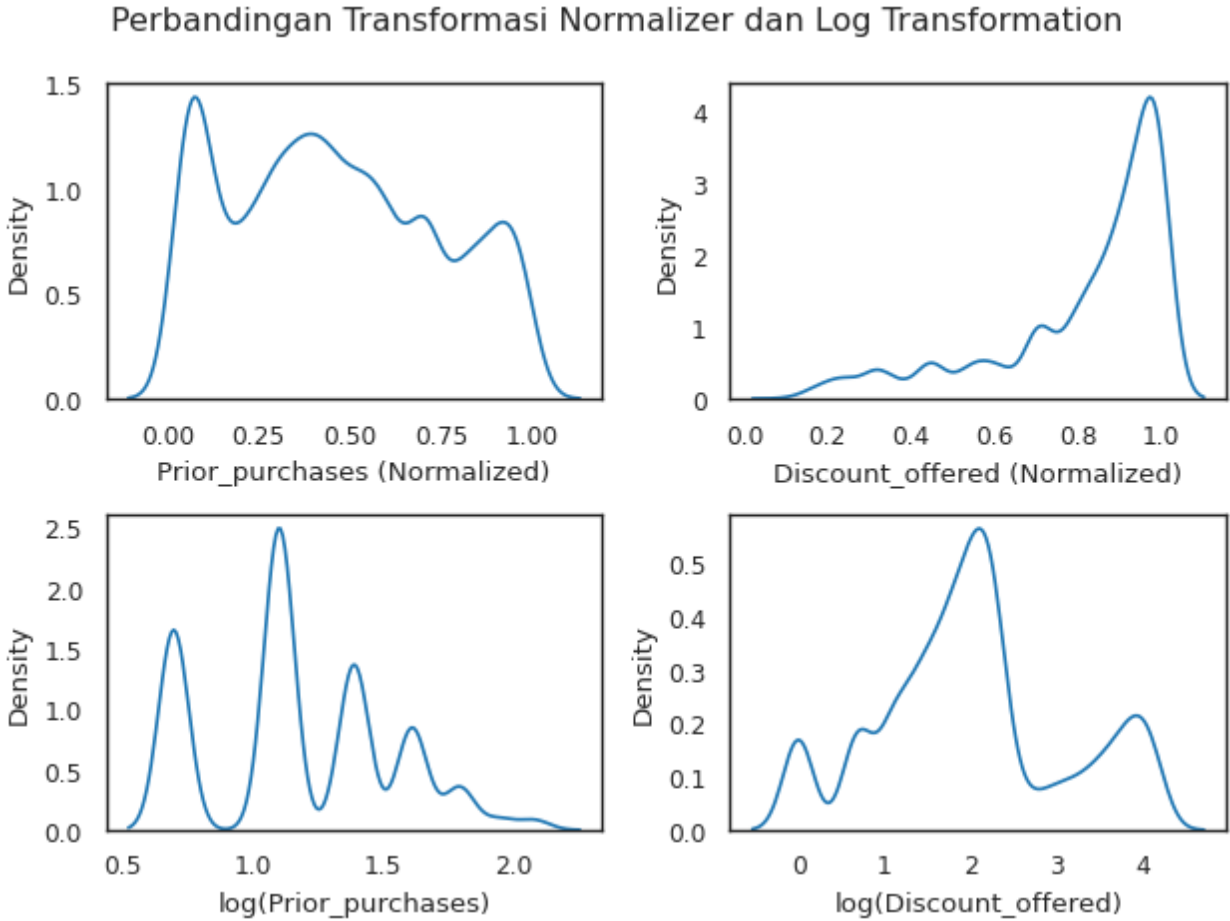


Pada pembentukan model, barang dipisah berdasarkan tiga kategori baru: **apakah barang diberikan diskon di atas 10% atau tidak**, kategori **berat barang**, dan **kategori harga**

Data Preprocessing (Pipeline)

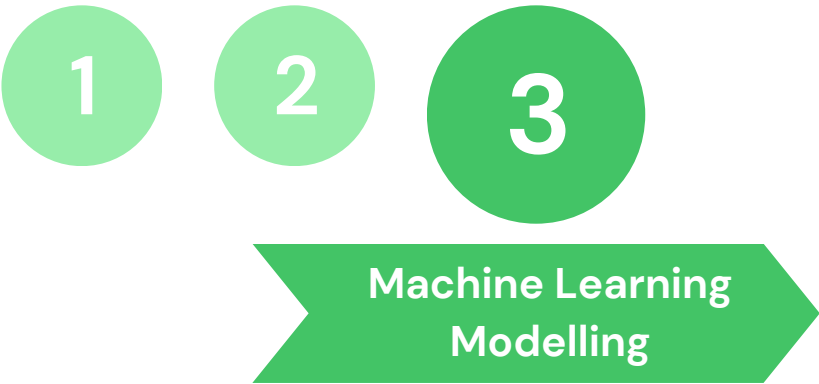


Split Data Train & Data Test	Missing Values & Duplicate Data	Outliers	Feature Transformation	Feature Encoding	Handle Imbalance
80 : 20	None	z-score	MinMaxScaler	Ordinal & OHE	None



Model

Cross Validation and Tuning



Metrik yang digunakan : Recall

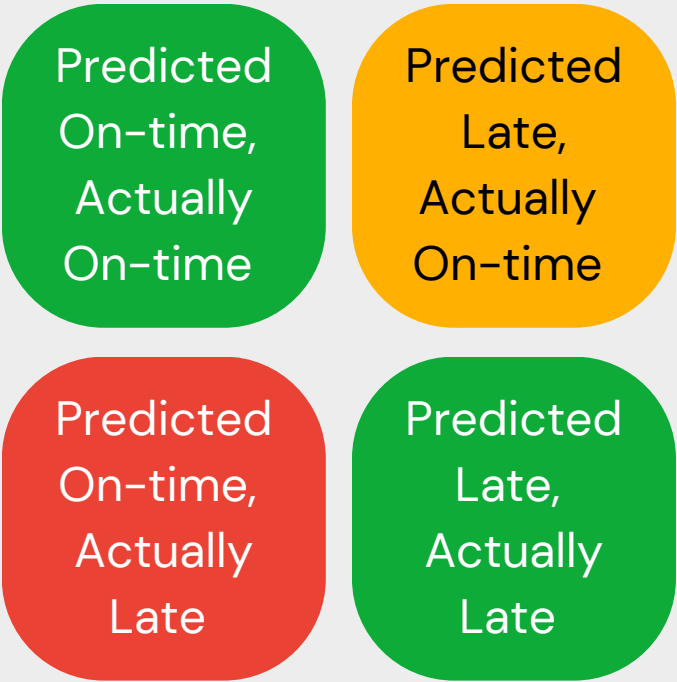
Model	Training Accuracy	CV Accuracy (mean)	CV Accuracy (std)	Training Precision	CV Precision (mean)	CV Precision (std)	Training Recall	CV Recall (mean)	CV Recall (std)
Decision Tree	1.000	0.644	0.006	1.000	0.702	0.004	1.000	0.703	0.014
Random Forest	1.000	0.653	0.007	1.000	0.753	0.003	1.000	0.624	0.013
Logistic Regression	0.679	0.675	0.011	0.841	0.835	0.021	0.570	0.568	0.024
KNN	0.777	0.646	0.009	0.836	0.718	0.009	0.779	0.671	0.010
XGBoost	0.891	0.654	0.016	0.950	0.736	0.022	0.863	0.658	0.017
SVM	0.692	0.681	0.011	0.963	0.931	0.021	0.503	0.504	0.020

Setelah Hyperparameter Tuning Decision Tree:

```
Recall score data TRAIN : 64.343 %
Recall score data TEST  : 62.833 %

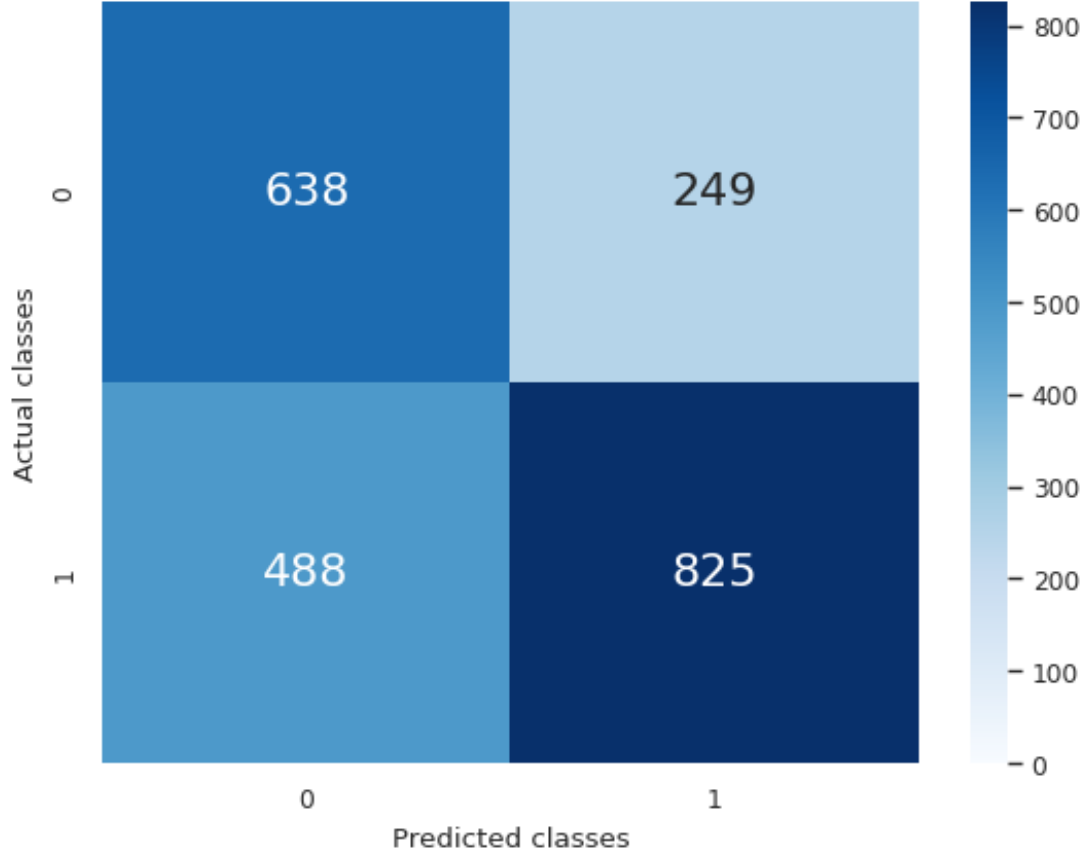
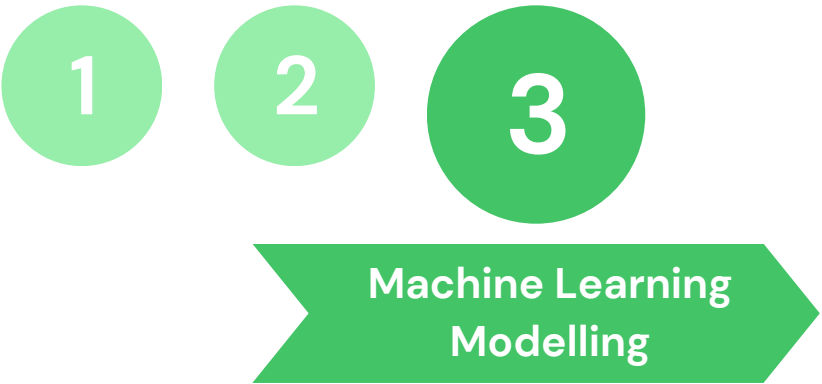
Other Scores (based on test data):

Accuracy score : 66.5 %
F-1 score      : 69.124 %
Precision score : 76.816 %
```

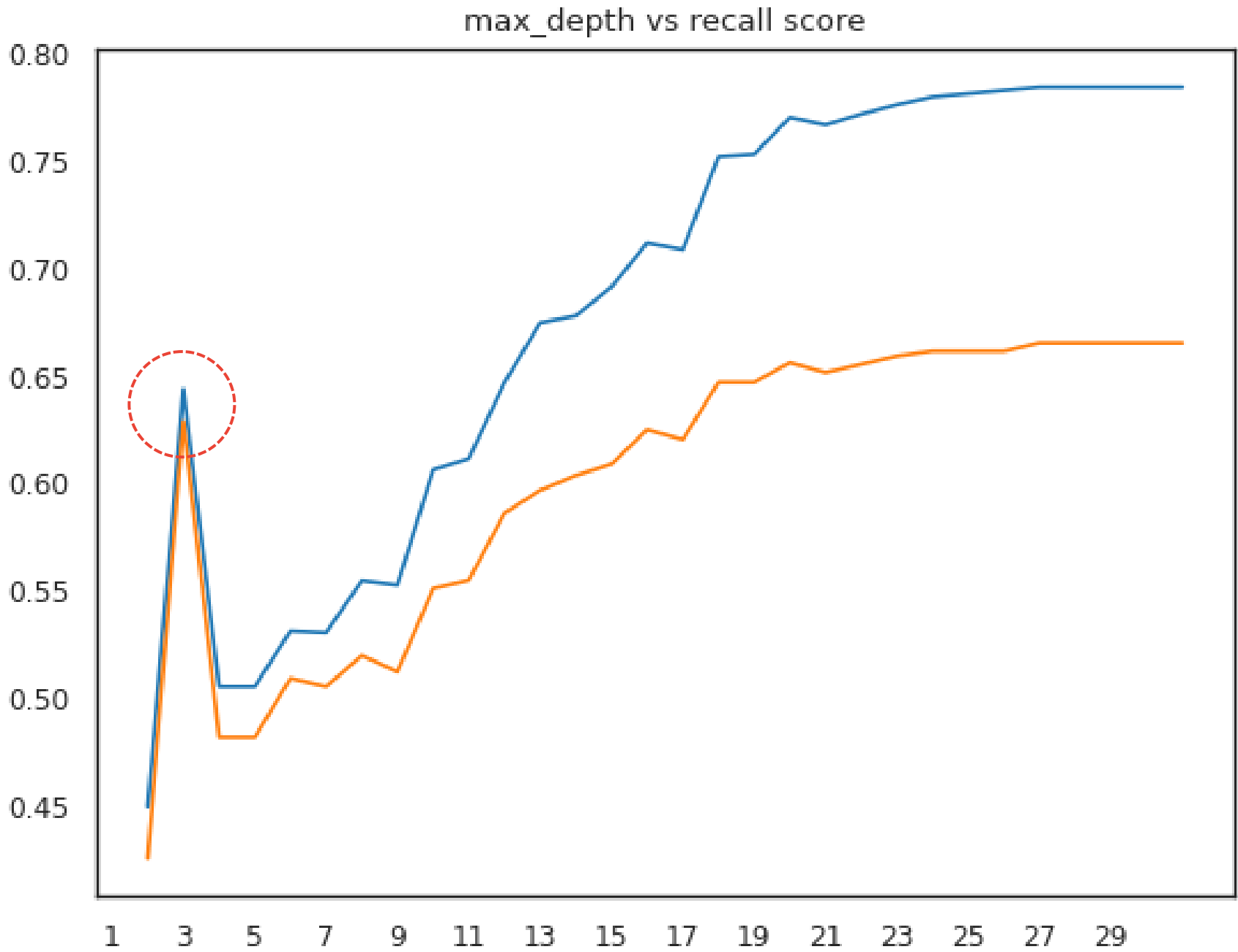


Model Results

Recall Score Evaluation



Total Pengiriman	10999
Persentase Keterlambatan	59,67%
Total Pengiriman Terlambat	6563
Recall	63%
Total Pengiriman Terlambat yang Bisa Diprediksi	4135



Model

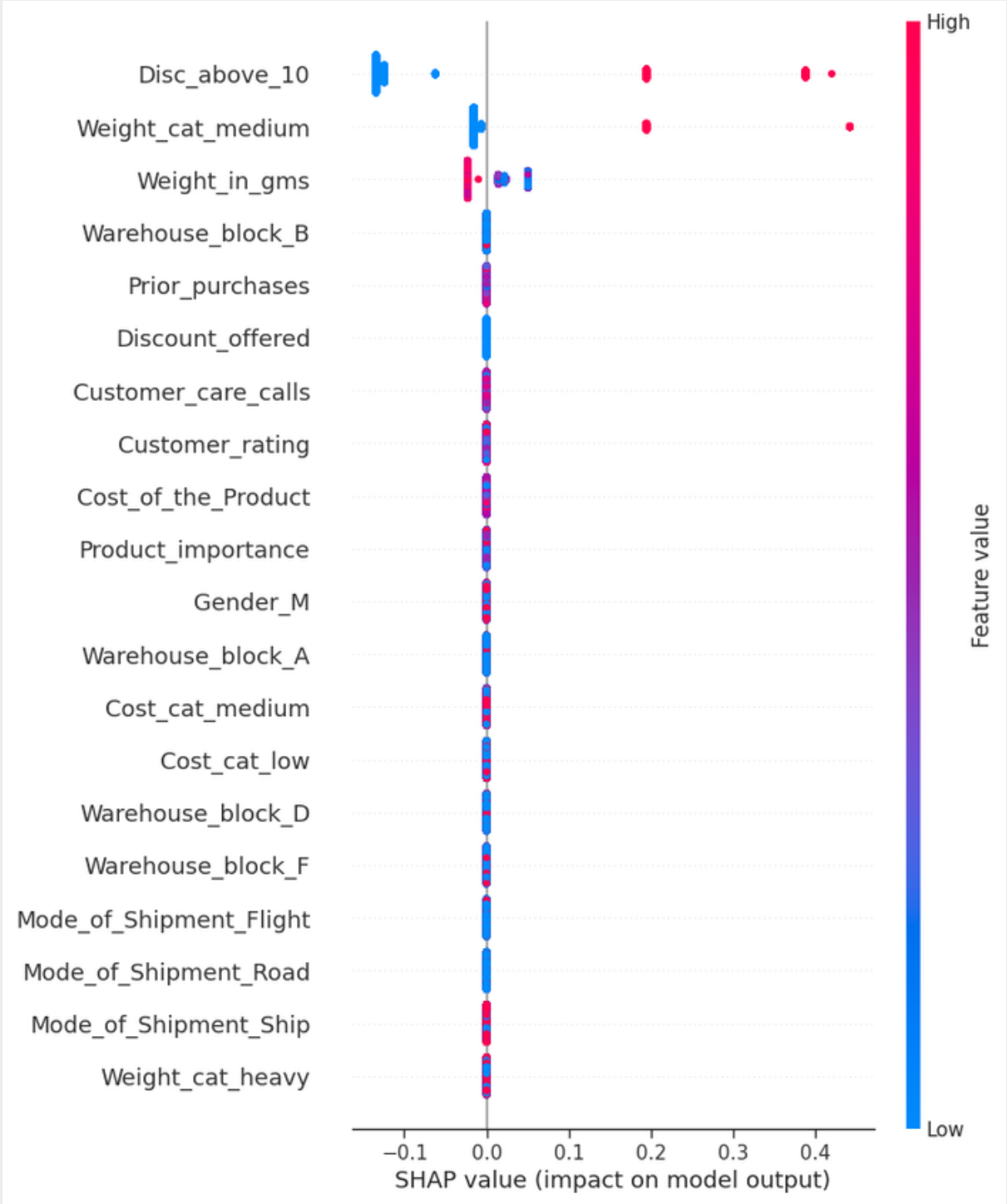
Feature Importance



Machine Learning
Modelling

Besar Diskon

Berat Barang



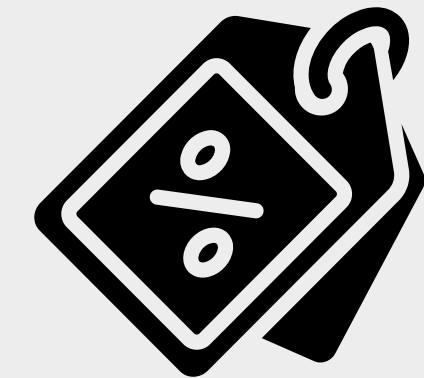


Insights & Recommendations

Insights



- Pengiriman barang tepat waktu hanya terjadi pada barang yang beratnya ada di kategori ringan (1000–2000 gram) dan berat (4000–6000 gram)
- Pengiriman barang tepat waktu hanya terjadi pada barang yang memiliki nilai diskon yang ditawarkan sebesar 10% atau lebih kecil
- Fitur yang paling penting pada model yang dibentuk adalah fitur-fitur terkait besar diskon yang ditawarkan dan berat barang



"Correlation does not imply causation"

Insights, Compared to Literatures



What Causes Delay In Delivery?



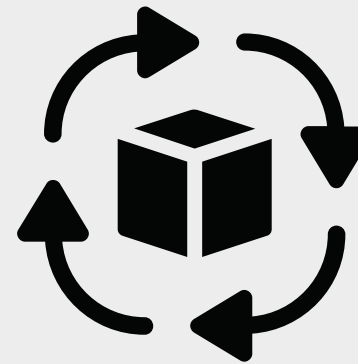
Last-Mile Logistics

(Source: Deloitte, "Last-Mile Delivery: The Future of eCommerce," 2017)



Peak Demand Periods

(Source: Parcel Monitor, "4 Strategies to Prevent Late Deliveries During Peak Season," 2018)



Supply Chain Disruptions

(Source: McKinsey & Company, "Supply-chain resilience: the missing link in disaster recovery," 2013)



Customer Address Issue

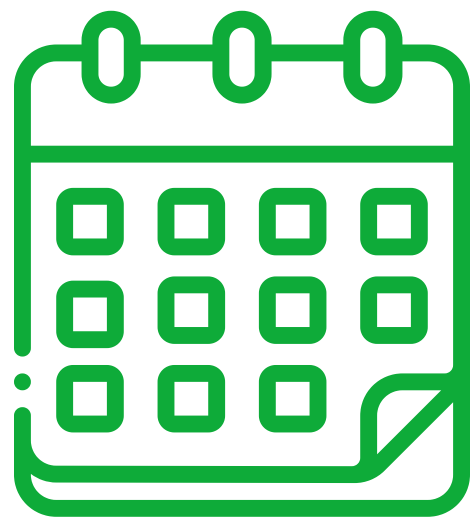
(Source: Multichannel Merchant, "3 Common Reasons for Late Deliveries and How to Prevent Them," 2019)



Warehouse Operations

(Source: Supply Chain Dive, "5 warehouse management issues causing shipping delays," 2019)

Recommendations



Pengumpulan Data Tambahan

- Waktu Keterlambatan
- Tanggal Pengiriman
- Wilayah Pengiriman
- Jenis/Jasa Kurir
- Jenis promo/diskon yang ditawarkan
- Tipe/kategori barang elektronik yang dijual

Deployment Model ke Sistem E-Commerce

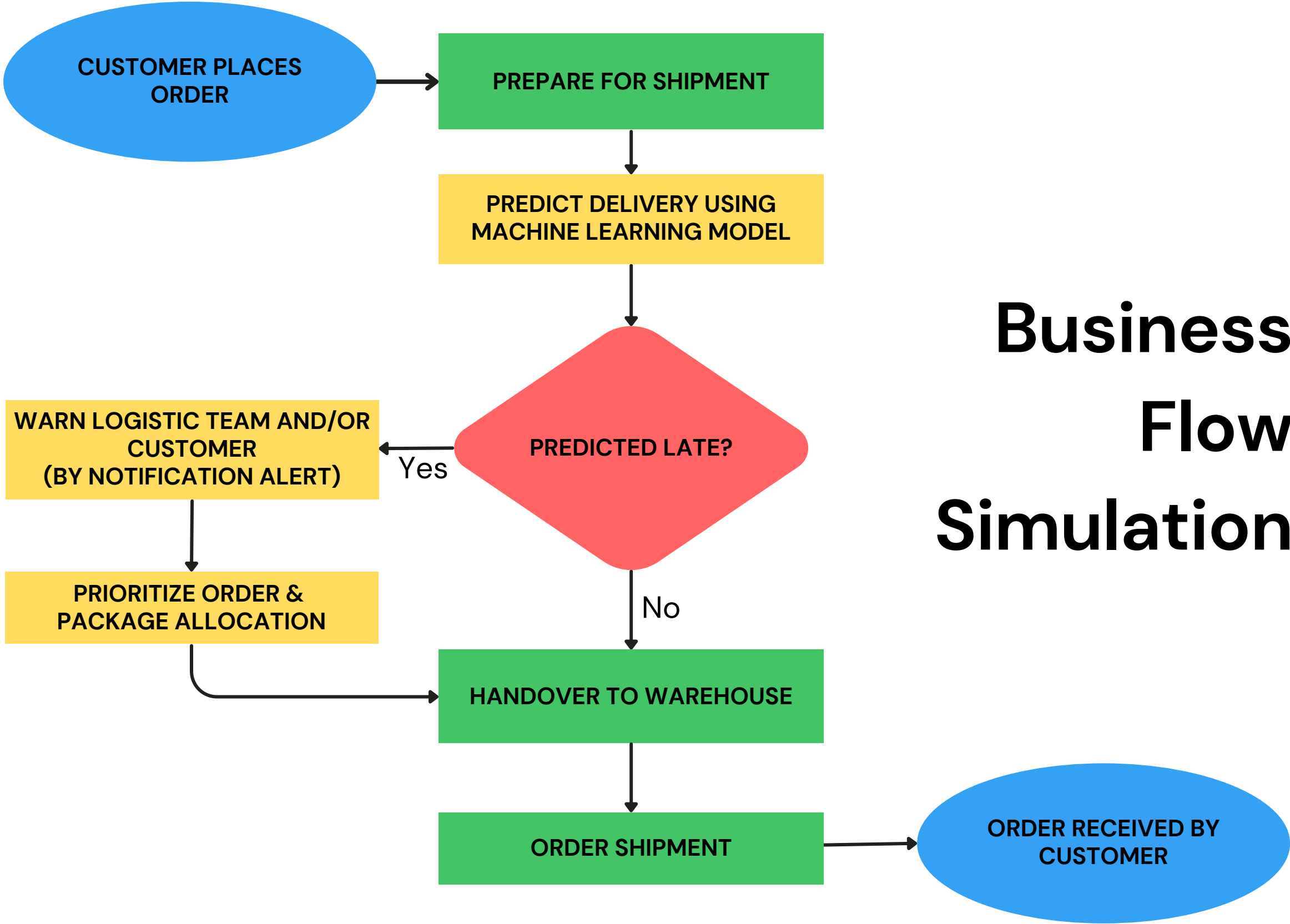
- Model dapat digunakan untuk mendeteksi keterlambatan sebelum kejadian
- Penanganan dapat dilakukan untuk barang yang diprediksi akan terlambat
- Bentuk penanganan: pelanggan dan tim logistik akan diberikan notifikasi, dan pengiriman dapat diprioritaskan



Before Model Deployment



After Model Deployment



Business Flow Simulation

Thank you for
participating. Have a
great day ahead.

We're done!