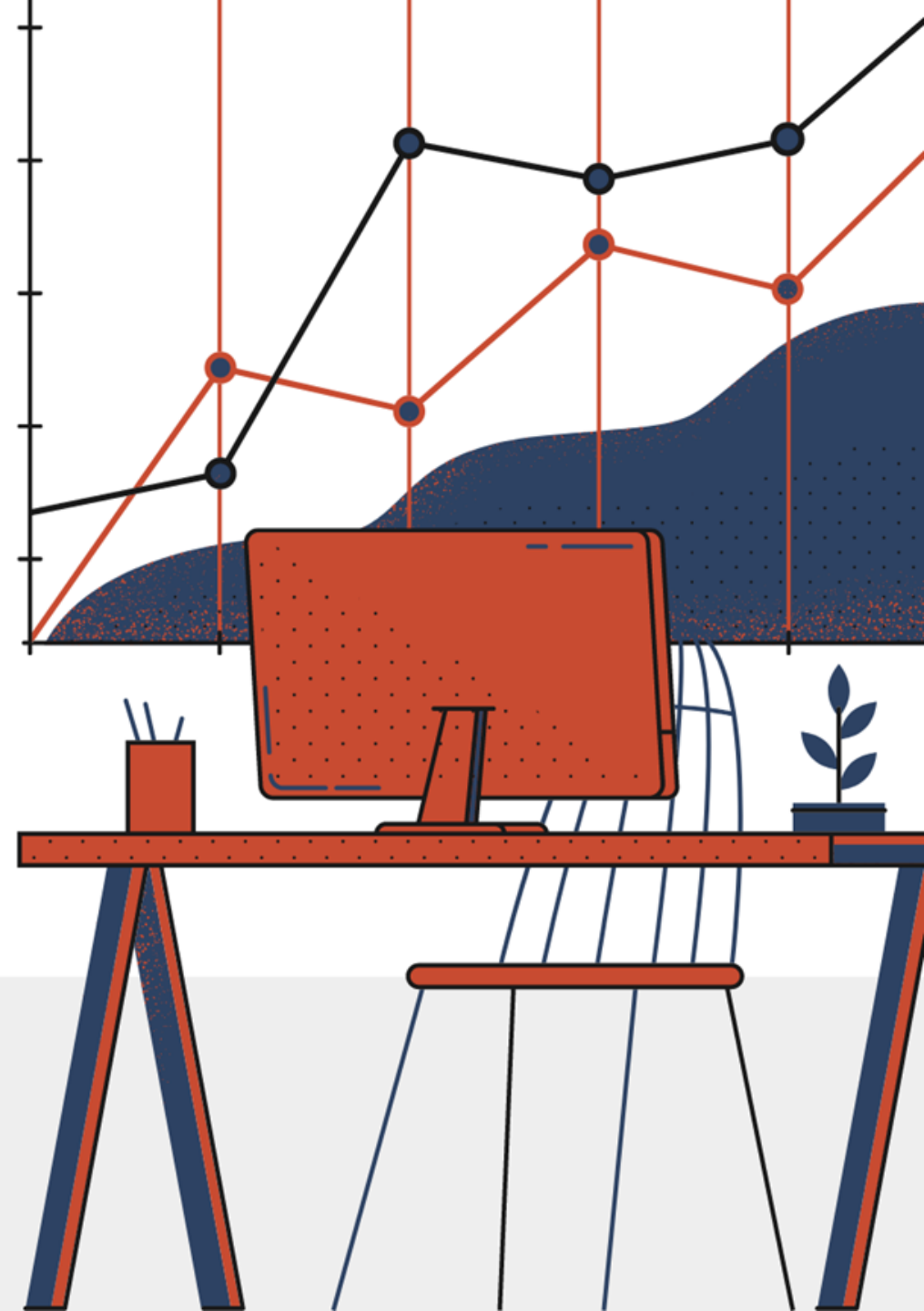


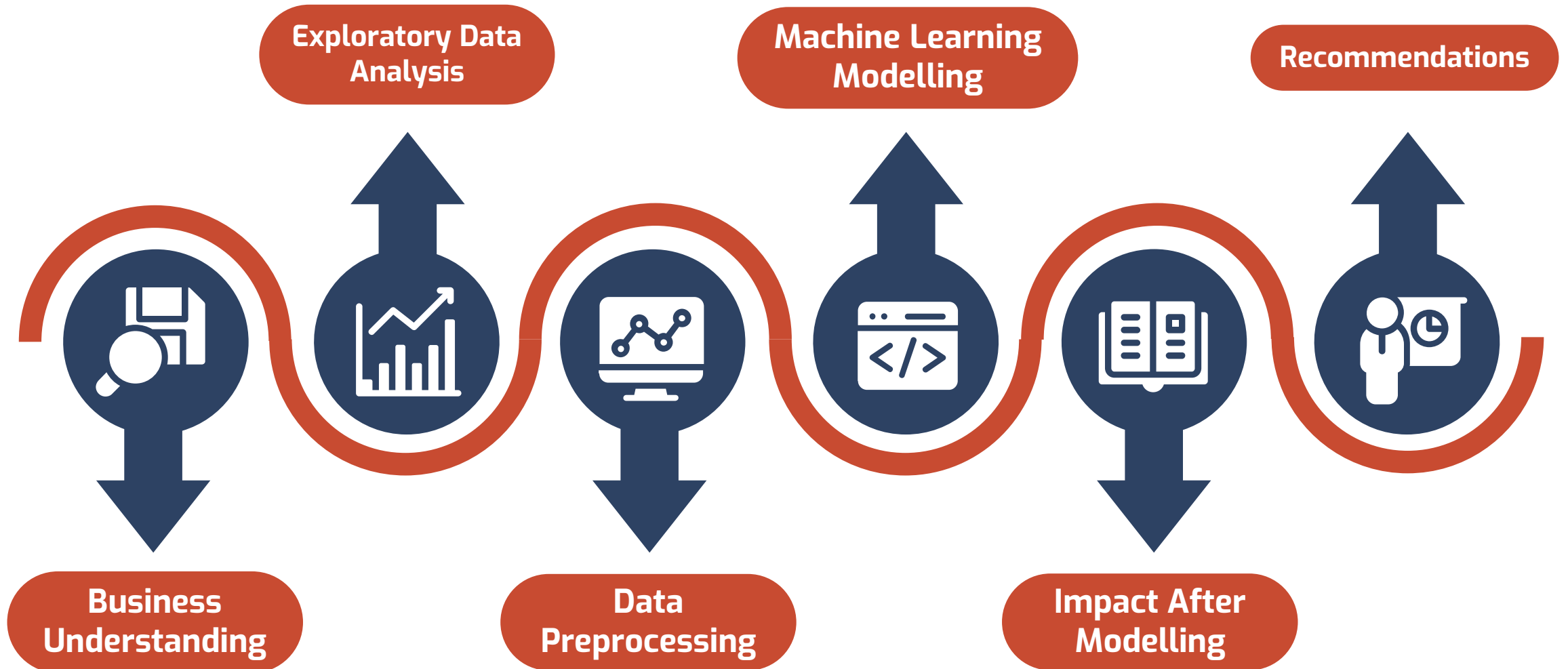
# Shipping Arrival Prediction

`DATASET{'SAT_SET'}`

- Nur Cahyanti
- Utlia Rahma
- Indra Laksana
- Handika
- Fajar Nurdiono
- Refanie Fajrina



# OUTLINE





## Data Science Team



Refanie



Indra



Yanti



Fajar



Rahma Handika

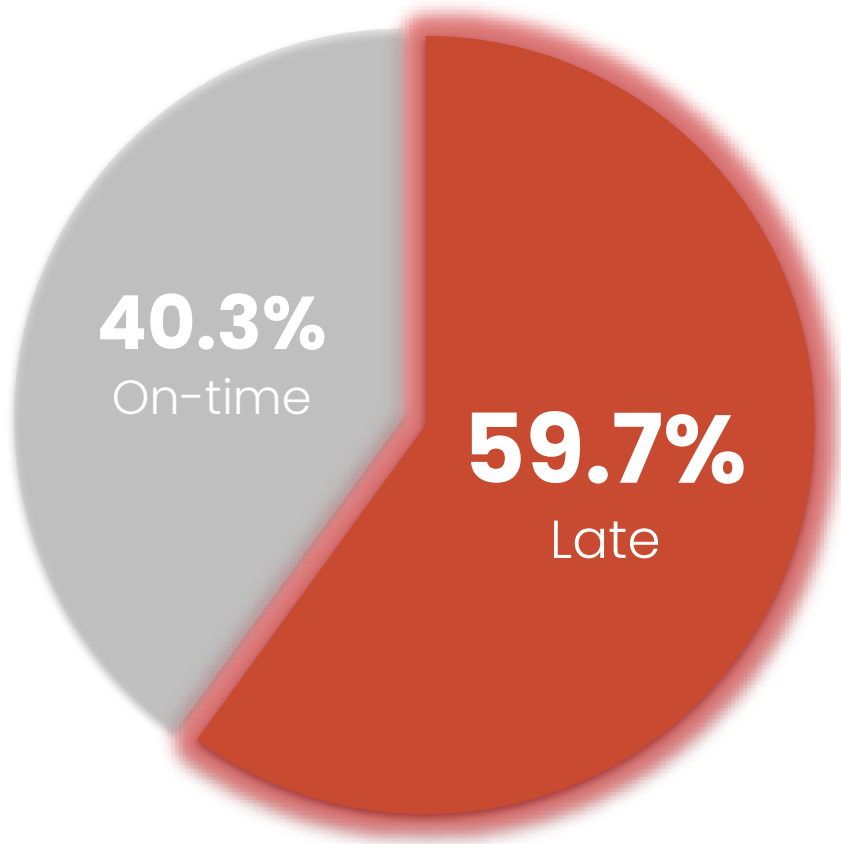


E-commerce Elektronik



Memiliki ekspedisi sendiri

## PERSENTASE KETERLAMBATAN PENGIRIMAN

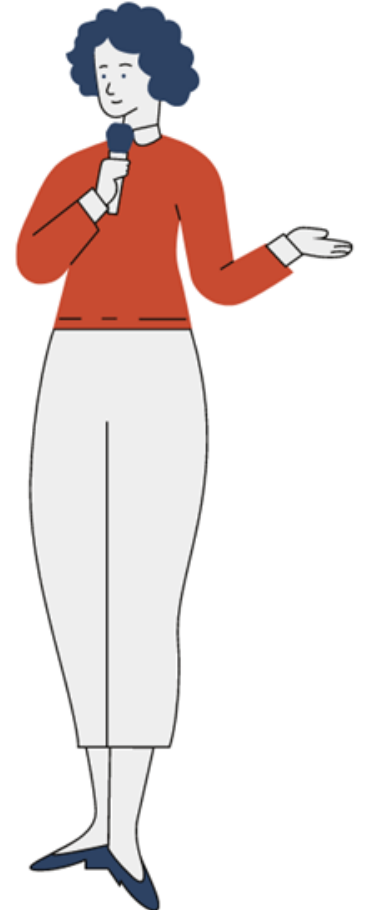


**69.7%** customer berpotensi tidak akan kembali membeli di suatu toko apabila produk yang mereka beli **datang terlambat namun tidak diberi tahu.**

**shopify.co.id, 2020**

**56%** konsumen berbelanja di toko lain karena toko langganannya **tidak dapat mengirimkan produk tepat waktu.**

**shopify.co.id, 2020**





## Goals

Mengurangi paket datang terlambat



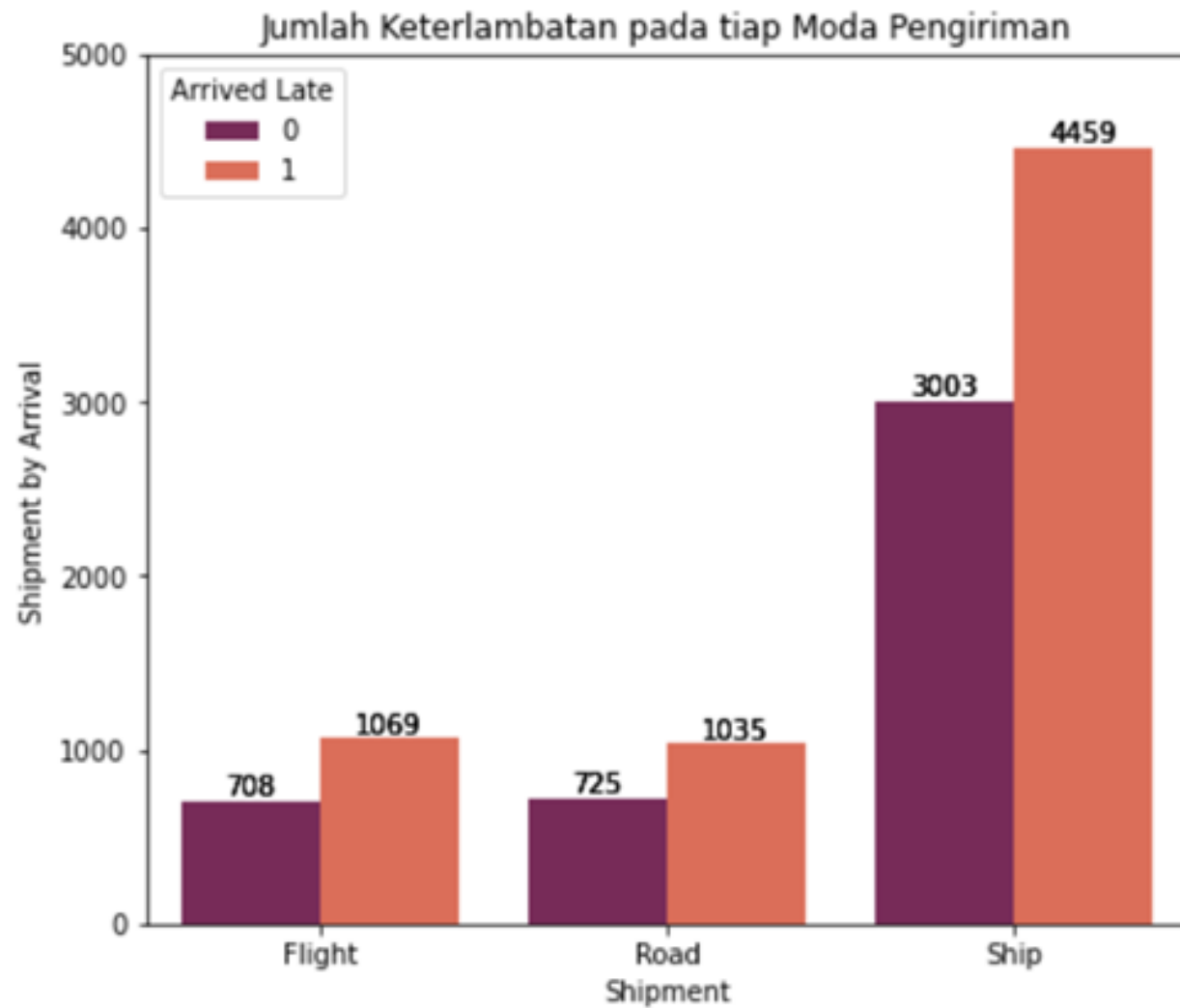
## Objectives

Membuat model machine learning untuk memprediksi paket akan datang terlambat atau tidak.

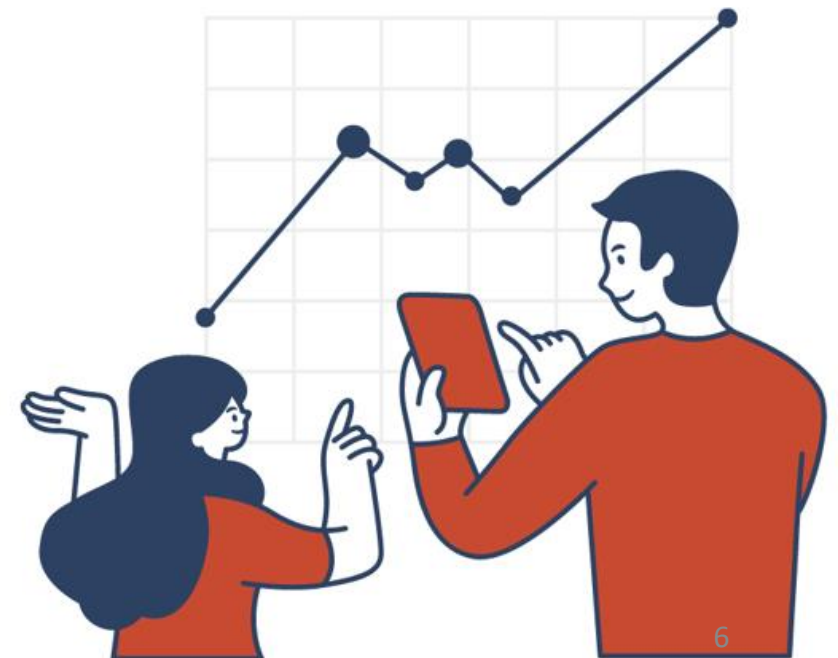


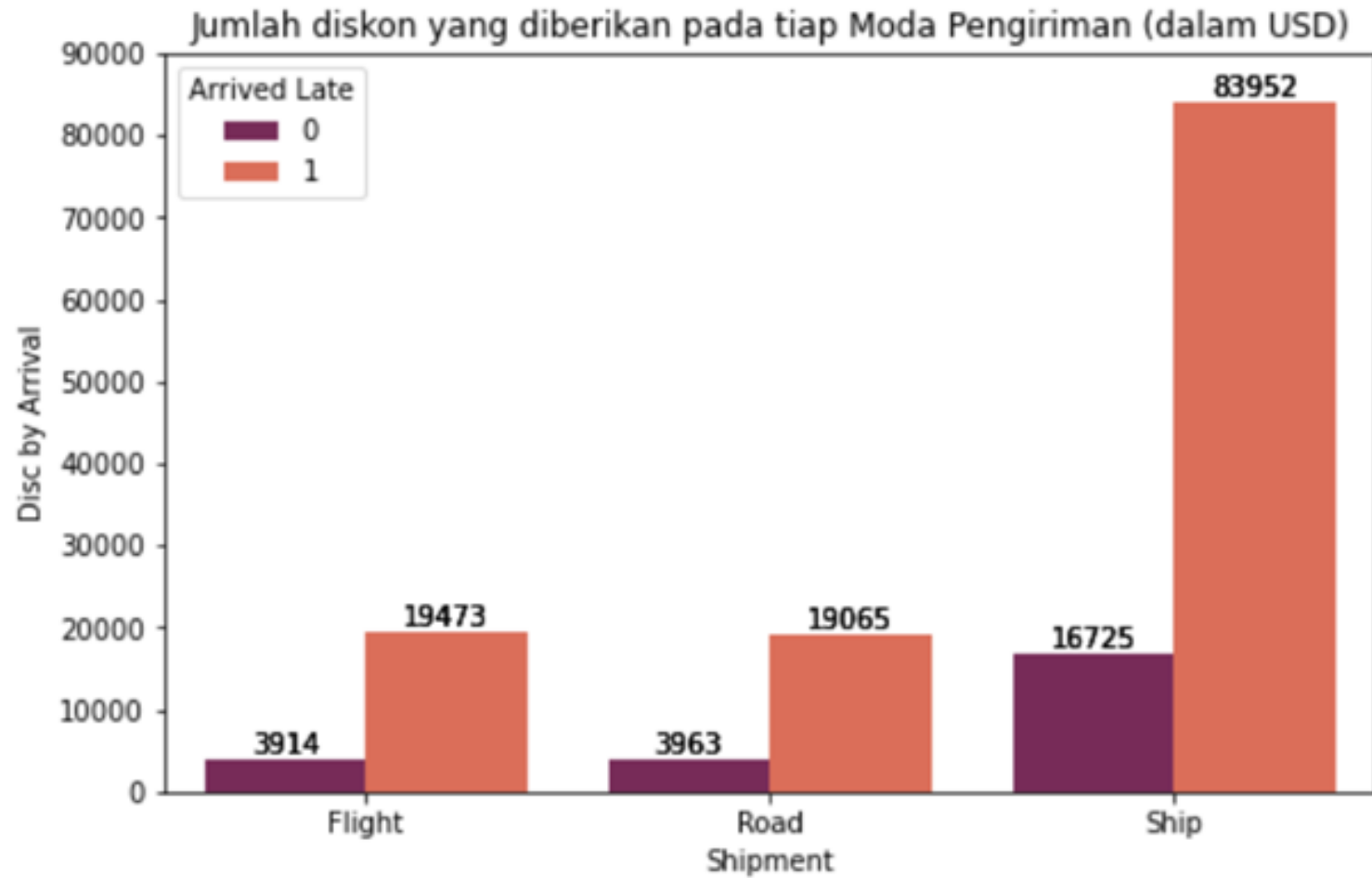
## Business Metrics

Persentase paket sampai tepat waktu



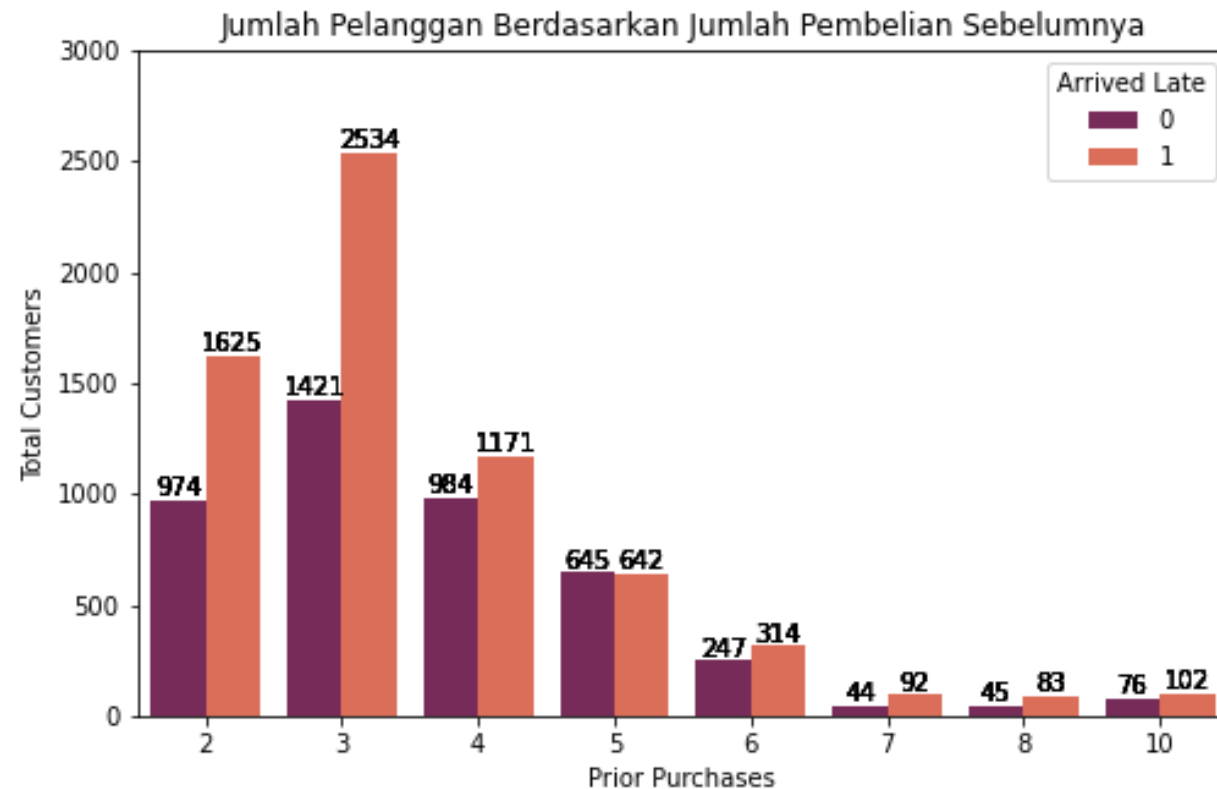
Mayoritas produk dikirim menggunakan kapal, terlepas dari produk datang terlambat atau tidak





- Produk yang diberikan diskon besar cenderung datang terlambat
- Produk yang diberikan diskon < \$10 cenderung datang tepat waktu

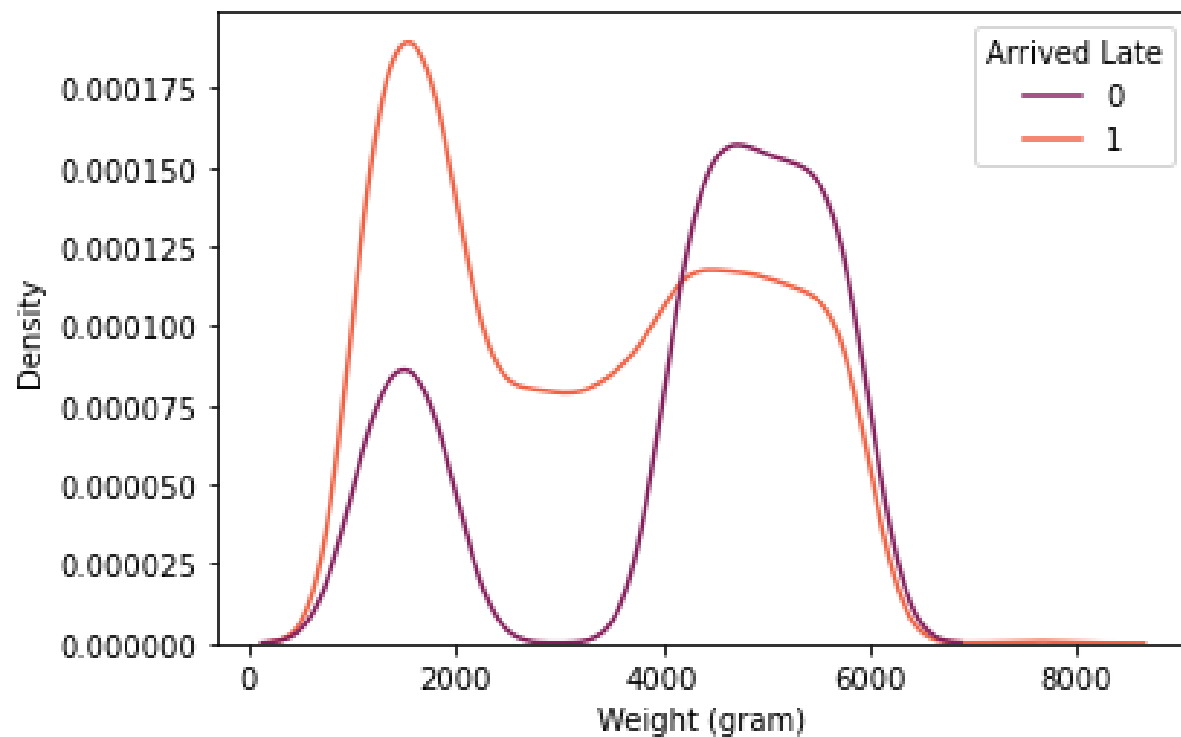




Mayoritas pelanggan telah melakukan pembelian sebanyak 3 kali sebelumnya, terlepas dari produk datang terlambat atau tidak







- Produk yang memiliki berat **4000 sampai 5000**, cenderung datang tepat waktu
- Produk yang terlambat cenderung memiliki berat lebih ringan

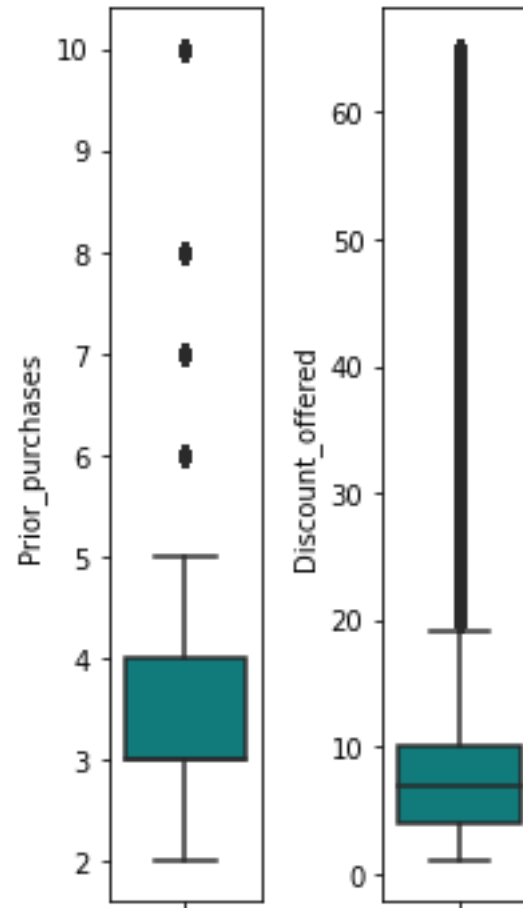


## Handling Outliers

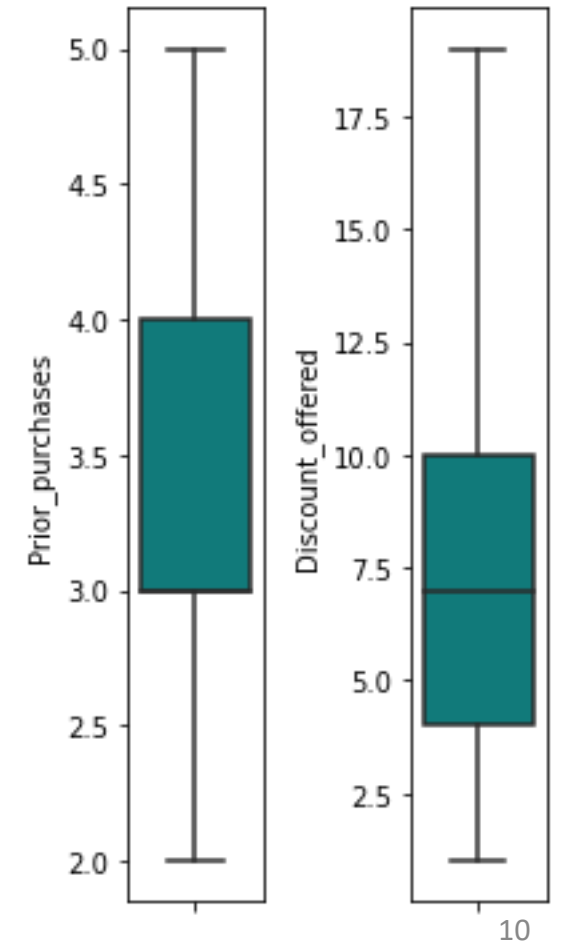
Mengganti outliers pada feature **Discount\_offered** dan **Prior\_purchase** dengan nilai  $Q_3 + 1,5(IQR)$

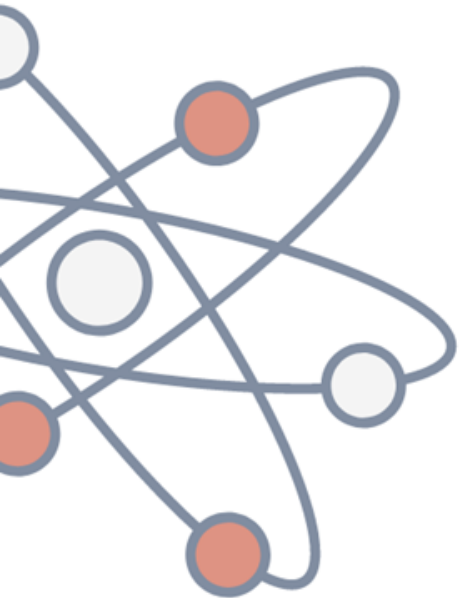


**BEFORE**



**AFTER**





## Transformasi (Normalization)

- Discount\_offered
- Cost\_of\_the\_Product
- Prior\_purchase
- Weight\_in\_gms



## Encoding

### Onehot Encoding

- Product\_importance
- Gender

### Label Encoding

- Warehouse\_block
- Mode\_of\_Shipment





## DROP KOLOM

'ID', 'Customer\_care\_calls', dan 'Customer\_rating'



### Before Preprocessing



- 11 Feature
- 1 target



### After Preprocessing



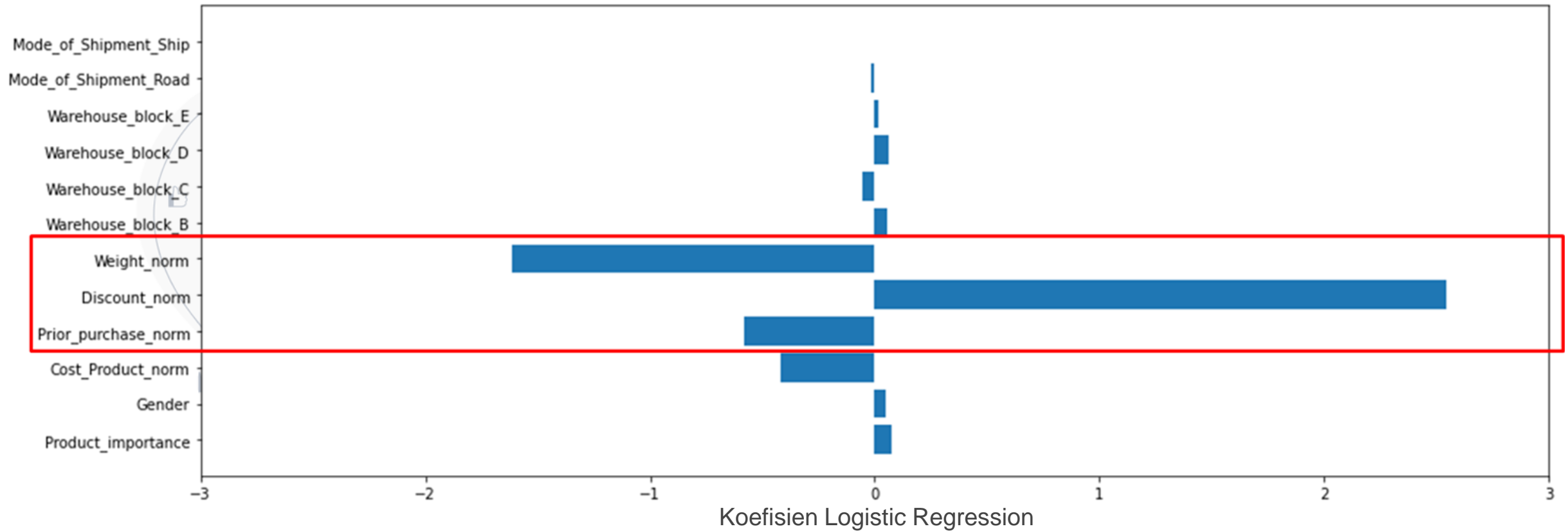
- 12 Feature
- 1 target



## SPLIT DATA

**Train Data = 70 | Test Data = 30**

Metric	LOGISTIC REGRESSION	K-NEAREST NEIGHBOR	DECISION TREE	RANDOM FOREST	XG BOOST	GRADIENT BOOSTING
Accuracy	0.6376	0.6370	0.6394	0.6645	0.6515	0.6685
Precision	0.6924	0.7194	0.7430	0.7968	0.7310	0.7760
Recall	0.7064	0.6420	0.6049	0.5876	0.6582	0.6247
F1 – Score	0.6993	0.6785	0.6669	0.6764	0.6927	0.6922
ROC-AUC (test)	0.7208	0.7047	0.7123	0.7315	0.7341	0.7420
ROC-AUC (train)	0.7209	0.7919	0.7803	0.8391	0.9802	0.7488



### Asumsi Pengaruh Feature Importance terhadap Hasil Prediksi

Perusahaan menerapkan event jangka pendek yang cukup masif untuk meningkatkan penjualan, adapun program event dapat diikuti dengan syarat pelanggan sbb:

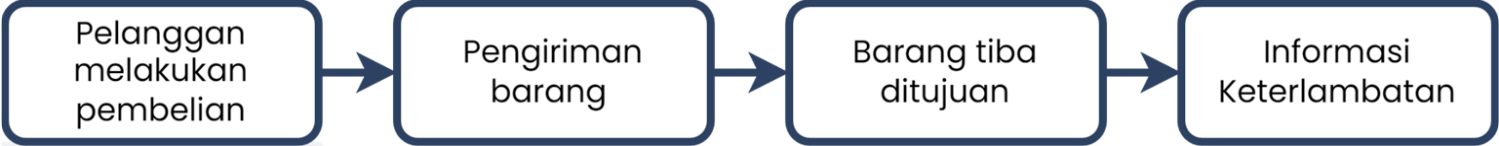
1. Setiap pelanggan mendapatkan diskon khusus;
2. Program ini khusus barang dengan berat < 5000 gram;
3. Tidak berlaku untuk pelanggan lama

## Rekomendasi:

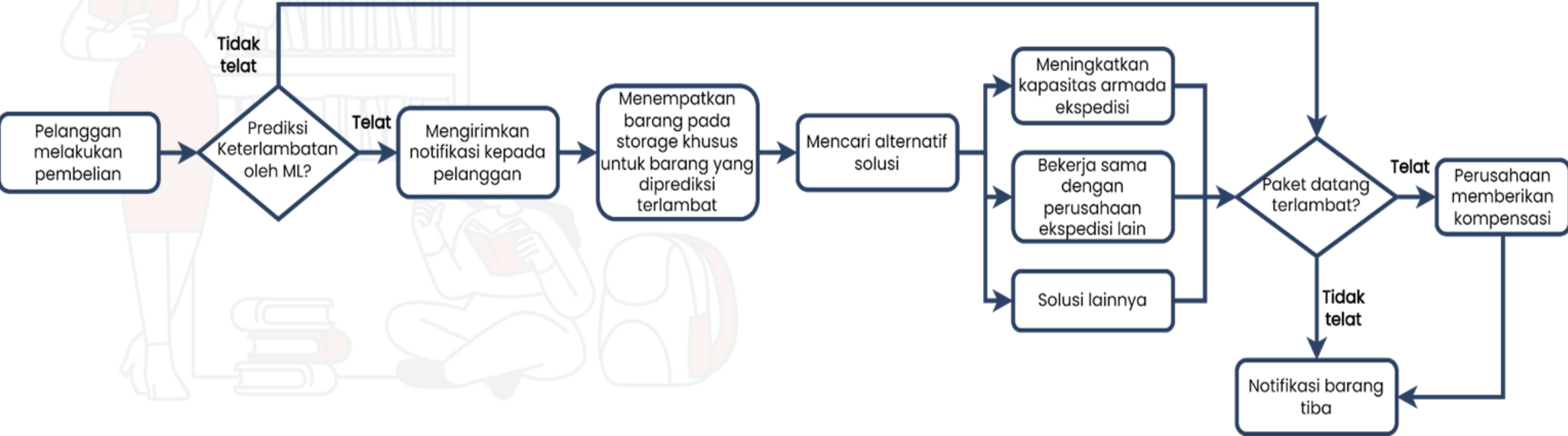
1. Perencanaan program diskon yang matang.
  - limitasi pada jumlah pelanggan
  - limitasi pada berat barang
  - personalisasi diskon dalam tier member
1. Kerja sama dengan pihak ketiga (jasa pengiriman) untuk meningkatkan kapasitas pengiriman saat peak season/program berlangsung.
2. Mengoptimalkan utilisasi armada pengiriman.



Sebelum Perusahaan Menerapkan Machine Learning

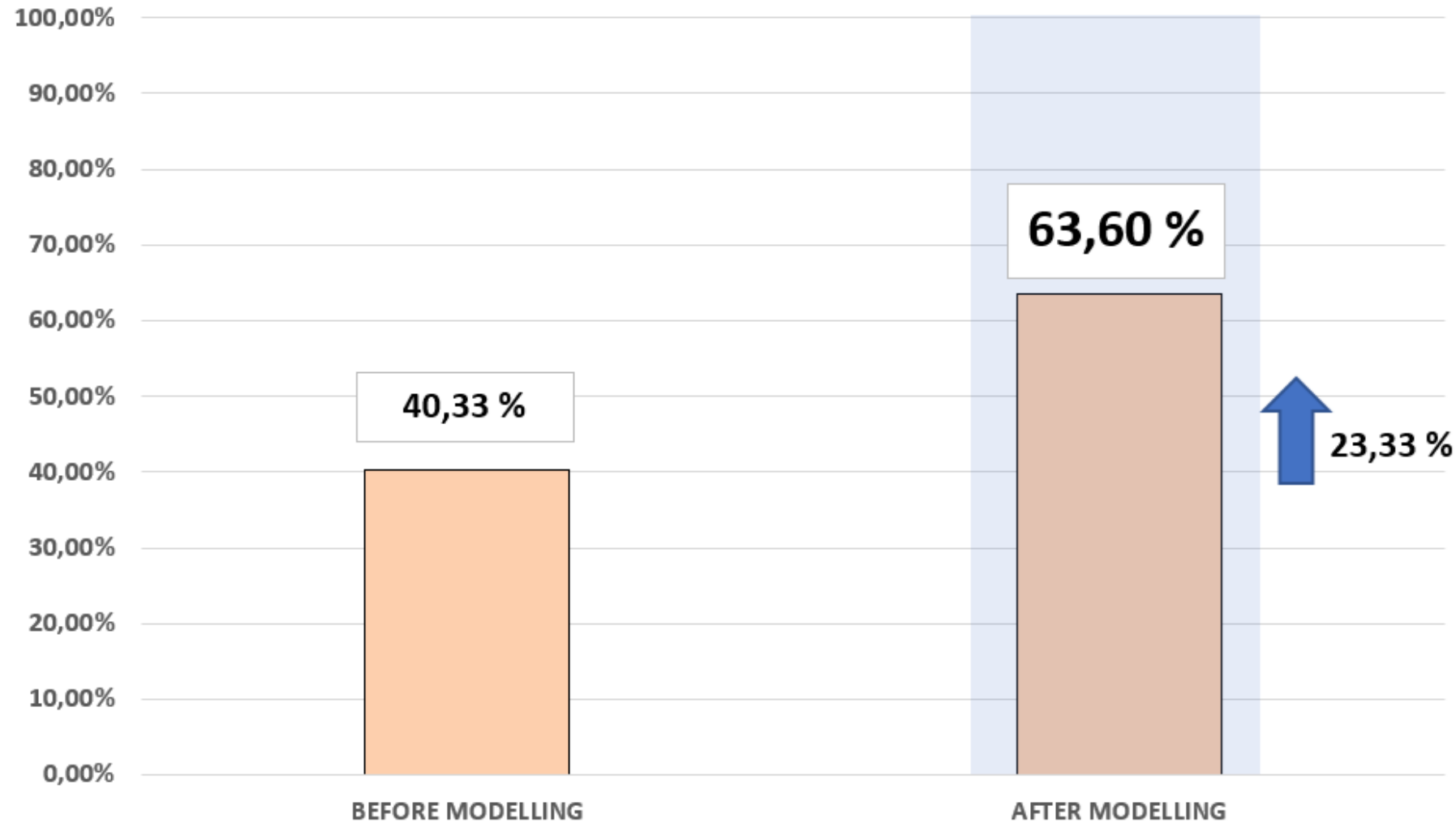


Setelah Perusahaan Menerapkan Machine Learning





## Persentase Barang Datang Tepat Waktu



Ada kenaikan sebesar

**23,27%**

barang datang tepat waktu setelah dilakukan pemodelan

# TERIMA KASIH

