

Final Project Rakamin Batch 35

# Kelompok 1 ShipTime Forecasters Shoppedia Delivery Prediction



<https://www.kaggle.com/datasets/prachi13/customer-analytics/>

# The Team

## -SHIPTIME FORECASTERS-



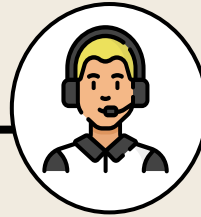
**Alvin Kurniawan  
Hanafie**  
Project Leader



**Berliana Ayu  
Aprilia**  
Data Analyst



**Gladys Olivia**  
Data Analyst



**Kevin Munandar**  
Data Scientist



**Aisyah Nur  
Hasanah**  
Data Scientist



**Awalsyah Erlangga**  
Business  
Intelligence



# LINE - UP THE PROBLEM



Background Problem

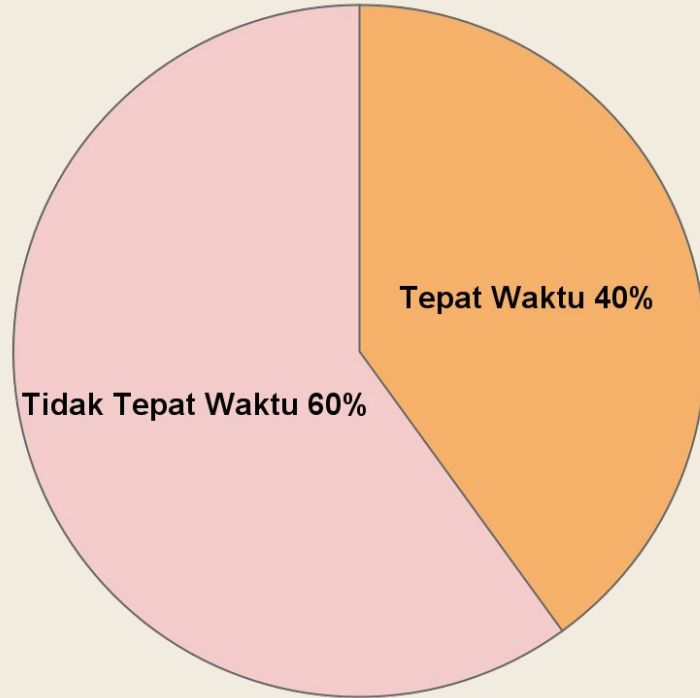


Business Recommendation



Business Simulation

# Background



**Status Pengiriman Barang**



Rendahnya persentase pengiriman barang tepat waktu pada e-commerce Shoppedia sebesar 40%. Kondisi ini dapat berdampak pada *customer retention*, yang akan mempengaruhi jumlah *sales* (*potential revenue loss* sebesar ~\$1.400.000).



# Faktanya...

84,1%

**Amazon**



Per Mei 2020, Amazon *Delivery On-Time Rate* mencapai 84,1%.  
(**statista.com**)

13%

Customer tidak akan membeli kembali suatu merek jika pengirimannya terlambat. (**Oracle**)



# Background

A

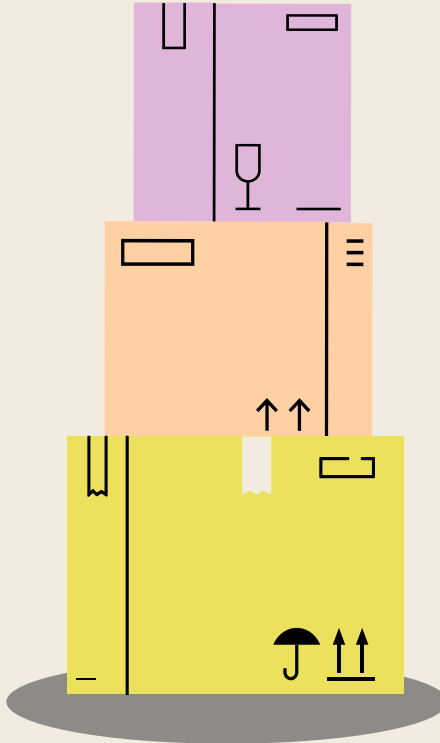
## Goals

Meningkatkan ***Delivery on Time (DoT) rate*** dari **40%** ke **60%**.

C

## Business Metrics

*Delivery on Time (DoT) Rate*



B

## Objective

- Membuat model yang dapat memprediksi apakah barang dapat sampai tepat waktu atau tidak.
- Membuat rekomendasi bisnis berdasarkan hasil prediksi untuk menurunkan *potential revenue loss*.

# Project Timeline

## Start

Preparation



1

## Stage 1

EDA, Insights & Visualization

## Stage 2

Data  
Pre-processing

2

3

## Stage 3

Modelling & Evaluation

## Stage 4

Final  
Presentation

4



## Final

The Day For  
Final



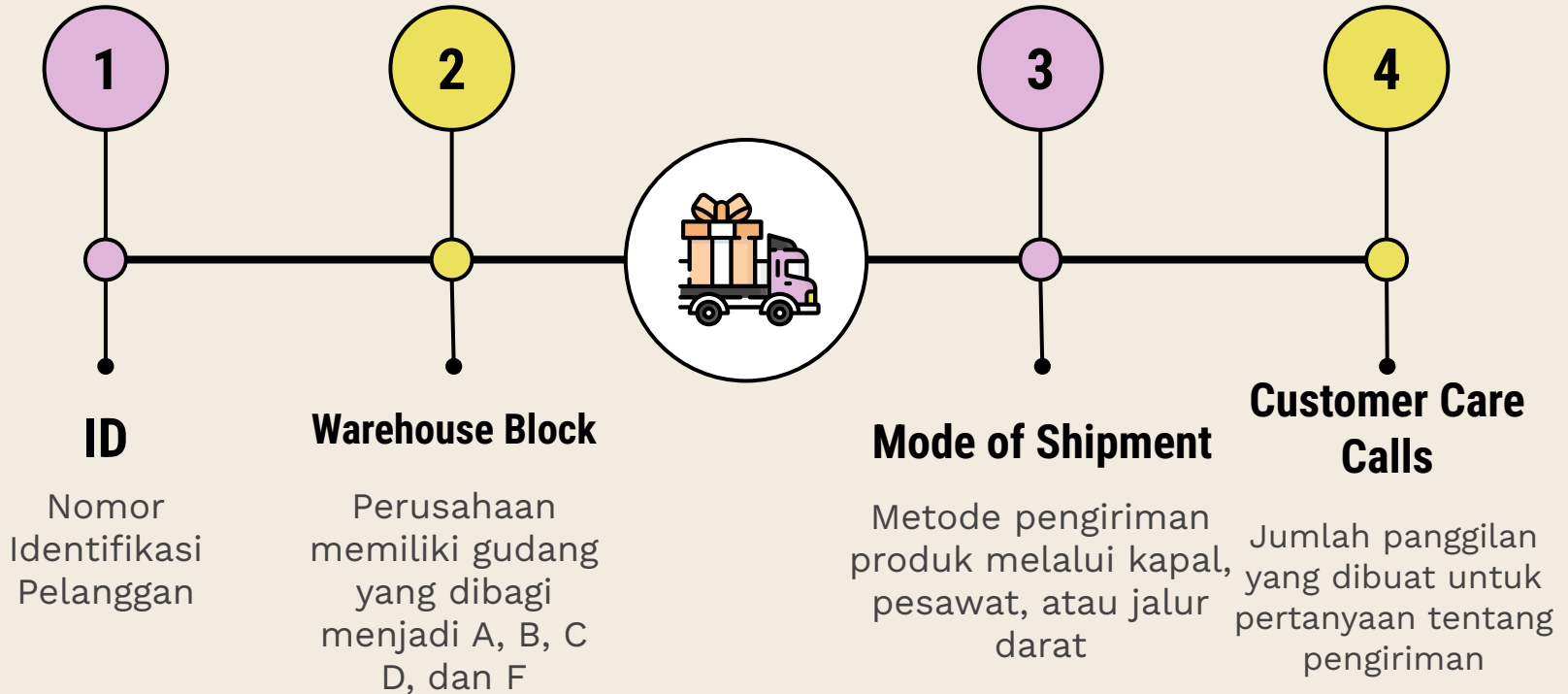
# Exploratory Data Analysis





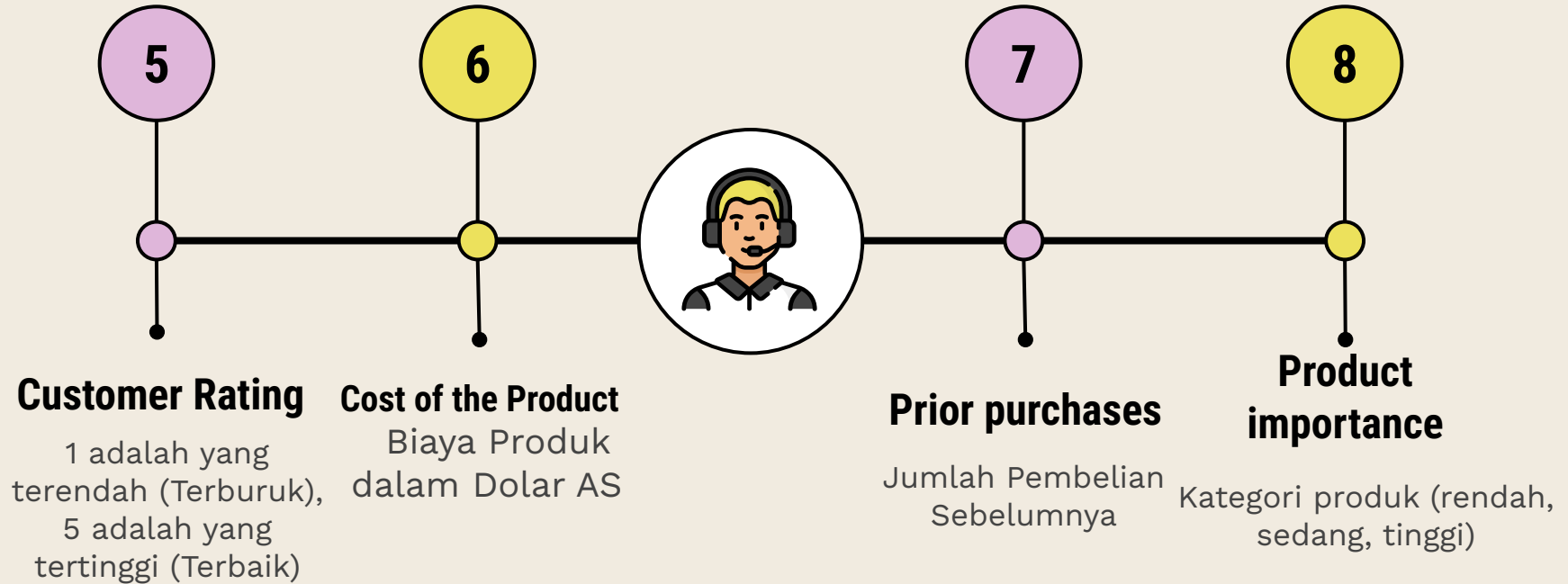
# Data Set

12 Columns and 10999 Rows



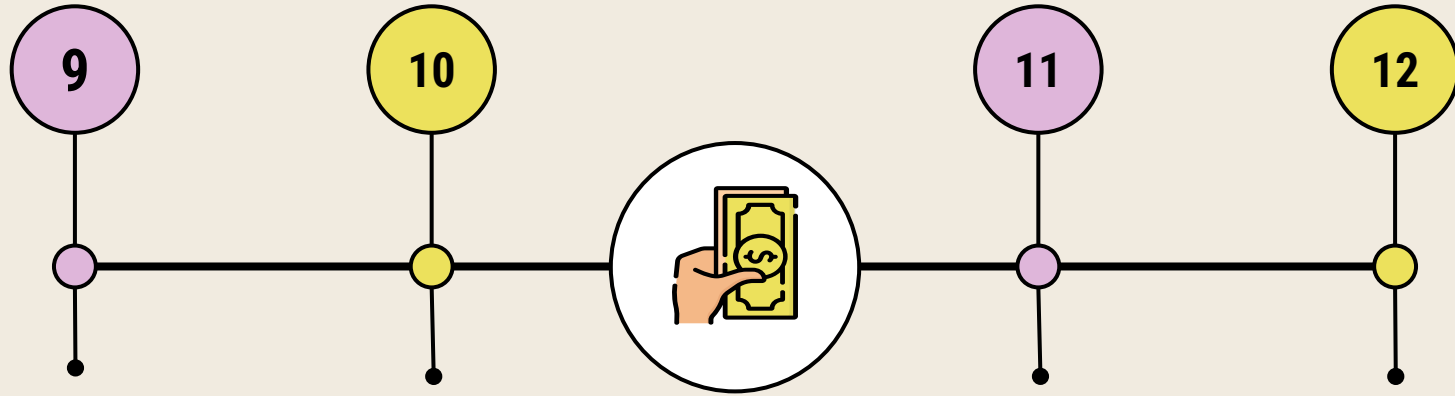
# Data Set

12 Columns and 10999 Rows



# Data Set

12 Columns and 10999 Rows



**Gender**

Laki-laki dan  
Perempuan

**Discount offered**

Diskon yang  
diberikan

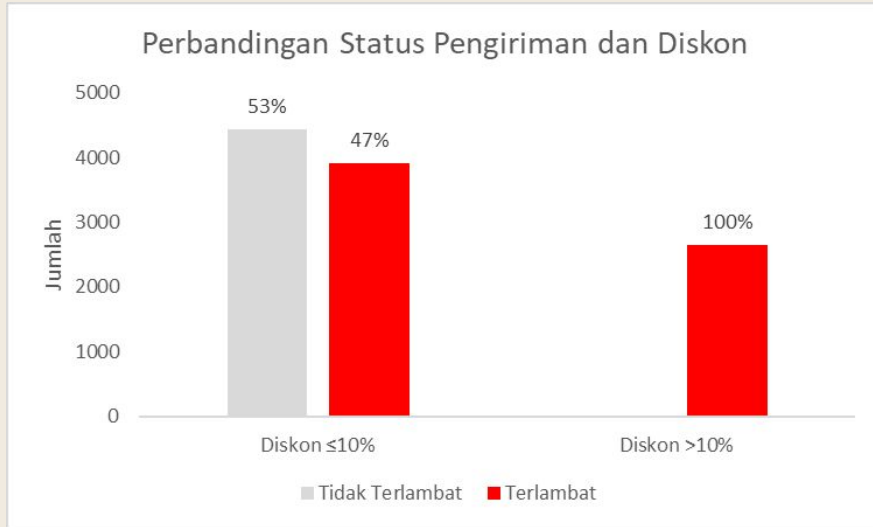
**Weight in gms**

Berat dalam  
gram

**Delivered on time**

1 menunjukkan bahwa  
barang TIDAK tepat  
waktu dan 0  
menunjukkan bahwa  
barang tersebut tepat  
waktu

# Data Insight (1)



“Pengiriman semua barang dengan Diskon  $> 10\%$  terlambat”

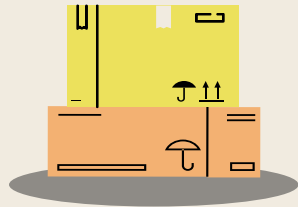
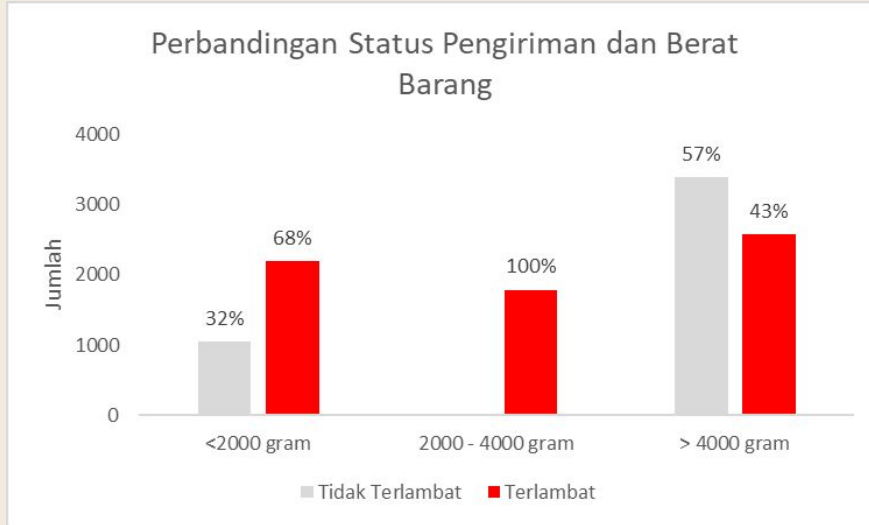
## Asumsi:

1. Barang dengan diskon besar tidak diperhatikan performa pengirimannya.
2. Pengurangan *delivery cost* pada barang dengan diskon besar sebagai penyebab keterlambatan barang.

**Rekomendasi:** Melakukan penyesuaian *discount rate* dan penelusuran lebih lanjut penyebab keterlambatan barang dengan diskon besar.



# Data Insight (2)



“Pengiriman semua barang dengan berat 2000 - 4000 gram terlambat”

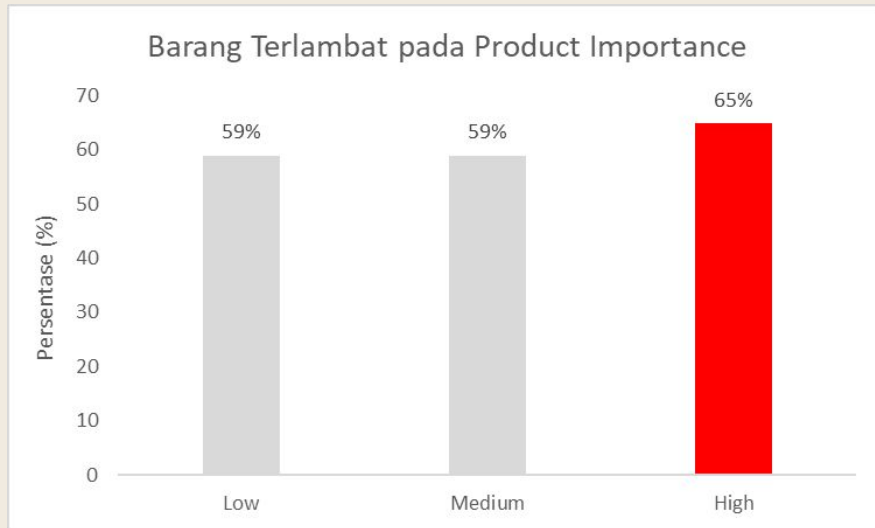
## Asumsi:

1. Berat barang “tangung” (kategori 2000 - 4000 gram tidak optimal jika digabung dengan kategori berat ringan)
2. Jumlah barang sedikit dibanding kelompok berat lain, jadi tidak terlalu diperhatikan performa pengirimannya.

## Rekomendasi:

- Analisa lebih dalam penyebab keterlambatan pada berat 2000 - 4000 gram
- *Tracking delivery* untuk memastikan status pengiriman barang.

## Data Insight (3)



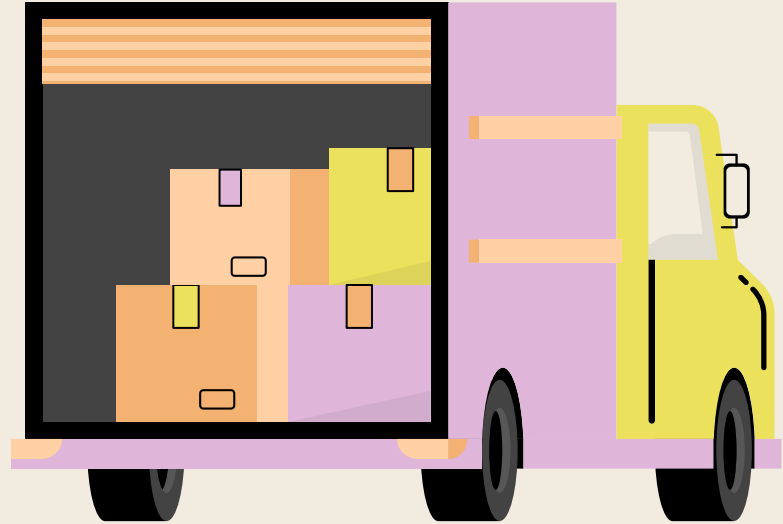
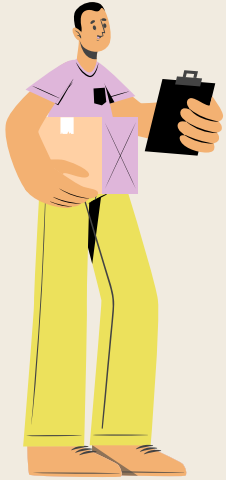
“Barang dengan *importance* tinggi paling banyak yang terlambat”

**Asumsi:** barang dengan *importance* ‘*high*’ memiliki proses *outbound* yang rumit sehingga rasio keterlambatannya lebih tinggi dibanding tingkat *importance* yang lain.

**Rekomendasi:** Diperlukan data tambahan terkait dasar penggolongan kategori *product importance*.

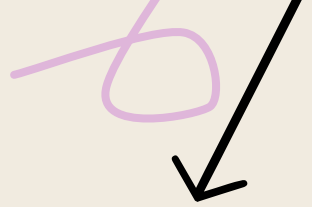


# Classification Model





# Data PreProcessing



## Handling Missing Value, Duplicate, dan Invalid Value

Tidak ditemukan *missing value*, *duplicated data*, dan *invalid value* pada dataset.

## Feature Extraction, Feature Selection, dan Feature Encoding

- Dilakukan pembuatan feature *Weight\_class* untuk kategorisasi barang berdasarkan berat menjadi 4 kategori.
- Dilakukan *drop* pada fitur *ID*, *Customer\_rating*, *Warehouse\_block*, *Mode\_of\_Shipment*, dan *Gender*.
- Dilakukan *feature encoding* pada fitur *Product\_importance*.

## Handling Outlier dan Data Transformation

*Handling outlier* dilakukan dengan data transformation yang *robust* terhadap *outlier* (*Quantile Transformation*) dan *standardization*, untuk menjaga agar dataset tetap utuh dan stabil pada saat proses modelling.



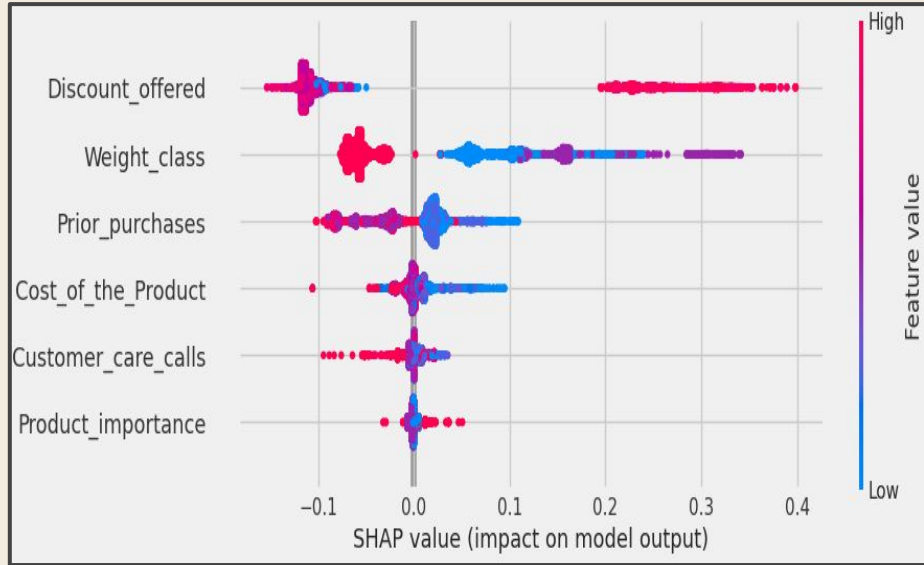
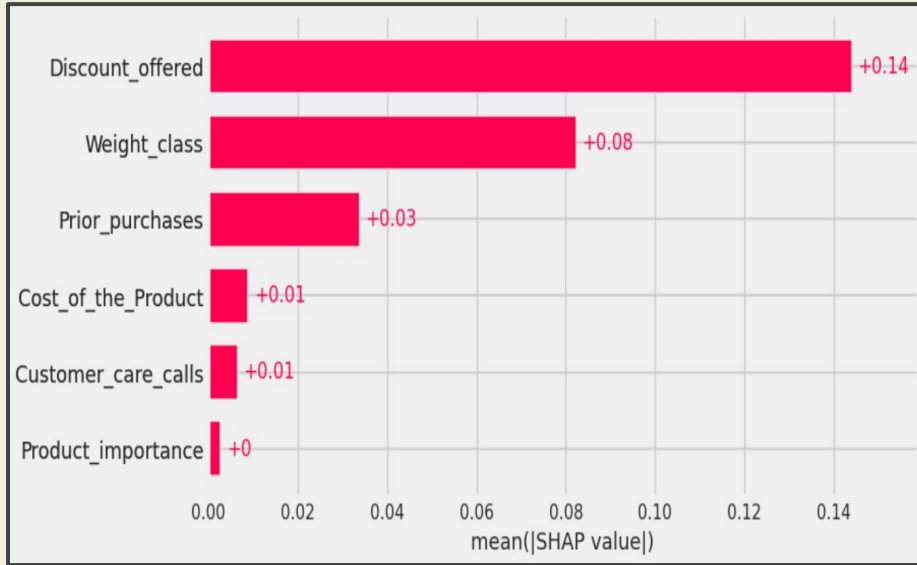


# Modelling

Classifier Metrics	Logistic Regression	KNN	Decision Tree	Random Forest	AdaBoost	XGBoost	CatBoost
Accuracy	0.63	0.67	0.67	0.63	0.68	0.68	0.68
Precision	0.65	0.78	0.83	0.65	0.83	0.81	0.81
Recall	0.80	0.61	0.55	<b>0.81</b>	0.58	0.60	0.61
ROC - AUC	0.71	0.74	0.75	0.74	0.74	0.74	0.75

**Recall:** digunakan untuk meminimalkan kesalahan prediksi pada barang yang terlambat dianggap sebagai barang tidak terlambat.

# Feature Importance



- 2 Fitur yang paling besar pengaruhnya adalah **Discount\_offered** dan **Weight\_class**.
- Semakin tinggi nilai diskon, semakin tinggi kemungkinan barang terlambat.
- Semakin rendah berat barang, semakin tinggi kemungkinan barang terlambat.



# Business Recommendation (1)





## Penyesuaian Rate Diskon

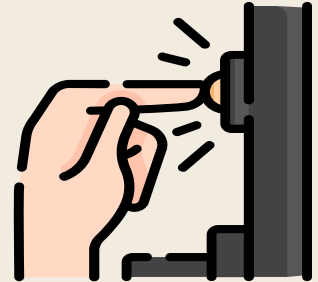
Melakukan penyesuaian pemberian diskon kepada pada kelompok barang dengan diskon  $>10\%$

Rata-rata diskon awal (barang yang terlambat): **38%**  
Rata-rata penyesuaian : **10%**



**! Penyesuaian Rate Diskon pada barang yang terlambat akan mengakibatkan *Customer Loss* !**

Asumsi *Customer Loss*: 25%  
(hanya *customer* yang mengalami keterlambatan)





# Business Simulation (1)



# Pre - Condition

1

# Delivery Terlambat

529

2

Rata-rata Diskon Awal

38%

3

Rata-rata Diskon Penyesuaian

10%

4

Potensi Customer Loss

25%

5

Rata-rata Harga Barang

\$196

# Penyesuaian Diskon pada Transaksi yang **Terlambat**

**Revenue Awal**

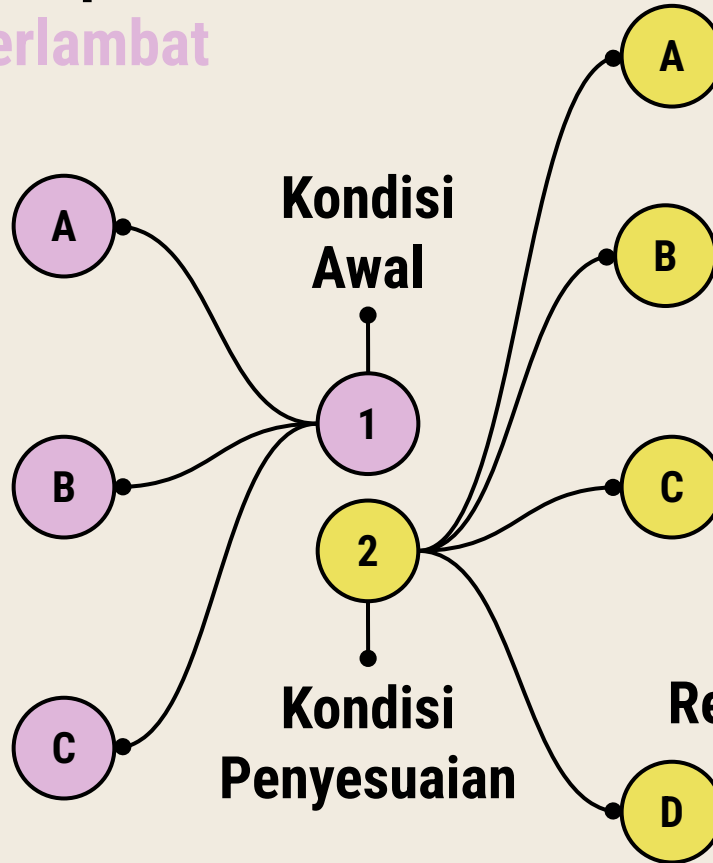
\$103.684

**Anggaran Diskon Awal**

\$39.400

**Revenue Awal Setelah Diskon**

\$64.284



**Potensial Revenue Loss**

\$25.921

**Revenue Penyesuaian**

\$77.763

**Anggaran Diskon Penyesuaian**

\$7.776

**Revenue Penyesuaian Setelah Diskon**

\$69.987

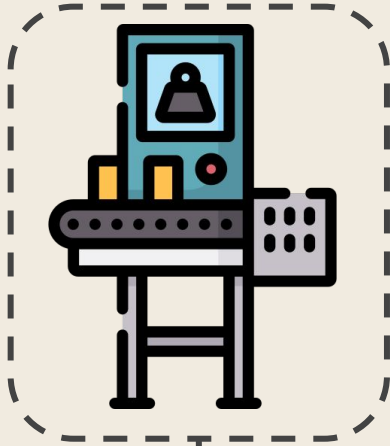


**Ada penambahan Revenue setelah menerapkan penyesuaian diskon, sebesar \$ 5.703, meskipun mengalami Customer Loss.**



# Business Recommendation (2)

Melakukan investasi untuk *improvement* operasional pengiriman (berfokus pada berat barang yang mempengaruhi keterlambatan pengiriman)



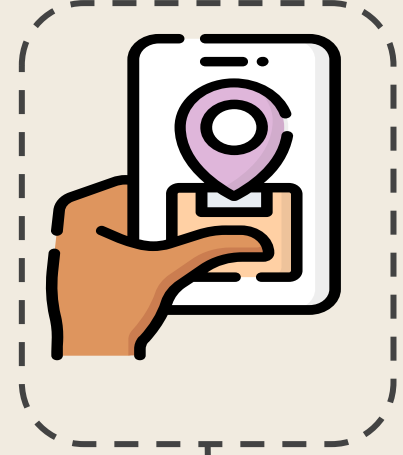
Checkweigher



Warehouse khusus Weight\_class tertentu



WMS yang efisien



Tracking Delivery

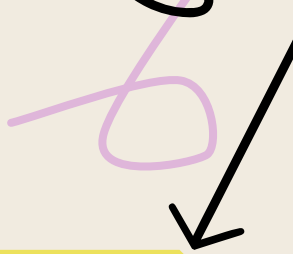


# Business Simulation (2)





# The Urgency



## Customer Loss

Keterlambatan barang berpotensi menyebabkan *customer* tidak kembali melakukan transaksi di masa depan.

## Potential Revenue Loss

Potensi *revenue* yang hilang karena *customer* tidak kembali melakukan transaksi.

## Forecasting Pengiriman Barang

Simulasi dilakukan untuk memberikan *treatment* yang tepat pada barang setelah melakukan prediksi keterlambatan barang.

# Pre-Condition

	Prediksi Tidak Terlambat	Prediksi Terlambat	Total Aktual
Aktual Tidak Terlambat	340	555	895
Aktual Terlambat	252	1053	1305
Total Prediksi	592	1608	2200

**Tidak dilakukan *modelling*:**

1.305 dari 2.200 barang terlambat dan tidak diberikan *treatment*

***Potential Revenue Loss:***

$1.305 \times \$213$  (rata-rata harga barang) = ~**\$280.000**

***Forecasting dengan Machine Learning!***

# Get The Fact

1 Untuk meningkatkan ***Delivery on Time rate***, diperlukan biaya tambahan.

2 Rata-rata biaya pengiriman barang sebesar **10 - 15%** dari total harga barang (**hopstack.io**).

3 Total harga barang dari 2.200 transaksi yang diprediksi: **\$460.926**

4 Biaya total pengiriman barang:  $12,5\% \times \$460.926 = \sim \mathbf{\$60.000}$

5 Biaya pengiriman per barang:  
 $\$60.000 / 2.200 = \mathbf{\$27 / barang}$

# The Resources

1

Shoppedia hanya berhasil mengirim **40% barang tepat waktu** dengan biaya pengiriman **\$60.000 (\$27 / barang)**.

2

Untuk mengirimkan **100% barang tepat waktu**, diperlukan biaya pengiriman **\$150.000** (2.5x kapasitas normal)

3

Biaya pengiriman per barang agar barang tepat waktu:  
 $\$150.000 / 2.200 =$   
**\$68 / barang**

4

Untuk meningkatkan *DoT rate*, Shoppedia diberikan **budget tambahan** sebesar **33% dari budget normal (\$20.000)**

# Calculating...

**Total budget** untuk meningkatkan *DoT rate*:

1. \$60.000 (budget awal)
2. \$20.000 (budget tambahan)
3. ~\$6,000 (tambahan budget dari penyesuaian diskon)
4. Total: **\$86.000**

**Alokasi** pemberian budget berdasarkan hasil prediksi *machine learning*:

Barang diprediksi **tepat waktu**:

**Budget normal** (\$27 / barang)

$$592 \times \$27 = \sim \$16.000$$

Barang diprediksi **terlambat**:

**Budget tambahan** (\$68 / barang)

$$\text{Sisa budget: } \$86.000 - \$16.000 = \$70.000$$

Barang yang dapat diberikan budget tambahan dari sisa budget:  
 $\$70.000 / \$68 = \mathbf{1.030 \text{ barang}}$

# The Impact

**Prediksi barang yang akan terkirim tepat waktu:**  
**340** (budget normal) + **1.030** (budget tambahan)  
**= 1.370 barang**



***Delivery on Time (DoT) rate* =**  
 $1.370 / 2.200 \times 100\% = \mathbf{62\%}$





# Potential Revenue Loss Salvaged

## **Potential Revenue Loss (before treatment):**

1.305 x \$213 (rata-rata harga barang) =  
**~\$280.000**



## **Potential Revenue Loss (after treatment):**

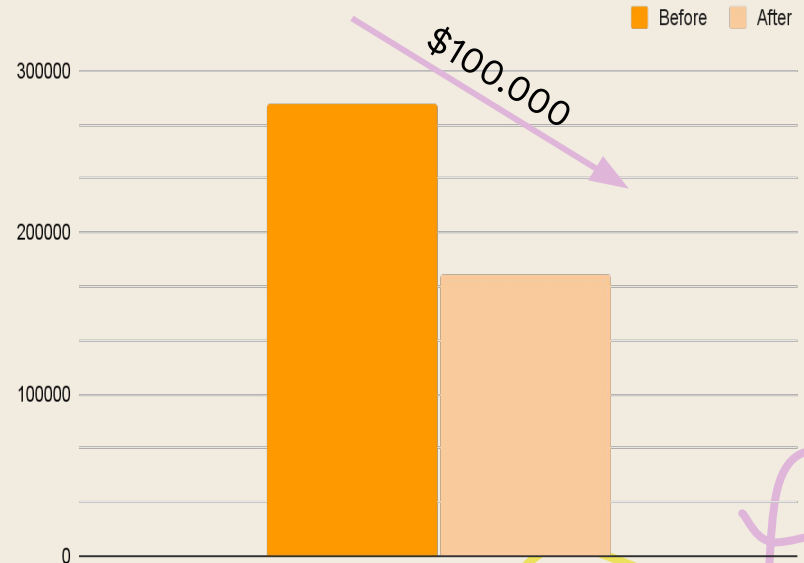
830 x \$213 (rata-rata harga barang) =  
**~\$180.000**



## **Potential Revenue Loss Salvaged:**

\$280.000 - \$180.000 = **\$100.000**  
**(~36% potential revenue salvaged)**

Potential Revenue Loss



# Summary

Hasil prediksi model *machine learning* dan simulasi rekomendasi bisnis (penyesuaian *discount rate* dan *improvement* operasional pengiriman barang) dapat meningkatkan *Delivery on Time rate* dari 40% menjadi **62%**.



**Terima**  
Kasih

