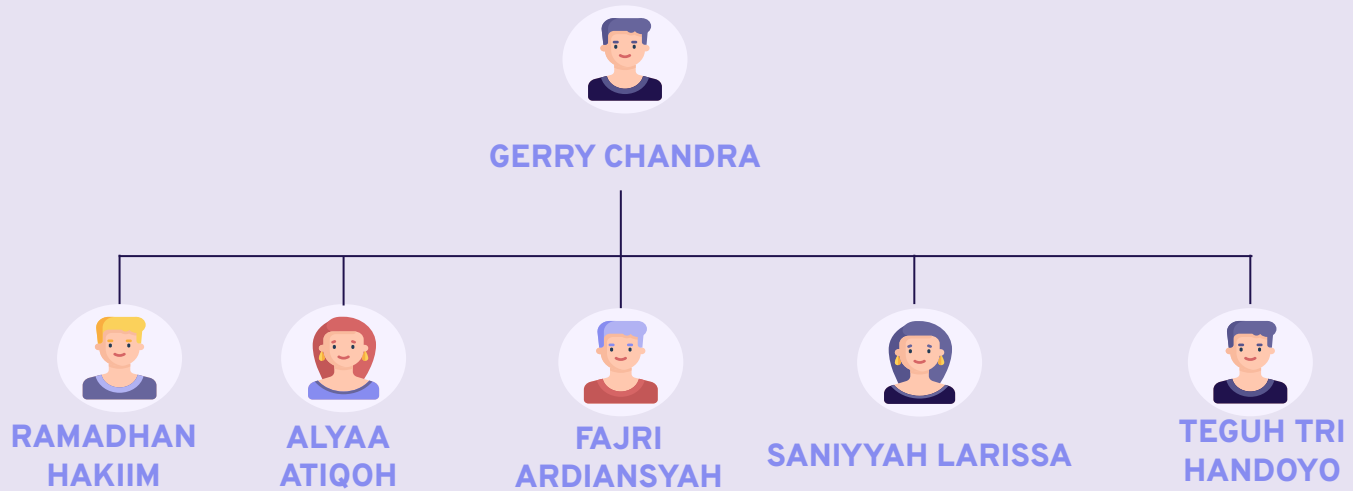




# FINAL PRESENTATION

SMELLS LIKE TEAM SPIRIT

# DATA SCIENTIST TEAM

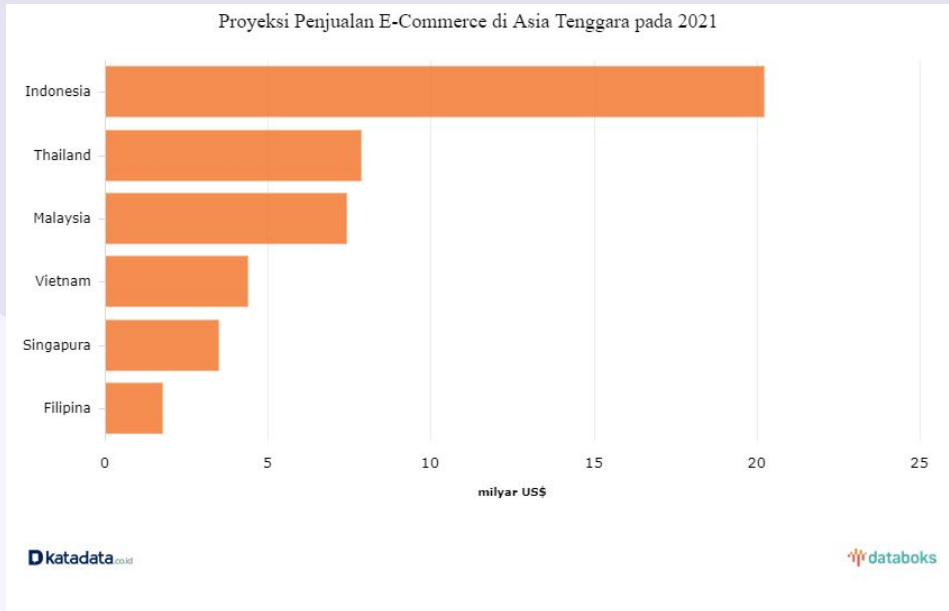




## OUR COMPANY

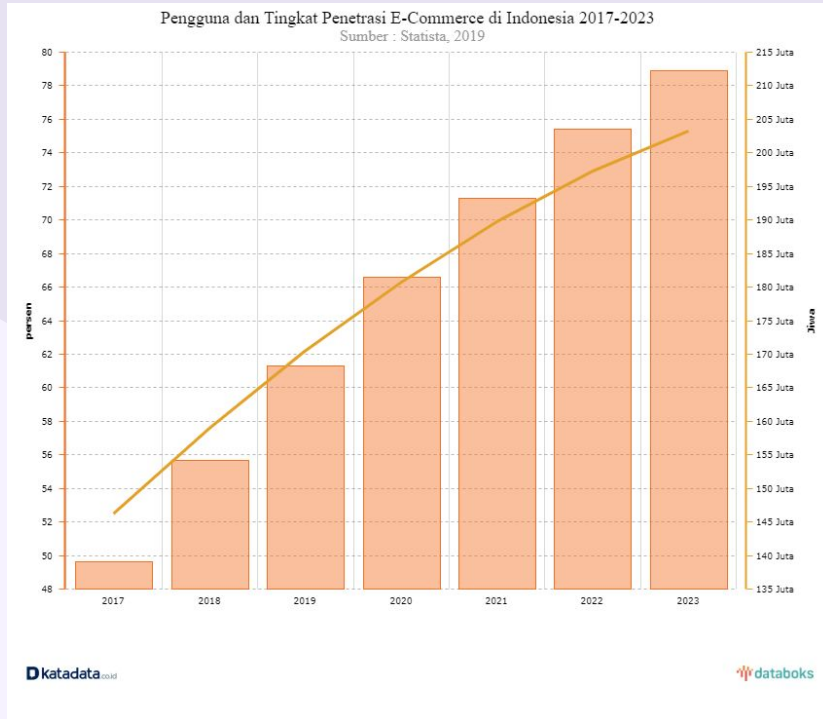
ANTERIN DONG adalah sebuah perusahaan e-commerce Indonesia yang berdiri sejak tahun 2018. Kami memiliki 3 jalur pengiriman barang, yaitu melalui jalur udara, laut, dan darat. Motto kami adalah ***"Your order is our happiness, and your happiness is our pride"***.

## Proyeksi Penjualan E-Commerce di Asia Tenggara



Berdasarkan data dari *katadata.co.id* social distancing menyebabkan penjualan retail melalui *e-commerce* dari enam negara di **Asia Tenggara** diperkirakan **mencapai US\$ 45,07 miliar** pada 2021. Dari jumlah itu, **Indonesia** menempati posisi pertama dengan proyeksi penjualan senilai **US\$ 20,21** atau sekitar **44.8%**

## Tren Pengguna E-Commerce



Statista mencatat jumlah pengguna *e-commerce* di Indonesia naik secara signifikan setiap tahunnya dan diproyeksikan akan mencapai **212,2 juta pada 2023**.

**APAKAH ADA DAMPAK TREN TERSEBUT  
PADA PERUSAHAAN ANTERIN DONG?**

**ADA DONG**

## LATE SHIPMENT





# BUSINESS UNDERSTANDING

## Problem

Anterin dong memiliki rate keterlambatan pengiriman barang sebanyak 59,6% dari total jumlah pengiriman barang.

## Business Metrics

- Persentase keterlambatan
- Customer Care Calls (Jumlah telepon ke customer care)

## Goals

Mengurangi rate keterlambatan pengiriman barang menjadi 20% dari total jumlah pengiriman barang.

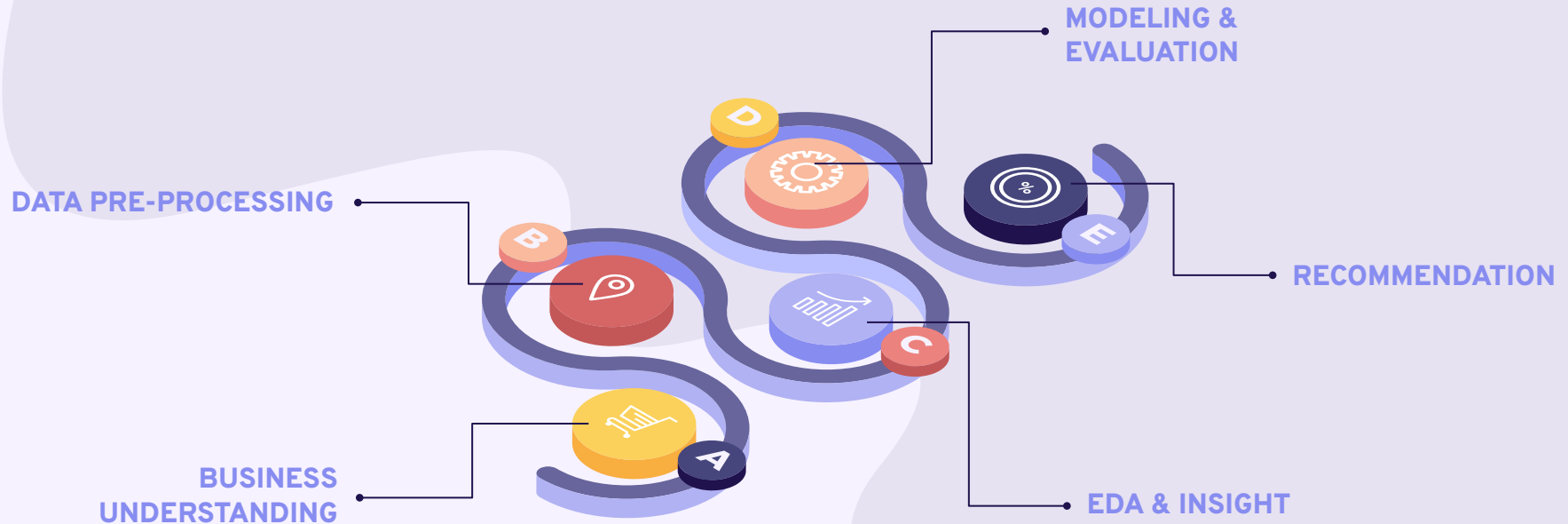
## Objective

- Membuat model machine learning untuk memprediksi ketepatan pengiriman barang.

# Impact of Poor eCommerce Shipping & Logistics Experience

- Company may Lose Customer's Loyalty
- A Negative Impact on the Brand Image
- This Will Lead to Increased Logistics Cost
- More Pressure on the Customer Support Team
- Increase in Negative Reviews on the Website
- Bad Reputation on Social Media Platforms

# DATA SCIENCE PROCESS



# DESCRIPTIVE INSIGHTS



## Dengan menggunakan reached on time sebagai target kolom, berikut beberapa insight yang kami dapatkan dari fitur numerik dan categorical

```
df[nums].describe()
```

	ID	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Discount_offered	Weight_in_gms	Reached.on.Time_Y.N
count	10999.00000	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000
mean	5500.00000	4.054459	2.990545	210.196836	3.567597	13.373216	3634.016729	0.596691
std	3175.28214	1.141490	1.413603	48.063272	1.522860	16.205527	1635.377251	0.490584
min	1.00000	2.000000	1.000000	96.000000	2.000000	1.000000	1001.000000	0.000000
25%	2750.50000	3.000000	2.000000	169.000000	3.000000	4.000000	1839.500000	0.000000
50%	5500.00000	4.000000	3.000000	214.000000	3.000000	7.000000	4149.000000	1.000000
75%	8249.50000	5.000000	4.000000	251.000000	4.000000	10.000000	5050.000000	1.000000
max	10999.00000	7.000000	5.000000	310.000000	10.000000	65.000000	7846.000000	1.000000

```
df[cats].describe()
```

	Warehouse_block	Mode_of_Shipment	Product_importance	Gender
count	10999	10999	10999	10999
unique	5	3	3	2
top	F	Ship	low	F
freq	3666	7462	5297	5545

Mean dan median pada customer care calls dan customer rating sudah sama (bell distribution)

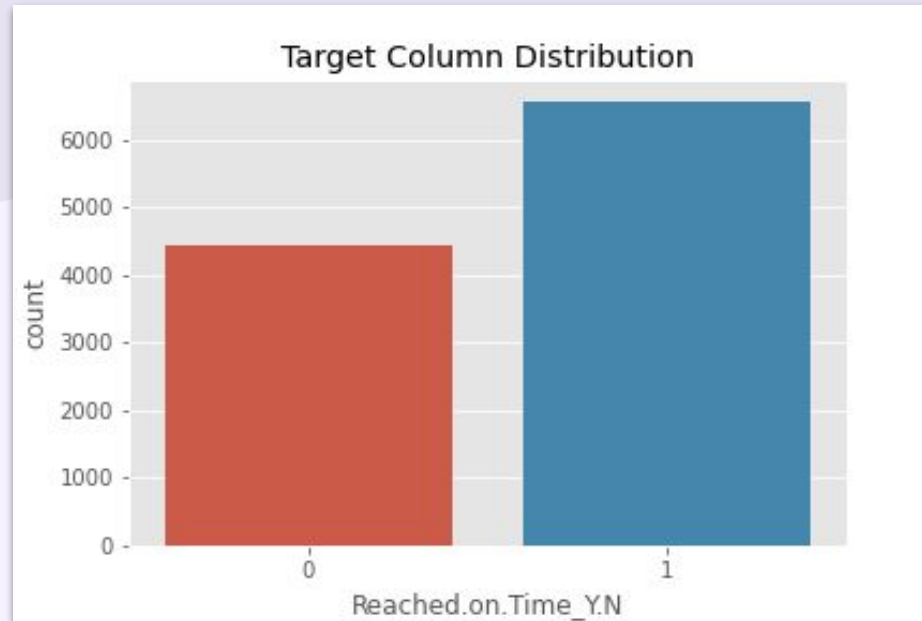
Tidak ada missing value dan duplicated value

Tidak terdapat kejanggalan dalam nilai minimum setiap kolom

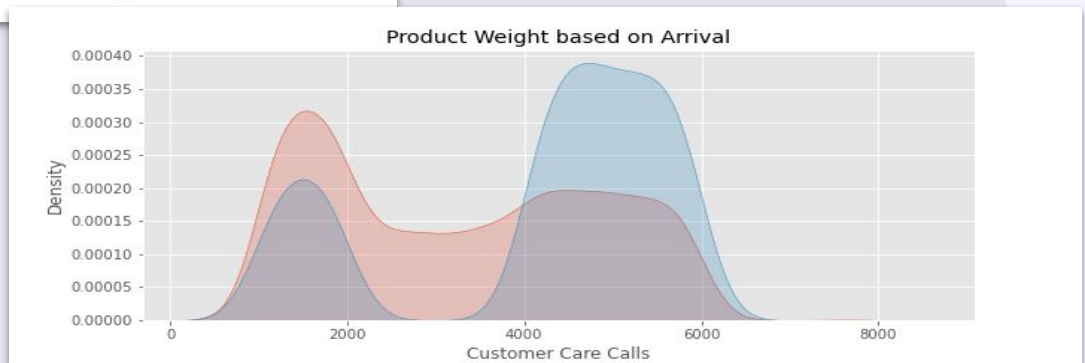
Ditemukan gap nilai yang cukup jauh antara percentile 75% dan 100% pada kolom Prior\_purchase dan Discount\_offered

Mode of shipment tertinggi dengan ketimpangan adalah **"Ship"** dengan frekuensi 7462 dari total 10999

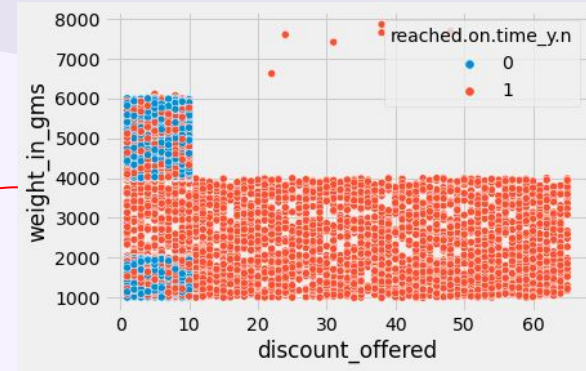
Pada kasus ini, kami menjadikan reached on time sebagai target kolom kami, berikut distribusinya



## Perbandingan **discount offered** dan **product weight** terhadap target kolom.



# Hubungan Numerik ke Numerik



## Produk kemungkinan tepat waktu

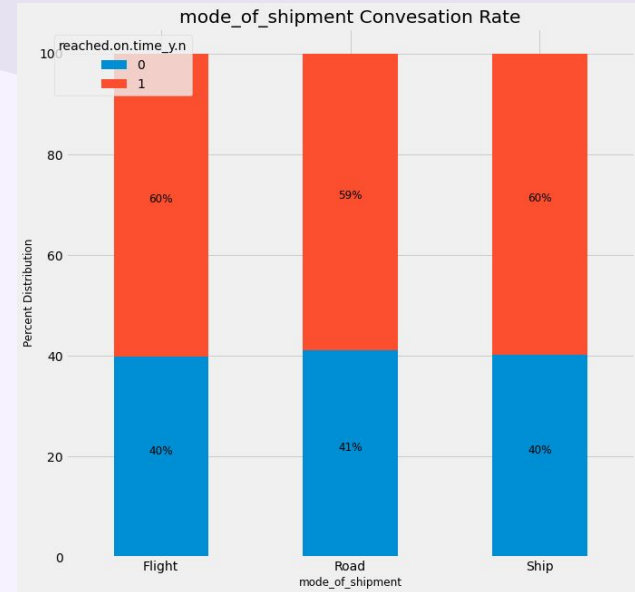
- Diskon produk berada pada 0% - 10 %
- Berat berada pada 1kg - 2kg dan 4kg - 6kg.



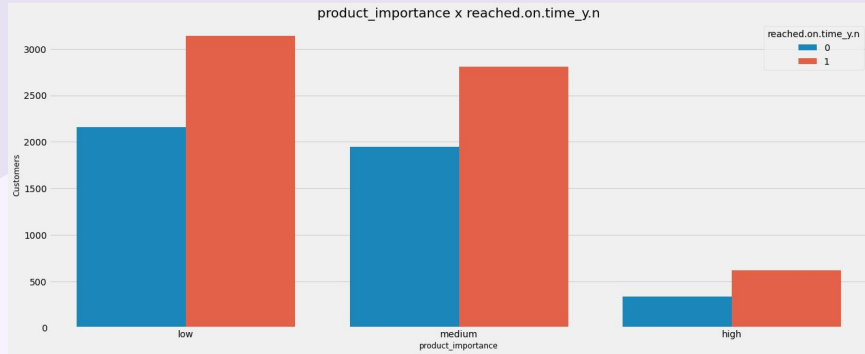
# Hubungan Mode of Shipment ke Target



Jalur pengiriman (mode\_of\_shipment) yang paling banyak digunakan adalah **ship**.

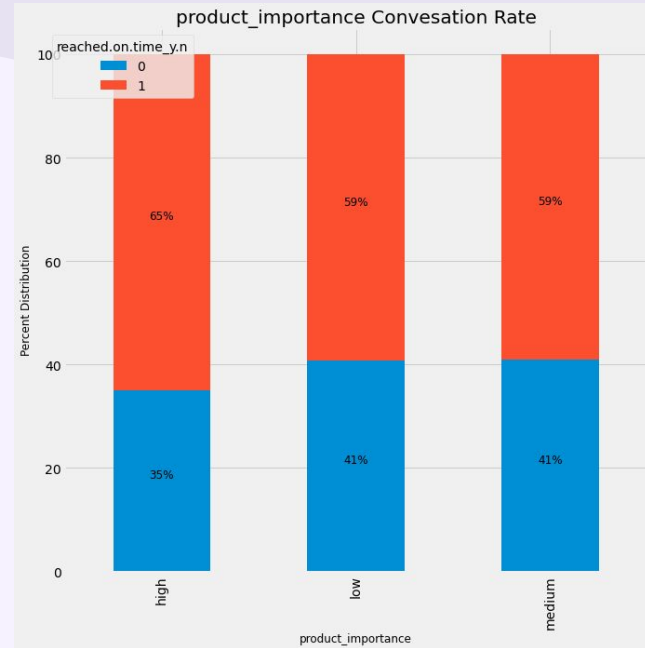


# Hubungan Product Importance ke Target



Mayoritas tingkat kepentingan produk adalah **low** dan **medium**.

Seharusnya tingkat kepentingan produk high memiliki persentase keterlambatan yang lebih kecil.



# Insight



- (Asumsi 1) Event besar yang memberikan **diskon tinggi** berdampak pada pembelian produk yang banyak pada waktu yang bersamaan. **Solusi:** Persiapan seluruh tim ketika ada event diskon yang tinggi.
- (Asumsi 2) **Perbedaan penanganan** untuk produk di atas 2kg dan di bawah 4kg membuat pesanan dengan kategori berat tersebut tidak ada yang sampai tepat waktu. **Solusi:** Penanganan produk 2kg - 4kg harus lebih diperhatikan
- (Asumsi 3) **Kepentingan produk seharusnya diutamakan** agar sampai tepat waktu. **Solusi:** Adanya perbaikan pada manajemen pengiriman untuk memperhatikan kepentingan produk.

# DATA PRE-PROCESSING



Data set : E-Commerce Shipping Data  
 Jumlah Feature : 12  
 Jumlah Baris Data : 10.999

Data Cleansing = 0  
 Missing Value = 0  
 Duplicate Value = 0

 = Skewed Distribution  
 = Outlier

No	Feature	Information	Action
1	ID	Number of Customers	Drop
2	Reached.on.Time_Y.N	It is the target variable, where 1 Indicates that the product has NOT reached on time and 0 indicates it has reached on time	Rename: ' <b>Late</b> ' 0 = On Time, 1 = Late
3	Weight_category1	It is the weight in grams	Transfrom Log, Lamda, Label Encoding
4	Gender	Male and Female	Label Encoding
5	Product_importance	The company has categorized the product in the various parameter such as low, medium, high	Label Encoding
6	Discount_offered	Discount offered on that specific product	Transfrom Log
7	Prior_purchases	The Number of Prior Purchase	Transfrom Log
8	Customer_care_calls	The number of calls made from enquiry for enquiry of the shipment	Transfrom Log
9	Mode_of_Shipment	The Company Ships the products in multiple way such as Ship, Flight and Road	One-hot Encoding
10	Warehouse_block	The Company have big Warehouse which is divided in to block such as A,B,C,D,E	One-hot Encoding
11	Cost_of_the_Product	Cost of the Product in US Dollars	Numeric int ke Numeric float
12	Customer_rating	The company has rated from every customer. 1 is the lowest (Worst), 5 is the highest (Best)	Numeric int ke Numeric float

# MODELING EVALUATION

Pemodelan machine learning untuk  
memprediksi keterlambatan produk.



# MODEL SELECTION



**Logistic  
Regression**



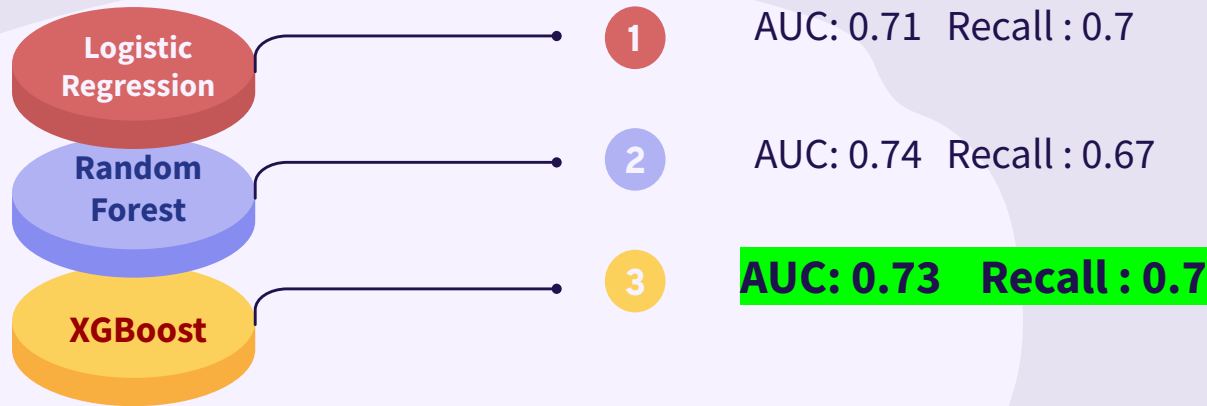
**Random Forest**



**XGBoost**

Pemodelan dilakukan dengan membandingkan 3 algoritma machine learning: Logistic Regression, Random Forest Classifier, dan XGBoost.

## MODEL PERFORMANCE



Metrics yang akan menjadi tolak ukur pada kasus ini adalah AUC dan recall. Jadi, berdasarkan performa di atas model yang direkomendasikan adalah **XGBoost**.



# POTENTIAL IMPACT

## Confusion Matrix

<b>23.4 %</b> True Neg	<b>16.4 %</b> False Pos
<b>17.9 %</b> False Neg	<b>42.4 %</b> True Pos

Late = 1 (pos) ; On Time = 0 (neg)

## Scenario

- 1 Model diterapkan untuk internal perusahaan.
- 2 Setiap shipment diprediksi apakah akan sampai tepat waktu atau tidak.
- 3 Asumsi: Barang yang diprediksi terlambat akan diantisipasi sehingga jadi tepat waktu.

# POTENTIAL IMPACT

## Scenario

1

Model diterapkan untuk customer service.

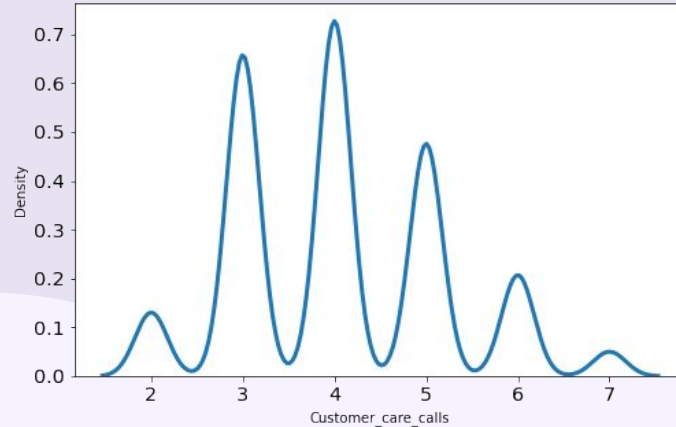
2

Ketika customer menelepon, CS sudah bisa memberitahu kemungkinan barang sampai tepat waktu atau tidak.

3

Asumsi threshold pelanggan tidak puas:  
 $\text{customer\_care\_calls} > 3$ .

## Customer Care Calls



# POTENTIAL IMPACT

Metrics	Before	After
Persentase produk on time	40.4%	82.2%
Potensi churn pelanggan	59,6%	17.8%
Potensi Kerugian	\$ 412,772	\$123,806



**Penghematan = \$288,966**



**Penurunan rate keterlambatan = 41,77%**

# BUSINESS RECOMMENDATION

Rekomendasi untuk kemajuan  
Anterin Dong kedepannya.



# OUR RECOMMENDATION

- Penerapan model machine learning ke sistem pengiriman perusahaan, sehingga perusahaan **dapat mengantisipasi produk-produk yang diprediksi akan sampai terlambat** sehingga dapat sampai tepat waktu.
- Buat sistem menggunakan model machine learning di mana **pelanggan dapat selalu up-to-date dengan status pengiriman** sesuai konteks model machine learning yang dibuat. Output yang diharapkan: pelanggan tahu apakah barangnya sampai tepat waktu atau tidak.
- Buat perbaikan pada sistem estimasi pengiriman, karena mayoritas pengiriman tidak sesuai dengan jadwal. Ada kemungkinan bahwa estimasi yang digunakan sudah tidak valid.
- **Penambahan feature/data seperti lokasi tujuan /zona wilayah, tanggal order, estimasi waktu dll** dengan tujuan agar mempermudah analisa selanjutnya yang lebih akurat.

# THANKS!

Any questions?

