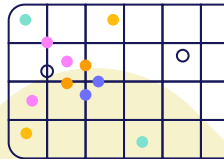
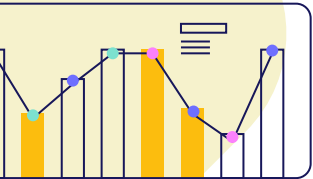


# PREDICTING PACKAGE STATUS: IN E-COMMERCE LOGISTICS

Nama Kelompok : Last Voyage  
Dataset : Amazon Sales Dataset  
Mentor : Dedi Irwanto Samosir



# MEMBER TEAM



**Mardio Edana Putra**

Project Manager



**Mochamad Rizqi**

Data Engineer



**Muhammad Gusti  
Maulana Alfin**

Data Scientist



**Nena Herdiani**

Data Analyst



# OUTLINE PRESENTASI

**01.**

**BACKGROUND**

**02.**

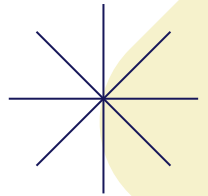
**EDA & INSIGHT**

**03.**

**DATA PRE-PROCESSING**

**04.**

**MODELING & REKOMENDASI**

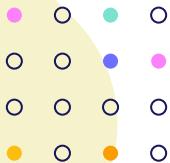




**01.**

# BACKGROUND


Latar Belakang, Dataset  
Review, Identifikasi  
Permasalahan, Goal & Objektif,  
Business Metrics





## Problem Statement

Tingkat pengembalian produk kepada seller selama periode bulan April-Juni 2022 yaitu **6.38%** dengan **total lost sales** **INR 1.275.310**.



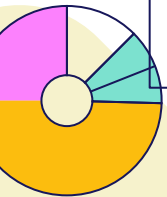
Meskipun tingkat pengembalian berada dalam **kisaran standar industri** untuk kategori umum, **dampaknya terhadap revenue dan efisiensi operasional** perlu dievaluasi lebih dalam. Oleh karena itu, diperlukan analisis mendalam untuk mengidentifikasi **akar penyebab pengembalian produk**, dan mengurangi **lost sales**.

## Goal & Objektif

1. Membangun **model machine learning** yang fokus pada memprediksi status paket dengan tingkat recall yang tinggi.
2. Mengidentifikasi **fitur-fitur utama** yang memengaruhi status pengembalian produk sehingga perusahaan dapat **mengembangkan strategi proaktif** untuk menangani pelanggan yang **berisiko returned**.

## Business Metrics

**Recall model** dan **tingkat persentase penurunan pengembalian produk**

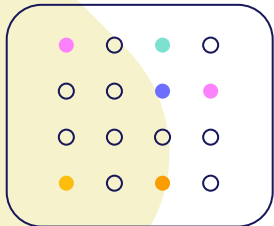




**02.**

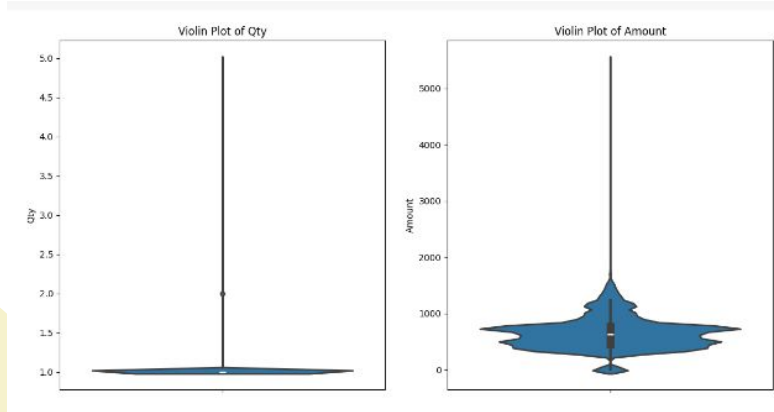
# EDA & INSIGHTS

Univariate & Multivariate  
Analysis, Business Insight





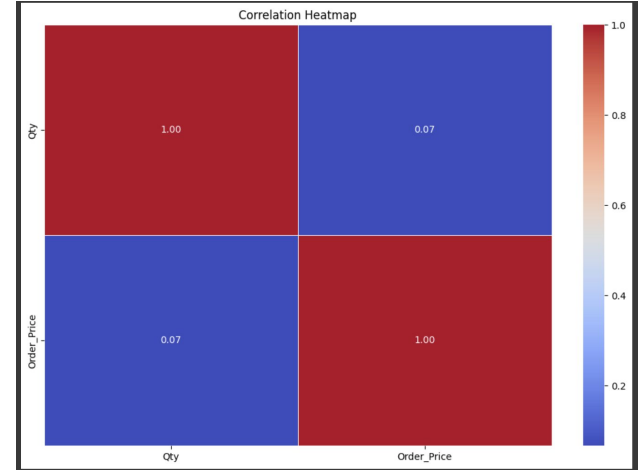
## Univariate Analysis



Fitur kuantitas berkumpul di 1 dan Fitur Amount(Harga) berdistribusi positively skewed



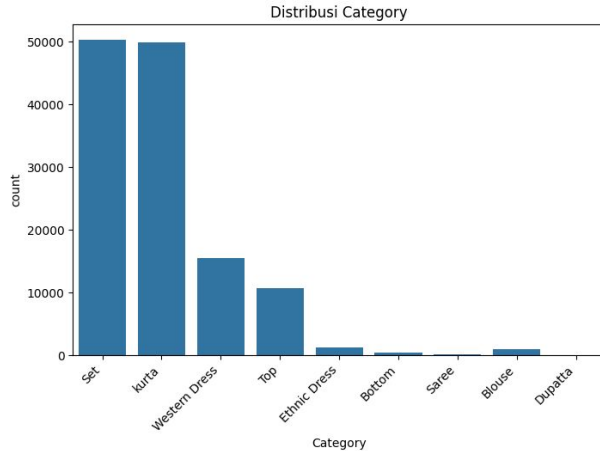
## Multivariate Analysis



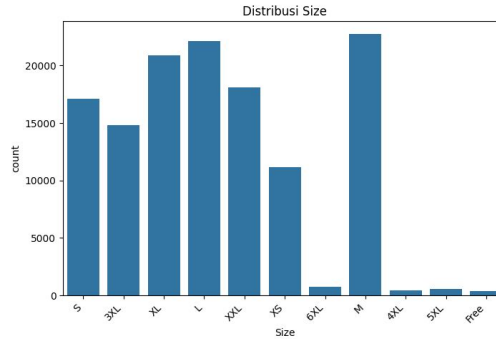
Nilai mendekati 0 menunjukkan bahwa tidak ada hubungan yang kuat antara Qty dan Amount (Harga)



# Business Insight (1)



Set dan Kurta adalah dua kategori produk yang paling banyak dilakukan pemesanan dengan total sekitar hampir 10000 baris pemesanan.



Jenis pakaian M, L, dan XL adalah 3 jenis pakaian dengan jumlah pemesanan terbanyak

ship-state	count
MAHARASHTRA	22260
KARNATAKA	17326
TAMIL NADU	11483
TELANGANA	11330
UTTAR PRADESH	10638
...	...
Mizoram	1
rajsthan	1
Punjab/Mohali/Zirakpur	1
Rajsthan	1
orissa	1

69 rows x 1 columns

State dengan pemesanan terbanyak adalah Maharashta, Karnataka, Tamil Nadu, Telangana



# Business Insight (2)

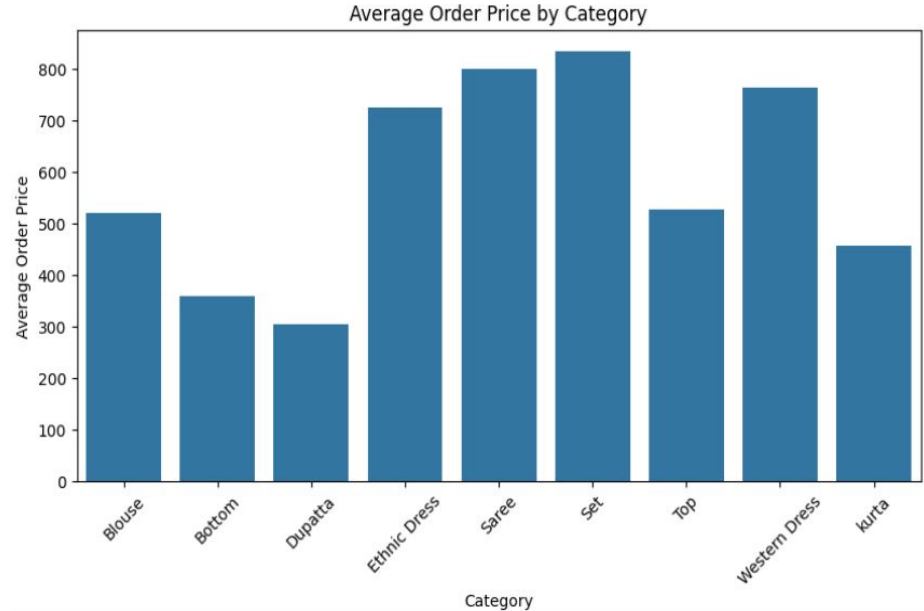
## Melihat average harga pesan per kategori produk

### Business Insight :

1. Kategori **Set**, **Saree**, dan **Western Dress** mungkin lebih diminati oleh pelanggan
2. Kategori **Dupatta** dan **Bottom** yang memiliki rata-rata harga lebih rendah

### Rekomendasi :

Fokus pada promosi dan pemasaran untuk kategori **Set**, **Saree**, dan **Western Dress**. Buat kampanye yang menonjolkan keunikan dan kualitas produk dalam kategori ini. Misalnya, gunakan influencer atau model untuk menunjukkan cara mengenakan produk dalam konteks gaya hidup yang menarik.



# Business Insight (3)

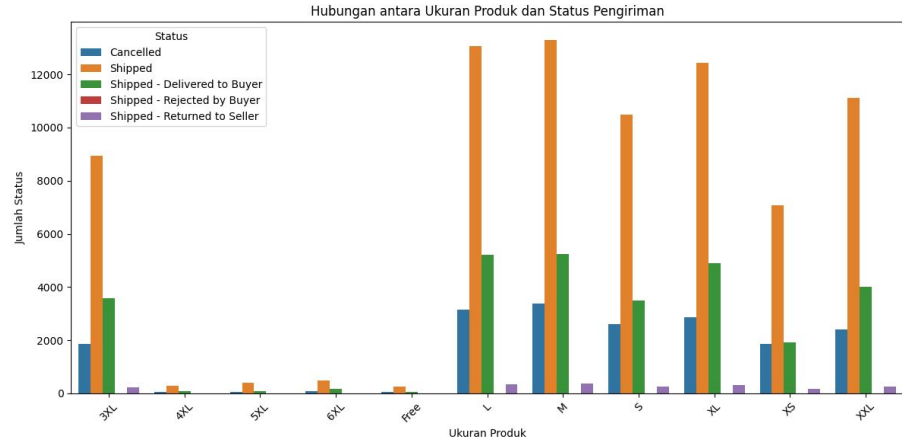


## Business Insight :

Ukuran produk yang memiliki jumlah pemesanan terbanyak seperti M, L, XL, XXL, S, dan XS memiliki jumlah status barang yang shipped dan diterima pembeli lebih banyak dibandingkan dengan jumlah status barang yang dikembalikan ataupun ditolak.

## Rekomendasi :

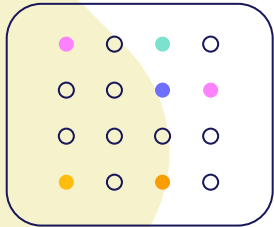
Pastikan stok untuk ukuran-ukuran ini selalu mencukupi dan pertimbangkan untuk meningkatkan variasi dalam ukuran, termasuk ukuran tambahan, untuk menarik lebih banyak pelanggan dan mengurangi kemungkinan kehabisan stok.



03.

# DATA PRE-PROCESSING

Data Cleansing & Feature  
Engineering



# ALUR DATA PRE-PROCESSING

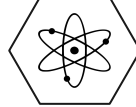
## Feature Selection

Change categorical data to numerical data



Drop Feature : Index, ship-postal-code, Sales Channel, Status

Create Feature : Is\_Weekend based On Date



## Standardization Data

Using Standard Scaler from scikit-learn to processing standardization



From ratio label 93.62:6.34:0.04,  
To ratio label 33.33:33.33:33.33  
With SMOTE



## Feature Encoding

One Hot Encoding Region dan Category

Label Encoding untuk Size

## Feature Extraction

**Missing Values** : Hapus

**No Duplicated Data**

**Outlier dipertahankan** karena qty besar atau harga tinggi adalah bagian alami dari transaksi di e-commerce

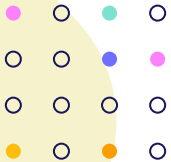
## Handle Imbalance



**04.**

# **MODELING & REKOMENDASI**

Melakukan pemodelan serta  
menentukan metriks evaluasi  
yang tepat



## Hasil Evaluasi Model

Nama Model	Recall (Train)	Recall (Test)
Logistic Regression	0.80	0.73
k-Nearest Neighbor	0.83	0.27
Random Forest	0.96	0.20
XGBoost	1.00	0.83

## Tuning Hyperparameter

Nama Model	Recall Train (Before)	Recall Train (After)	Recall Test (Before)	Recall Test (After)	Peningkatan (%)
Logistic Regression	0.80	0.88	0.73	0.84	15
XGBoost	1.00	1.00	0.83	0.95	14

Alasan pemilihan model:

**Logistic dan kNN** adalah pemodelan sederhana dan cocok untuk baseline, sementara

**Random Forest dan XGBoost** cocok untuk hubungan kompleks dan data tidak seimbang.

## Jumlah Data

Label 0	28754
Label 1	1958

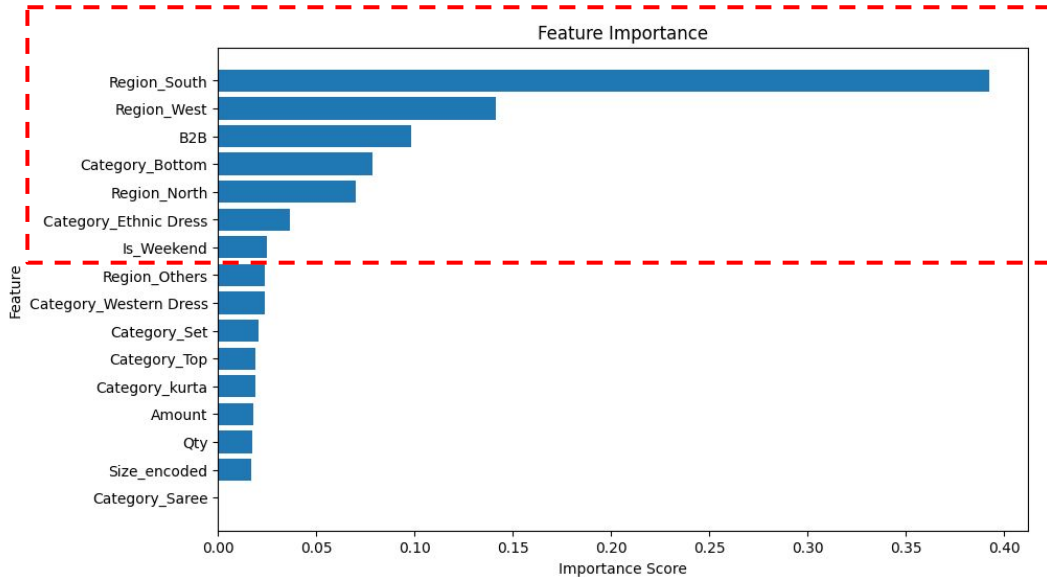
Dari keempat model yang telah dicoba, dipilih model **Logistic Regression dan XGBoost** yang akan dilakukan **tuning hyperparameter** untuk meningkatkan performa model dan mencari parameter optimal

**Model terpilih:** XGBoost

**Alasan :** Lebih robust terhadap outlier, memiliki recall tinggi

# Feature Importance

## XGboost (Model Terpilih)



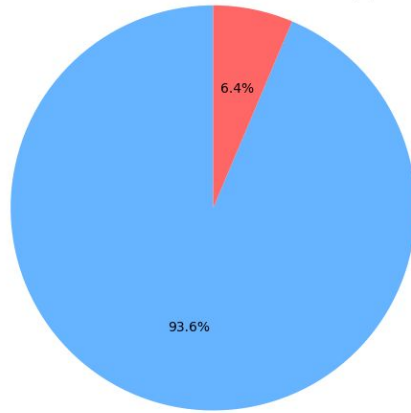
Wilayah (Region) terutama **West dan South** sangat mempengaruhi prediksi status paket

Kategori produk (seperti pakaian bawah atau pakaian etnik) memiliki tingkat **pengembalian lebih tinggi**.

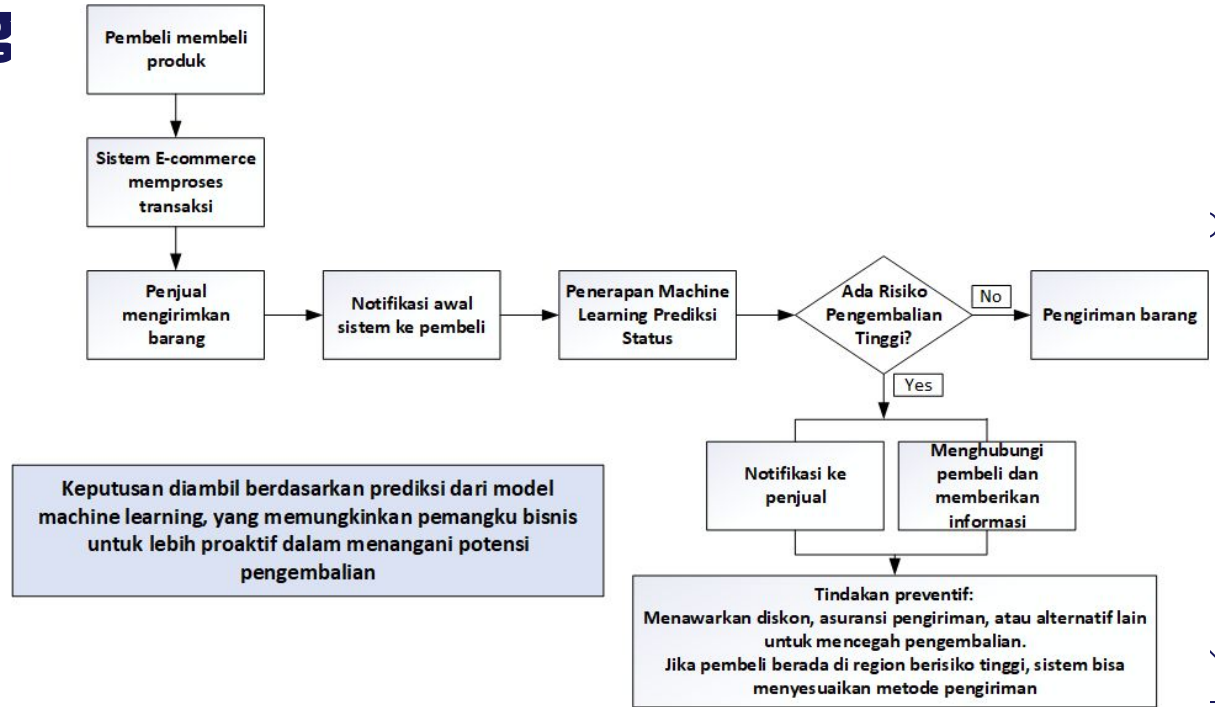
Waktu pengiriman (**akhir pekan**) mempengaruhi status paket, kemungkinan terkait dengan logistik atau jadwal pelanggan.

# Business Impacts on Machine Learning

Distribusi Status Pengiriman Barang  
Dikembalikan (1)



Diterima (0)





# Rekomendasi Bisnis

## Fokus Wilayah dengan Tingkat Pengembalian Tinggi:

Tingkatkan analisis logistik untuk wilayah seperti South dan West. Pertimbangkan optimalisasi jalur pengiriman dan komunikasi lebih baik dengan pelanggan di wilayah tersebut

## Optimalisasi Jadwal Pengiriman:

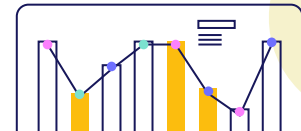
Hindari pengiriman di akhir pekan untuk wilayah atau kategori yang memiliki tingkat pengembalian tinggi. Berikan pilihan jadwal pengiriman yang lebih fleksibel kepada pelanggan.

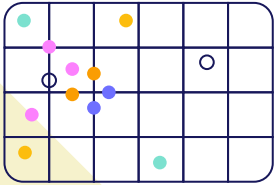
## Perbaikan untuk Kategori Produk Etnik dan Bottom:

Tawarkan panduan ukuran yang lebih jelas, deskripsi produk yang lebih detail, atau visualisasi produk untuk membantu pelanggan membuat keputusan yang tepat.

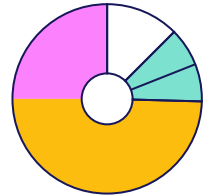
## Kampanye Edukasi Pelanggan:

Untuk wilayah dengan tingkat pengembalian tinggi, jalankan kampanye edukasi untuk pelanggan mengenai proses pengembalian, manfaat menerima paket, dan peningkatan komunikasi untuk memastikan kepuasan pelanggan.





# Terima Kasih!



CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution