

Final Project

Rakaplus

Meet our team

Gamal Abdul Naser

KETUA

Adji Putra Abriantero

Desy Rizky Amalia

Farrell Gaffar

Ferry Irawan

M Farhan Sidik

A decorative graphic at the bottom of the slide consisting of several overlapping, wavy lines in a light purple color, creating a sense of movement and depth.

Table of Content

1. Background

**2. Exploratory Data
Analysis (EDA)**

3. Data Pre-Processing

4. Modelling

5. Recommendation

Background

Rakaplus

E-commerce Shipping Data

Business Metrics = On Time Delivery (OTD) Rate

1

E-commerce RAKAPLUS ingin menyelesaikan permasalahan keterlambatan pengiriman yang dialami oleh customer, karena pengiriman penting dalam memberi kepuasan terhadap customer

2

Meningkatkan keputusan pembelian customer dalam e-commerce tersebut sehingga dapat mempertahankan customer bahkan meningkatkan jumlah customer

3

Mengembangkan suatu model untuk memprediksi barang yang dipesan oleh customer tiba dengan tepat waktu atau tidak dan mencari penyebab keterlambatan pada pengiriman

E-commerce Shipping Data

4

Jika model prediksi ketepatan waktu sudah selesai, dilanjutkan dengan saran rekomendasi memberikan notifikasi kepada customer apabila adanya keterlambatan pada pengiriman barang.

5

Business metrics yang digunakan yaitu OTD Rate

Exploratory Data Analysis (EDA)

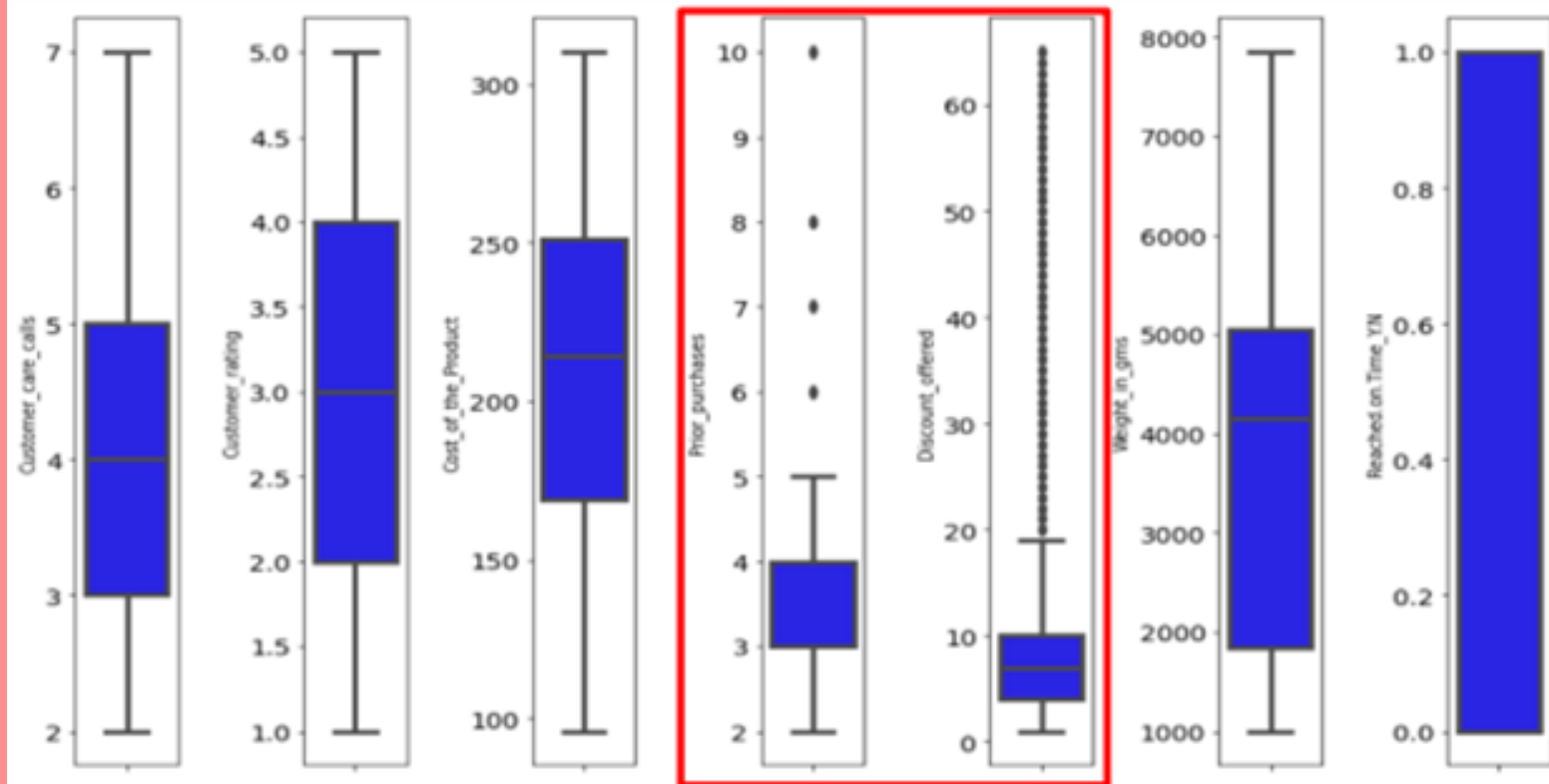
Rakaplus

EDA

```
df.isna().sum()
```

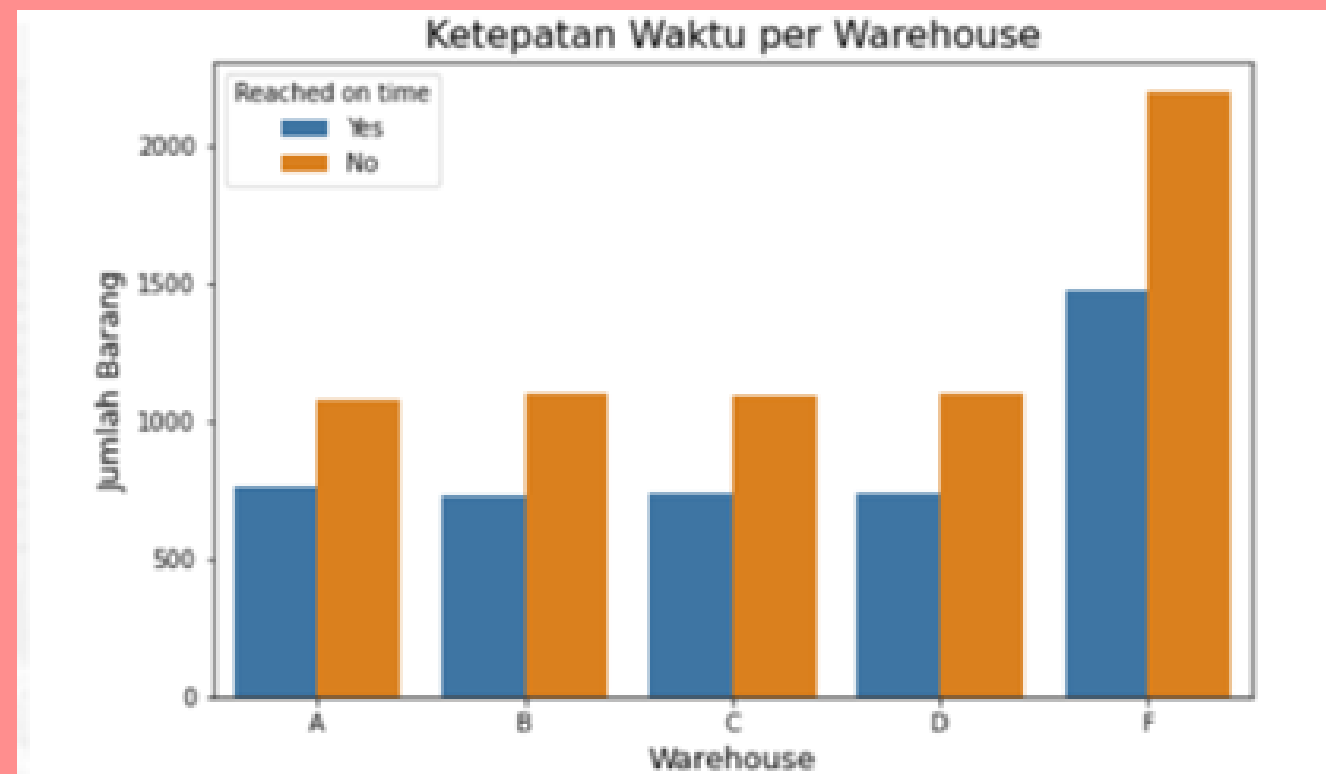
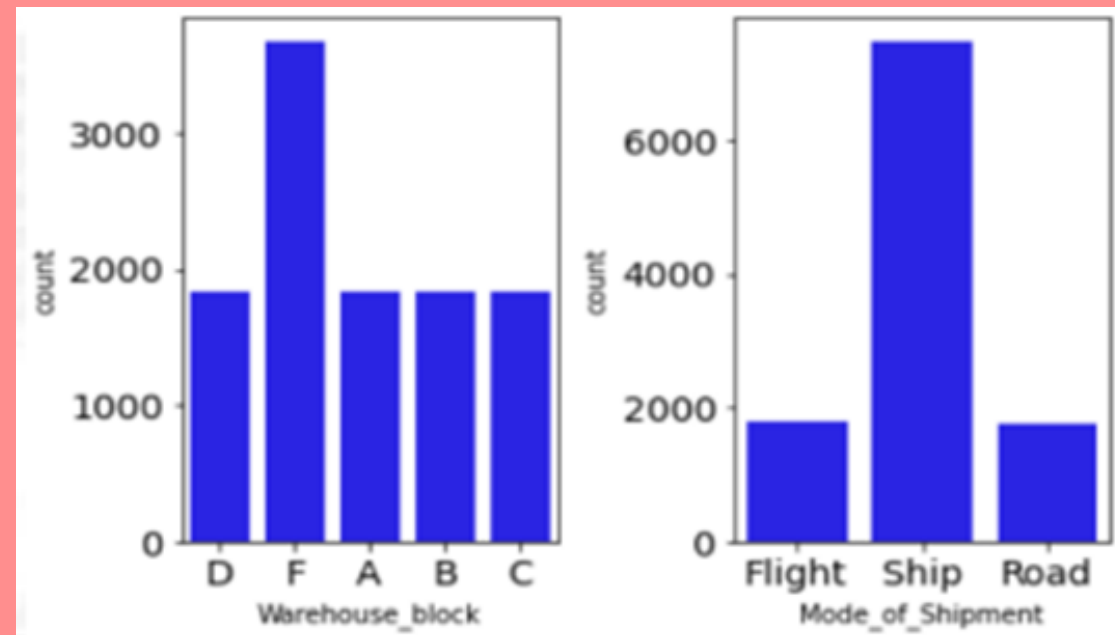
ID	0
Warehouse_block	0
Mode_of_Shipment	0
Customer_care_calls	0
Customer_rating	0
Cost_of_the_Product	0
Prior_purchases	0
Product_importance	0
Gender	0
Discount_offered	0
Weight_in_gms	0
Reached.on.Time_Y.N	0

dtype: int64



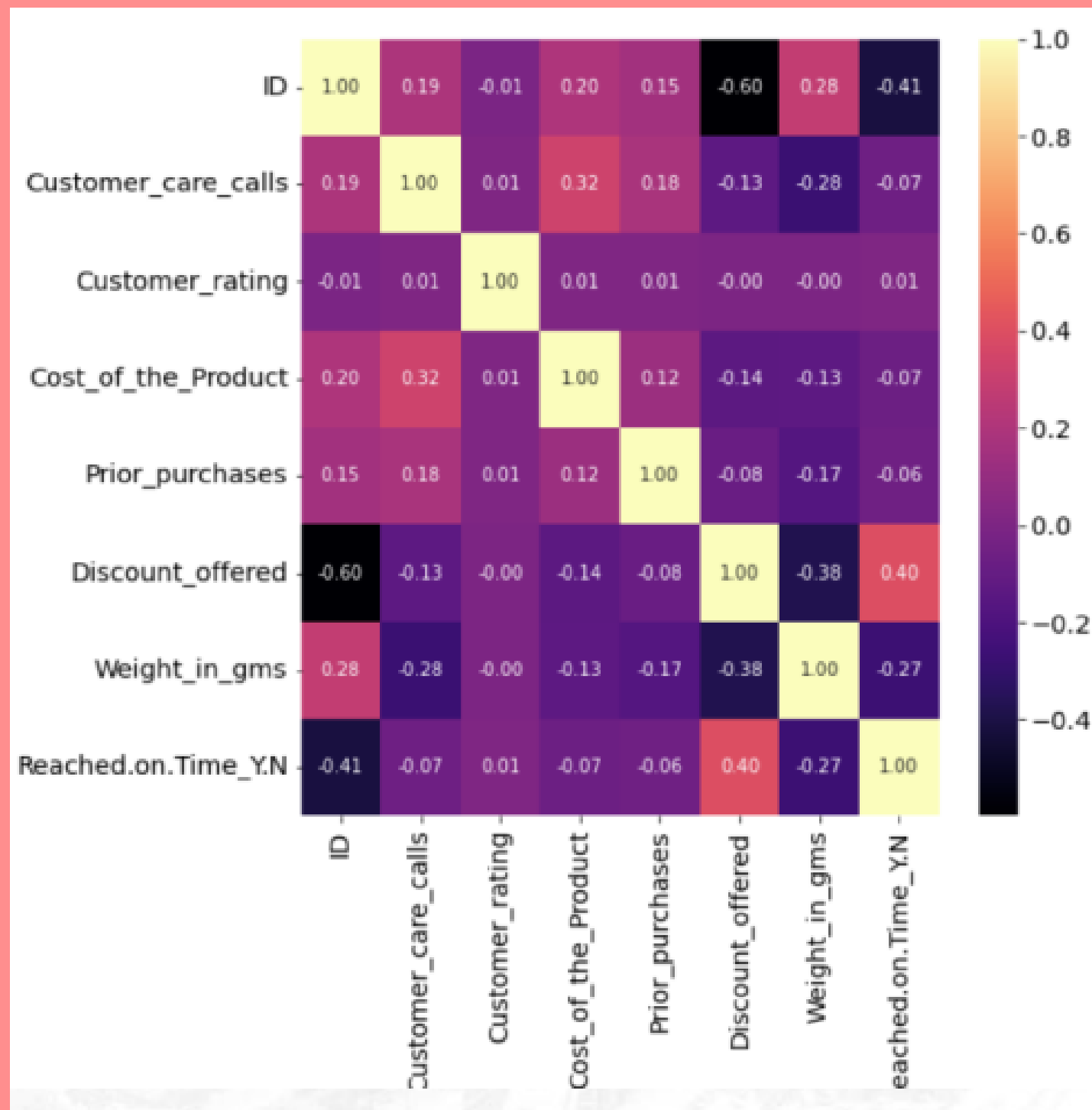
- Data berkaitan dengan Customer Satisfaction (Pengiriman)
- Fitur = “Reached on Time”
- Label = “Discount Offered” dan “Weight in gms”
- No Missing Value and Duplicated Data
- Distribusi Positive Skewed berada pada kolom “Discount Offered” dan “Prior Purchase”

EDA



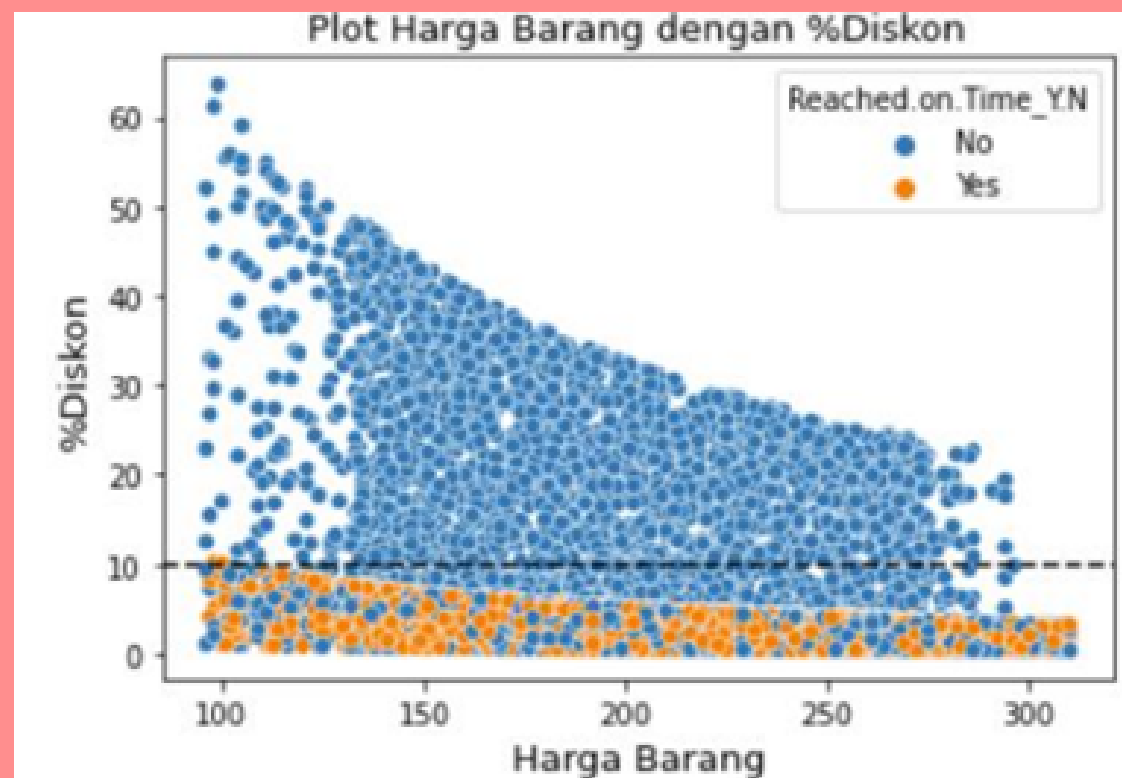
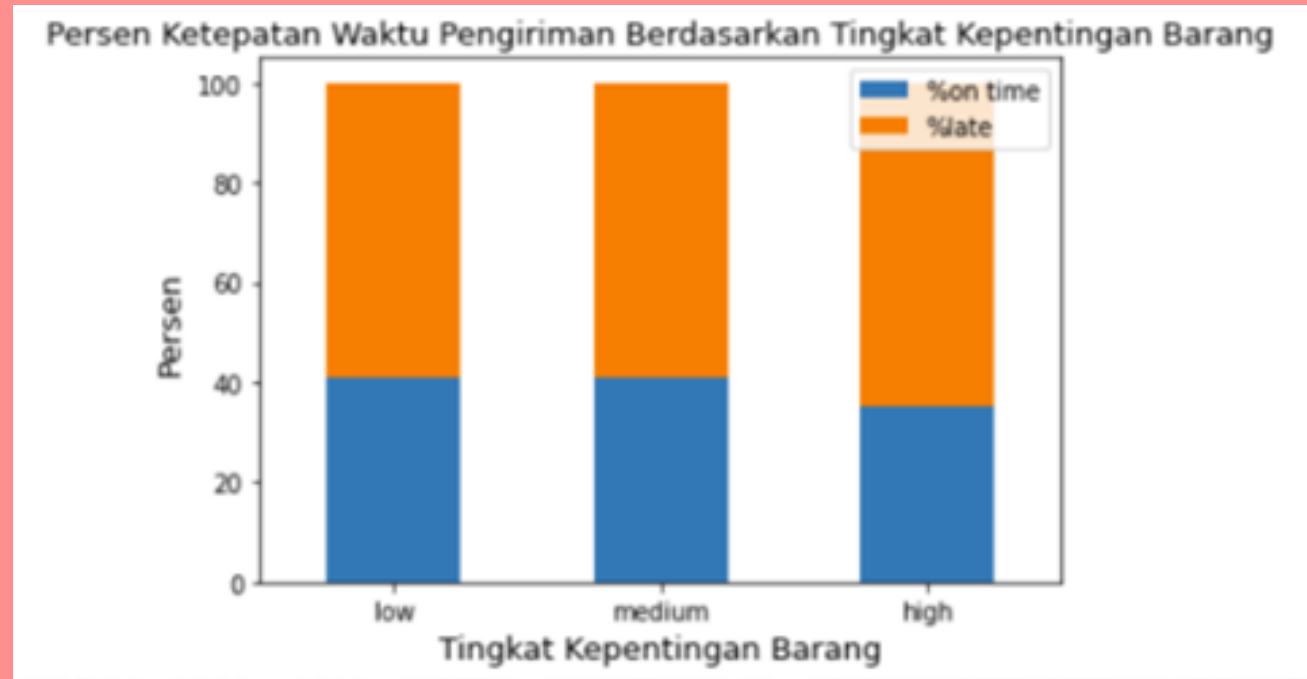
- Terdapat nilai dominan pada "Warehouse F" di Kolom Warehouse dan Mode "Ship" di Kolom Mode of Shipment
- Warehouse F memiliki jumlah barang paling banyak namun memiliki jumlah keterlambatan paling tinggi

EDA



- "Reached on Time" dan "Discount Offered" memiliki nilai korelasi Positif 0.40
- "Reached on Time" dan "Weight in gms" memiliki nilai korelasi Negatif -0.27

EDA



- Terjadi keterlambatan pada setiap Tingkat Kepentingan Barang (Tidak berkorelasi dengan Fitur)
- Keterlambatan mayoritas terjadi dengan barang yang memiliki diskon lebih dari 10%

Data Pre-Processing

Rakaplus

Data Pre-Processing

Handling Data

- ✓ No Missing Value
- ✓ No Duplicate Data

Handling Outlier

- ✓ Prior_purchases
Dapat di handle dengan IQR
- ✓ Discount_offered
Masih terdapat outlier -> log transformation

Feature Transformation

- ✓ Normalization
- ✓ Standardization
- ✓ Log Transformation:
 - ✓ Discount_offered.

Feature Encoding

- ✓ Label Encoding:
 - ✓ Product_importance
 - ✓ Gender
- ✓ One Hot-Encoding:
 - ✓ Warehouse_block
 - ✓ Mode_of_Shipment

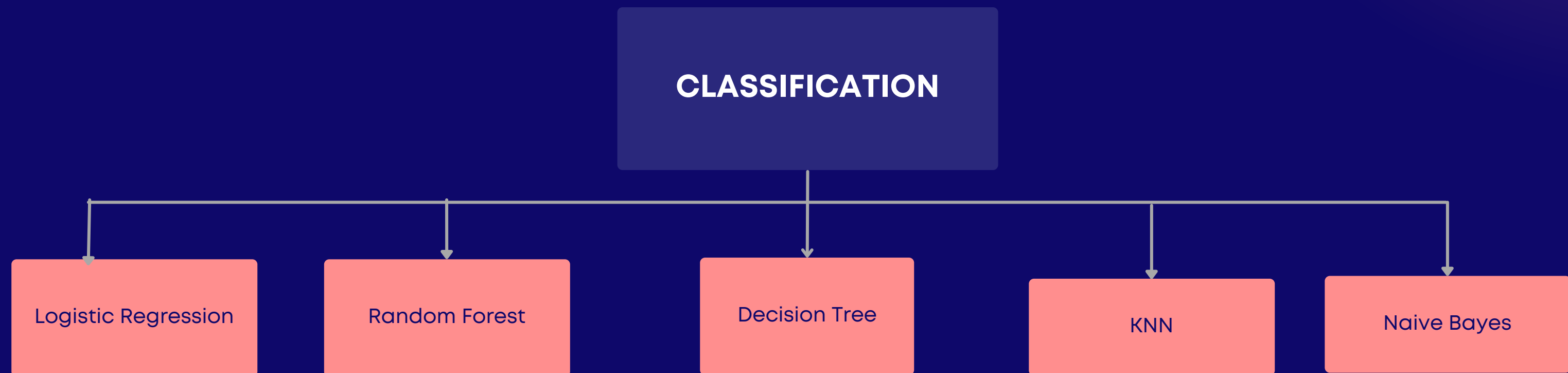
Drop Columns

- ✓ ID:
 - ✓ Warehouse_block
 - ✓ Mode_of_Shipment

Modelling

Rakaplus

MODELLING



Kita ingin membagi / mengklasifikasi proses pengiriman barang menjadi TEPAT WAKTU atau TELAT

Parameter Evaluasi Model

Aktual	Prediksi	Keterangan
Tepat Waktu	Tepat Waktu	True Positive (TP)
Telat	Telat	True Negative (TN)
Tepat Waktu	Telat	False Positive (FP)
Telat	Tepat Waktu	False Negative (FN)

Tepat Waktu : **Positif**

Telat : **Negatif**

Jadi, **False negative** :

Target yang nilai aktualnya positif, namun nilai diprediksi negatif.

GOOD, prediksi tepat

GOOD, prediksi tepat

NOT GOOD, perlu dihindari

BAD, harus dihindari

Mengapa FN harus dihindari di kasus ini?

When my online order
hasn't arrived



Customer diinfokan bahwa
barang akan diantar tepat waktu



Aktualnya barang akan telat
diantar



Customer kecewa dan memberi review buruk

Model	Accuracy	Precision	Recall
Logistic Regression	0.64	0.69	0.73
Decision Tree	0.65	0.70	0.72
Random Forest	0.66	0.76	0.63
KNN	0.64	0.71	0.67

Evaluasi Model

Skor Sebelum Tuning

Nilai-nilai yang perlu diperhatikan:

- **Accuracy** (Ketepatan memprediksi / TP dan TN)
- **Precision** (Ketepatan menghindari FP)
- **Recall** (Ketepatan menghindari FN)
- **AUC** (Kombinasi ketiga nilai)

Skor antara masing-masing model tidak berbeda signifikan.

Evaluasi Model

Skor Setelah Tuning

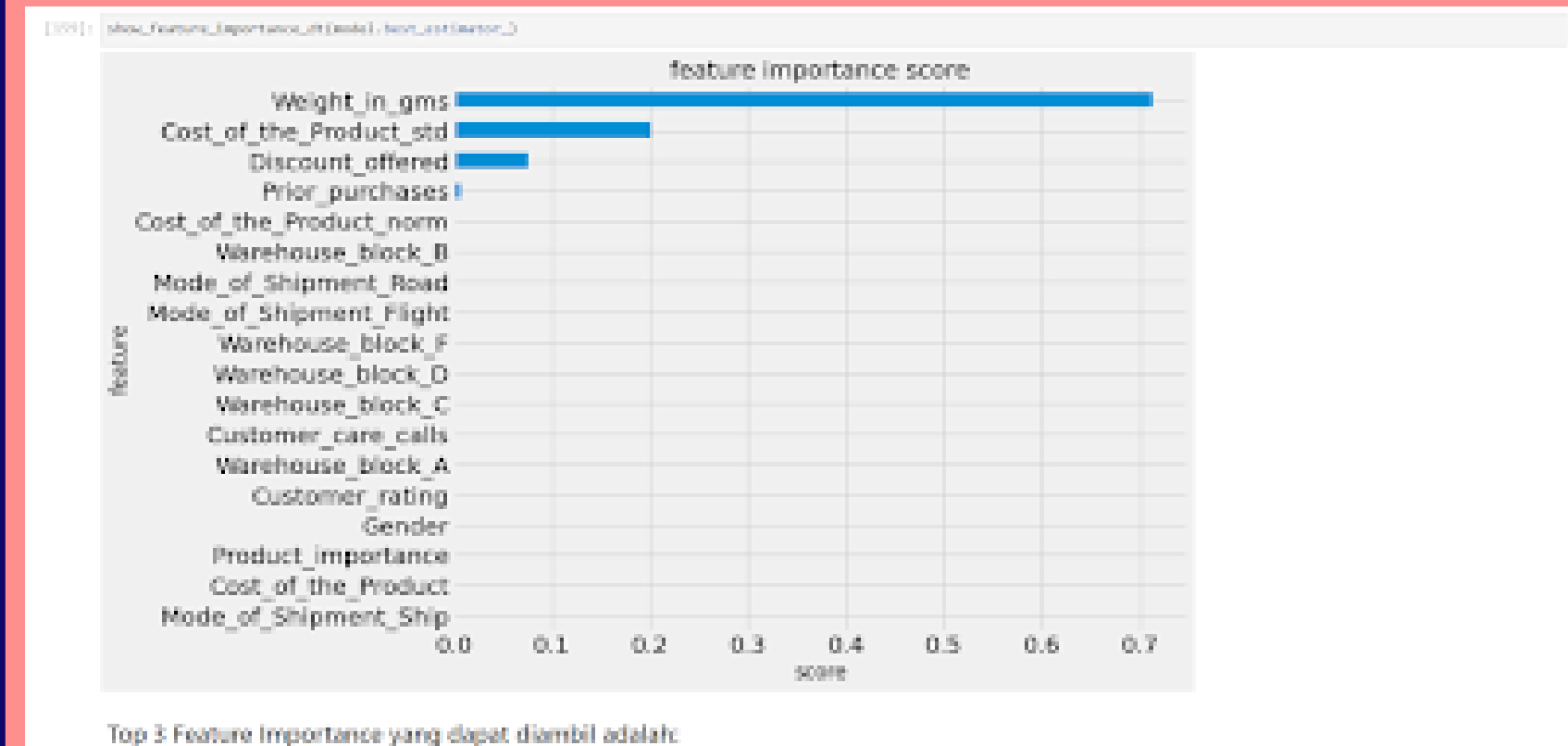
Secara overall, Decision tree dengan tuning menunjukkan hasil paling baik sehingga dipilih sebagai model untuk prediksi.

Model	Accuracy	Precision	Recall
Logistic Regression	0.59	0.59	0.75
Decision Tree	0.69	0.96	0.65
Random Forest	0.66	0.76	0.69
KNN	0.63	0.69	0.69

Feature Importance

3 fitur paling penting / utama yang digunakan model untuk prediksi adalah:

- Berat barang
- Harga barang
- Diskon yang diberikan

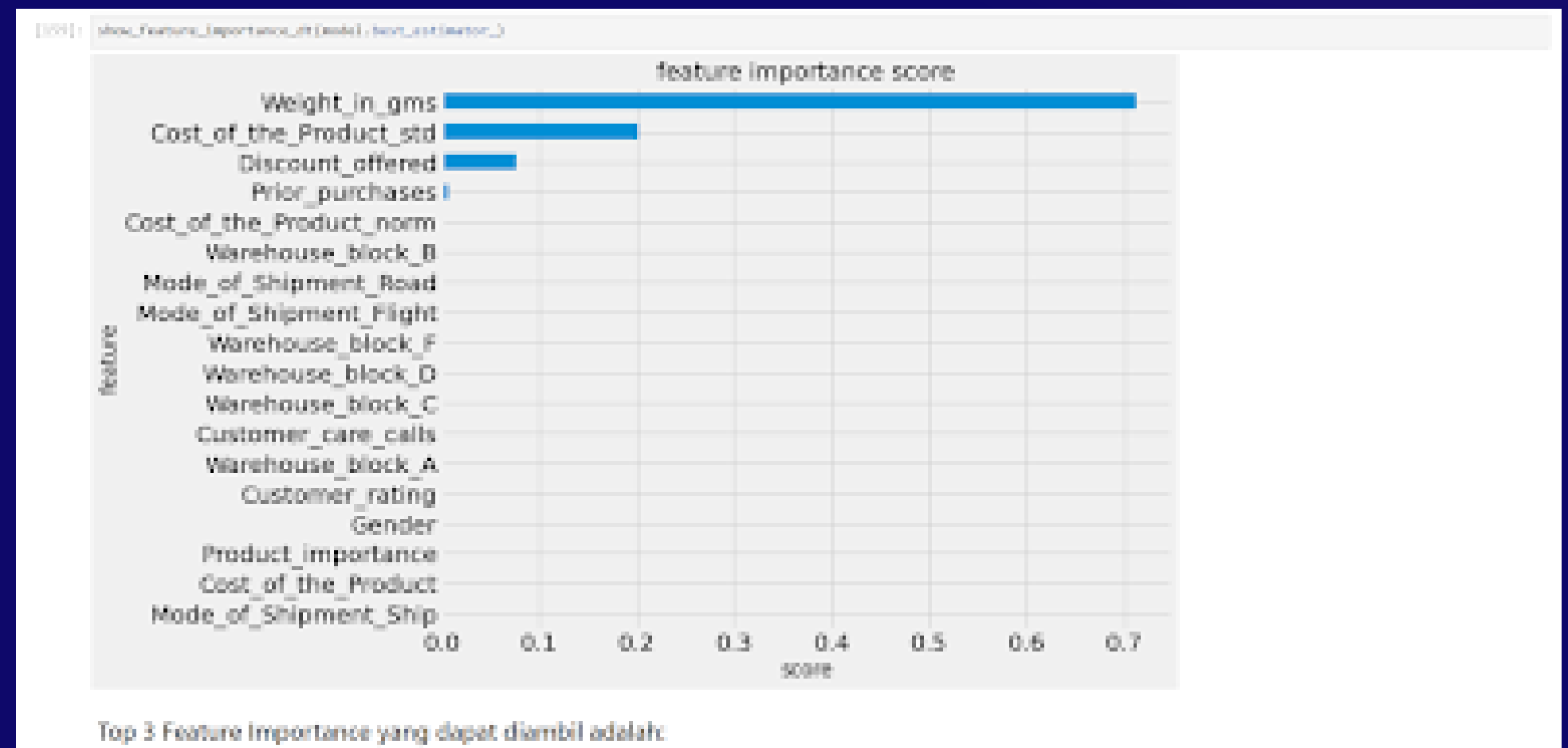


Recommendation

Rakaplus

Business Insight

Untuk feature importance yang paling mempengaruhi keterlambatan pengiriman barang terletak pada fitur 'Berat Produk'. Dapat dilihat pada proses data preprocessing yang sudah dilakukan, terlihat bahwa nilai korelasi antara feature 'Berat Produk' dengan target 'Keterlambatan' memiliki nilai korelasi negatif. Jadi dapat disimpulkan bahwa berat produk mempengaruhi proses pengiriman.



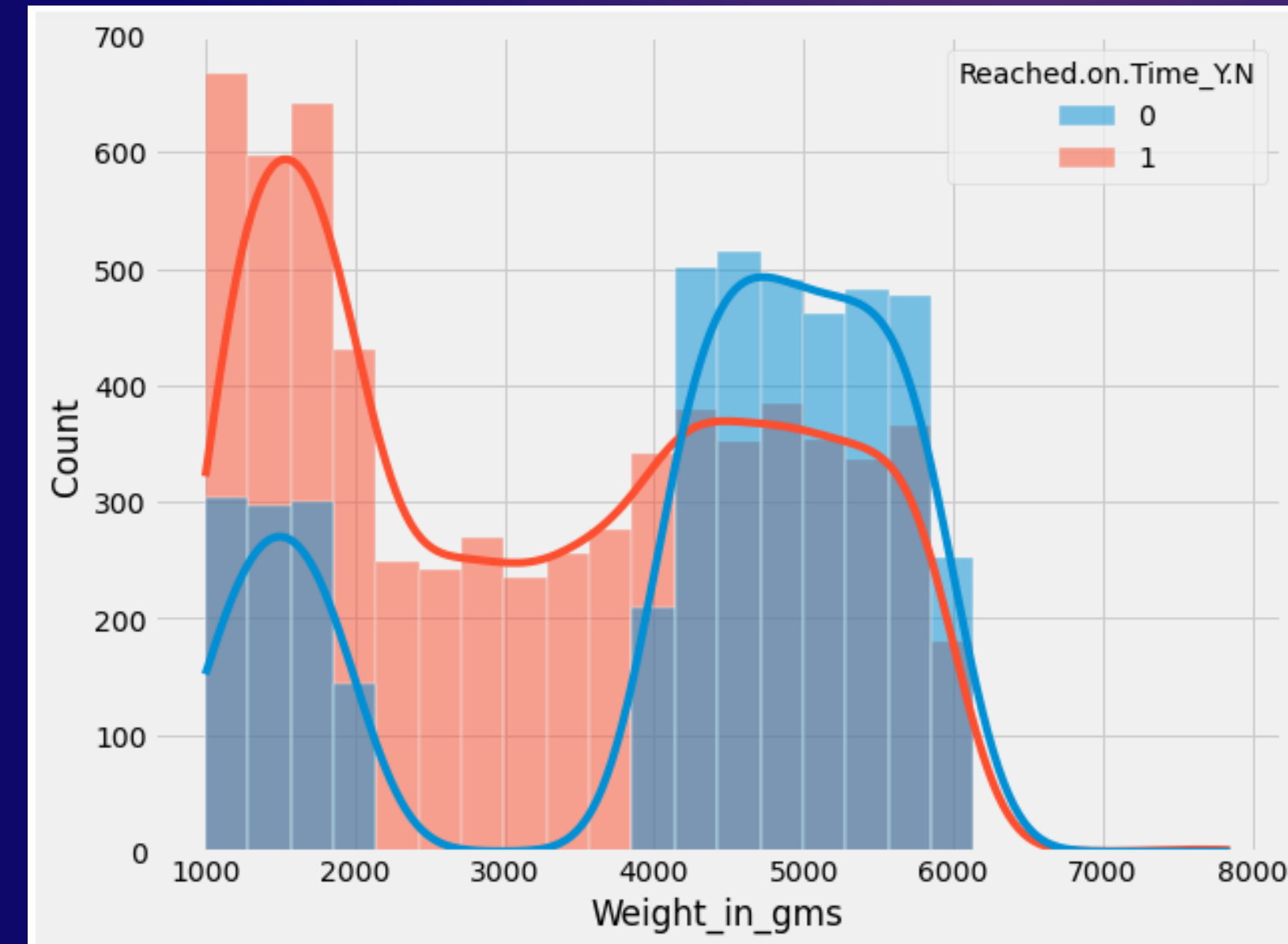
Business Insight

Weight in grams,

Sebagai *Data Analyst* kami dapat mengkategorikan prioritas suatu barang yang dibeli oleh *customers*, sehingga apabila sudah mengetahui kategori suatu barang pesanan melalui berat/bobot (gram) suatu barang maka proses pengiriman dapat disesuaikan dengan moda pengiriman.

Hal tersebut juga dapat terealisasi dengan baik apa bila menambahkan feature berupa 'Jarak Lokasi Pengiriman', agar dalam menentukan mode pengirimannya dapat sesuai dengan jenis barang dan jarak lokasi pengiriman.

Apabila suatu barang telah diketahui kategori berat dan lokasi pengirimannya, maka setelahnya dapat dengan mudah merekomendasikan mode jenis pengiriman yang digunakan (*Flight/Ship/Road*).



Pengiriman barang dipastikan terlambat apabila berat produk antara 2-4 kg.

Business Recommendation

- Membuat estimasi waktu pengiriman barang ke lokasi *customers* sebagai pencegahan keterlambatan.
- Mengacu pada *feature "Prior Pirchases"*, maka kami merekomendasikan untuk membuat program peringkat kepada *customers* sehingga pemberian diskon kepada *customers* akan lebih efektif dan tepat sasaran.
- Memberikan notifikasi pelacakan pengiriman barang kepada *customers* sehingga dapat mengurangi *customer care calls*.

Thank you!

Rakaplus