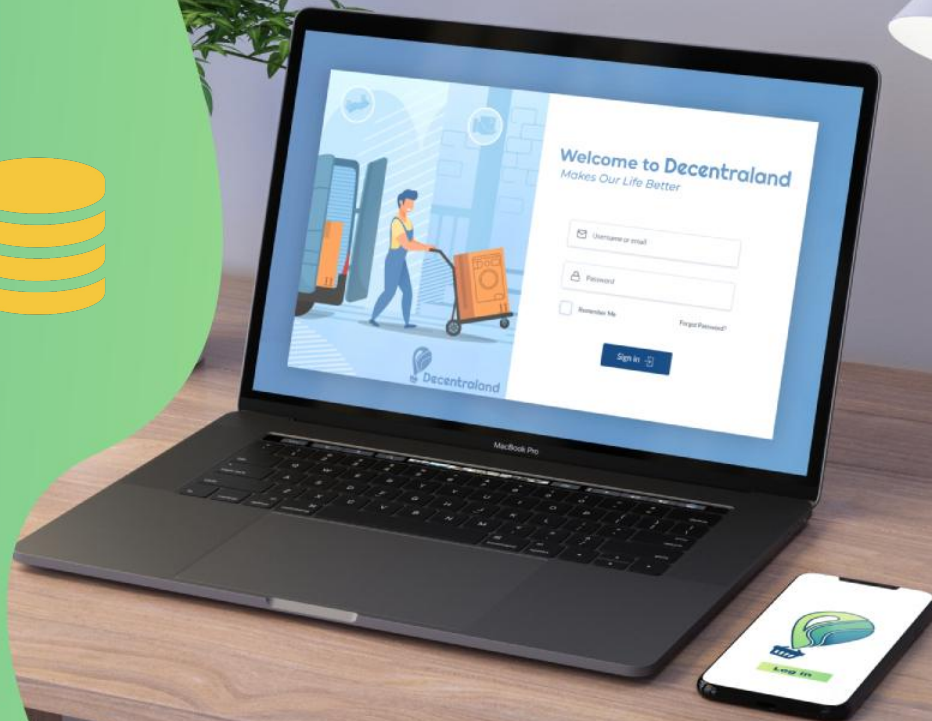


Decentraland E-Commerce Shipping Data



PREDICTING LATE SHIPMENT MODEL

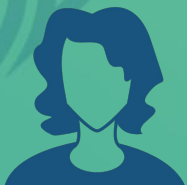


DATA SCIENTIST

EXTERNAL CONSULTANT



Gerry Chandra



Ahmad Ilham H.



Dharma Setiawan



Fikri Diva S.

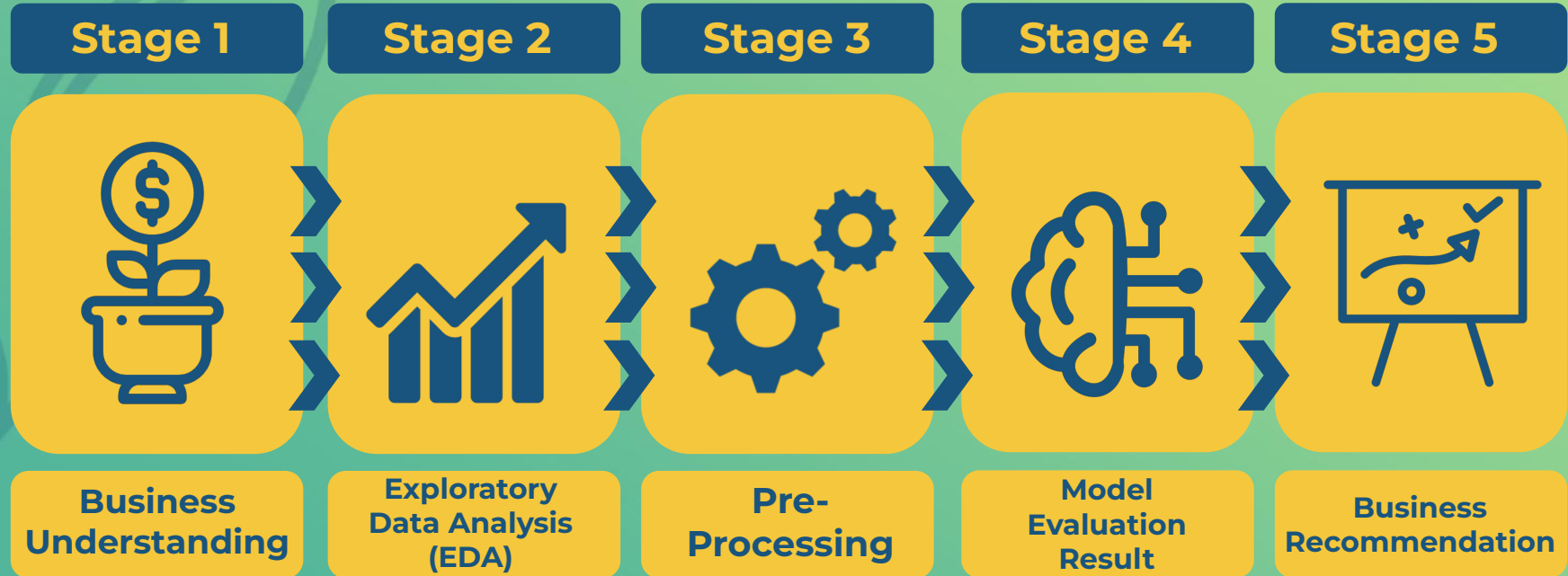


Muhammad Farhan A.



Ilham Ibnu A.

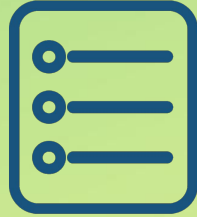
WORKFLOW



Stage 1

Business Understanding

CONTENT LIST



- Business Background
- Problem Statement
- Goal dan Objective



Business Background

Decentraland adalah sebuah perusahaan e-commerce yang menjual berbagai produk elektronik.

Sebagai perusahaan data science consultant independen, kami di-hire oleh **Decentraland** untuk mengolah suatu data set shipment. Kami diminta untuk mencari solusi bagaimana cara mengatasi permasalahan yang mereka punya melalui data set tersebut.

[Decentraland E-Commerce Shipping Data](#)



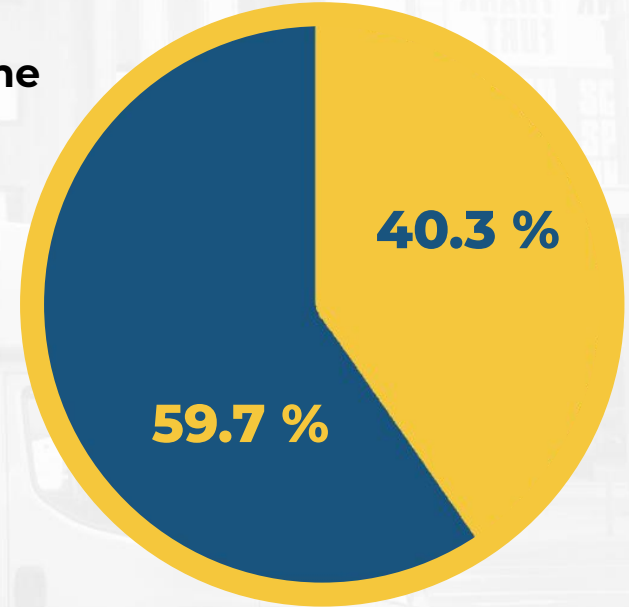
Click this link to
see the data set



Problem Statement

Data set E-Commerce Shipping menunjukkan bahwa dari 10.999 sampel transaksi pembelian online, **59.7%** transaksi diantaranya masih terjadi **keterlambatan** waktu pengiriman barang.

Keterlambatan yang tinggi ini tentunya berkaitan dengan **customer happiness dan potential revenue**. Customer akan mendapatkan experience yang kurang baik dalam menggunakan layanan **Decentraland**. Keterlambatan ini juga menyebabkan company mengalami potential revenue loss.



**Late Shipment
Percentage**

Mengapa Late Shipment yang Tinggi Harus Diperhatikan?



Late Shipment

Hasil penelitian Rajendran (2020) terhadap 1000 lebih review online yang dilakukan customer terhadap 4 perusahaan shipping terbesar, yaitu DHL, FedEx, UPS, dan USPS menunjukkan bahwa Late Deliveries merupakan salah satu topik teratas yang muncul. Topik-topik tersebut merupakan fitur penjelas mengapa terjadi dissatisfaction. [Link Sumber](#)



Low Rating

Ketika customer merasa tidak puas, mereka akan memberikan low rating. Hasil penelitian Posselt dan Gerstner (2005) yang menguji variabel target overall rating menggunakan model OLS menemukan hasil bahwa On Time Shipment berkorelasi positif terhadap target dengan koefisien 0.37. [Link Sumber](#)



Stop Shopping

Survey yang dilakukan Voxware terhadap 500 responden pada tahun 2020 menunjukkan hasil bahwa 30% responden customer yang menerima kiriman telat tidak akan melakukan pesanan lagi. Persentasenya meningkat dari penelitian yang sama pada tahun 2018 (23%). [Link Sumber](#)



Potential Revenue Loss

Setelah pelanggan berhenti berbelanja, penghasilan perusahaan akan menurun. Menurut penelitian MHL news & Last Mile, 52% customers mengharapkan refund maupun voucher diskon untuk mengurangi customer pain point. [Link Sumber](#)

Objective, Business Impact, Goal

Goal yang ingin dicapai adalah mengurangi persentase keterlambatan sebesar 20 persen, meningkatkan customer rating sebesar 10 persen, dan potential revenue sebesar 20-25 persen

Goal

Rekomendasi atau solusi bisnis.
Metrics yang terdampak: On Time Rate, Customer Rating, dan Potential Revenue Loss

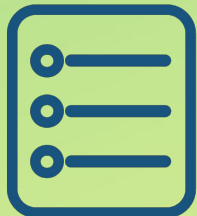
Business Impact

Membuat predictive model yang mampu memprediksi ketepatan waktu pengiriman barang dan mencari tahu fitur-fitur apa yang paling besar mempengaruhi suatu pengiriman barang akan telat/tidak.

Objective



CONTENT LIST



- General Info
- Distribusi Variabel Target
- Korelasi Fitur dan Target



Exploratory Data Analysis (EDA)

General Info



10.999

Rows

6 Numerical

10

Feature

1 Target

0

Duplicated

0

Missing Value

4 Categorical

Customer
Care Calls

Prior
Purchases

Customer
Rating

Discount
Offered

Cost of The
Product

Weight
in gram



Late Shipment

Warehouse
Block

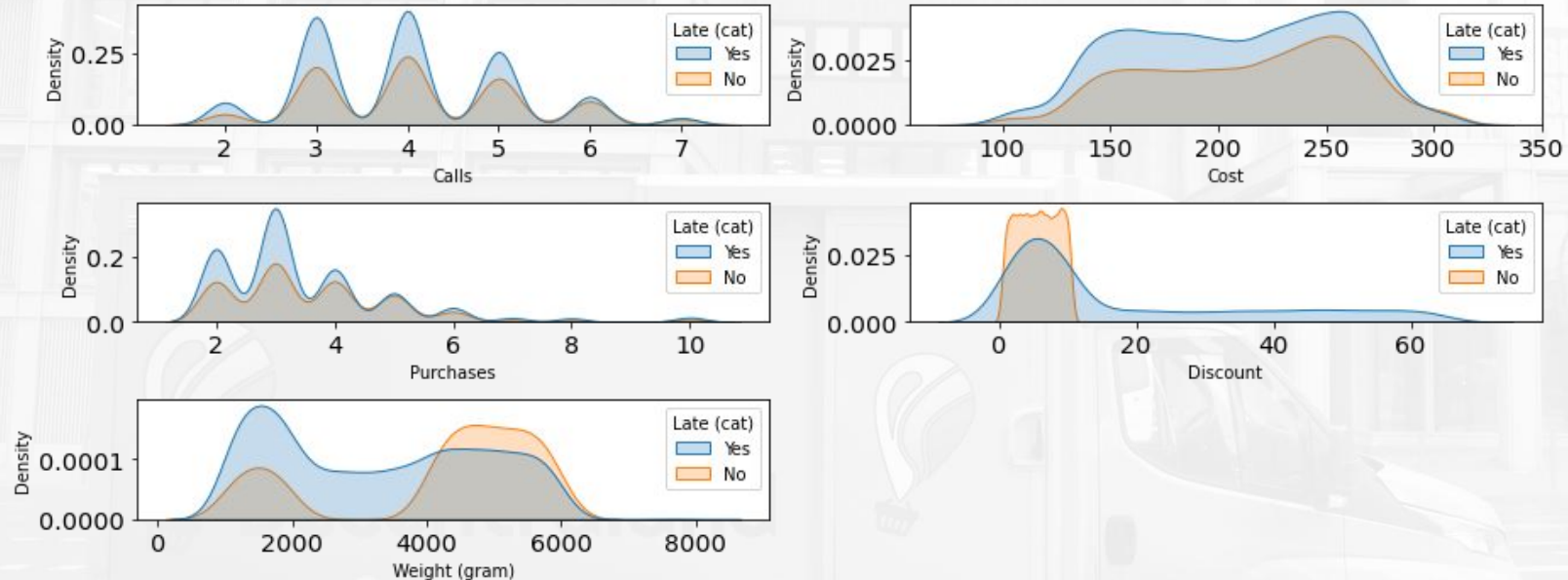
Product
Importance

Mode of
Shipment

Gender

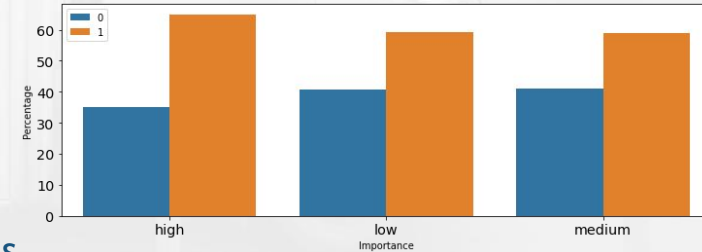
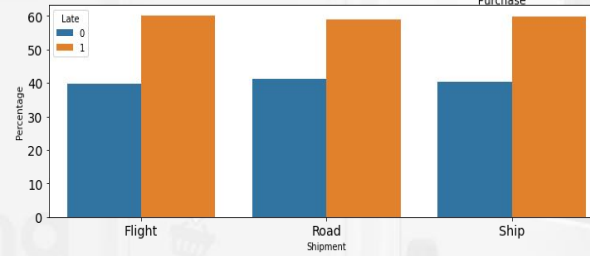
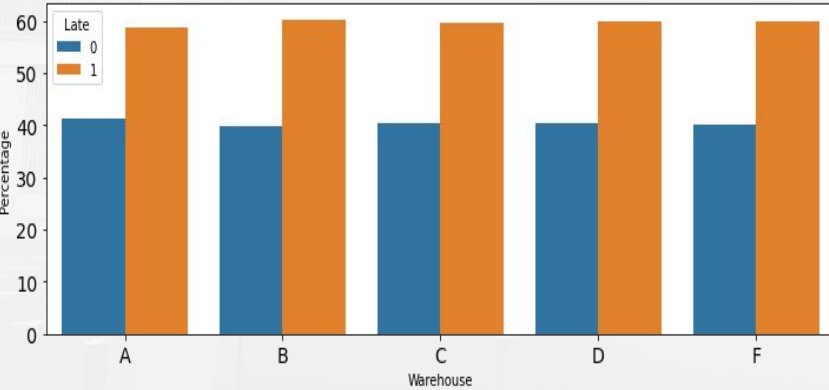
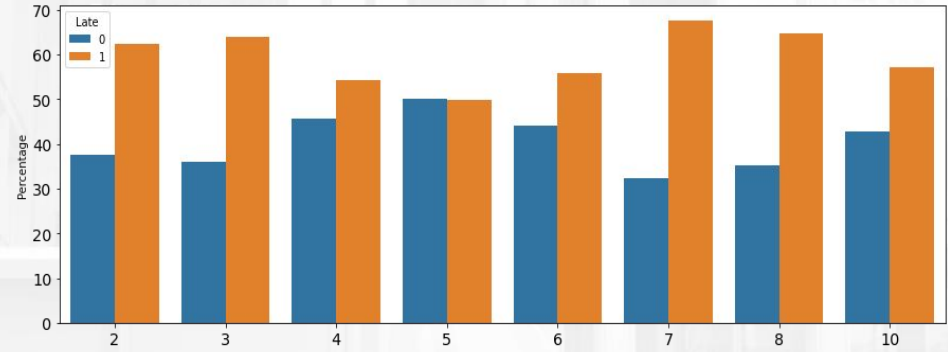
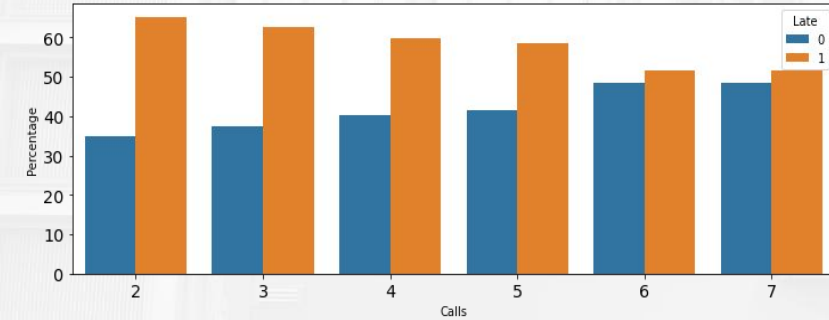
Exploratory Data Analysis (EDA) Univariate Analysis

Distribusi Data Variabel Numerik



- Variabel Calls dan Cost sudah cukup simetrik distribusinya mendekati distribusi normal (mean dan median tidak berbeda jauh).
- Variabel Discount pola persebarannya membentuk positively skewed (mean>median), sedangkan variabel Weight (gram) membentuk negatively skewed (median>mean).

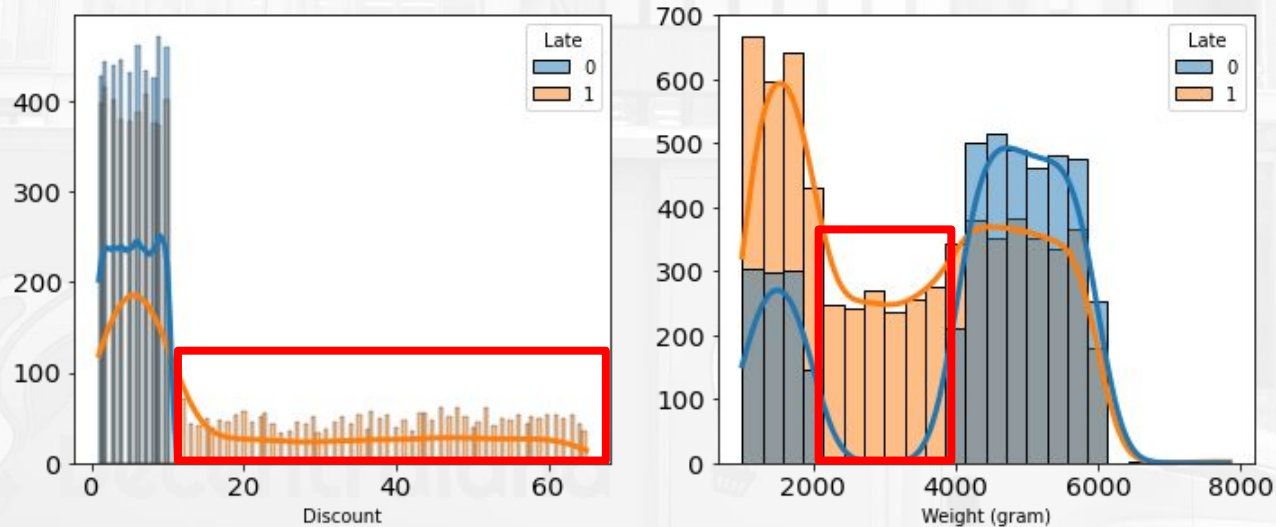
Fitur yang Tidak Mempengaruhi Keterlambatan



Customer care calls, shipment, warehouse block, product importance, dan prior purchase memiliki ratio yang kurang lebih sama. Maka dari itu, kami tidak menganalisis lebih lanjut. Permasalahan keterlambatan tidak disebabkan oleh fitur-fitur diatas

Fitur-Fitur yang Berpengaruh terhadap Keterlambatan

Hubungan variabel target dengan variabel numerik



- Discount yang lebih dari 10 persen terkonfirmasi terlambat. Hal ini bisa mengindikasikan beberapa bulan dimana promo diberikan besar-besaran sehingga pengiriman mengalami keterlambatan.
- Berat barang yang berkisar 2-4 kg terkonfirmasi terlambat. Hal ini mengindikasikan bahwa ada kemungkinan banyak customer membeli barang-barang kategori ringan.

Exploratory Data Analysis (EDA) Multivariate Analysis



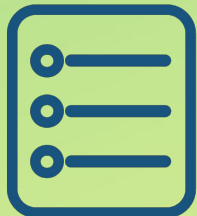
Feature Discount dan Weight adalah 2 fitur yang paling berkorelasi dengan feature target yaitu Late.

Tidak ada feature yang redundant di dataset ini.

Stage 3

Pre- Processing

CONTENT LIST



- Outlier Handling
- Feature Encoding
- Feature Engineering

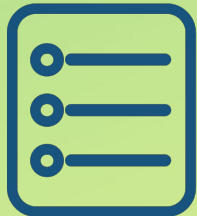


Data Treatment	Dataset 1 (n=10.999)	Dataset 2 (n=10.642)	Dataset 3 (n=8.790)
Warehouse Block	One Hot Encoding, Drop Block F		
Mode of Shipment	One Hot Encoding, Drop Moda Pengiriman Ship		
Customer Care Calls	Standarisasi		
Customer Rating	Drop, agar tidak data leaked		
Cost of The Product	Standarisasi		
Prior Purchase	Standarisasi		
Product Importance	Standarisasi dan Label Encoding		
Gender	Drop, tidak make sense memprediksi barang telat/tidak		
Discount Offered	Standarisasi	Standarisasi dan Remove Outlier Z score	Standarisasi dan Remove Outlier IQR
Weight in gram	Standarisasi		

Stage 4

Model Evaluation Result

CONTENT LIST



- Alternatif Treatment Experimental
- Algoritma yang dicoba
- Matrix Scoring yang Dipakai
- Hasil dari Model Terbaik



- Logistic Regression
- k-Nearest Neighbor
- Decision Tree
- Random Forest
- AdaBoost
- XGBoost



Dari hasil eksperimen kami terhadap 3 dataset, kami memperoleh hasil bahwa prosedur **Boosting menggunakan AdaBoost** menghasilkan performa terbaik (AUC Data Training lebih besar daripada AUC Data Testing dengan gap perbedaan yang dekat dan nilainya relatif lebih tinggi dibandingkan hasil di algoritma lain).

Scoring Matrix

Primary: ROC AUC

Secondary: Average Precision

Dataset 1	Logistic Regression		k-Nearest Neighbor		Decision Tree		Random Forest		Adaboost		XGBoost	
	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision
Train	0.734	0.678	0.748	0.858	0.746	0.856	0.740	0.855	0.752	0.851	0.742	0.856
Test	0.706	0.832	1	1	0.809	0.885	0.900	0.936	0.747	0.852	0.850	0.909

Lalu, bagaimana perbandingan hasil penerapan adaboost pada dataset yang lain?

Model Evaluation Hasil Model Terbaik

Dataset 1: Dataset original

Dataset 2: Dataset yang menggunakan remove outlier (Z-score)

Dataset 3: Dataset yang menggunakan remove outlier (IQR)

Dataset 1 boosting hyperparameter Adaboost menghasilkan performa terbaik (AUC Data Training lebih besar daripada AUC Data Testing dengan gap perbedaan yang dekat dan nilainya relatif lebih tinggi dibandingkan hasil di dataset 2 dan 3).

	Dataset 1		Dataset 2		Dataset 3	
	AUC	Average Precision	AUC	Average Precision	AUC	Average Precision
Train	0.752	0.851	0.748	0.845	0.629	0.667
Test	0.747	0.852	0.741	0.838	0.615	0.655

Confusion Matrix

Hasil Model Terbaik

Adaboost Dataset 1

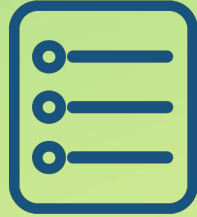
Actual Label	Predicted Label	
	Late (1)	On time (0)
Late (1)	TRUE POSITIVE (TP) 1223 37.06%	FALSE NEGATIVE (FN) 765 23.18%
On time (0)	FALSE POSITIVE (FP) 296 8.97%	TRUE NEGATIVE (TN) 1016 30.79%

Late = 1 (Positive), On time = 0 (Negative)

Stage 5

Business Recommendation

CONTENT LIST



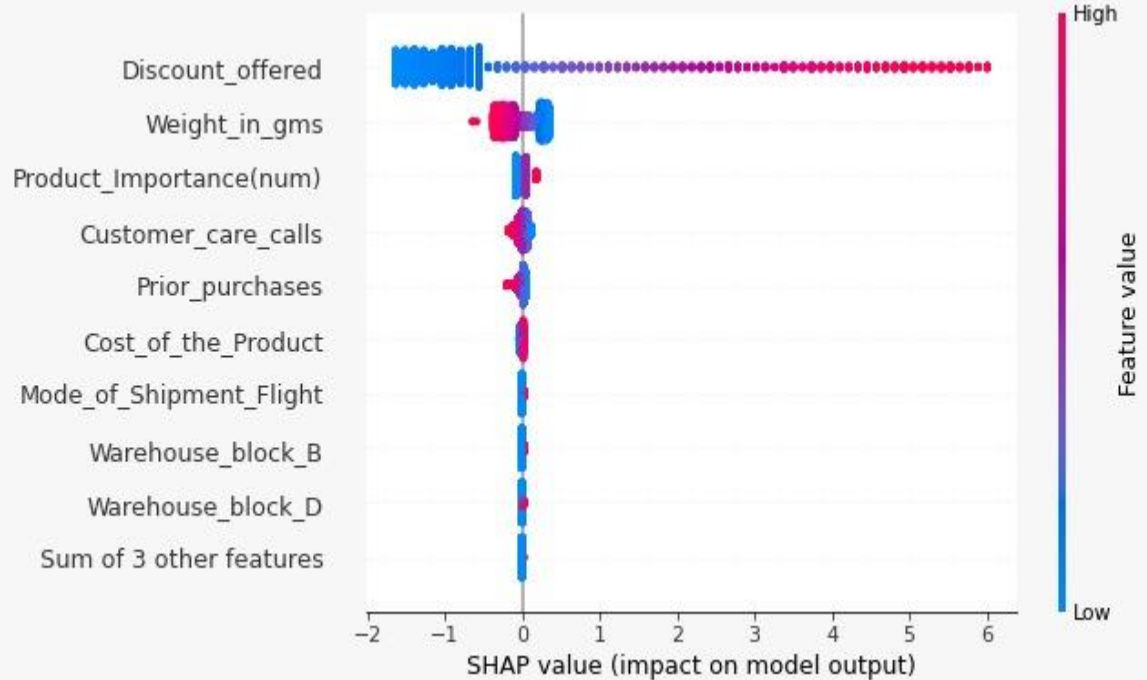
- Comparison
- Business Insight
- Business Recommendation



Fitur yang Harus Diperhatikan

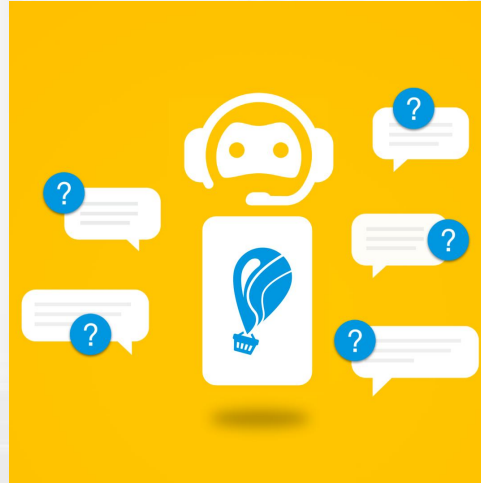
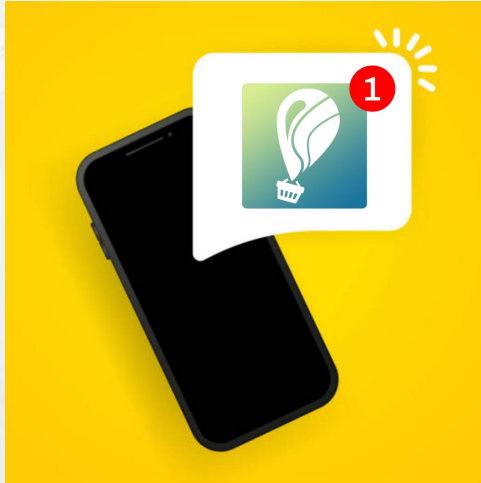
Dengan menggunakan shap value didapatkan bahwa ada 2 fitur yang paling perlu untuk segera diambil tindakan. Kedua fitur tersebut adalah:

1. Discount_offered
2. Weight_in_gms



Business Insight & Recommendation

1. Banyak sekali delays yang terjadi pada order dengan diskon diatas 10%. Hal ini dapat diantisipasi dengan memberikan notifikasi kepada pelanggan ketika terjadi diskon yang besar seperti “Harbolnas” yang memungkinkan barang tidak terkirim dengan tepat waktu. Kami juga menyarankan untuk membuat chatbot khusus untuk menangani pertanyaan-pertanyaan dari customer. Chatbot ini dapat menjawab pertanyaan-pertanyaan dari customer. Customer akan tetap puas karena pertanyaan-pertanyaan mereka terjawab.



Business Insight + Recommendation

2. Berat barang antara 2-4 kg mengalami keterlambatan yang signifikan. Pihak decentraland perlu untuk memberi notifikasi kepada pembeli di rentang tersebut seperti barang akan terlambat x hari sebelum pelanggan menekan tombol buy. Selanjutnya, menurut Desy (2019), pihak e-commerce perlu memberikan voucher sebagai pengganti keterlambatan. Pihak e-commerce juga dapat memberikan promo voucher/kupon apabila barang yang terprediksi on time ternyata late. Hal ini dapat mengurangi customer pain point dan membuat customer tetap puas dan tetap berbelanja di e-commerce Decentraland. Percobaan voucher ini bisa dibilang lumayan risky dan membuat customer ketergantungan sehingga kami memutuskan untuk melakukan AB Testing.



Profit Calculation



Key Assumption:

Revenue Decentraland = 3%

Voucher = \$25

False Negative = 765

Profit = Revenue-Cost

Gross Profit = Profit/Revenue * 100%

There is no fixed cost

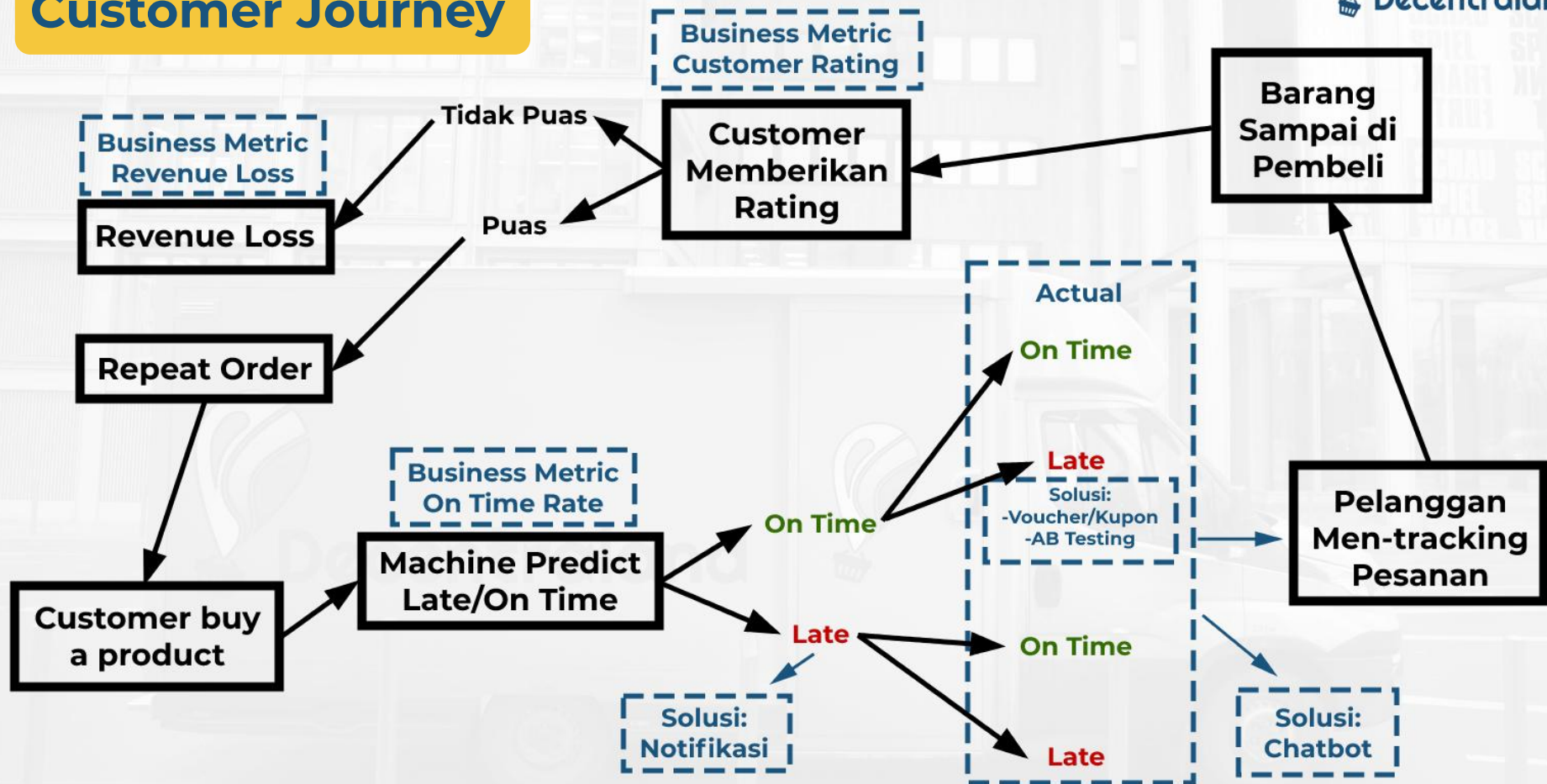
Variabel	Total
Total Sales-Discount	\$2,014,620.38
Revenue Decentraland (3% fee)	\$ 60,438.6114
Voucher (Variable Cost)	\$19,125
Profit	\$ 41313.6114
Gross Profit	68.35%

A/B Testing




- Nama experiment: AB testing kepuasan user setelah pemberian voucher terhadap user yang terprediksi on time tetapi late.
- Hipotesis: Voucher meningkatkan customer rating dan dapat mempertahankan customer
- Partisipan: Customer Decentraland
- Variabel yang diuji: User yang puas dan user yang tidak puas
- Metrics yang digunakan: Customer rating and revenue loss
- Sample size : 386 orang
- Durasi experiment: 2 minggu - 1 bulan



Customer Journey



Comparison

	Metrics	Before	After	Growth Rate	Our Target
	On Time Rate	40.3%	80.655%	50.03%	20%
	Customer Rating	2.99	3.77	20.69%	10%
	Revenue Loss	\$1,202,079.08	\$898,033.48	25.29%	20%-25%

Key Assumption:

1. Kenaikan on time rate setelah diprediksi mencapai 80.655%. Kami menerapkan margin of error sebesar 5% untuk mengatasi seandainya ada error di deployment.
2. Setiap customer yang late dan terpredicted benar menggunakan model menaikkan 1 rating pada customer rating.

On Time Rate Growth Calculation

Existing


	Count	Percentage
Delivery	10999	
Late	6563	59.7%
On Time	4436	40.3%



After Modelling

	Count	Percentage
Delivery	10999	
Late	6563	59.7%
Predicted Late	4903	74.7%
Predicted On Time	1660	25.3%
Late After Predict	1660	15.1%
On Time	4436	40.3%
On Time After Predict	9339	80.665%
On Time Growth Rate	50.0%	

Customer Rating Calculation



Description	Count
Total Rating	32.893
Average Rating	2.99
Predict Late Rating	$32.893 + 3.922 = 36.815$
Predicted Late Estimated Rating goes up by 1 (except Rating = 5)	
Average Rating after predict	3.77
Customer Rating Growth Rate	20.69%

Key Assumption:

Customer yang late dan terprediksi late menaikkan 1 rating berjumlah 973 orang.

Potential Revenue Loss Calculation



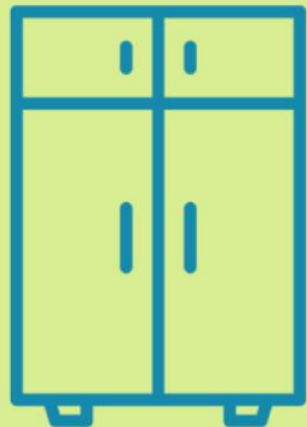
Assumption:
 $\text{Pred Late} = \text{Late} * 0.747$

Variabel	Total
Income	\$ 2,311,955
Discount	\$ 297,334.62
Revenue	\$2,014,620.38
Average Revenue	\$183.16

Description	Potential Revenue
Late (Before Modelling)	\$1,202,079.08
Predicted Late (After Modelling)	\$898,033.48



Decentraland



QnA



APPENDIX

Exploratory Data Analysis (EDA)

General Info



Warehouse Block	Kategorik
Mode of Shipment	Kategorik
Customer Care Calls	Numerik
Customer Rating	Ordinal
Cost of The Product	Numerik
Prior Purchase	Numerik
Product Importance	Ordinal
Gender	Biner
Discount Offered	Numerik
Weight in gram	Numerik
Late Shipment	Biner

- Data terdiri dari 10.999 sampel (baris).
- Terdapat 10 fitur (variabel independen) dan 1 variabel target (variabel dependen), yaitu Late Shipment.
- 5 fitur kategorik (Warehouse Block, Mode of Shipment, Customer Rating, Product Importance, dan Gender).
- 5 fitur numerik (Customer Care Calls, Customer Rating, Cost of The Product, Prior Purchases, Discount Offered, dan Weight in gram).



Numerik :

- Customer Care Calls
- Customer Rating
- Cost of The Product
- Prior Purchases
- Discount Offered
- Weight in gram

Kategorik::

- Warehouse Block
- Mode of Shipment
- Customer Rating
- Product Importance
- Gender

- Data terdiri dari 10.999 sampel (baris).
- Terdapat 10 fitur (variabel independen) dan 1 variabel target (variabel dependen), yaitu Late Shipment.
- 5 fitur kategorik (Warehouse Block, Mode of Shipment, Customer Rating, Product Importance, dan Gender).
- 5 fitur numerik (Customer Care Calls, Customer Rating, Cost of The Product, Prior Purchases, Discount Offered, dan Weight in gram).



Target:
Late Shipment

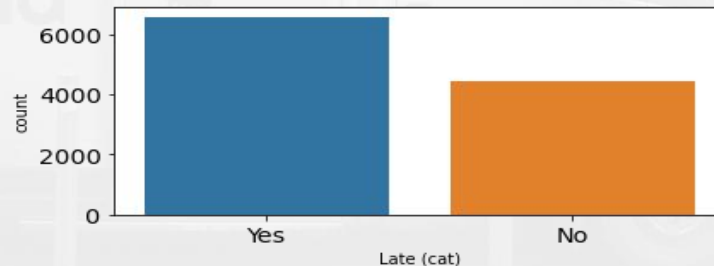
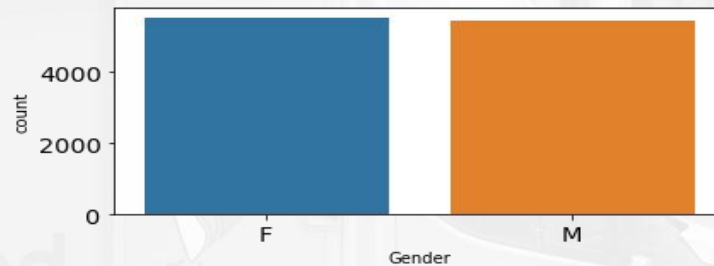
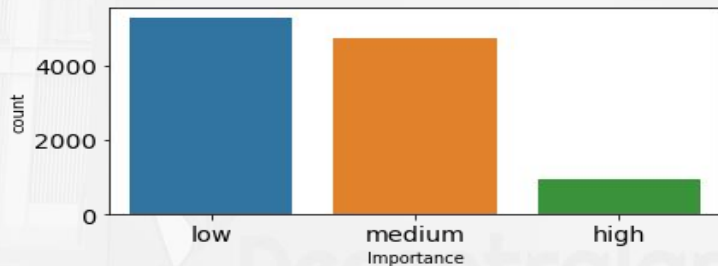
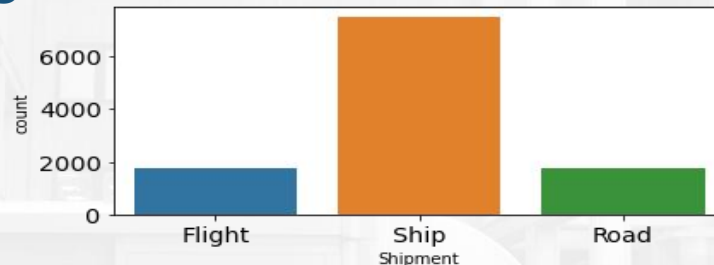
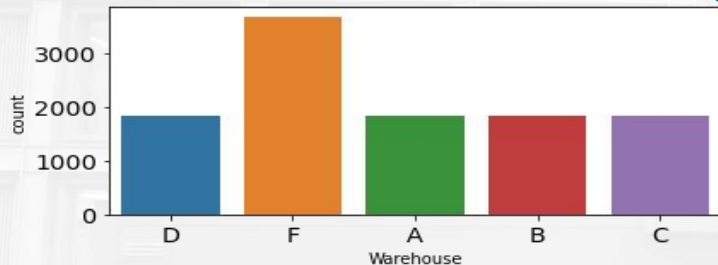
Exploratory Data Analysis (EDA) Descriptive Statistic

	Calls	Rating	Cost	Purchases	Discount	Weight (gram)
count	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000	10999.000000
mean	4.054459	2.990545	210.196836	3.567597	13.373216	3634.016729
std	1.141490	1.413603	48.063272	1.522860	16.205527	1635.377251
min	2.000000	1.000000	96.000000	2.000000	1.000000	1001.000000
25%	3.000000	2.000000	169.000000	3.000000	4.000000	1839.500000
50%	4.000000	3.000000	214.000000	3.000000	7.000000	4149.000000
75%	5.000000	4.000000	251.000000	4.000000	10.000000	5050.000000
max	7.000000	5.000000	310.000000	10.000000	65.000000	7846.000000

- Dilihat dari nilai min, max, median, dan mean, variabel Purchases dan Discount tampaknya memiliki nilai outlier di titik max.
- Disini terlihat bahwa variabel Rating bisa juga dikategorikan sebagai variabel kategorikal, 1-5 merupakan kategori rating. Untuk analisis statistik deskriptif, lebih pas untuk melihat variabel Rating ini sebagai variabel kategorik.

Exploratory Data Analysis (EDA)

Distribusi Data Variabel Kategorik



Model Evaluation

Matrix scoring yang dipakai

Matrix utama yang kami gunakan sebagai scoring adalah **ROC-AUC Score**

Kenapa?

Kami ingin meminimalisir false negative dan false positive value. Minimalisasi false negative akan lebih kami prioritaskan karena sifatnya lebih fatal daripada false positive.

Meet the customer expectation!

False Negative

Model **memprediksi** pengiriman akan sampai **tepat waktu**, kenyataannya pengiriman **telat**.

Failed to meet the customer expectation.

Sangat fatal, ibaratnya disini kita menjanjikan kepada customer bahwa barang akan sampai tepat waktu. Mereka sudah mempunyai ekspektasi yang tinggi, tetapi kita malah mengecewakan mereka dengan miss prediksi ini. **Harus dihindari!**

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

False Positive

Model **memprediksi** pengiriman akan **telat**, ternyata **faktanya** pengiriman terkirim **tepat waktu**.

Failed to meet the customer expectation.

Tidak sefatal false negative, tetapi kalau bisa tetap juga harus diminimalisir.

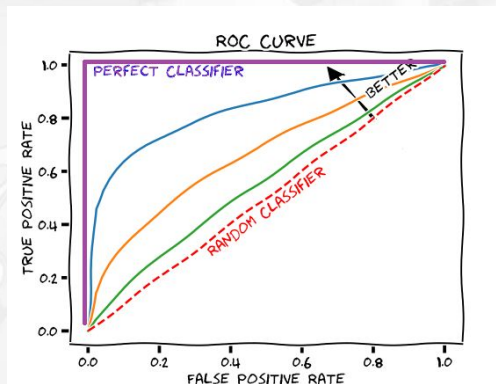
Matrix utama yang kami gunakan sebagai scoring adalah **ROC-AUC Score**

Kenapa?

ROC-AUC bagus bagi model kami karena telah mengkonsiderasi value false negative dan false positive dalam perhitungannya. AUC juga cocok pada data set kami karena sifatnya yang robust pada imbalanced target. Target kami (Late Shipment) sedikit imbalance, (59% sample telat, 41% sample tepat waktu). Score AUC yang lebih tinggi menunjukkan hasil yang lebih baik.

$$\begin{aligned} \text{True Positive Rate (TPR)} &= \frac{TP}{P} = \frac{TP}{TP + FN} \\ \text{also called sensitivity/recall/hit rate} \end{aligned}$$

$$\begin{aligned} \text{False Positive Rate (FPR)} &= \frac{FP}{N} = \frac{FP}{FP + TN} \\ \text{also called fall out} \end{aligned}$$



Kami juga akan memakai matrix scoring average precision sebagai secondary opinion.

Data Pre-processing

Data Treatment Setiap Fitur Berdasarkan Hasil EDA



Fitur	Special Case	Solusi
Warehouse Block	Kategorik	Perlu di-encoding agar bisa dikuantifikasi (One hot encoding)
Mode of Shipment	Kategorik	Perlu di-encoding agar bisa dikuantifikasi (One hot encoding)
Customer Care Calls	Sudah mendekati distribusi normal	Rescaling agar kinerja machine learning optimal dan model dapat memberlakukan semua fitur secara seimbang
Customer Rating	Ordinal	Rescaling agar kinerja machine learning optimal dan model dapat memberlakukan semua fitur secara seimbang
Cost of The Product	Sudah mendekati distribusi normal	Rescaling agar kinerja machine learning optimal dan model dapat memberlakukan semua fitur secara seimbang
Prior Purchase	Positively skewed dan Outlier	Log Transformation agar bisa mendekati distribusi normal dan outlier filter
Product Importance	Ordinal	Perlu di-encoding agar bisa dikuantifikasi (Label encoding)
Gender	Biner	Perlu di-encoding agar bisa dikuantifikasi (Label encoding)
Discount Offered	Positively skewed dan Outlier	Log Transformation agar bisa mendekati distribusi normal dan outlier filter
Weight in gram	Negatively skewed	Rescaling Standardization agar bisa mendekati distribusi normal

Model Evaluation

Hyperparameter yang dipakai

Log Reg	kNN	Dec Tree	RF	AdaBoost	XGBoost
Best Penalty	Best n_neighbors	max_depth	n_estimators	n_estimators	max_depth
Best C	Best p	min_samples_split	bootstrap	learning_rate	min_child_weight
	Best algorithm	min_samples_leaf	criterion	algorithm	gamma
	weights	max_features	max_depth		tree_method
		criterion	min_samples_split		colsample_bytree
		splitter	min_samples_leaf		eta
			max_features		lambda
			n_jobs		alpha

Dataset 2	Logistic Regression		k-Nearest Neighbor		Decision Tree		Random Forest		Adaboost		XGBoost	
	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision
Train	0.717	0.832	0.785	0.858	0.802	0.879	0.890	0.931	0.748	0.845	0.874	0.921
Test	0.725	0.833	0.722	0.821	0.715	0.824	0.730	0.842	0.741	0.838	0.731	0.843

Model Evaluation

Hasil Scoring Dataset 3

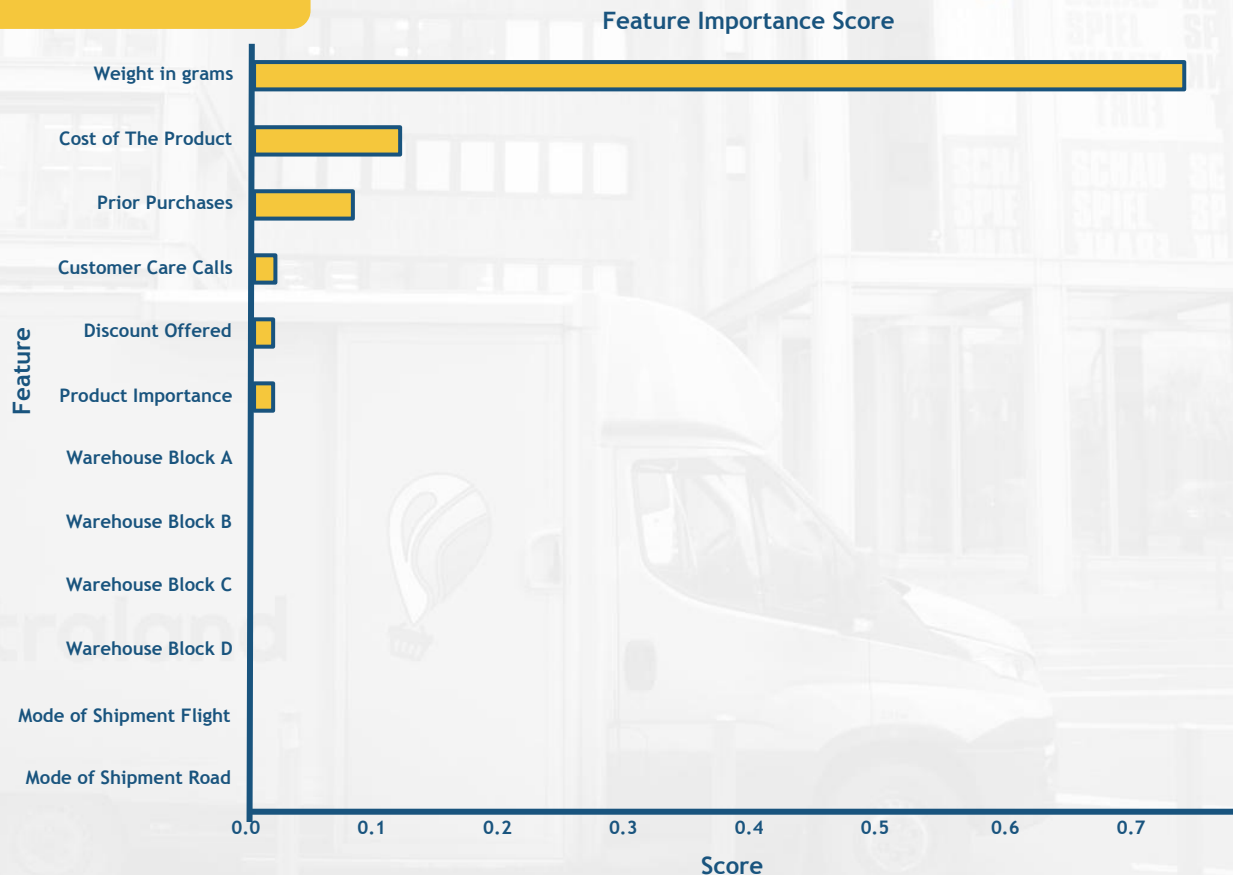
Dataset 3	Logistic Regression		k-Nearest Neighbor		Decision Tree		Random Forest		Adaboost		XGBoost	
	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision	AUC	Av. Precision
Train	0.585	0.632	1	1	0.646	0.672	0.619	0.677	0.629	0.667	0.984	0.983
Test	0.596	0.636	0.602	0.637	0.603	0.649	1	1	0.615	0.655	0.617	0.670

Feature Importance Adaboost

Melakukan Iterasi berdasarkan Feature Importance, menggunakan metode Adaboost pada dataset 1.

Diperoleh fitur-fitur yang paling penting sebagai berikut:

1. Weight in grams
2. Cost of The Product
3. Prior Purchases
4. Customer Care Calls



Cross Validation dengan Model Adaboost

K-cross validation	Mean Score
3	0.707
5	0.721
10	0.723