

E-COMMERCE

• CUSTOMER CHURN

TABLE OF CONTENT

01 Business Problem

02 Data Understanding

03 Exploratory Data Analysis

04 Data Preparation

05 Modeling

06 Conclusion & Recommendation



Business Problem



BUSINESS PROBLEM

- Tim Marketing E-Commerce X ingin melakukan evaluasi terhadap terjadinya "Customer Churn".
- E-Commerce X ingin memberikan promo untuk mencegah customer churn.



01

Problem Statement

promosi khusus harus diprioritaskan kepada pelanggan yang berpotensi akan churn

02

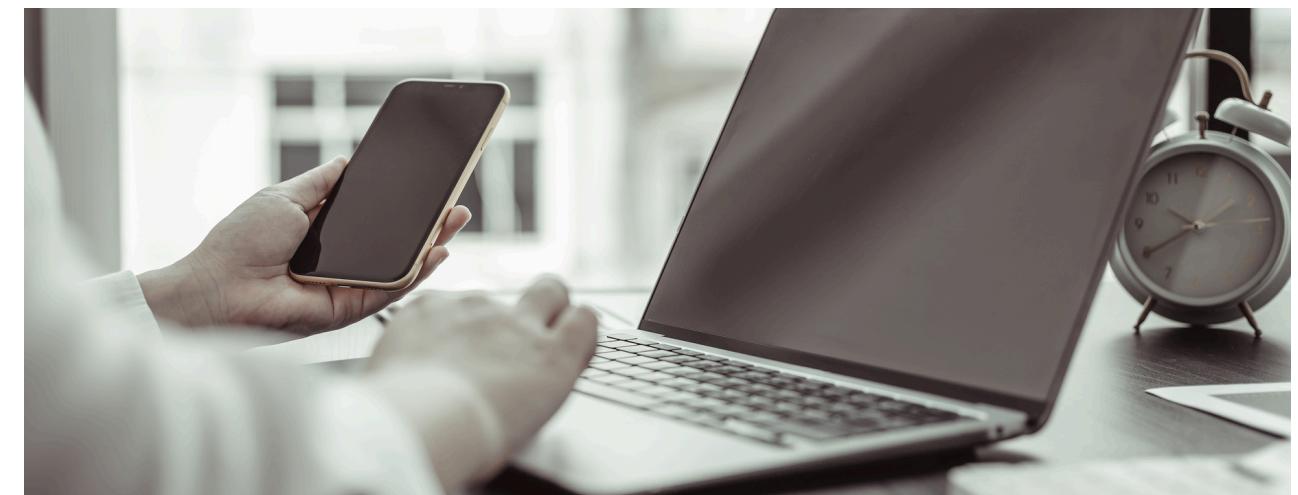
Goals

Memprediksi kemungkinan pelanggan akan churn dan apa penyebabnya.

03

Analytic Approach

Membuat, mengevaluasi, dan mengimplementasikan model machine learning untuk memprediksi pelanggan akan churn



METRICS EVALUATION

		Predicted	
		Negative	Positive
		True Negatives (TN)	False Positive (FP)
Actual	Negative	Prediciton: Tidak Churn Actual: Tidak Churn	Prediciton: Churn Actual: Tidak Churn
	Positive	False Negatives (FN) Prediciton: Tidak Churn Actual: Churn	True Positives (TP) Prediciton: Churn Actual: Churn

Konsekuensi:
Kehilangan pelanggan

Konsekuensi:
Biaya promosi
terbuang

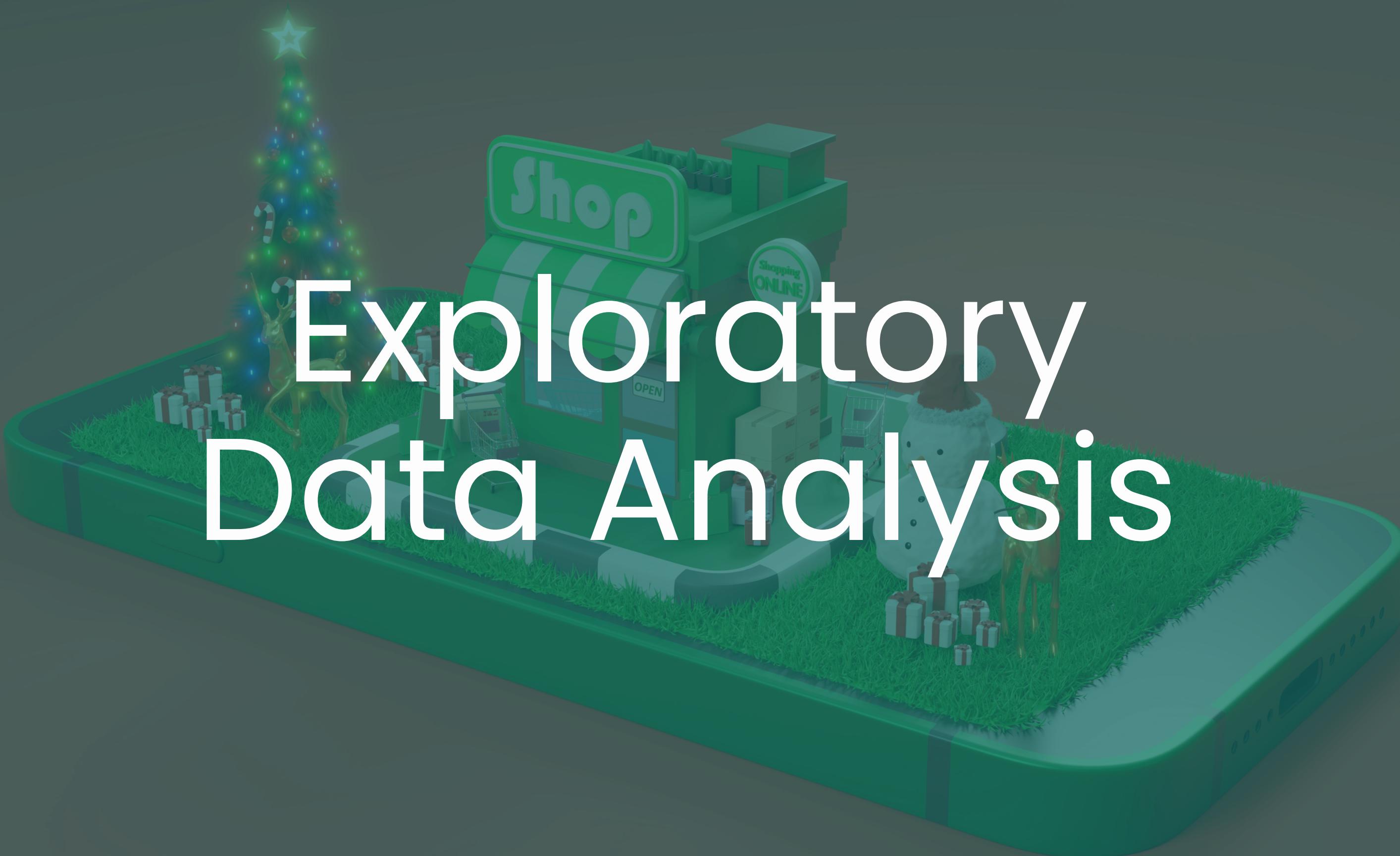
Data Understanding



Data terdiri dari 3941 baris dan 11 kolom

Attribute	Data Type	Description
Tenure	Float	Masa pelanggan menggunakan layanan
Warehouse to Home	Float	Jarak antara gudang ke rumah pelanggan
Number of Device	Int	Jumlah perangkat yang terdaftar pada satu pelanggan
Prefered Order Cat	Object	Kategori yang disukai pelanggan
Satisfaction Score	Int	Skor kepuasan pelanggan pada satu pelanggan
Marital	Object	Status pernikahan pelanggan
Number of Address	Int	Jumlah alamat yang terdaftar pada satu pelanggan
Complain	Int	Komplain pelanggan. 0-tidak komplain, 1-komplain
Day Since Last Order	Float	Hari sejak pemesanan terakhir
Cashback Amount	Float	Rata-rata cashback yang diterima satu pelanggan
Churn	Int	0 - Tidak <i>churn</i> , 1 - <i>Churn</i>

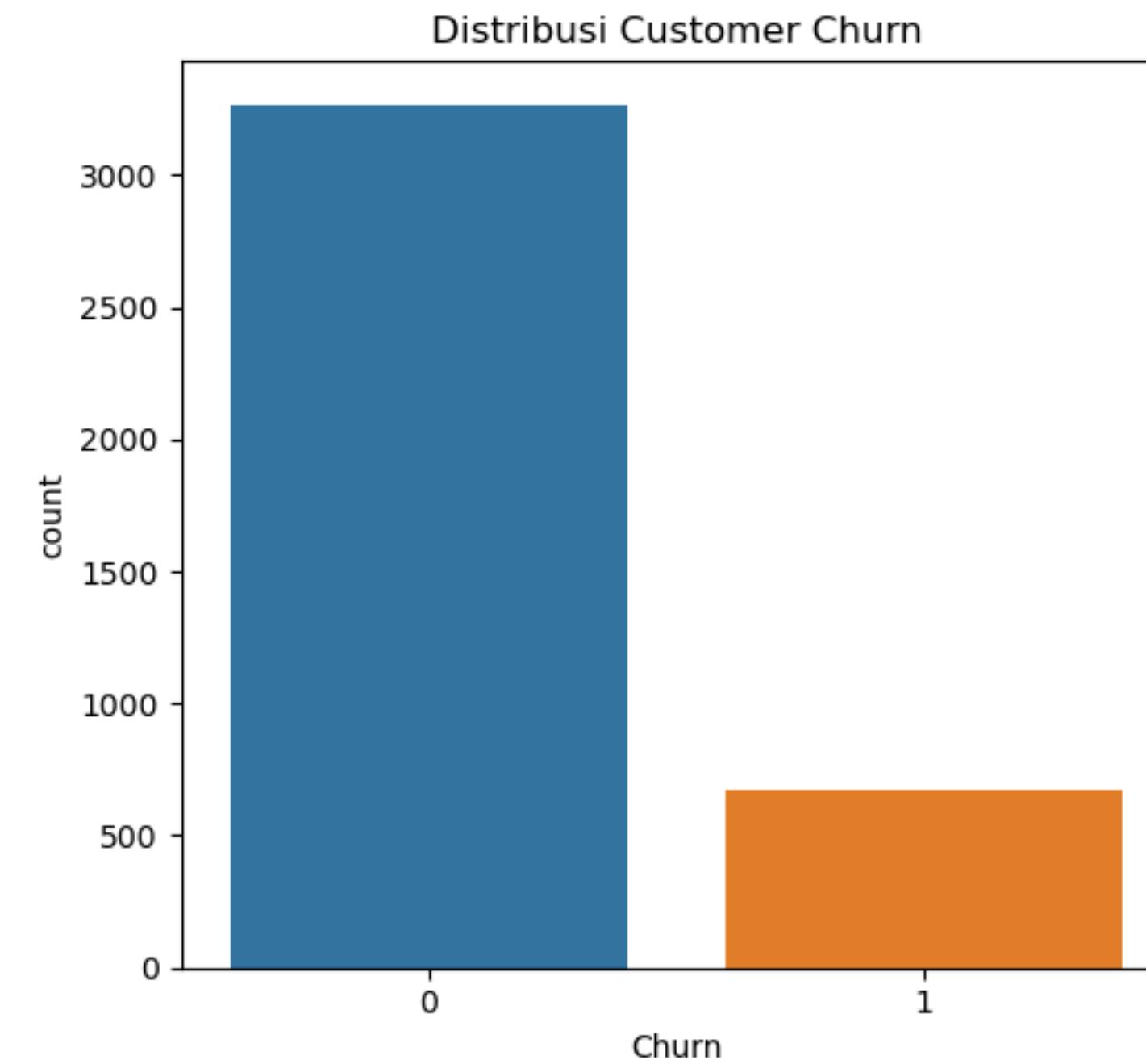
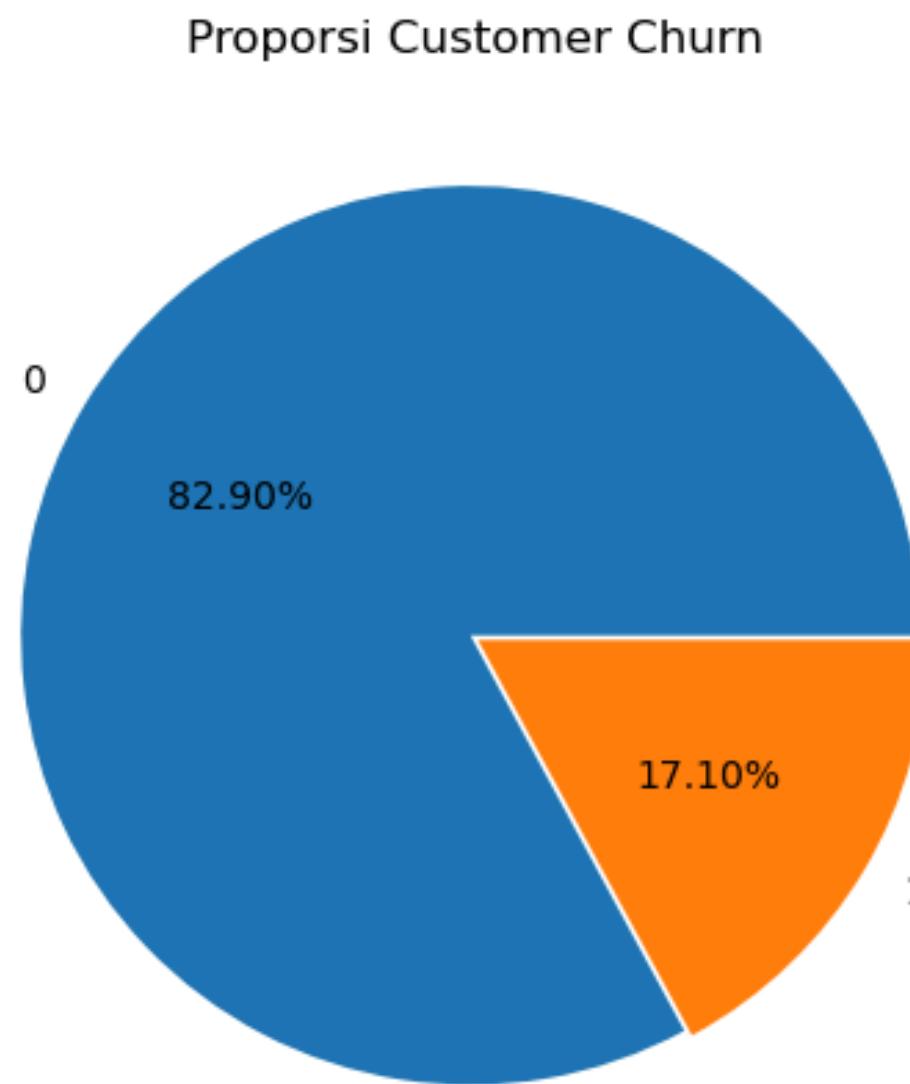
Exploratory Data Analysis



Antar feature tidak ada yang memiliki hubungan yang kuat

	Tenure	WarehouseToHome	NumberOfDeviceRegistered	SatisfactionScore	NumberOfAddress	Complain	DaySinceLastOrder	CashbackAmount	Churn
Tenure	1.000000	-0.053618	-0.017633	-0.048987	0.277656	-0.041638	0.198661	0.431859	-0.420157
WarehouseToHome	-0.053618	1.000000	0.026200	0.023106	0.023512	0.039878	0.025546	0.023663	0.074552
NumberOfDeviceRegistered	-0.017633	0.026200	1.000000	0.000564	0.124153	0.013035	0.035833	0.203806	0.103913
SatisfactionScore	-0.048987	0.023106	0.000564	1.000000	0.041714	-0.033045	0.019757	-0.009212	0.106293
NumberOfAddress	0.277656	0.023512	0.124153	0.041714	1.000000	0.008248	-0.070876	0.259277	0.025071
Complain	-0.041638	0.039878	0.013035	-0.033045	0.008248	1.000000	-0.048045	0.018085	0.261450
DaySinceLastOrder	0.198661	0.025546	0.035833	0.019757	-0.070876	-0.048045	1.000000	0.375708	-0.188466
CashbackAmount	0.431859	0.023663	0.203806	-0.009212	0.259277	0.018085	0.375708	1.000000	-0.182730
Churn	-0.420157	0.074552	0.103913	0.106293	0.025071	0.261450	-0.188466	-0.182730	1.000000

Target (churn) pada dataset ini masih berdistribusi timpang



A woman with long dark hair tied back, wearing a cable-knit sweater, is smiling and looking down at her smartphone. She is holding a pen in her right hand and writing on the top book of a stack of four books. The books are light-colored and stacked vertically. In the background, there are large green plants.

Data Preparation

PROBLEMATIC DATA

HANDLING

Handling Missing Value

Terdapat missing value pada kolom Tenure, WarehouseToHome, dan DaySinceLastOrder dengan persentase ~ 5%

KNN Imputer

Duplicated Data

Terdapat 671 (17%) data yang sama / duplikat

Drop Duplicate

Inconsistent Data

Pada kolom PreferredOrderCat terdapat data yang bermakna sama dengan data yang lain -> Mobile dan Mobile Phone

Mobile digabung dengan mobile phone

Outlier

Ada beberapa kolom yang memiliki outlier. Namun dalam kasus ini, outlier bersifat natural karena mungkin terjadi, bukan karena ada kesalahan

Membiaran data Outlier

PREPROCESSING

Encoding

Transformasi data kategorik menjadi data numerik (Marital Status dan PreferredOrderCat)

Scaling

Transformasi data numerik agar antar variable memiliki skala yang sama

Balancing

Menyeimbangkan data antara kelas churn dan tidak

HANDLING

One Hot Encoding

Robust Scaler

Class Weight

Modeling



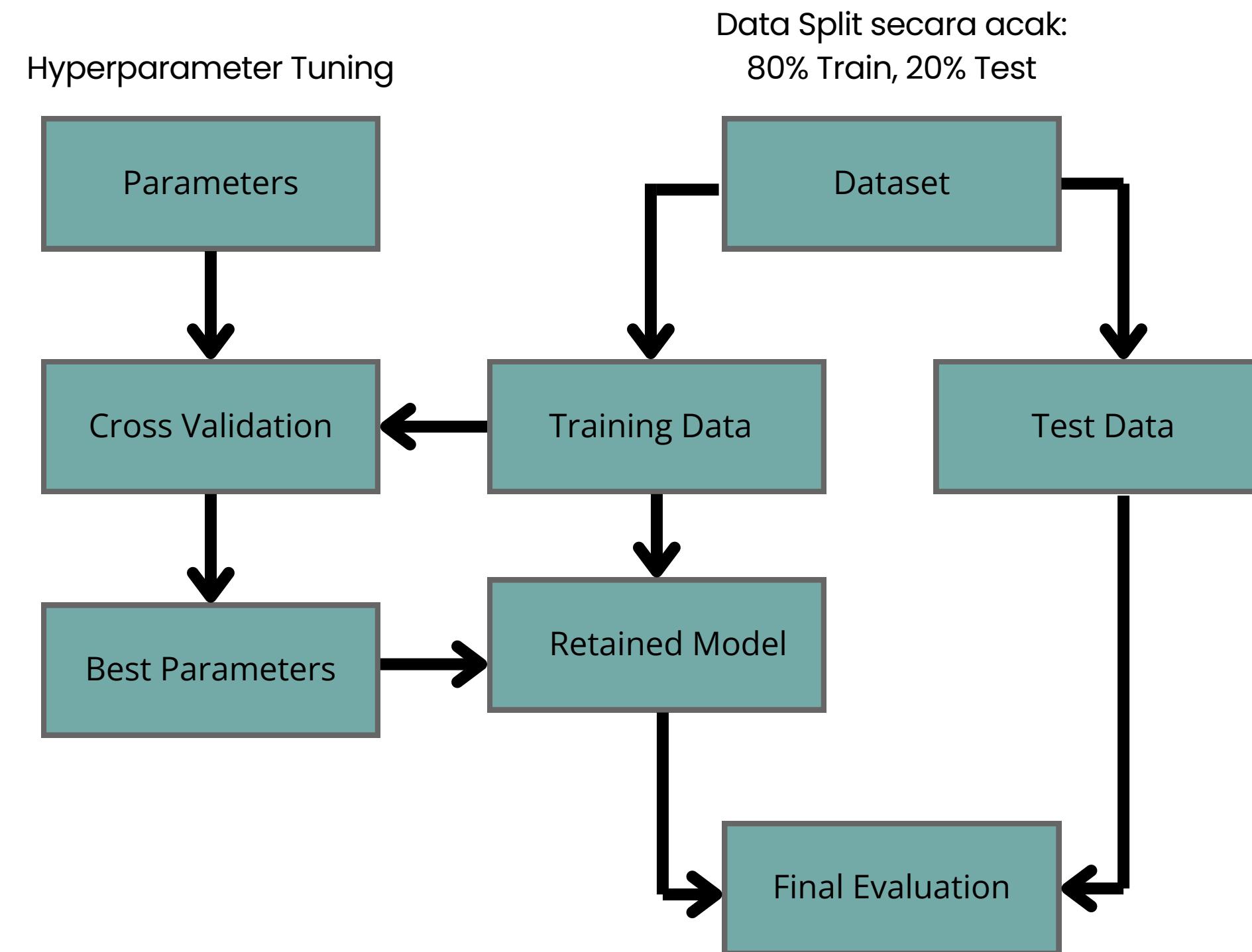


MACHINE LEARNING

MODELING

- Mendapatkan model dengan F2 Score tertinggi pada data train maupun data test
- Semakin tinggi score, semakin baik model dalam melakukan prediksi
- Mengetahui kualitas model (terlalu kompleks atau terlalu sederhana)
- Tune model yang sudah dipilih untuk mendapatkan score yang lebih baik

Mencari model terbaik F2 dengan score tertinggi



Best Model: **LightGBM**

Model terbaik ini sudah optimal
karena perbedaan nilai pada train
dan test yang kecil

Score Train Score Test

79%

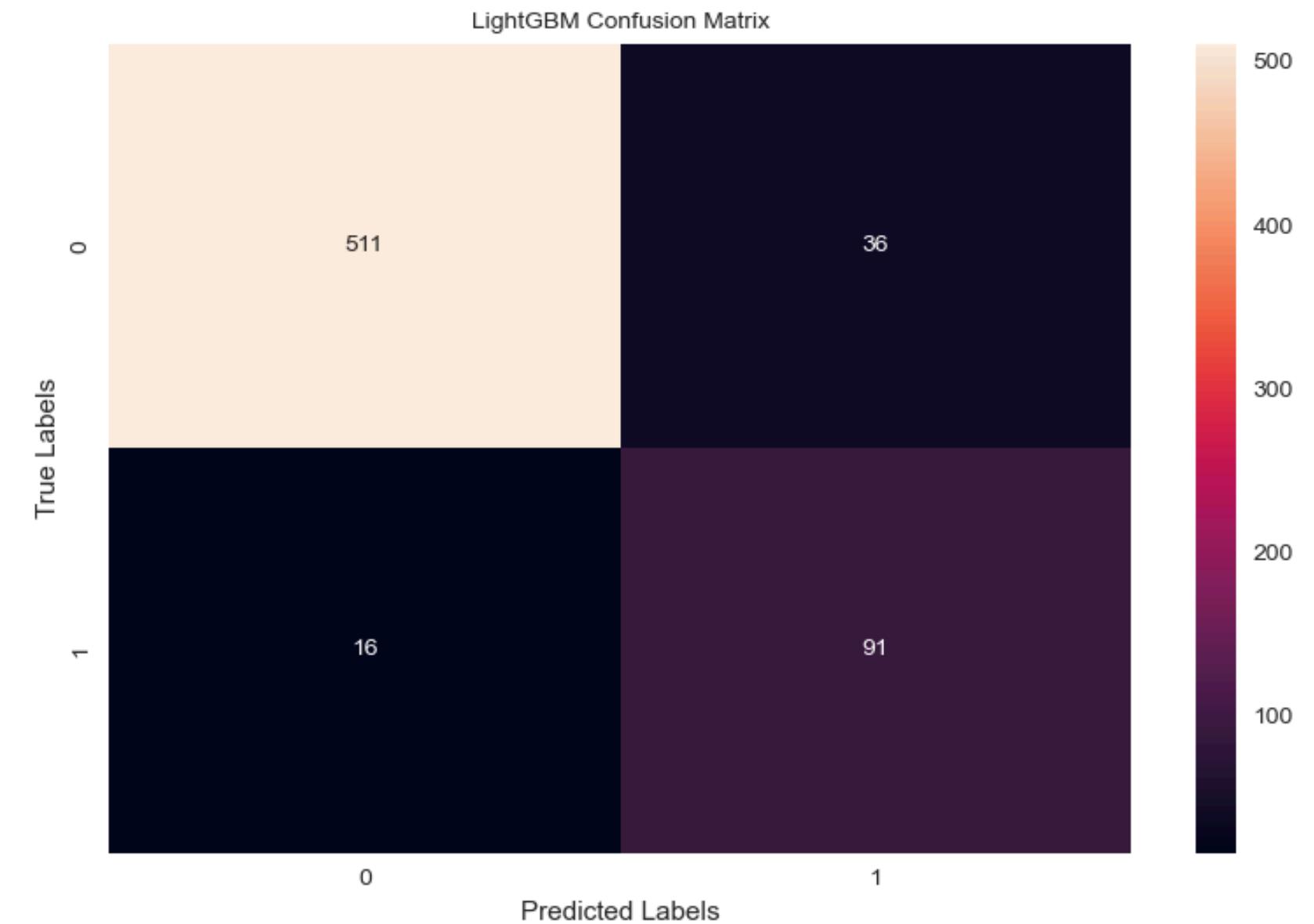
81%

Model dapat memprediksi pelanggan yang tidak churn dengan akurasi 93% dan pelanggan yang churn dengan akurasi 85%

Classification Report Tuned LGBM :				
	precision	recall	f1-score	support
0	0.97	0.93	0.95	547
1	0.72	0.85	0.78	107
accuracy			0.92	654
macro avg	0.84	0.89	0.86	654
weighted avg	0.93	0.92	0.92	654

Dari 654 pelanggan yang dijadikan data test model mampu mengklasifikasikan;

- 511 pelanggan yang tidak Churn (True Negative)
- 91 pelanggan Churn (True Positive)
- Meski masih ada sekitar 52 orang terkласifikasi secara kurang tepat (False Positive & False Negative)



Terdapat 654 pelanggan dalam data set. Asumsikan pemberian promo sebesar \$10 per pelanggan.

Tanpa machine learning:

$$654 \times \$10 = \$6540$$

Dengan machine learning:

- True Positive Churn customer: $91 \times \$10 = \910

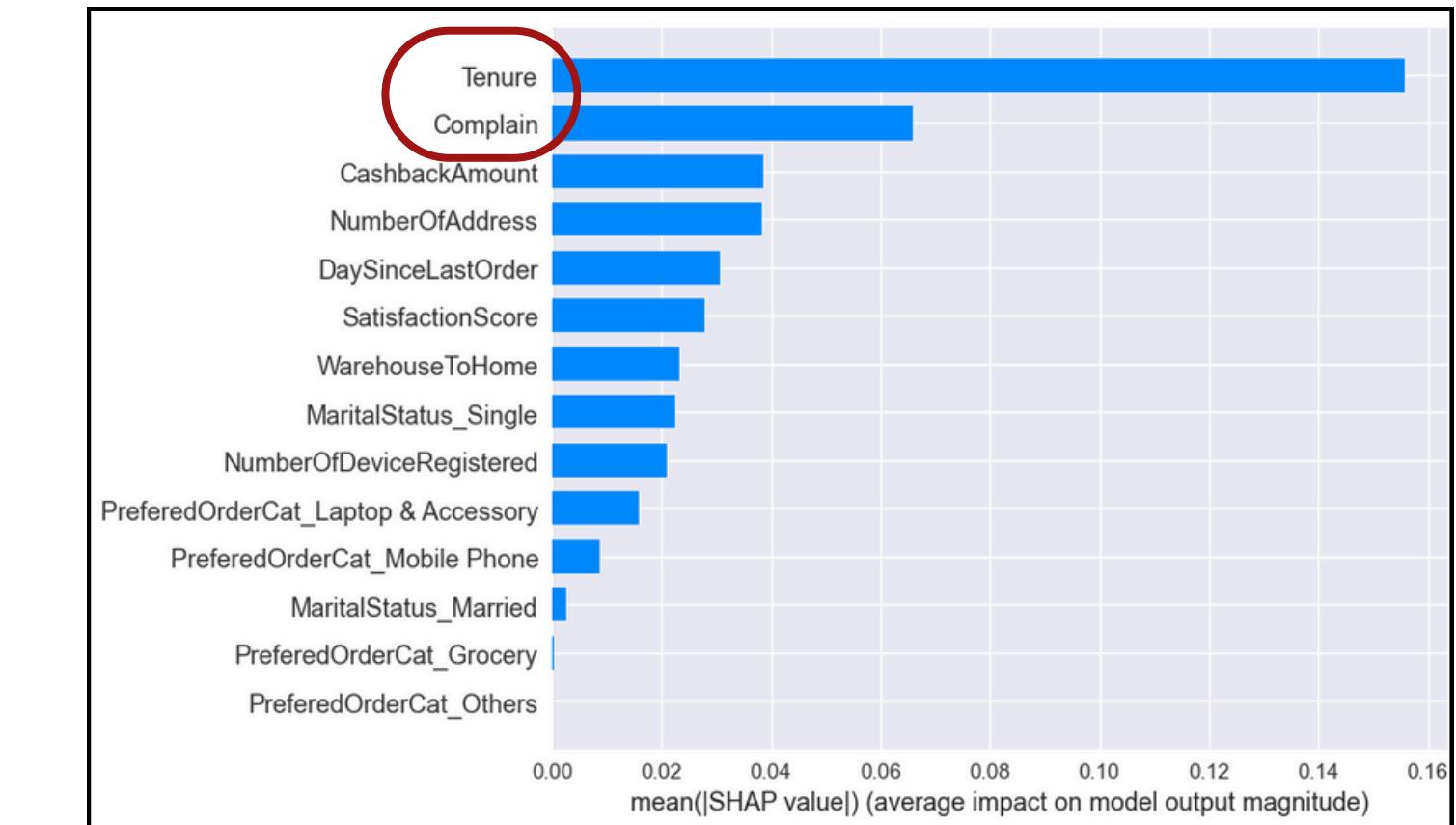
Kesalahan prediksi

- False Positive : $36 \times \$10 = \360
- False Negative: $16 \times \$50 = \800
- Total: $\$910 + \$360 + \$800 = \2070

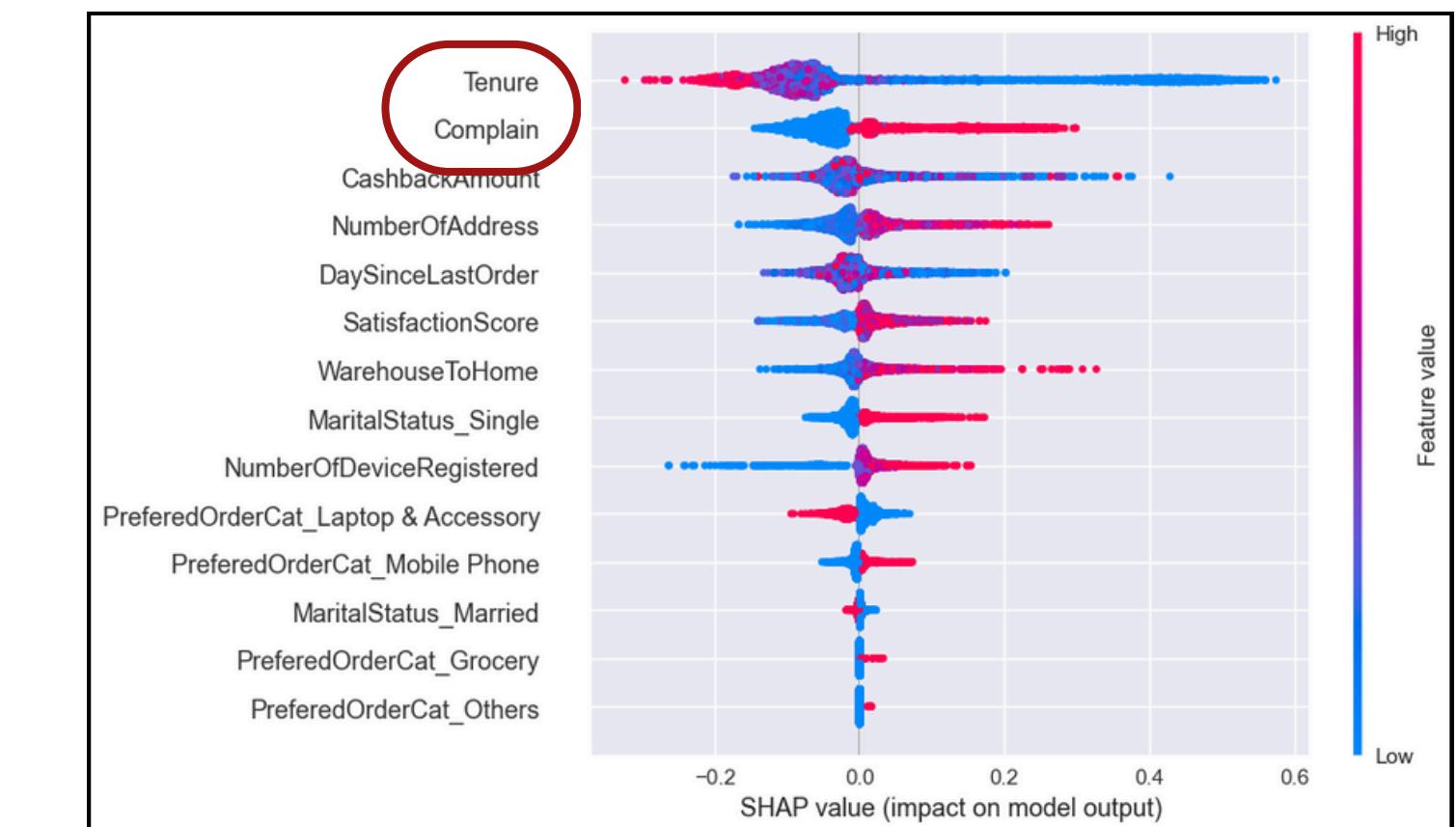
Manfaat monetary yang didapatkan e-commerce dengan adanya model machine learning:
 $\$6540 - \$2070 = \$4470$

E-commerce X dapat **menghemat sebesar \$4470** dari biaya yang perlu dikeluarkan untuk promo.

Tenure: Semakin kecil nilai Tenure artinya semakin baru customer menggunakan layanan, semakin besar kemungkinan Customer akan churn.



Complain: Semakin besar nilai Complain, semakin besar kemungkinan Customer akan Churn dan sebaliknya.



5 Customer yang paling berpotensi churn

	Tenure	WarehouseToHome	NumberOfDeviceRegistered	PreferedOrderCat	SatisfactionScore	MaritalStatus	NumberOfAddress	Complain	DaySinceLastOrder	CashbackAmount	Churn	Prediction(Prob->1)
2244	1.0	9.0	5	Mobile Phone	3	Single	8	1	1.0	145.90	1	0.993216
1088	0.0	30.0	5	Mobile Phone	2	Single	2	1	0.0	111.18	1	0.992210
2800	0.0	30.0	5	Mobile Phone	1	Single	2	1	0.0	111.18	1	0.992210
148	1.0	20.0	5	Mobile Phone	5	Married	6	1	1.0	148.35	1	0.991954
487	1.0	17.0	5	Mobile Phone	5	Single	8	1	1.0	149.04	1	0.991410

Conclusion & Recommendation



Conclusion

Hasil Model

- Dengan hasil best model LightGBM dapat mendeteksi pelanggan yang akan churn dan pelanggan yang tidak churn dengan prediksi sebesar lebih dari 80%

Manfaat menggunakan ML bagi perusahaan

- Tanpa machine learning: biaya yang dikeluarkan **\$6540**
- Dengan machine learning: biaya yang dikeluarkan **\$2070**

Faktor yang mempengaruhi customer churn

- Tenure
- Complain



Recommendation

For Business:

- Memberikan promo dan layanan maksimal, khususnya pada pelanggan baru.
- Memperhatikan complain dan segera menanganinya
- Gunakan machine learning yang telah dibuat untuk melihat customer yang berpotensi churn.

For Model:

- Menambahkan lebih banyak data, khususnya pada kelompok churn.
- Menambahkan fitur-fitur lain dari interaksi pelanggan.
- Meningkatkan kualitas data

Thank You