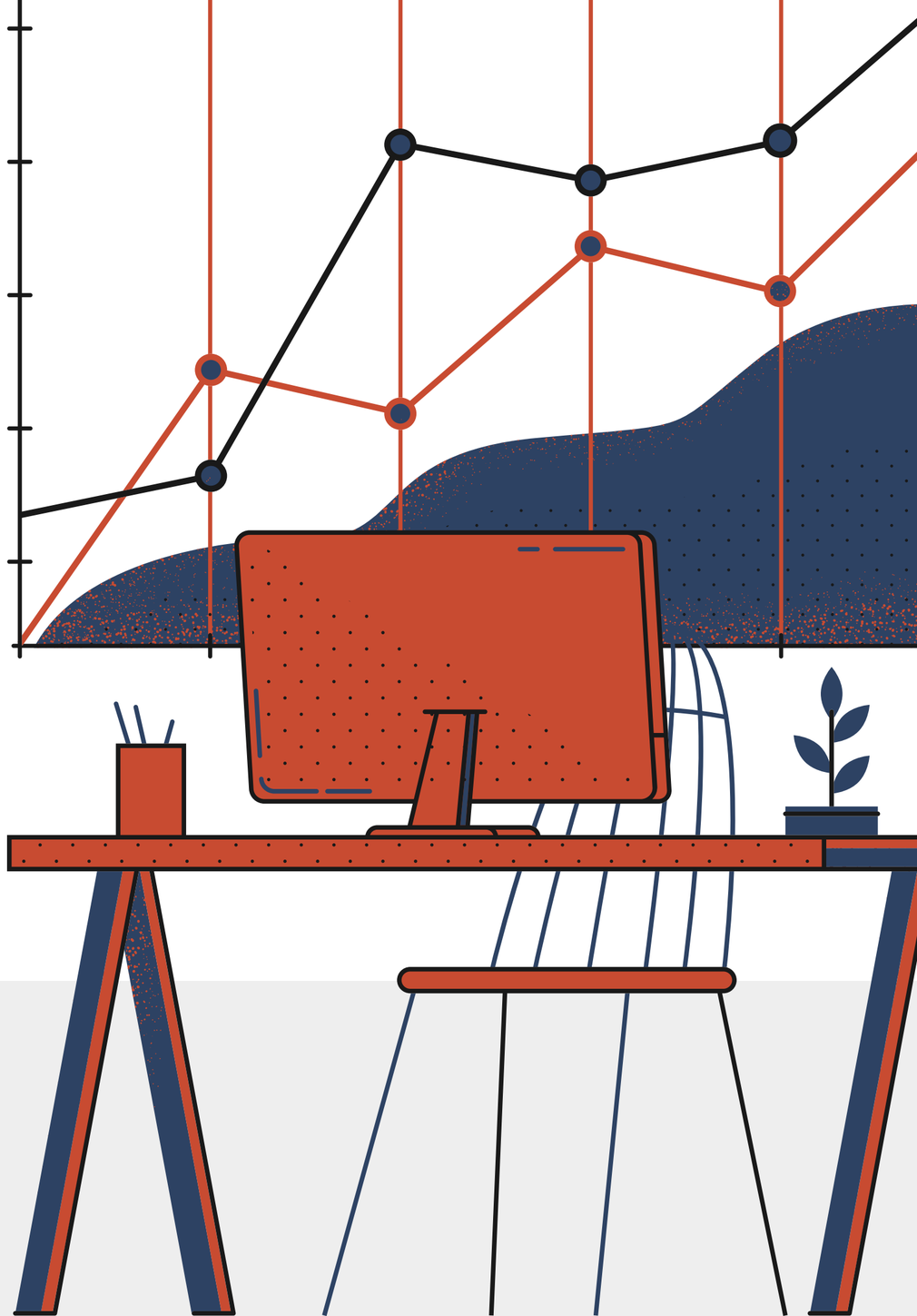
Ecommerce Customer Churn Prediction



INTELLEGEND
CONSULTANT

Kami dan Klien kami

Intellegend Consultant

Kami merupakan tim analis dari Intellegend Consultant. Perusahaan jasa profesional di bidang konsultasi bisnis

Shopful

Sebuah perusahaan B2C (Business to Customer) eCommerce di US yang menjual berbagai macam produk

Meet Our Team

Bryan Tamin

Maysarah Ulfah

Vicken

Wildan R R

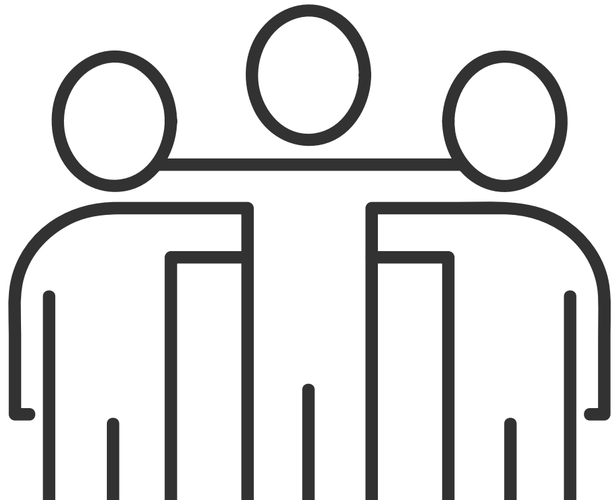
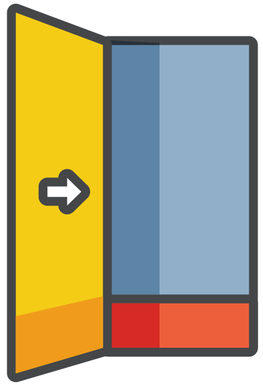


Table of Contents



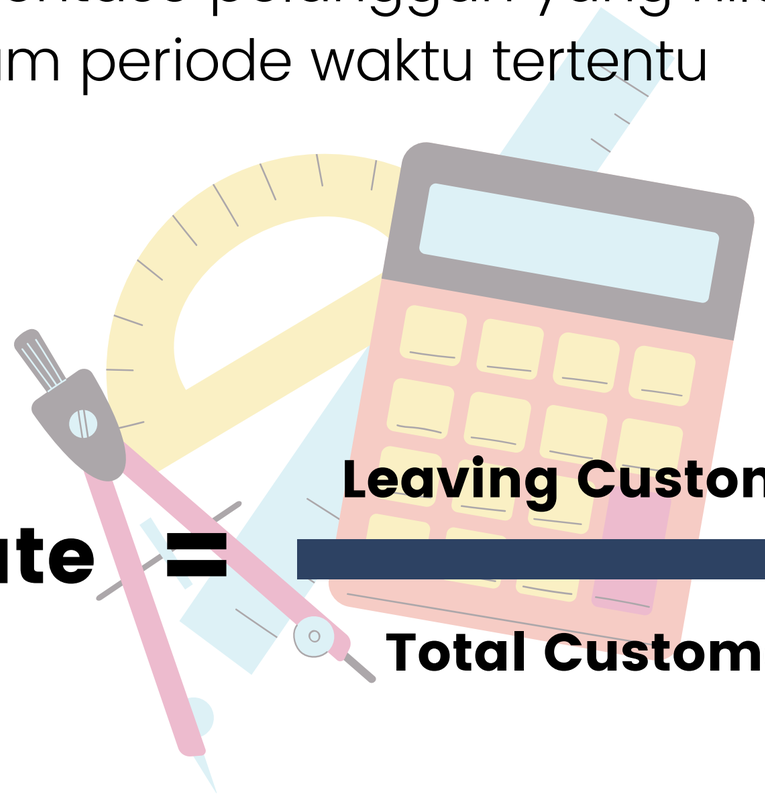
16.83% Customer Churn

Churn adalah kondisi dimana customer berhenti berlangganan (menutup/menghapus akun)

biaya untuk mempertahankan customer **5 sampai 25 kali lipat** lebih kecil dibandingkan biaya untuk mencari customer baru.

Churn rate

persentase pelanggan yang hilang dalam periode waktu tertentu



$$\text{Churn rate} = \frac{\text{Leaving Customer}}{\text{Total Customer}}$$

Source: https://hbr.org/2014/10/the-value-of-keeping-the-right-customers#:~:text=Depending%20on%20which%20study%20you,the%20one%20you%20have%20happy.

Problem Statement

Terjadinya customer churn yang tinggi yaitu 16.83% dari total customer

Goal

Menurunkan jumlah customer churn

Objective

Memberikan insight serta membuat model machine learning untuk memprediksi customer churn

Business Metrics

Customer Churn Rate

Potential profit

Potential Profit = revenue - expenses

Perhitungan Kerugian

Total Belanja = 50 x rata-rata Cashback amount x Total order

Dengan asumsi cashback adalah 2% dari **total belanja** maka didapatkan total pembelanjaan sebesar

$154,339,328

Dan dari 16.83% customer yang churn potensi **kerugian** diperhitungan sebesar

$23,134,478

CustomerID
Tenure
PreferredLoginDevice
CityTier
WarehouseToHome
PreferredPaymentMode
Gender
HourSpendOnApp
NumberOfDeviceRegistered
PreferedOrderCat
SatisfactionScore
MaritalStatus
NumberOfAddress
Complain
OrderAmountHikeFromlastYear
CouponUsed
OrderCount
DaySinceLastOrder
CashbackAmount
Churn 🎯

About this dataset

Dataset ini berisi data dari satu bulan penjualan Ecommerce Shopful

Shape

5630 baris dan 20 kolom. Churn adalah target.

Duplicated and Missing

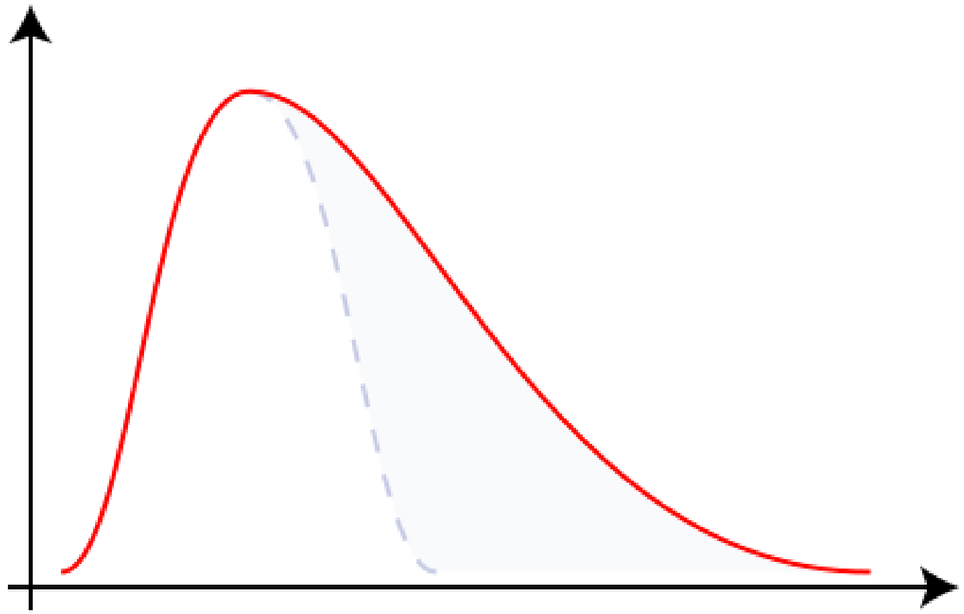
Tidak ada data yang duplikat, namun terdapat 7 kolom yang memiliki missing data

Mislabeled value

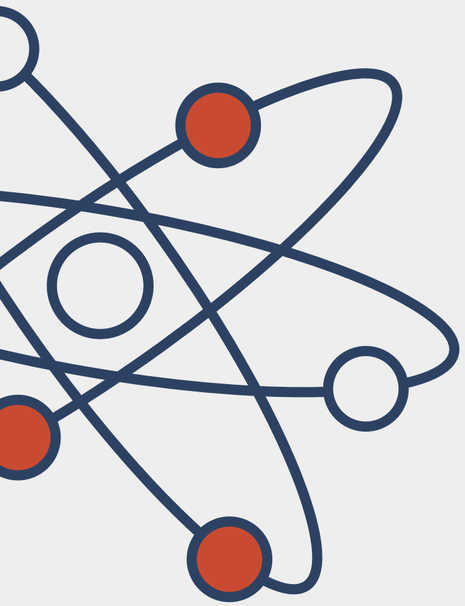
PreferredLoginDevice, PreferredPaymentMode

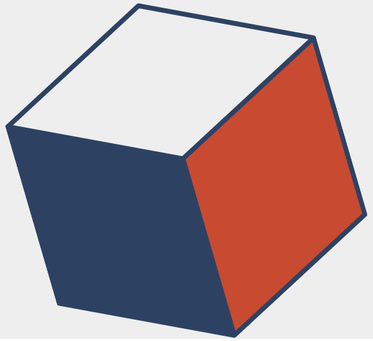
Distribution
1 Uniform distribution

Distribution
7 right-skewed distribution

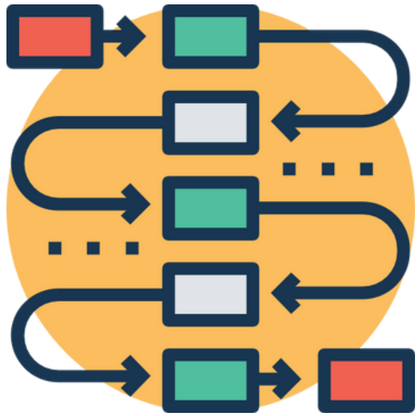


Correlation
2 feature highly correlated



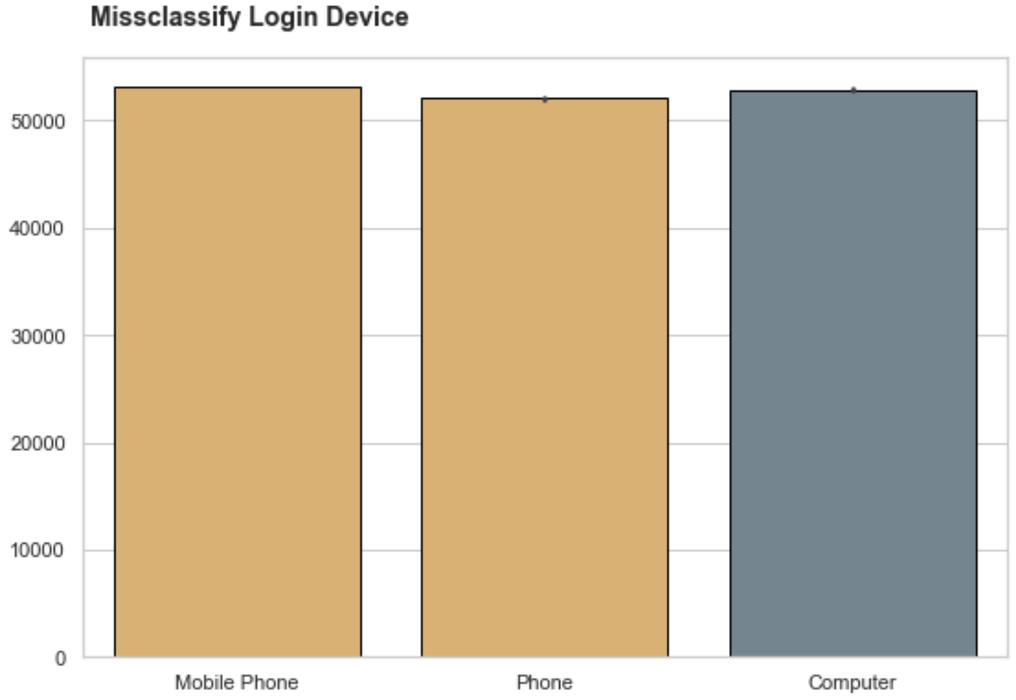


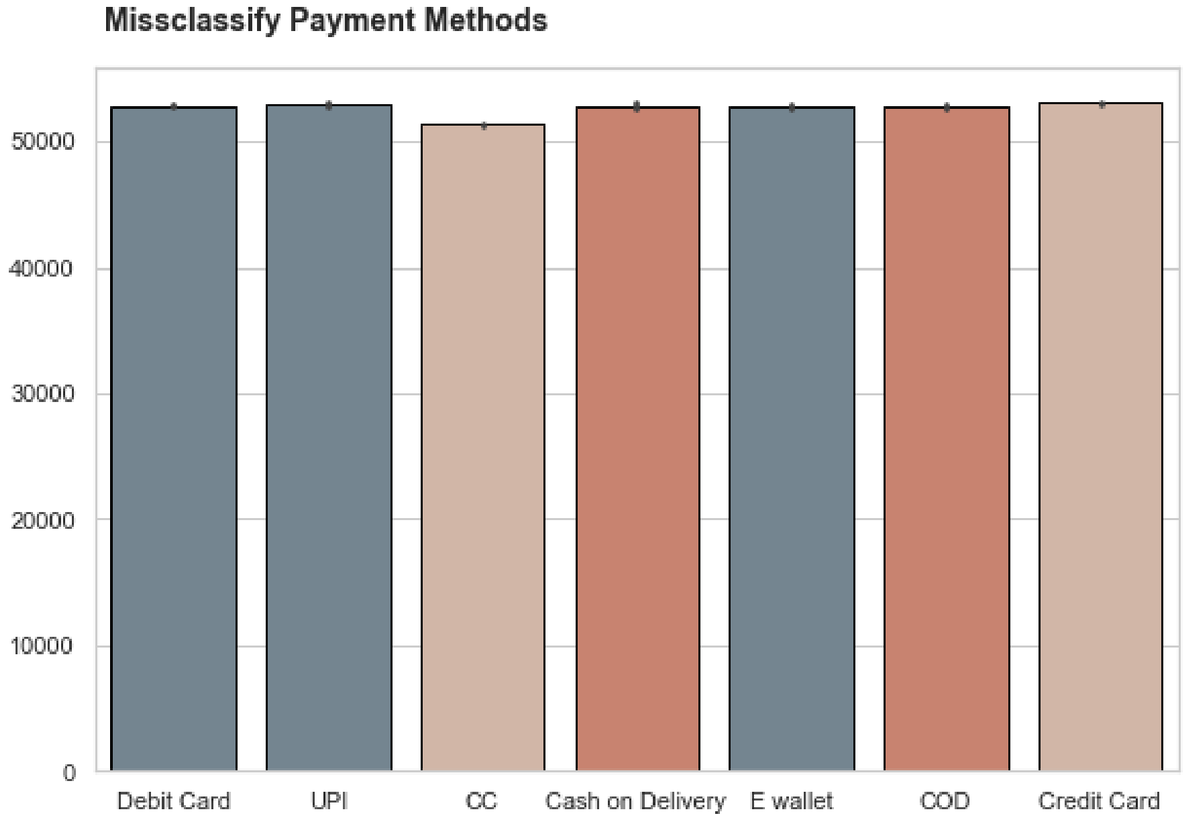
Data Preprocessing

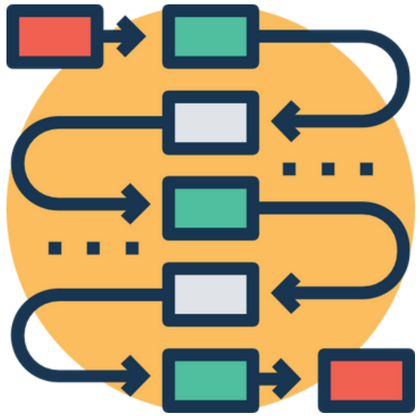


Split Data

Pembagian data 80: 20







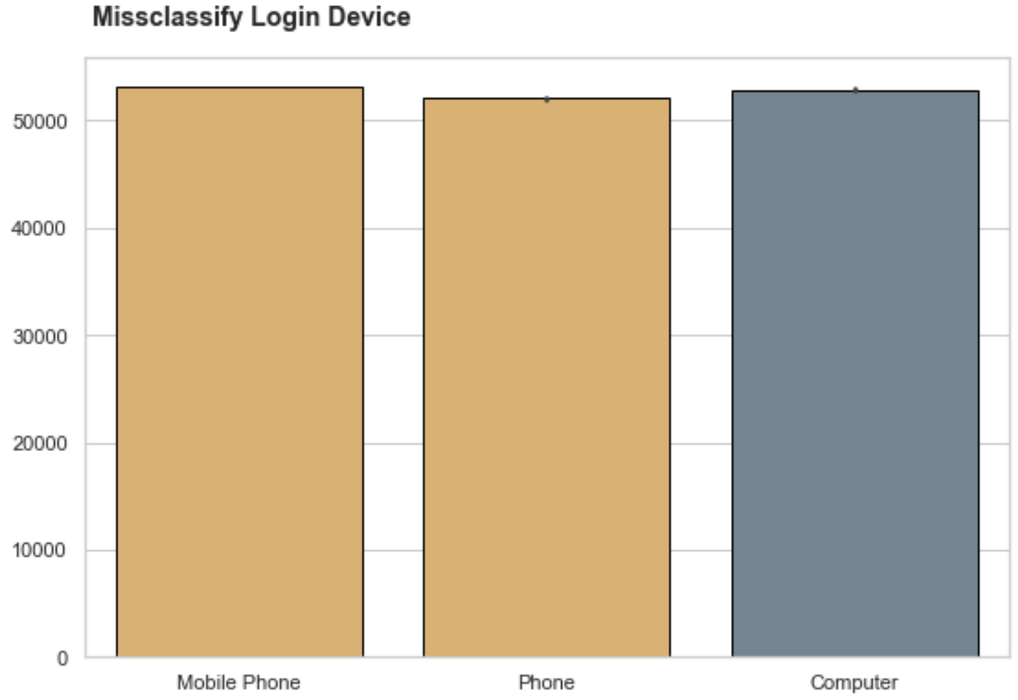
Split Data

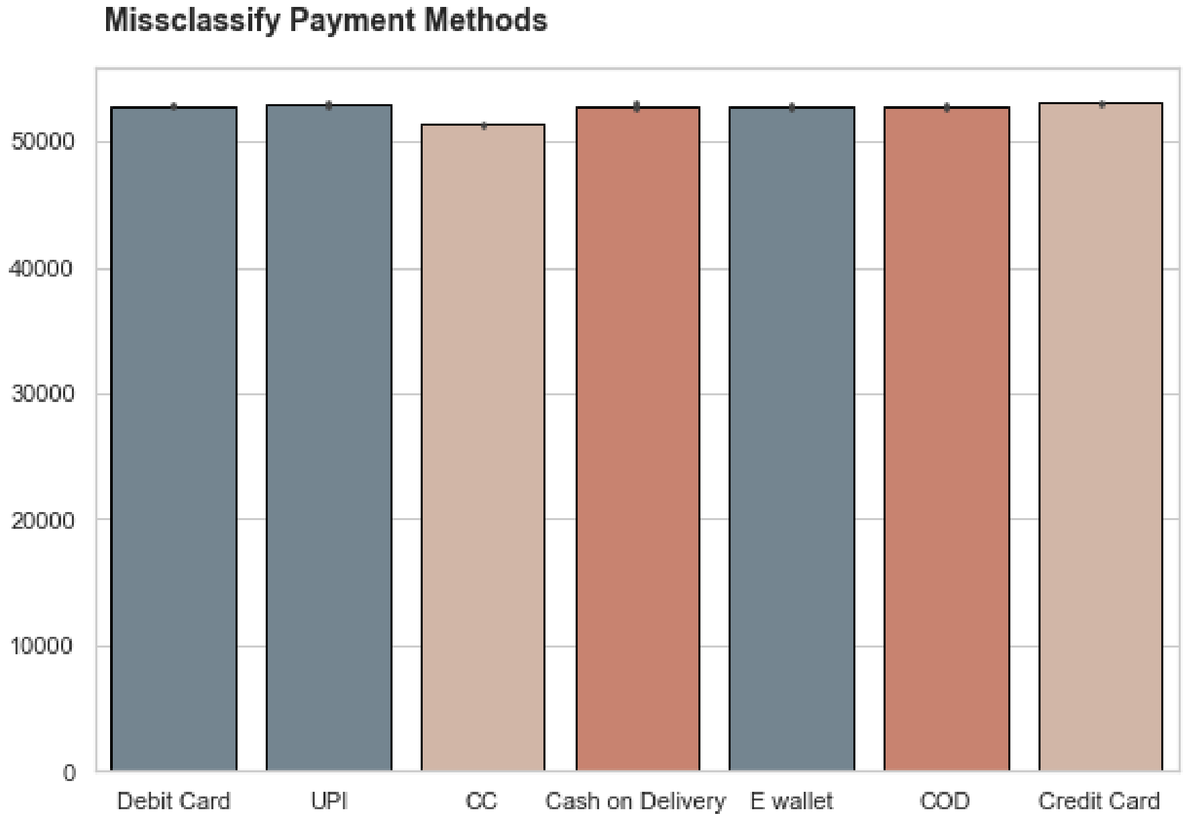
Pembagian data 80: 20

Handling missing values

Categorical : Modus

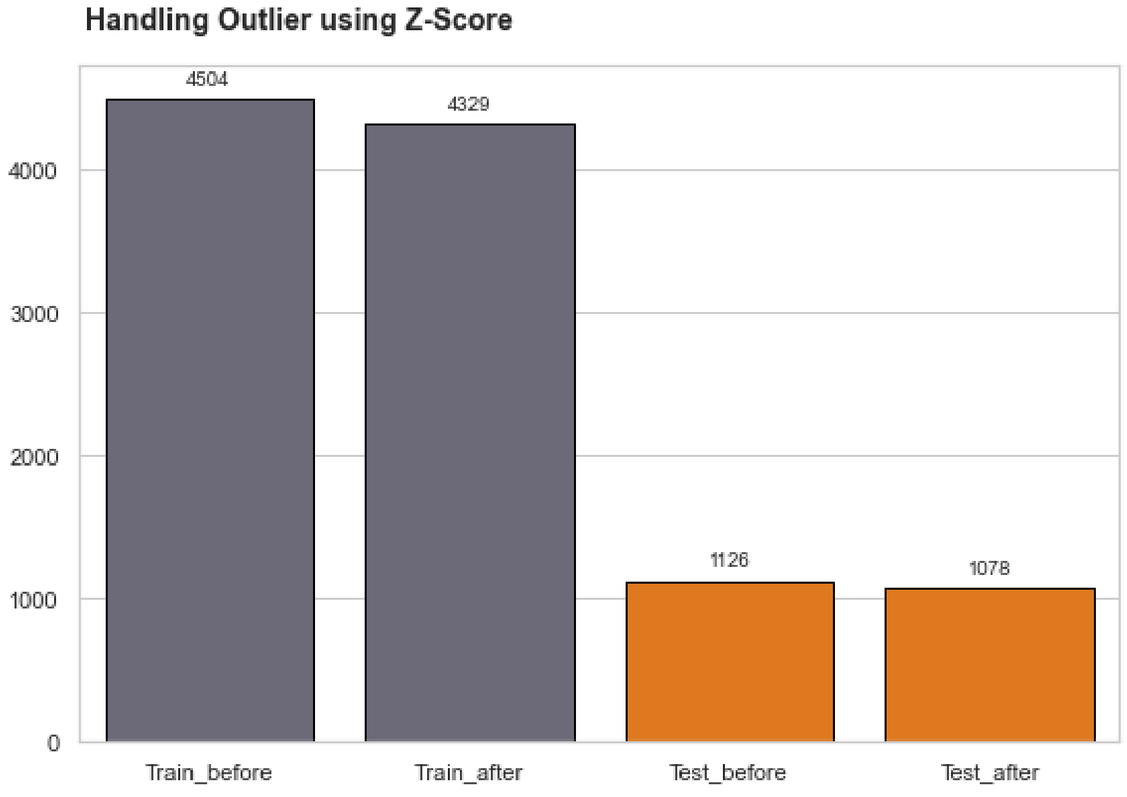
Numerical : Median





Handling Outlier

Z-Score

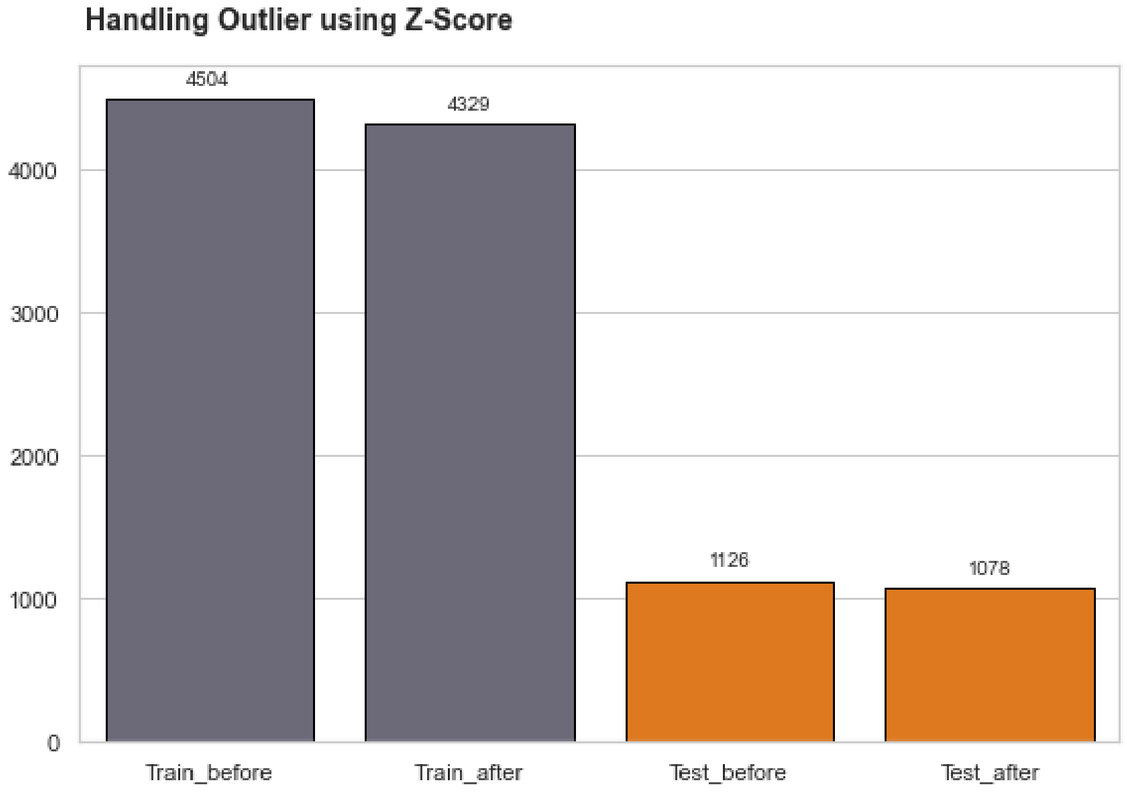


Handling Outlier

Z-Score

Transformasi

Boxcox

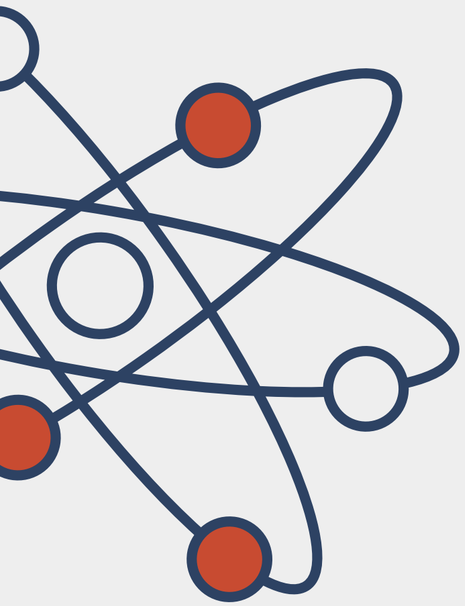


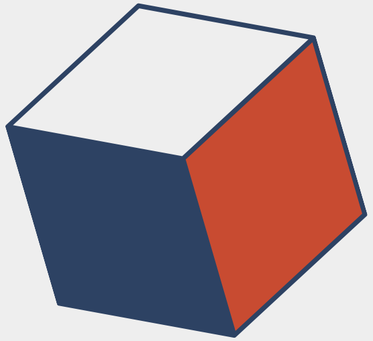
Feature Encoding

PreferedLoginDevice dan **Gender** akan menggunakan label Encoder, sisanya OneHotEncoder

Scaling

Normalization





Modeling and Evaluation

Model

- Logistic Regression (Baseline)
- Decision Tree
- KNN
- SVM
- Random Forest
- Gradient Boosting
- XGBoost
- AdaBoost
- CatBoost

Evaluation Metrics

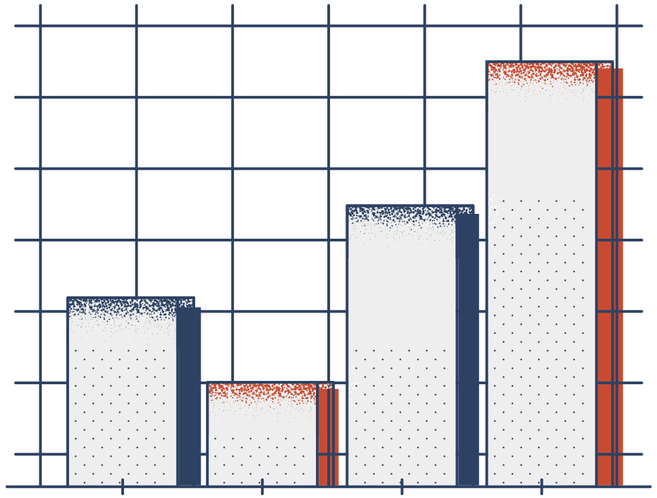
- Recall

fokus pada seluruh customer yang berpotensi untuk churn

- ROC_AUC

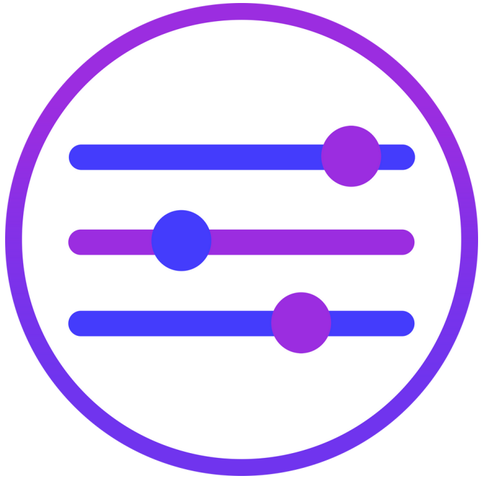
Memastikan model dapat membedakan kelas dengan baik (tidak semua data diprediksi positif / negatif)

Model Evaluation



Model	Train		Test	
	Recall	ROC_AUC	Recall	ROC_AUC
Logistic Regression	0.562	0.761	0.652	0.808
Decision Tree	1.000	1.000	0.876	0.906
KNN	0.667	0.823	0.472	0.718
SVM	0.650	0.819	0.663	0.825
Random Forest	1.000	1.000	0.820	0.907
Gradient Boosting	0.705	0.842	0.680	0.829
XGBoost	1.000	1.000	0.888	0.936
AdaBoost	1.000	1.000	0.831	0.907
CatBoost	0.974	0.982	0.815	0.900

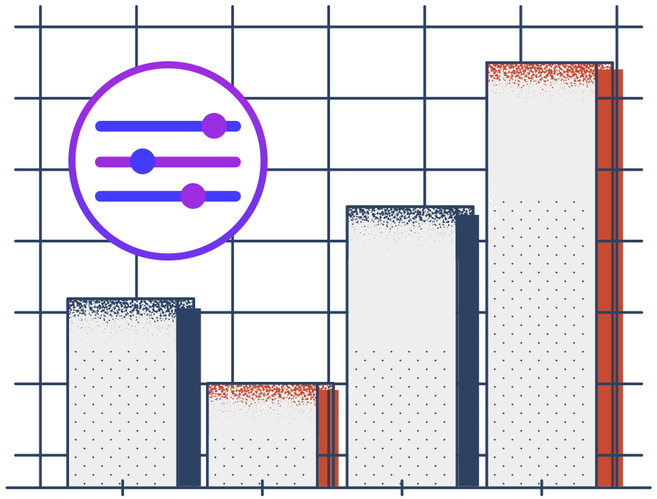
Hyperparameter Tuning



Hyperparameter yang digunakan untuk mentuning best model adalah:

- eta: penyusutan step size untuk mencegah overfitting (a.k.a. learning_rate)
- gamma: minimum loss reduction yang dibutuhkan untuk membuat partisi selanjutnya pada leaf node
- max_depth: kedalaman maksimum tree
- min_child_weight: jumlah weight minimum pada sebuah "child" (partisi), semakin tinggi parameter ini, model semakin konservatif
- colsample_bytree: rasio subsample pada konstruksi tree
- lambda: regularisasi L2, semakin tinggi parameter ini, model semakin konservatif
- alpha: regularisasi L1, semakin tinggi parameter ini, model semakin konservatif
- tree_method: algoritma konstruksi tree yang digunakan pada model XGBoost

Hyperparameter Tuning



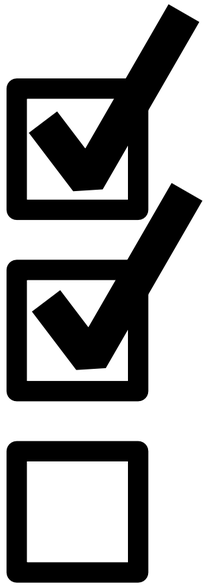
Model	Train		Test	
	Recall	ROC_AUC	Recall	ROC_AUC
Logistic Regression	0.562	0.761	0.652	0.808
Decision Tree	0.997	0.999	0.893	0.925
KNN	0.667	0.823	0.842	0.906
SVM	0.650	0.819	0.899	0.885
Random Forest	1.000	1.000	0.916	0.955
Gradient Boosting	0.705	0.842	1.000	0.500
XGBoost	1.000	1.000	0.955	0.960
AdaBoost	1.000	1.000	0.930	0.940
CatBoost	0.853	0.924	0.764	0.870

- Dari model score di atas, dapat disimpulkan bahwa XGBoost memiliki performa yang paling baik.
- Recall 95%: Dari seluruh customer yang churn, hanya 5% yang tidak terdeteksi churn (false negative), artinya potensi customer yang akan churn tidak terdeteksi sangat kecil.

Predicted

Actual		Negative	Positive
	Negative	**868**	32
	Positive	8	**170**

Precision = 84.2%



Feature Selection

Predicted

Actual		Negative	Positive
	Negative	**868**	32
	Positive	8	**170**

- **Recall = 95.5%**
- ROC_AUC = 96.0%
- Precision = 84.2%

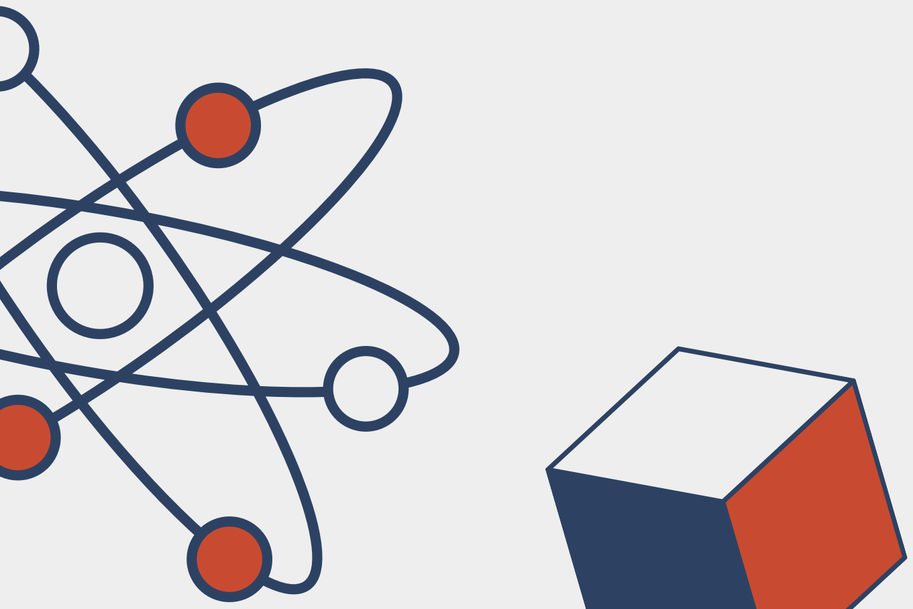
after selection

Predicted

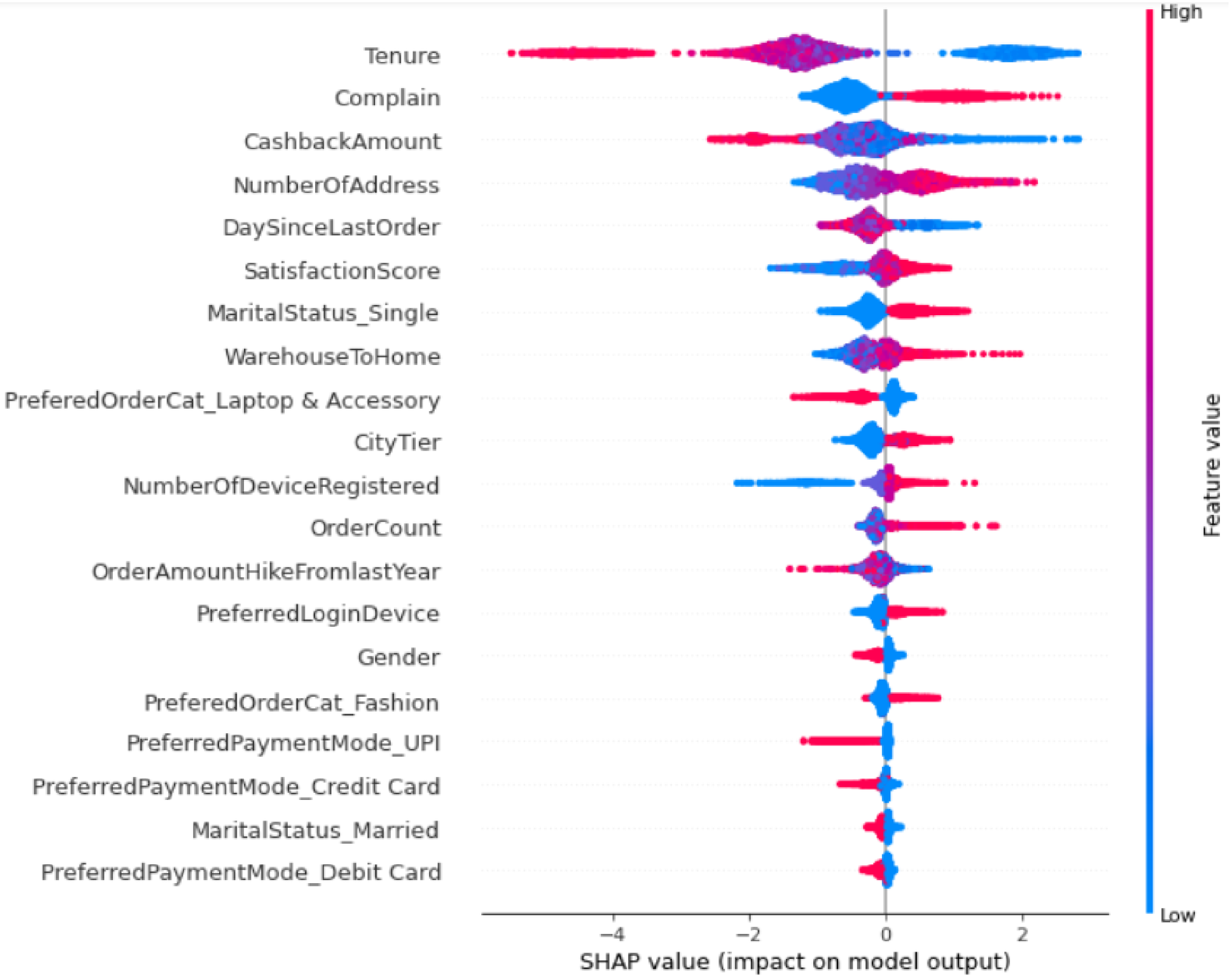
Actual		Negative	Positive
	Negative	**893**	7
	Positive	6	**172**

- **Recall = 96.6%**
- ROC_AUC = 97.9%
- Precision = 96.1%

Beberapa feature yang di drop adalah feature yang redundant dan berpotensi mengakibatkan multicolinearity. Karena pada saat One-Hot encoding baiknya kolom terakhir di drop. Dengan melakukan feature selection dapat terlihat bahwa score dari model menjadi naik



Business Insight dan Rekomendasi



SHAP Value

Pelanggan baru butuh diberikan cashback 1.6x lebih besar sampai setidaknya 2 bulan

Mendahulukan penyelesaian complain bagi customer yang terdeteksi berpotensi Churn

Business Recommendation

(31.67% Churn rate)

Complain

(10.9% Churn rate)

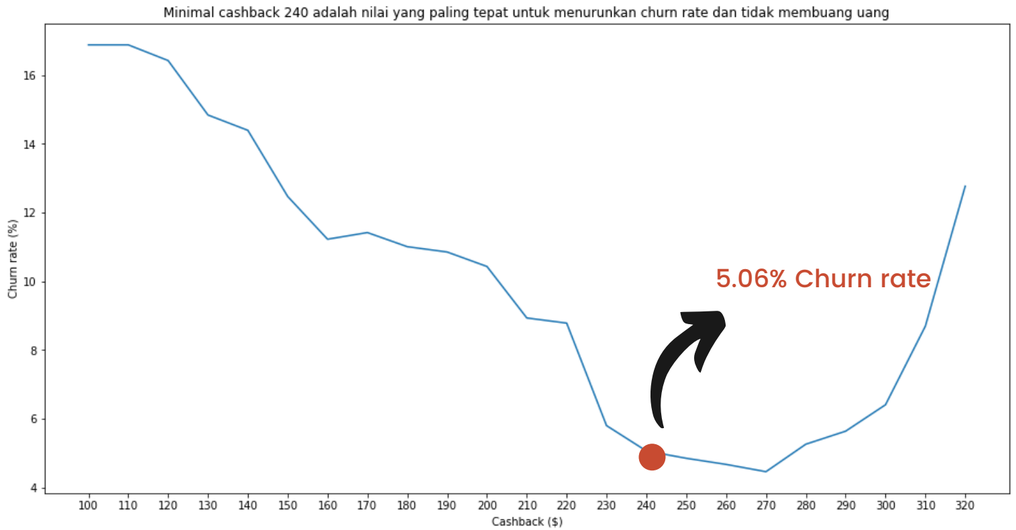
No complain

Turun hampir 3 kali lipat

Mendahulukan penyelesaian **complain** bagi customer yang **terdeteksi** berpotensi **Churn**

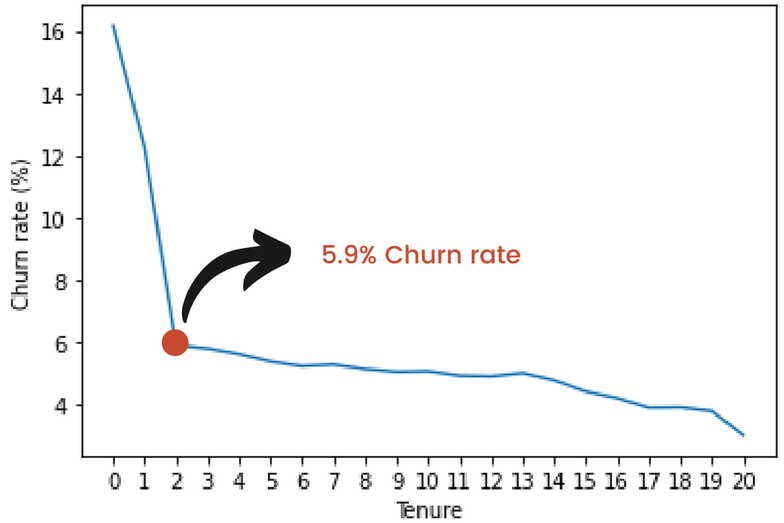
Business Recommendation

Di saat cashback minimal $240, churn rate berkurang lebih dari 3x lipat



96.4% customer yang churn berasal dari customer yang memiliki cashback dibawah $240

Churn rate turun secara signifikan saat tenure minimal 2 bulan



Churn rate pada customer dengan tenure dibawah 2 bulan sebesar 51.83%
65.5% customer yang churn berasal dari customer yang memiliki tenure dibawah 2 bulan

Rata-rata cashback awal = $150
$150 -> $240 = kenaikan 1.6x

Business Recomandation :

Pelanggan baru butuh diberikan **Cashback 1.6x** lebih besar sampai setidaknya **2 bulan**

Peningkatan jumlah cashback dapat meningkatkan total transaksi dan kemungkinan customer untuk bertransaksi kembali (Vana et al., 2018)

Source: Vana, P., Lambrecht, A., & Bertini, M. (2018). Cashback is cash forward: Delaying a discount to entice future spending. Journal of Marketing Research. https://doi.org/10.1509/jmr.15.0125

Profit Calculation

Dengan asumsi Shopful mendapatkan profit 5% dari setiap barang yang dijual

Total Cashback yang perlu diberikan kepada pengguna baru selama 2 bulan pertama	=	normal cashback * 1.6	=	2% * 1.6	=	3.2%	
Profit dari customer dengan cashback 2%	=	profit awal - cashback	=	5% - 2%	=	3%	
Profit dari customer dengan cashback 3.2%	=	profit awal - cashback	=	5% - 3.2%	=	1.8%	

Berdasarkan data bulan ini, Shopful mendapatkan 500 customer baru dan mempunyai 690 pelanggan dengan tenure 1 bulan dan 4000 pelanggan tetap (tenure di atas 2 bulan)

Simulasi

Before

Persentase customer baru (tenure < 2) yang akan tinggal = 100% - 51.83% = 48.17%

Persentase customer tenure minimal 2 yang akan tinggal = 100% - 5.9% = 94.1%

Bulan 1	tenure 0 bulan	tenure 1 bulan	
	500	690	
Profit	3%	3%	
	48.17%	48.17%	
Bulan 2		tenure 1 bulan	tenure >= 2 bulan
		241	332
Profit		3%	3%
		48.17%	94.1%
Bulan 3			tenure >= 2 bulan
			116 + 312 = 428
Profit			3%

Profit Calculation

Bulan 1 = 500 * 3% + 690 * 3% = 3570%

Bulan 2 = 241 * 3% + 332 * 3% = 1719%

Bulan 3 = 428 * 3% = 1284%

6573% +

After

Persentase pengguna baru (tenure < 2) yang akan tinggal = 100% - 5.06% = 94.94%

Persentase customer tenure minimal 2 yang akan tinggal = 100% - 5.9% = 94.1%

Bulan 1	tenure 0 bulan	tenure 1 bulan	
	500	690	
Profit	1.8%	1.8%	
	94.94%	94.94%	
Bulan 2		tenure 1 bulan	tenure >= 2 bulan
		475	655
Profit		1.8%	3%
		94.94%	94.1%
Bulan 3			tenure >= 2 bulan
			451 + 616 = 1067
Profit			3%

Profit Calculation

Bulan 1 = 500 * 1.8% + 690 * 1.8% = 2142%

Bulan 2 = 475 * 1.8% + 655 * 3% = 2820%

Bulan 3 = 1067 * 3% = 3201%

8163% +

24% profit increase

Simulasi keuntungan penggunaan machine learning dan insight

16.83%
Customer
Churn Rate

5.06% Churn rate (Cashback)

5.9% Churn rate (Tenure)

Turun hampir 3 kali lipat (Complain)

< 6%
Customer
Churn Rate

Simulasi keuntungan penggunaan machine learning dan insight

Shopful → Tanpa Prediksi & insight → 16.83% Customer Churn Rate → Potensi kerugian **$23,134,478**

Shopful → Churn prediction & insight → < 6% Customer Churn Rate → Potensi kerugian **$8,247,585**

Potensi penurunan kerugian = $23,134,478 - $8,247,585 = **$14,886,893**

The End

Terima kasih