DATA SCIENCE

Faisal Dhio Saputra
940266

Case Study

Prediksi pelanggan yang akan membayar tagihan Indihome tepat waktu. Sesuai kebijakan perusahaan, disebut tepat waktu apabila pelanggan membayar sebelum tanggal 21 untuk tagihan bulan berjalan. Jika pelanggan membayar di antara tanggal 21 sampai akhir bulan, maka pelanggan akan berstatus isolir dengan konsekuensi layanan Indihome akan diputus sementara hingga pelanggan melakukan pembayaran. Adapun jika pelanggan tidak membayar hingga akhir bulan, maka pada tagihan berikutnya pelanggan akan berstatus PraCT0/PraNPC.

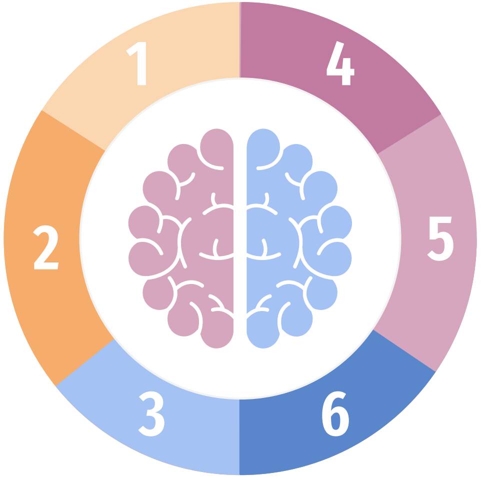
Prediksi pelanggan yang akan membayar tagihan Indihome tepat waktu akan bermanfaat untuk memetakan pelanggan yang low-risk dan high-risk, sehingga dapat dilakukan program loyalty yang tepat sasaran untuk meningkatkan *cash collection monthly ratio* (C3MR).

Flowchart

Importing Data To Python 1

Data Preprocessing 2

Exploratory Data Analysis and Feature Selection 3



4 Training Machine Model Learning

Mercury is the closest planet to the Sun

5 Evaluation

6 Deployment

Importing Data to Python

Load Data from Sql

Rename Column per month

Join table

Variable Dataset

MONTH	INET_PAY		POTS_PAY		INET_REV	POTS_REV
M1	Pay_status_inet_M1	Payment_inet_M1	Pay_status_pots_M1	Payment_pots_M1	Revenue_billing_inet_M1	Revenue_billing_pots_M1
M2	Pay_status_inet_M2	Payment_inet_M2	Pay_status_pots_M2	Payment_pots_M2	Revenue_billing_inet_M2	Revenue_billing_pots_M2
M3	Pay_status_inet_M3	Payment_inet_M3	Pay_status_pots_M3	Payment_pots_M3	Revenue_billing_inet_M3	Revenue_billing_pots_M3
M4	Pay_status_inet_M4	Payment_inet_M4	Pay_status_pots_M4	Payment_pots_M4	Revenue_billing_inet_M4	Revenue_billing_pots_M4
M5	Pay_status_inet_M5	Payment_inet_M5	Pay_status_pots_M5	Payment_pots_M5	Revenue_billing_inet_M5	Revenue_billing_pots_M5
M6	Pay_status_inet_M6	Payment_inet_M6	Pay_status_pots_M6	Payment_pots_M6	Revenue_billing_inet_M6	Revenue_billing_pots_M6

Variable Dataset

MONTH	INET_TICKET			POTS_TICKET		
M1	Tipe_gangguan_inet_M1	N_ticket_ticket_M1	Mttr_inet_M1	Tipe_gangguan_pots_M1	N_ticket_ticket_M1	Mttr_pots_M1
M2	Tipe_gangguan_inet_M2	N_ticket_ticket_M2	Mttr_inet_M2	Tipe_gangguan_pots_M2	N_ticket_ticket_M2	Mttr_pots_M2
M3	Tipe_gangguan_inet_M3	N_ticket_ticket_M3	Mttr_inet_M3	Tipe_gangguan_pots_M3	N_ticket_ticket_M3	Mttr_pots_M3
M4	Tipe_gangguan_inet_M4	N_ticket_ticket_M4	Mttr_inet_M4	Tipe_gangguan_pots_M4	N_ticket_ticket_M4	Mttr_pots_M4
M5	Tipe_gangguan_inet_M5	N_ticket_ticket_M5	Mttr_inet_M5	Tipe_gangguan_pots_M5	N_ticket_ticket_M5	Mttr_pots_M5
M6	Tipe_gangguan_inet_M6	N_ticket_ticket_M6	Mttr_inet_M6	Tipe_gangguan_pots_M6	N_ticket_ticket_M6	Mttr_pots_M6

Variable Dataset

MONTH	POTS_USAGE									
M1	Call_lokal_M1	Call_sljj_M1	Call_mobile_M1	Call_sli_M1	Call_other_M1	Duree_lokal_M1	Duree_sljj_M1	Duree_mobile_M1	Duree_sli_M1	Duree_other_M1
M2	Call_lokal_M2	Call_sljj_M2	Call_mobile_M2	Call_sli_M2	Call_other_M2	Duree_lokal_M2	Duree_sljj_M2	Duree_mobile_M2	Duree_sli_M2	Duree_other_M2
M3	Call_lokal_M3	Call_sljj_M3	Call_mobile_M3	Call_sli_M3	Call_other_M3	Duree_lokal_M3	Duree_sljj_M3	Duree_mobile_M3	Duree_sli_M3	Duree_other_M3
M4	Call_lokal_M4	Call_sljj_M4	Call_mobile_M4	Call_sli_M4	Call_other_M4	Duree_lokal_M4	Duree_sljj_M4	Duree_mobile_M4	Duree_sli_M4	Duree_other_M4
M5	Call_lokal_M5	Call_sljj_M5	Call_mobile_M5	Call_sli_M5	Call_other_M5	Duree_lokal_M5	Duree_sljj_M5	Duree_mobile_M5	Duree_sli_M5	Duree_other_M5
M6	Call_lokal_M6	Call_sljj_M6	Call_mobile_M6	Call_sli_M6	Call_other_M6	Duree_lokal_M6	Duree_sljj_M6	Duree_mobile_M6	Duree_sli_M6	Duree_other_M6

Variable Dataset

TARGET		POP						SPEED
Id_customer	Y	Length_of_stay	Divre_id	Technology	Kw	Indihome_type	Total_Minipack	Speed

Telkom Athon
Digital Talent Ready

AMOEBA Telkom Indonesia
the world in your hand

Data Preprocessing

Drop Duplicate	Check Missing Value	Remove ID
Separating X and Y	Test and Train Split	Select Numerical & Categorical variables
Outlier Handling for Numerical Data	Numerical Imputation	Categorical Imputation
Categorical Dummy Variables	Join Numerical & Categorical Data	

Exploratory Data Analysis & Feature Selection

Check Statistical Summary

Check Distribution

Check Relation between Independent Variables and Dependent Variables

Check Correlation

Select K Best

Training Machine Model Learning

Scaling

Regresion Logistic Model

Evaluation

Performance Training without Hyperparameter Tuning

Accuracy : 0.808500
Sensitivity : 0.542331
Spesificity : 0.921574
Precision : 0.746045
ROC AUC Score : 0.840675

Performance Testing without Hyperparameter Tuning

Accuracy : 0.809500
Sensitivity : 0.543981
Spesificity : 0.925638
Precision : 0.761890
ROC AUC Score : 0.843997

Performance Training with Hyperparameter Tuning

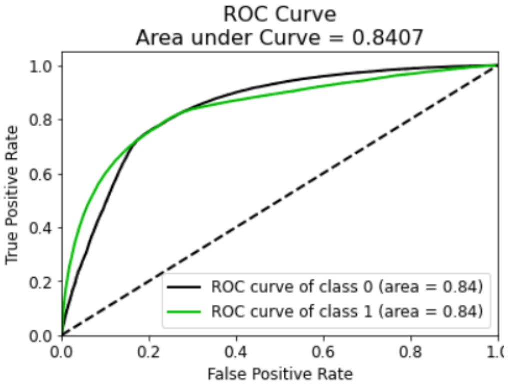
Accuracy : 0.808500
Sensitivity : 0.541565
Spesificity : 0.921899
Precision : 0.746565
ROC AUC Score : 0.840657

Performance Testing with Hyperparameter Tuning

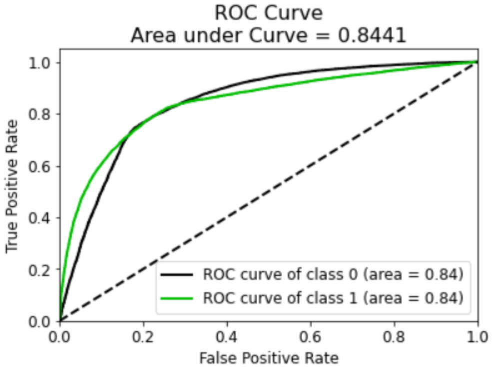
Accuracy : 0.809700
Sensitivity : 0.543543
Spesificity : 0.926118
Precision : 0.762915
ROC AUC Score : 0.844134

Evaluation

ROC Curve



ROC Training Set



ROC Testing Set

Telkom Athon
Digital Talent Ready

AMOEBA Telkom Indonesia
the world in your hand

Evaluation

Feature Importance

Pay_status_inet_M2_NOT_LATE	-0.255484426896891
Pay_status_inet_M4_NOT_LATE	-0.2385198772575212
Pay_status_inet_M1_LATE_SAME_MONTH	0.2322490724125231
Pay_status_inet_M3_NOT_LATE	-0.21614527291683056
Pay_status_inet_M1_NOT_LATE	-0.1858010402451609
Pay_status_inet_M6_NOT_LATE	-0.1643566394711911
Pay_status_inet_M6_LATE_SAME_MONTH	0.15776140383629983
Pay_status_inet_M5_NOT_LATE	-0.1426051132368522
Length_of_stay	-0.13771730965892082
Pay_status_inet_M3_LATE_SAME_MONTH	0.09766587437124673

Deployment

Importing Data Submission to Python

Get ID

Select Numerical & Categorical variables

Outlier Handling for Numerical Data

Numerical Imputation

Categorical Imputation

Get Selected Features

Scaling

Make Prediction

Make a DataFrame

Save Prediction

Thank You