



Churn prediction of telcom internet company

Lamzahhera Berinpalla

Education

Dibimbing

Trisakti School of Management

Experience

Ma Little Be

EY Indonesia

Other Project

Data Analyst & Data Science

- Beecycle Customer Segmentation
- A/B Testing
- Customer Satisfaction
- People Analytics
- Marketing Channel Analysis
- Churn Analysis on e-Commerce
- etc,



Overview

Background:

Churn pelanggan di industri telekomunikasi mengakibatkan penurunan pendapatan dan meningkatnya biaya akuisisi pelanggan baru. Dengan Memahami pola churn dapat membantu dalam mengambil tindakan pencegahan dan meningkatkan strategi churn.

Problem

Perlunya identifikasi pelanggan berisiko Churn untuk mengurangi mengakibatkan pendapatan tidak optimal dan inefisiensi operasional.

Statement:

Goals & Objectives:

- Mengidentifikasi faktor utama penyebab churn
- Membangun model prediktif untuk mengklasifikasikan pelanggan yang berisiko churn

Business Understanding

Problem

- Perusahaan kesulitan memprediksi pelanggan yang berisiko melakukan churn

Objective

- Mengembangkan model untuk memprediksi probabilitas churn pelanggan berdasarkan data dan riwayat transaksi

Goal

Memahami karakteristik pelanggan dan pola Churn.

Metrics

Precision, Recall, F1-Score, ROC-AUC

Data Info

Customer_Info.csv

Location_Data.csv

Online_Services.csv

Payment_Info.csv

Service_Options.csv

Status_Analysis.csv

KEY FEATURES ANALYSIS

Demographic Features:

- Age, Gender, Senior Citizen, Married, Dependents, ect

Service Features:

- Internet Service (No, DSL, Fiber, Cable)
- Online Security, Backup, Tech Support
- Streaming Services (TV, Movies, Music), ect

Financial Features:

- Contract Type, Payment Method
- Monthly Charge, Total Charges
- Tenure in Months, ect

Behavioral Features:

- Number of Referrals, Satisfaction Score
- Avg Monthly GB Download, ect

Data Cleaning

Drop Column

Unnecessary Features:

- Country, state, zip_code, city, latitude, longitude

Duplicate column

- phone_service_y and internet_service_y

Related Features:

- Churn_reason, churn_category, churn_score, churn_label, churn_status.





Data Cleaning

Missing Value

Internet_type 21.67%

Missing value to no

Offer 55.05%

Missing value to no offer

Handle Duplicated

No duplicate data found

Handle Outlier

No outlier found

PRE-PROCESSING DATA

Train-Test Split

80% Train: 20%
Test

Engineering & Feature Selection

encoding categorical
Standarisasi (minmaxscaler)

Multicollinearity check

Checking multicollinearity
and drop the variable



Model Selection

MODEL	DATA	METRICS			
		Precision	Recall	F1-Score	Roc-Auc
Decission Tree	Train	100%	100%	100%	100%
	Test	90.05%	87.40%	88.71%	91.96%
Random Forest	Train	100%	100%	100%	100%
	Test	96.16%	87.40%	91.57%	98.54%
XGBoost	Train	100%	100%	100%	100%
	Test	93.48%	88.47%	90.91%	99.08%
LightGBM	Train	100%	100%	100%	100%
	Test	94.32%	89.01%	91.59%	99.19%

LightGBM memiliki Skor F1 tertinggi pada set pengujian (91.59%) dengan tetap mempertahankan skor ROC AUC yang tinggi (99%). LightGBM memiliki selisih terkecil antara proses Train dan Test dibandingkan model lainnya. Hal ini menunjukkan keseimbangan presisi-recall yang kuat dan kemampuan klasifikasi keseluruhan yang sangat baik.

Optimizing LightGBM Parameters

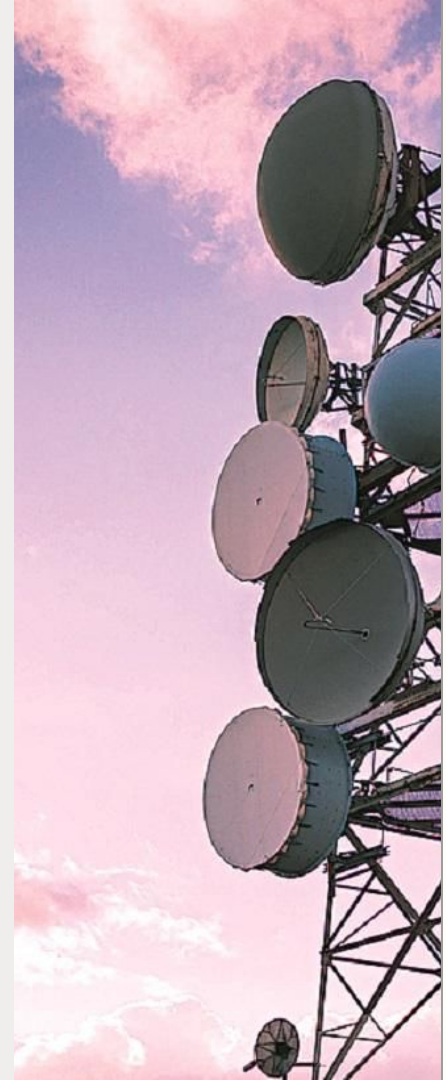


Model	Metrics			
	Precision	Recall	F1-Score	Roc-Auc
Base LightGBM (O=4138 I=1496)	94.32%	89.01%	91.59%	99.19%
Tuned LightGBM (Without SMOTE)	95.68%	89.01%	92.22%	99.24%
Tuned LightGBM With SMOTE	93.01%	92.76%	92.89%	98.93%

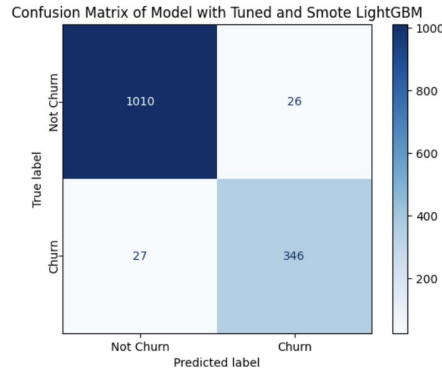
Parameter yang di tune: n_estimators, max_depth, learning_rate, subsample, colsample_bytree, reg_alpha, reg_lambda, num_leaves.

Dampak SMOTE: SMOTE kemungkinan menimbulkan kenaikan pada f1-score model.

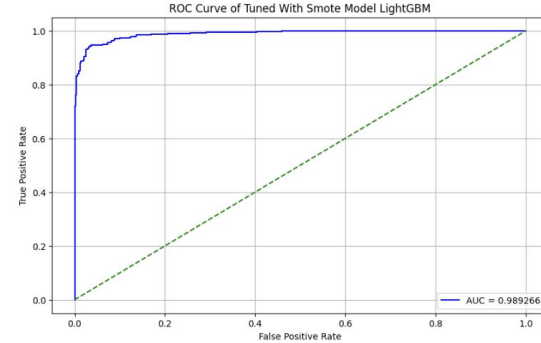
Model Terbaik: Model LightGBM yang Dituned (dengan SMOTE) memberikan kinerja tertinggi dan paling seimbang



MODEL ANALYSIS



- **True Positive (TP):** 1010 customers were correctly identified
- **False Positives (FP):** 26 customers were predicted to churn but turn out not churn
- **True Negatives (TN):** 346 customers were correctly predicted to churn
- **False Negatives (FN):** 27 customers were predicted to not churn but actually did churn



Model tersebut memiliki AUC sebesar 98,92% yang menunjukkan bahwa model tersebut secara efektif membedakan antara pelanggan yang cenderung membatalkan dan mereka yang tidak.

FEATURES IMPORTANCE

1. Monthly Charges (Biaya Bulanan)

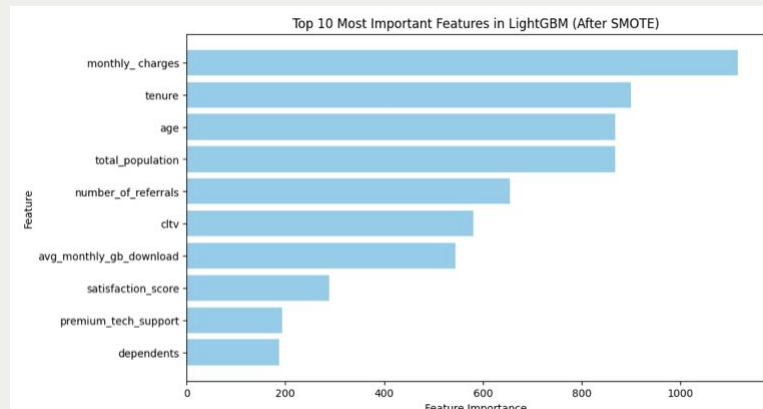
Pelanggan sangat sensitif terhadap kenaikan biaya bulanan. Pelanggan cenderung membandingkan value for money akan pergi jika merasa tidak mendapatkan nilai yang setara dengan uang yang dibayarkan.

2. Total Population (Kepadatan Penduduk)

Daerah dengan kepadatan penduduk tinggi menunjukkan pola churn yang berbeda dengan daerah jarang penduduk. Di area padat, persaingan lebih ketat sehingga pelanggan mudah berpindah ke provider lain. Sementara di daerah jarang penduduk, keterbatasan pilihan justru membuat pelanggan lebih bertahan meski dengan layanan yang kurang optimal.

3. Number of Referrals (Jumlah Referensi)

Pelanggan yang puas akan mereferensikan layanan kita ke orang lain. Sehingga ketika seorang pelanggan berhenti mereferensikan atau jumlah referensinya menurun, ini adalah early warning bahwa mereka mulai kehilangan antusiasme dan berpotensi meninggalkan layanan kita dalam waktu dekat.



Kesimpulan

Model **LightGBM** terpilih sebagai model terbaik dengan kinerja optimal (**F1-Score 92.89%**, **AUC 98.92%**) setelah melalui proses tuning dan penanganan data imbalance dengan SMOTE.

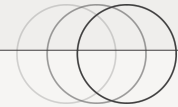
Analisis feature importance mengungkap **3 faktor utama** yang memengaruhi churn pelanggan:

1. **Biaya Bulanan (Monthly Charges)** - Sensitivitas harga pelanggan
2. **Jumlah Referral (Number of Referrals)** - Indikator early warning loyalitas
3. **Masa Langganan (Tenure)** - Stabilitas hubungan pelanggan

Temuan ini memberikan dasar yang kuat untuk menyusun strategi retensi yang tepat sasaran.



Recommendations



Strategi Penyesuaian Biaya Bulanan

Tinjau struktur harga untuk pelanggan dengan biaya tinggi dan ketidakpuasan. Tawarkan paket bernilai lebih melalui penambahan fitur premium atau bundling yang menguntungkan untuk meningkatkan loyalitas.

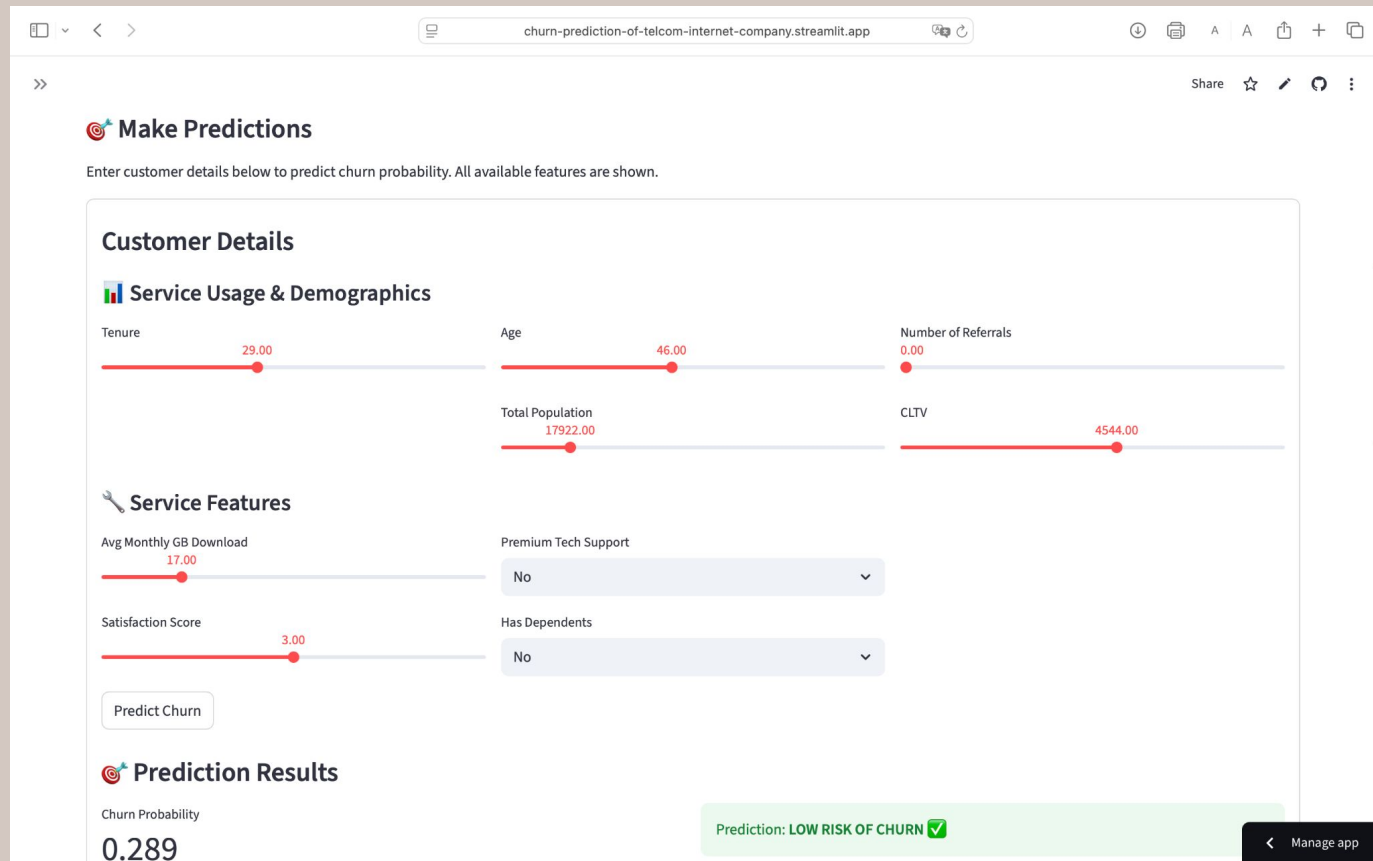
Optimalisasi Program Referral dan Data

Perkuat program referral dengan insentif menarik dan sesuaikan paket data untuk pengguna berat. Berikan reward proporsional untuk mempertahankan pelanggan aktif sebagai ambassador merek.

Paket Keluarga

Kembangkan paket keluarga komprehensif. Fokus pada harga kompetitif di daerah padat, serta keandalan layanan di wilayah jarang penduduk.

Streamlit





THANKYOU