

# PROPOSAL

## Uplift Modelling and Churn Prediction for Sleeping Customers

Practice Case 3 (Bussiness)  
by : Dhea Fajriati Anas



# FLAWS AND SUGGESTION

## FLAWS

- Dataset yang dibutuhkan tidak secara dijelaskan
- Model yang digunakan hanya random forest
- Pendekatan secara teknis tidak dijelaskan
- Melakukan analisa hanya secara prediktif
- Tidak memenuhi 6 elemen pembuatan proposal

## SUGGESTION

- Dataset yang dibutuhkan adalah data pribadi konsumen aktif dan tidak, data informasi transaksi konsumen aktif dan tidak, indikator churn, dan variabel penentu apakah pelanggan ditargetkan kampanye retensi atau tidak
- Model yang digunakan disarankan beberapa untuk menemukan yang terbaik seperti, xgboost, logistic regression, naive bayes, dan lainnya
- Pendekatan secara teknis perlu dijelaskan
- Melakukan analisa preskriptif juga yaitu, uplift modelling
- Pembuatan proposal perlu memenuhi 6 elemen

# TABLE OF CONTENTS

**01**

**PROJECT  
SUMMARY**

**02**

**PROPOSED  
TECHNICAL  
APPROACH**

**03**

**DATA SETS**

**04**

**SOFTWARE**

**05**

**MILESTONE**

# PROJECT SUMMARY

Telcozee adalah penyedia layanan komunikasi nirkabel terbesar di suatu negara yang memiliki jutaan pelanggan. Salah satu pelanggan potensial yang dapat meningkatkan profit adalah sleeping customers. Customers retention telah terbukti sangat menguntungkan bagi perusahaan karena (1) menarik klien baru menghabiskan biaya 5-6 kali lebih banyak daripada mempertahankan pelanggan yang ada, (2) pelanggan jangka panjang lebih menguntungkan, cenderung kurang sensitif terhadap kegiatan pemasaran yang kompetitif, cenderung lebih murah dalam pelayanan, dan dapat menghasilkan rujukan positif atau negatif kepada orang lain, (3) kehilangan pelanggan menyebabkan kerugian karena penjualan mengalami penurunan (Devriendt *et al.* 2019). Oleh karena itu, peningkatan kecil dalam retensi pelanggan dapat menghasilkan pengembalian yang signifikan.

Tim analisis mengusulkan untuk menganalisa secara prediktif dan preskriptif. Model Uplift adalah salah satu model machine learning yang tidak hanya memprediksi suatu hasil tetapi juga memberikan solusi. Model prediksi customers churn pada umumnya adalah memprediksi pelanggan mana yang akan churn dan tidak, kemudian menargetkan pelanggan yang paling berpotensi untuk diberikan kampanye retensi. Namun, apakah pelanggan yang diberikan kampanye retensi sudah tepat sehingga dapat dipertahankan atau sebaliknya. Model Uplift digunakan untuk memecahkan masalah terkait target yang tepat untuk diberikan kampanye retensi dari para pelanggan yang berpotensi berhenti menggunakan layanan. Penelitian yang dilakukan Devriendt *et al.* (2019) menghasilkan performa preskriptif yang lebih baik daripada model churn prediction.

# PROPOSED TECHNICAL APPROACH

## ALGORITMA

Algoritma yang digunakan pada prediksi churn dan uplift adalah extreme gradient boosting (XGBoost) dan random forest.

- **Churn Prediction (CP)** : memprediksi 2 target kelas yaitu, konsumen akan berhenti menggunakan layanan (churn) atau tidak (not churn)
- **Uplift Modelling** : memprediksi 4 target kelas yang merupakan tipe konsumen yaitu, control non-responder (CN), control responder (CR), treatment non-responder (TN), dan treatment responder (TR).

Metode yang digunakan pada preskripsi uplift modeling adalah Lai's Generalized Uplift Method (LGWUM). Metode tersebut dipilih karena menurut Kane *et al* , dari kedua model lainnya yaitu, two-model, dan treatment dummy, metode tersebutlah yang terbaik untuk menghasilkan skor uplift.

## PENGOLAHAN DATA

### • Pembersihan Data

- Periksa apakah ada missing values, duplicate, tipe data yang tidak sesuai dan lainnya
- Sesuaikan format data yang tidak sesuai. Label encoding dan one hot encoding digunakan untuk mengubah data non-numerik menjadi numerik.

### • Feature Engineering

Pengembangan model dari prediktif ke preskriptif dengan menambahkan fitur baru bernama Target Class, sehingga kurva Qini dapat dihasilkan untuk menghitung nilai uplift.

### • Resampling

Kelas target yang tidak seimbang diatasi dengan Synthetic Minority Over Sampling Technique (SMOTE).

# PROPOSED TECHNICAL APPROACH

- **Resampling**

Kelas target yang tidak seimbang diatasi dengan Synthetic Minority Over Sampling Technique (SMOTE).

- **Data Visualization**

Lakukan visualisasi data untuk mendapatkan insight.

- **Korelasi Treatment**

Fungsi yang digunakan adalah Pearson Correlation untuk menganalisa hubungan korelasi antara variabel customers treatment retention dengan customers churn.

## EVALUASI

Evaluasi terbagi dua yaitu metrik akurasi prediksi, dan preskripsi

- **Akurasi Prediksi** menggunakan confusion matrix untuk mengetahui akurasi dan mengamati hasil prediksi. Berikut rumus confusion matrix untuk CP dan uplift,

$$\text{Akurasi model CP} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$\text{Akurasi model Uplift} = \frac{\sum_{i=0}^3 \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{4} \times 100\%$$

- **Evaluasi Preskripsi** menggunakan kurva Qini. Analitik preskriptif suatu model dinilai berhasil apabila kurva uplift yang dihasilkan lebih tinggi dibandingkan dengan kurva Random Model. Random Model adalah garis yang menghubungkan nilai awal dan akhir pada kurva Qini, garis ini menandakan nilai uplift yang dihasilkan tanpa menggunakan model.

# DATA SETS

Dataset yang dibutuhkan terdiri atas,

- Data pribadi konsumen aktif dan tidak
- Data informasi transaksi aktif dan tidak
- Indikator churn
- Variabel penentu apakah pelanggan ditargetkan kampanye retensi atau tidak

Pelanggan yang dianggap churn adalah pelanggan yang tidak melakukan transaksi selama 90 hari setelah pemutusan kontrak terakhir.

# SOFTWARE

## Programming Language

Bahasa pemrograman yang digunakan adalah Python 3.5 dan library yang terdiri atas,

- Scikit Learn : library model algoritma xgboost, confusion matrix
- Seaborn : library untuk menampilkan visualisasi data
- Matplotlib : library untuk menampilkan visualisasi data
- Pandas : library untuk manipulasi data tabel
- Numpy : library untuk komputasi data numerik

## Tools

- Anaconda Navigator
- Jupyter Notebook
- Microsoft Excel
- Microsoft Power Point
- Database ex. Postgresql, mysql, dbeaver, dll
- Tableau



# MILESTONE

## **Bulan 1**

- Bussiness Understanding
- Data Understanding
- Data Preparation

## **Bulan 2**

- Modelling
- Evaluation

## **Bulan 3**

- Deployment
- Maintenance



**THANK  
YOU !**