Telco Customer Churn

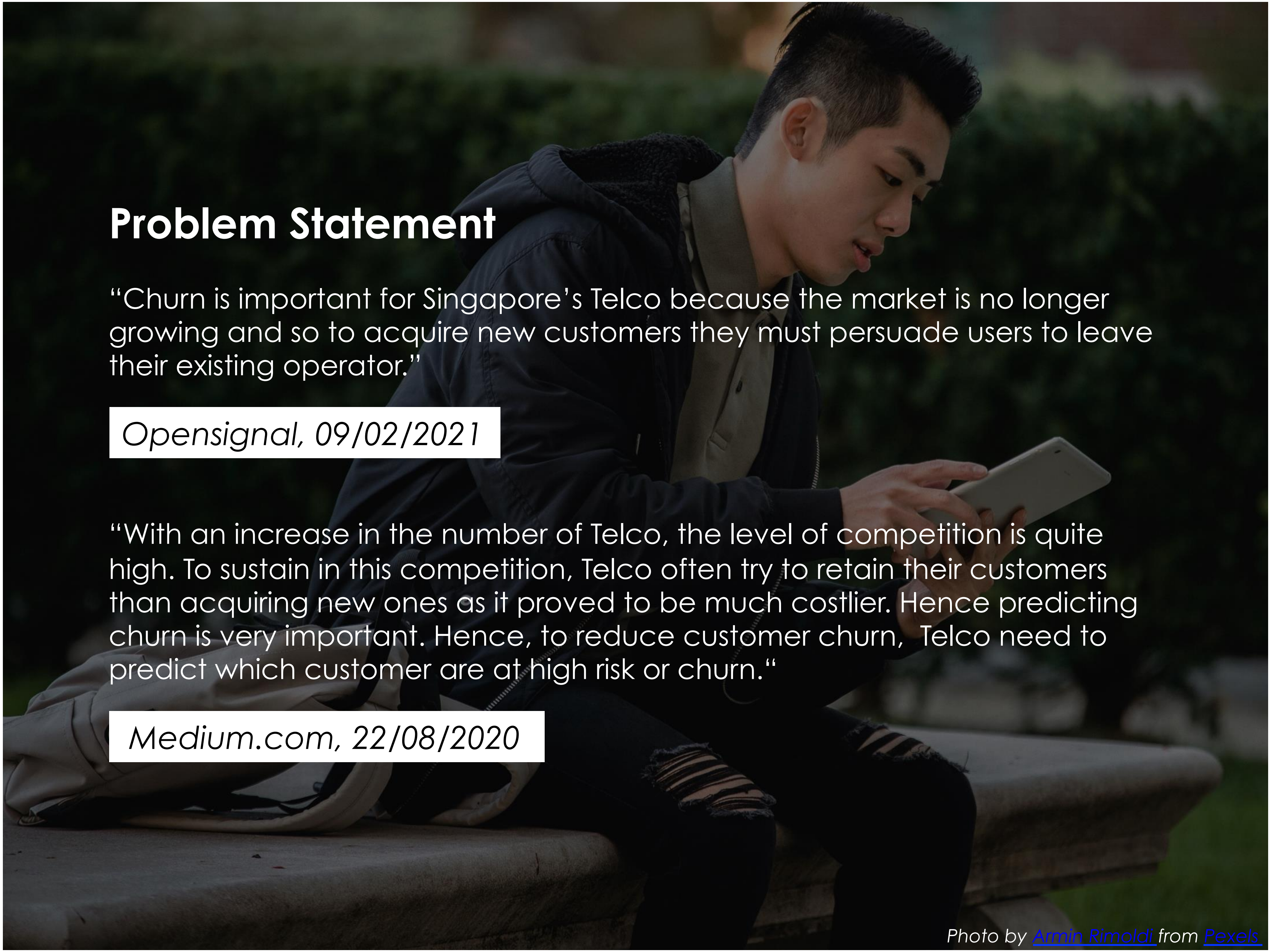
DS106 – Capstone Project Proposal

By Jia Yang

Photo by [Cristian Dina](#) from [Pexels](#)

Contents

Photo by Ketut Subiyanto from Pexels



Problem Statement

"Churn is important for Singapore's Telco because the market is no longer growing and so to acquire new customers they must persuade users to leave their existing operator."

Opensignal, 09/02/2021

"With an increase in the number of Telco, the level of competition is quite high. To sustain in this competition, Telco often try to retain their customers than acquiring new ones as it proved to be much costlier. Hence predicting churn is very important. Hence, to reduce customer churn, Telco need to predict which customer are at high risk or churn."

Medium.com, 22/08/2020

Photo by Armin Rimoldi from Pexels

Goal

To find out the most striking behaviour of customers through EDA and train the most predictive machine learning (ML) model to determine the customers who are most likely to churn.

Questions

- What's the % of churn and customers that still in active services?
- Can we see different patterns in churn customers based on the type of service provided?
- Any difference pattern of churn between demographics?
- Any difference between customers that pay monthly and by year?
- Any other questions will be raise through EDA…
- What‘s the best classification ML model for prediction?
- What’s the best hyper-parameters for selected ML model?

Dataset

File :
Dataset_(Jia Yang).csv

From :
www.kaggle.com

Description :
Each row represents a customer, each column contains customer's attributes such as Churn/No-Churn, Signed-up Services, Account Info and Demographic Info

21 Columns & 7043 Rows

Dataset

Total: 21 Columns & 7043 Rows

Churn/No-Churn

Target
D.Type: **Character**
Var.Type**: Categorical**
Column: **1**

Churn

Services

Predictor
D.Type: **Character**
Var.Type**: Categorical**
Column: **9**

Phone Service
Multiple Lines
Internet Service
Online Security
Online Backup
Device Protection
Tech Support
Streaming TV
Streaming Movies

Account Info

Predictor
D.Type: **Num./Char.**
Var.Type**: Cont./Cat.**
Column: **7**

Customer ID
Tenure
Monthly Charges
Total Charges
Contract
Paperless
Payment Method

Demographics

Predictor
D.Type: **Character**
Var.Type**: Categorical**
Column: **4**

Gender
Partners
Dependents
Senior Citizen

All columns are important as they are used as predictor and target

Dataset

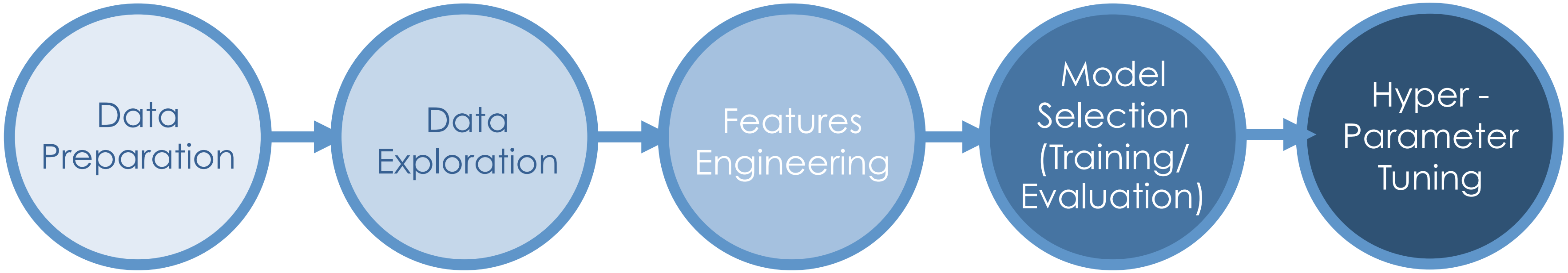
Snapshot

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No

5 rows × 21 columns

DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
Yes	No	No	No	One year	No	Mailed check	56.95	1889.5	No
No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
Yes	Yes	No	No	One year	No	Bank transfer (automatic)	42.30	1840.75	No
No	No	No	No	Month-to-month	Yes	Electronic check	70.70	151.65	Yes

Approach & Challenges



- Null values treatment
- Duplicate records
- Joining of datasets

- Relationship between variables
- Answer most of the identified questions

- Create new meaningful features
- Reduce unused features
- Data unbalance.
- Encode categorical features
- Scale overall dataset

- Choose the best Classification ML model based on different evaluation method

- Choose the best hyper parameter

Approach & Challenges

Classification Models Selection

Speed & Accuracy → (Speed) → Explainable → (No) → Data is Too Large →

Accuracy
- Random Forest
- XG boost

Yes
- Logistic Regression

Yes
- Naïve Bayes

No
- Linear SVM

Evaluation Methods

- Confusion Matrix
  - Accuracy
  - F1-Score
  - ROC / AUC

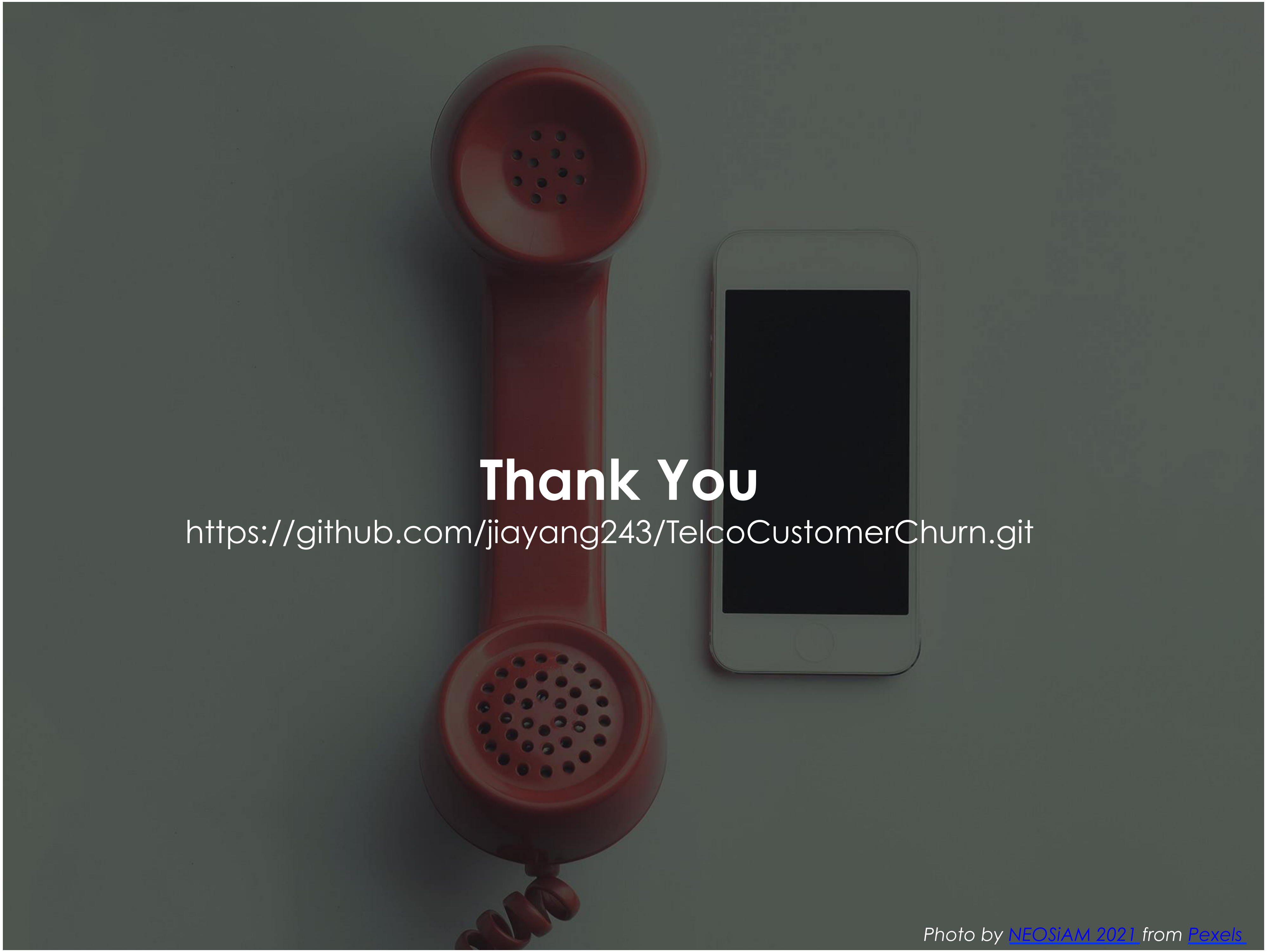


Photo by [NEOSiAM 2021] from [Pexels]