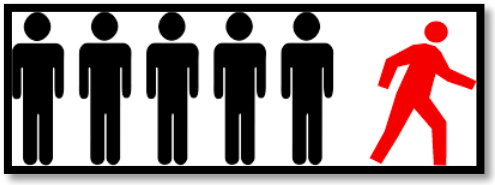
# IOD Capstone Project



# Customer Churn Prediction

# Cohort: ds-sg-pt-19jul21

| 0 | 12.1.2022 | Kong Wei Shen | Sifat Khan/<br>Gregory Murray |
|---|---|---|---|
| Revision | Date | Prepared by | Review By |

# Contents

# 1. Background

This project is discussed about customer churn in telco company. Churn means leaving the company or terminating a contract/service; thus, churn is a revenue loss to company. Nowadays customer retention is one of the primary KPI for company with subscription-based business model.

# 2. Scope

## 2.1 Problem Statement

- How to identify which are the customer who going to churn/ terminating the contract/services.
- What are the characteristics of customer who is going to churn?

## 2.2 Objective

- To develop a model to predict customer is churn.
- To identify features which lead to customer churn in a specific dataset.
- Forecast minimum retaining rate for surplus profit in a event a marketing campaign to be conducted.

# 3. Folder Structure

## 3.1 Readme – Documentation

- 1. Operation Manual.docx - Current document
- 2. Presentation -KWS.pptx - Summary of this document in presentation format.
- 3. Archive_Original_Notebook_All.ipynb - An Achieve of Jupiter notebook consists of full stack codes

## 3.2 Data

- The data set is collected from Kaggle
- Source: (https://www.kaggle.com/blastchar/telco-customer-churn)
- Telco-Customer-Churn.csv - For training and testing)
- Telco-Customer-Churn - Deployment.csv - For deployment purpose
- Telco-Customer-Churn - Deployment (Batch).csv - For deployment purpose in simplified contain.

## 3.3 EDA (Exploratory Data Analysis)

- EDA.ipynb - Jupiter notebook for data profiling and exploration purpose.

## 3.4 Pre-process & Model Development

- Preprocess.ipynb - Jupiter notebook contain all pre-process in a function.

- Modelling.ipynb - Jupiter notebook for model development.
- model.sav - Model used for prediction purpose.

### 3.5 Deployment

- Deployment.ipynb - Jupiter notebook use for prediction. (Only activate this notebook for production)
- Preprocess.py - Python file created from Preprocess.ipynb. This python file needs to be in this directory for deployment usage.

## 4. Data Input & Output

### 4.1 Input File

- The input file to be store in folder structure as in 3.2
- Format shall be csv, xls any version readable by python.

### 4.2 Default File

- Telco-Customer-Churn - Deployment.csv - Serve as default file for deployment.

### 4.3 Output File

- Prediction output such as Results.csv shall be save to folder structure 3.5.

## 5. Workflow

### 5.1 Data Exploration

- Activate notebook state in 3.3.

### 5.2 Prediction Deployment

- Activate notebook state in 3.5

### 5.3 Modification/ Troubleshooting

- Activate notebook state in 3.4 (Preprocess.ipynb) for alteration in any pre-processing. Upon completion, download the python file and save into directory in 3.5.
- Activate notebook state in 3.4 (Modelling) for modifying the predictive model. If you wish to save the new file in another name, rename the file not in 'model.sav'.

## 6. Exploratory Data Analysis

EDA contain of data profiling of target and features. All contain shall be listed as below:

- Data correlation
- Churn distribution.

- Bar plot of all features category by Churn
- Scatter plot using 2 features with category,
- Count plot using create new features consists of all additional services.

## 7. Pre-process

This section included data cleaning, pre-processing, and feature engineering. Refer below list

- Drop the rows with missing values.
- Consolidating wording with same meaning
- Change data type from object to float
- Check columns with two classes
- Label Male & Female to 1 & 0
- Convert Yes & No to 1 & 0
- Creating dummy columns for features with multiple classes and non continues datapoint
- Rename column to presentable format
- Create updated columns
- Convert all columns to float type
- Convert the data object into NumPy array otherwise you will not be able to impute
- Define the criteria for dealing with the missing values
- Transform to impute data
- Scale data between 0 & 1 and normalisation
- Move the data back to a dataframe

## 8. Model Development

Several models are developed to compare the predictive performance. Hyperparameter tuning by using Grid Search CV. Ensembled method by using Ada Boosting. Models developed as below:

- Logistic Regression
- Support Vector Machine
- K Nearest Neighbours
- Random Forest
- Decision Tree
- AdaBoosting

# 9. Evaluation

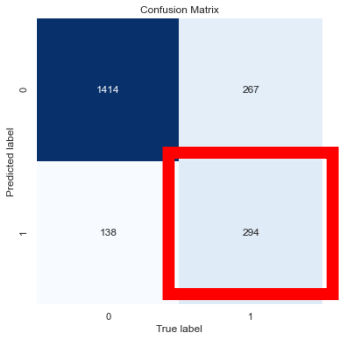
## 9.1 Model Evaluation

Evaluation metrics as below:

- Accuracy
- F1 Score
- Precision
- Recall
- ROC_AUC
- Confusion Matrix

## 9.2 Business Evaluation

Forecast minimum retaining rate base on confusion matrix.




Assumption:

Using median charges for churn customer as sales monthly sales per person.

Campaign cost 20% of the monthly sales.

Monthly Charges = **$79.25**

Campaign Cost = $79.25 x 20% = **$15.85**

Total cost for campaign is False Negative + True Negative = 138+294 x (15.85)

Profit at 20% Customer NOT Churn

Sales - Cost

= $ (79.25 x 294 x 20%)- (138+294)x(15.85)

=$ 4659.6- 6847.2 = - 2187.6

Profit at 29.39% Customer NOT Churn

Sales - Cost

= $ (79.25 x 294 x 29.39%)- (138+294)x(15.85)

=$ 6847.2 - 6847.2 = 0

Profit at 80% Customer NOT Churn

Sales - Cost

= $ (79.25 x 294 x 80%)- (138+294)x(15.85)

=$ 18639.6- 6847.2 = 11792.4

# 10. Deployment

Activate notebook in structural folder state in 3.5.