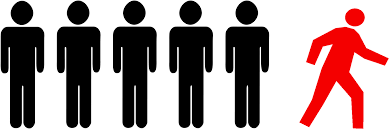
**IOD Capstone Project**



**Customer Churn Prediction**

**Cohort: ds-sg-pt-19jul21**

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

Contents

[**1.** **Background** 3](#_Toc92841011)

[**2.** **Scope** 3](#_Toc92841012)

[2.1 Problem Statement 3](#_Toc92841013)

[2.2 Objective 3](#_Toc92841014)

[**3.** **Folder Structure** 3](#_Toc92841015)

[3.1 Readme – Documentation 3](#_Toc92841016)

[3.2 Data 3](#_Toc92841017)

[3.3 EDA (Exploratory Data Analysis) 3](#_Toc92841018)

[3.4 Pre-process & Model Development 3](#_Toc92841019)

[3.5 Deployment 4](#_Toc92841020)

[**4.** **Data Input & Output** 4](#_Toc92841021)

[4.1 Input File 4](#_Toc92841022)

[4.2 Default File 4](#_Toc92841023)

[4.3 Output File 4](#_Toc92841024)

[**5.** **Workflow** 4](#_Toc92841025)

[5.1 Data Exploration 4](#_Toc92841026)

[5.2 Prediction Deployment 4](#_Toc92841027)

[5.3 Modification/ Troubleshooting 4](#_Toc92841028)

[**6.** **Exploratory Data Analysis** 4](#_Toc92841029)

[**7.** **Pre-process** 5](#_Toc92841030)

[**8.** **Model Development** 5](#_Toc92841031)

[**9.** **Evaluation** 6](#_Toc92841032)

[9.1 Model Evaluation 6](#_Toc92841033)

[9.2 Business Evaluation 6](#_Toc92841034)

[10. Deployment 7](#_Toc92841035)

# **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.

# **Scope**

## 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?

## Objective

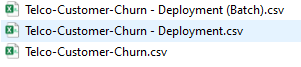
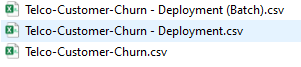
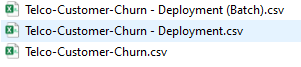
* To develop a model to predict customer is churn.
* To identify features which lead to customer churn in a specific dataset.
* Forecast minimum retaining rate for surplus profit in a event a marketing campaign to be conducted.

# **Folder Structure**

## Readme – Documentation

* - Current document
* - Summary of this document in presentation format.
* - An Achieve of Jupiter notebook consists of full stack codes

## Data

* The data set is collected from Kaggle
* Source: (<https://www.kaggle.com/blastchar/telco-customer-churn>)
* - For training and testing)
* - For deployment purpose
* - For deployment purpose in simplified contain.

## EDA (Exploratory Data Analysis)

* - Jupiter notebook for data profiling and exploration purpose.

## Pre-process & Model Development

*  - Jupiter notebook contain all pre-process in a function.
* - Jupiter notebook for model development.
*  - Model used for prediction purpose.

## Deployment

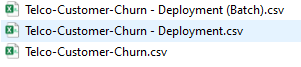
*  - Jupiter notebook use for prediction. (Only activate this notebook for production)
*  - Python file created from . This python file needs to be in this directory for deployment usage.

# **Data Input & Output**

## 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.

## Default File

*  - Serve as default file for deployment.

## Output File

* Prediction output such as shall be save to folder structure 3.5.

# **Workflow**

## Data Exploration

* Activate notebook state in 3.3.

## Prediction Deployment

* Activate notebook state in 3.5

## 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’.

# **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.

# **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

# **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

# **Evaluation**

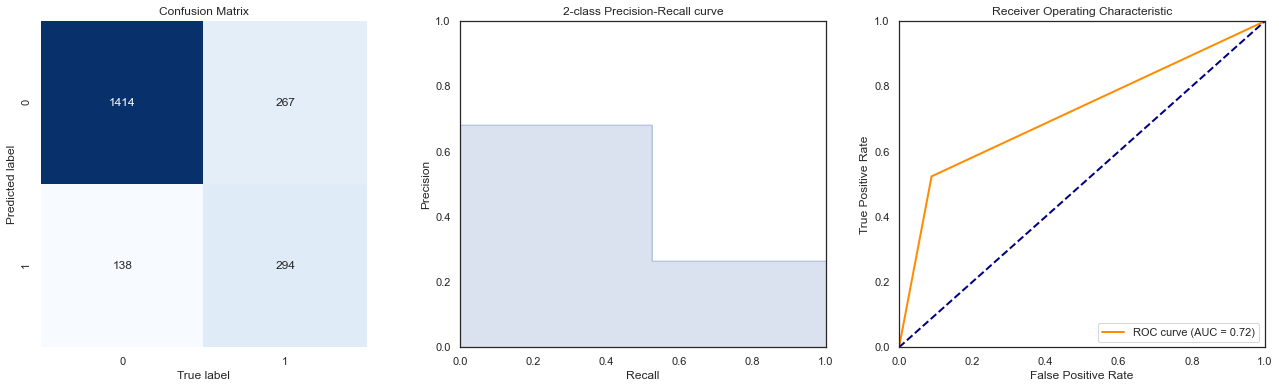
## Model Evaluation

Evaluation metrics as below:

* Accuracy
* F1 Score
* Precision
* Recall
* ROC\_AUC
* Confusion Matrix

## 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

# **Deployment**

Activate notebook in structural folder state in 3.5.