

1102_資料分析與學習基石 FUNDAMENTALS OF DATA ANALYTICS

Personal Term Project – Proposal

姓名：何子安

系級：心理 110

學號：E44065020

About Dataset

Dataset

Telco Customer Churn (Customer Churn Prediction)

Context

"Predict behavior to retain customers. You can analyze all relevant customer data and develop focused customer retention programs." [IBM Sample Data Sets]

Content

Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:

- Customers who left within the last month – the column is called Churn.
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies.
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges.
- Demographic info about customers – gender, age range, and if they have partners and dependents.

Goal

This is a binary classification task. Given the customer's personal data, (try using different models from classical to modern ML models and compare between those) predict whether the customer churned or not (Yes or No).

Dedicate

If our models can predict well on whether the customer will churn or not, company can make use of this on their advertising/marketing to offer promotion for the customers that are going to leave.

Before Model

I am going to do EDA to see if there are any way to represent the data better for our models to fit. Furthermore, I will consider doing normalization and deal with data imbalanced.