[客戶續約金額預測]

設計文件

隊伍: [TT]

成員：[PAN KUAN LUN]

# 摘要

特徵主要採用統計的特徵，使用 LightGBM 做模型，用 cv 的方式來選取最佳參數。

# 環境

Win10、32GB、i7、Python、Pandas、lightgbm

# 特徵

Premium

Insured\_Amount1/2/3

Insured\_Amount1+Insured\_Amount2+Insured\_Amount3

Coverage\_Deductible\_if\_applied

統計sum、min、max、mean、std。

count\_insurance\_coverage

count\_cancellation\_1

previous\_total\_premium

ratio\_cancellation\_1

ratio\_insured\_premium

information of each insurance’s coverage

# 訓練模型

用lightgbm regression，cv找最佳參數。預測值是Next\_Premium/total\_Premium。

# 訓練方式及原始碼

如Github。

# 結論

比較晚繳交文件，能夠先看一下別人的做法檢討自己的不足哈。這次的比賽中，我使用的方法沒有太特別的地方，後來嘗試做些特徵轉換和降維，但並未有大幅改善。受限於時間的關係，也就沒有多嘗試其他方法。觀察前幾名的文件，發現我應是主要有幾點的不足：

1. 沒有訓練分類模型，來判斷是否會續約後，再用迴歸模型預測值。
2. 沒有注意到隱藏在資料中的一些特殊特徵
3. 特徵工程做的很基礎，也未做特徵的篩選