

資料科學與人工智慧競技

Final Project Presentation 2

Predict Future Sales

組員：製造所 P96104170 林柏呈



Rank

1854

Johnson0930



0.88610



Your Best Entry!

Your most recent submission scored 0.88610, which is an improvement of your previous score of 0.90484. Great work!



Johnson0930

[Add occupation](#)

[Add organization](#)

[Add location](#)

Joined 7 months ago · last seen in the past day

[Home](#) [Competitions \(2\)](#) [Datasets](#) [Code \(2\)](#) [Discussion](#) [Followers](#) [Notifications](#) [Account](#)

Account (User ID 8688618)

User Name

johnson0930

Your username cannot be changed.

Email Address

p96104170@gs.ncku.edu.tw

Phone Verification

Verified

Result Observation

	XGBoost	LightGBM	CatBoost
Score	0.91285	0.88610	0.89494

My Thought

資料處理

- 刪除離群值

- 3511 項商品單日交易筆數是小於 0 的
- 7356 筆資料交易筆數是小於 0 的
- 2 筆資料是離群值 (交易筆數 > 750)

- 特徵添加

- 嘗試添加銷售資料 滯後 1, 2, 3, 6, 12 個月
- 應 test.csv 提供的 shop_id、item_id，將 商店與商品 類別分類 (經過分析發現關聯度較小，故移除)

- 商店與商品相關時間編號

- 初次銷售、最後一次銷售 等

My Thought

選用的特徵：

- 'date_block_num'
- 'shop_id'
- 'item_id'
- 'item_cnt_month'
- 商店地區/編號
- 月銷售滯後 1, 2, 3, 6 個月
- 日商品銷售滯後 1, 2, 3, 6 個月
- 日商店銷售滯後 1 個月
- 月份/ 日期
- 初次銷售、最後一次銷售

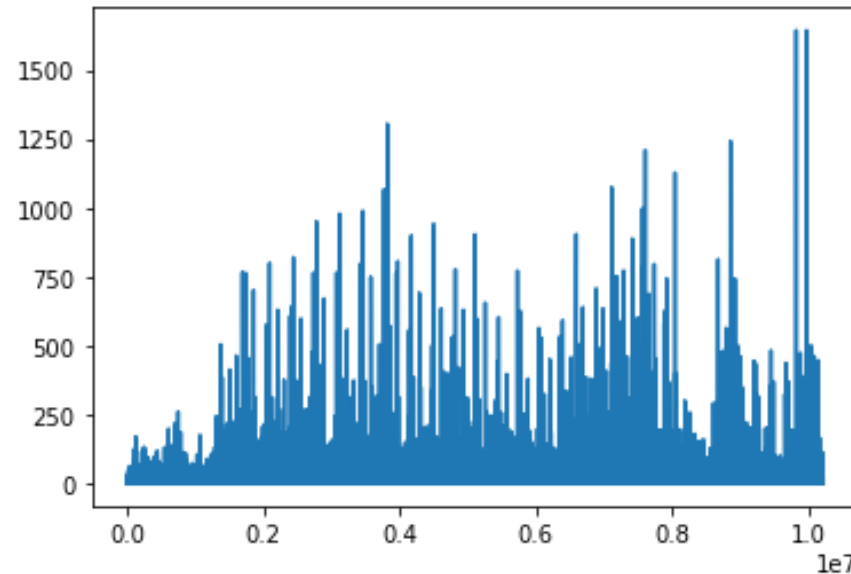
My Thought

模型訓練

- 選用 XGBoost、LightGBM、CatBoost 個別訓練
- 使用 sklearn gridsearchcv 窮舉，協助調正 hyperparameter

Try & Error Experience

- sales_train.csv 提供的 item_cnt_day 欄位在加總計數轉換為每個月的銷售量的數值後，在使用之前要把數值調整為設置的下限和上限。



Thanks for listening