



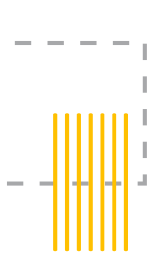
MeBike

共享單車使用量 預測系統

Group 6

廖宇凡、黃顥

賴品均、陳怡誼



Outline

1

背景介紹
與問題討論

2

研究方法

3

個案研究

4

網頁設計

5

結論與未來展望





1

背景介紹與 問題討論



背景介紹

- 「YouBike 微笑單車」是一個提供24小時甲租乙還租賃服務的電子無人自動化管理公共自行車系統，此系統的特點包含高品質的營運服務，舒適好騎的腳踏車，以及方便註冊及使用。



現況分析-5W1H





研究方法

模型選擇

條件：

1. 擅長處理時間序列數據
2. 準確的預測能力
3. 高維特徵處理



XGBoost

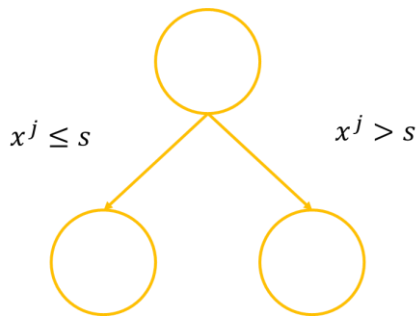
- 強大的預測能力：XGboost是boosting**集成學習**的其中一種，是將許多樹模型(弱分類器)集成在一起，形成一個很強的分類器。而所用到的樹模型則是**CART回歸樹模型**。
- 高效的計算性能：可有效計算**大規模數據**和**高維特徵**。

集成學習

- Boosting流派：各分類器之間有依賴關係，例如：XGboost
- Bagging流派：各分類器之間沒有依賴關係，例如隨機森林 (Random Forest) 。

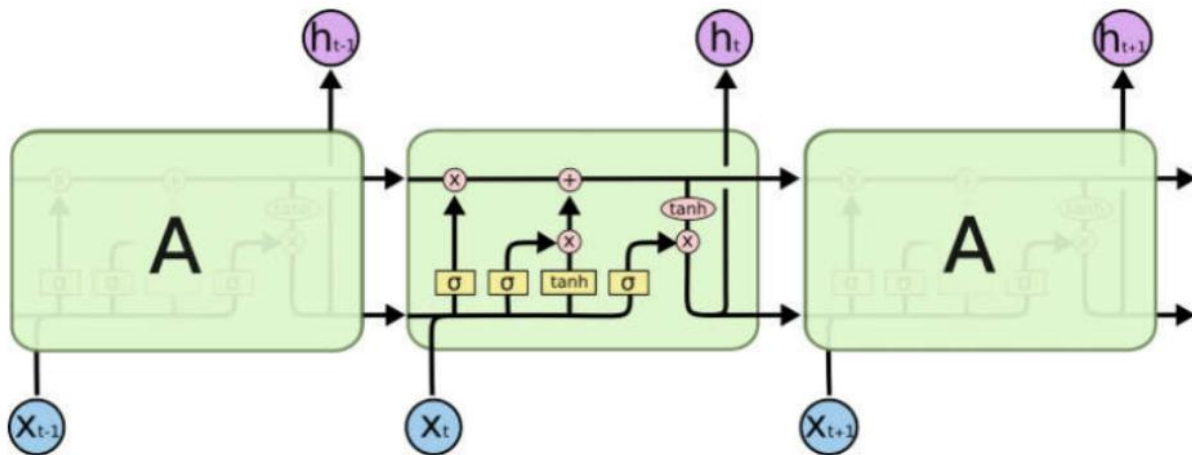
CART回歸樹模型

- CART回歸樹是假設樹為二叉樹，不斷將特徵行分裂。



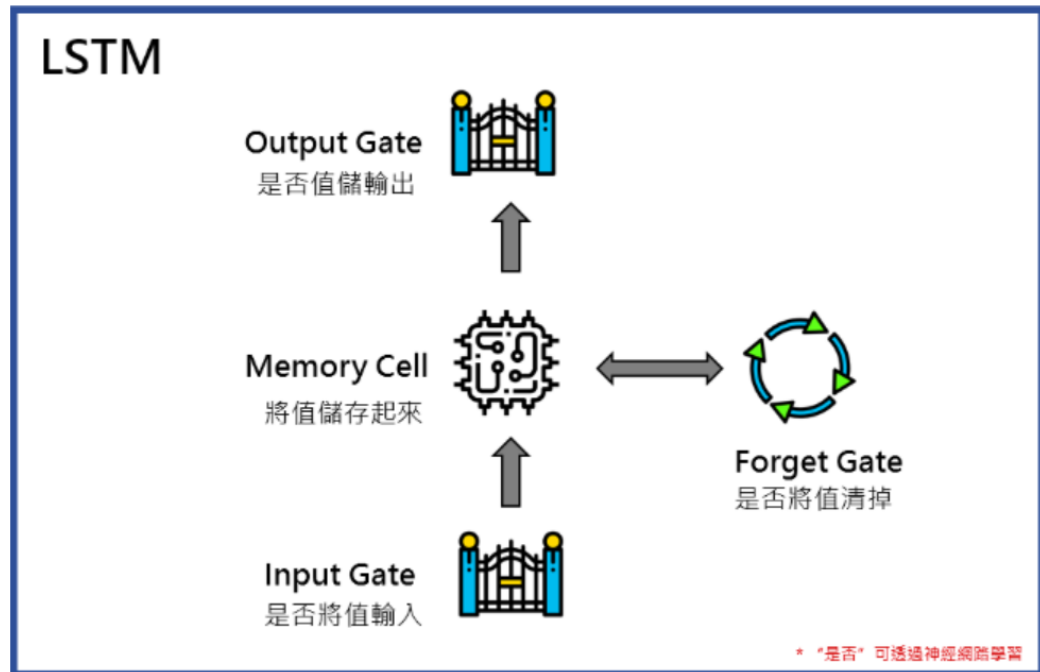
LSTM

- 特殊的循環神經網路
- 擅長處理**序列數據**：可捕捉時間相關性或長期依賴關係
- 具有記憶單元：可解決RNN在長期序列訓練過程中**梯度消失**或**梯度爆炸**的問題



LSTM

- **忘記階段**→針對輸入值執行選擇性遺忘
- **選擇記憶階段**→根據重要程度決定記憶重點
- **輸出階段**→決定輸出因子





個案研究

資料集-倫敦共享單車

- 資料來源：倫敦交通局公開資料
- 時間範圍：2015.1~2017.1
- 資料筆數：17415

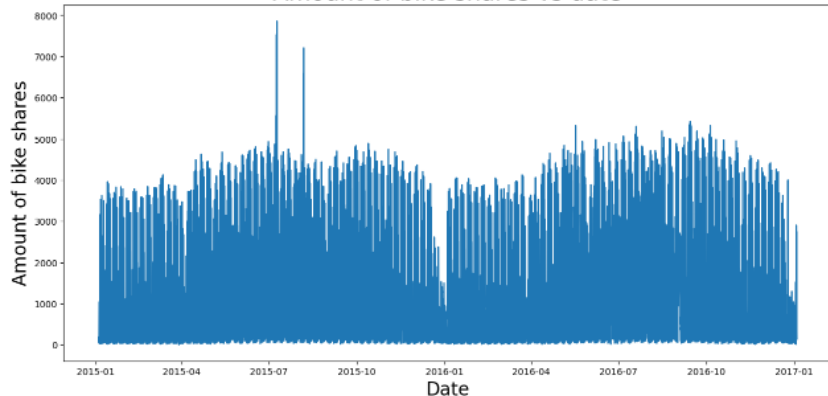
```
df = pd.read_csv("/content/drive/MyDrive/Youbike dataset/london_merged.csv")
df.head()
```

	timestamp	cnt	t1	t2	hum	wind_speed	weather_code	is_holiday	is_weekend	season
0	2015-01-04 00:00:00	182	3.0	2.0	93.0	6.0	3.0	0.0	1.0	3.0
1	2015-01-04 01:00:00	138	3.0	2.5	93.0	5.0	1.0	0.0	1.0	3.0
2	2015-01-04 02:00:00	134	2.5	2.5	96.5	0.0	1.0	0.0	1.0	3.0
3	2015-01-04 03:00:00	72	2.0	2.0	100.0	0.0	1.0	0.0	1.0	3.0
4	2015-01-04 04:00:00	47	2.0	0.0	93.0	6.5	1.0	0.0	1.0	3.0

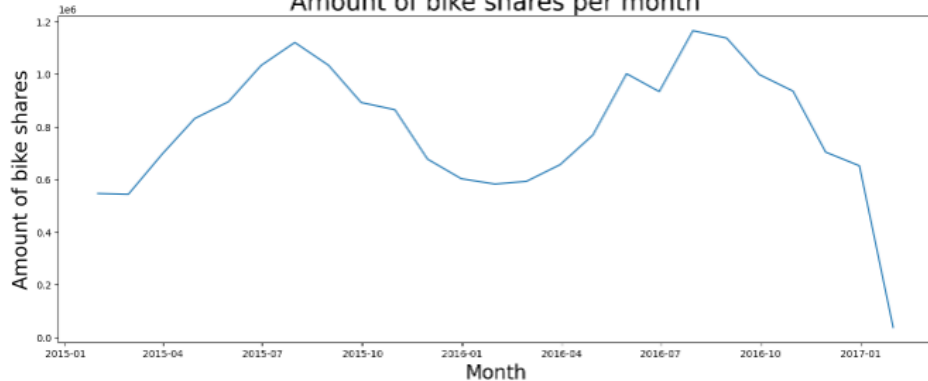
資料欄位	說明
timestamp	時間戳記
cnt	共享單車數量
t1	實際溫度(攝氏)
t2	體感溫度(攝氏)
hum	濕度百分比
wind_speed	風速(公里/小時)
weather_code	天氣類型
is_holiday	是否為節日
is_weekend	是否為周末
season	季節

資料集-倫敦共享單車時間趨勢

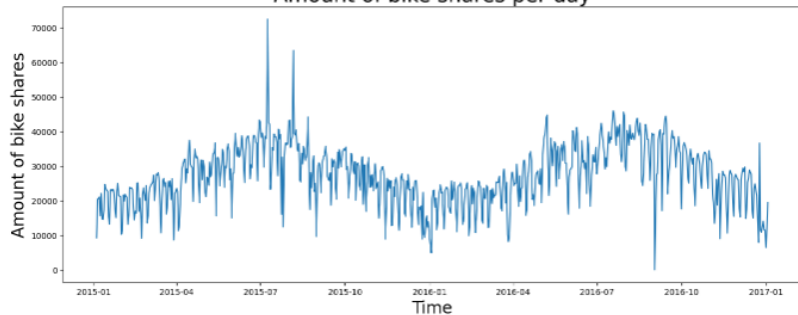
Amount of bike shares vs date



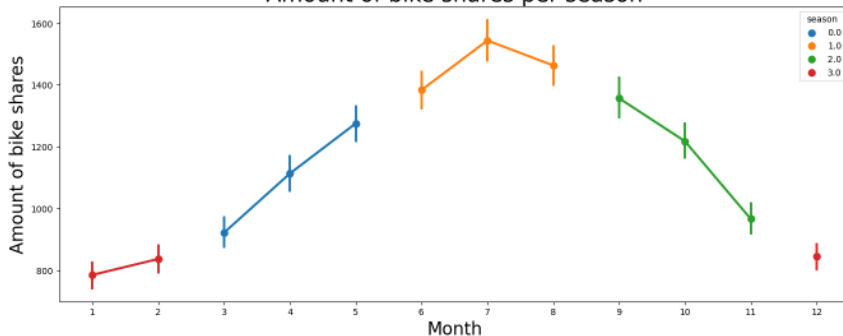
Amount of bike shares per month



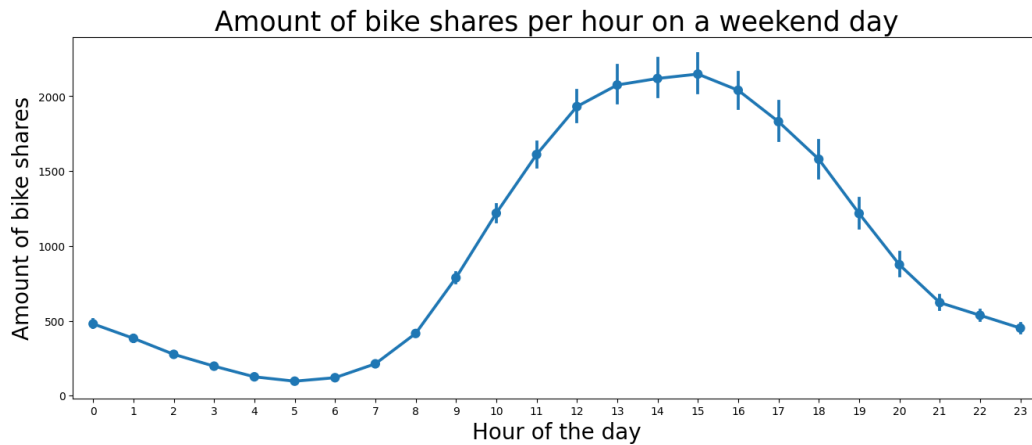
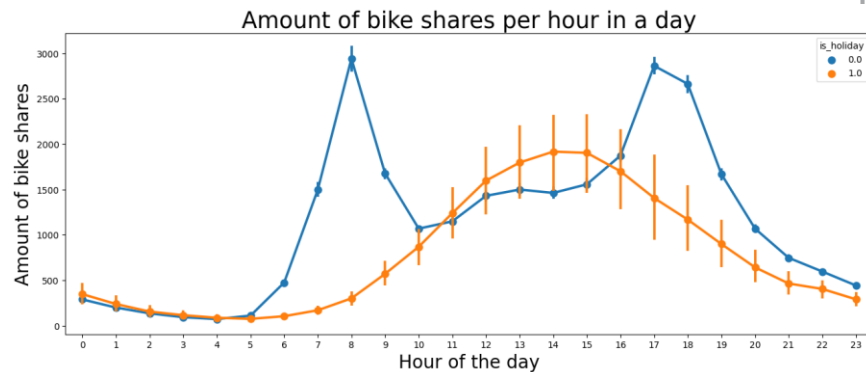
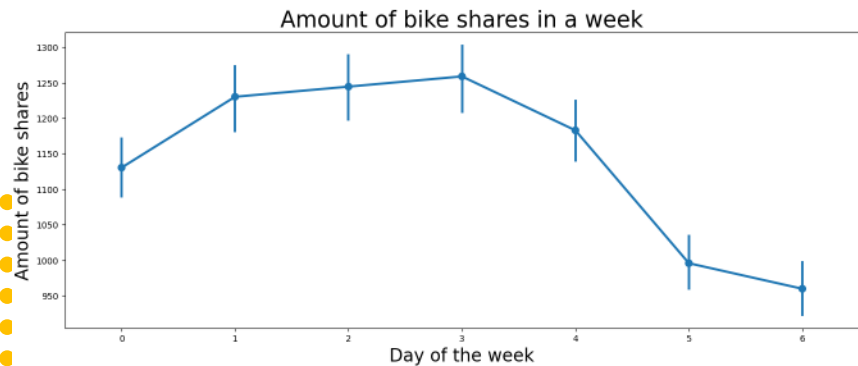
Amount of bike shares per day



Amount of bike shares per season

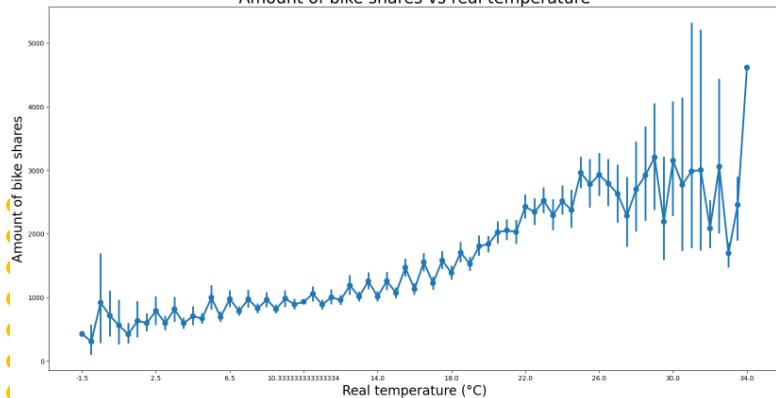


資料集-倫敦共享單車時間趨勢

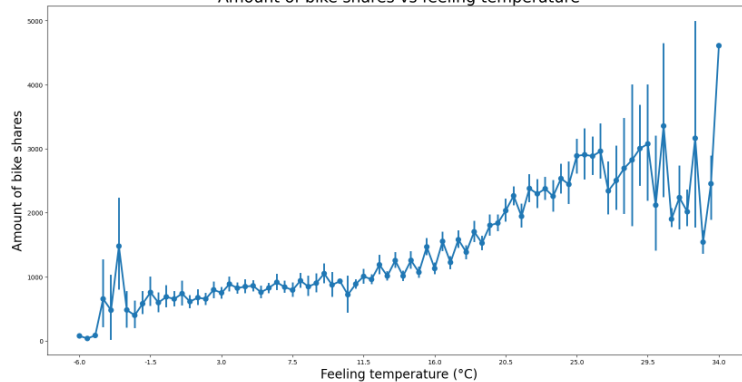


資料集-倫敦共享單車屬性關係

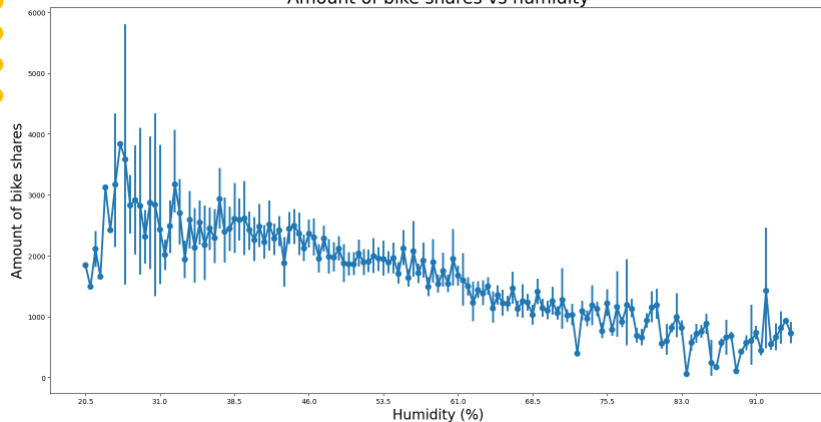
Amount of bike shares vs real temperature



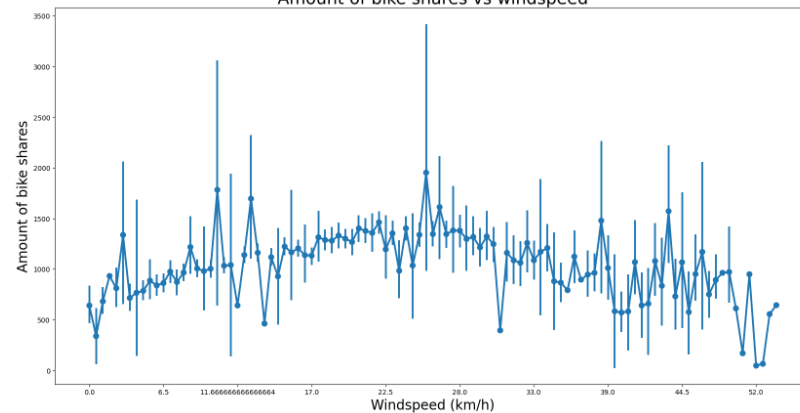
Amount of bike shares vs feeling temperature



Amount of bike shares vs humidity

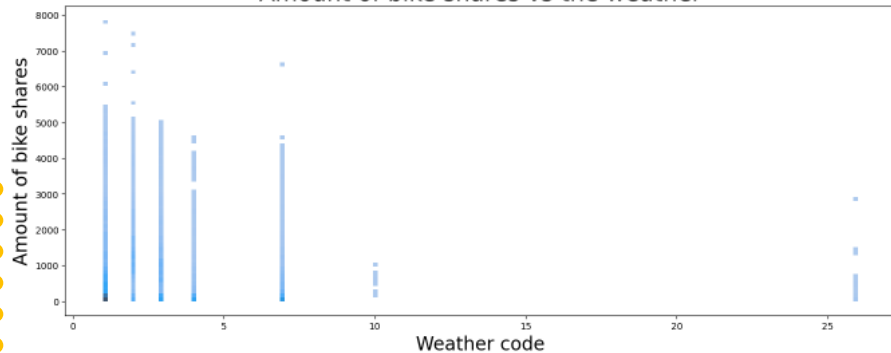


Amount of bike shares vs windspeed



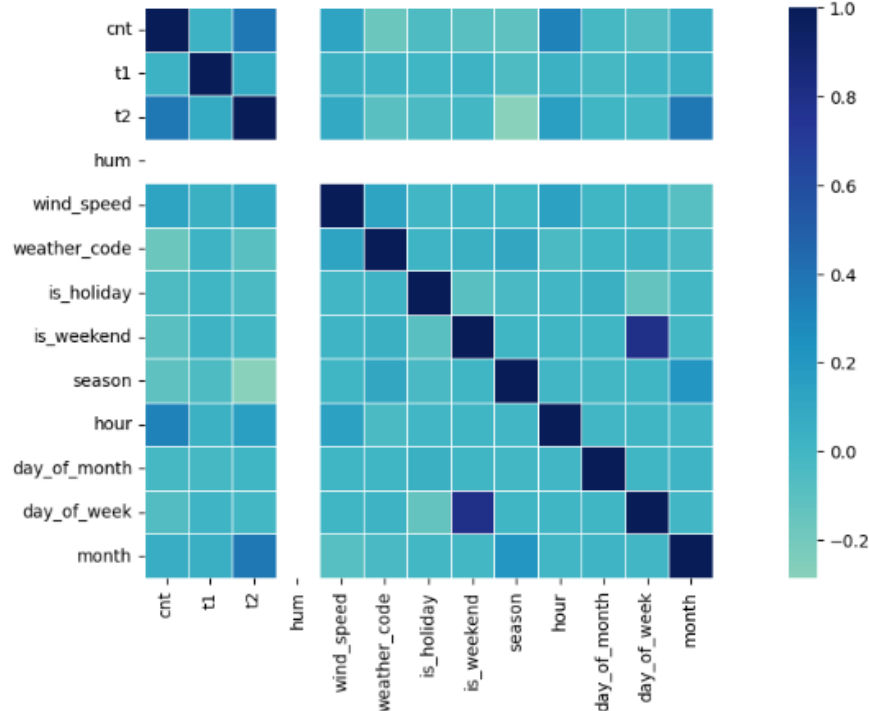
資料集-倫敦共享單車屬性關係

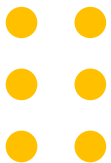
Amount of bike shares vs the weather



代號	說明
1	晴朗
2	分散的雲層
3	破碎的雲層
4	多雲
7	下雨
10	雷雨
26	下雪

<Axes: >





XGBoost



XGBoost-資料前處理

- 決定輸入特徵x和目標變量y

```
X = data.drop(['timestamp', 'cnt'], axis=1)
y = data['cnt']
```

X：去除時間戳記和共享單車數量的其他元素
Y：共享單車數量

- 分割訓練集及測試集(7:3)

```
def df_split(df, train_percent):
    split_index = int(train_percent * len(df))
    train = df.iloc[:split_index]
    test = df.iloc[split_index:]
    return train, test
```

```
[ ] X_train, X_test = df_split(X, 0.7)
    y_train, y_test = df_split(y, 0.7)
```

XGBoost-模型建立

- 使用XGBRegressor創建回歸模型xgbmodel

```
from xgboost.sklearn import XGBRegressor
from sklearn.metrics import mean_absolute_error, r2_score, mean_squared_log_error, mean_squared_error, make_scorer
```

```
[26] xgbmodel = XGBRegressor()
      xgbmodel.fit(X_train, y_train)

      preds = xgbmodel.predict(X_test)
```

- 使用RMSE來評估實際結果和預測結果之間的差異

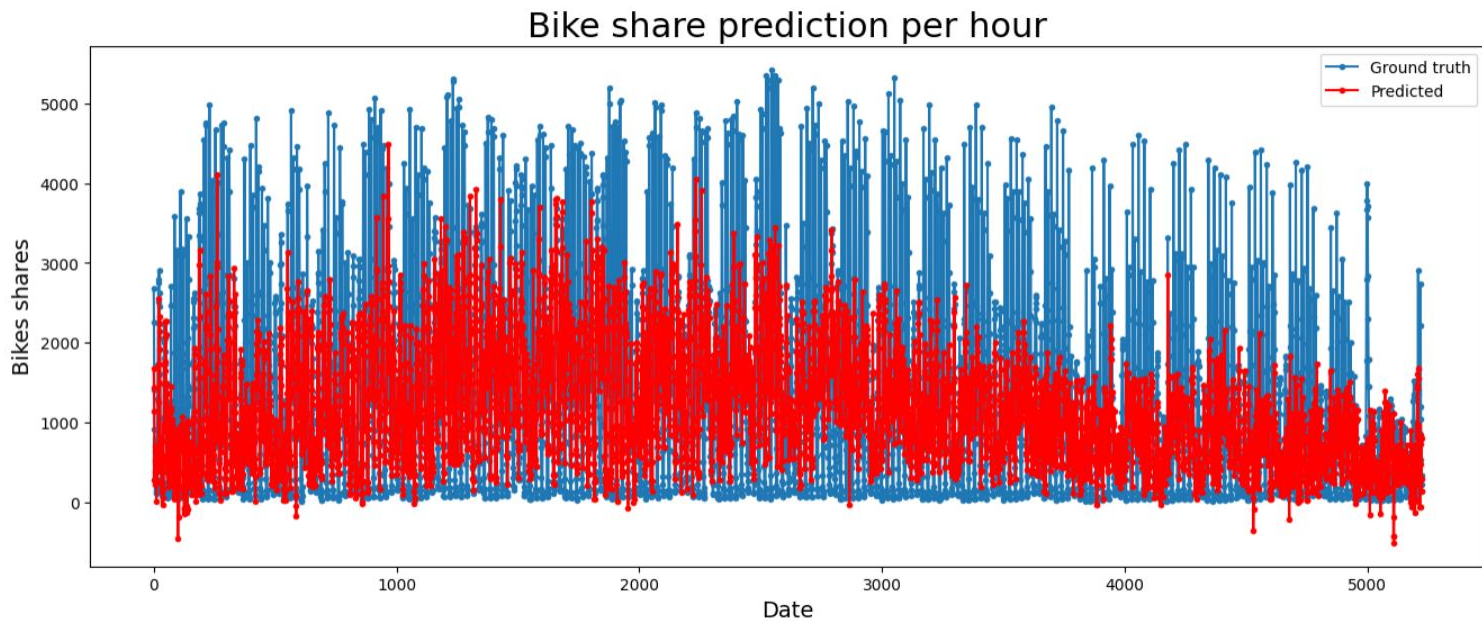
```
import numpy as np
rmse = np.sqrt(mean_squared_error(y_test, preds))
print("RMSE: %f" % (rmse))
```

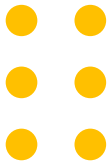
RMSE: 1015.137644

- 使用平方項，將誤差轉為正數，同時放大較大誤差的懲罰值
- 獲得能解釋的單位

XGBoost-預測結果

- 折線圖比較-實際值和預測值





LSTM



LSTM –資料前處理

- 分割訓練集和測試集(9 : 1)

```
import math
from sklearn.preprocessing import RobustScaler

training_data_len = math.ceil(len(df) *.9) # taking 90% of data to train and 10% of data to test
testing_data_len = len(df) - training_data_len

time_steps = 24
train, test = df.iloc[0:training_data_len], df.iloc[(training_data_len-time_steps):len(df)]
print(df.shape, train.shape, test.shape)
```

```
(17414, 13) (15673, 13) (1765, 13)
```

- 用RobustScaler進行特徵縮放

```
[ ] # Scale the all of the data from columns ['t1', 't2', 'hum', 'wind_speed']
train_trans = train[['t1', 't2', 'hum', 'wind_speed']].to_numpy()
test_trans = test[['t1', 't2', 'hum', 'wind_speed']].to_numpy()

scaler = RobustScaler() # Handles outliers
train.loc[:, ['t1', 't2', 'hum', 'wind_speed']] = scaler.fit_transform(train_trans)
test.loc[:, ['t1', 't2', 'hum', 'wind_speed']] = scaler.fit_transform(test_trans)

#Scale the all of the data from columns ['cnt']
train['cnt'] = scaler.fit_transform(train[['cnt']])
test['cnt'] = scaler.fit_transform(test[['cnt']])
```

處理每個特徵值域與單位不同的問題

LSTM – 模型建立

- 建立LSTM模型

```
from keras.preprocessing import sequence
from keras.models import Sequential
from keras.layers import Dense, Dropout, LSTM, Bidirectional

model = Sequential()
model.add(Bidirectional(LSTM(50, input_shape=(x_train.shape[1], x_train.shape[2]))))
model.add(Dropout(0.2))
model.add(Dense(units=1))

model.compile(optimizer="adam", loss="mse")

# prepared_model = model.fit(X_train, y_train, batch_size=32, epochs=100, validation_data=[X_test, y_test])

history = model.fit(x_train, y_train, epochs=150, batch_size=32, validation_split=0.3, shuffle=True)
```

LSTM –超參數優化



	Level 1	Level 2	Level 3
Dropout	0.1	0.2	0.3
Batch size	16	24	32
Validation split	0.1	0.2	0.3



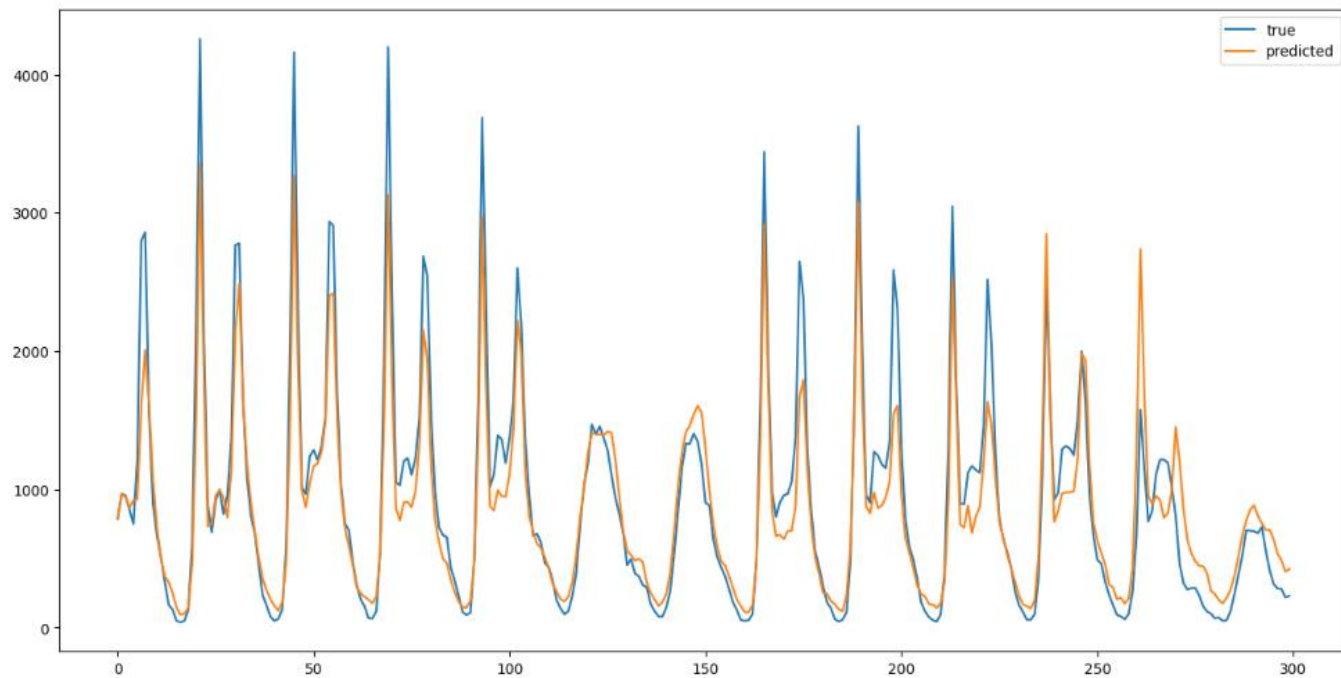
	Dropout	Batch size	Validation split	R-sq
1	0.1	16	0.1	0.8189
2	0.1	24	0.2	0.78
3	0.1	32	0.3	0.8385
4	0.2	16	0.2	0.83
5	0.2	24	0.1	0.8
6	0.2	32	0.3	0.845
7	0.3	16	0.2	0.796
8	0.3	24	0.3	0.83
9	0.3	32	0.1	0.82

LSTM – 模型訓練

```
Epoch 136/150
343/343 [=====] - 5s 14ms/step - loss: 0.0118 - val_loss: 0.0712
Epoch 137/150
343/343 [=====] - 6s 18ms/step - loss: 0.0116 - val_loss: 0.0672
Epoch 138/150
343/343 [=====] - 5s 14ms/step - loss: 0.0121 - val_loss: 0.0619
Epoch 139/150
343/343 [=====] - 6s 17ms/step - loss: 0.0139 - val_loss: 0.0645
Epoch 140/150
343/343 [=====] - 5s 15ms/step - loss: 0.0118 - val_loss: 0.0656
Epoch 141/150
343/343 [=====] - 5s 14ms/step - loss: 0.0112 - val_loss: 0.0633
Epoch 142/150
343/343 [=====] - 6s 18ms/step - loss: 0.0111 - val_loss: 0.0731
Epoch 143/150
343/343 [=====] - 5s 14ms/step - loss: 0.0113 - val_loss: 0.0623
Epoch 144/150
343/343 [=====] - 6s 18ms/step - loss: 0.0131 - val_loss: 0.0635
Epoch 145/150
343/343 [=====] - 5s 14ms/step - loss: 0.0116 - val_loss: 0.0615
Epoch 146/150
343/343 [=====] - 5s 15ms/step - loss: 0.0111 - val_loss: 0.0763
Epoch 147/150
343/343 [=====] - 6s 16ms/step - loss: 0.0123 - val_loss: 0.0639
Epoch 148/150
343/343 [=====] - 5s 14ms/step - loss: 0.0175 - val_loss: 0.0616
Epoch 149/150
343/343 [=====] - 6s 18ms/step - loss: 0.0126 - val_loss: 0.0630
Epoch 150/150
343/343 [=====] - 5s 14ms/step - loss: 0.0110 - val_loss: 0.0564
```

LSTM –預測結果

- 折線圖比較-實際值和預測值



LSTM – 評估指標

- 使用**RMSE**來評估實際結果和預測結果之間的差異
- 使用 **R^2 score**來衡量模型的解釋能力

```
[ ] from sklearn.metrics import mean_squared_error, r2_score  
    rmse_lstm = np.sqrt(mean_squared_error(y_test, y_pred))  
    rmse_lstm
```

RMSE

```
365.1102269162983
```

```
[ ] r2_lstm = r2_score(y_test, y_pred)  
    r2_lstm
```

R^2 score

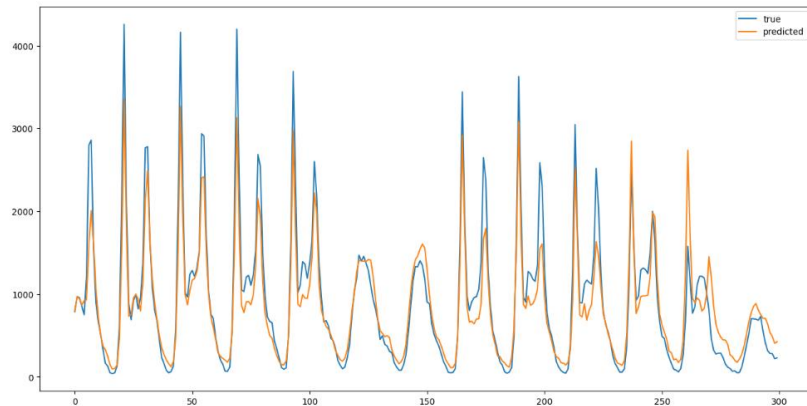
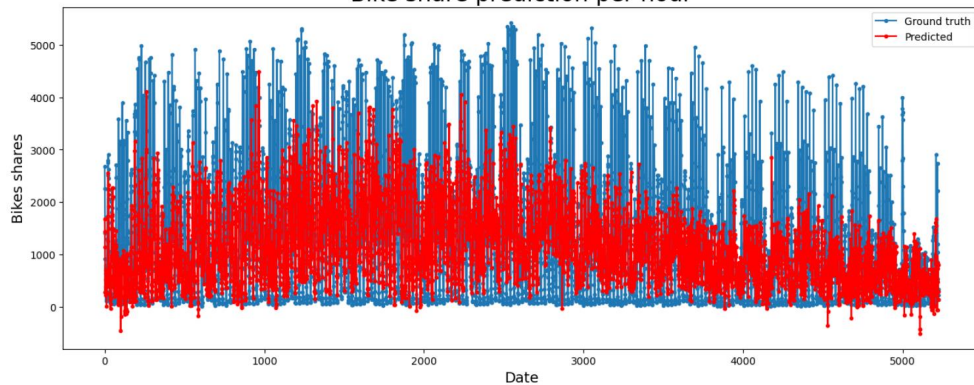
```
0.8456010779991341
```

特徵x可以解釋目標變量y的比例

結果比較

	XGBoost	LSTM
RMSE	1015.13	365.11

Bike share prediction per hour





網頁設計

資料集-Youbike2.0台北市公共自行車即時資訊

- 資料來源：政府資料開放平臺
- 時間範圍：即時資訊

df_combined

	sno	sna	tot	sbi	sarea	mday	lat	lng	ar	sareaen	snaen	aren	bemp	act	srcUpdateTime	updateTime	infoTime	infoDate
0	500101001	YouBike2.0_捷運科技大樓站	28	0	大安區	2023-05-31 00:32:13	25.02605	121.54360	復興南路二段235號前	Daan Dist.	YouBike2.0_MRT Technology Bldg. Sta.	No.235, Sec. 2, Fuxing S. Rd.	25	1	2023-05-31 09:16:22	2023-05-31 09:16:26	2023-05-31 00:32:13	2023-05-31
1	500101002	YouBike2.0_復興南路二段273號前	21	0	大安區	2023-05-30 23:50:09	25.02565	121.54357	復興南路二段273號西側	Daan Dist.	YouBike2.0_No.273, Sec. 2, Fuxing S. Rd.	No.273, Sec. 2, Fuxing S. Rd. (West)	21	1	2023-05-31 09:16:22	2023-05-31 09:16:26	2023-05-30 23:50:09	2023-05-30
2	500101003	YouBike2.0_國北教大貴小東側門	16	1	大安區	2023-05-31 00:24:13	25.02429	121.54124	和平東路二段96巷7號	Daan Dist.	YouBike2.0_NTUE Experiment Elementary School (...)	No. 7, Ln. 96, Sec. 2, Heping E. Rd	13	1	2023-05-31 09:16:22	2023-05-31 09:16:26	2023-05-31 00:24:13	2023-05-31
3	500101004	YouBike2.0_和平公園東側	11	1	大安區	2023-05-30 23:06:14	25.02351	121.54282	和平東路二段118巷33號	Daan Dist.	YouBike2.0_Heping Park (East)	No. 33, Ln. 118, Sec. 2, Heping E. Rd	10	1	2023-05-31 09:16:22	2023-05-31 09:16:26	2023-05-30 23:06:14	2023-05-30
4	500101005	YouBike2.0_辛亥復興路口西北側	16	3	大安區	2023-05-30 23:10:20	25.02153	121.54299	復興南路二段368號	Daan Dist.	YouBike2.0_Xinhai Fuxing Rd. Intersection (Nor...	No. 368, Sec. 2, Fuxing S. Rd.	13	1	2023-05-31 09:16:22	2023-05-31 09:16:26	2023-05-30 23:10:20	2023-05-30

資料集-Youbike2.0台北市公共自行車即時資訊

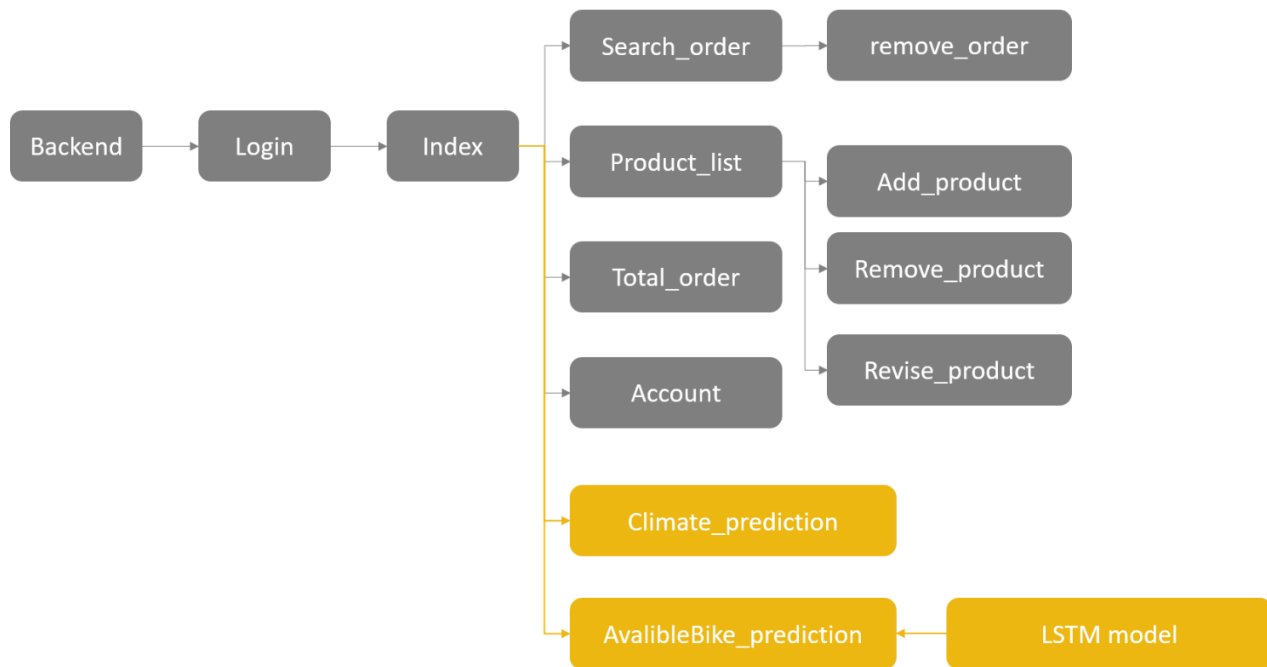
資料欄位	說明	資料欄位	說明
sno	站點代號	sareaen	場站區域英文
sna	場站中文名稱	snaen	場站名稱英文
tot	場站總停車格	aren	地址英文
sbi	場站目前車輛數量	bemp	空位數量
sarea	場站區域	act	全站禁用狀態
mday	資料更新時間	srcUpdateTime	YouBike2.0系統發布資料更新的時間
lat	緯度	updateTime	大數據平台經過處理後將資料存入DB的時間
lng	經度	infoTime	各場站來源資料更新時間
ar	地點	infoDate	各場站來源資料更新時間

資料集-天氣

- 資料來源：open weather公開資料
- 時間範圍：即時資訊

資料欄位	說明
weather.description	天氣類型
main.temp	實際溫度(攝氏)
main.feels_like	體感溫度(攝氏)
main.humidity	濕度
wind.speed	風速(公里/小時)

網頁架構圖-後台





Mbike

員工後臺系統

請輸入帳號密碼

帳號

密碼

登入



5

結論與未來 展望



針對這次改善成果，透過LSTM模型設計**站點使用量預測系統**，可以幫助調度人員更靈活的安排單車後勤作業，不再像過往作業需要高度依賴人員巡邏，更可以大幅降低站點滿車以及空車率，提供客戶更好的使用體驗。**未來可再結合線上預約系統**，可以將即時資訊及預約數量結合，預測每個站點的流量，進而超前部署，達到利用率最大化，提升顧客滿意度。



Thank you

