

國立清華大學

工業工程與工程管理學系

智慧化企業整合

MeBike 微笑單車使用量預測系統

指導老師：邱銘傳 教授

組別：第 6 組

組員：廖宇凡

黃顥

賴品均

陳怡誼

中 華 民 國 一 一 二 年 六 月 十 六 日

目錄

目錄.....	I
圖目錄.....	II
表目錄.....	III
一、 簡介.....	1
1.1 背景動機與目的.....	1
1.2 5W1H 分析.....	1
二、 研究方法.....	1
2.1 XGBoost.....	1
2.2 LSTM.....	1
三、 實驗設計與結果.....	2
3.1 資料集介紹.....	2
3.2 模型建立.....	8
3.2.1 XGBoost 模型.....	8
3.2.2 LSTM 模型.....	9
3.3 LSTM 模型訓練結果.....	11
四、 網頁設計與後臺管理系統.....	12
4.1 數據來源.....	12
4.2 後臺管理系統.....	13
4.3 網頁設計.....	13
五、 結論與未來展望.....	14
六、 參考資料.....	15

圖目錄

圖 1 以日期的單車租借量.....	2
圖 2 以月為單位的單車租借量.....	3
圖 3 以日為單位的單車租借量.....	3
圖 4 以季為單位的單車租借量.....	3
圖 5 一週的單車租借量.....	4
圖 6 一天內各小時的單車租借量.....	4
圖 7 假日各小時的單車租借量.....	4
圖 8 實際溫度對單車租借量的影響.....	5
圖 9 體感溫度對單車租借量的影響.....	5
圖 10 濕度對單車租借量的影響.....	6
圖 11 風速對單車租借量的影響.....	6
圖 12 天氣類別對單車租借量的影響.....	7
圖 13 所有特徵類別相關性.....	7
圖 14 資料匯入.....	8
圖 15 決定輸入特徵 x 和目標變量 y.....	8
圖 16 分割訓練集及測試集.....	8
圖 17 模型訓練與 RMSE 評估.....	8
圖 18 可視化實際值和預測值.....	9
圖 19 資料匯入.....	9
圖 20 分割訓練集和測試集.....	10
圖 21 RobustScaler 進行特徵縮放	10
圖 22 建立訓練模型.....	10
圖 23 RMSE 和 R-sq 評估	10
圖 24 可視化實際值和預測值.....	11
圖 25 後臺架構圖.....	13
圖 26 天氣預報.....	13
圖 27 各站點即時使用量.....	14
圖 28 未來使用量預測圖.....	14

表目錄

表 1 資料欄位說明.....	2
表 2 天氣類別代號與說明.....	7
表 3 調整超參數水準.....	11
表 4 實驗組合直交表與 R-sq.....	12
表 5 YouBike2.0 臺北市公共自行車即時資訊欄位說明.....	12

一、簡介

1.1 背景動機與目的

「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。鼓勵民眾外出善用大眾運輸系統，以減少交通壅塞問題，同時也能保護環境、節省能源消耗，達到永續智慧交通環境。

1.2 5W1H 分析

What：民眾會遇到長時間無車可借或是無車位可還的情況。

Who：YouBike 員工、顧客。

When：日常交通尖峰時刻。

Where：學校周遭、運輸量大的地區、交通轉乘站點。

Why：YouBike 員工需要開著車輛調派車巡邏各站點。

How：運用 AI 方法配合網路資訊，設計車輛使用量即時預測系統。

二、研究方法

2.1 XGBoost

XGBoost 全名為 eXtreme Gradient Boosting，它是一種強大的梯度提升機（Gradient Boosting Machine）演算法，在分類、回歸、排序、推薦系統等領域中廣泛應用。通過結合多個弱學習器來構建一個更強大的預測模型。它通過迭代的方式，逐步改進預測結果，將新的學習器關注於之前預測錯誤的樣本，從而提高模型的性能。XGBoost 使用梯度提升的技術，以最小化損失函數來訓練模型，使模型能夠更好地適應訓練數據，也在訓練過程中引入了正規化方法，避免過擬合，提高模型的泛化能力。

2.2 LSTM

LSTM（Long Short-Term Memory）是一種特殊的循環神經網絡（Recurrent Neural Network, RNN），被廣泛應用於處理序列數據的領域，如自然語言處理和時間序列預測。LSTM 模型在處理長期相關性和記憶問題方面表現優異。相比於傳統的 RNN，LSTM 通過引入門控機制，可以更好地捕捉和保持序列中的長期相關性。LSTM 模型的結構包含了一個或多個 LSTM 層，每層都有一個細胞狀態（cell state）和一個隱藏狀態（hidden state）。細胞狀態負責記住長期的信息，而隱藏狀態則根據當前輸入和前一個隱藏狀態進行計算，提供模型的預測結果。LSTM 模型的訓練過程通常使用反向傳播算法和梯度下降法。在每個時間，模型根據當前輸入和目標輸出計算預測值，然後通過比較預測值和真實值的誤差來更新模型的權重，以使預測結果更接近真實結果。

三、實驗設計與結果

3.1 資料集介紹

「London bike sharing」是一個與倫敦市自行車共享系統相關的資料集。這個資料集收集了倫敦市自行車共享系統的每小時租借數量、時間戳記以及各種天氣特徵，例如：溫度、風速、降雨情況等。這些資料可以用於進行自行車租借量的預測、分析與探索。資料集的時間範圍涵蓋了 2015-2017 年的資料，讓我們可以觀察到不同季節、天氣條件和時間段對自行車租借量的影響，資料欄位說明如下表 1。下圖 1—圖 12 顯示 2015-2017 年間，共享單車租借量在不同時間單位(日期、月、日、季)和不同天氣特徵(類別、溫度、風速)的影響與變化，共 17415 筆資料。

表 1 資料欄位說明

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

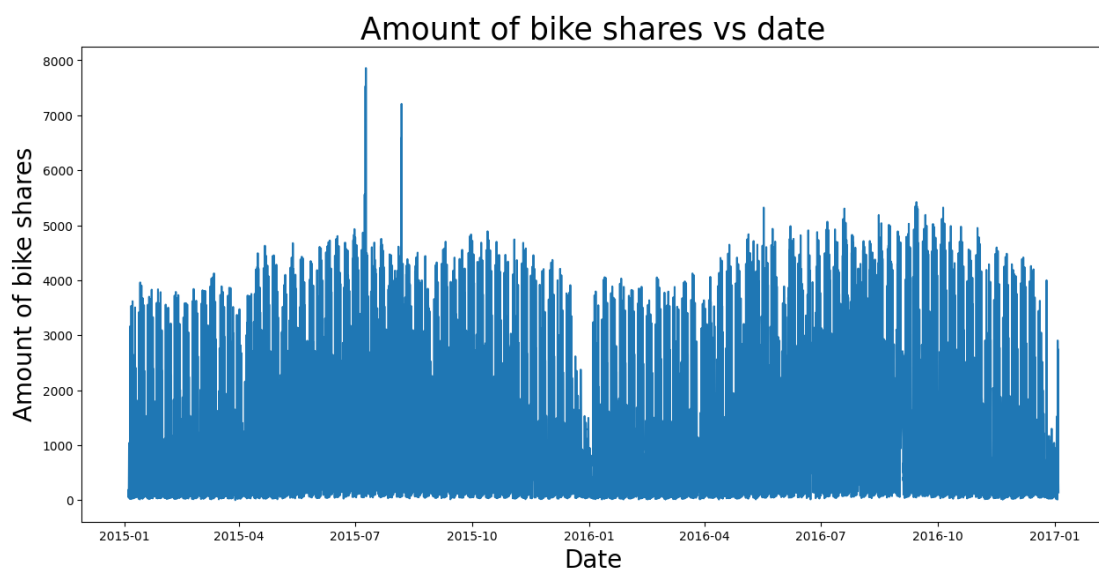


圖 1 以日期的單車租借量

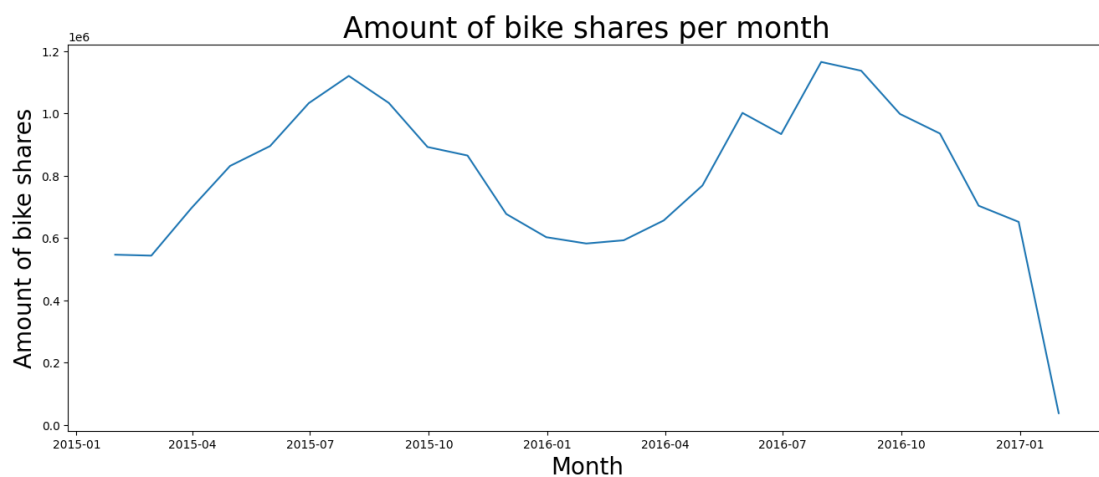


圖 2 以月為單位的單車租借量

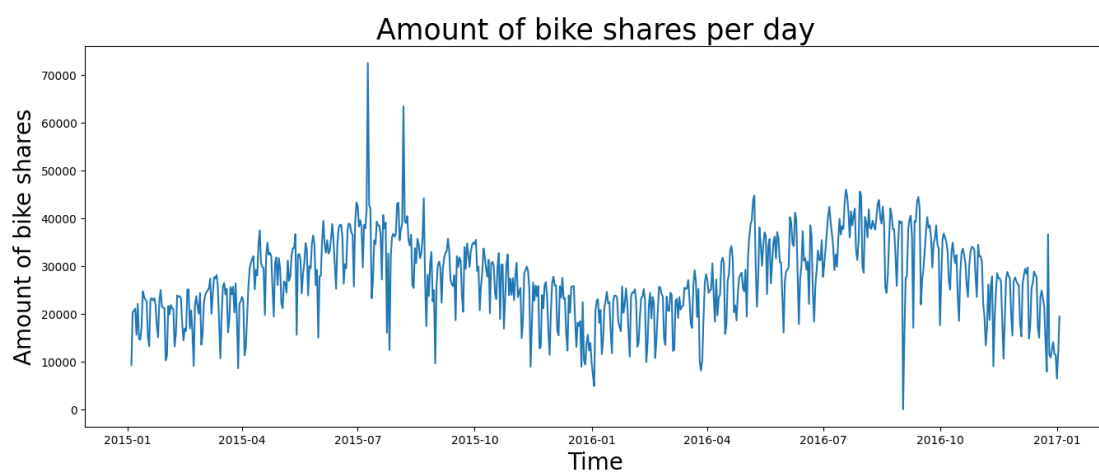


圖 3 以日為單位的單車租借量

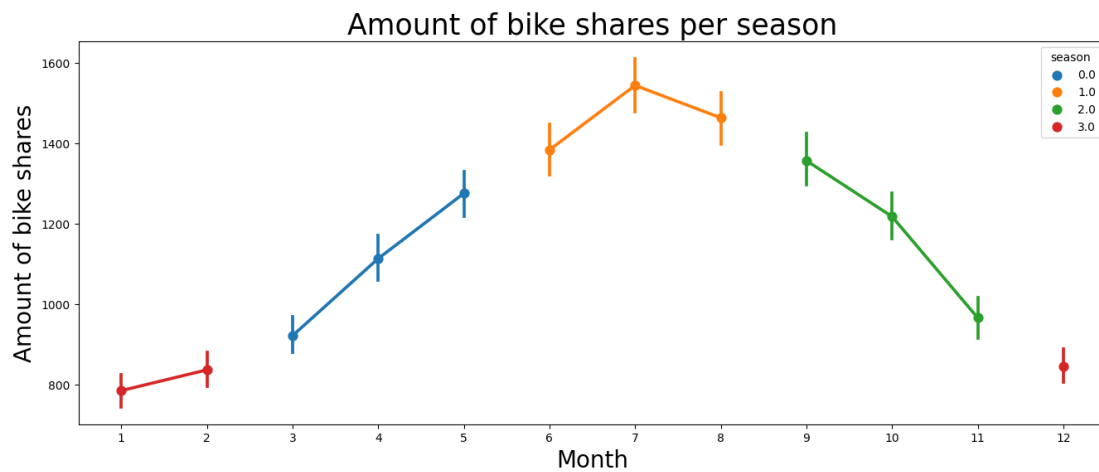


圖 4 以季為單位的單車租借量

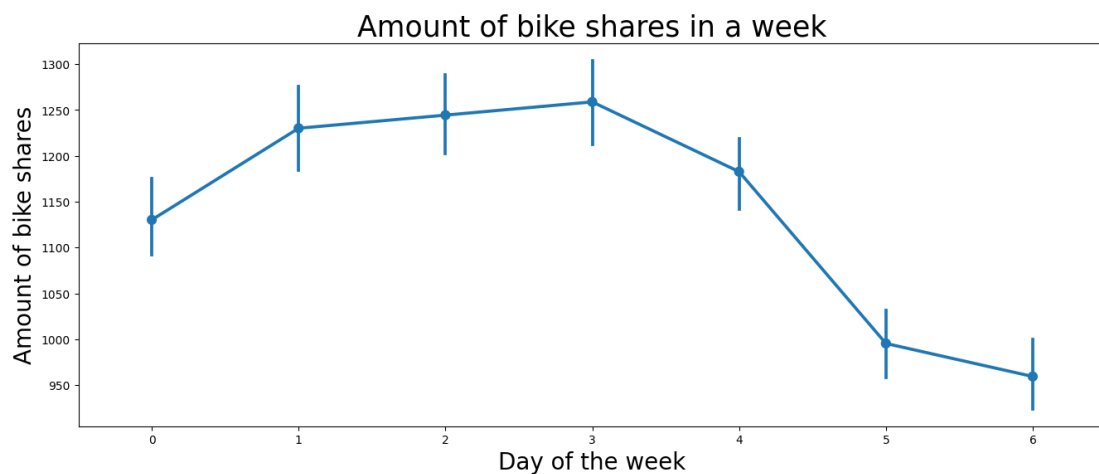


圖 5 一週的單車租借量

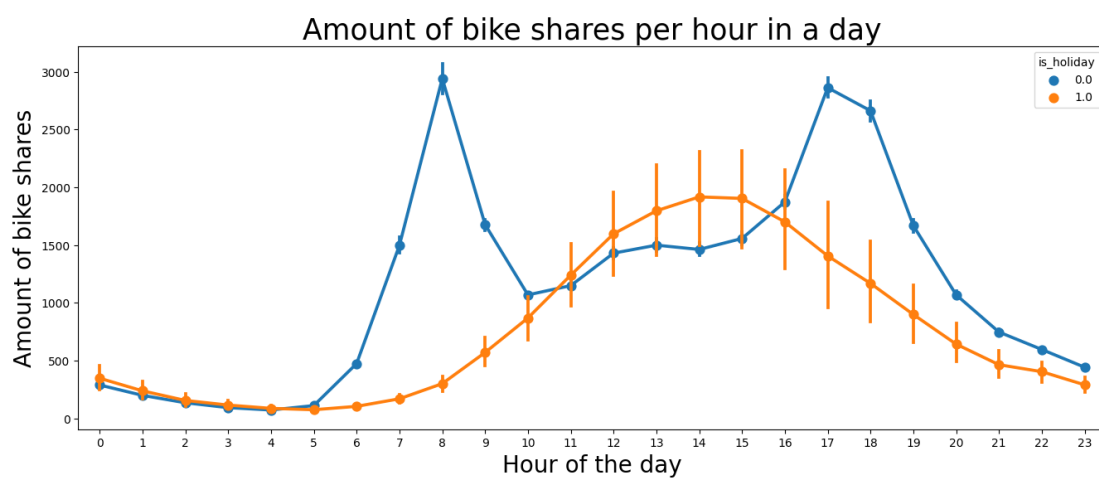


圖 6 一天內各小時的單車租借量

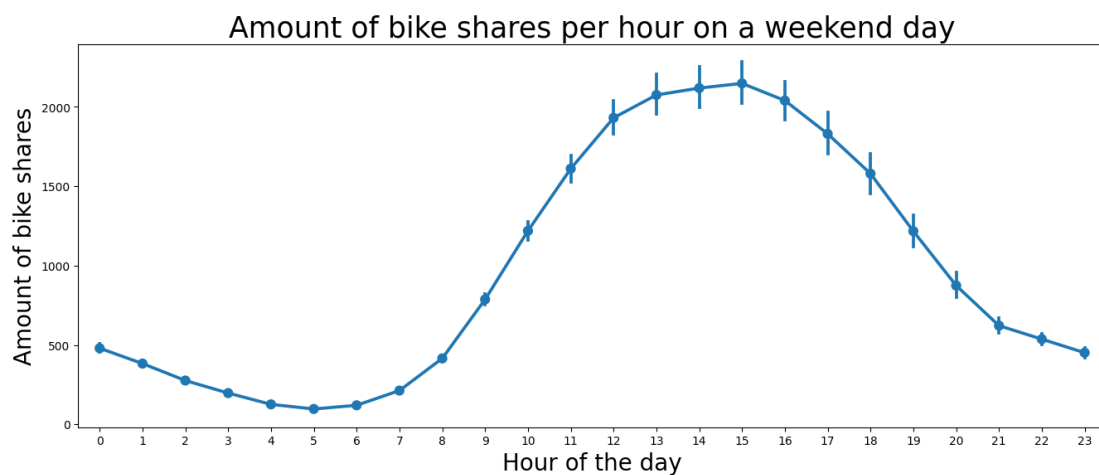


圖 7 假日各小時的單車租借量

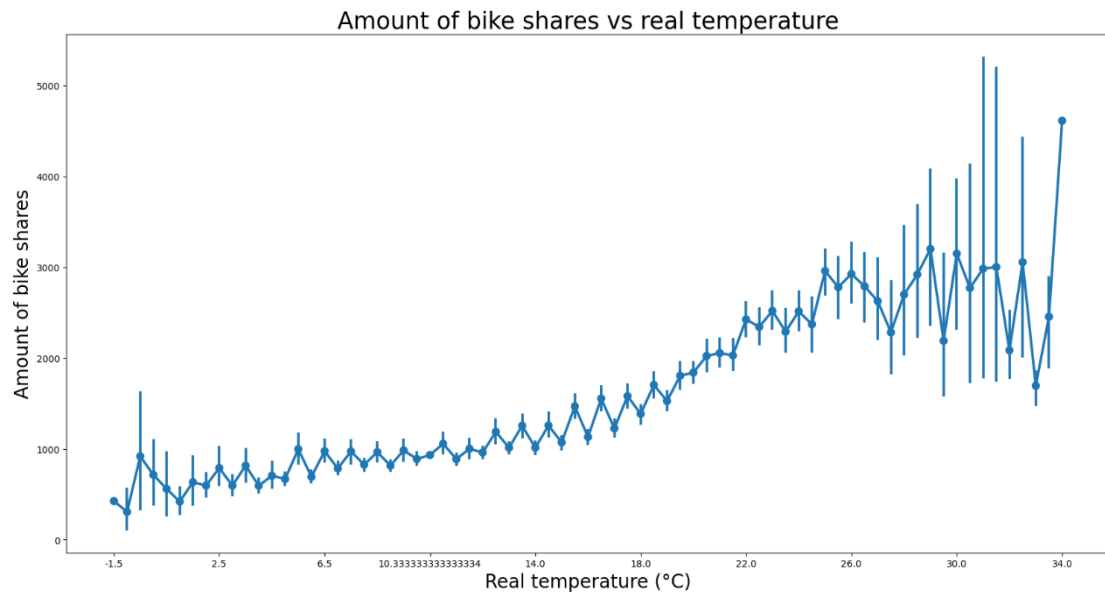


圖 8 實際溫度對單車租借量的影響

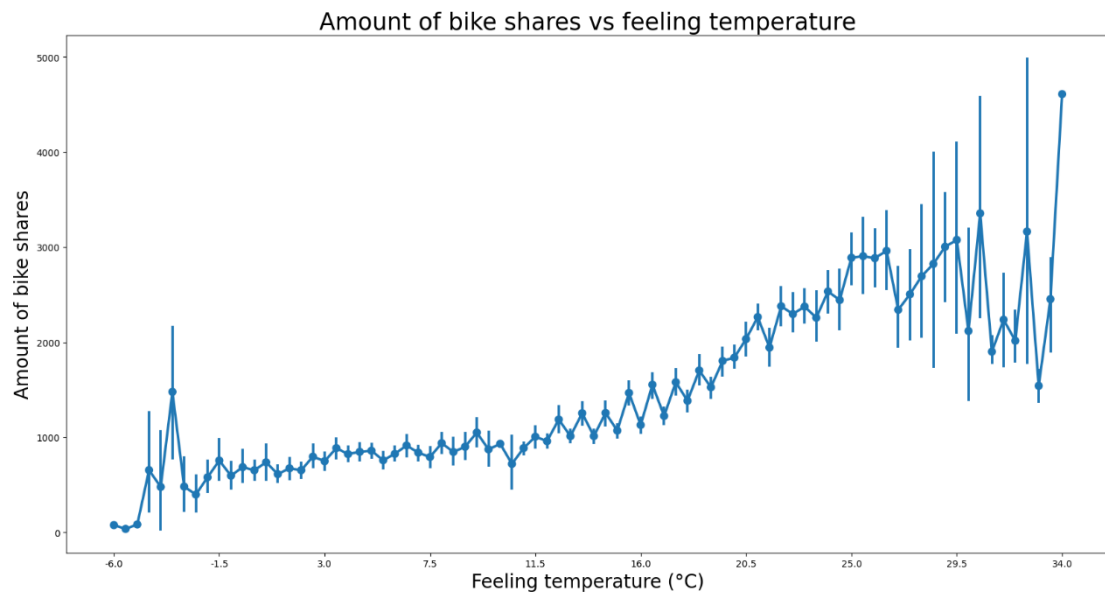


圖 9 體感溫度對單車租借量的影響

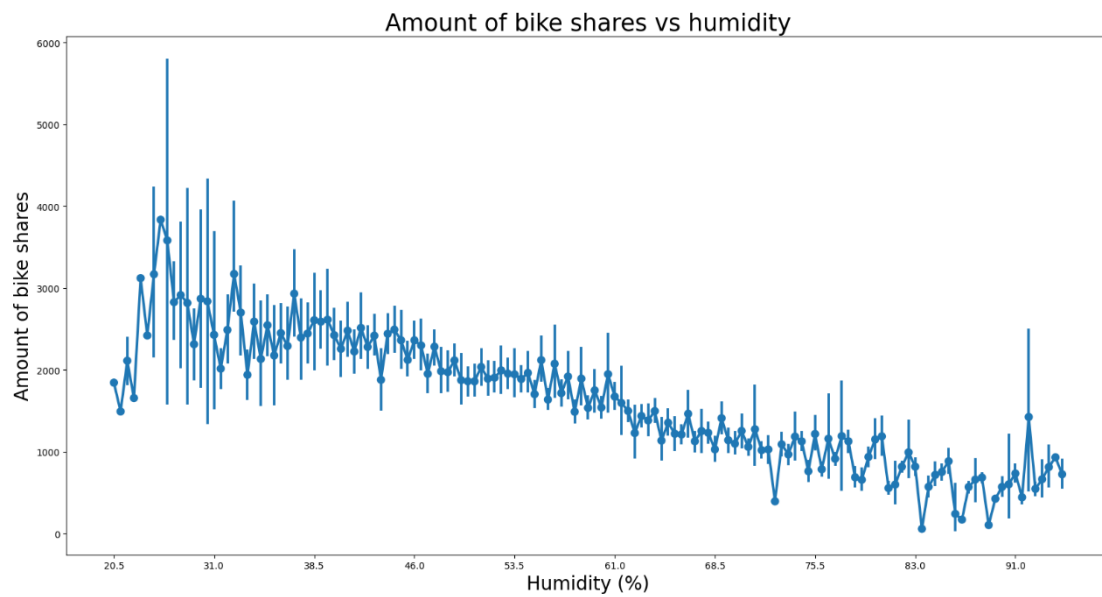


圖 10 濕度對單車租借量的影響

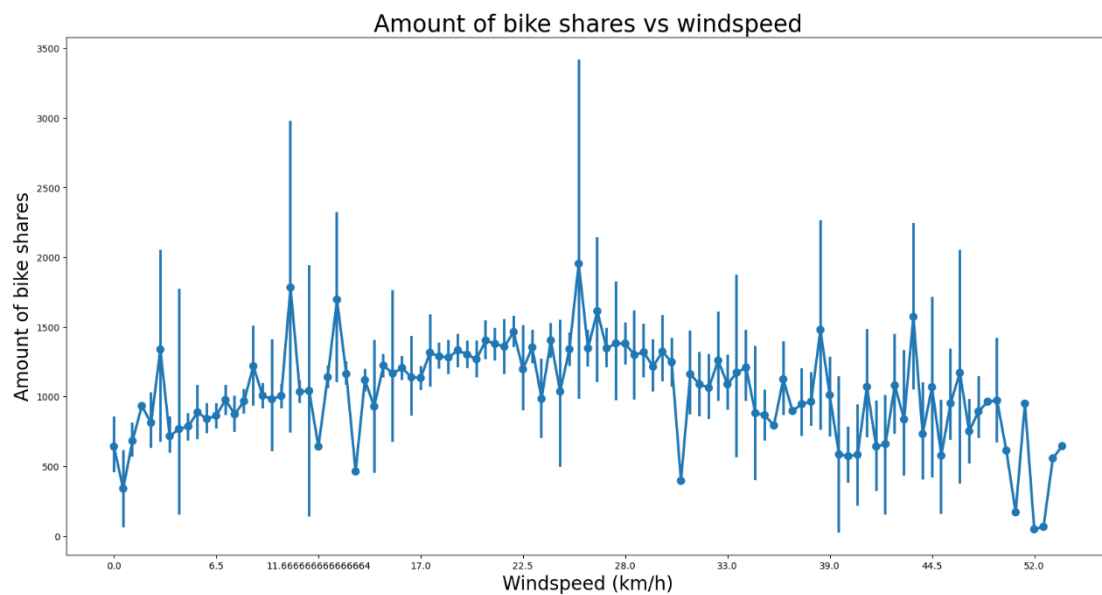


圖 11 風速對單車租借量的影響

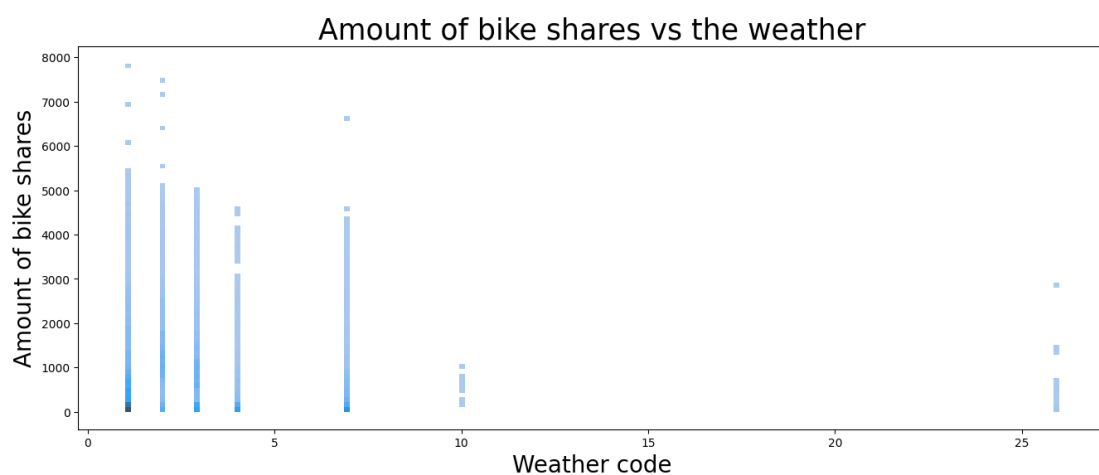


圖 12 天氣類別對單車租借量的影響

表 2 天氣類別代號與說明

天氣代號	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

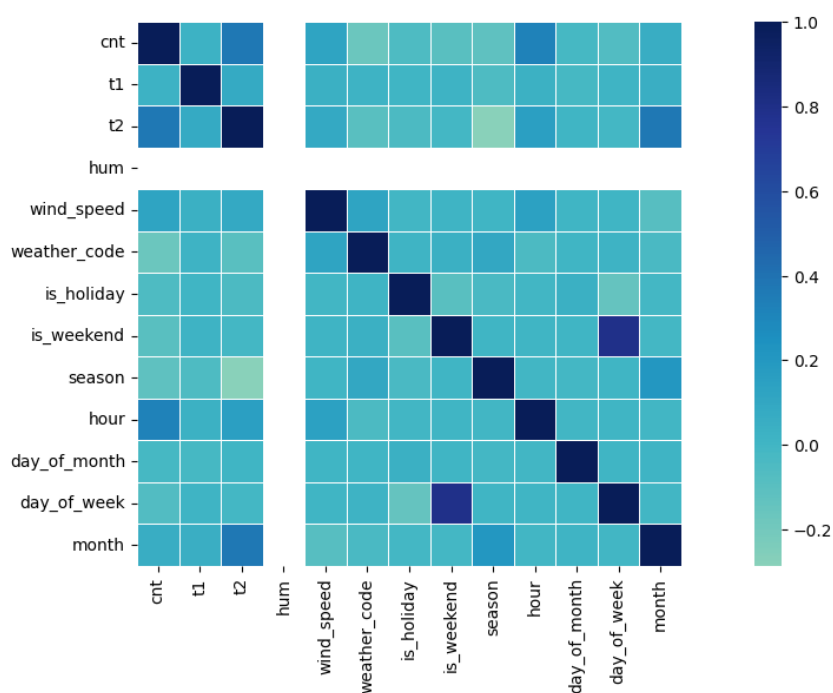


圖 13 所有特徵類別相關性

3.2 模型建立

3.2.1 XGBoost 模型

data										
	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
...	...	...	...	...	...	...	...	...	...	...
17409	2017-01-03 19:00:00	1042	5.0	1.0	81.0	19.0	3.0	0.0	0.0	3.0
17410	2017-01-03 20:00:00	541	5.0	1.0	81.0	21.0	4.0	0.0	0.0	3.0
17411	2017-01-03 21:00:00	337	5.5	1.5	78.5	24.0	4.0	0.0	0.0	3.0
17412	2017-01-03 22:00:00	224	5.5	1.5	76.0	23.0	4.0	0.0	0.0	3.0
17413	2017-01-03 23:00:00	139	5.0	1.0	76.0	22.0	2.0	0.0	0.0	3.0
17414 rows × 10 columns										

圖 14 資料匯入

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

圖 15 決定輸入特徵 x 和目標變量 y

```
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)
```

圖 16 分割訓練集及測試集

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

```
xgbmodel = XGBRegressor()
xgbmodel.fit(X_train, y_train)

preds = xgbmodel.predict(X_test)
```

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

```
RMSE: 1015.137644
```

圖 17 模型訓練與 RMSE 評估

```
import matplotlib.pyplot as plt
plt.figure(figsize=(16,6))
plt.plot(y_test.values,marker=".",label="actual")
plt.plot(preds,marker=".",label="prediction",color="r")

plt.title('Bike share prediction per hour', fontsize=22)
plt.xlabel(xlabel='Date', fontsize=14)
plt.ylabel(ylabel='Bikes shares', fontsize=14)
plt.legend(['Ground truth', 'Predicted'])

plt.show()
```

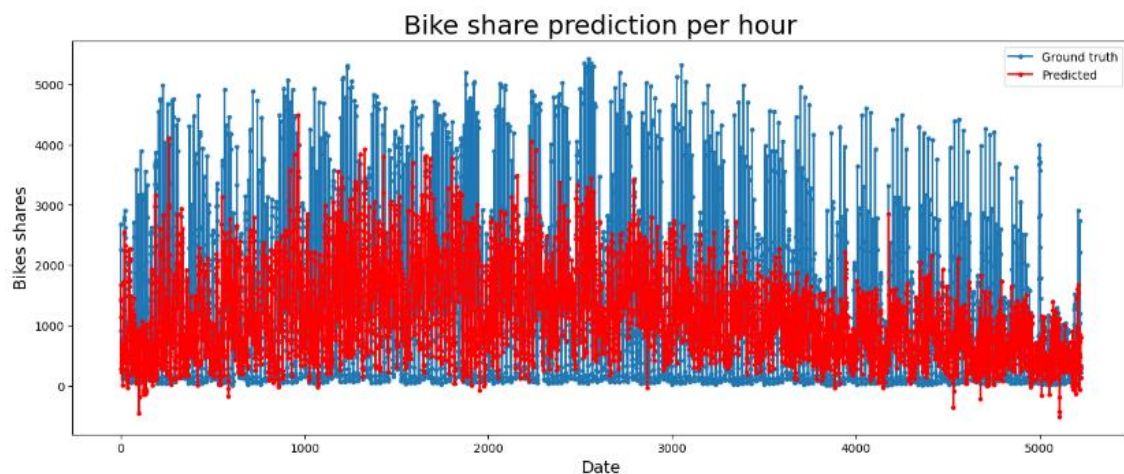


圖 18 可視化實際值和預測值

3.2.2 LSTM 模型

df													
timestamp	cnt	t1	t2	hum	wind_speed	weather_code	is_holiday	is_weekend	season	hour	day_of_month	day_of_week	month
2015-01-04 00:00:00	182	3.0	2.0	93.0	6.0	3.0	0.0	1.0	3.0	0	4	6	1
2015-01-04 01:00:00	138	3.0	2.5	93.0	5.0	1.0	0.0	1.0	3.0	1	4	6	1
2015-01-04 02:00:00	134	2.5	2.5	96.5	0.0	1.0	0.0	1.0	3.0	2	4	6	1
2015-01-04 03:00:00	72	2.0	2.0	100.0	0.0	1.0	0.0	1.0	3.0	3	4	6	1
2015-01-04 04:00:00	47	2.0	0.0	93.0	6.5	1.0	0.0	1.0	3.0	4	4	6	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2017-01-03 19:00:00	1042	5.0	1.0	81.0	19.0	3.0	0.0	0.0	3.0	19	3	1	1
2017-01-03 20:00:00	541	5.0	1.0	81.0	21.0	4.0	0.0	0.0	3.0	20	3	1	1
2017-01-03 21:00:00	337	5.5	1.5	78.5	24.0	4.0	0.0	0.0	3.0	21	3	1	1
2017-01-03 22:00:00	224	5.5	1.5	76.0	23.0	4.0	0.0	0.0	3.0	22	3	1	1
2017-01-03 23:00:00	139	5.0	1.0	76.0	22.0	2.0	0.0	0.0	3.0	23	3	1	1

17414 rows x 13 columns

圖 19 資料匯入

```
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)
```

圖 20 分割訓練集和測試集

```
# 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']])
```

圖 21 RobustScaler 進行特徵縮放

```
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)

# 保存模型
model.save("/content/drive/MyDrive/Youbike dataset/lstm_model.h5")
```

圖 22 建立訓練模型

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

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

圖 23 RMSE 和 R-sq 評估

```
plt.figure(figsize=(16, 8))
plt.plot(y_test[1200:1500], label='true')
plt.plot(y_pred[1200:1500], label='predicted')
plt.legend()

plt.plot(history.history["loss"], label="loss")
plt.plot(history.history["val_loss"], label="val_loss")
plt.legend(loc="best")
plt.xlabel("No. Of Epochs")
plt.ylabel("mse score")
```

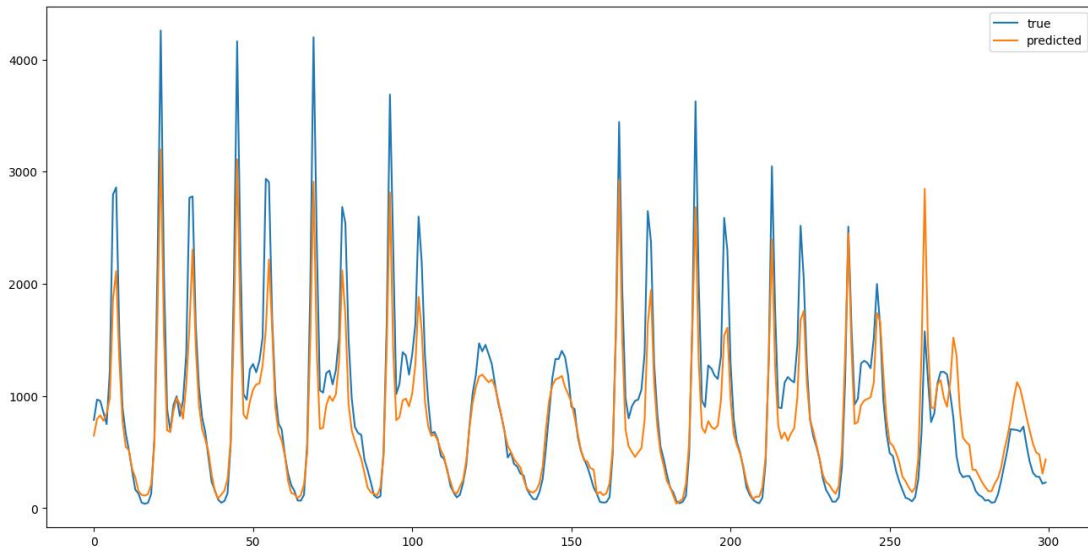


圖 24 可視化實際值和預測值

經過比較兩種模型的訓練成果後，觀察到 LSTM 模型表現優於 XGBoost 模型，因此在本次改善專案中，我們選擇採用 LSTM 模型來建立共享單車租借量預測系統。

3.3 LSTM 模型訓練結果

表 3 為 LSTM 模型超參數調整之水準，表 4 為實驗組合直交表與 R-sq，可知組合 6 的參數結果 R-sq = 0.845，訓練成果最好。

表 3 調整超參數水準

	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

表 4 實驗組合直交表與 R-sq

組合	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

四、網頁設計與後臺管理系統

4.1 數據來源

預測系統之資料集為「YouBike2.0 臺北市公共自行車即時資訊」，該資訊集的更新頻率為每分鐘一次，資料內容之欄位與說明如下表 5 所示。

表 5 YouBike2.0 臺北市公共自行車即時資訊欄位說明

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

4.2 後臺管理系統

後臺管理系統結合使用量預測功能，新增了天氣預報圖、各站點即時使用量列表、未來使用量預測圖。

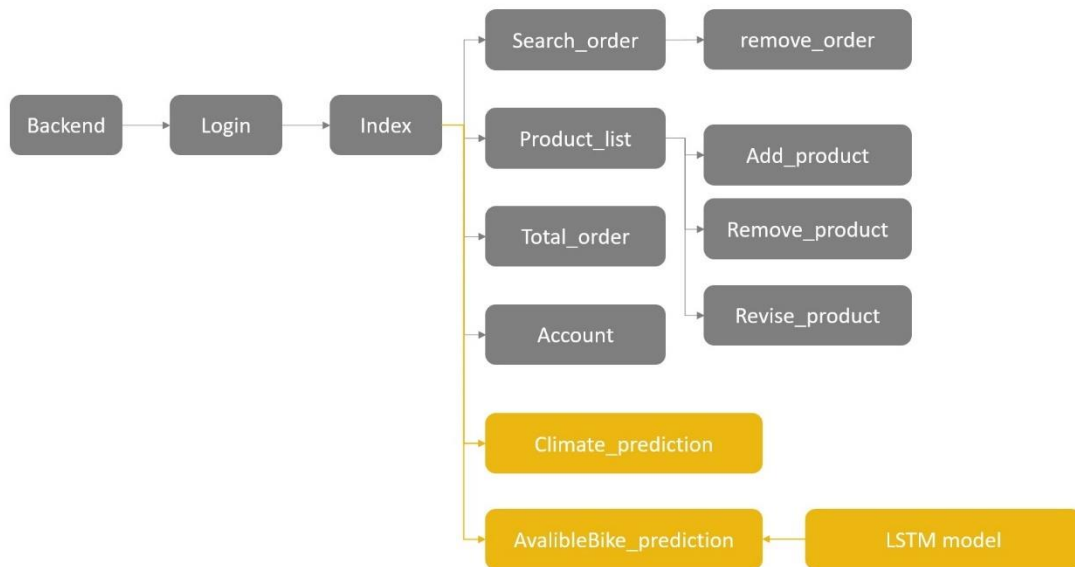


圖 25 後臺架構圖

4.3 網頁設計

後臺系統首頁所呈現的網頁如下圖 26 — 圖 28，隨著人員更新網站，預測系統將匯入即時資訊，以調整使用量預測結果。

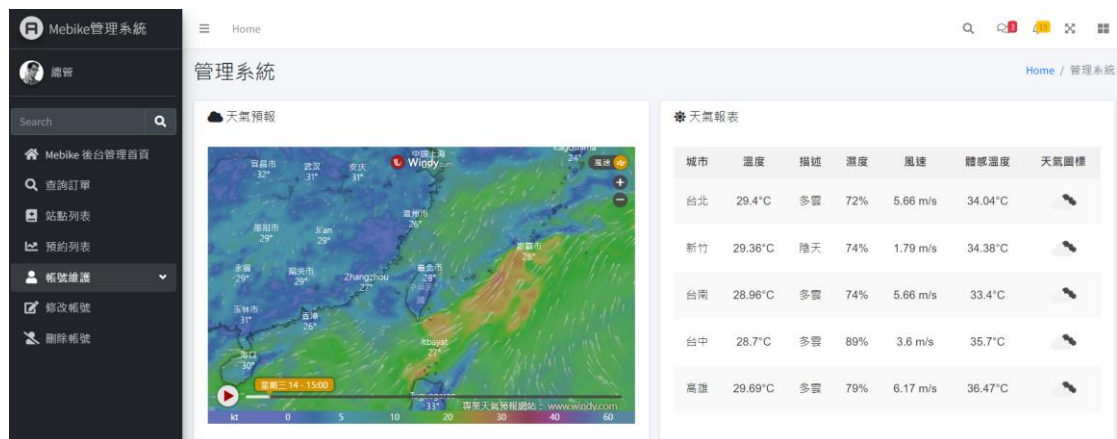


圖 26 天氣預報

Stations				Filter
站點	空位數量	站點使用率	More	
捷運科技大樓站	28台	100%	Q	
復興南路二段273號前	20台	95%	Q	
國北教大實小東側門	14台	87%	Q	
和平公園東側	4台	36%	Q	
辛亥復興路口西北側	15台	93%	Q	
復興南路二段280號前	5台	45%	Q	
復興南路二段340巷口	8台	72%	Q	
新生南路三段52號前	9台	52%	Q	

圖 27 各站點即時使用量

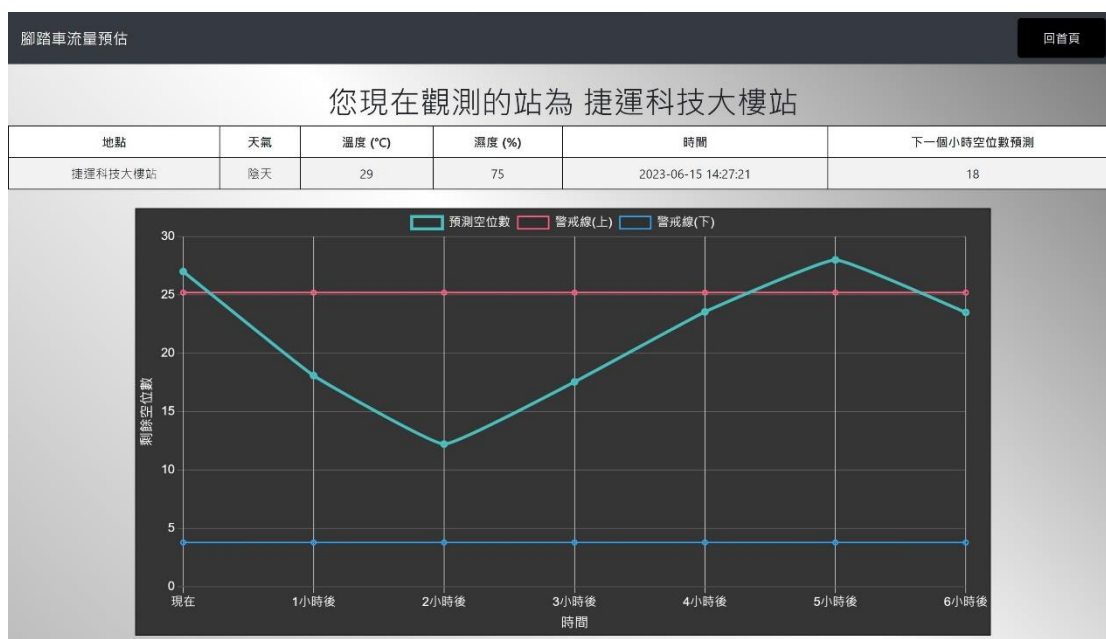


圖 28 未來使用量預測圖

五、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。

六、參考資料

[1] XGBoost 演算法預測報酬(上)

<https://www.tejwin.com/insight/%E3%80%90%E8%B3%87%E6%96%99%E7%A7%91%E5%AD%B8%E3%80%91xgboost-%E6%BC%94%E7%AE%97%E6%B3%95%E9%A0%90%E6%B8%AC%E5%A0%B1%E9%85%AC%E4%B8%8A/#XGBoost%E7%B0%A1%E4%BB%8B>

[2] [Day 15] 機器學習常勝軍 - XGBoost

<https://ithelp.ithome.com.tw/articles/10273094>

[3] [Day-16] RNN - LSTM 介紹

<https://ithelp.ithome.com.tw/articles/10223055>

[4] YouBike2.0 臺北市公共自行車即時資訊

<https://data.gov.tw/en/datasets/137993>

[5] London bike sharing dataset

<https://www.kaggle.com/datasets/hmavrodiev/london-bike-sharing-dataset>

[6] ChatGPT

[7] Open weather

<https://openweathermap.org/>