



以 **LSTM** 建立
番石榴價格預測模型

109034508 楊怡芳

2020 IIE_final



1 研究背景&5W1H

2 資料蒐集&前處理

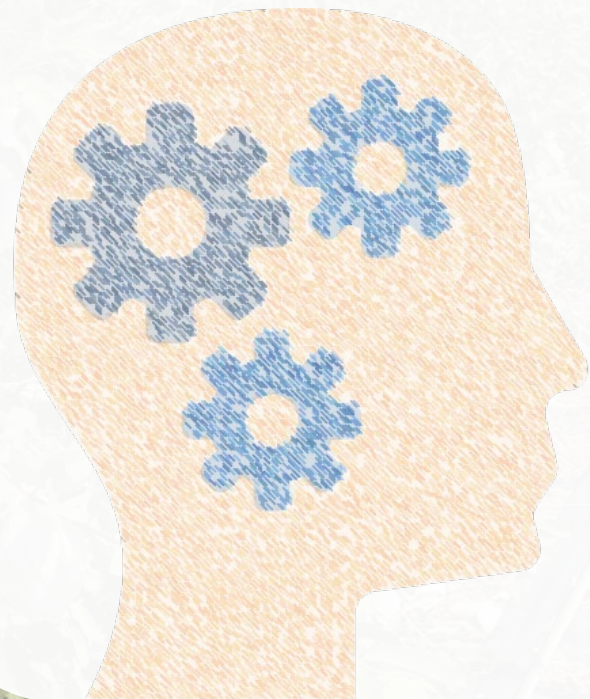
3 模型介紹&超參數調整

4 結論&未來展望

研究背景

&

5W1H



1

水果產銷失衡 賴清德：已要求農委會檢討原因

朱冠諭

+ 追蹤

2018-07-05 21:20 927 人氣

〈中部〉芭樂天價 拍賣價 每公斤近60元

2012-09-04

北中南產期一致 芭樂價崩盤

04:10 2017/04/25 | 中國時報 | 高錦如、嘉縣

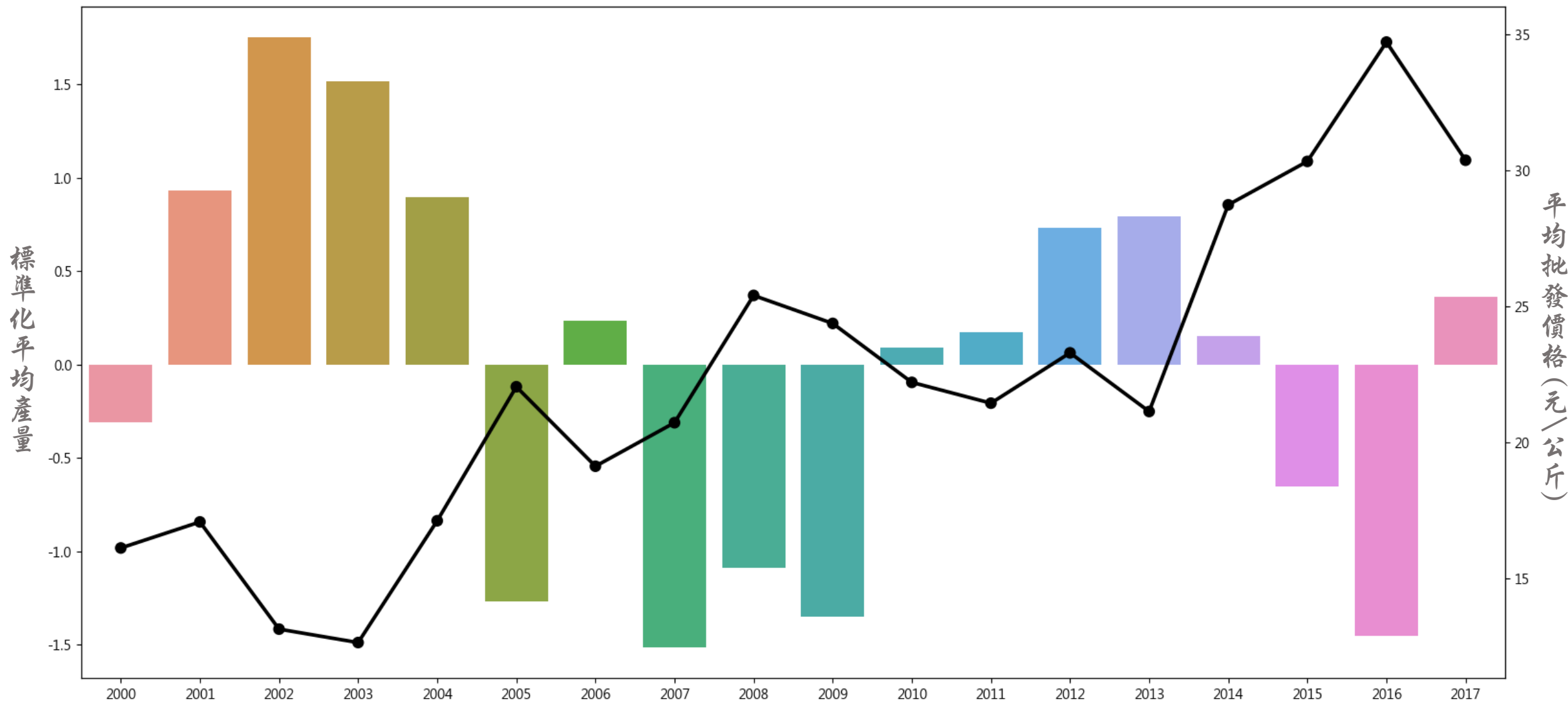
芭樂價慘跌

出版時間：2003/08/07

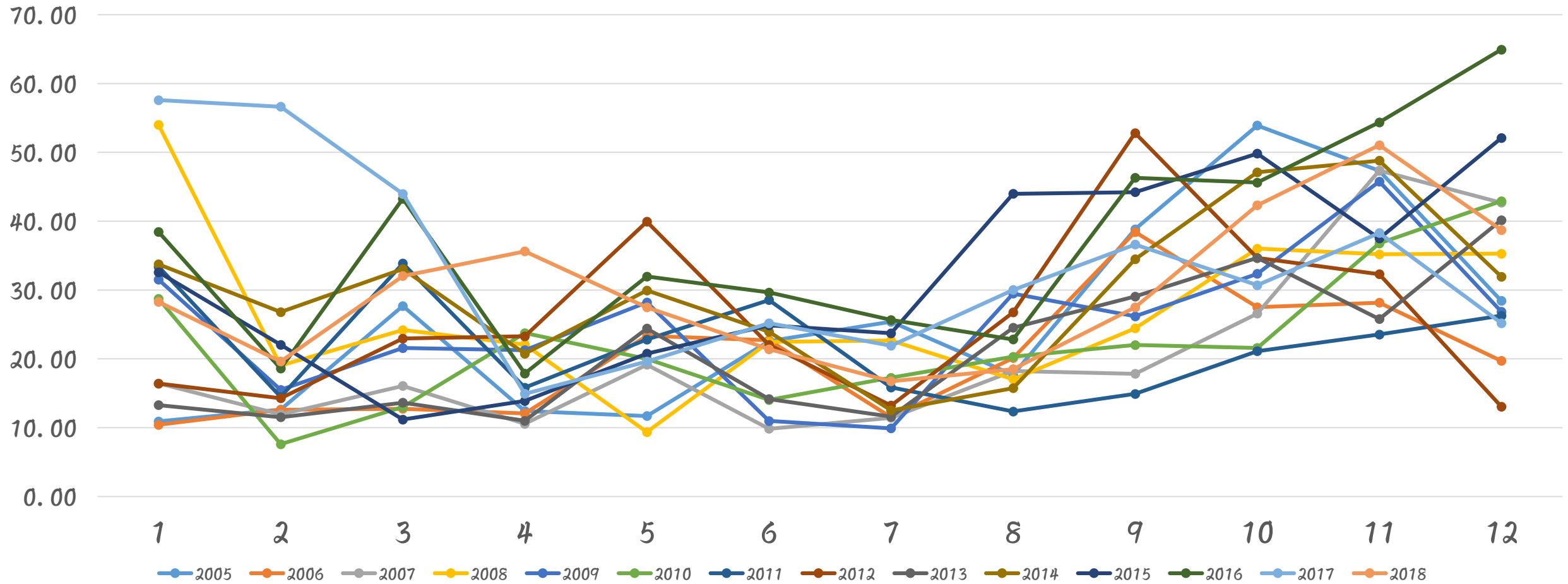
2016/09/17, 社會

全臺農損失破8億：蜜棗農民明年春節難過，金門高粱4成沒了

番石榴歷年平均批發價及產量變動圖



番石榴台北二市場歷年平均批發價變動圖



➤ 具季節性

➤ 價格差距大

5W1H

What

番石榴
批發價預測

Who

農民
消費者

Why

降低價格波動
改善產銷失衡

When

欲購買、販賣
番石榴時

Where

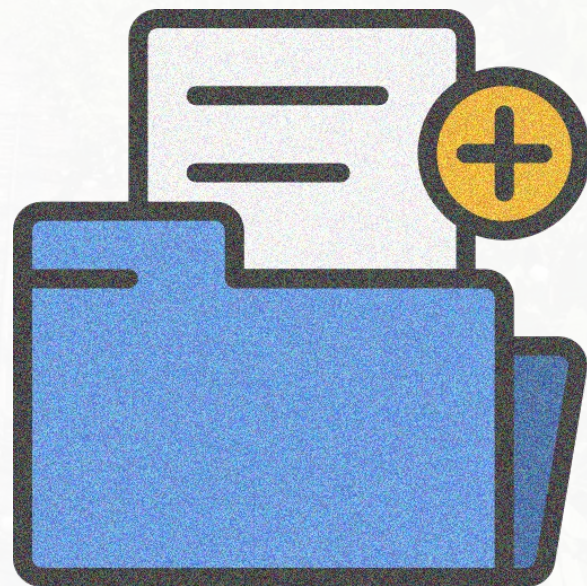
台北二
批發市場

How

資料預處理
LSTM預測

資料蒐集 & 前處理

2

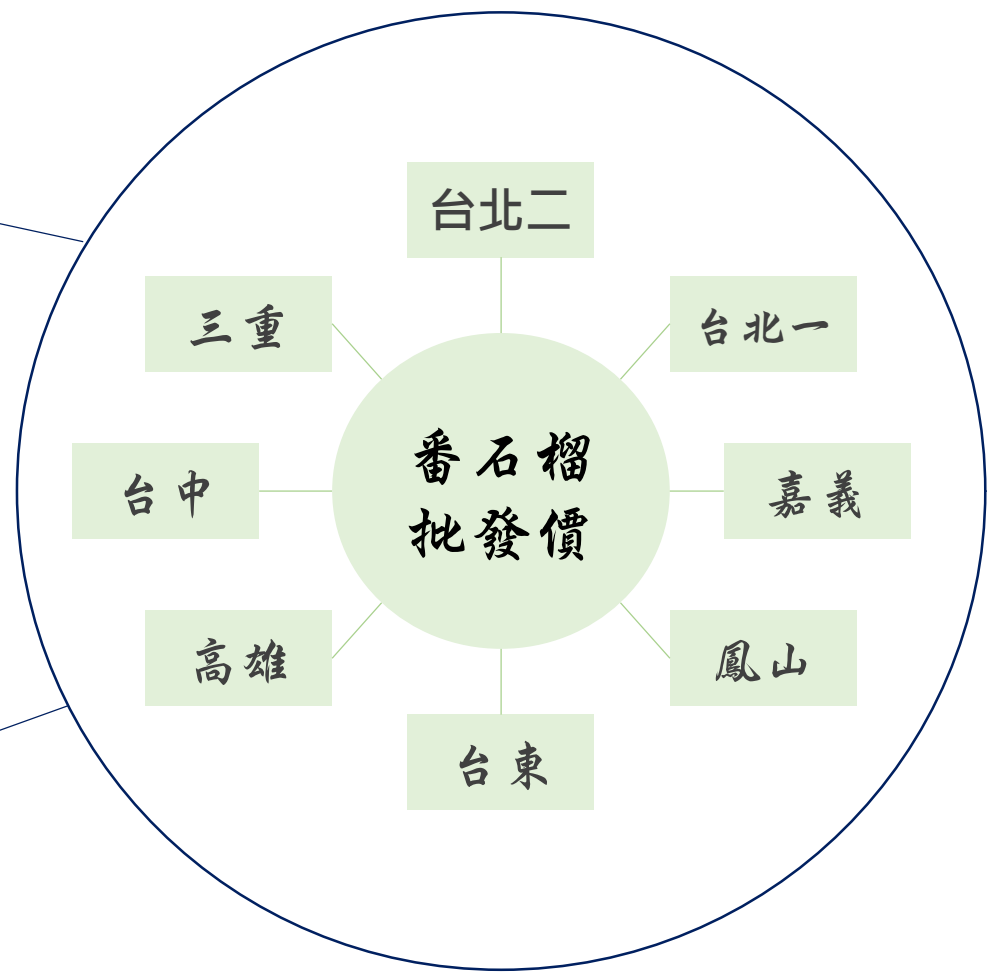


水果交互影響

香蕉、鳳梨、木瓜
芒果、葡萄

批發市場因子

平均價、上價
中價、下價
交易量、休市日



氣象因子

平均氣溫
平均最高氣溫
平均最低氣溫
平均相對溼度
平均降水量
平均風速
颱風警報
大雨特報
低溫特報

From 政府公開資料平台

批發市場因子

依據各果菜批發市場公布之每年休市日及因颱風假造成休市之情形，補上休市日之欄位

資料補值

氣象因子

多個測站在2015年以前有嚴重缺值，為保留資料完整性，即使用鄰近測站資料進行補值。

相關性分析

Pearson

相關性	特徵數
$r \geq 0.4$	150
$r \geq 0.7$	29

以「中度線性相關」為標準，挑選所需變數

相關性分析

表 22 對平均價(P)相關特徵分析表

	104_P_芭
104_P_芭	1
109_P_芭	0.98877
241_P_芭	0.97754
400_P_芭	0.96596
600_P_芭	0.51273
800_P_芭	0.83518
830_P_芭	0.56518
930_P_芭	0.61566
104_P_蕉	0.69974
109_P_蕉	0.68979
241_P_蕉	0.71003
400_P_蕉	0.69446

600_P_蕉	0.30219
800_P_蕉	0.65316
830_P_蕉	0.62655
930_P_蕉	0.57115
104_P_木	0.68453
109_P_木	0.67943
241_P_木	0.67101
400_P_木	0.67737
600_P_木	0.25651
800_P_木	0.62518
830_P_木	0.56941
930_P_木	0.49102
104_P_鳳	0.62818

109_P_鳳	0.64787
241_P_鳳	0.59771
400_P_鳳	0.70334
600_P_鳳	0.12286
800_P_鳳	0.61454
830_P_鳳	0.47313
930_P_鳳	0.45528
104_P_芒	0.18474
109_P_芒	0.20016
241_P_芒	0.09118
400_P_芒	0.01026
600_P_芒	-0.15943
800_P_芒	-0.0316

830_P_芒	-0.06338
930_P_芒	-0.02964
104_P_葡	0.59113
109_P_葡	0.59977
241_P_葡	0.64877
400_P_葡	0.75086
600_P_葡	0.16101
800_P_葡	0.59519
830_P_葡	0.34619
930_P_葡	0.39574

相關性分析

表 27 對天氣相關特徵分析表

	104_P_芭	rain_蕉	-0.07166	MaxT_芒	-0.08141	temp_鳳	-0.10083
low temp_芭	-0.00089	RH_蕉	-0.00168	MinT_芒	-0.09819	typhoon_鳳	-0.01001
MaxT_芭	-0.06591	temp_蕉	-0.09653	RA_芒	-0.05526	low temp_葡	-0.00292
MinT_芭	-0.09539	typhoon_蕉	-0.01001	rain_芒	-0.0657	MaxT_葡	-0.07861
RA_芭	-0.0608	low temp_木	0.00189	RH_芒	0.05207	MinT_葡	-0.09484
rain_芭	-0.06482	MaxT_木	-0.08758	temp_芒	-0.10167	RA_葡	-0.06816
RH_芭	0.06037	MinT_木	-0.10104	typhoon_芒	-0.01001	rain_葡	-0.07055
temp_芭	-0.09732	RA_木	-0.06278	low temp_鳳	-0.0066	RH_葡	0.01044
typhoon_芭	-0.01001	rain_木	-0.07262	MaxT_鳳	-0.08872	temp_葡	-0.09515
low temp_蕉	0.00189	RH_木	0.0494	MinT_鳳	-0.09648	typhoon_葡	-0.01001
MaxT_蕉	-0.08923	temp_木	-0.10346	RA_鳳	-0.06745		
MinT_蕉	-0.09153	typhoon_木	-0.01001	rain_鳳	-0.08639		
RA_蕉	-0.07156	low temp_芒	-0.00364	RH_鳳	0.01717		

天氣相關特徵與台北二番石榴價格相關性極低，故不納入模型訓練因子

特徵標準化

```

38 x1 = guava[['109_P_芭', '109_PD_芭',
39            '109_PM_芭', '109_PU_芭', '241_P_芭', '241_PD_芭', '241_PM_芭', '241_PU_芭',
40            '400_P_芭', '400_PD_芭', '400_PM_芭', '400_PU_芭', '600_P_芭', '600_PM_芭',
41            '600_PU_芭', '800_P_芭', '800_PD_芭', '800_PM_芭', '800_PU_芭', '830_P_芭',
42            '830_PD_芭', '830_PM_芭', '830_PU_芭', '930_P_芭', '930_PD_芭', '930_PM_芭',
43            '930_PU_芭']]
44 x2 = banana[['104_P_蕉', '104_PD_蕉', '104_PM_蕉', '104_PU_蕉', '109_P_蕉', '109_PD_蕉',
45             '109_PM_蕉', '109_PU_蕉', '241_P_蕉', '241_PD_蕉', '241_PM_蕉', '241_PU_蕉',
46             '400_P_蕉', '400_PD_蕉', '400_PM_蕉', '400_PU_蕉', '800_P_蕉', '800_PD_蕉',
47             '800_PM_蕉', '800_PU_蕉', '830_P_蕉', '830_PD_蕉', '830_PM_蕉', '930_P_蕉',
48             '930_PD_蕉', '930_PM_蕉', '930_PU_蕉']]
49 x3 = papaya[['104_P_木', '104_PD_木', '104_PM_木', '104_PU_木', '109_P_木', '109_PD_木',
50             '109_PM_木', '109_PU_木', '241_P_木', '241_PD_木', '241_PM_木', '241_PU_木',
51             '400_P_木', '400_PD_木', '400_PM_木', '400_PU_木', '800_P_木', '800_PD_木',
52             '800_PM_木', '800_PU_木', '830_P_木', '830_PD_木', '830_PM_木', '830_PU_木',
53             '930_PU_木']]
54 x4 = pineapple[['104_P_鳳梨', '104_PD_鳳梨', '104_PM_鳳梨', '104_PU_鳳梨', '109_P_鳳梨', '109_PD_鳳梨',
55                '109_PM_鳳梨', '109_PU_鳳梨', '241_P_鳳梨', '241_PD_鳳梨', '241_PM_鳳梨', '241_PU_鳳梨',
56                '400_P_鳳梨', '400_PD_鳳梨', '400_PM_鳳梨', '400_PU_鳳梨', '800_P_鳳梨', '800_PD_鳳梨',
57                '800_PM_鳳梨', '800_PU_鳳梨']]
58 x5 = grape[['104_P_葡', '104_PD_葡', '104_PM_葡', '104_PU_葡', '109_P_葡', '109_PD_葡',
59            '109_PM_葡', '109_PU_葡', '241_P_葡', '241_PD_葡', '241_PM_葡', '241_PU_葡',
60            '400_P_葡', '400_PD_葡', '400_PM_葡', '400_PU_葡', '800_P_葡', '800_PM_葡',
61            '800_PU_葡']]
62 x6 = guava[['104_rest', '109_rest', '241_rest', '400_rest', '800_rest', '830_rest', '930_rest']]
63 x = pd.concat([x1, x2, x3, x4, x5, x6], 1)
64 y1 = guava[['104_P_芭']]

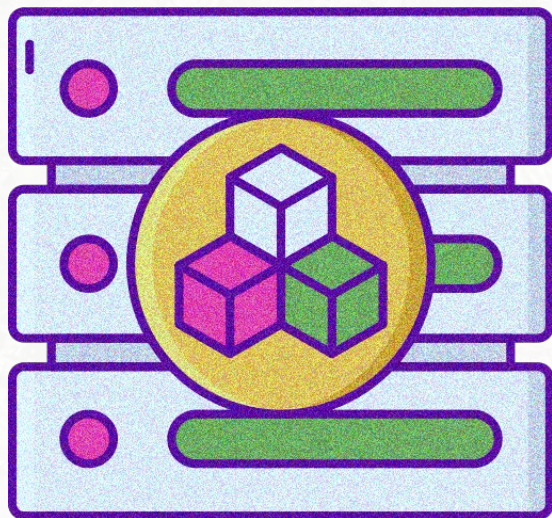
```

```

65
66 # 做MinMax Scaler
67 scaleX = MinMaxScaler()
68 scaleY = MinMaxScaler()
69
70 scale_x = scaleX.fit_transform(x)
71 scale_y1 = scaleY.fit_transform(y1)
72

```

因價格不會有負值，
所以採用**MinMax Scaler**對變數做標準化



模型介紹 & 超參數調整

3

2005-2018年共**5388**筆日資料
&
多數市場特徵數超過**100**種
&
時間序列特性

輸入訓練資料



LSTM類神經網路



Dense串接所有神經元



Dense輸出預測價格



MSE衡量模型配適度

建置方法：Python
Keras & Tensorflow

超參數調整

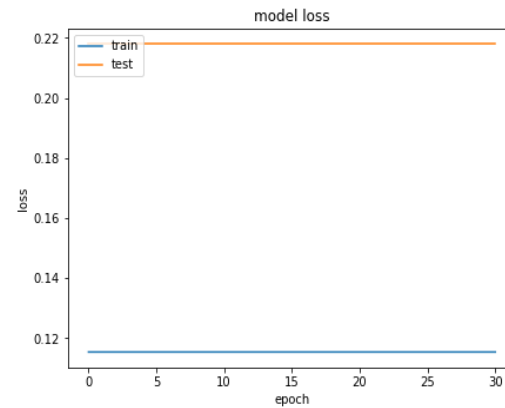
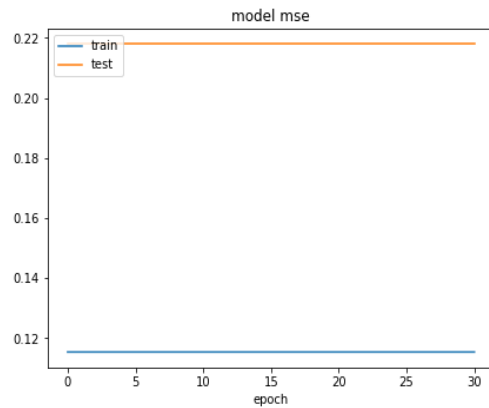
1. epochs迭代次數(100 vs. 500)

```
# 增加callbacks
from keras.callbacks import ReduceLRonPlateau
from keras.callbacks import EarlyStopping
reduce_lr = ReduceLRonPlateau(monitor='val_loss', factor=0.5, patience=5,
                              mode='auto', min_delta=0.0001, cooldown=0, min_lr=0)
es=EarlyStopping(monitor='val_loss', patience=30,mode='auto')
```

因本模型有設定earlystopping的機制，當validation loss超過30個epochs都沒有下降的話，即停止訓練以節省時間，故最後以100作為最終設定。

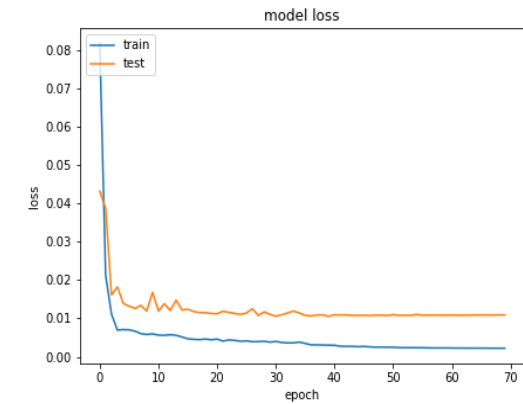
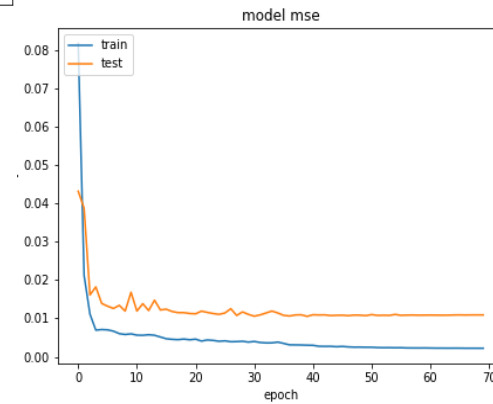
超參數調整

2. Adam learning rate (0.01 vs. 0.005)



lr = 0.01

lr = 0.005



超參數調整

3. Dense layer層數 (1 vs. 2)

```
48/48 [=====] - 0s 2ms/step - loss: 0.0129 - mse: 0.0129  
Test loss: 0.0128784766420722  
Test mse: 0.0128784766420722
```

1層dense layer

```
48/48 [=====] - 0s 3ms/step - loss: 0.0145 - mse: 0.0145  
Test loss: 0.014507105574011803  
Test mse: 0.014507105574011803
```

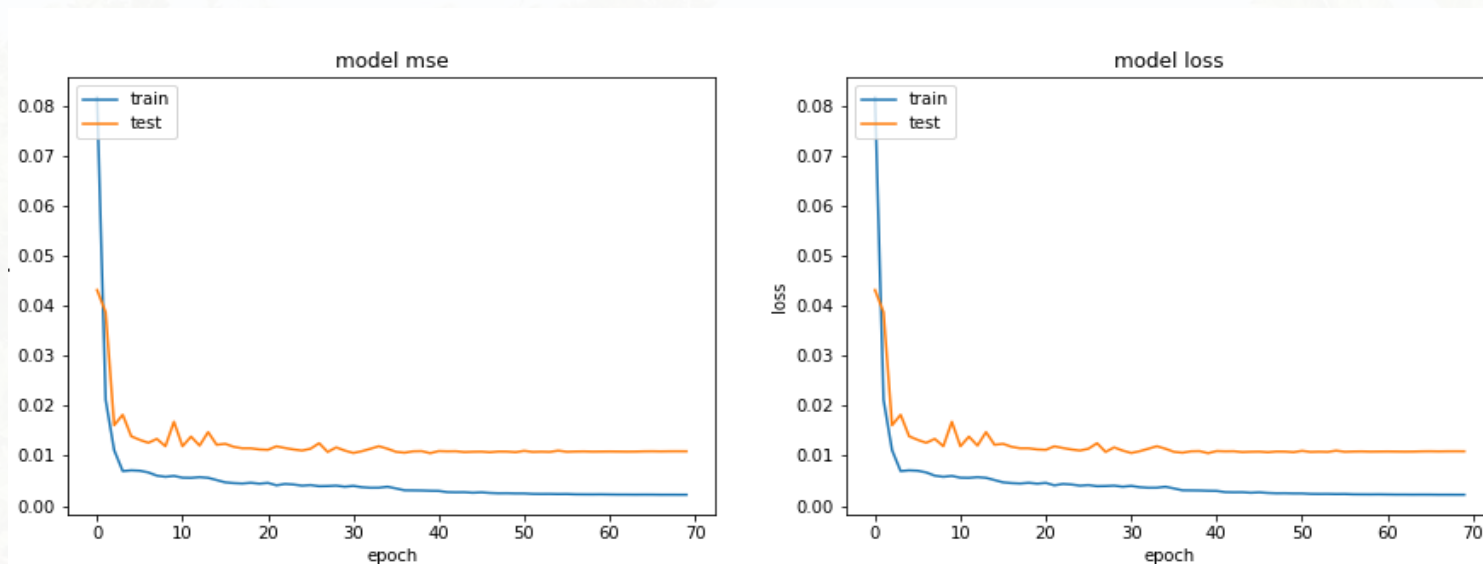
2層dense layer

結論 & 未來展望

4

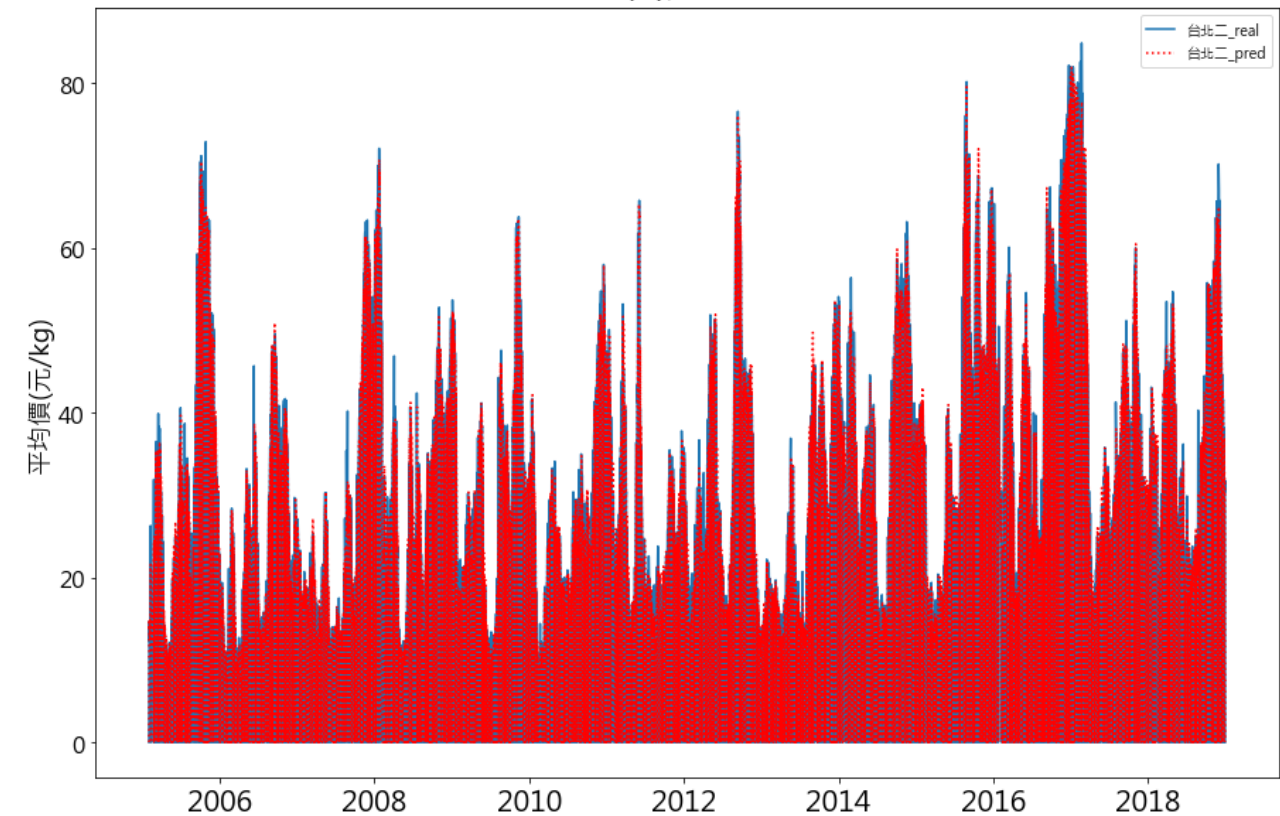


模型訓練結果



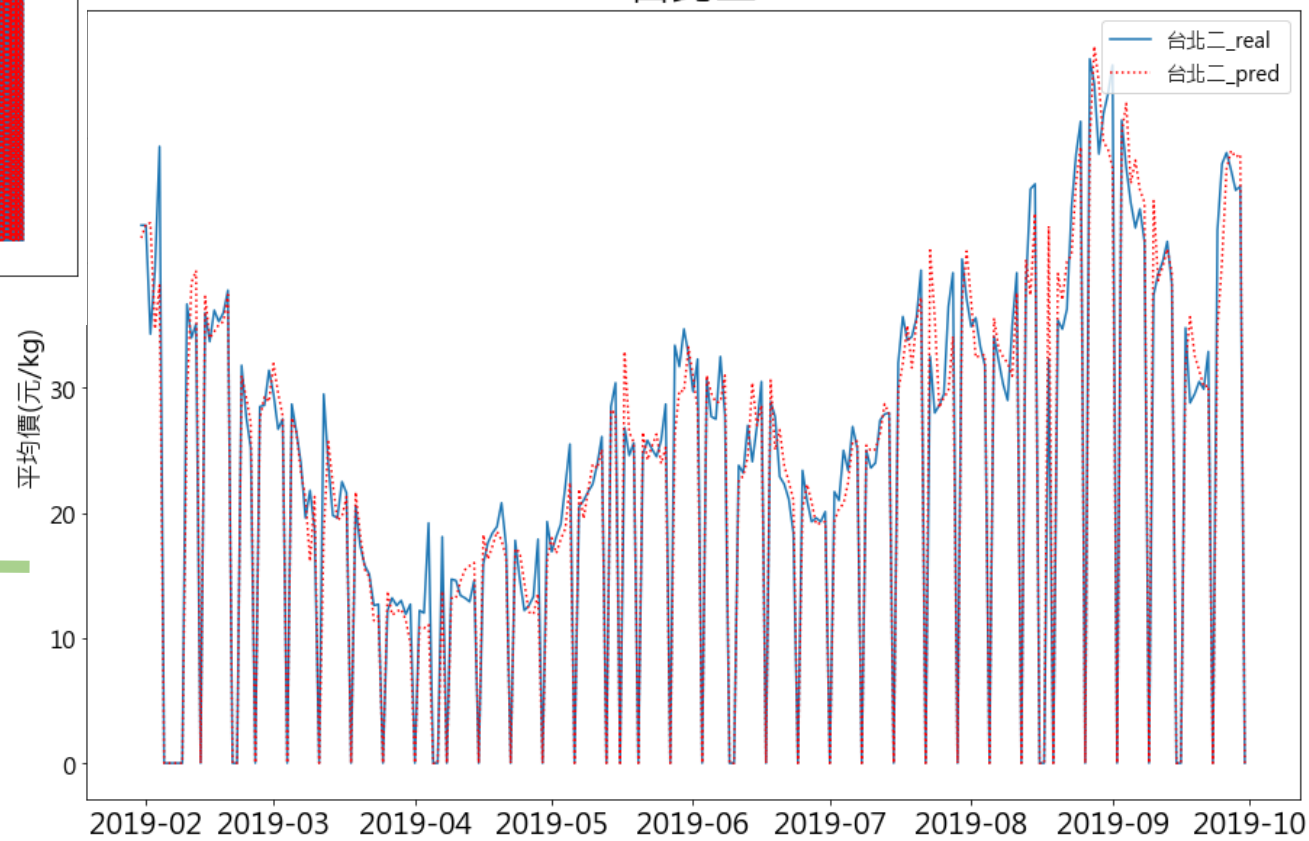
模型mse、loss皆有隨著訓練epoch增加而下降，雖偶有小幅震盪但訓練集(train)的表現始終優於驗證集(test)，表示模型有成功收斂且無出現過擬合(overfitting)的情況。

台北二



訓練之MSE	6.36
R-squared	0.98

台北二



預測之MSE	8.32
預測之RMSE	2.88

未來展望

1. 將所有大宗水果價格做關聯性分析，以找出符合現今社會之真正互補品及替代品，再投入模型訓練，並可將模型擴大為預測所有大宗水果價格以供更全面的價格預測及示警。
2. 目前的預測模型僅能預測隔一天之平均價格，若能取得逐筆交易資料和水果分級，便能預測更詳細的分類與更準確的價格，且若將時間線拉長至5天或7天，或許能將預測實際價格轉換為預測漲跌幅度，可以在維持預測準確率的狀況下提供政府更早的示警以做因應。



Thank you

Q & A