

# 2022力成科技 暑期數據分析實習成果

張祐瑋,系統支援課

# 工作內容

幫忙蒐集機台資料

數據清理、視覺化

建Hadoop

建模

跟AI廠商確認進度

預測、準確度分析



系統支援課  
工作內容

# 成果

1.

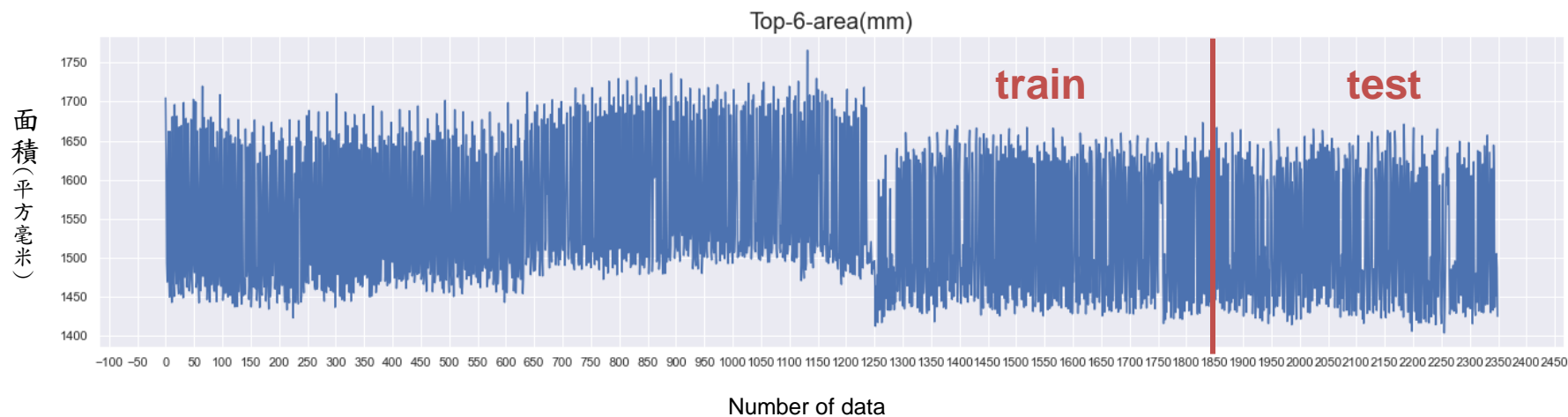
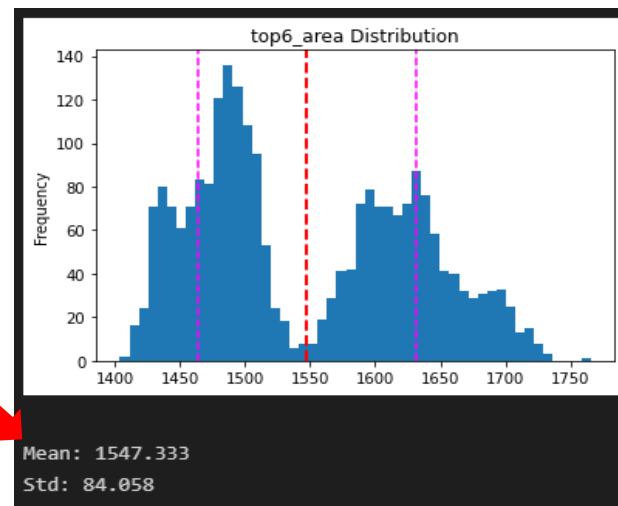
預測下一個時間點數值  
(預防機台異常情況發生)

2.

預測產品pass fail  
(輔助人工判斷產品製作成功與否)  
與AI廠商結果做比較

# 上蓋點膠點位

成果1



# 建模方法

成果1

- 訓練資料範圍

以全部6個點位的面積數據訓練，預測下一個時間的第6點位面積

以第6個點位的面積數據訓練，預測下一個時間的第6點位面積

- 訓練時間單位

每6個時間點為一次訓練單位

每12個時間點為一次訓練單位

- 模型


Linear Regression

XGBoost

XGBoost + Random grid search

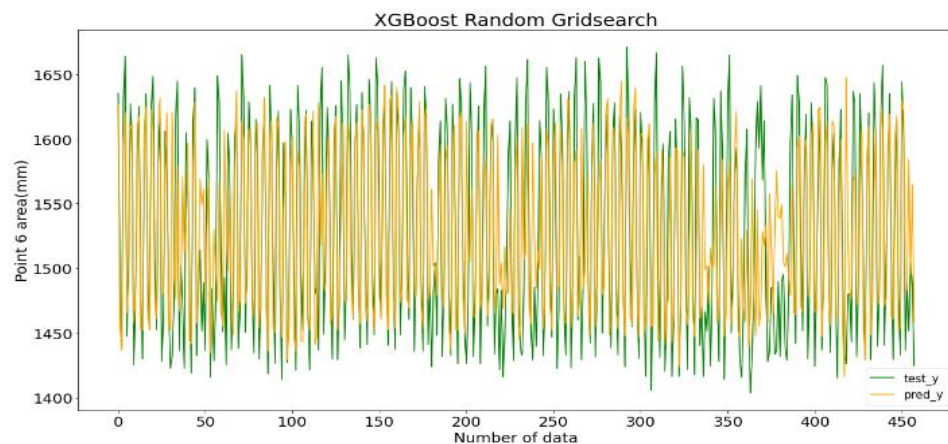
- 評估指標

MAE、MAPE、 $R^2$

預測第六點面積(mm <sup>2</sup> )	Linear Regression			XGBoost			XGBoost + Random GS		
	MAE	MAPE	$R^2$	MAE	MAPE	$R^2$	MAE	MAPE	$R^2$
全部點位+ 每6個時間點	43.593	2.899%	49.19%	38.741	2.539%	49.014%	 <b>34.29</b>	<b>2.261%</b>	<b>63.127%</b>
全部點位+ 每12個時間點	43.55	2.897%	49.669%	41.929	2.744%	49.506%	34.831	2.292%	62.135%
第六個點位+ 每6個時間點	45.886	3.051%	44.241%	38.306	2.519%	53.519%	36.316	2.389%	59.502%
第六個點位+ 每12個時間點	45.257	2.998%	46.648%	39.974	2.618%	52.245%	37.031	2.427%	58.993%

# 模型誤差

成果1

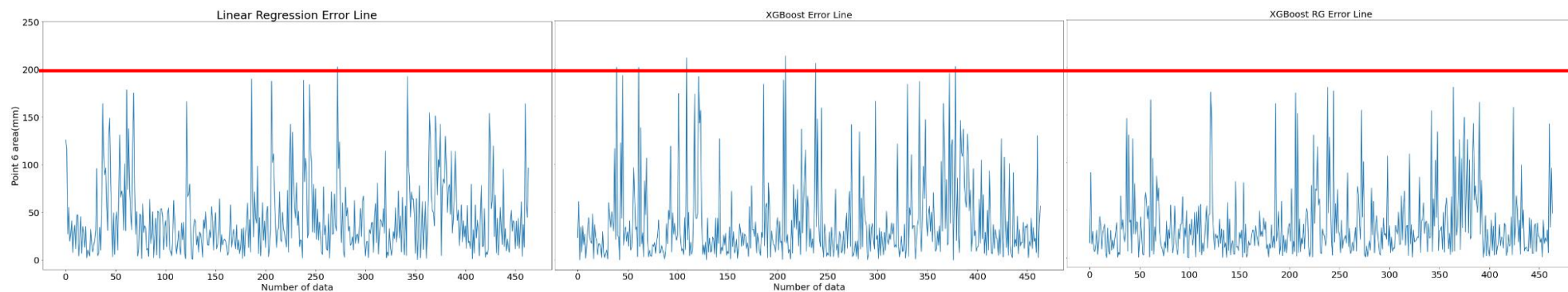


小補充

綠色：實際面積數值

黃色：預測面積數值

藍色：|綠色 - 黃色|



# 成果1結論

成果1

- 結果

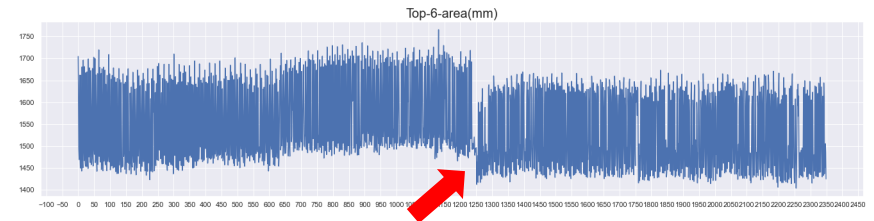
1. 以全部6個點位面積前期數據訓練的誤差 < 只以第6個點位面積為前期數據訓練的誤差。
2. 以每6個時間點為一次訓練單位的誤差 < 以每12個時間點為一次訓練單位的誤差。
3. 調整過超參數的XGBoost模型的預測誤差最小。

- 改善方向

1. 加入下蓋數據進行預測。
2. 增加數據量，避免過程中有機台維護的資料
3. 嘗試其他模型進行預測，ex: LSTM...

- 應用

根據模型預測下一個時間的數值，若下一個時間的數值脫離正常範圍，可立即停機進行維護，以預防機台的異常情況發生。





# 成果

1.

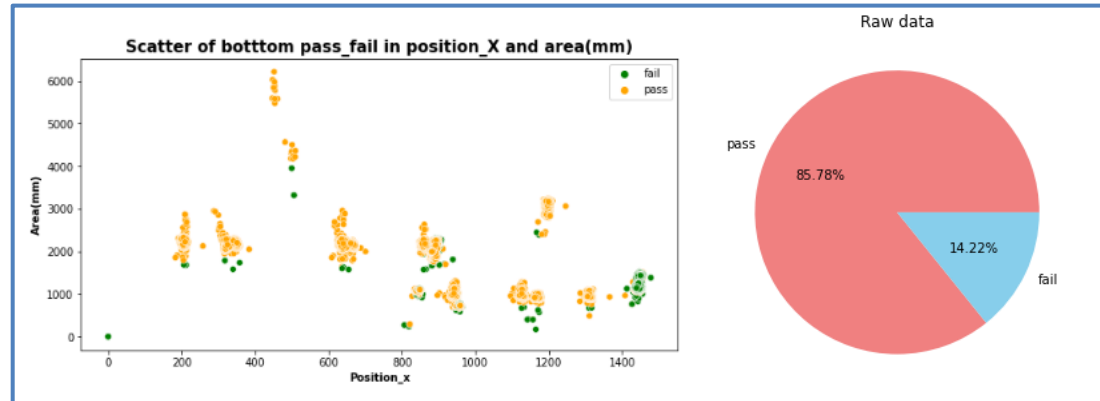
預測下一個時間點數值  
(預防機台異常情況發生)

2.

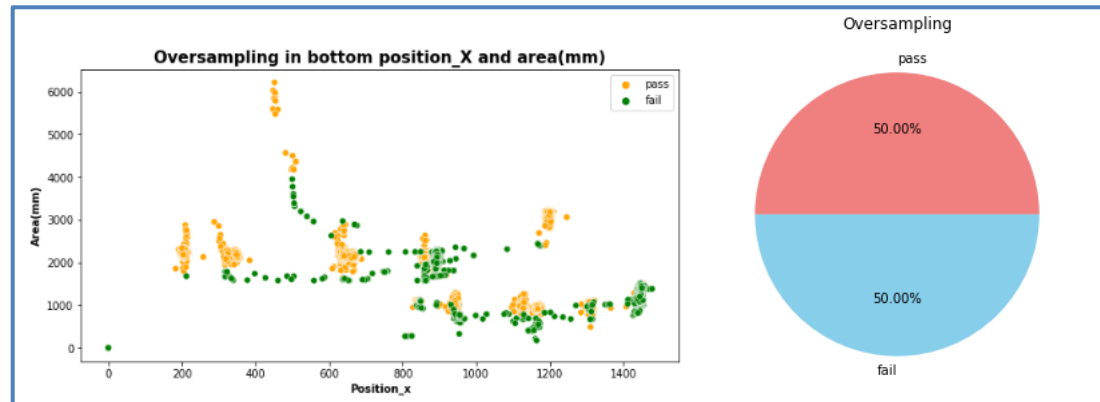
**預測點膠pass fail**  
(輔助人工判斷產品製作成功與否)  
與AI廠商結果做比較

# 下蓋資料

成果2




為了不讓機器盲猜



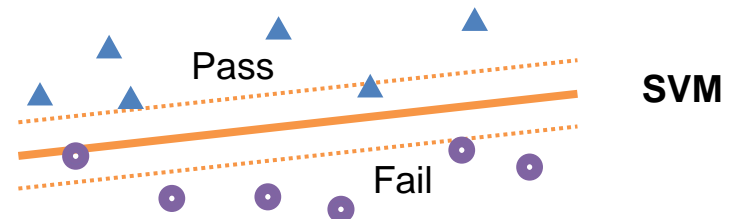
# 預測產品製作pass fail

成果2

	Logic Regression	 SVM
採樣方式	Oversampling(SMOTE)	Oversampling(SMOTE)
特徵	X位置、Y位置、產品長、產品寬、面積	X位置、Y位置、產品長、產品寬、面積
Precision	66%	<b>99%</b>
Recall	82%	<b>97%</b>
F1score	65%	<b>98%</b>
Accuracy	73%	<b>99%</b>

## 小補充

邊界最大化讓少數位於模糊地帶的資料較不容易影響模型的判斷標準



# Confusion matrix

成果2

	真實狀況	
	TP 6516	FP 61
預測狀況	FN 19	TN 1001

排除異常相連的  
fail資料，真正fail  
的100筆資料中只  
有5筆判斷錯誤

# 與AI廠商結果比對

成果2

針對產品為良品或不良品建模 分數：**0.992**

算法順序	準確率	混淆矩陣				召回率	精確率	f1	ROC 曲線下面積	準確率 (非CV)	混淆矩陣 (非CV)
		真陽性	假陽性	假陰性	真陰性						
0_Super_Stacked	0.9924	129	386	12	51732	0.9149	0.2505	0.3933	0.9537	0.9932	0.9966

預測pass fail					
模型	採樣方式	Precision	Recall	F1score	Accuracy
SVM	Oversampling/ Undersampling	100%	<b>62%</b>	70%	100%

預測fail的成功率 **25%**

# 成果2結論

成果2

- 結果

1. SVM模型判斷點位的產品是否異常的準確率達99%，較Logic Regression的模型準確度高。說明在此模型中輸入產品位置、面積等特徵，可判斷產品製作結果是否異常。
2. 與廠商預測結果部份相符，可能原因在於產品的fail資料過少，造成廠商在沒有分類或整理數據的情況下進行預測，而有模型precision較差的結果。

- 改善方向

1. 加入產品其他部分數據進行預測
2. 目前fail的資料依然過少，oversampling的採樣方式可能造成模型偏誤，未來還需增加fail的資料量，並嘗試用其他較不容易有偏誤的採樣方法。
3. 結合成果1，若要預測下一個時間點是否會Fail，需增加fail的時間註記。

- 應用

將此模型結合影像辨識可輔助人工判斷點膠成功與否，可減少人力。

可結合成果1，判斷下一個時間點的產品製作是否會fail。

**Thank you for your listening !**