

# 團隊測驗報告

報名序號：111924 (報名序號(格式:111XXX)已寄至隊長email)

團隊名稱：\_\_\_放寬心\_\_\_\_\_

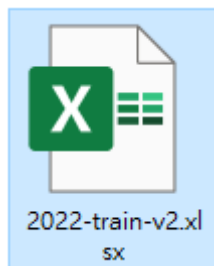
註1：請用本PowerPoint 文件撰寫團隊程式說明，請轉成PDF檔案繳交。

註2：依據競賽須知第七條，第4項規定：

測試報告之簡報資料不得出現企業、學校系所標誌、提及企業名稱、學校系所、教授姓名及任何可供辨識參賽團隊組織或個人身分的資料或資訊，違者取消參賽資格或由評審會議決議處理方式。

# 一、資料前處理(說明資料前處理過程)

Raw data



資料筆數:635  
資料欄位:131

Step1.缺失值處理

處理方式: 刪除缺失值比率過高之欄位

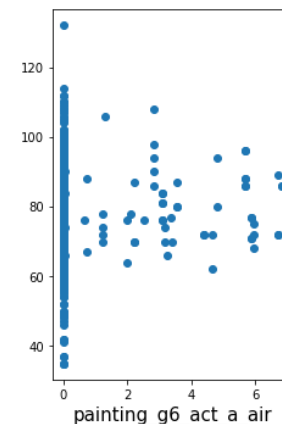
	Column Name	Missing Count	Percentage(%)
0	clean_ph5	286	0.450394
1	clean_ph7	286	0.450394
2	clean_ph4	286	0.450394
3	painting_g9_act_t_air	0	0.000000
4	painting_g9_act_hvv	0	0.000000
5	painting_g9_act_hvc	0	0.000000

資料筆數:635  
資料欄位:128

Step2.處理數值變化較小特徵不明顯之欄位

觀察各欄位數值變化，當數值變化較小，我們視為特徵不明顯，假定此欄位對預測值影響較小

處理方式: 刪除標準差 $<0.01$ 之欄位  
(ex: 資料大部分落於0之欄位，如下圖所示)



資料筆數:635  
資料欄位:110

# 一、資料前處理(說明資料前處理過程)



## 二、演算法和模型介紹(介紹方法細節)

將6個預測值分別進行訓練建立模型，以下以[sensor\_point5\_i\_value]為例，其餘以此類推

Step1. 經上一章節資料前處理完，X資料有61欄位，Y資料先取sensor\_point5\_i\_value對其進行建模

Step2. 採用blending作法 (將不同模型的預測值加權合成，權重和為1)。所使用模型如下所示:  
(XGBRegressor, SVR, GradientBoostingRegressor, CatBoostRegressor, RandomForestRegressor, KNeighborsRegressor, StackingRegressor)  
共7種模型，分別對其調參，取得模型

Step3. 採用SLSQP作法，計算各模型之權重，將多個模型之預測值分別乘以權重相加

	model	optimised_wts
0	xgb	0.696675
1	svr	0.050000
2	gbr	0.050877
3	catb	0.050192
4	rfr	0.050000
5	skr	0.050445
6	knn	0.051811

各模型權重



```
def blending_pred(X):  
    return (  
        0.70*xgbmodel.predict(X)+\  
        0.05*gbrmodel.predict(X)+\  
        0.05*svrmodel.predict(X)+\  
        0.05*catbmodel.predict(X)+\  
        0.05*rfrmodel.predict(X)+\  
        0.05*knnmodel.predict(X)+\  
        0.05*skrmodel.predict(X)  
    )
```

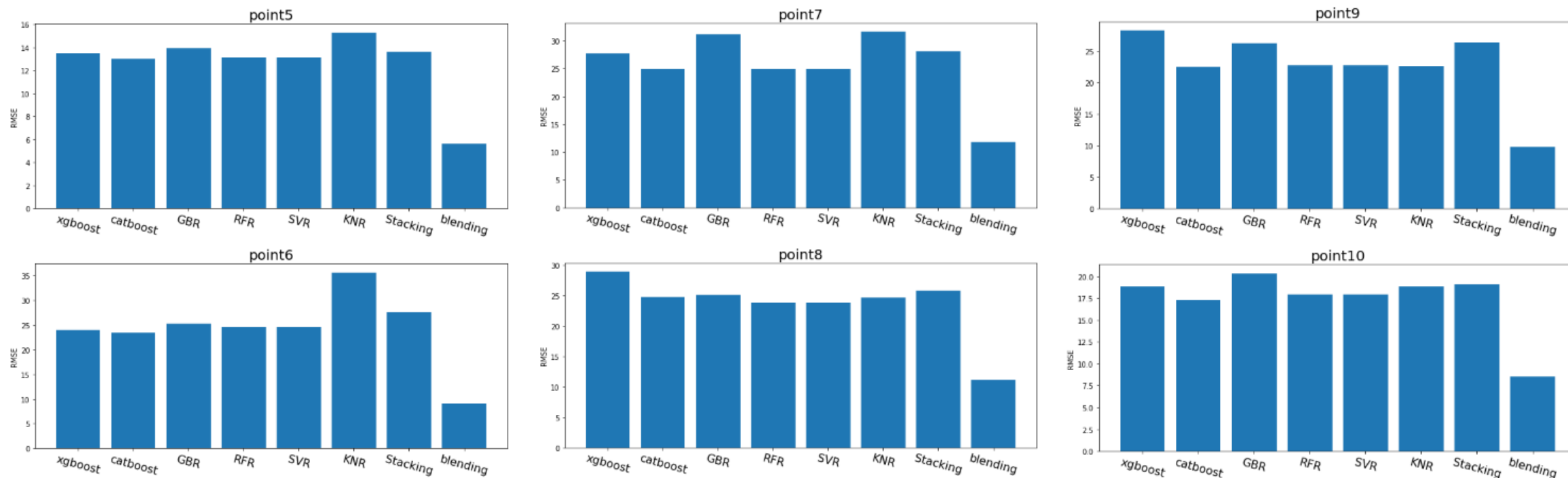
多個模型之預測值分別乘以權重相加

Step4. 對其餘5個預測值，重複步驟1~3，以取得個別之blending模型

### 三、預測結果

不同模型之RMSE結果如下圖所示，可觀察到針對不同的sensor\_point\_i\_value，blending做法都能有效降低RMSE，取得較好的成果。而此方法我們也使用於預測驗證資料集(2022-test-v1.xlsx)

模型簡稱( XGBRegressor : xgboost , CatBoostRegressor : catboost, GradientBoostingRegressor : GBR, RandomForestRegressor : RFR , KNeighborsRegressor : KNR, StackingRegressor : Stacking)



## 四、補充說明(其他或自行定義項目)