**PM2.5**

**線性回歸**

Mean Absolute Error (MAE)代表平均誤差，公式為所有實際值及預測值相減的絕對值平均。

**5.694068343810315**

Mean Squared Error (MSE)比起MSE可以拉開誤差差距，算是蠻常用的指標，公式所有實際值及預測值相減的平方的平均

**54.77458000595562**

Root Mean Squared Error (RMSE)代表MSE的平方根。比起MSE更為常用，因為更容易解釋y。

**7.400985069972484**

**K相鄰(KNN)**

1448筆資料時，準確度knn.score(X\_test, y\_test) ------>**0.9287356321839081**

可用 For迴圈，選擇最佳的K值

可給予一筆資料，判定是否對人體有危害

KNN用於分類，也可使用KNN進行迴歸計算

**隨機森林**

很多棵的決策樹，屬於整體學習的一種。

可繪製出決策樹、進行迴歸計算

**可找出特徵重要程度**  
Variable: NO2 Importance: 0.48

Variable: TEMP Importance: 0.22

Variable: CO Importance: 0.13

Variable: O3 Importance: 0.06

Variable: SO2 Importance: 0.02

Variable: NMHC Importance: 0.02

Variable: Humidity Importance: 0.02

Variable: Nox Importance: 0.01

Variable: NO Importance: 0.01

Variable: THC Importance: 0.01

Variable: CH4 Importance: 0.01

Variable: WindSpeed Importance: 0.01

Mean Absolute Error: 7.32 degrees.

**邏輯迴歸**

判斷0~1之間的機率有多少

precision recall f1-score support

0 0.96 0.98 0.97 560

1 0.98 0.96 0.97 526

micro avg 0.97 0.97 0.97 1086

macro avg 0.97 0.97 0.97 1086

weighted avg 0.97 0.97 0.97 1086

Confusion Matrix 預測 (No) 預測（Yes）

實際（No） TN=549 FP=11

實際（Yes） FN=20 TP=506

TN=True Negative; FP=False Positive, FN=False Negative; TP=True Positive

precison = TP / (TP + FP)

recall = TP / (TP + FN)

accuracy = (TP + TN) / (TN + FP + FN + TP)

F-beta score可視為precision和recall加權平均的值，數值介於0-1，最好是1。

support代表在測試組y實際值的發生次數。\*\*

**多種演算法比較**

Scores:

**Linear regression score**: 0.9101721045818417

**Neural network regression** score: 0.8532397997401182

**Lasso regression** score: 0.8873130463758103

**ElasticNet regression** score: 0.888862421794889

**Decision forest** score: 0.9023960871039343

**Extra Trees** score: 0.9167571974098775

**Boosted decision tree** score: 0.892229896203619

**XGBoost** score: 0.8902505097601974

RMSE:

**Linear regression** RMSE: 3.74

**Neural network** RMSE: 4.78

**Lasso** RMSE: 4.19

**ElasticNet** RMSE: 4.16

**Decision forest** RMSE: 3.90

**Extra Trees** RMSE**:** 3.60

**Boosted decision tree** RMSE: 4.09

**XGBoost** RMSE: 4.13

**TensorFlow神經網路迴歸**

一般迴歸預測PM2.5

The Explained Variance: 0.90

The Mean Absolute Error: 2.94

The Median Absolute Error: 2.19

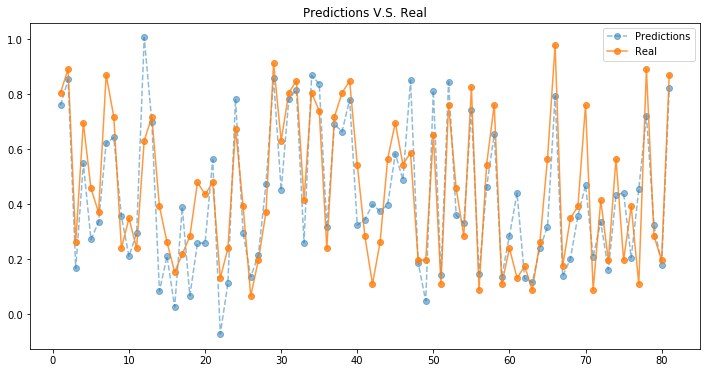
神經網路預測PM2.5

The Explained Variance: 0.96

The Mean Absolute Error: 2.17

The Median Absolute Error: 1.55

**Keras深度學習**

****

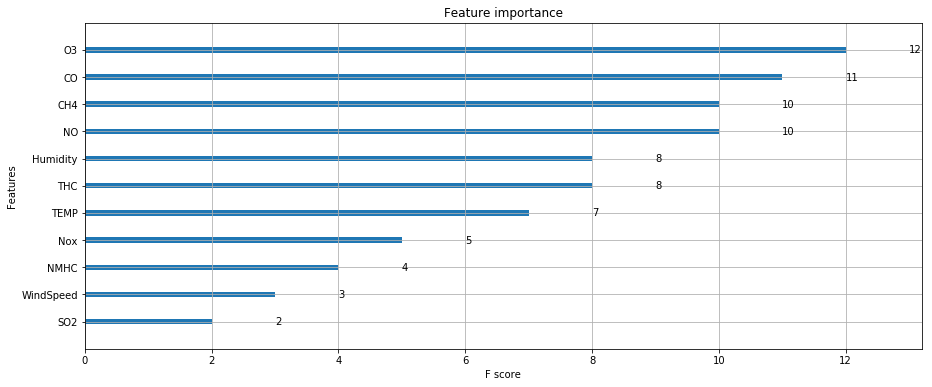
**統計學預測分析方法(移動平均、指數平滑法……)**

幾乎沒有準確率可言，因此不擺上成果

**XGBoost**

號稱為資料科學比賽中最強的演算法

RMSE: 9.892907

****

**PM2.5與風向之關聯性**

**The Impact of Meteorological Factors on PM2.5 Variations in Hong Kong**

The north wind in winter increased the PM2.5 in Hong Kong while **the south wind in summer decrease the PM2.5.**<https://iopscience.iop.org/article/10.1088/1755-1315/78/1/012003>

**Effects of Wind Direction on Coarse and Fine ParticulateMatter Concentrations in Southeast Kansas**

The effect of wind direction was found to be **statistically significant** for both PM10and PM2.5.   
  
<https://www.tandfonline.com/doi/pdf/10.1080/10473289.2006.10464559>

**PM2.5與溫度之關聯性**

溫度確實會影響，但是，是因為風向所造成。

例如:印度溫度也很高，但是空氣汙染也很嚴重，因此，溫度並非主要指標。