**國立雲林科技大學資訊管理系**

**資料探勘**

專案作業二

以SVM、AdaBoost

分析混泥土抗壓強度及通過物理測量鮑魚年齡

M10823022蔡成霖

M10823029李承融

M10823040古力維

M10823041郭勁宏

指導教授：許中川

2019 年 11 月

**摘要**

本研究從UCI資料庫中，取自一個Concrete Compressive Strength Data Set和一個Abalone Data Set做為分析資料集，我們使用混泥土的抗壓強度的數據希望能跟了解鮑魚如果是吸附在混泥土上而不是岩石上，那會不會影響鮑魚的生活環境。其中包含八個特徵屬性及一個類別屬性，本研究利用SVM與AdaBoost去分析混泥土抗壓強度及通過物理測量鮑魚年齡。最後我們得出結果，AdaBoost的準確度較SVM的準確度高。

關鍵字：SVM、AdaBoost、混泥土抗壓強度、通過物理測量鮑魚年齡

1. **緒論**

**1.1 動機**

現今是一個科技蓬勃發展的年代，那因為氣候的變遷很大原因跟人為汙染有關，希望透過鮑魚的研究 去了解每年能捕獲的鮑魚的數量，藉此了解鮑魚在現代有人為因素環境下平均年齡分布，是否因為人為因素使鮑魚數量降低，鮑魚是[中國](https://zh.wikipedia.org/wiki/%E4%B8%AD%E5%9C%8B)「四大海味」之一，那我們也發現自古以來人類就有捕撈鮑魚的行為，那為何到了21世紀的現在才發現對鮑魚有很大影響 ?

那根據生物學專家指出鮑魚喜歡棲息在水流暢通、天然餌料豐富、水質清晰、水深度為1.5-21m範圍的海區，喜歡平躺著生活，晝伏夜出。足部的附著力很強，殼長15cm，一旦充分附著岩石後，需用100Kg左右力方可拔掉，所以能在大風大浪中生存。受驚動或襲擊時迅速收縮，移動速度慢，1分鐘爬行50～80厘米。那我們使用混泥土的抗壓強度的數據希望能跟了解鮑魚如果是吸附在混泥土上而不是岩石上，那會不會影響鮑魚的生活環境，因為根據近年來人類發展，已經開發很多填海造陸的設施，混泥土等人工陸地是否會影響鮑魚棲息的環境已成為生態問題。

**1.2目的**

過去研究可能都只是在探討一個物種的數量問題，但很多時候沒有考慮到其他因素，比如過去只有談到北極熊是因為什麼樣的氣候而造成生態問題，那本次研究主要是加入人為因素做考量，正好用鮑魚的生態為例，鮑魚是從古至今都有的一個物種，為何以前人們大量捕食鮑魚都沒有造成生態威脅，而在現在卻造成鮑魚生態上的威脅甚至到汙染，這很大的原因跟人為加工和不當排放汙染物有關，利用混泥土可以看出當抗壓強度越強時，是否對於鮑魚的數量和年齡有很大的關係，因為抗壓程度可以看出一個地方人為加工的多寡，當加工量一大，是否會影響鮑魚生活在種環境的數量，是不是當有混泥土填海造陸的地方鮑魚的數量和年齡會減少? 還是增加?

1. **方法**

本研究將採用SVM與AdaBoost進行資料分析，目的在於確認在SVM與AdaBoost的績效差異為何，在此研究中我們使用了兩種不同的分類方式，SVM以及AdaBoost去進行分類。最後探討這兩種分類方式哪一種方式較佳。



讀取資料

前處理

切分訓練及測試資料

建立分類器

預測

績效計算

輸出結果

圖 **1** 方法流程架構圖

1. **實驗**

**3.1資料集**

本研究資料集來自於UC Irvine (UCI) Machine Learning Repository，資料集為Contraceptive Method Choice Data Set，總共有九個欄位，其中包含八個特徵屬性及一個類別屬性，資料筆數為1030筆，資料集型態個別為混合型態以及純數值型態，並以下接續介紹。

表1混泥土抗壓強度所得資料集

|  |  |  |
| --- | --- | --- |
| 欄位名稱 | 譯名 | 型態 |
| Cement | 水泥 | Float |
| Blast Furnace Slag | 高爐渣 | Float |
| Fly Ash | 粉煤灰 | Float |
| Water | 水 | Float |
| Superplasticizer | 高效減水劑 | Float |
| Coarse Aggregate | 粗聚合 | Float |
| Fine Aggregate | 細骨料 | Float |
| Age | 年齡 | Float |
| Concrete compressive strength | 混凝土抗壓強度 | Float |

表2通過物理測量鮑魚年齡

|  |  |  |
| --- | --- | --- |
| 欄位名稱 | 譯名 | 型態 |
| Sex | 性別 | String |
| Length | 長度 | Float |
| Diameter | 直徑 | Float |
| Height | 高度 | Float |
| Whole weight | 重 | Float |
| Shucked weight | 剝皮重量 | Float |
| Viscera weight | 內臟重量 | Float |
| Shell weight | 殼重量 | Float |
| Rings | 環 | Int |

**3.2實驗設計**

我們使用Python3.7版本，先繪製實驗流程圖如圖3.2.1，匯入資料並把部分資料做前處理，且確認檢查資料是否有遺失如圖3.2.2，把數據標準化的同時，如圖3.2.3，切分訓練測試(70% 和30%)資料，把測試的資料分別使用SVM與AdaBoost做訓練，並從中觀察資料特性，最後算出正確率。

### 混泥土抗壓強度 SVM

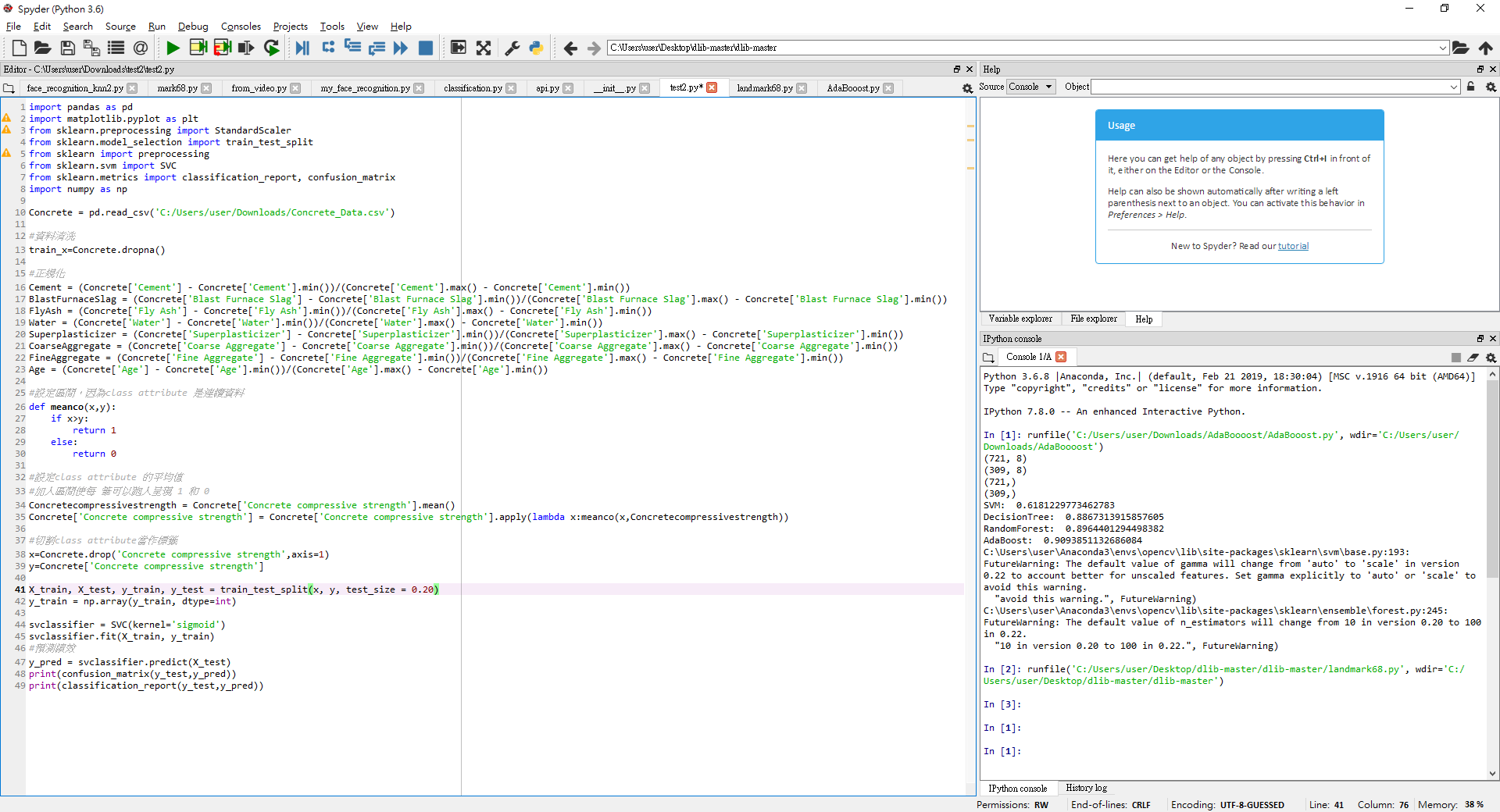


圖 2 混泥土抗壓強度 SVM 程式碼

### 混泥土抗壓強度 AdaBoost

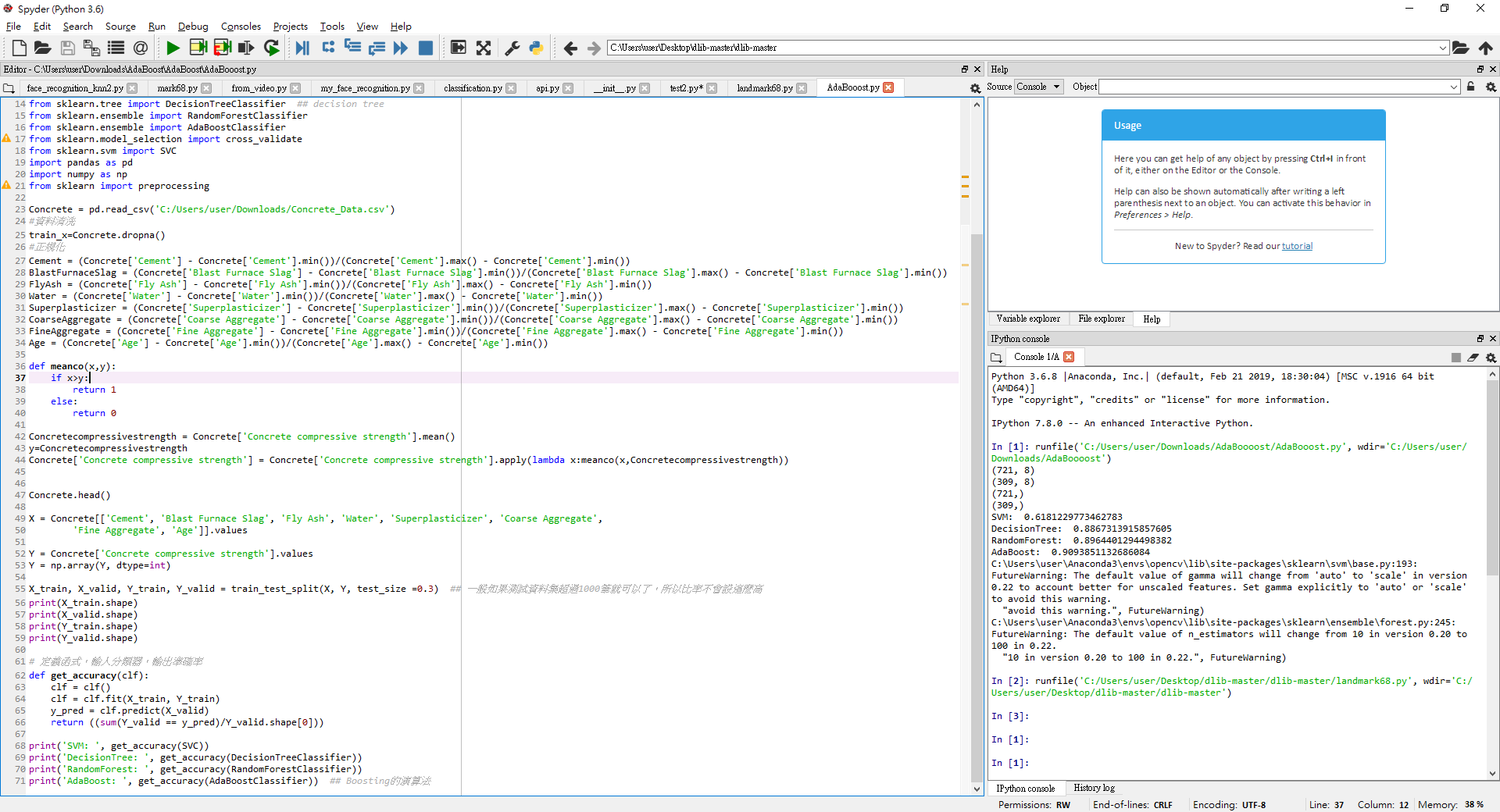


圖 3 混泥土抗壓強度 AdaBoost 程式碼

### 通過物理測量鮑魚年齡 SVM

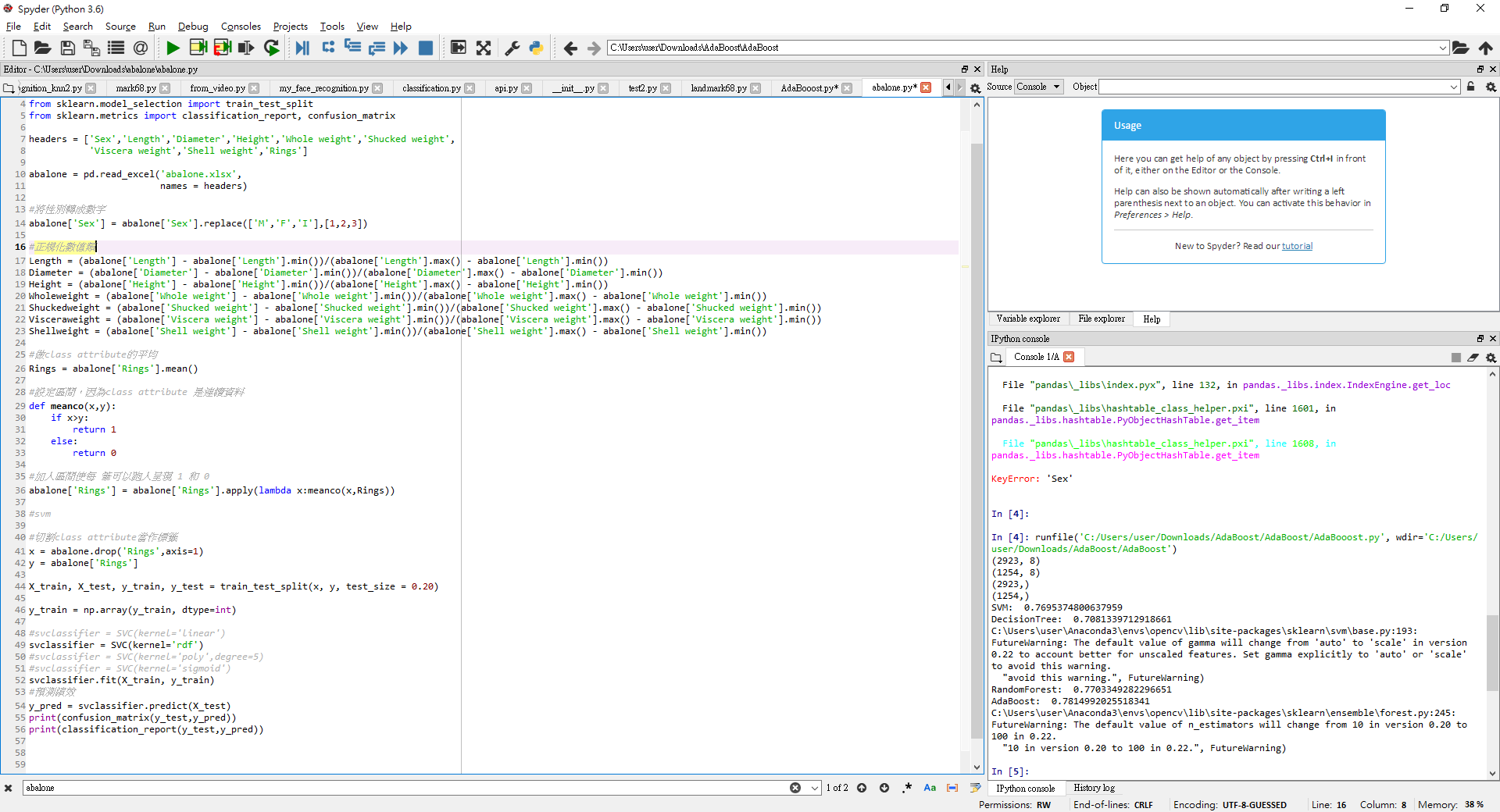


圖 4 通過物理測量鮑魚年齡 SVM

### 通過物理測量鮑魚年齡 AdaBoost

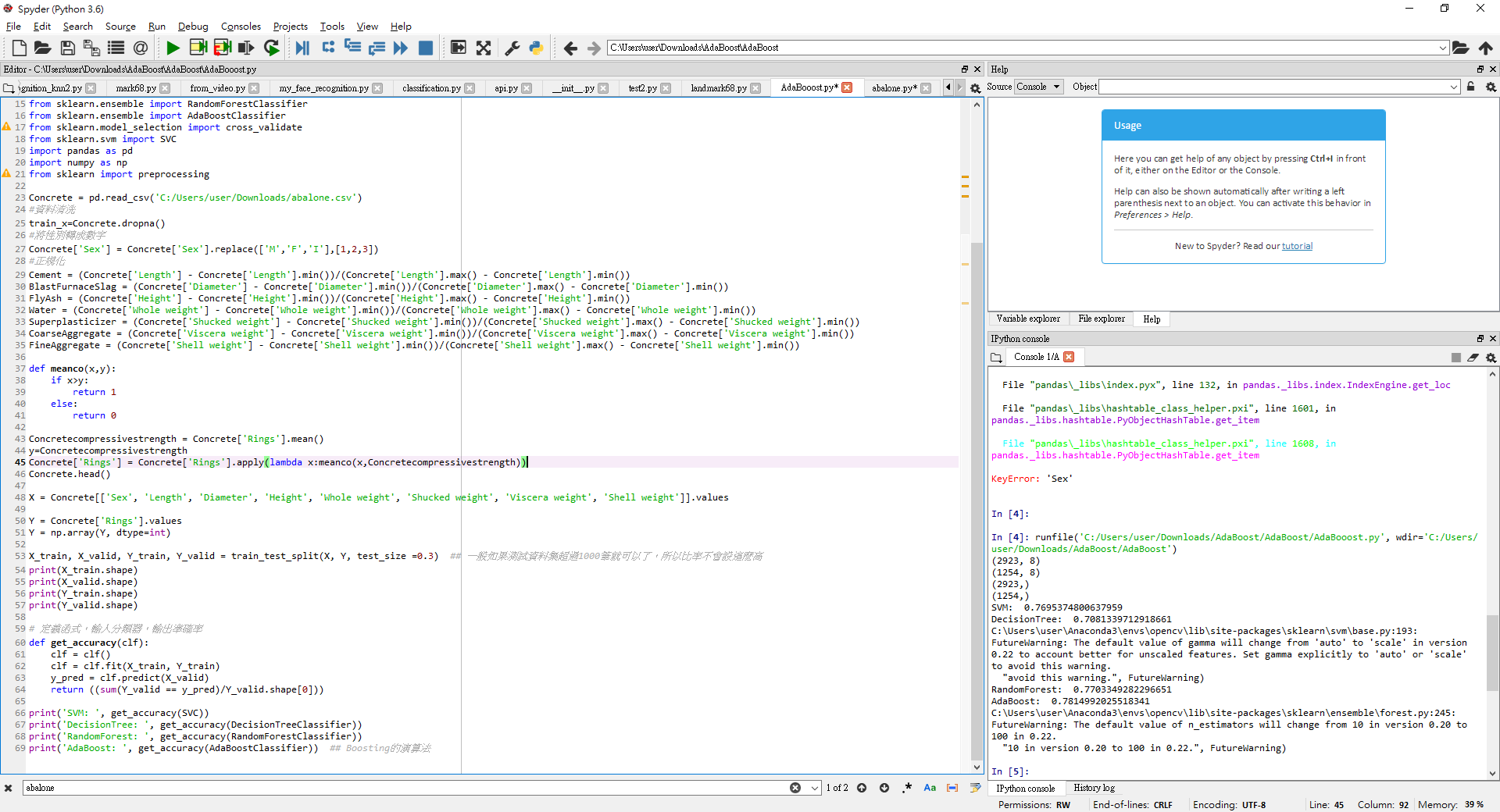


圖 5 通過物理測量鮑魚年齡 AdaBoost

**3.4** 實驗結果

本研究小組利用兩種分類方式對資料集進行分析的結果如下：

1. SVM：此方法利用 SVM 來進行分類，最後得出混泥土抗壓強度準確度為 0.87，通過物理測量鮑魚年齡準確度為 0.88。

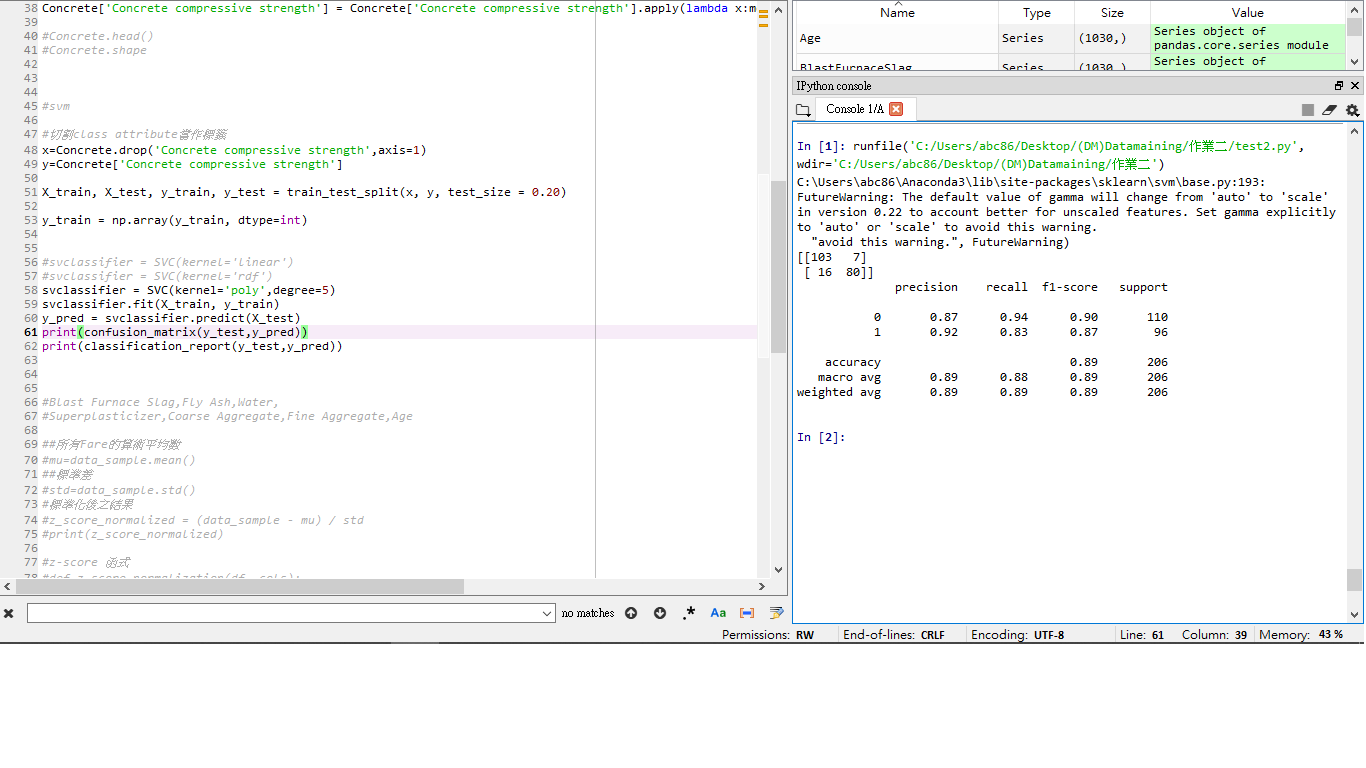


圖 **6** SVM 混泥土抗壓強度所得程式結果畫面

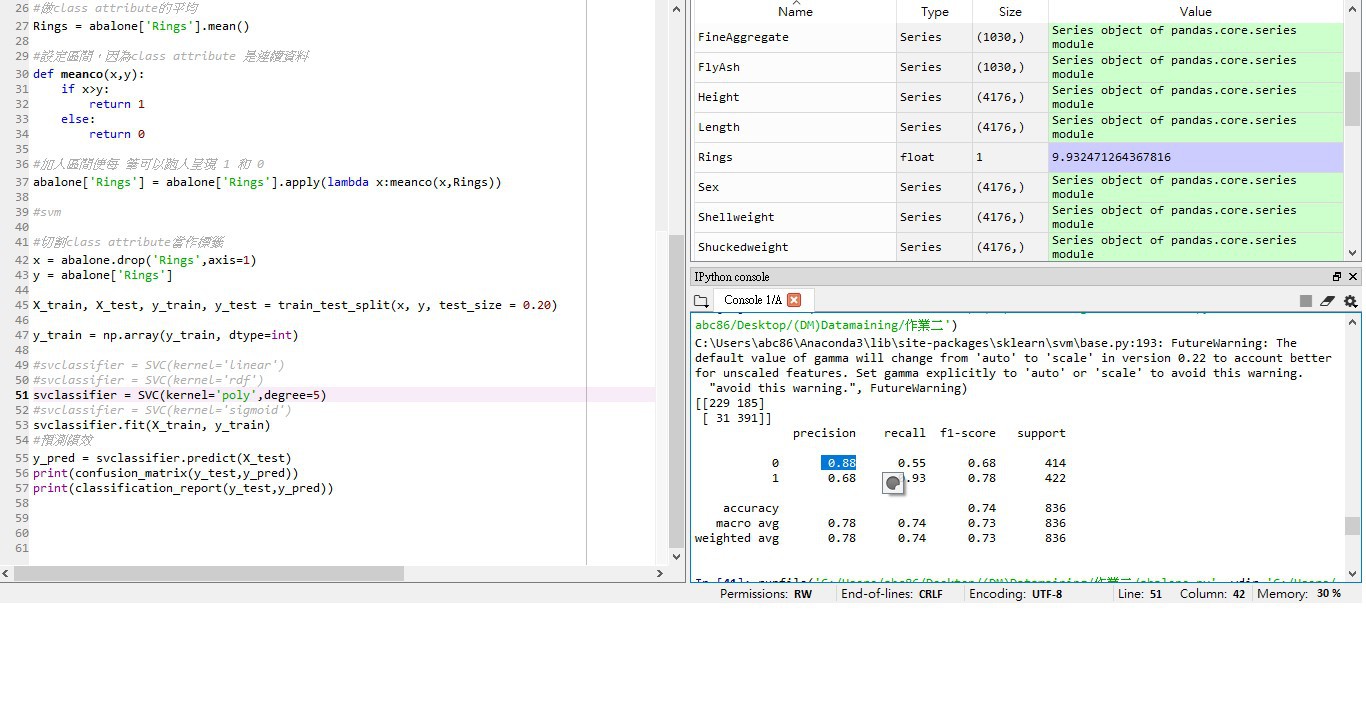


圖 **7** SVM 通過物理測量鮑魚年齡程式結果畫面

1. Adaboost：此方法利用使用 Adaboost 進行分類，最後得出混泥土抗壓強度準確度為 0.88，通過物理測量鮑魚年齡準確度為 0.76。

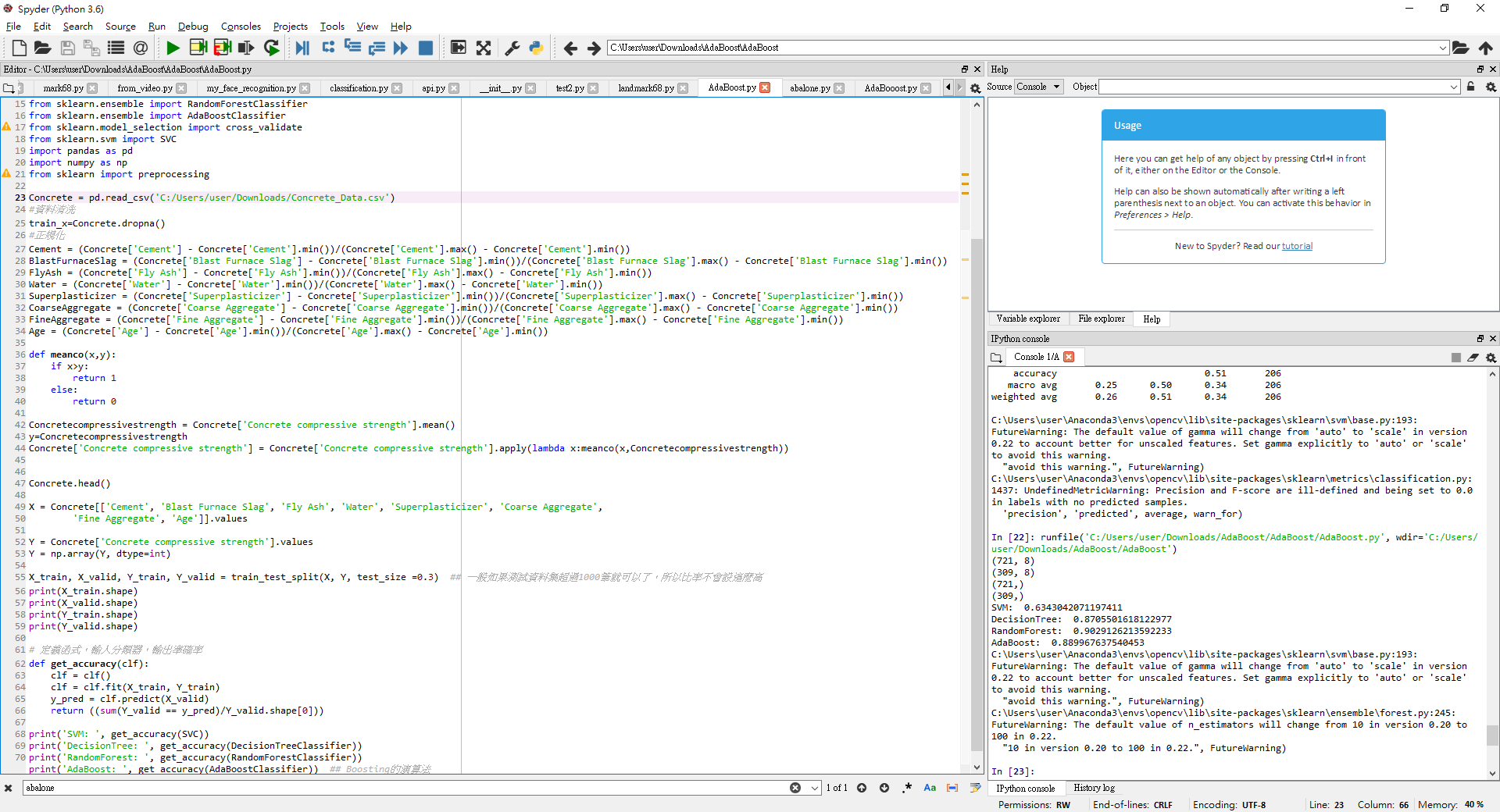


圖 **8** Adaboost混泥土抗壓強度所得程式結果畫面

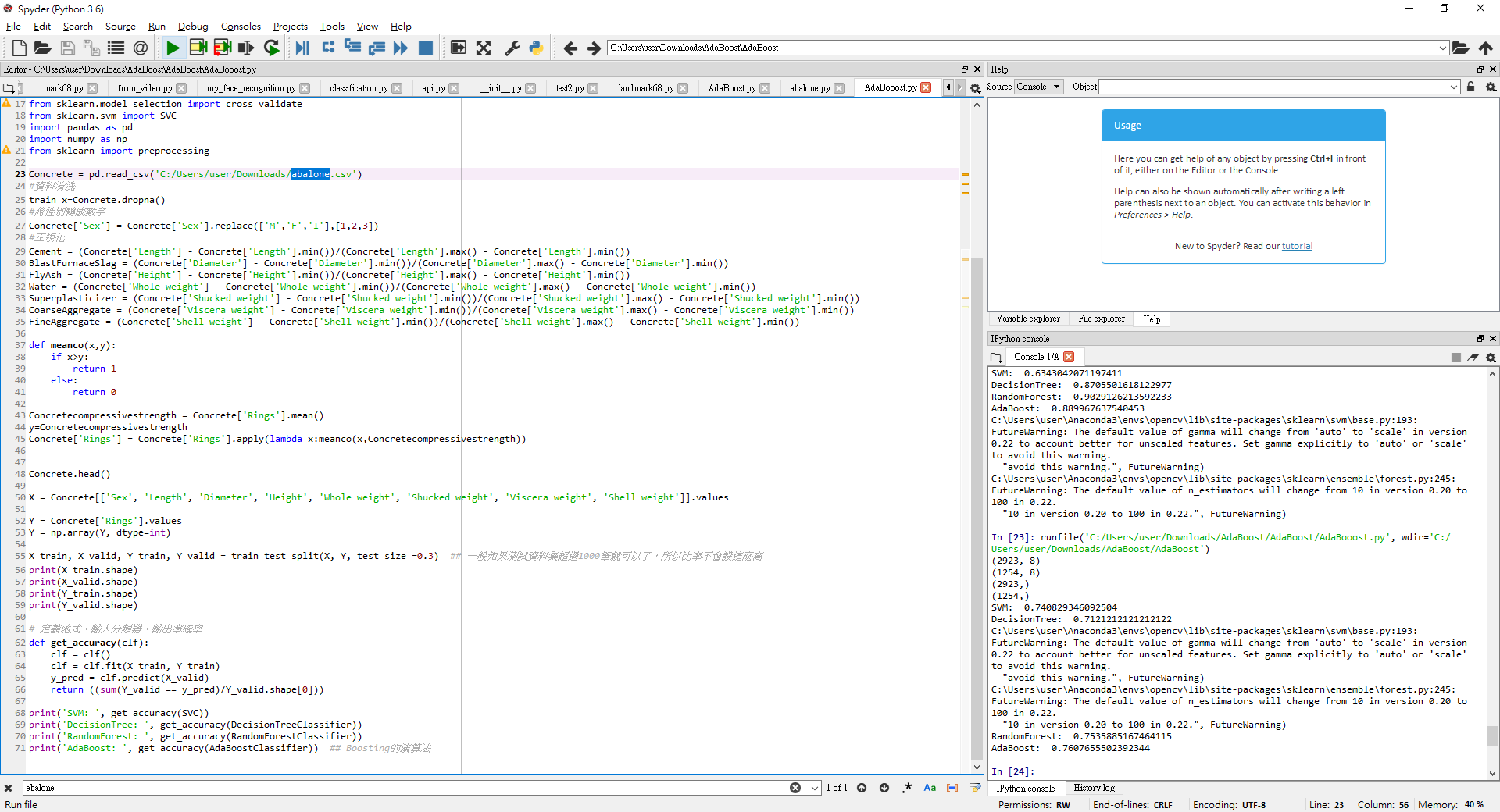


圖 **9** Adaboost 通過物理測量鮑魚年齡程式結果畫面

# 四、結論

本研究是以Concrete Compressive Strength Data Set 與 Abalone Data Set 為主的數據資料，並使用 SVM 及 AdaBoost 去進行分類並評估兩者的績效及準確度，最終我們得出的結果個別為 SVM：0.87 及 0.88，AdaBoost 為 0.88 及 0.76。因此，藉由以上的兩種資料分類法，得知 SVM的準確度較高。

# 參考文獻

<https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>

<http://archive.ics.uci.edu/ml/datasets/Abalone>

<http://ir.lib.ntust.edu.tw/retrieve/50882/Compared+Artificial+Neural+Networks+with+Experimental+Formulae+in+Predicting+Strength+of+Concrete.pdf>