**kaggle Driver Detect**

1. **參與者資訊:**

101308016 風管四 周俊川

101701016 應數四 杜靖愷

103304042 統計二 劉乃慈

**貳、情境說明:**

許多車輛意外原始於駕駛者的分心。他們可能一邊駕駛一邊滑著演書，或是一邊轉頭跟後座的朋友談話。根據統計資料顯示，高達五分之一的車禍原因是駕駛者的分心。為了改善這個狀況，我們利用監視器中拍攝到的畫面，試著分析出駕駛者現在的駕駛狀況，他是否正專注於其他事物，是否正在從事其他事情......，並希望藉由分析結果警告有高事故危險性的駕駛者，讓他們能夠提高警覺。

**Input：**監視器的圖片(csv檔)

**Output：**預測的駕駛者狀態，分為下列10種狀況

c0: 安全駕駛者(safe driving)

c1: 正在用右手打字駕駛者(texting – right)

c2: 正在用右手講電話的駕駛者(talking on the phone – right)

c3: 正在用左手打字駕駛者(texting – left)

c4: 正在用左手講電話的駕駛者(talking on the phone – left)

c5: 正在操作收音機的駕駛者(operating the radio)

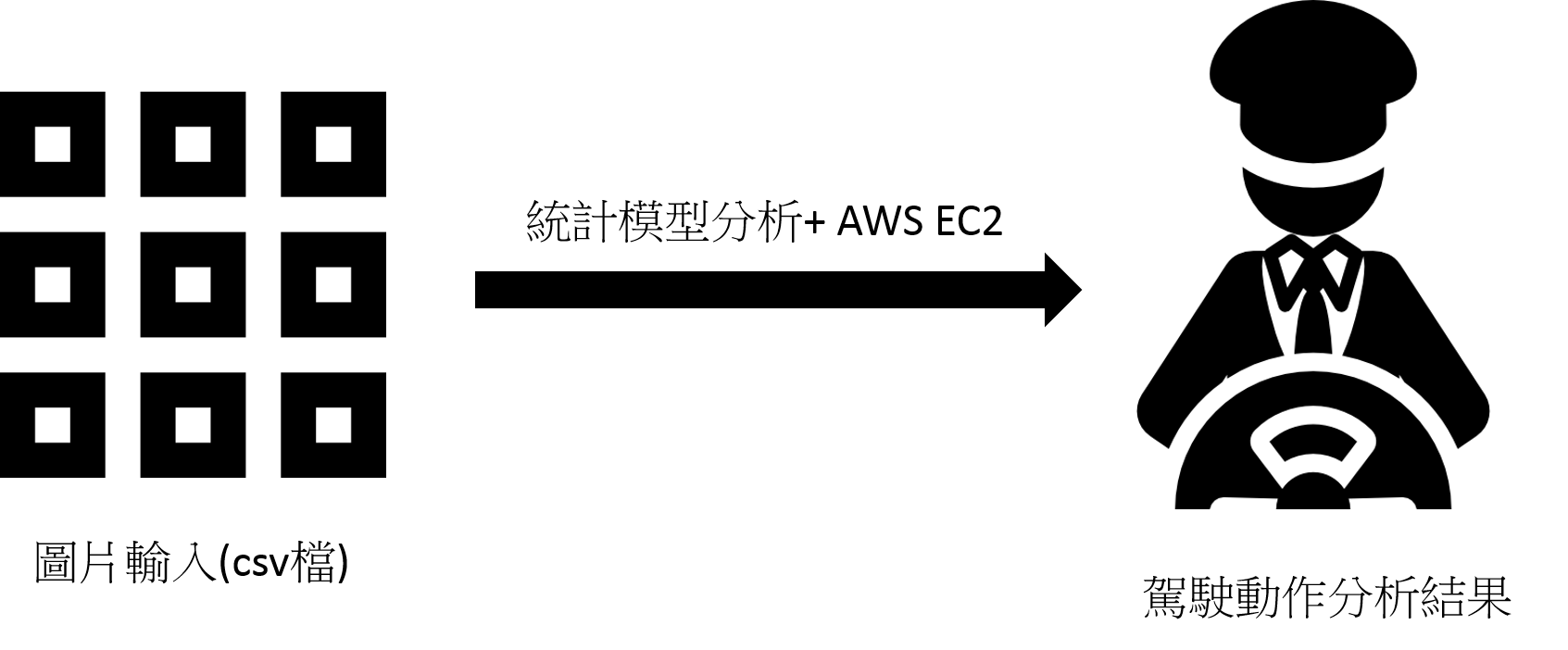
c6: 正在喝飲料的駕駛者(drinking)

c7: 正在向後座拿東西的駕駛者(reaching behind)

c8: 正在化妝的駕駛者(hair and makeup)

c9: 正在談話駕駛者(talking to passenger)

**情境表示圖像:**

****

1. **資料集來源與處理**
   1. 資料來源

kaggle 網站( <https://www.kaggle.com/c/state-farm-distracted-driver-detection>)

* 1. 資料處理

為了讓資料量減少，以便於後續分析執行速度，在執行前，我們對資料進行了三個動作的處理:彩色轉黑白、格數下降、轉成 sparsevector型式，以下就各點進行說明。

資料輸入時為彩色格式，且分別有(紅，綠，藍) 三原色的數值。為了使資料量減少，且顯然地，圖片為彩色或黑白並不影響分析結果，因此我們使用了python裡內建的函式Canny，在不影響圖片分析下，將圖片轉為黑白。

另外我們試著將照片格數降低，已到資料量減少的效果。在本專案中，一個pixel(像素)即為一個feature，原輸入資料的格數共有，精簡化後，我們順利將資料格數降為。

最後，我們將每張圖片的pixel 轉為一列的feature，再轉為spark Mlilb 中的型態 sparsevector，表示為SparseVector(3072,{11:1.0,……})，代表總共有3072個feature，並且第11個feature值為1(黑)。因此在降低資料量的同時，也有助於後續匯入spark中進行分析。

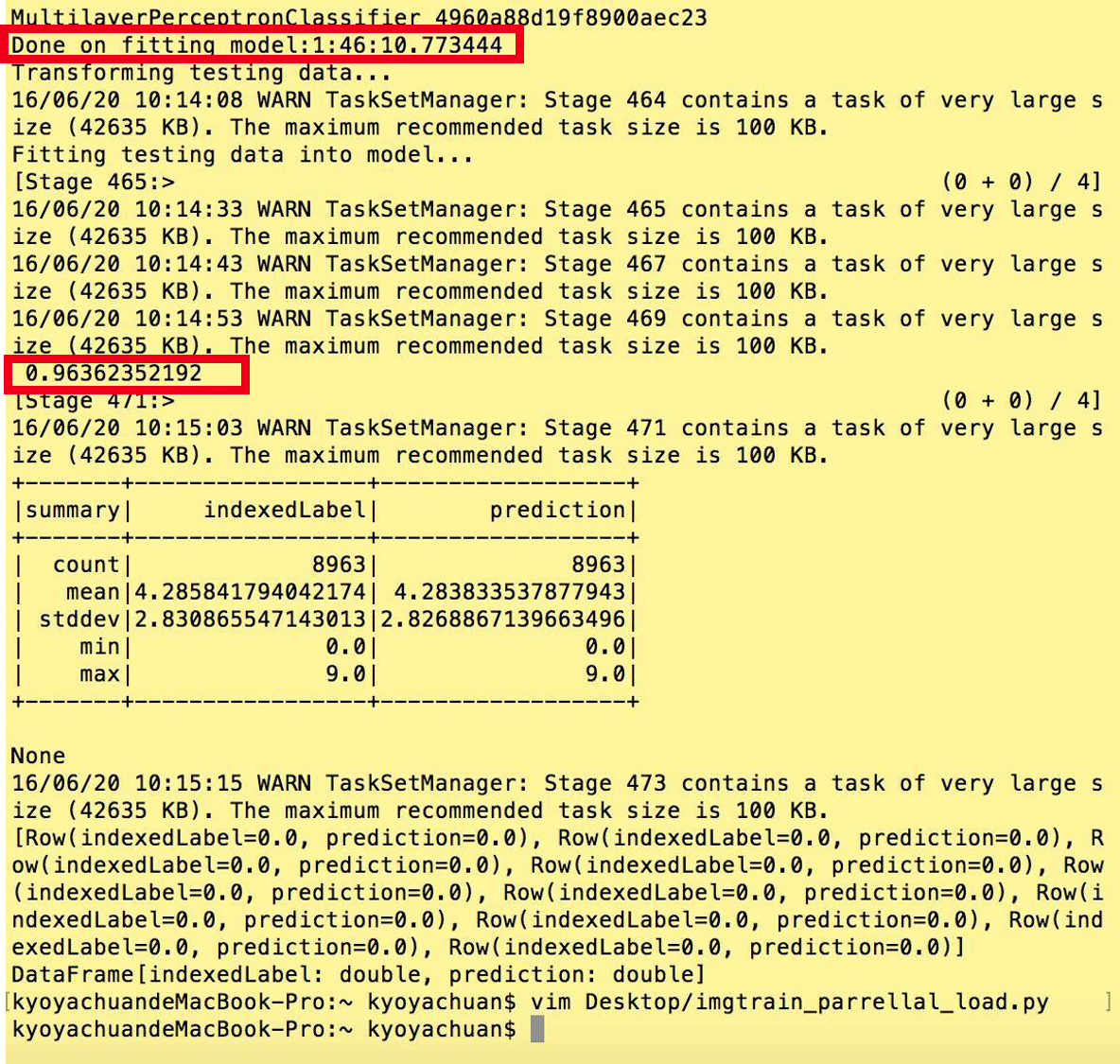
* 1. 特徵值的選取策略
  2. 分析模型演算法的挑選

我們使用了multilayer的演算法進行分析。

* 1. 建模、參數與超級參的優化處理
  2. 最佳模型的確認
  3. 測試資資料集的評量準則

1. **成果圖形化與解讀**

下圖為執行結果，以最佳模型進行最後的資料測試後，我們總共執行了1個小時46分10秒，準確率高達96%

****

1. **應用的系統平台說明與Spar函式庫的使用**

在本次的專案中，我們使用的工具有Spark、Slack以及github。我們利用Spark中的Mllib和Ml進行資料分析，並利用Slack 進行組內溝通，最後使用github 進行專案程序管理。

1. **系統效能（含記憶體使用量與運算效能）**