學號 ：R07323035 系級 ： 經濟系碩二 姓名 ： 王嚴

1.

Logistic model 的準確度較佳，可能是因為這筆資料的分配並沒有很像高斯分配所以 generative 的表現沒有那麼好

2.

由於原始資料中，各個 feature 的數字大小差異頗大，如果直接跑 model 的話，會使得數字較大的 feature 對於預測結果會有比較大的影響，所以做標準化能解決這個問題

3.

我使用 gradient boosting classifier 的 model

資料處理方面，由於我發現 fnlwgt,capital-gain,capital-loss 都存在有離群值，所以我將 fmlwgt 的數值範圍限定在[0,800000]，然後將另外兩項 label encoder 其中原本為 0 的仍為零，原本大於 0 的變為 1，之後再對整組資料做標準化