

1. 此次在 training 的部分，首先將資料用 pandas 讀進後，針對非 01 項做 hot-encode 且對用 robustscaler()將資料 normalize，使得預測比較準確，之後便是先測試過 Naive_Bayes、DecisionTreeClassifier、linear regression 但是測試後發現準確率和要求差異過大，所以最後使用 catboostclassifier 作為 training model，在使用後，就達到兩個 line 的需求了
2. 基本上給的參數中比較重要的一點是，需要將一些原本並非 01 表示的數據變成 01 表示，像是 education、婚姻狀態，都可以藉由將 1~4 分別對應的類別，改變成為多項的 yes or no 的是非題，而基本上將這些參數改變成這樣的表達方式後，基本上參數都是被我所使用的。
3. 此次 feature 中，個人認為最重要的便是 pay bill 和 limit_bal 需要做前處理，我們可以知道一個人的 pay and bill 是和他的 limit 有絕對關係，所以我認為在這邊需要消除 limitbal 和 pay、bill 之間的關係，所以便決定 pay&bill 都要除以 limit_bal(舉例來說 15 萬上限的人不會借到 20 萬)所以其實一個人 pay and bill 是會被 Limit_bal 做限制。在直觀上來說，原本沒有對這邊做處理時，準確率只有 0.77(未達 simple line)但加入後便上升到 0.79(兩條線都過

了)

4. 此次使用 python 3.6 和 catboost 的原因，第一 python2.7 在 win10 上面安裝時會有 ascii 的 error 所以上網查了以後發現，3.6 板就沒這個問題，所以就使用了。

第二 catboost 是一個頗新的 opensource 的 machine learning method 裡面，是用 gradinet 提升的方式來做學習，那雖然他和 sklearn 的幾種學習方式做比較後，花的時間多蠻多的，但是準確率 (accuracy) 上升很多，那選用的原因，也是因為在前面五次的嘗試中，發現諸如 Naive_Bayes、DecisionTreeClassifier、linear regression 表現真的不好，所以在多方嘗試下最後才選擇使用 catboost