



國立台南大學資訊工程學系109級畢業專題 應用資料探勘演算法於信用卡客戶違約資 料的分析與預測

第三組

陳銘濡 蘇政穎

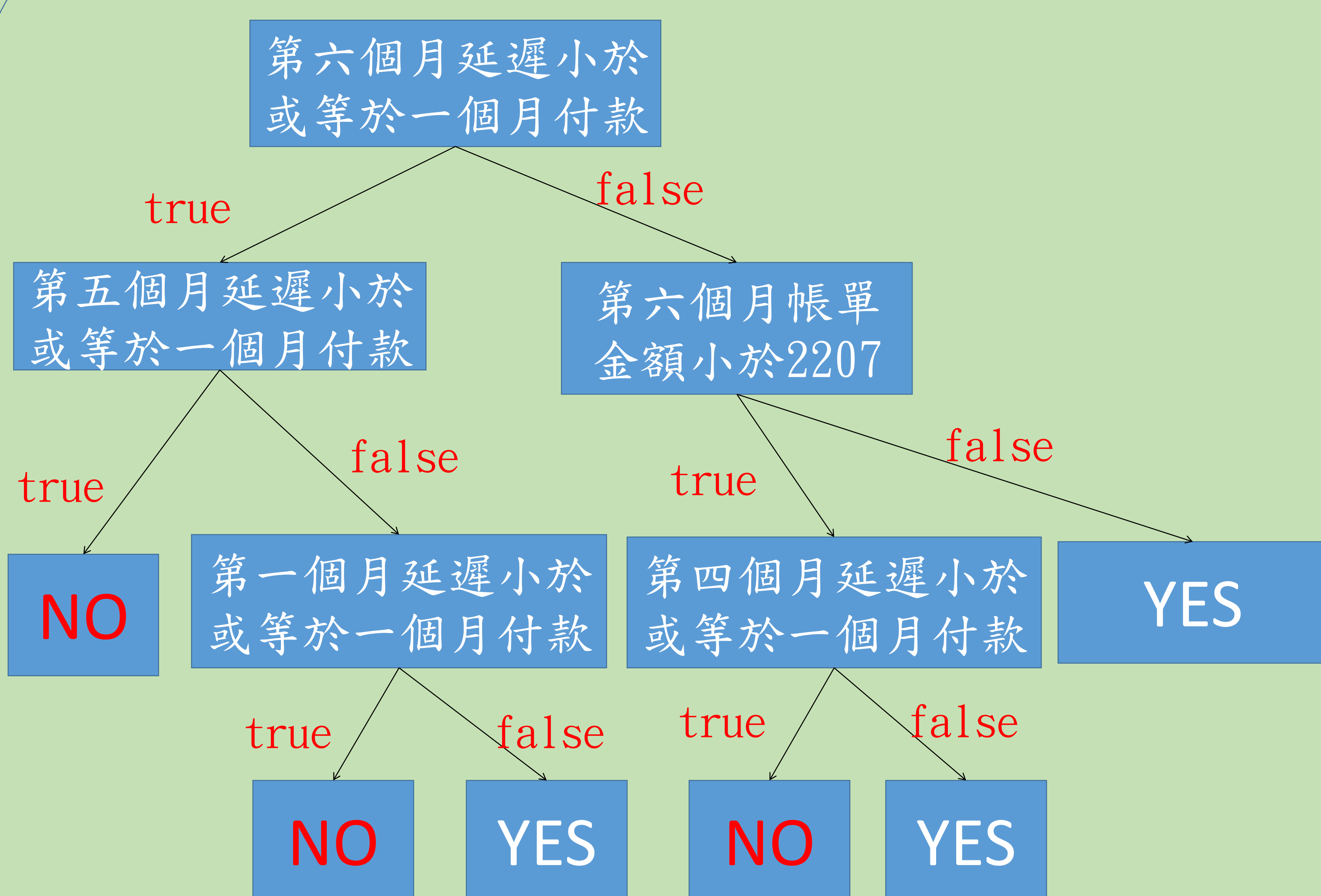
摘要

我們將運用從UCI資料平台得到的信用卡客戶默認支付資料集，我們希望對它進行分析並找出擁有什麼樣特徵的客戶會違約；我們利用Python語言，並運用Cart決策樹演算法、KNN演算法、貝氏演算法、ANN演算法 針對不同性別、年齡、教育程度、過去數月付款情況……等等來對客戶是否違約來進行分析，經過各項研究進行參數調整分析並得到最佳結果的演算法，同時找出資料的規則，藉以供相關人士作參考，提高他們的便利性。

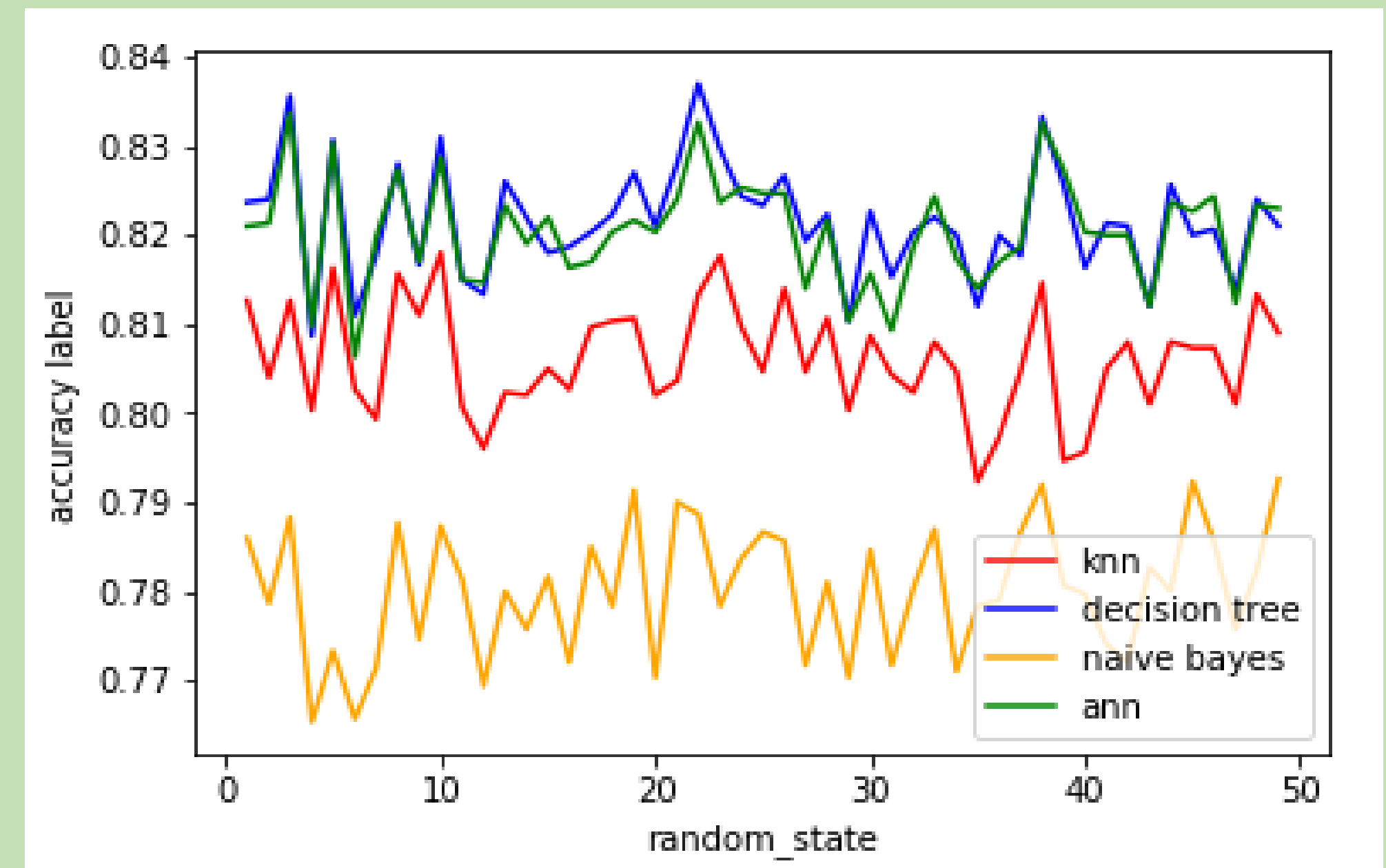
演算法介紹

1. KNN: 選出K個與測試資料最接近的訓練資料，進行投票，得到結果。
 2. BAYES: 根據測試資料的條件帶入訓練資料算出是否會違約的機率，機率較高的作為輸出。
 3. Decision Tree: 把訓練資料轉換成二元樹的模型，並將測試資料依據規則進行判斷，得出結果。
 4. ANN: 先建立神經元模型，再將訓練資料重複訓練，讓他變成像生物大腦中的突觸一樣，可以將信號傳輸到其他的神經元，再將測試資料放入，得到結果。
- Random state: 確保每次取樣的資料都會一樣，方便對於各演算法的執行時間與準確率做比對。

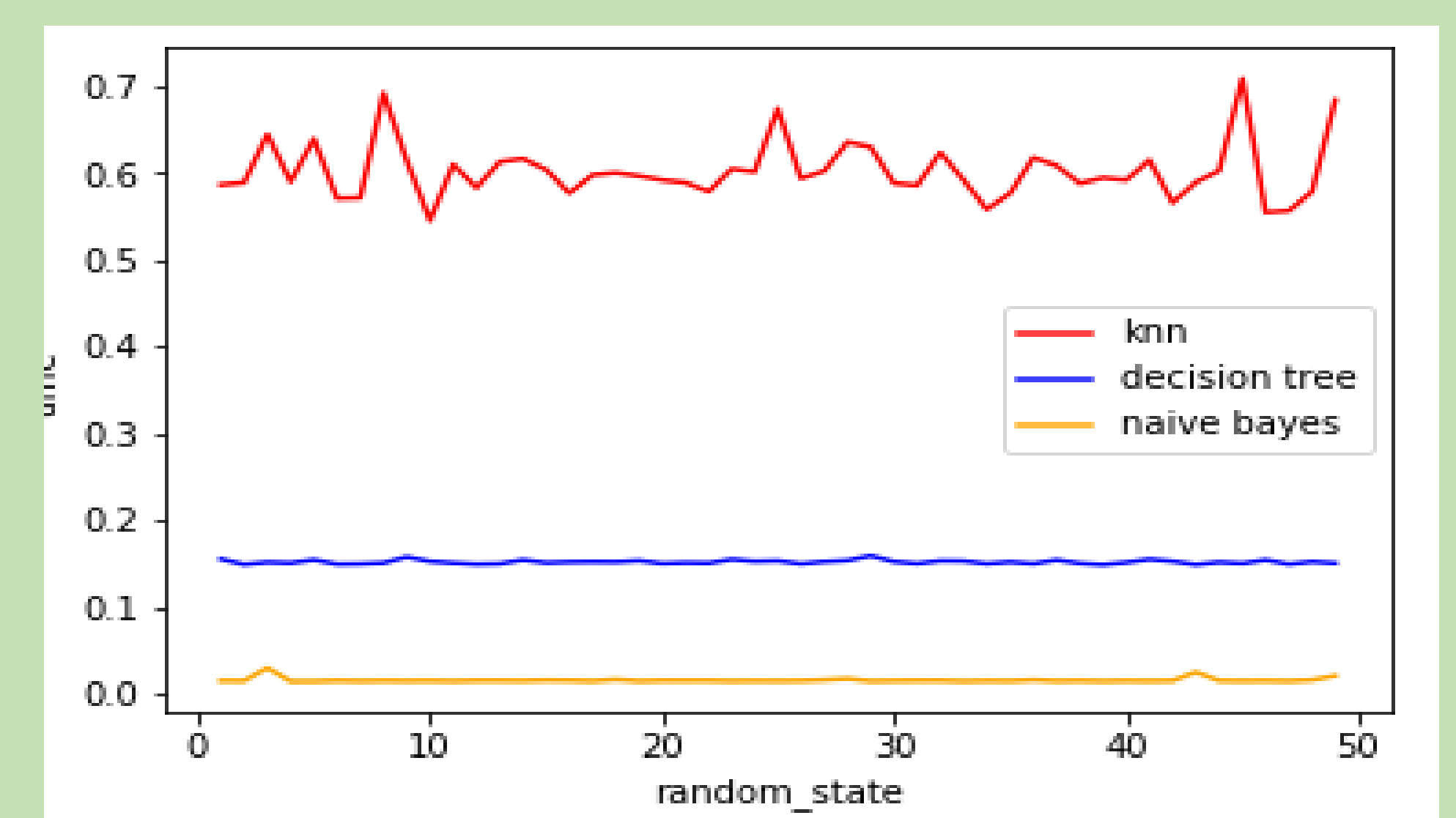
成果展示



圖(a)決策術規則，(延遲小於一個月付款即為準時付款)，yes代表違約，no則否



圖(b)四種演算法準確度比較



圖(c)三種演算法執行時間比較，ANN因為執行時間相對過長的關係，因此不列入比較

未來展望

在資料方面，雖然資料的筆數夠多，能夠有充分訓練樣本來做出模型，讓預測準確率有不錯的成果，但資料來源時間可能不夠新，無法充分表達當下社會的情況，也無法把通膨的元素考慮進去，因此未來希望有機會能取得更及時的資料，確切反映當下的情形，隨時代變動，也希望能有更佳的演算法或資料處理方式，或找出更佳的模型參數，讓訓練出來的模型分析能更加精確。