



P2P違約風險預測

羅吉斯迴歸模型與隨機森林模型之比較

指導教授：李美杏 教授 組員：統計三 吳啟玄、張逸昇、王薇淳、陳怡升、胡郁唯



金融與科技的結合，產生新的借貸方式：P2P 網路借貸。本研究旨在預測P2P 借貸的違約與否，使投資人可以做參考，進而降低違約風險。利用羅吉斯迴歸與隨機森林建模，找出個別模型中的重要變數，再相互比較兩模型之間的差異。最終希望能以APP的方式呈現，使投資人能更快速且方便的以此做為投資依據。

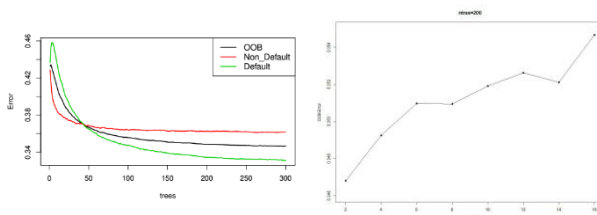
羅吉斯迴歸實證結果分析：

各模型的解釋變數為：

- 模型一：借款利率、借款人等級、貸款目的、借貸金額
- 模型二：借款利率、借款人等級、貸款目的、借貸金額、貸款週期、信用範圍最低值
- 模型三：全部變數

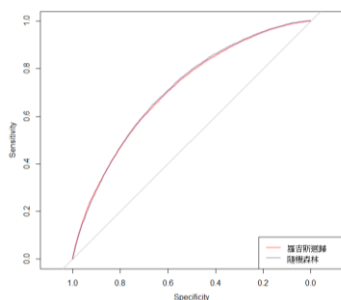
模型	準確度		敏感度		特異度	
	平均數	標準差	平均數	標準差	平均數	標準差
模型一	0.622	0.018	0.671	0.0017	0.607	0.0023
模型二	0.635	0.0015	0.655	0.0027	0.628	0.002
模型三	0.657	0.0016	0.648	0.0029	0.659	0.0023

隨機森林實證結果分析：



準確度		敏感度		特異度	
平均數	標準差	平均數	標準差	平均數	標準差
0.650	0.0019	0.669	0.0035	0.644	0.0028

比較模型差異：



比較項目	羅吉斯迴歸			隨機森林			P 值
	次數	平均數	標準差	次數	平均數	標準差	
準確度	10	0.657	0.0016	10	0.646	0.0023	< 0.0001
敏感度	10	0.648	0.0029	10	0.675	0.0033	< 0.0001
特異度	10	0.659	0.0023	10	0.640	0.0011	0.0009

結論：

1. 綜合指標外，我們選用羅吉斯迴歸模型，因為羅吉斯迴歸模型也較隨機森林模型快速許多
2. 當資料量大且解釋變數多時，解釋變數中的變異愈大且噪聲變量愈多時，羅吉斯迴歸之準確率會優於隨機森林

建議：

1. 同步資料，更能幫助投資人
2. 搭配經濟指標（例如：失業率、初次請領就業金人數）
3. 用不同的增強學習，提升準確率（例如：AdaBoost, GDBT, XGBOOST）

APP設計：

