

## **Peer review on group 12 by Wu Jiajun (20666111)**

### **1. Summary of the report**

The report uses gradient boosting decision tree to predict the clients' repayment abilities. It also shows the improvement of the performance by adding additional input data. The performance is measured by the area under the ROC curve. The private score is 0.74 using the original data while it is 0.75 when adding additional information.

### **2. strengths of the report**

It is nice to explain how to deal with the categorical data in the data process part. Also, this report not only shows the performance of the model using the main table data but also compares it with the one adding additional data.

### **3. weaknesses of the report**

The data description part could be more detailed. For instance, how many input variables are included in the model? What is the sample size of the train and test data? That would make the data description part be more straight forward. Also, for classification problem, it is better to include a confusion matrix or other performance score such as specificity. In addition, there is lack of interpretation of the performance to the real world situation.

### **4. Evaluation on Clarity and quality of writing: 5**

### **5. Evaluation on Technical Quality: 3**

### **6. Overall rating: 3**

### **7. Confidence on your assessment: 2**