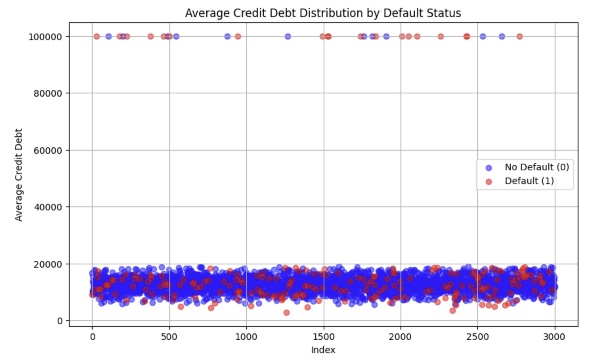
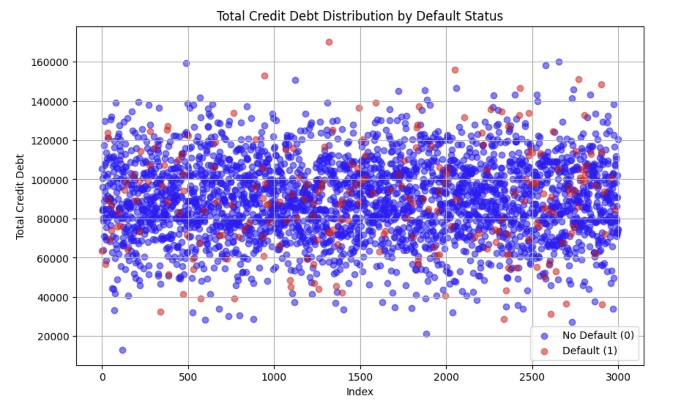
【Final Result】

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Base model | Improve 1 | Improve 2 | Improve 3 |
| model | **logistic regression** | **Piecewise Logistic Regression** | **Random Forest** | **Neural Networks** |
| Accuracy | 0.937 | 0.9386666666666666 | 0.941 | 0.9353333333333333 |
| ROC AUC Score | 0.7912848053236823 | 0.7912848053236823 | 0.8508730112401819 | 0.8367595457228287 |
| Bad Debt Amount base on tot\_credit\_debt | 14845364.21 | 13063494.829999998 | 6323830.040000001 | 10862684.170000002 |
| Bad Debt Amount base on avg\_card\_debt | 3224567.8899999997 | 1824581.89 | 1051177.2 | 2648471.33 |
| Confusion Matrix | [[2755 23]  [ 166 56]] | [[2746 32]  [ 152 70]] | [[2763 15]  [ 162 60]] | [[2730 48]  [ 146 76]] |
| Rate of being predicted as loanable for high avg\_card\_debt user | 0.80 | 0.80 | 0.93 | 0.53 |
| Rate of being predicted as loanable for low avg\_card\_debt user | 0.98 | 0.98 | 0.98 | 0.96 |
| Rate load to user has account | 0.98 | 0.98 | 0.98 | 0.96 |
| Rate load to user has no account | 0.97 | 0.97 | 0.97 | 0.95 |

【Important Process】

1. Visualize the distribution of Default\_ind



We can see there is a higher risk on "user who has a high avg\_card\_debt", and since they load more money, they have higher risk.

1. If we are trying to avoid risk that can’t get money back, logically we should be careful on loading big amount of money.

And Random Forest did the best job on that.

【Describe how you would use it to make decisions on future credit card applications.】

By looking at the last two row in Final Result Table, we can see that if bank use my model, they will have more welling to load money to those who do NOT usually load a lot of money. If a user always load a lot of money, bank might won’t load again.

And Random Forest model did shows a super high importance on avg\_card\_debt.

