IOE 373 Final Project

Authors: Eric Xu , Frederick Kusumo , Joseph Dib

Introduction

Our group was tasked with modeling customer behavior from Universal Bank's last campaign, which had a 9% success rate. We wanted to analyze this behavior to determine which factors contribute most to the likelihood of a customer accepting a personal loan. To do so, we created a heatmap and a logistic regression model on the data.

Analysis

```
Accuracy: 95.44%
```

Fig. 1 Model Accuracy

	precision	recall	f1-score	support
0.0	0.96	0.99	0.98	2244
1.0	0.88	0.66	0.75	256
accuracy			0.96	2500
macro avg	0.92	0.82	0.87	2500
weighted avg	0.95	0.96	0.95	2500

Fig. 2 Classification Report

```
[[2221   23]
 [  87  169]]
```

Fig. 3 Confusion Matrix

Evaluating Model

The logistic regression model we created has an accuracy of 95.44%. This can be seen from Figure 1. The accuracy represents the overall correct prediction made over the total prediction of the model.

Figure 2 and Figure 3 show the classification report and confusion matrix of the model when validating using the test set.

Based on the classification report, it can be seen that the precision of predicting a person getting a personal loan (1) and not (0) are 0.88 and 0.96, which is pretty high. But the values of (1) decreases in the recall, and f1-score, as it can be seen based on the support, there are generally lesser values of 1 than 0, meaning that fewer people are getting personal loan than the ones getting. This then made the model hard to catch positive predictions as can be seen from the values of recall and f1-score.

From the confusion matrix, it can be seen that the value of True Positive which is in the top left quadrant, and the value of True Negative which is in the bottom right quadrant are higher than the values of False Positive which can be seen from the top right quadrant and False Negative that can be seen from the bottom left quadrant. The accuracy would be 2390/2500 which brings us to 0.956 accuracy which means that our model predicted the well.

Significant Factors

Figure 4 shows the coefficient values of all of the factors based on our model when being fit to our train set. It can be seen that CD Account has the highest coefficient value of 3.12 followed by Family and Professional Education Level. The positive coefficient values mean that if a person has these factors then it would be a higher chance for them to get personal loans.

	0
CDAccount	3.115922
Family	0.652652
EducProf	0.212800
CCAvg	0.192729
EducGrad	0.188051
Experience	0.096157
Income	0.050755
Mortgage	0.000874
Age	-0.082807
OnlineAccount	-0.611556
SecAccount	-0.620531
CreditCard	-0.722755
EducUgrad	-2.858425

Fig.4 Coeff df

	0
CDAccount	22.554211
Family	1.920627
EducProf	1.237137
CCAvg	1.212554
EducGrad	1.206895
Experience	1.100932
Income	1.052065
Mortgage	1.000874
Age	0.920529
OnlineAccount	0.542506
SecAccount	0.537659
CreditCard	0.485413
EducUgrad	0.057359

Fig.5 Odds Ratio

The lowest coefficient can be seen to be from Undergraduate students (EducUgrad), followed by Credit Card holders and people with Securities Accounts. The negative coefficient generally means that if a person has these factors, there is a lower chance of them getting or needing a loan.

Figure 5 shows the odds ratio of the factors based on our model that was run with our train data. The odds ratio > 1 suggests that it is more likely that they get a loan with these factors; a CD Account, higher family number, Professional Education, higher spending on average using credit card, Graduate level education, more years of professional experience, higher income, and mortgage they are more likely to get loans. While odds ratio < 1, like age, Online Account, Securities Account, Credit Card account, and Undergraduate education level would mean they are more likely not to get a personal loan.

Conclusion

In conclusion, based on our findings of the coefficient values of each factor as well as the odds ratio on our model when being fit to our train data set, it can be said that the top five most important factors are CD Account, Family, Professional Education, average spending with a credit card, and Graduate level education. These factors would be best for Universal Bank to take into account as these factors have high positive coefficient values and odds ratio higher than 1, especially CD Accounts so Universal Bank can easily target people who have these factors to get a higher chance of them wanting to get loans from them. Factors that are least important would be Undergraduate level education as the coefficient value is in the negative extreme and pretty low in general, as well as having an odds ratio dramatically lower than the other factors, which means that in general undergraduates won't take loans.