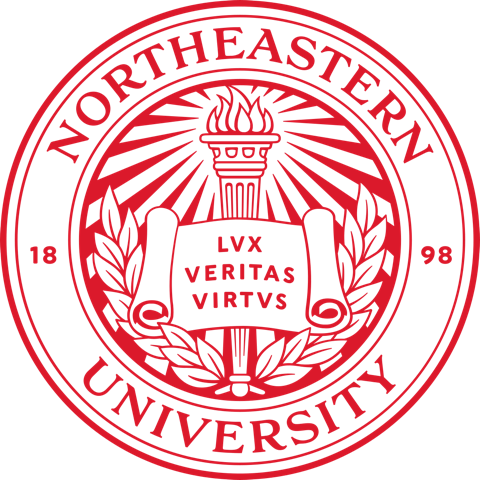
**Module 3 Midweek Project**

**Loan Approval**

Sunil Raj Thota



College of Professional Studies, Northeastern University ALY 6020 – Predictive Analytics

Prof. Justin Grosz

November 18, 2021

**Introduction:**

My Bank's management intends to look for ways to convert its liability customers to personal loan customers in this case (while retaining them as depositors). Last year, the bank ran a campaign targeting liability customers that resulted in a solid conversion rate of over 9%. This has prompted the retail marketing department to create campaigns with better target marketing to raise the success rate while spending as little money as possible. My bank has received a list of loans and will not be able to approve or reject them until the following day.

**Analysis:**

I've installed all of the required libraries and imported the data. The panda's library is being used. Data about 5000 clients is contained in the file Bank.xls. The dataset is rectangular (5000, 14). The dataset contains no missing or NaN values. Although the dataset contains both category and numerical features, all categorical data is represented by numbers. We can observe that the "Age" feature is relatively evenly distributed, with the majority of clients being between the ages of 30 and 60. We can also see that the median is the same as the mean. The distributions for "Income," "CCAvg," and "Mortgage" are all positively skewed. In the case of "Income," the mean is higher than the median. We can also confirm that the majority of the consumers earn between 45 and 55 thousand dollars. For "CCAvg," the majority of clients spend less than 2.5K, with an average of 0-10K. For "Mortage," we can observe that about 70% of consumers have a mortgage value of less than 40K, with a maximum value of 635K. The distribution of "Family" and "Education" is balanced.

The dataset contains information regarding the bank's storage of the customer's personal information. It also includes information on bank terms such as prior bank loans, securities accounts, and check deposit accounts, among others. We'll create a model to predict if a loan will be accepted or rejected. We'll do so by looking at the Experience column, which details the customer's professional background. This factor is highly crucial when deciding whether or not to take a loan because clients with more professional experience have more income, securities, mortgage value, and so on. The dependent variable will be the acceptance of personal loans in the previous campaign.

**Data Modelling:**

When there are several explanatory variables, logistic regression is used to calculate the odds ratio. With the exception that the response variable is binomial, the approach is quite similar to multiple linear regression. The impact of each variable on the odds ratio of the observed event of interest is the end outcome.

1. **What were the three most significant variables?**

CD Account, Education, and Family are the three most important elements. The significant variable's coefficient is as follows: (See Appendix for Figure 1) p-value of car width is 0.000. education p-value as 0.000, The P-value of the family is 0.000, which are the significant elements in overall performance.

1. **Of those three, which had the most negative influence on loan acceptance?**

Because it is not the most important concern among the three categories, the family has the most negative impact on loan acceptance. The combination of a CD account and education has a significant favorable influence.

1. **How accurate was the model overall and what was the precision rate?**

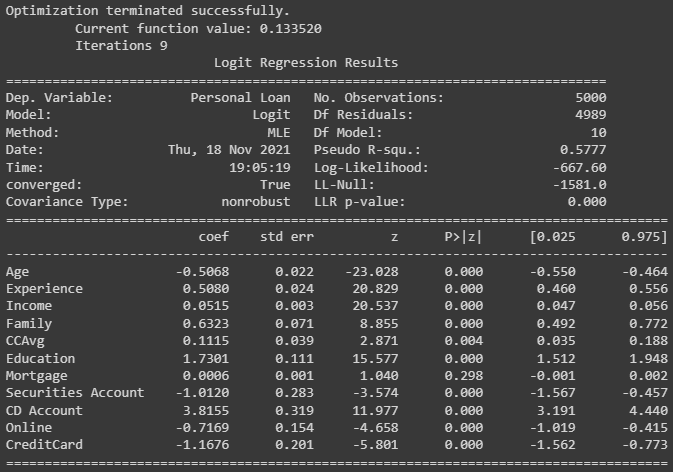
The model's accuracy is 95 percent, with an 81 percent precision rate for accepting loans. This suggests that 81% of our results are relevant, which is a very high percentage. (see Figure 2 in Appendix).

**Conclusion:**

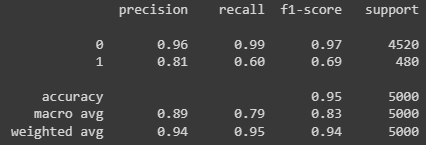
Without taking experience into account, the model's accuracy for loan acceptance is 95% with an 81 percent precision rate. The major variables to examine are the CD account, education, and family size.

**Appendix:**

**Figure 1: Logit Regression Results**

****

**Figure 2: Classification Report Results**

****

**Figure 3: Confusion Matrix Results**

****