Credit Risk Classification Analysis

Parker Breaux, John Dennis, Ryan McElvain

University of Colorado - Boulder

Applied Mathematics

COLLEGE OF ARTS AND SCIENCES

Introduction

Evaluating whether a potential customer is at risk of defaulting on a loan is critical for banks and other credit lending institutions when making lending decisions. Estimating credit risks will benefit the institution as it's worse to label a person as good when in fact they're bad than to label a customer as higher risk or bad when they are good. Through multiple logistic regression, a model can be established that takes predictor variables of both categorical and numerical nature, such as the Age of the loan applicant, the Amount requested, Credit history, etc. This will be used to predict whether a customer will be flagged as a credit risk (Credit Risk) or will be able to repay (Non-Credit Risk). This subject is important in real-world financial problems that carry high stakes and providing insights that have potential to be highly valuable to banks and lending institutions for maintaining a sustainable business model.

Methods

We got the data from the UCI Machine Learning Repository, which is a collection of datasets for the purpose of empirical research. The Machine Learning Repository is hosted by the University of California, Irvine for the purpose of advancing the fields of algorithm development and machine learning. The dataset was created by Dr. Hans. Hoffman, a professor of Statistics and Econometrics at the University of Hamburg. The wrangled data set contains 1000 observations and 8 predictor variables. We reformatted some of the categories for our ordinal predictors to simplify the logistic regression process. One observation in the set represents a singular loan application, and we plan to use a total of 4 numerical and 4 categorical predictors in our model. We could not verify the exact bank that the data comes from, so we had to generalize how someone is deemed a credit risk. We used a forward stepwise selection process to obtain the best model based on AIC, which had 5 predictor variables: 'Age', 'Duration', 'Credit_History', 'Purpose', and 'Housing'.

Variable	Units	Min	Median	Max
Credit Risk	1: Bad, 0:Good	0: Good	N/A	1: Bad
Amount	DM	250	2319.5	18,424
Duration (Of the Loan)	Months	4	18	72
Age (Loan Applicant)	Years	19	33	75

Exploratory Results

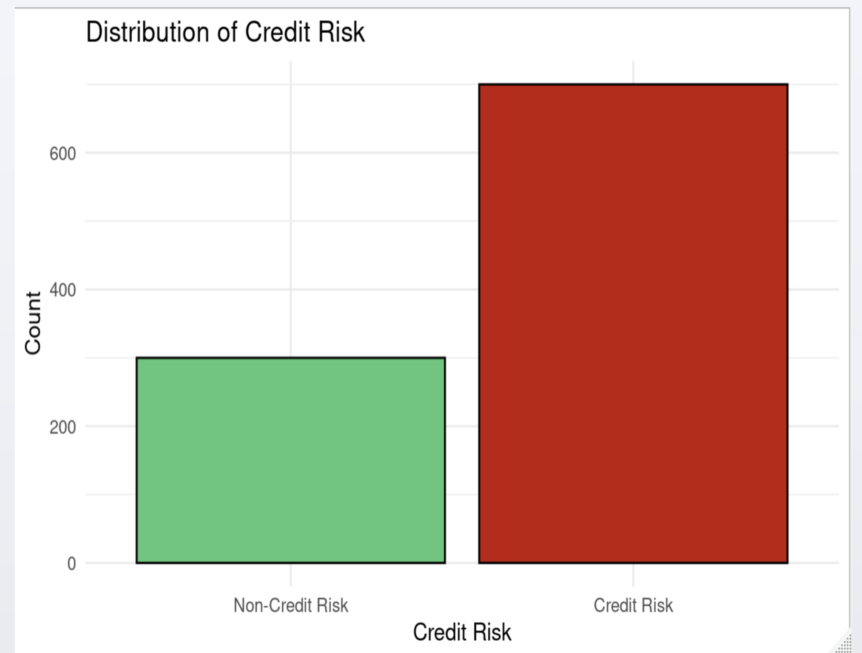

Figure 1

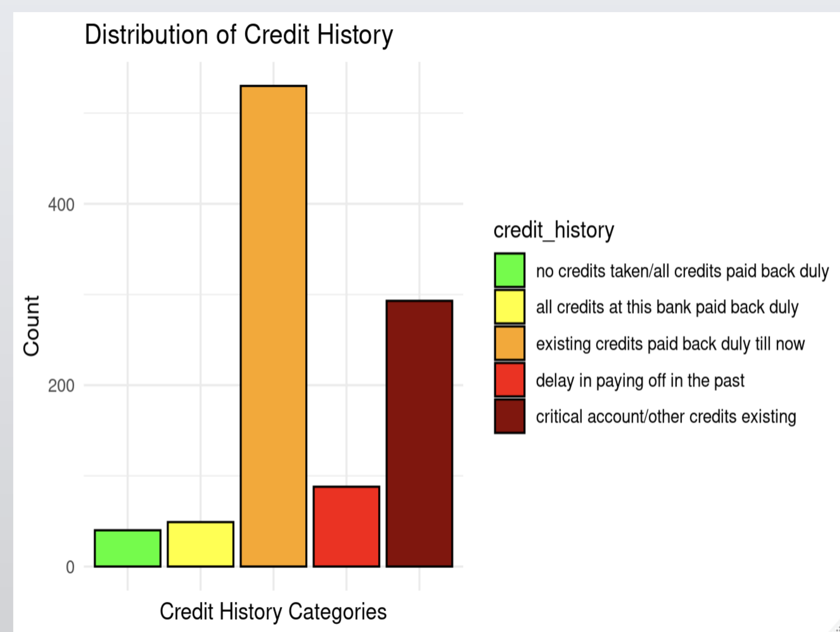

Figure 3

- (Figure 1) Most applicants were labeled as a credit risk. A bank would consider most loan applicants as a credit risk
- (Figure 2) As the applicant's credit history worsened, they were more often labeled as a credit risk. It is intuitive that a worse history would suggest the applicant being a risk.
- (Figure 3)The most common credit history for applicants was 3 and 5
- (Figure 4) Applicants with more savings were more often labeled a credit risk (A savings category of 1 is the best). Most applicants in every savings category were labeled as a credit risk.
- A good credit history was more important than savings in being deemed a non-credit risk

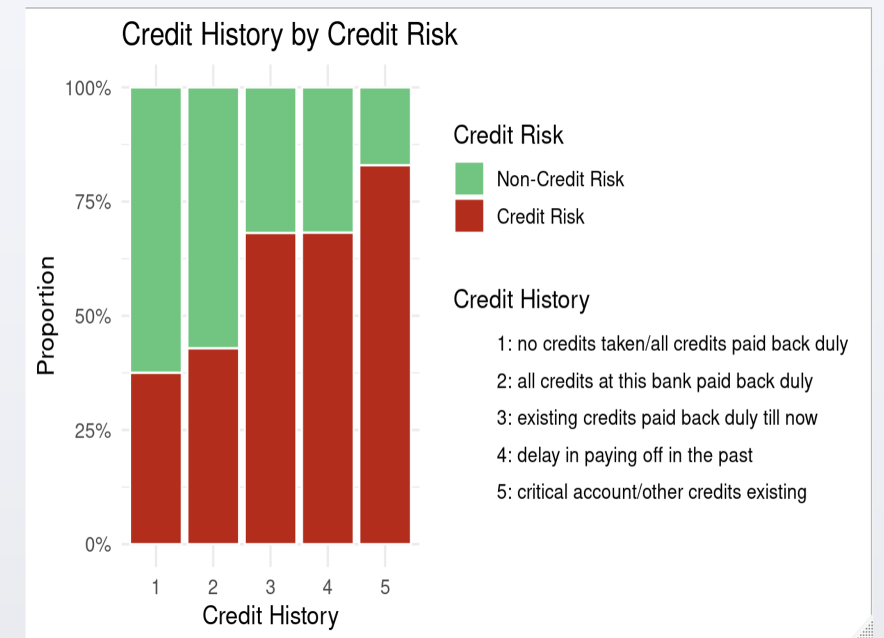

Figure 2

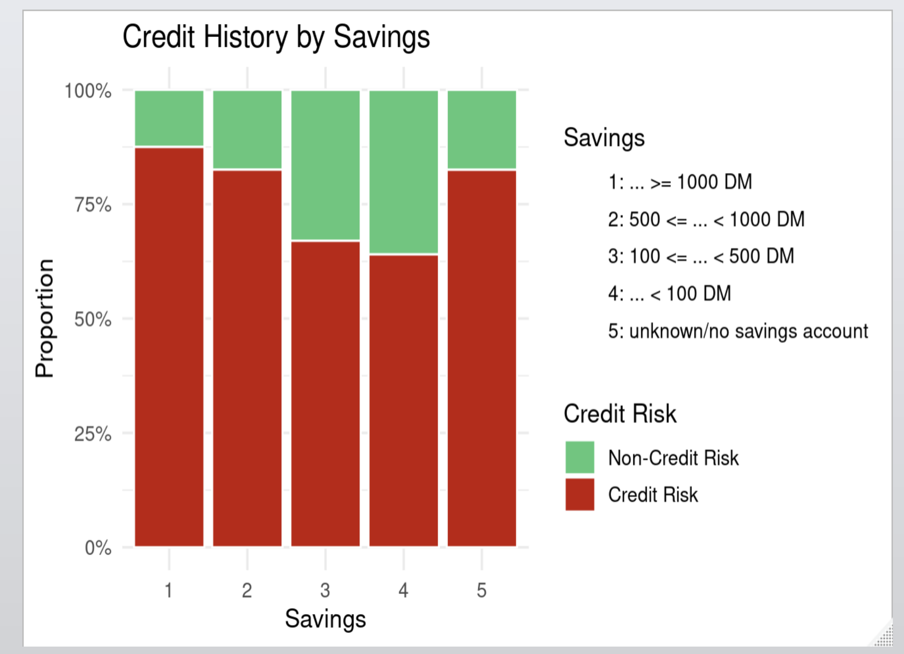

Figure 4

Predictive Results

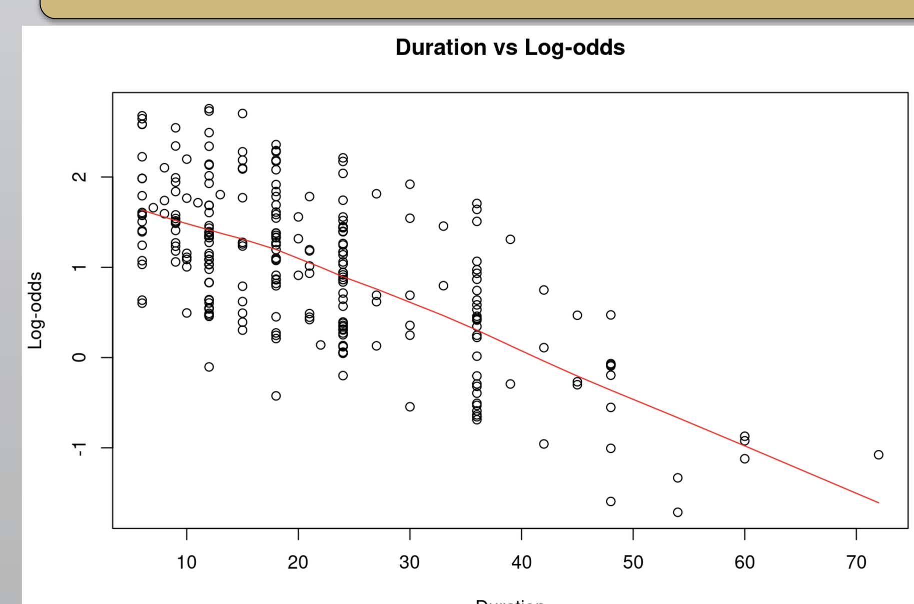

Figure 5

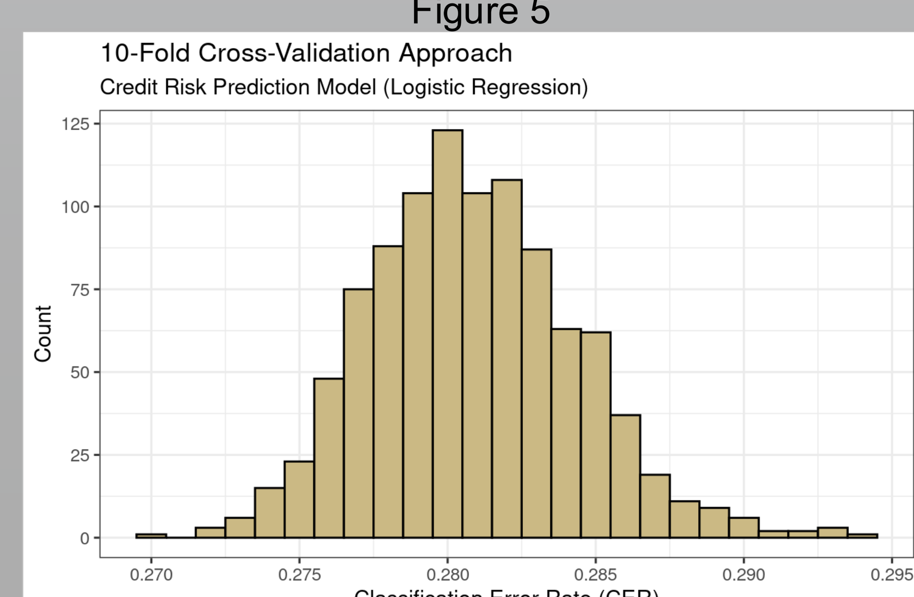

Figure 7

- (Figure 5) Duration against log-odds is roughly linearity, so linearity in the log-odds is satisfied for 'Duration'.
- (Figure 6) Adding predictors past four yields minimal further reduction in the CER; we chose 5 predictors, but started with 8 originally
- (Figure 7) We expect our model to misclassify new loan applicants around 28% of the time
- (Figure 8) A low FNR is most desirable in relation to the task at hand, so a classification threshold of 0.5 and below would be the most desirable: it is worse to label someone as a non-credit risk, when they are in fact a credit risk.

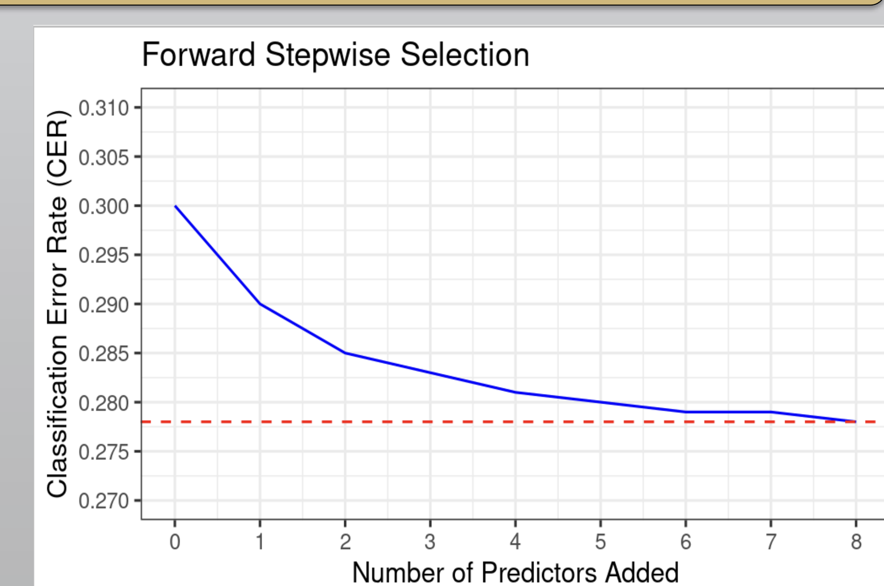

Figure 6

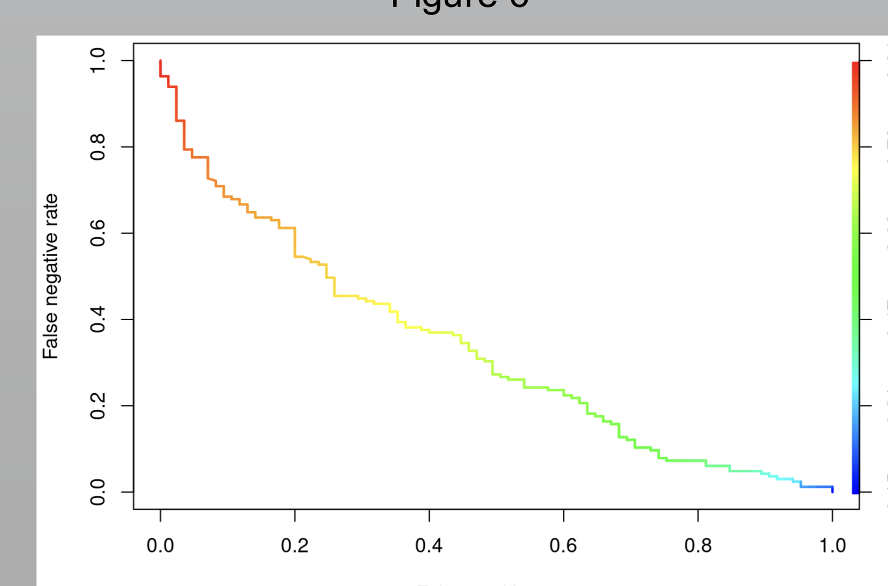

Figure 8

Discussion

To understand the comparison of our model to the null, we observe the differences within a few criteria: our model's AIC, which is penalized for the additional 5 variables it includes, outperforms the null which is basically its deviance due to having no variables. Without the penalty, the fitted model will naturally have a lower deviance than that of the null model, because adding predictor variables can only make the deviance decrease or remain constant.

Metric	Null Model	Fitted Model
AIC	1223.7	1125.2
Deviance	1221.7	1105.2
Log-Likelihood P-Value	N/A	0
Cross-Validated Error (5)	0.3	0.283
FNR (0.5)	0	0.049
FNR (0.75)	1	0.543

The Log-Likelihood P-Value from the likelihood ratio test is the probability, under the null model, of observing a log-likelihood at least as extreme as the one from the fitted model. Naturally for the null it is N/A as it's the base line and for the fitted model is zero which suggests that the observed reduction in deviance is too great to be attributed to chance alone. This suggests there is strong statistical evidence that at least one predictor is significantly associated with if someone was deemed a credit risk. Lastly, to analyze CV error, the average CER performed by 5-fold cross validation was 0.283. The null on the other hand will misclassify roughly 30% of the time or 0.3. Although not a tremendous difference, the model does outperform the Null in CER. Figures aside, the social and ethical concerns to consider are how the bank places value in maximizing profit for a given level of risk, as well as balancing who they might lend to. Since false negatives are more costly for a bank, a threshold such as 0.5 that reduces the FNR is preferred. Banks may wish to utilize a classification threshold below 0.5 to reduce their FNR, even if it meant a higher a FPR: using our model, a bank will need less evidence to label someone as a credit risk and to deny giving them a loan.

Conclusions

Throughout this project, we built a logistic regression model on the German Credit dataset to classify applicants as a credit risk or not. Our model outperformed the null model and helps banks choose a threshold to minimize risk and net losses. As a next step, we could develop a function that takes in features like a lender's size, total loan volume, and risk tolerance to automatically choose an optimal rejection threshold, so our model can be applied to lenders of many sizes