Credit One Final Report

Bailey Keel

2/19/2023

What prompted this investigation into Credit One's defaulting model was due to an increase in customer default rates. This is not ideal for Credit One as they are the business approving customers for loans and assigning the amount. So, from this they could potentially start to lose customers. They tasked us find a better way to understand how much credit to allow someone to use, or at the very least, if someone should be approved or not. We approached this with Data Science and used the following steps: Cleaning and Pre-Processing, EDA, Regression, and Model Evaluation.

First off, to start this undertaking the model needed to be cleaned. This meant the data was to be made presentable and nice for the analysis done later. The duplicates in the data were removed along with the null/non complete data entries. Finally, the data types were converted to the correct types, for example all the data was considered objects. But we needed most of them to be of int type. Once all of this was completed the next step is to start analyzing the data.

For the analysis portion we looked for any correlations between the data entries given. A Heat Map was made to help better visualize the relationships and can be seen in Figure 1. Some of the stronger correlations found were between payment history and whether they will default, their bill amount and their limit balance, graduate school and limit balance, and their limit balance and whether they will default.

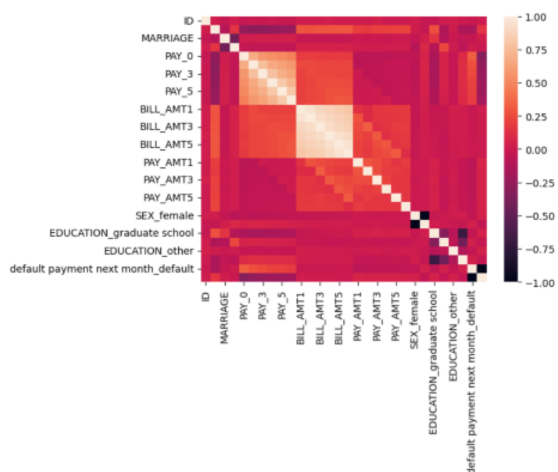

Figure 1

Once realizing there was a good correlation between defaulting and payment history a table/graph were made, Figure 2, to compare how late a payment is to how likely the customer will default by being in that payment zone. From looking at the results, the most likely people to default are 7 months late on their payment, followed by 3 months, 2 months, and 4 months. What was surprising though is the fact that people with 5/6/8-month late payments are less likely to default then those with 7/3/2-month late payments.

Payment History	Percentage of Default (%)
(7) 7 Months late	77.78
(3) 3 Months late	75.78
(2) 2 Months late	69.14
(4) 4 Months late	68.42
(8) 8 Months late	57.89
(6) 6 Months late	54.54
(5) 5 Months late	50.00
(1) 1 Month late	33.95
(-1) Paid in Full	16.78
(-2) No Consumption	13.23
(0) Revolving Credit	12.81

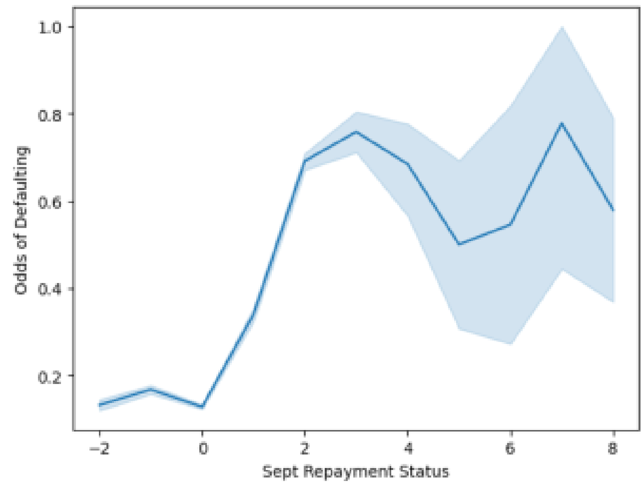

Figure 2

Once we were finished analyzing the data for trends, we could move onto the machine learning portion. When tackling this step, we used various methods of machine learning that included Regression with limit Balance as dependent, Classification with discretized Limit Balance as the dependent, and Classification with defaulting as the dependent. For the Regression models we found Random Forest Regressor to be the most accurate model and for Classification, Random Forest Classifier was the most accurate model. The accuracy result for each model ran can be found below in Figure 3. One thing that is interesting is that when discretizing the age into decades (0-9, 10-19, 20-29, etc) the accuracy of the model slightly decreases. This was interesting because typically grouping data results in more accurate models since there are less unique values. Also, it should be noted that while Limit Balance being grouped by $500,000 was included in this report, it will not be considered as that is way too large of a bin with the max Limit Balance being $1M.

Method	Accuracy (%)	Accuracy (%) with Age Discretized by Decade
Regression with Limit Balance as Dependent	46.8	45.4
Classification with Limit Balance as Dependent (Grouped by $100,000 Intervals)	46.05	45.38
Classification with Limit Balance as Dependent (Grouped by $200,000 Intervals)	48	47.59
Classification with Limit Balance as Dependent (Grouped by $250,000 Intervals)	43.98	43.58
Classification with Limit Balance as Dependent (Grouped by $500,000 Intervals)	87.15	86.42
Classification with Defaulting as Dependent	81.58	81.33

Figure 3

Now that all the data has been analyzed we can look at the questions proposed to us by Credit One. Can we approve customers with high certainty? How do you ensure that customers can/will pay their loans? In regard to ensuring a customer can/will pay their loans, there is no way for us to ensure that. Customers spending habits can not be controlled and people cannot be controlled either, the best you can do as a company is to loan money only to the people that have low odds of defaulting and hope for the best. This now leads us to the question of can customers be approved with high certainty. The answer is yes, when looking at the data from Figure 3, we have a model that can predict whether a customer will default or not with 81.58% accuracy. Though when it comes to predicting a limit balance that should be assigned to the customer, we are only about 46% sure. So, after running all the data, the recommendations we can make to Credit One are to use our models to evaluate whether to approve a customer for a loan in the first place, but do not use these models to predict limit balance to assign as they are not as accurate based on the data given for evaluation.