**Specify the problem**

An US bank would like to improve its financial performance with regards to approval of new loans. There are probably too many people that defaulted on the loan and might cost the bank money in certain scenarios. but not every defaulter is a loss.

**What can we look at in a person to know if he/she would most likely be charged off**

To access a person’s charged off probability, we can consider these factor when issuing them loans.

**Summary Plot**

Most probably be charged off

Graphical user interface

Description automatically generated with low confidence

In the summary plot, the one at the top are what is considered the more important features. Features are sorted by the sum of the SHAP value magnitudes across all samples.

1. Int Rate: The higher the interest rate is, the higher the probability of one being charged off. (Red color to the right means high, therefore high chance)



1. Grade → LC assigned loan grade

The higher the grade is (A-1, G-7), the higher the probability of one being charged off.



1. The lower his mbbXnart (max\_bal\_bc \* num\_actv\_rev\_tl), more likely he is to be charged off
   * 1. max\_bal\_bc: Maximum current balance owed on all revolving accounts
     2. num\_actv\_rev\_tl: Number of currently active revolving trades



Graphical user interface, text

Description automatically generated

1. Installment \* term: The higher it is, the more likely he is to be charged off. Make sense because the higher ur installment is and when your term is 36/60, it adds up.



1. acbOloanamnt → The lower it is, the more likely he is to be charged off (avg\_cur\_bal / loan\_amnt)
   * 1. avg\_cur\_bal: Average current balance of all accounts
     2. loan\_amnt: The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.



1. acc\_open\_past\_24mths → The more a person open an account, the more likely they are to be charged off.



1. Loan amount → The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
   1. The more they loan, the more likely they are to be charged off (probably inability to pay off the large sum of money, also relates to installment\*term, if they borrow more, they probably have a larger installment to pay with their terms given)
   2. 

Graphical user interface, text, application

Description automatically generated

1. open\_rv\_24m: Number of revolving trades opened in past 24 months
   1. The lower the number of revoling trades opened, the more likely they are to be charged off
   2. The higher the number of revolving trades opened, the more likely they are to full pay
   3. 
2. Total Account Over Open Account (credit line ratio)
   1. Total Account: The total number of credit lines currently in the borrower's credit file
   2. Open Account: The number of open credit lines in the borrower's credit file.
   3. 
3. DTI → A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower’s self-reported monthly income.
   1. The higher the dent and the lower the income, the higher the ratio (eg. 4D(debt)/1Inc = 4:1, they may not have the ability to pay compared to someone where the ratio is 1D/1Inc = 1:1
   2. 
4. Inq\_last\_12m → Number of credit inquiries in past 12 months
   1. Makes sense since the more they call to ask (assuming they dont ask redundant questions like how’s your day etc, but would probably call to extend the installment by a month), they will probably be charged off due to inability to pay.
   2. 
5. Annual\_inc → The self-reported annual income provided by the borrower during registration.
   1. Makes sense as those with low income may not be able to repay the installment, which leads to them being charged off
   2. 

**Force Plot**

Graphical user interface, Word

Description automatically generated

The red are what we call the positive push, and the blue the negative push.

In this plot, you can see that the grade and int\_rate is pushing the prediction for this particular row, but max\_bal\_bc and open-rv\_24m is pushing it back. (but you can’t really say that these predictors are bad, as it might push down when the person is NOT A defaulter)

This actually could be seen from the **Summary Plot** too, where grade and int\_rate played a huge part in the model performance, which in this case was to detect charged off.

**Waterfall Plot**

Chart

Description automatically generated

From this waterfall plot, you can see how are the predictors doing in the model and row. I prefer this chart than force plot as it tells you what predictors are bringing the results down/up. In this case, we see that max\_bal\_bc is brining the predictions down, same with open\_rv\_24m as specificed in the force plot. However, we see that int\_rate, acbOloanamnt,grade, avg\_cur\_bal and installment\*term is pushing the model performance up. In case you are wondering, the model baseline prediction was at 0.22, in which after all the predictors have pushed up/down, it ended up with 0.135. Which in this case, it was saying the probability of the person being charged off is 0.135, which means he probably will full pay off the loan. Which in this case, he did fully pay off. (Note that my value have been rounded off in the diagram ontop)

