C2T3 Writeup: Credit One Final Submission

E. Jarrett (FT cohort)

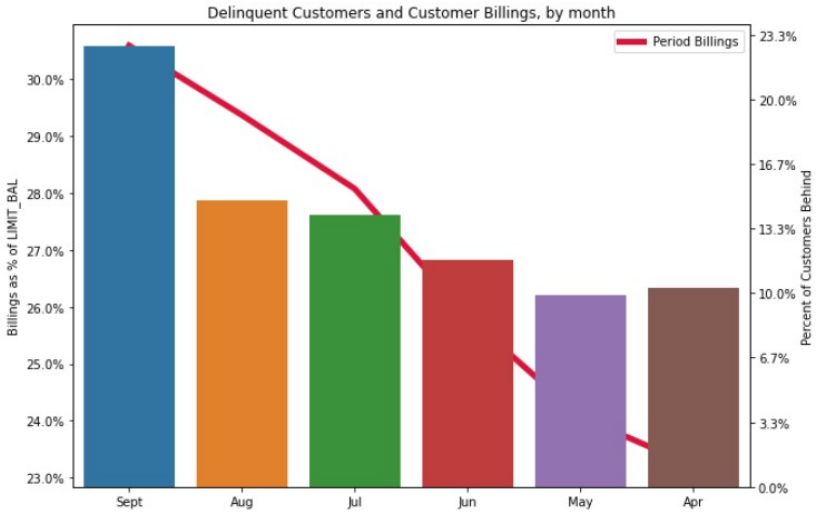
Project Intro: Credit One ('CO') customers (debtors) have begun defaulting at higher rates, posing significant risk to the financial stability/credibility of its credit-scoring business

Goals:

1. Identify individuals most likely to default ("classification")
2. Determine optimal credit limit for new (approved) loan customers ("regression")
3. Identify any demographic or behavioral factors highly important for 1 & 2

Problem Background:

- The financial picture of CO, and it customers ,worsened significantly during the April – September time period; customer billings grew 32%, but payments received by CO have increased only 8.5%

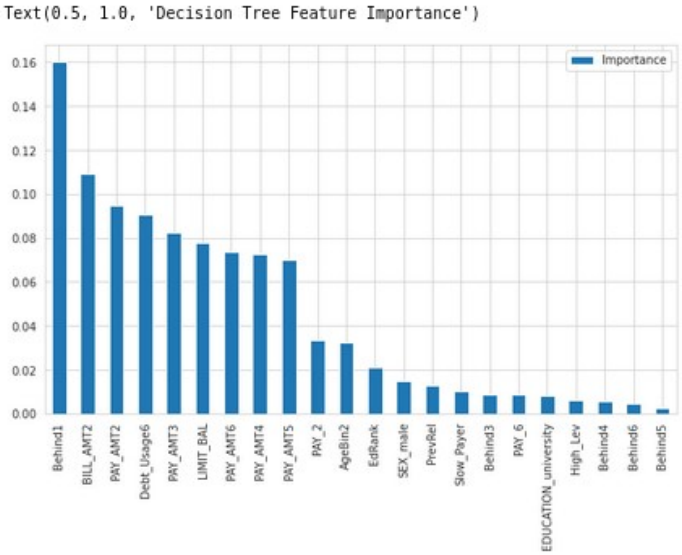


Topline Business Findings:

1. *Customer default can be forecast with high accuracy – models scored between 75% to 80% depending on the specific algorithm used.* These weighted precision scores however mask an important detail -- <u>every model performed significantly better at predicting non-default than default,</u> despite efforts to correct dataset imbalance through model parameters.

2. *Customer credit limit can also be predicted quite well, with top model 'Random Forest Regression' achieving ~ 71% accuracy.* While almost every classifier was similarly effective, Random Forest was bar far the clear winner for regression, almost 50% more accurate than runner-up 'Bayesian Ridge'.

3. *Demographic factors (age, sex, relationship/living status, education background) are NOT important features for either default classification or credit limit* regression. In fact, they typically rate as the <u>least</u> important, far behind spending/payment history.



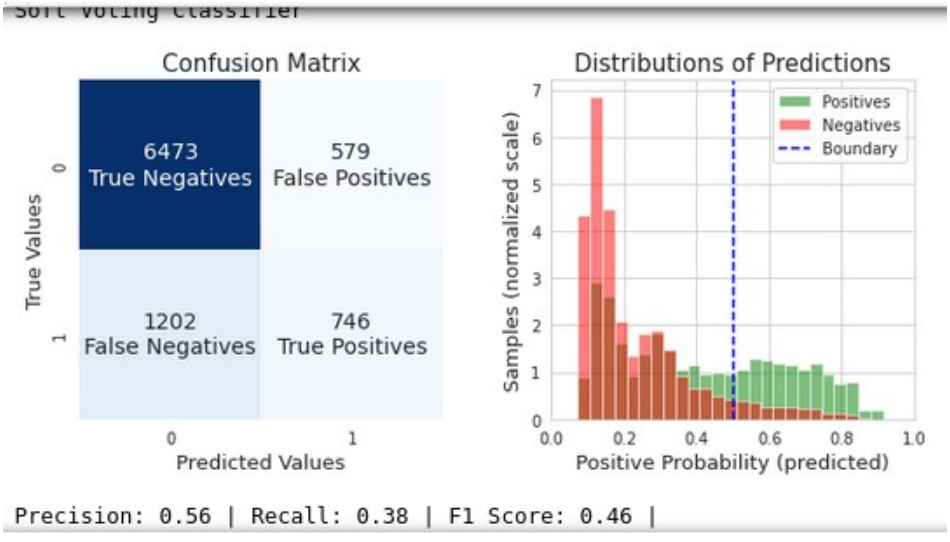
With minimal correlation to default, analyst's initial hypothesis (C2T1) that education and relationship/living status were good proxies for 'responsible mindset', thus affecting default appears disproven. It turns out the best indicator of a borrower's fiscal responsibility is *demonstrated directly* by prior action (i.e., stable and timely payment history).

4. *Model construction/outputs pass basic reality check against real world dynamics.* Features rated as highly important for classifying default risk were: **BEHIND1 (**a derived Boolean of September late status), **BILL_/PAY_AMT2** (amt billed/paid in Aug), **DEBT_USAGE6** (derived float for Month 6, i.e., consumer's preexisting leverage).

   In narrative form, that reads -- "Customers currently late or behind, especially those with sizable April or August billings are more likely to default; that chance declines amongst those who've typically made steady large AMT payments in the past" ← Matches how credit cards "work", i.e., someone who signs on with pre-existing debt and increases their credit utilization % further (via new monthly purchases/billings) is risky!

Business Discussion and Recommendations:

- *In terms of operations risk, it's more dangerous for CO to rely on a classification model with high-degree of false negatives than one with false positives,* i.e., failing to predict a customer *who does* default is worse than mistakenly predicting someone would default (and denying them) when they wouldn't. Scoring metrics like F1, Recall, or ROC-AUC may be preferable (over precision) to compare models while accounting for default class imbalance.



(**‘Soft’ Voting Classifier:** Confusion Matrix / Prediction Histogram, **Logistic Regression** (not pictured) produced fewest ‘False Negatives’)

- However, *there’s substantial opportunity cost from being TOO conservative in lending decisions.* Future comprehensive analysis should take into account CO’s potential revenues, generated as merchant fees plus other service charges like late fees, i.e., basic customer LTV. Otherwise, concern about customer default may preclude CO accepting extremely valuable segments like those who (consistently) pay late, but in full with interest.

- *Default likelihood should be explored further via “probabilistic classification” techniques*. Logistic Regression was already tested and produced the fewest false negatives,but other algorithms, g metrics like log_loss/Brier, calibration plots, etc. may reveal new insights. Treating an individual’s default risk as a probability better reflects reality’s uncertainty AND may optimize credit limit decision (I.e, customer with low default risk can safely be granted higher limit, while those CO is ‘on-the-fence’ about can still be approved with less downside.

- *Enhance CO’s customer dataset with additional source*s – Longer card payment history, payment history of housing/utility/vehicle bills, customer income/savings, and details on other debt held should all improve current models even further.

Process Thoughts / Learnings

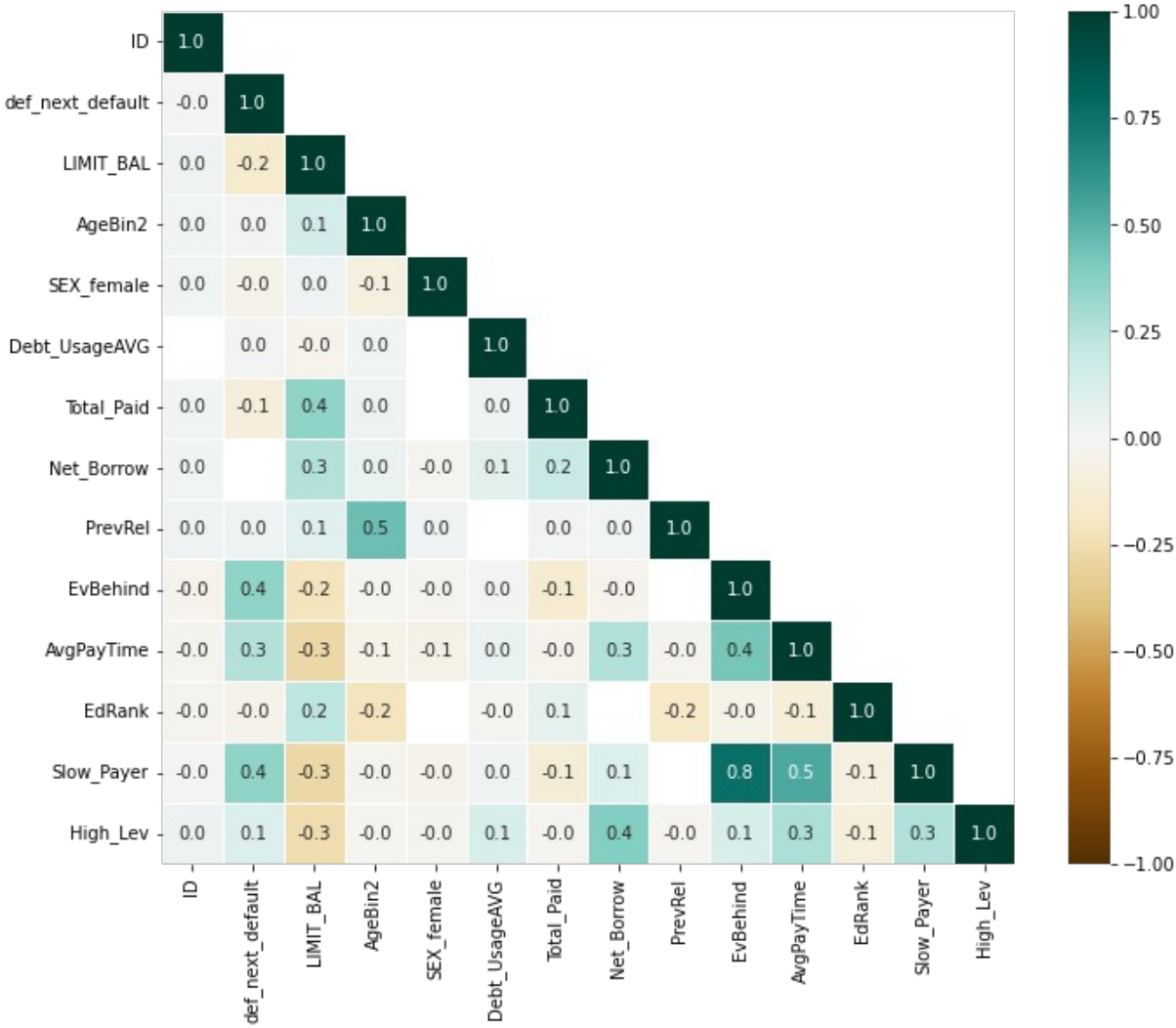
- *Specific pre-processing function appeared unimportant*, with only minimal differences in accuracy depending on which (if any) were applied against X. I’m guessing this step becomes necessary/more important for other datasets with different distributions, or that there are some ‘best practices’ to achieve the best results..

- *There was only minimal boost to accuracy (+ ~0.02) from hyper-tuning parameters* using GridSearchCV/RandomSearchCV. Given CPU resources required to fit a model 150 times (5 folds x 30 iterations of parameters), the cost:benefit seems dubious; I also question the robustness/validity of any model requiring hypertuning of already ‘black box’ algorithms.

Addendum Chart Outputs:

1 – Correlation Heatmap of 'Summary' Dataframe

- VIF threshold 5.0
- Masked cells (white-out) are NOT statistically significant at p_value .05



2 – Correlation Heatmap of Apr-Sept related variables

- Same statistical mask applied
- Note the diagonal-right band of 0.3 correlation starting with PAY_AMT1:BILL_AMT2 near the center… This suggests customers typically pay on lagging basis (as real world), although there are also customers who pay before statement periods end, (also real world).