UNDERWRITING WITH MACHINE LEARNING

GROUP 3

GOAL OF PRESENTATION

To give a high-level overview of our team's process and considerations behind designing a machine learning model for underwriting.

THE OBJECTIVE

Our main objective in this project is to build a set of machine learning models to determine whether we should lend money to a customer, given various constraints expressed by the management of our bank (budget, risk, etc).

This objective involves answering several questions:

1. What is the probability of default for each customer?
2. What is the loss given default created by each customer?
3. Given the risk & budget of each scenario, which customers do we approve for a loan?

STRATEGY

As the underwriting team, our strategy is to break down our model development into three pieces that, when utilized in sequence, can accurately answer the questions we've addressed in our objective.

These three pieces are:

1. Probability of Default model
2. Loss Given Default model
3. Combination phase which determines risk and constrains to budget from outputs of the first two models.

PROBABILITY OF DEFAULT MODEL

Our first piece of strategy required building a model that correctly predicts probability of default for each customer and classifies them given a 50% forgiveness threshold.

Process Steps:

1. Clean data; remove duplicate columns, impute missing values, remove zero variance/noisy variables, standardize and scale, balance test data to similar proportion of default/non-defaulters
2. Try several methods of modeling and compare results
   - Basic logistic regression
   - Lasso regression
   - Random Forest
3. Determine best model from performance metrics
   - Best model was ***Lasso regression***

LOSS GIVEN DEFAULT MODEL

Our second piece of strategy required building a model that correctly predicts loss given default for each customer.

Process Steps:

1. Same pre-processing steps as previous, with addition of removing statistically insignificant variables
2. Try several methods of modeling and compare results
   - No-penalty linear regression
   - Lasso-penalty linear regression
   - XGBoost
3. Determine best model from performance metrics
   - Best model was ***Linear regression with Lasso penalty***
   - ***Note: XGBoost was making false assumption and basic linear didn't zero out coefficients of variables.

COMBINATION OF MODELS

Our final piece of strategy was combining our previous efforts to perform the required analysis we needed to answer our posed questions. We took the results of each model and assimilated them along the following parameters:

1. **Risk**

   Risk = Loan value * LGD * PD

2. **Gain**

   Gain = Loan value * Interest Rate * (1-PD) * Number of years

3. **Delta**

   Delta = Gain - Risk

***Note: Loan value - defined as requested loan amount by each customer for 5 years

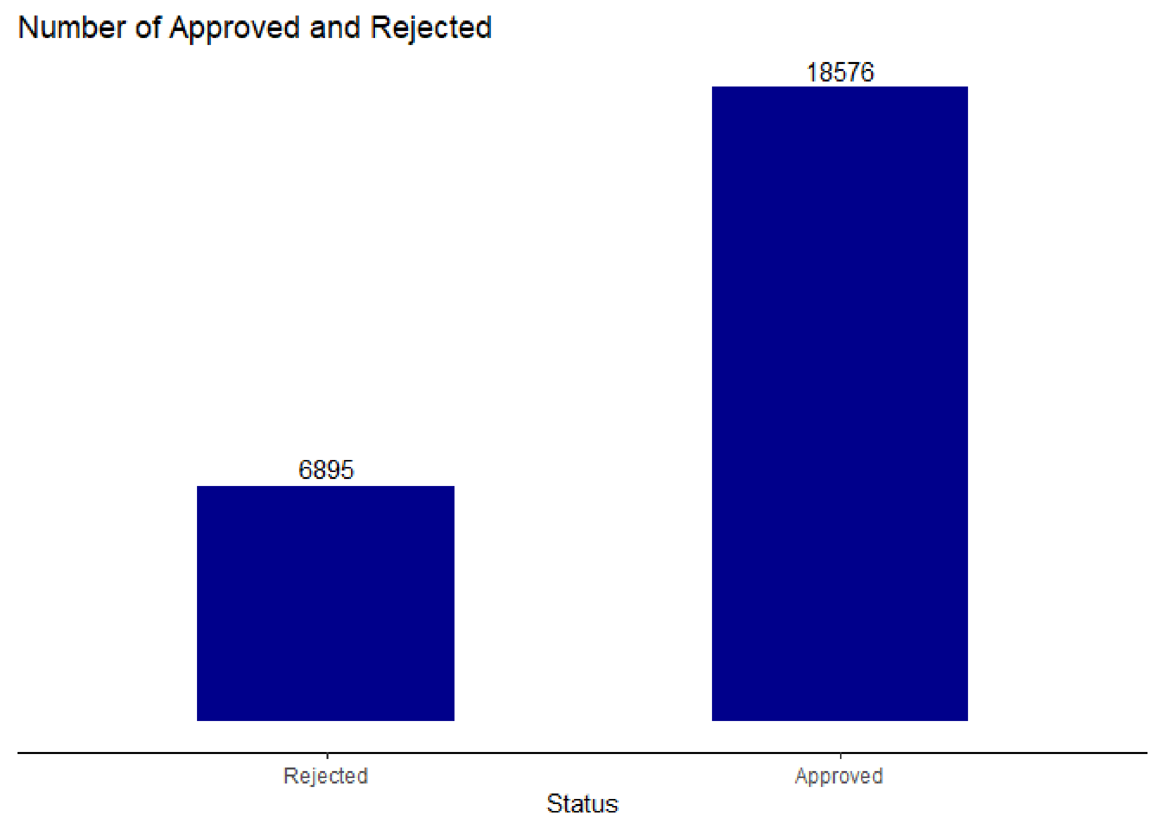
LGD - Loan Given Default in percentage predicted by the model on test scenario

PD - Probability of Default s 1(defaulter) & 0(non-defaulter) predicted by the model on test scenario

SCENARIO 1

Given our capital budget of 1.4 billion dollars, we used the **RISK** parameter to accept or reject customers for a loan.

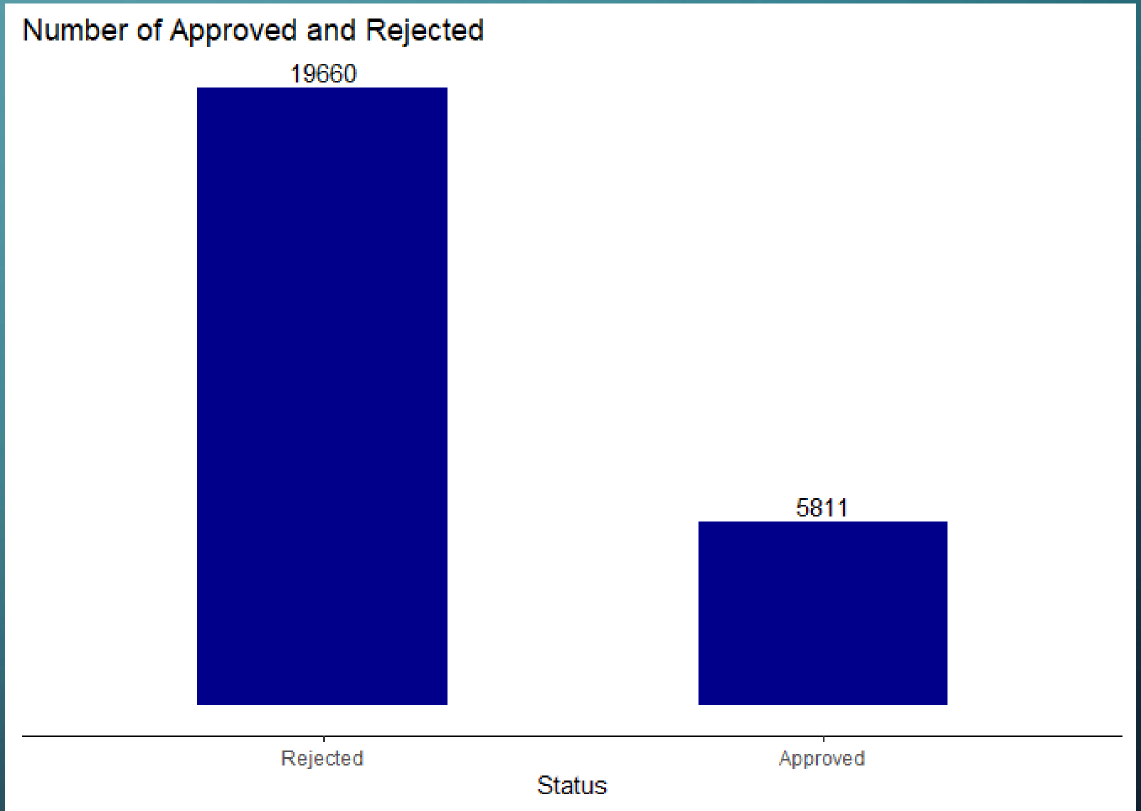
Considering **Risk < Gain**, our model outputted the number of customers fitting this constraint and rejected the rest. We were able to make 1 billion dollars of loans with this method and have 0.4 billion of our budget left over.



SCENARIO 2

Given our capital budget of 450 million dollars, we used the **DELTA** parameter to accept or reject customers for a loan.

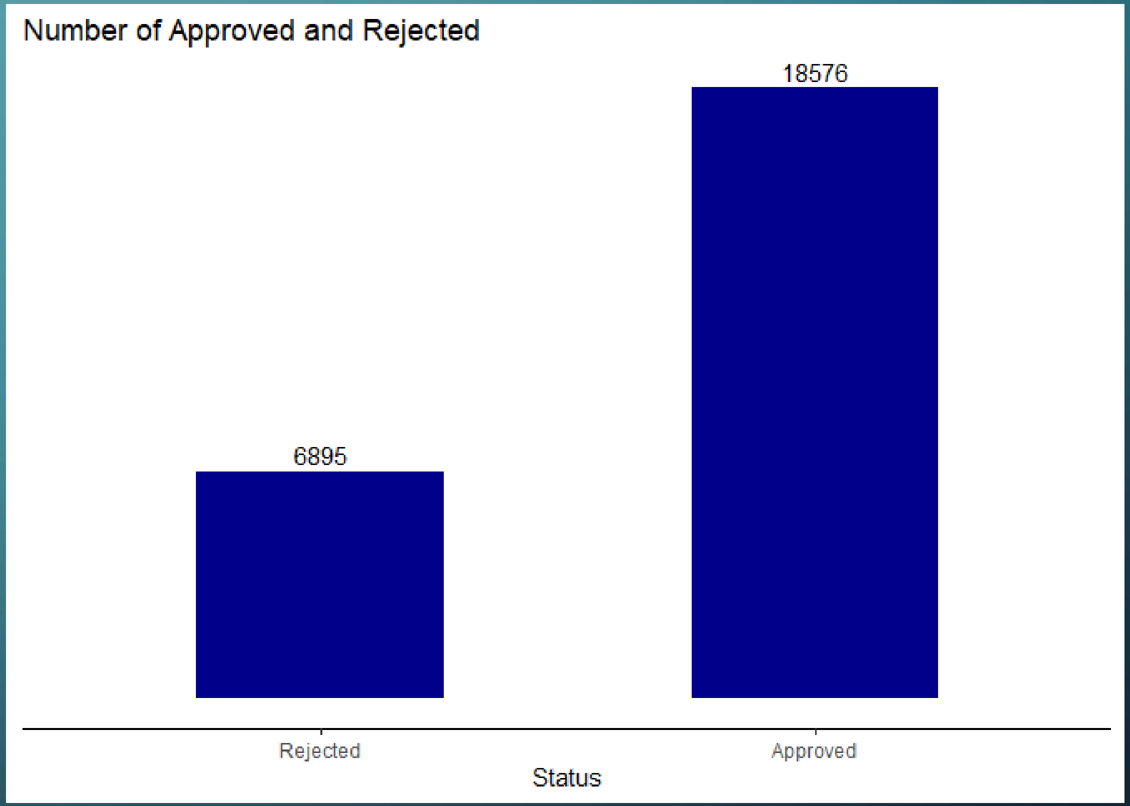
Considering **Delta > 0,** we ranked the output of our model by delta values in descending order and chose to accept customers for loan by reading down the output list until the cumulative sum of all accepted customers reached 450 million dollars.



SCENARIO 3

Given our capital budget of 1.4 billion dollars, we used the **DELTA** parameter to accept or reject customers for a loan.

Considering **Delta > 0** along with each customer's **proposed interest rate,** we receive different results from scenario one. The results from prediction classify for approval given difference of gain vs risk for each customer, where **GAIN** is newly determined by the customer's proposed rate. The gain is variable here and a different configuration of customers meeting the capital budget was achieved.



OBJECTIVE - REVISITED

From our 3-parted strategy, our underwriting team was able to successful predict both the probability of default and loss given default of each customer and leverage that information to classify for loan approval given the delta between gain and risk of each given customer. We achieved this successful model by utilizing concepts of imputation, Lasso penalizing, logistic and linear regression, and algebraic methods.

Given the continual addition of more customer loan data, this modeling process could be fine-tuned to serve as an accurate decision-maker at our bank. We were able to stay at-or-under budget while minimizing risk and maximizing gain for our bank.