UGA Data Science Competition: Information Session

Industry Data Scientists

March 24 & 25, 2021

Overview

- Credit Decisions: Background and Modeling
- Description of Data
- Data Science Competition
- Q & A

Credit Lending

- Financial institutions offer a variety of credit products:
  - consumers (retail) and companies (wholesale)
- Types of products:
  - Credit cards – regular, prime, student, etc.
  - Auto-loans – direct to customer or indirect through car dealership
  - Mortgages – residential (primary, home equity loan); commercial
  - Personal Loans – small amount of loans for personal use
- Secured and Unsecured Loans
- Most institutions use their own models to approve or deny application for credit

  (details discussed in the next few slides)
- Once approved: Different products ("loans") have different terms of payment
- Possible outcomes over time:
  - minimum payment made, delinquent, default, pre-paid, bankruptcy
- For credit cards:
  - Account is OK if min payment that is due is paid each month
  - Delinquent if min payment is not made that month
  - Default: different rules; for this project: default = delinquent for three consecutive months.

How Are Credit Decisions Made: Retail Banking

- Decisions are made based on the “creditworthiness” of the applicant.
- Data-based models are used to determine creditworthiness:
  - Based on historical information on applicant’s credit-related financial transactions
    - Example: An applicant for a new credit card may be declined due to multiple late payment history on his car loan or default on a personal loan initiated two years ago
  - Income may or may not play a role in credit decisions, depending on the product.
    - For mortgages, income needs to be verified.
    - For cards, income may not be verified.
- Primary source of historical data: credit bureaus (next page)
- There are also other sources of data:
  - From internal relationships with the bank (checking or savings accounts)
  - Alternative sources of data accumulated from online lending platforms (such as Lending tree)

What are Credit Bureaus?

- A credit reporting company that collects and researches credit information on individuals and sells it for a fee to potential creditors.
- Financial companies also report default/delinquency behaviors to the bureaus.
- Several such bureaus but top ones in the U.S. are Experian, Equifax, and TransUnion.

Models for Credit Decisions

- Individual models are developed for a specific portfolio within a product type:
  - Example: Product type: mortgage; Portfolio: home equity line of credit
- Data for developing a model:
  - Data for approved accounts for that portfolio and product type
    - Response: Typically it is binary (for example, account defaulted or not)
    - Covariates/Predictors: Credit bureau information that was available for the approved accounts available at time of credit application plus a few others (we will see specific examples soon)
- Typical steps:
  - Fit a binary regression model to the data for predicting probability of default (between 0 and 1);
  - Come up with appropriate criteria to select a threshold;
  - Apply it to the predicted default probability to get a 0-1 decision rule for accept/decline decision
    - Example: Threshold is determined as 0.2; if predicted default prob is < 0.2, accept; decline otherwise
  - For new application, use the model to develop the predicted probability, and the 0-1 decision rule to make accept/decline decision
- Traditional statistical techniques, such as logistic regression, used extensively
- More recently, machine learning algorithms have also become popular

Data Science Competition: Credit card decision algorithm

Task: Develop an algorithm for making accept/decline decisions for credit cards

- Use historical data to develop two models for predicting probability of default
- Compare and decide which model to use
- Use selected model to develop a decision algorithm for new customers

- Data: Simulated data for a hypothetical bank XYZ
  - Three datasets: training data with 20,000 accounts; validation dataset with 3,000 accounts, and test dataset with 5,000 accounts.
    - Validation dataset is used for hyper-parameter tuning
    - Test data set is used to assess performance
  - Response: binary (whether the account defaulted or not – see next page)
  - Twenty predictors/covariates the accounts (see next page)

Variables: More information

- Response variable is binary: approved/active account defaulted or not
  - Default means minimum payment was not made for 3 consecutive months
  - 1 = account defaulted within 18 months window;
  - 0 = did not default

- Covariates/predictors: (more details next page)
    - Credit debt
    - Credit tenure/age
    - Past credit delinquency information
    - Credit inquiries
    - Number of Accounts, Balances, and Utilization
    - Additional Information

Information on Predictors (1)

- **Credit Debt:**
  - *tot_credit_debt:* Total debt (amount owed by applicant at the time of application) on all of their credit products (credit cards, auto-loans, mortgages, etc.)
  - *avg_card_debt:* Average monthly debt (amount owed by applicant) on all of their credit cards over last 12 months
- **Credit tenure/age:**
  - *credit_age*: Age in months of first credit product (credit cards, auto-loans, mortgages, etc.) obtained by the applicant
  - *credit_good_age*: Age in months of first credit product obtained by the applicant that is currently in "good" standing (no past due payments)
  - *card_age:* Age in months of first credit card obtained by the applicant

- **Past Credit Delinquency Information:**
  - *non_mtg_acc_past_due_12_months_num:* Number of non-mortgage credit-product accounts by the applicants that are 30 or more days delinquent within last 12 months
  - *non_mtg_acc_past_due_6_months_num:* Number of non-mortgage credit-product accounts by the applicant that are 30 or more days delinquent within last 6 months
  - *mortgages_past_due_6_months_num*: Number of mortgages by the applicant that are delinquent within last 6 months
  - *credit_past_due_amount*: Total amount of money that is currently past due on all credit accounts

Information on Predictors (2)

- **Credit Inquiries**: An inquiry occurs when the applicant's credit history is requested by a lender from the credit bureau. This occurs when a consumer applies for credit.
  - *inq_12_month_num*: Number of credit inquiries in last 12 months
  - *card_inq_24_month_num*: Number of credit card inquiries (on applicant's credit) in last 24 months

- **Number of Accounts, Balances, and Utilization:** Utilization is ratio of balance divided by credit limit
  - *card_open_36_month_num*: Number of credit cards opened by applicant in last 36 months
  - *auto_open_36_month_num*: Number of auto loans opened by applicant in last 36 months
  - *uti_card*: Utilization on (all currently available) credit card accounts
  - *uti_50plus_pct*: Percentage of open credit products (accounts) with over 50% utilization
  - *uti_max_credit_line*: Utilization of credit product (account) with highest credit limit
  - *uti_card_50plus_pct*: Percentage of open credit cards with over 50% utilization

Additional Information:

- *ind_acc_XYZ*: Indicator: 1 if applicant already has some account (checking/savings, etc.) with the bank XYZ; 0 otherwise
- *rep_income*: annual income (self-reported by applicant and not verified)
- *States*: Reported state of residence of applicant (AL, FL, GA, LA, MS, NC, SC)

VARIABLE NAMES USED IN THE DATASET	DESCRIPTION OF VARIABLES
Response: Default_Ind	Indicator of Default: Binary: 1 = account defaulted after an account was approved and opened with bank within 18 months performance window; 0 = not defaulted; (Default means no payments for 3 consecutive months)
Predictors: Applicant's attributes derived from information available from credit bureaus at the time of application	
tot_credit_debt	Total debt (amount owed by applicant at the time of application) on all of their credit products (credit cards, auto-loans, mortgages, etc.)
avg_card_debt	Average monthly debt (amount owed by applicant) on all of their credit cards over last 12 months
credit_age	Age in months of first credit product (credit cards, auto-loans, mortgages, etc.) obtained by the applicant
credit_good_age	Age in months of first credit product obtained by the applicant that is currently in "good" standing (no past due payments)
card_age	Age in months of first credit card obtained by the applicant
non_mtg_acc_past_due_12_months_num	Number of non-mortgage credit-product accounts by the applicants that are 30 or more days delinquent within last 12 months (Delinquent means payment not made)
non_mtg_acc_past_due_6_months_num	Number of non-mortgage credit-product accounts by the applicant that are 30 or more days delinquent within last 6 months
mortgages_past_due_6_months_num	Number of mortgages by the applicant that are delinquent within last 6 months
credit_past_due_amount	Total amount of money that is currently past due on all credit accounts
inq_12_month_num	Number of credit inquiries in last 12 months (An inquiry occurs when the applicant's credit history is requested by a lender from the credit bureau. This occurs when a consumer applies for credit.)
card_inq_24_month_num	Number of credit card inquiries (on applicant's credit) in last 24 months
card_open_36_month_num	Number of credit cards opened by applicant in last 36 months
auto_open_36_month_num	Number of auto loans opened by applicant in last 36 months
uti_card	Utilization on (all currently available) credit card accounts (Utilization is ratio of balance divided by credit limit)
uti_50plus_pct	Percentage of open credit products (accounts) with over 50% utilization
uti_max_credit_line	Utilization of credit product (account) with highest credit limit
uti_card_50plus_pct	Percentage of open credit cards with over 50% utilization
ind_acc_XYZ	Indicator: 1 if applicant already has some account (checking/savings, etc.) with the bank XYZ; 0 otherwise
rep_income	annual income (self-reported by applicant and not verified)
States	Residence state of applicant (AL, FL, GA, LA, MS, NC, SC)

Data Science Competition (Recap)

- You are expected to work with you teammates to develop credit decision models using the simulated datasets, and to interpret the results, document the findings, and give a presentation to the judges.
- Specifically, you are asked to
  - Conduct an exploratory data analysis
  - Develop and fit a logistic regression model
  - Develop an additional machine learning (ML) model: random forest, gradient boosting or feedforward neural network.
  - Compare the logistic regression model with ML model, select one for credit approval and describe the reason for your selection
  - Describe how you would use it to make decisions on future credit card applications.
  - Answer following questions:
    - do customers who already have an account with the financial institution receive any favorable treatment in your model?
    - Suppose a credit card application is rejected using your model, and the applicant asks you to provide an explanation on why it was rejected. How would you explain the results to the customer?

Deliverables

- The final deliverables include:
  - A report describing the key steps in your analysis, with no more than 15 pages of main body and 5 pages of appendix.
  - The codes for analysis, with brief but adequate annotations
  - A presentation deck summarized your results and conclusions, with no more than 12 slides
- Details can be found out in the Grad-UGA-Data-Analysis-Competition.docx and UGrad-UGA-Data-Analysis-Competition.docx document.
- Final rating will be based on the quality of data analysis, documentation, and presentation.

Questions