DREXEL UNIVERSITY
LeBow
College of Business

Understanding Loan Customers through Mortgage and Home Equity Datasets

Predicting the customer creditworthiness using Machine Learning

MGMT 715 | Group 4

Agenda

- Introduction
  - Executive Summary
  - Background Research
- Mortgage Loans
- Home Equity Loans
- Recommendations
- References

DREXEL UNIVERSITY
LeBow
College of Business

Executive Summary

Problem Statement:
Understanding loan customers and predicting their creditworthiness

Key Findings:
Understanding loan characteristics including customer information such as socio-economic factors as well as market information such as *GDP and employment rate* are important to fulfilling the project objective

Recommendations:
Predicting creditworthiness can play a significant role in determining factors like loan amount and interest rates to offer.

DREXEL UNIVERSITY
LeBow
College of Business

Background Research

Housing Bubble during the 2008 economic depression

Early stages of a substantial downshift?

Mortgage Loans

DREXEL UNIVERSITY
LeBow
College of Business

Data Exploration : Mortgage

- Total observations in the raw dataset: 622,489
- Unique observations: 50,000 customers
- After data aggregation: 8,146 updated observations
- New variable created based on Average and Standard Deviation

Types	Variables
Numerical Variable - 19	Maturity time, Outstanding Balance, Loan to Value, Interest Rate, Home Price Index Average of Outstanding Balance, Loan to Value, Interest Rate, Home Price Index, Gross Domestic Product, Unemployment Rate Standard Deviation of Outstanding Balance, Loan to Value, Interest Rate, Home Price Index, Gross Domestic Product, Unemployment Rate
Categorical Variable - 3	Real Estate Type, Investor, and Status time

DREXEL UNIVERSITY
LeBow
College of Business

Data Quality and Pre-processing : Mortgage

Missing Values:

- All missing values in standard deviation variables, count: 763
- Imputed missing variable with value '0'

Outliers:

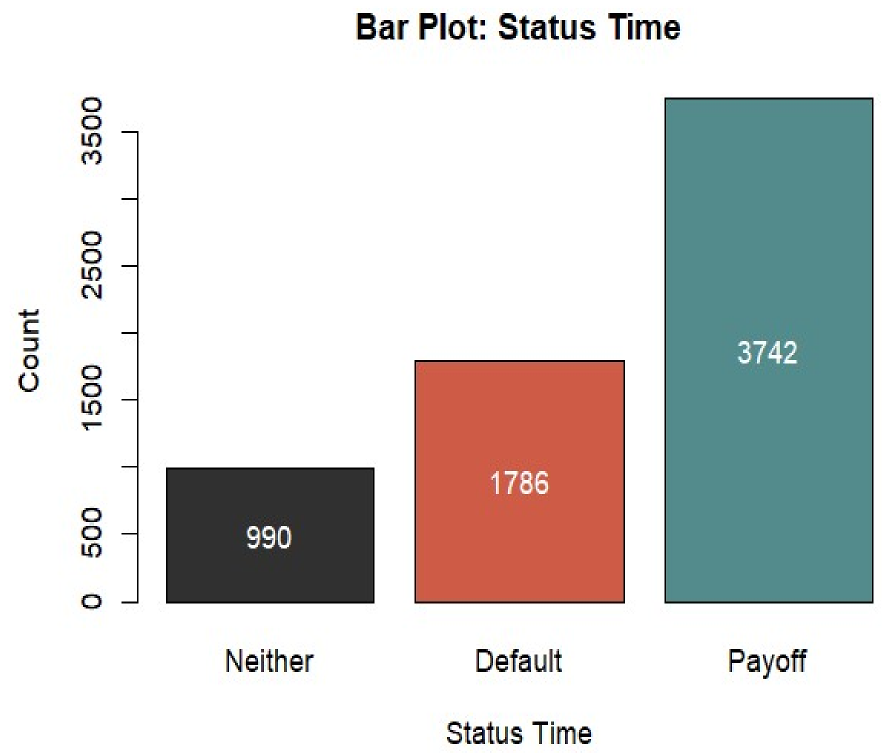
- 9% of the dataset, count of 752 outliers
- Outliers have been preserved

We applied min-max (range) normalization to the training data prior to modeling.

Training/Testing: Mortgage

Target Variable: *Status Time*

- Class imbalance in the target variable
- SMOTE (Synthetic Minority Oversampling Technique) used for up-sampling of the minority class
- Train and Test split in 80:20 ratio



DREXEL UNIVERSITY LeBow College of Business

Model Testing: Mortgage

Decision Tree

- Grid and control are used to tune complexity parameters
- Default class seems overfitting

Decision Tree	
Accuracy	0.7377
Kappa	0.5484

Naïve Bayes

- Yeo-Johnson used to normalize the dataset
- Laplace smoothing technique used

Naïve Bayes	
Accuracy	0.7359
Kappa	0.5447

Model Comparison: Mortgage

Comparing measures across different models

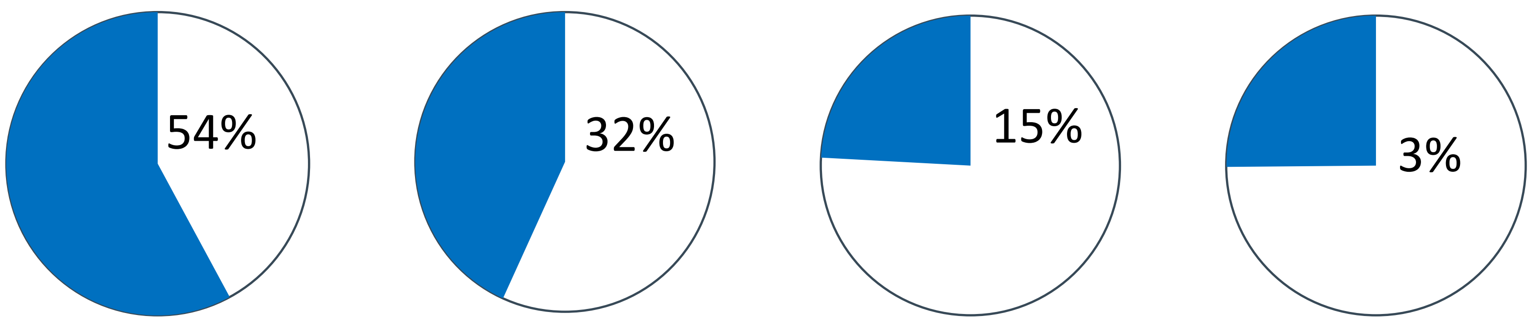
		Precision	Recall	F1
Neither	Decision Trees	0.9317	0.9393	0.9355
	Naïve Bayes	0.4857	0.8907	0.6286
	Deep Neural Network	0.8207	0.9701	0.8892
Default	Decision Trees	0.5359	0.5852	0.5595
	Naïve Bayes	0.6641	0.3857	0.4879
	Deep Neural Network	0.6474	0.6293	0.6382
Payoff	Decision Trees	0.7904	0.7540	0.7718
	Naïve Bayes	0.7828	0.7669	0.7747
	Deep Neural Network	0.8336	0.7929	0.8127

Home Equity Loans

Data Exploration : Home Equity

Reasons for Taking Home Equity Loans

- Homeowner study (Lending Tree) - 2.3 million House Prices – Jan'20
- House Prices Growth Jan 2019 (5.5%) – Jan 2022 (17.50)
- Home Remodeling - 4.7% Q1 to 9.5% Q2



Home Improvement | Debt Consolidation | Investment | Retirement Homes

Boston	54%
Philadelphia	53%

Las Vegas	32%
Phoenix	28%

San Jose	17%
Miami	14%

Las Vegas	3%
Los Angeles	2%

Source: CEICdata.com (2018)

Source: Kolomatsky (2022)

DREXEL UNIVERSITY
LeBow
College of Business

Data Exploration : Home Equity

- Total observations in the raw dataset: 5,960
- Observations across 13 variables
- Target Variable: *Default*

Types	Variables
Numerical - 9	Loan Amount, Existing Mortgage, Home Valuation, Home Valuation, Years on Job, Credit Line age, Credit Inquiries, Credit Lines, and Debt to Equity Ratio
Categorical Variable - 4	Reason, Occupation, Number of Derogatory Reports, Number of Delinquent Credit Lines,

Data Quality and Pre-processing : Home Equity

Missing/Duplicate Values:

- 5,283 missing values
- Imputed missing variable with the median of the variables
- Dropped 4 observed duplicate values
- High correlation between *Loan Amount* and *Mortgage Balance*

Outliers:

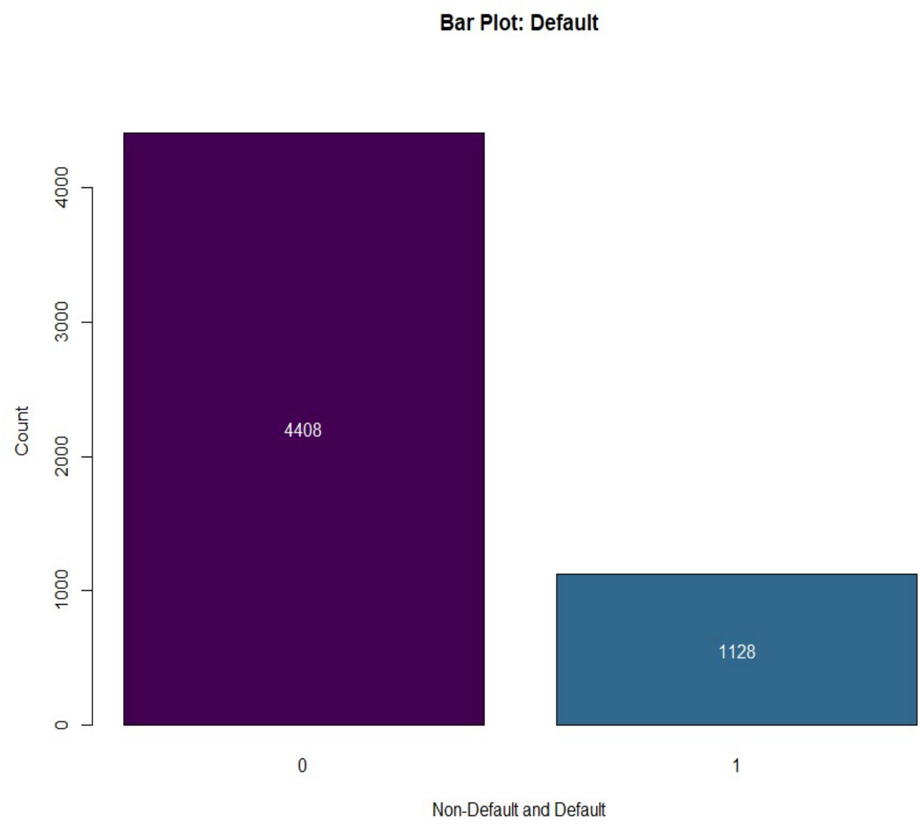
- 9% of the dataset, count of 752 outliers
- Outliers have been preserved
- Major outliers observed in *Mortgage balance* & *Age of oldest credit line*

Drexel University LeBow College of Business

Training/Testing : Home Equity

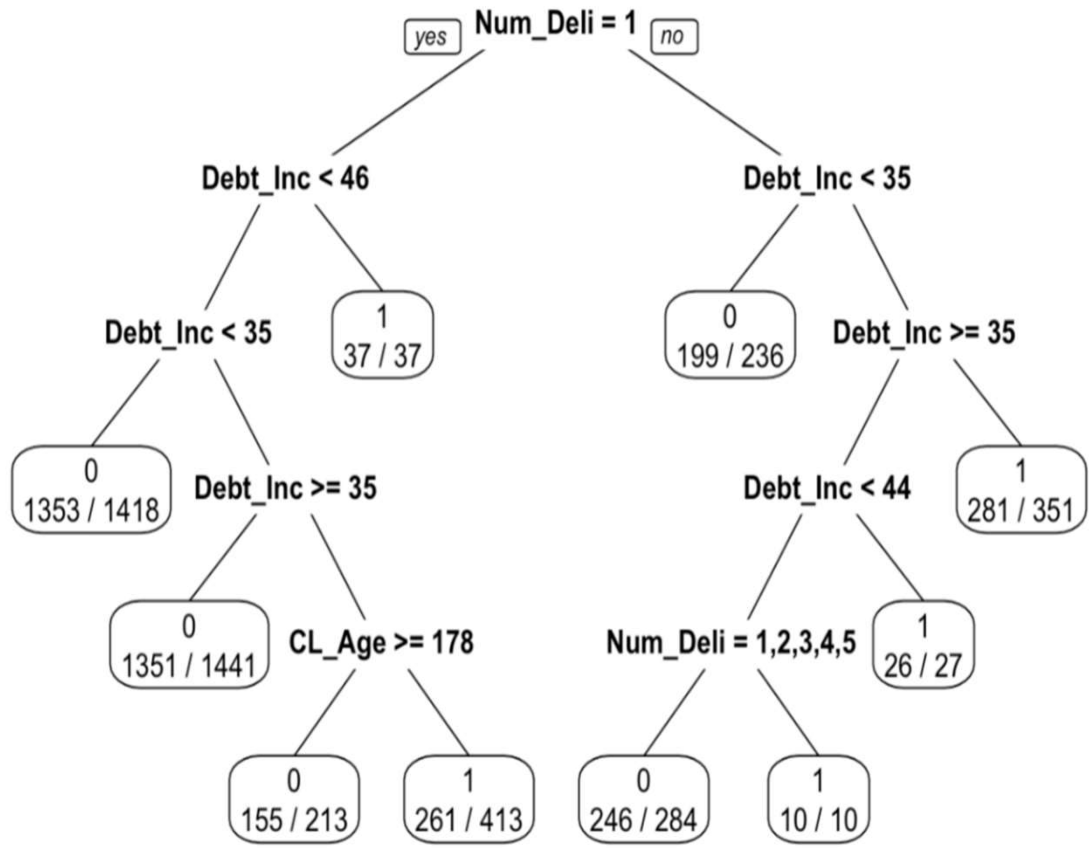
Target Variable: *Default*

- Class imbalance in the target variable
- Up-sampled the data to handle the imbalance
- We applied min-max (range) normalization to the training data prior to modeling.
- Train and Test split in 80:20 ratio



Decision Tree: Home Equity

- Highest accuracy observed with a cp value between 0 and 0.05
- Grids divided from 0 to 0.2 with a step size of 0.005
- Used repeated cv with 10 folds and 10 times
- Debt-to-income ratio is observed the most crucial among others



DREXEL UNIVERSITY LeBow College of Business

Deep Learning: Home Equity

- No significant loss between training and validation
- Up-sampled the data for dealing with class imbalance
- We used Keras tuning for fine-tuning results
- No overfitting or underfitting was observed in the training and testing results.

		Precision	Recall	F1
Not Default	Testing	0.7864	0.8454	0.8148
Default	Testing	0.8389	0.7781	0.8074

Model Comparison: Home Equity

Comparing measures across different models

		Precision	Recall	F1
Not Default	Decision Trees	0.9083	0.9557	0.9313
	Naïve Bayes	0.8484	0.9534	0.8978
	Deep Neural Network	0.7864	0.8454	0.8148
Default	Decision Trees	0.7821	0.6222	0.6930
	Naïve Bayes	0.6465	0.3333	0.4398
	Deep Neural Network	0.8389	0.7781	0.8074

Recommendations

- More data for better and more precise results
- Variables with high importance as compared to others. *Debt-to-income ratio* and *Number of delinquencies* are of very high importance
- Any delinquent record is a red flag, look for other important variables
- Customers with delinquency equal to 1 vs. customers with 2-5 delinquencies, likely to default
- Debt-to-income ratio of less than 35 is very less likely to default, can be used to allocate loan amount
- For a customer's lifetime age greater than 178 says that there is more than a 50% chance that the customer will default

References

- Ashwani Kumar et al. 2021 J. Phys.: Conf. Ser. 1998 012027). Credit Score Prediction System using Deep Learning and K-Means Algorithms. Credit Score Prediction System Using Deep Learning and K-Means Algorithms. https://iopscience.iop.org/article/10.1088/1742-6596/1998/1/012027/pdf
- Bachaud, N., & Lee, A. (2021) Rent Prices Soar Beyond Pre-Pandemic Projections (July 2021 Market Report). Retrieved from https://www.zillow.com/research/july-2021-zillow-market-report-29992/
- CEICdata.com. (2018, June 1). United States House Prices Growth. CEIC Data. https://www.ceicdata.com/en/indicator/united-states/house-prices-growth
- Consumer Finance Protection Bureau. (2021, August). *Data Point: 2020 Mortgage Market Activity and Trends*. https://files.consumerfinance.gov/f/documents/cfpb_2020-mortgage-market-activity-trends_report_2021-08.pdf
- Edwards, W. (2022, May 26). *What's next for the US housing bubble? industry experts break down the bull and bear cases for home prices in the months ahead.* Business Insider. Retrieved August 29, 2022, from https://www.businessinsider.com/housing-market-crash-home-prices-mortgage-rates-bubble
- Kolomatsky, M. (2022, January 27). How Are People Spending Their Home-Equity Loans? The New York Times. https://www.nytimes.com/2022/01/27/realestate/home-equity-loans.html
- Manhertz, T., & Lee, A. (2022). Renters at the tipping point of homeownership: Estimating the impact of telework. *Cityscape, 24*(1), 259-285. Retrieved from

  http://ezproxy2.library.drexel.edu/login?url=https://www.proquest.com/scholarly journals/renters-at-tipping-point- homeownership-estimating/docview/2665650520/se-2?accountid=10559
- Washington post. (2022, June) - *Mortgage Lenders Timed the Market Perfectly,* https://www.washingtonpost.com/business/mortgage-lenders-timed-the-market-perfectly/2022/06/21/d3fe397c-f11f-11ec-ac16-8fbf7194cd78_story.html

DREXEL UNIVERSITY
LeBow
College of Business

Thank You

Amaan Ghauri | Shivam Pandey
Utsav Pradhan | Amrut Prasade