

Team 43

Harris Ashraf Sharmila Baskaran Bo Chen Travis Jefferies Daniel Mower Cody Nguyen Ryan Wong
hashraf3@gatech.edu sbaskaran30@gatech.edu bchen354@gatech.edu tjefferies3@gatech.edu dmower3@gatech.edu cnguyen311@gatech.edu rwong33@gatech.edu

Motivation - Introduction

Lenders provide borrowers with a mortgage loan that is paid back monthly within a fixed time-frame. Lenders make profit by imposing interest, which decreases if the borrower introduces risk by defaulting, refinancing, or prepaying their mortgages.

Lenders want to understand influential factors associated with mortgage risk.

Borrowers want to reduce their interest rate when requesting a mortgage loan.

Data Source



Freddie Mac provides loan-level origination and performance data to the public for purchased mortgages

26.3 M
US mortgages

75 GB
large dataset

18 Years
1999 - 2017

53
viable features

Project Goal

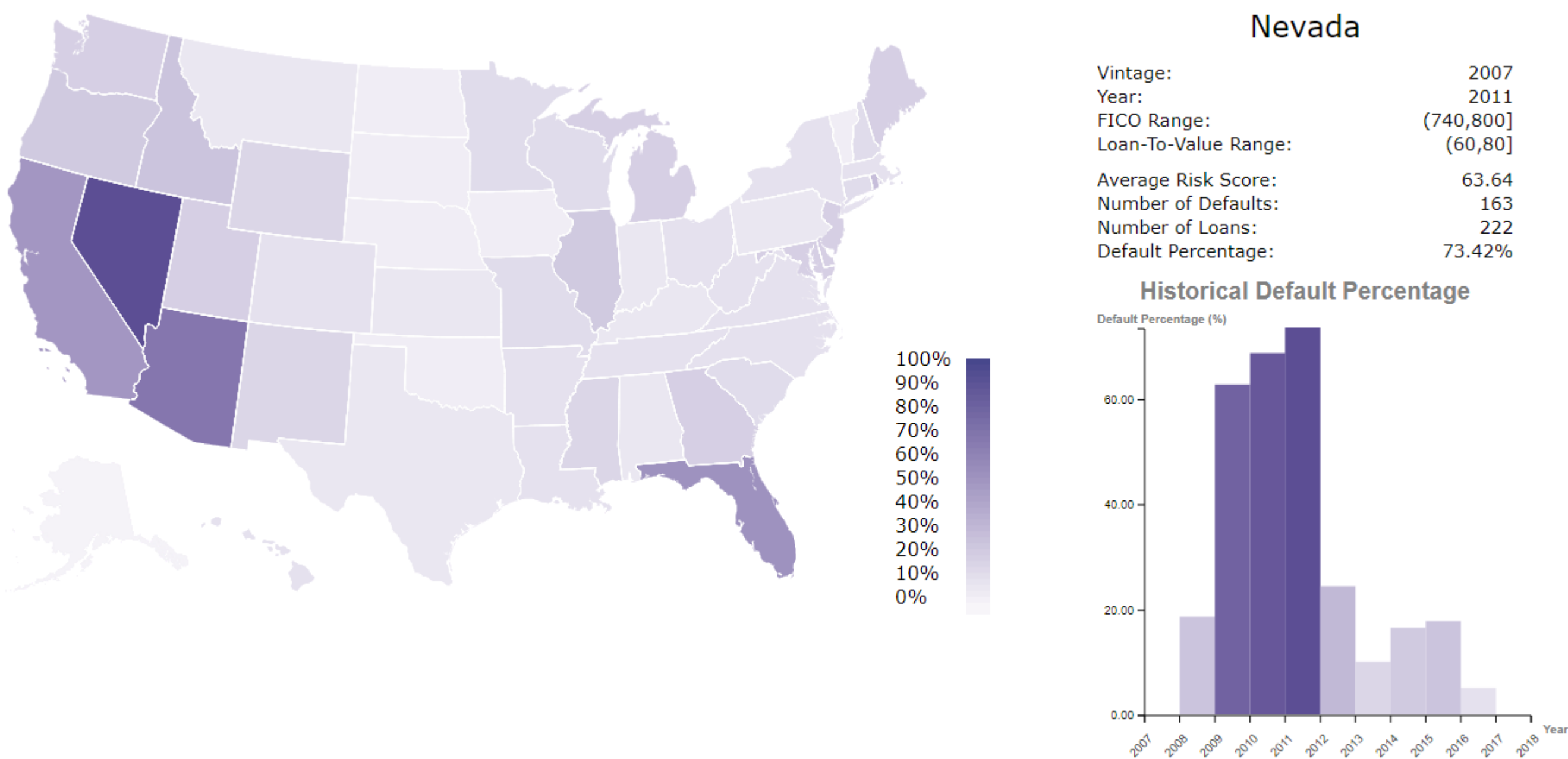
Create a tool that will allow both lenders and borrowers to **easily gauge the financial market and mortgage loan characteristics** to make more informed decisions.

We will do this by allowing users to:

- Easily access historical mortgage data
- Efficiently calculate the riskiness of specific loan characteristics

Our method is novel because it is analyzed on an individual-loan level, and conveniently integrates two important financial applications.

Historical Mortgage Heatmap



Users select an origination year, observation year, FICO score range, and a Loan-to-Value range to view default statistics across the United States. This provides users with 5,575 queries across 50 states and D.C.

Risk-Based Mortgage Rate Quote

FICO Score: 741-760

First Time Buyer: No

Occupancy Status: Primary Residence

Property Type: Single Family

Purpose for Loan: No Cash-out Refinanc

Term (Months): 30-yr

Number of Borrowers: One

Is there current a recession?: No

Purchase Price(\$): \$0-\$50,000

Down Payment: 5.0%

Monthly Income before taxes (\$): 6250

Total Monthly Debt (\$): 2250

Get Quote

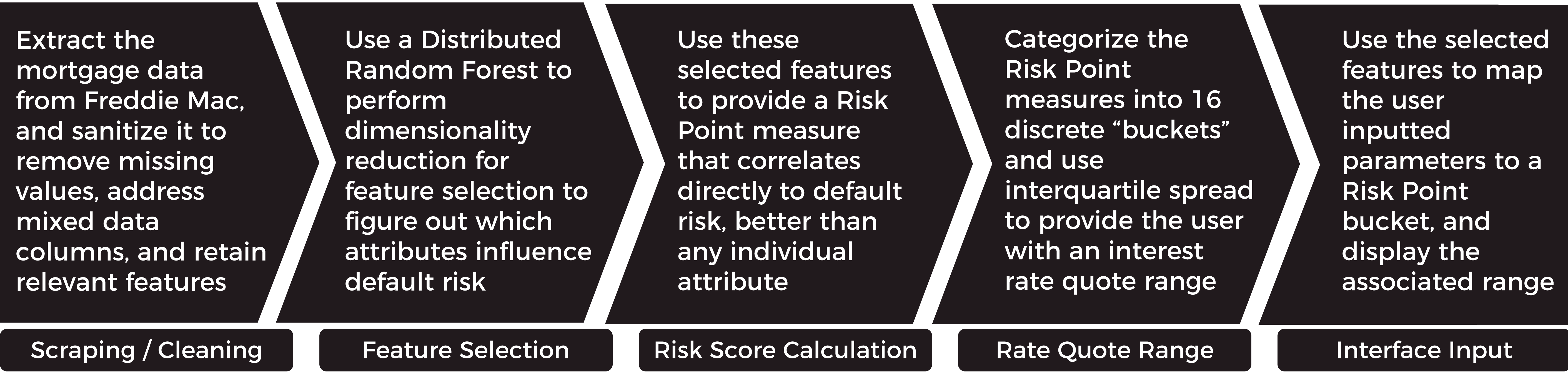
48.00% of borrowers are more risky than you.

Low Rate :	Median Rate :	High Rate :
4.78	4.98	5.22

Current Average 30-Year Mortgage Rate : 4.94

Users input various information about their financial profile and loan characteristics to see their riskiness compared to the market, and rates that resemble the spread appropriate for their situation.

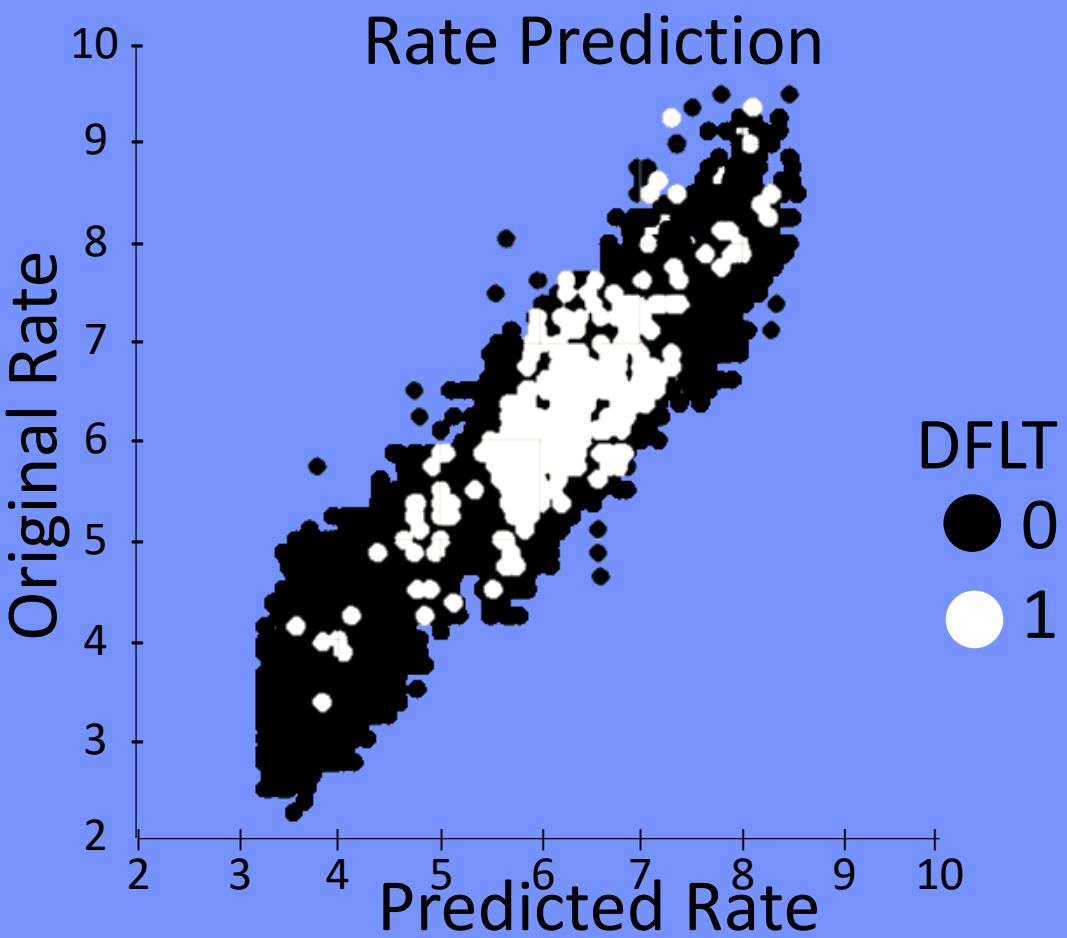
Model Algorithm



Experimentation and Evaluation

We found that traditional methods to classify mortgage default at the time of origination (e.g. K-NN, Random Forest classification) failed because of heavily skewed non-defaulted historical data. Survival Analysis techniques were also negative influenced by the limited natural of mortgage datasets.

By maintaining 15% of the dataset or 4 million historical mortgage data points as our test sample, we were able to consistently predict median interest rate quotes with a $\pm 0.25\%$ difference from historical averages. Our original vs predicted rate has a slope of 0.987 and R-squared value of 0.993, meaning our predictions are close to the market.



Based on our user studies, the most useful capability was the integration between historical data and personalized input analyses via both tools. Our respondents reported this integration provides the right useful information to make requesting a quote easier.

Innovation



Users can quickly understand mortgage risk without knowledge of machine learning



Lenders can easily understand how to maximize expected profits



Borrowers can effortlessly evaluate their mortgage rates



Users can conveniently analyze mortgage trends and access historical data