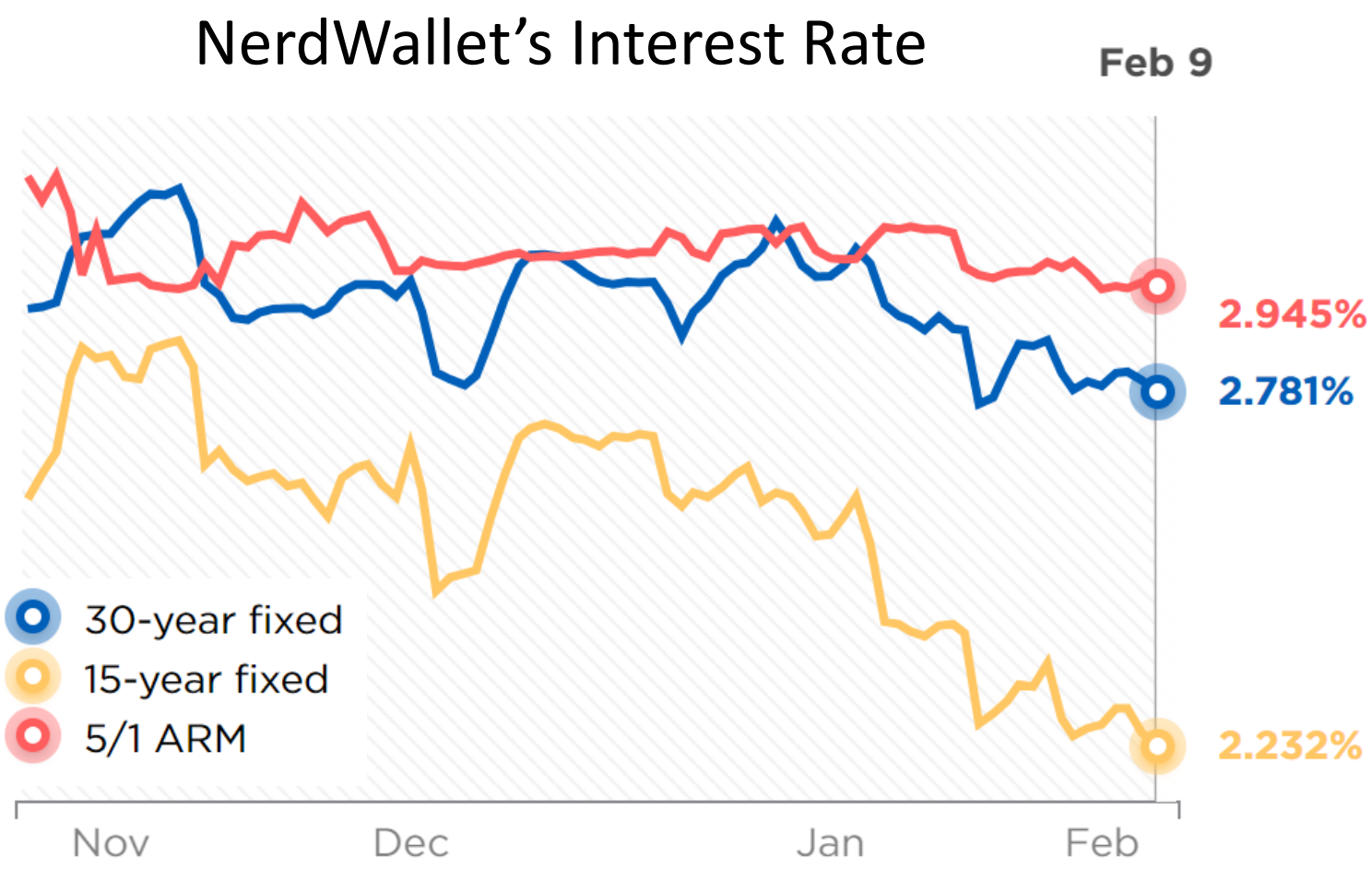
Predicting Lenders' Decision on Home Loan Refinancing Application

A Classification Analysis

Wei Zhao

Background



Good time to refinance your home loan

Understand what and how financial features are used to make decisions

Data Collection

Consumer Financial Protection Bureau

Data Summary

State:
MASSACHUSETTS

2019

Action Taken:
1 - Loan Originated
2 - Application approved but not accepted
3 - Application denied

Loan Purpose:
31 - Refinancing

Selected Variables	# of Records	$ Amount
1 - Loan Originated, 31 - Refinancing	63,212	25,012,190,000
2 - Application approved but not accepted, 31 - Refinancing	2,003	677,885,000
3 - Application denied, 31 - Refinancing	13,405	3,905,945,000

Approved

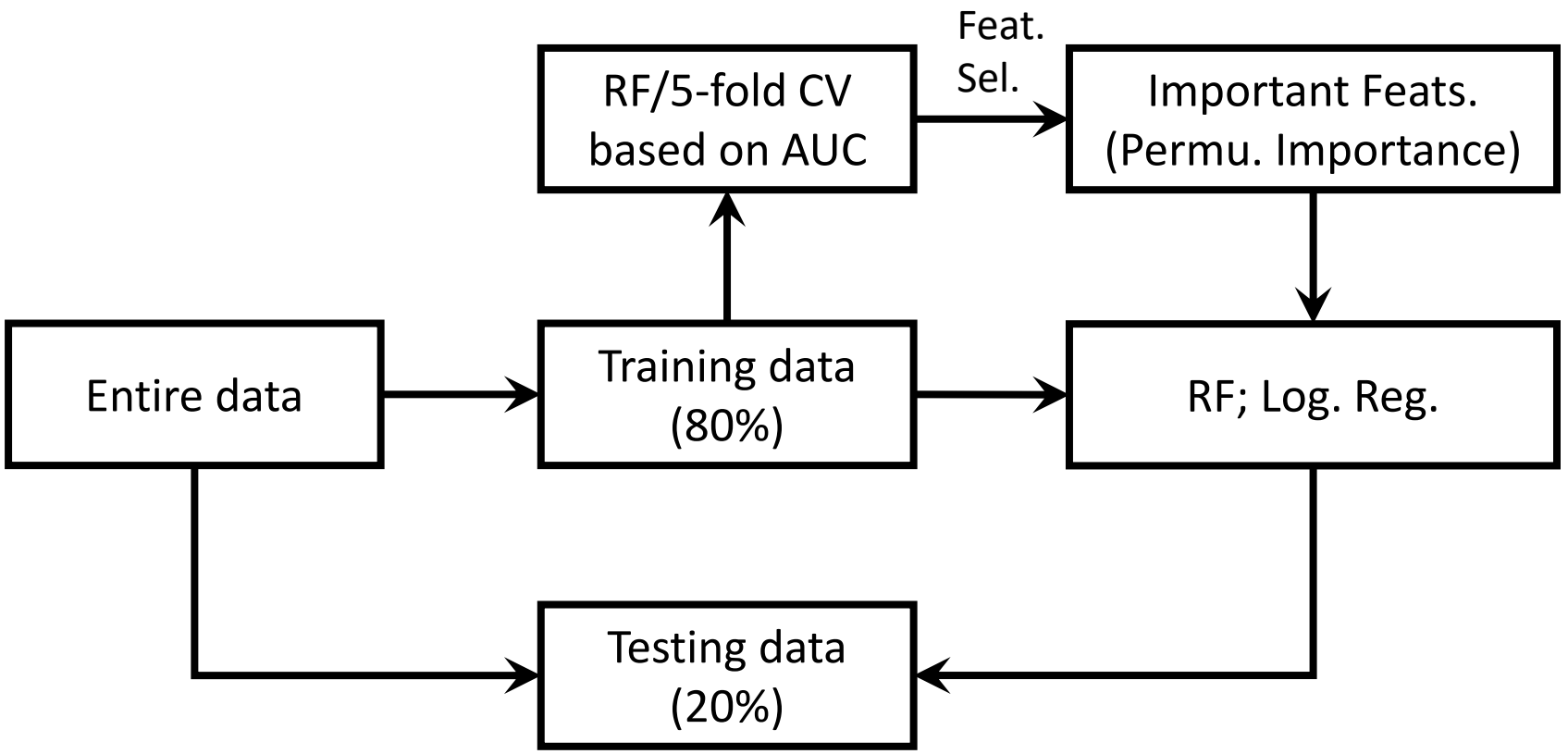
Clean data:

62310 applications:

- 51628 approvals
- 10682 denials

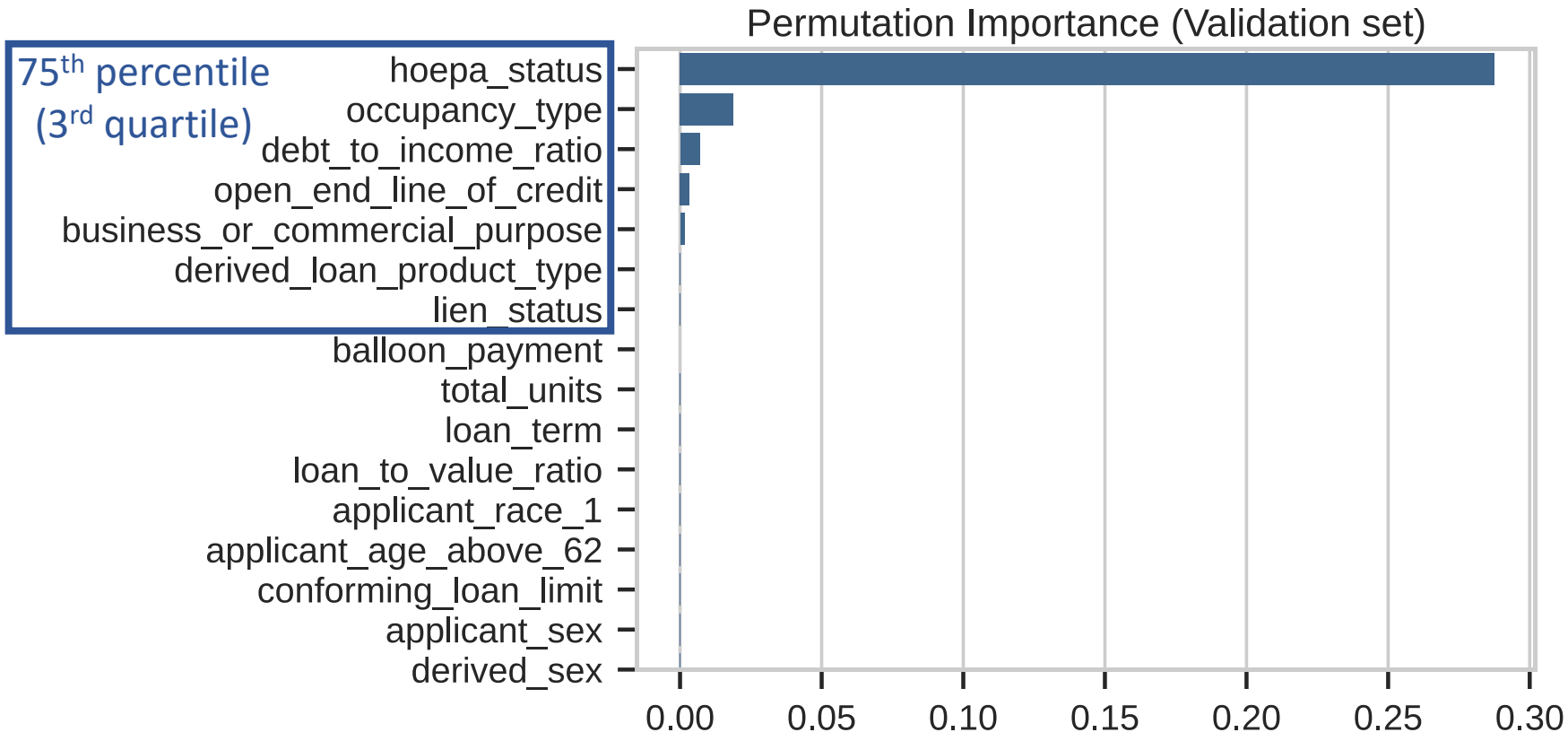
28 features

Modeling Workflow



Random Forest Feature Selection – 5-Fold Randomized Search CV

Tune hyperparams to maximize ROC-AUC (0.99):

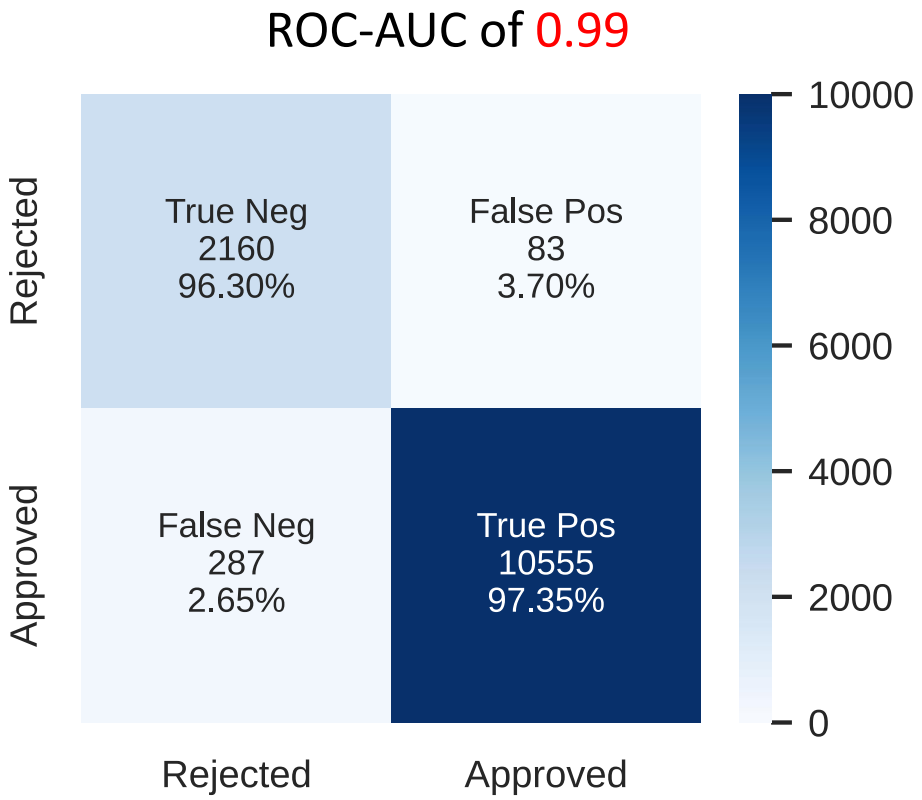


Need to know how these features play a role in decision making

Model Construction with Selected Features

All training data

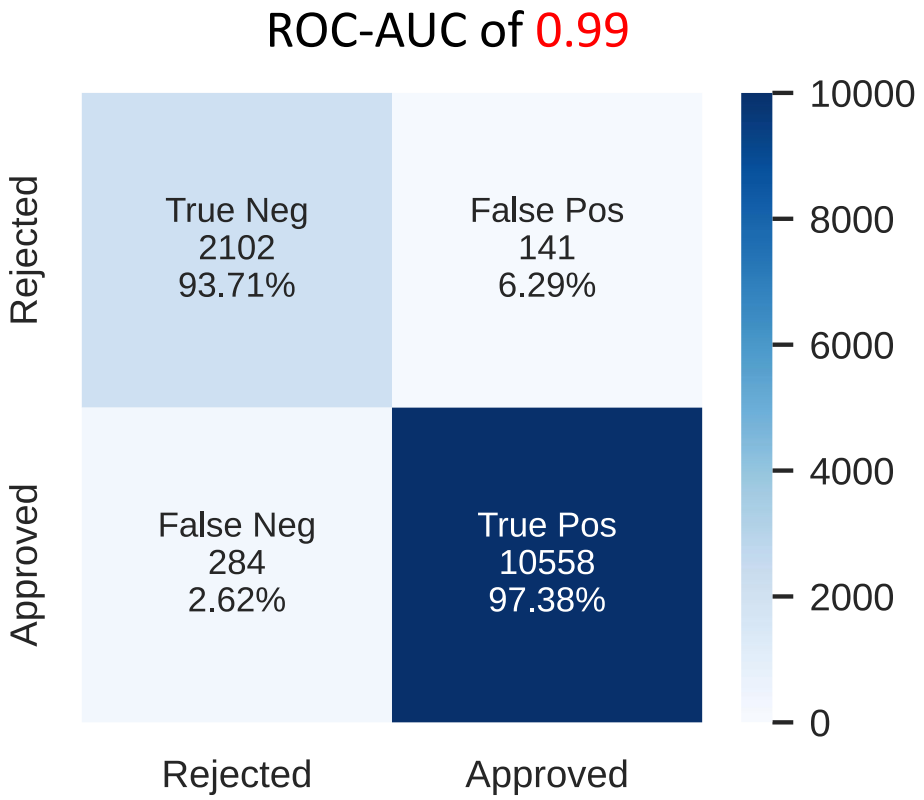
- Trained a random forest model



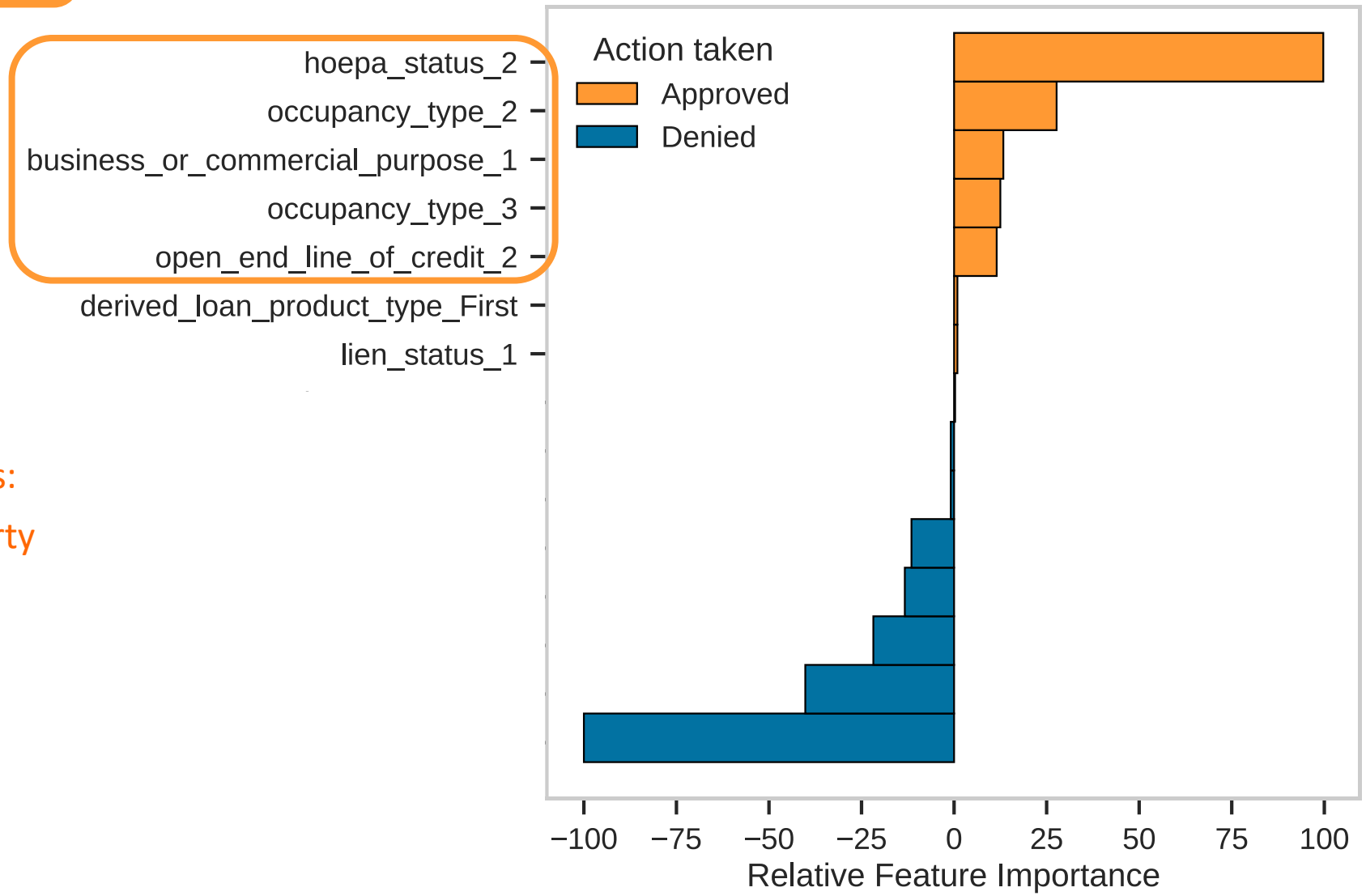
Model Construction with Selected Features

Existing training & validation data

- Tuned Log. Reg. model with 5-fold CV (regularization penalty strength, “C”)



More likely to approve



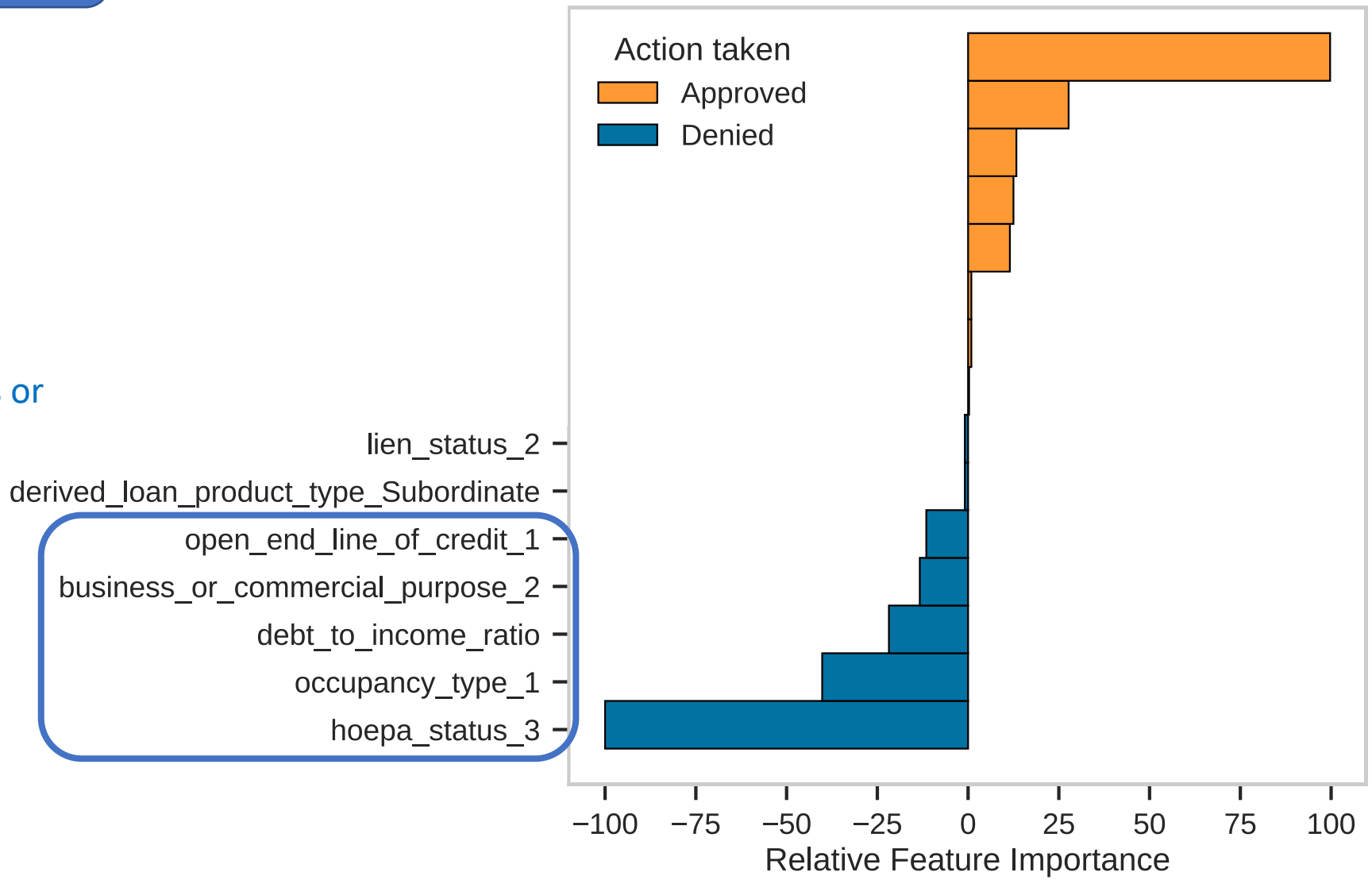
Approvals:

- Subject to HOEPA regulations:
- Second or investment property
- Primarily for a business or commercial purpose
- close-end line of credit

More likely to deny

Denials:

- Not subject to HOEPA regulations
- Principal residence
- High debt-to-income ratio
- Not primarily for a business or commercial purpose
- Open-end line of credit



Summary

- Understood how lenders make decisions on home loan refinancing
  - HOEPA Status
  - Occupancy Type
  - Business/commercial purpose
  - Whether if open-end line of credit
  - Debt to income ratio

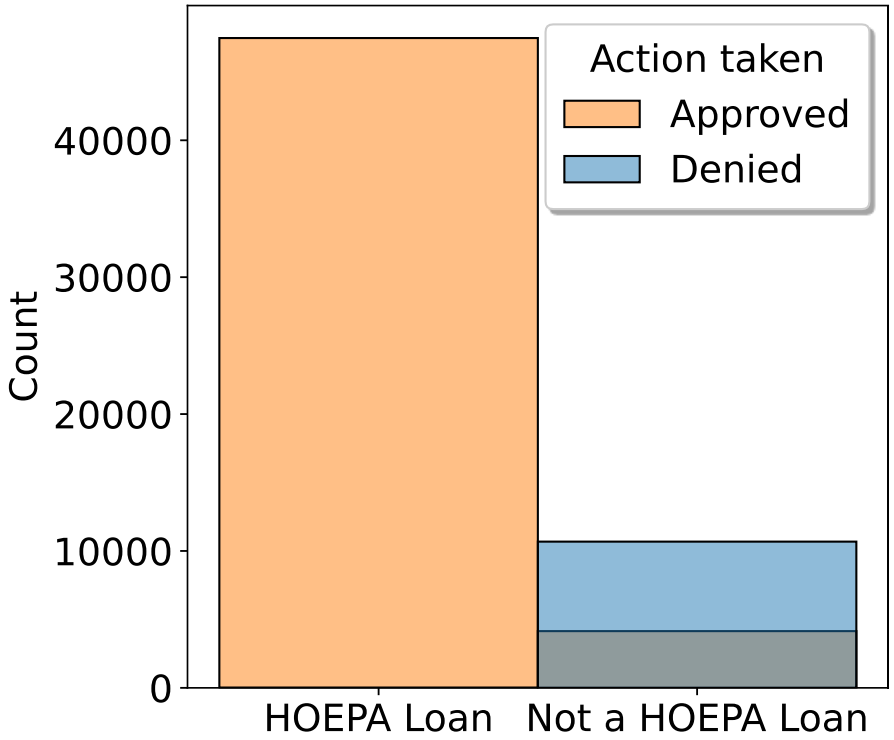
- Trained RF. and Log. Reg. models

Future work

- Investigate nationwide CFPB data
  - Including multiple loan purchases, e.g., home purchase, home improvement

- More advanced model
  - XGBoost
  - Deep neural network for more complex classification

Thank You

Appendix

HOEPA is highly separable, which explains why the AUC is as high as 0.99.