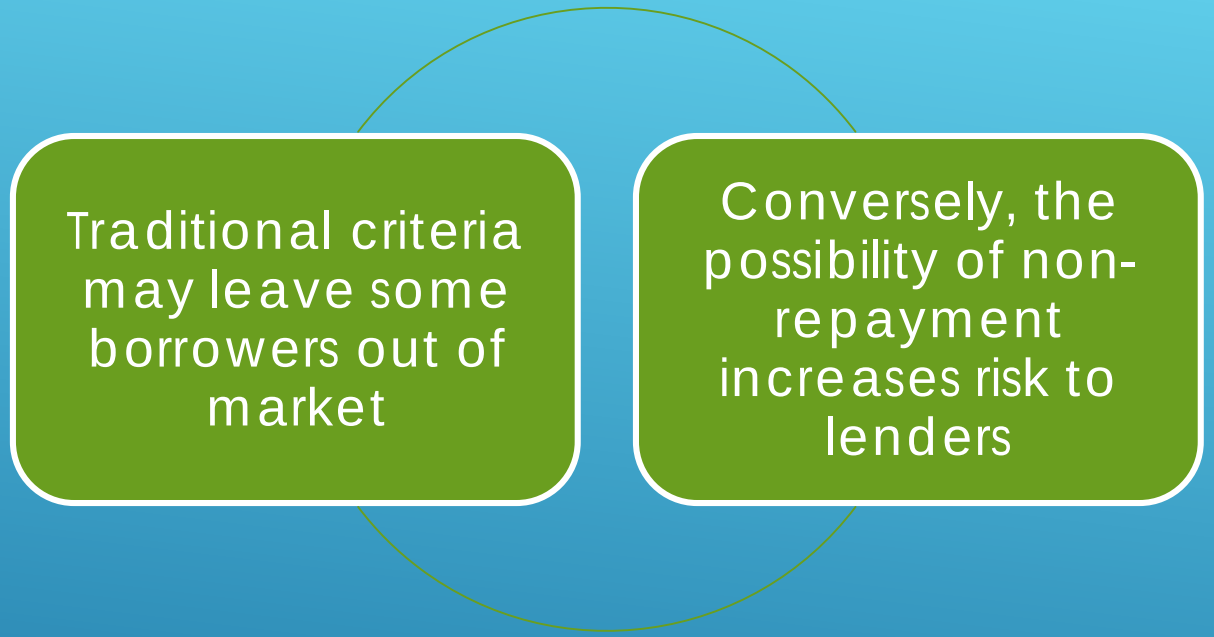
HOME CREDIT: FINAL REPORT

James Matheson
February 24, 2019

THE PROBLEM


Traditional criteria may leave some borrowers out of market
Conversely, the possibility of non-repayment increases risk to lenders

THE SOLUTION

USE MACHINE LEARNING ALGORITHMS TO SUPPLEMENT TRADITIONAL CRITERIA IN DETERMINING LOAD ELIGIBILITY

THE TASK

- Utilize Statistical Analysis on Past Client Data to Determine Which Client Data May Be Relevant To Clients' Repayment Ability
- Subject This Data to Machine Learning Algorithms to Develop a Predictive Model That Can Be Utilized to Assist in Determining Loan Eligibility

WHAT WE DID

- To determine the best algorhithm, we trained four different machine learning models:
  - Generalized Linear Model
  - Naïve Bayes Model
  - K Nearest Neighbor Model
  - Random Forest Model

WHAT WE FOUND

- Using A Random Forest Model Trained on 80% of the Test Data We Were Able to Achieve an Area Under a ROC Curve Score of 0.7297375

WHAT DOES THIS MEAN?

- This Represents the Ability for the Model to Distinguish Between Debtors Likely to Repay.
- A Score of .5 Indicates That A Model Is Equally Likely As A Random Decision to Predict a Debtor Who May Have trouble With Repayment
- A Score of 0.7297375 Represents a Significant Improvement Over Randomness

CONCLUSION

- Machine Learning Cannot Determine Debtors' Repayment Ability With Perfect Accuracy
- However, In Combination with Traditional Criteria For Determining Whether or Not to Make a Loan, the Right Machine Learning Model May Be Combined With Traditional Methods to Improve Upon These Methods, Making For Lower Risk Exposure