ML methods fail to adequately predict PPP loan forgiveness but succeed at classifying business based on their tax status.

Michael Lewis, Peter Magee, Ethan Wong, Margaret Willis
American University

Introduction/Context

- $525 Billion was disbursed as part of the federal government's Paycheck Protection Program (PPP). Some of these loans have been met with allegations of fraud. Today, the vast majority of the loans have been forgiven.
- Key Questions
  - How can we assess the relationship between Amount forgiven and other variables?
  - How can we best predict the Business Type?

Data

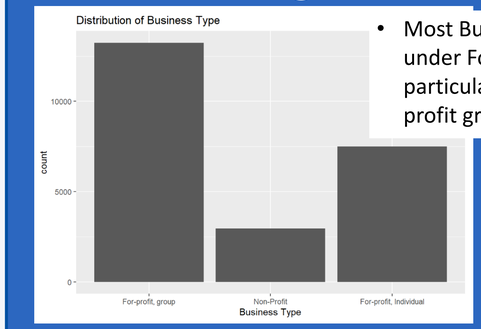
- PPP loan information for DC businesses
- ~23,000 rows, 2020 to present
- Work with same data as lenders
- Had to create `Business Type` variable` to streamline multiple levels

Approach

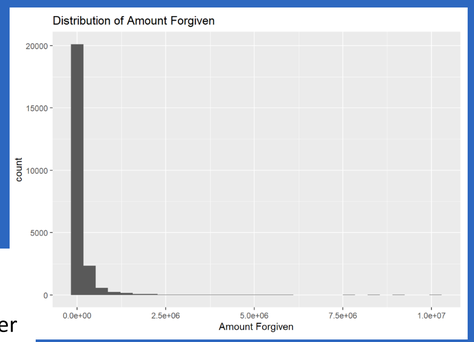
- Classification: Using KNN, LDA, QDA, and trees to predict business type.
- Regression: Using PCR, PLS, ridge, stepwise, and LASSO to predict the amount forgiven

Evaluation of PPP Loans Best Models

Understanding the Data



- Most Business fall under For-profit, particularly for-profit group



- Strong right skew
- Mostly smaller amounts forgiven

Best Models

Classification

Method	Error Rate
KNN	32.3%
LDA	35.76%
QDA*	55.88%
Tree	24.36%

*Removed `Race`, `Industry`, and `Age of Business` due to rank deficiency.

Regression

Method	MSE	Predictors
Stepwise	2106947766	12
Ridge	3777754837	40
LASSO	2002900584	4
PCR	2041329571	40
PLS	2040856908	11

Best models selected by having the lowest error. LASSO for Regression and Trees for Classification.

Selected Models

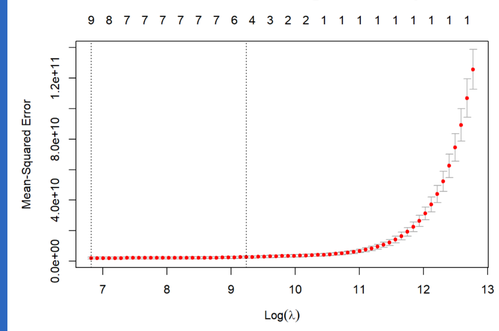
Tree

- Tree had 3 nodes which made decisions based on jobs reported and industry.
- Random first of 500 trees resulted in OOB error rate of 24.36%.

Classification Table

Yhat	For-profit, group	Non-profit	For-profit, individual
For-profit, group	4566	477	268
Non-profit	527	903	27
For-profit, individual	1494	91	3486

Cross Validation Curve along the λ Sequence



LASSO

- The Lambda.1se was 10260.3.
- Model has four predictors not including the intercept which are Payroll, Rent, Utilities, and Mortgage Interest.
- LASSO regression increases the accuracy of our predictions which is useful as our goal for this model is prediction.

$$\widehat{AmountForgiven} = 4845 + 0.959(Payroll) + 0.705(Rent) + 0.469(Utilities) + 0.242(MorgageInterest)$$

Results/Implications

- Most regression models suffer from high MSE that makes useful predictions impractical.
- The variables that were important between the two approaches differed (e.g., minority owned business, etc.)
- Using models to help build better loan forgiveness requirements
- Using our models to predict on more recent PPP loan forgiveness

Assumptions/Limitations/ or Secondary Results

- Assumed that the data collected by the SBA Is accurate for all loans
- Limited Data Availability
- Results not generalizable beyond D.C.
- Assumed Proxies were accurate representations of the enterprises
- Inherent barriers in filing for PPP forgiveness

Primary References

- House, The White. 2022. "FACT SHEET: New Data Show Progress on Biden-Harris Commitment to Equity in Emergency Small Business Relief." The White House. https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/11/fact-sheetnew-data-show-progress-on-biden-harris-commitment-to-equity-in-emergency-smallbusiness-relief/.
- Pfeiffer, Sacha, and Austin Fast. 2023. "How the Paycheck Protection Program Went from Good Intentions to a Huge Free-for-All." NPR, January. https://www.npr.org/2023/01/ 09/1145040599/ppp-loan-forgiveness. 3 "Some Firms Thrived During Covid and Then Got Their PPP Covid Relief Loans Forgiven." 2021. NBC News. https://www.nbcnews.com/news/firms-thrived-covid-got-ppp-covidrelief-loans-forgiven-rcna5697.