Gradient Boosting in OxyCotin Distribution

Joshua Gabino, Amrit Gill, Alex Jacobs, Riley Rizzo, Regan Schulman, and Burke St. Claire
Texas A&M University
April 24, 2024

Abstract

Gradient boosting can simplify analyzing non-linear relationships and high-dimensional data. Feature importance allows the inclusion of multiple variables to identify hidden correlations, strengthening a causal DiD model.

Introduction

We used Gradient Boosting to efficiently estimate the conditional average treatment effects (CATE) within subgroups by using Feature Importance to determine variables.

Literature Review

- Gradient Boosting is efficient in processing large volumes of pharmacy data (Paperspace; Analytics Vidhya).
- Limited research on ML systems for immediate intervention in drug abuse
- Addressing gaps can help the opioid epidemic in pharmacies

Methodology

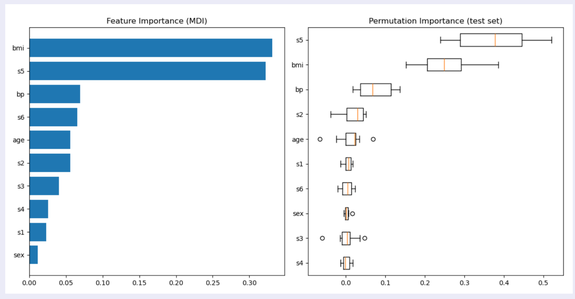
We replicated an OLS regression of DiD design to estimate OxyCotin dispensing. Using scikit-learns GradientBoostingRegressor model, we compare feature importance to and plot them against relative importance

Findings

We found the feature with the most importance was feature and the least was feature

Table: Regression Results Summary

Variable	OLS	Feature
DPre	5.099	xxx
Dpost	-9.303	xxx
Chain	-8.362	xxx
Constant	32.036	xxx
Observations	5,055,761	
R-squared	0.003	
Other features: zipcode, competition, all opiods		



Discussion

Our discussion

Conclusions

Using Gradient Boosting, researchers can limit intensive analysis of multiple regressors by using Feature Importance to determine the most relevant influences on the outcome.

References

Figure: Citations listed in QR Code

Appendix

Include supplementary material like detailed statistical analyses, additional graphs, etc.