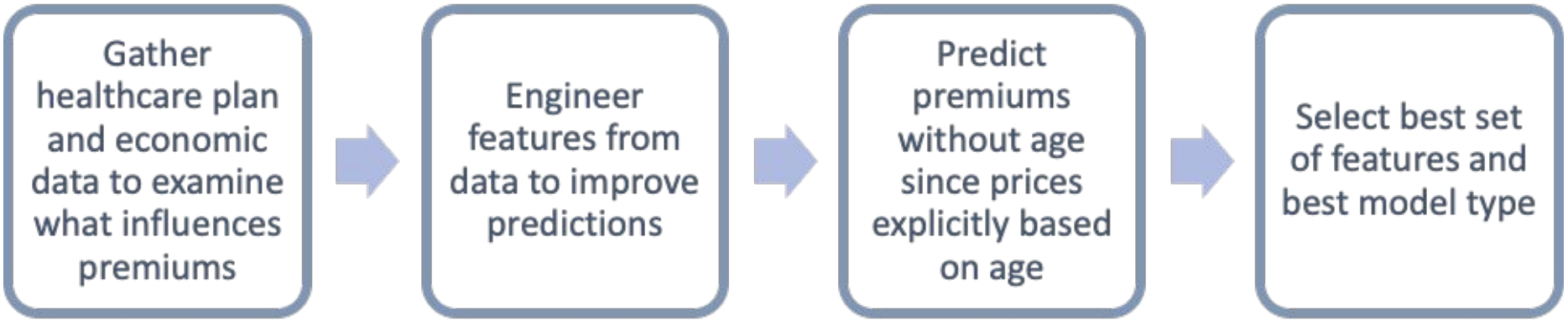
Predicting Iowa Healthcare Premium Prices

Luke Von Glan

Results

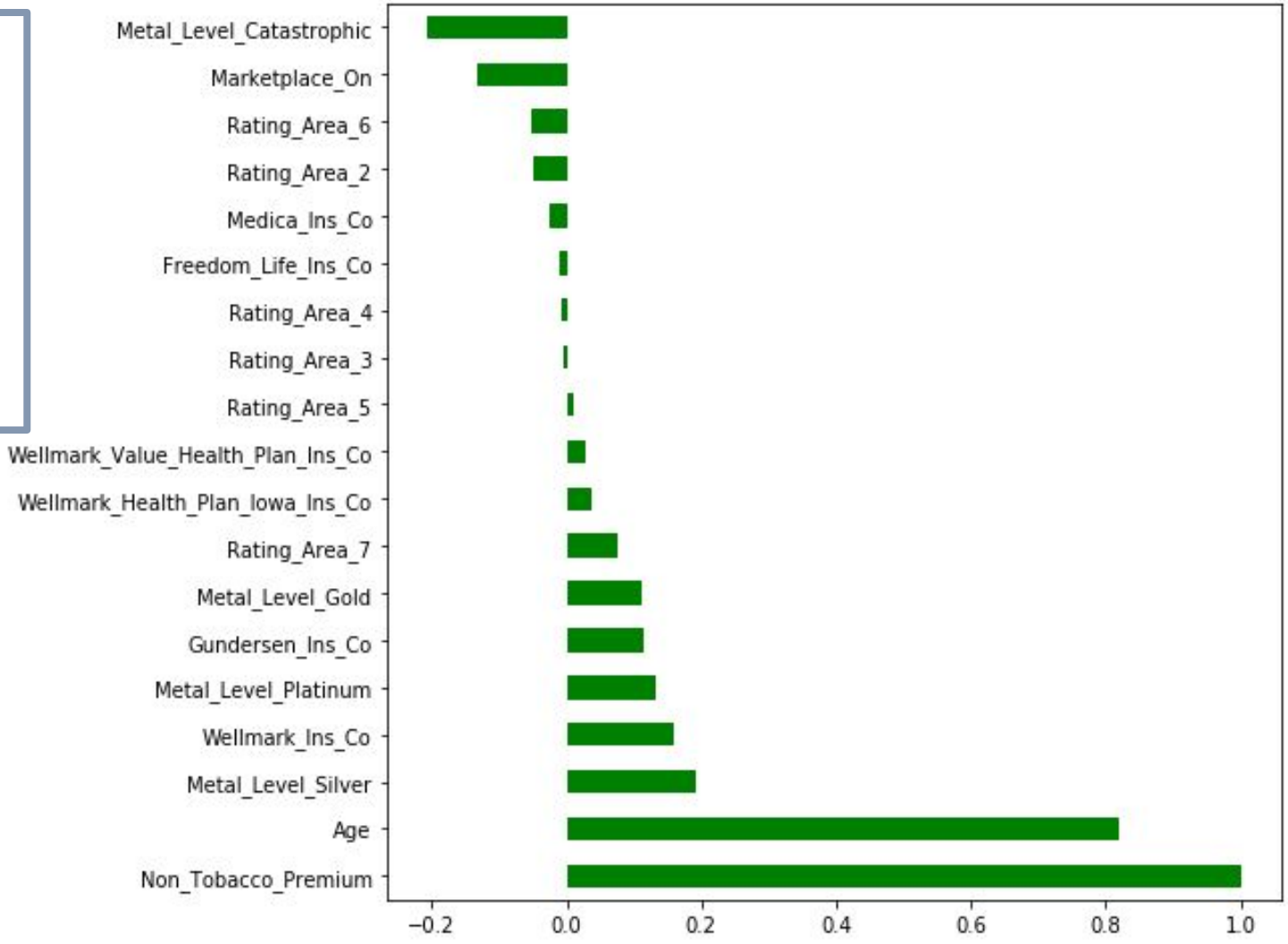
- Age is the overwhelming driver of healthcare premiums
- Economic data is unrelated to premium prices
- More data out there to better predict prices

Process


Gather healthcare plan and economic data to examine what influences premiums
Engineer features from data to improve predictions
Predict premiums without age since prices explicitly based on age
Select best set of features and best model type

Healthcare Premium Correlations

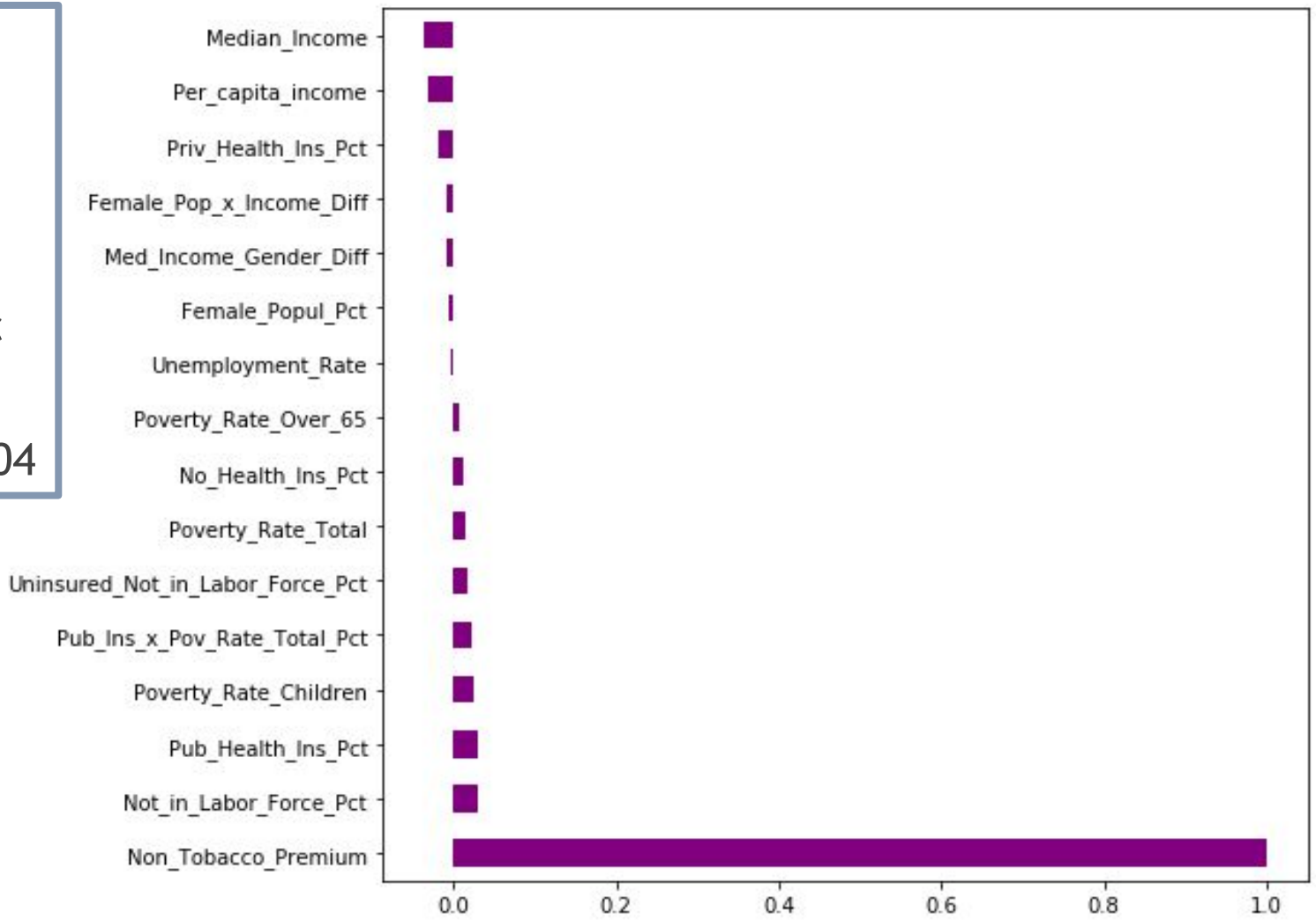
Healthcare plan data



Healthcare Premium Correlations

Census economic data

No correlation>.04



Final Model

- Included variables for insurance company, location, and plan quality
- No census data features improved model so none used
- Final model $r^2 = .17$
- If age included: model $r^2 >= .81$

Takeaways

- Model is only as predictive as the data you give it
- Healthcare plan data set has more data worth investigating
- Factors influencing individual premiums are not as complex as many believe