Income Classifier

Zach Brown, Albert Frantz, David Cortes

Table of Contents

Problem Statement

Build a model optimized for accuracy to predict if a person’s income is greater than or less than $50,000, limited to 20 features.

Cleaning and EDA

- Limited to 20 features
- Imputed the mode for missing values
- Engineered terms with linear relationship to income
- Chose most highly correlated terms, dropped collinear terms

Modeling

- Random Forest
- Logistic Regression
- SVM
- XGBoost
- Gaussian Naive Bayes

Results

XGBoost

Best Params:

max_depth = 2

max_features = None

n_estimators = 100

Best Score: 85.5%

Conclusion

The XGBoost model produced a score of 85.5%. Compared to the baseline accuracy of 76%, this model resulted in satisfactory predictions.

Thank you!

Questions?