

DECISIONS

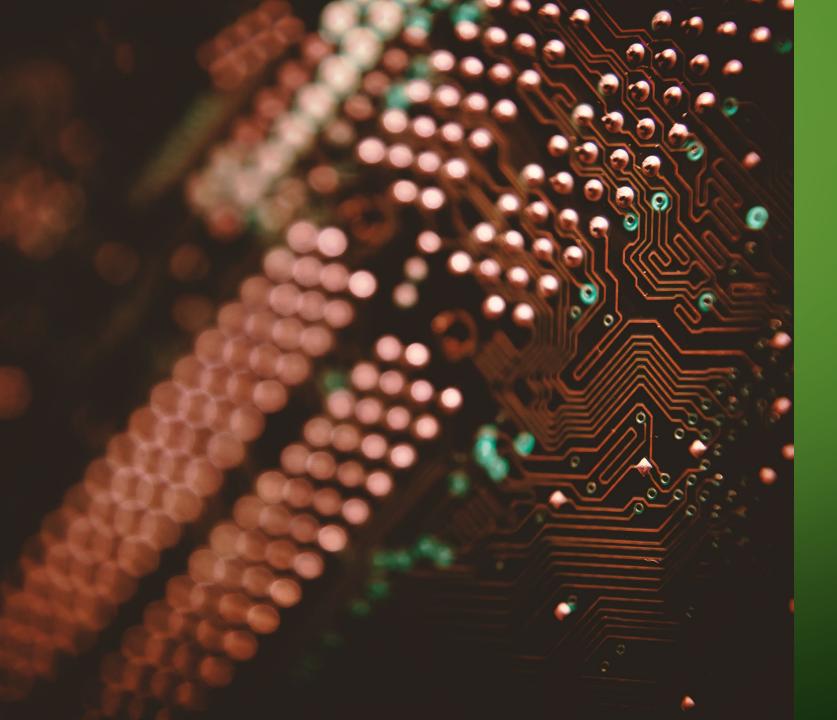
- Crossroad about going to college.
- Is the income after college worth it?
- Created a predicting model to forecast annual income for individuals with a college education.

VARIABLES

- •Target Variable: Income >50K
- •Education: Highest level of education individual achieved
- Occupation: Current work position
- •Race
- •Gender
- •Relationship: If individual is a husband, wife, single, etc.
- •Hours worked per week

A TOOL FOR EDUCATIONAL INSTITUTIONS

- With the predictive model, institutions can show students or potential students that in the long run a college education is the best route.
- The model will try to predict if the individual has the potential to make more than 50 thousand a year.
- Shows individuals a forecast of the future.



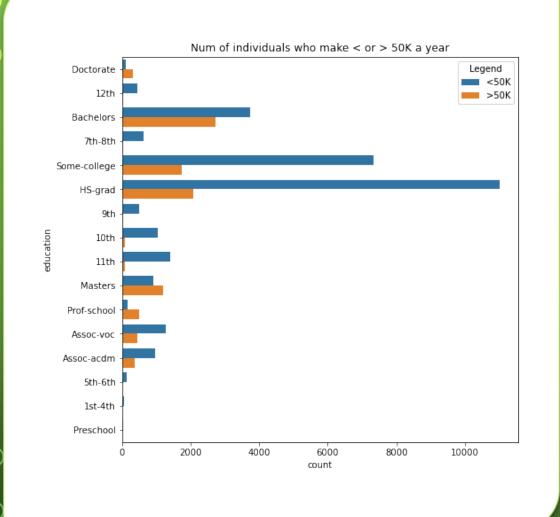
COLLEGE ENROLLMENT RATE DOWNFALL

- Nearly one million less students enrolled in college in 2022 then before 2020.
- From 2019-2021
 colleges seen
 undergraduate student
 enrollment fall by 6.6%.
 - Biggest decline in more than 50 years.

Americans choose jobs over college: NPR

SHORT-TERM BENEFITS VS. LONG-TERM BENEFITS OF A DEGREE.

- According to Georgetown University, bachelor's degree holders were found to "earn a median of \$2.8 million during their career, 75% more than if they had only a high school diploma."
- Many high school students go for short-term money rather than a long-term career and high wages.



DATA UNDERSTANDING

Machine learning model:

- Demonstrates who is making more money annually.
- Predictor takes in attributes such as education, relationship, hours worked, etc.
- In Dataset 9.6K/39.4K individuals made over 50K

Num of individuals who make < or > 50K a year Doctorate Legend 12th Bachelors 7th-8th Some-college HS-grad 10th 11th Masters Prof-school Assoc-voc Assoc-acdm 5th-6th 1st-4th Preschool 2000 4000 6000 8000 10000 count

DATA UNDERSTANDING CONT.

- Georgetown University study, HS-Grads are the ones earning less than 50K a year.
- There's more than 50% HS-grads that make less than 50K then someone that has their bachelor's degree.

Confusion Matrix - 6000 <50K 2428 - 5000 6489 Frue labels - 4000 - 3000 >50K - 2000 408 2504 - 1000 <50K >50K Predicted labels

PREDICTIVE MODEL

- Random Forest- is learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.
- Accuracy on dataset: individual's who make <50K: 82%
- Accuracy on dataset: individual's who make >50K: 64%

NEXT STEPS • Bigger dataset Handle imbalanced dataset • Refine model performance



THANK YOU, QUESTIONS?