The Old College Try:

Predicting Future Recipients of Bachelor's Degrees

Andrew Zhou
28 Oct 2020

Motivation and Objectives

Outreach

- Effective, early targeting for organizations
- Inform youth of
- opportunities

Fairness

- Avoid algorithmic biases
- Encourage diversity and representation

INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
PANEL STUDY OF INCOME DYNAMICS

Our Data

PANEL STUDY

1968 to present

24000 individuals, 1000 families

FROM CHILDHOOD

Children born into study

2002 survey, 2900 subjects

PRESENT DAY

Between 22 and 33

Bachelor's by 2017?

Features

Predicting Future College Graduation from Childhood Data

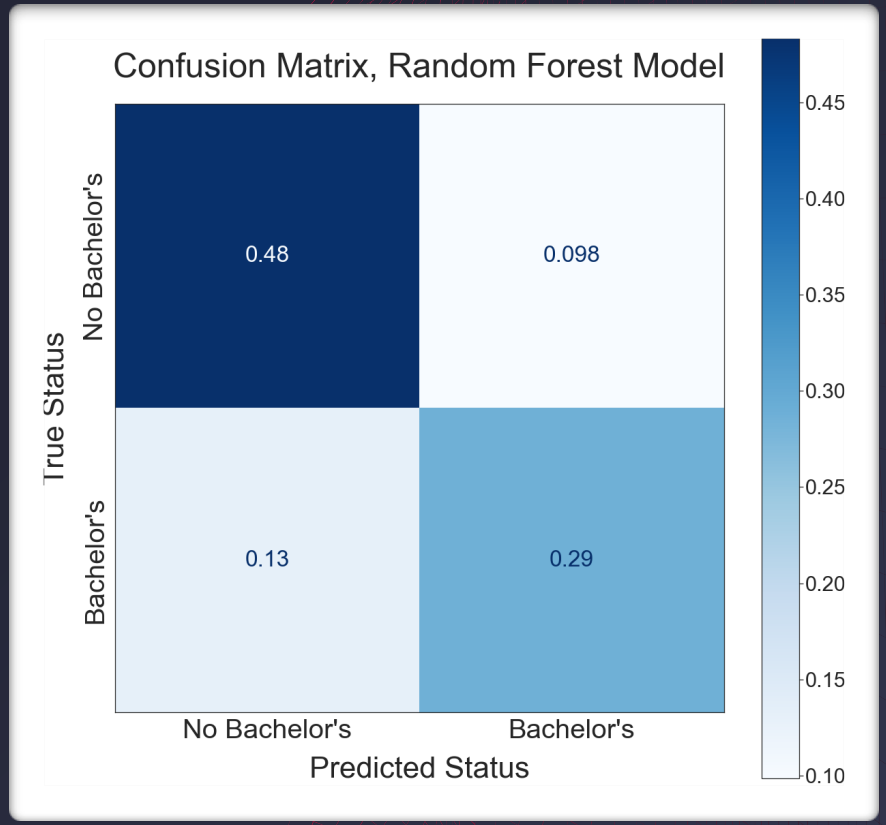
Graduation Prediction | Feature Exploration | Model Performance

Welcome! Please choose a page!

Model Metrics

F-score

- Focus on graduates, penalize false positives
- 0.744 precision, 0.686 recall
- F0.5 of 0.732



Model Performance

	$F_{0.5}$	Precision	Recall
Random Forest	**0.732**	**0.744**	**0.686**
Logistic	0.701	0.731	0.604
SVC	0.671	0.695	0.588
KNN*	0.703	0.757	0.546

* Data not weighted by survey weights

Fairness

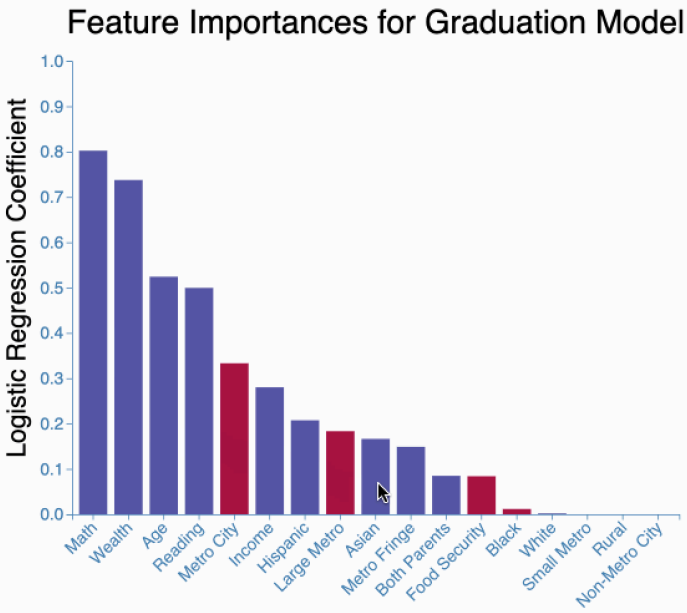
We can use race, gender, etc.

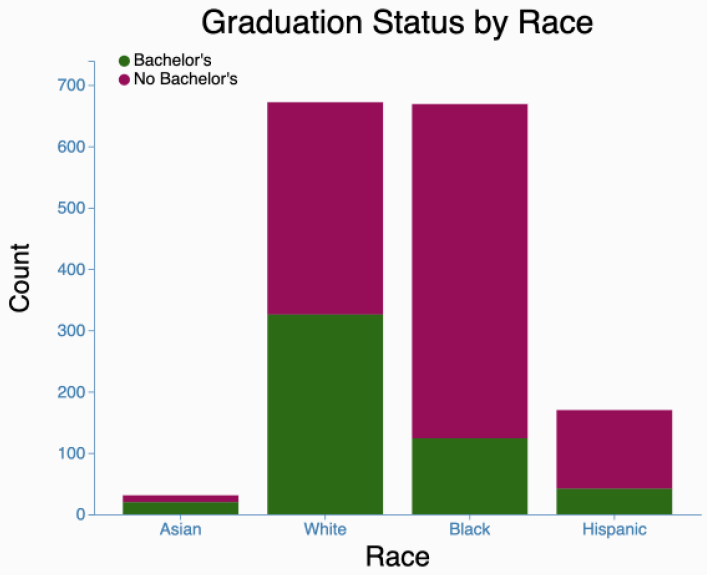
Should we?

Predicting Future College Graduation from Childhood Data

Graduation Prediction | Feature Exploration | Model Performance

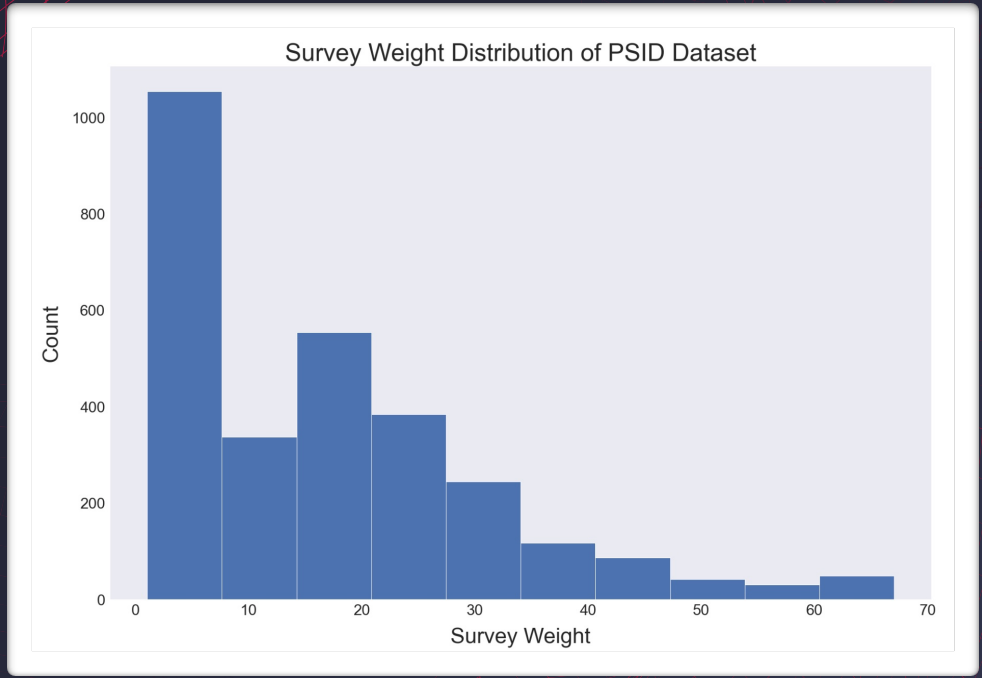
Click a feature on the left chart to show its relationship with graduation on the right.





Survey Weights

- Ensure representative sample
- Weight both training and validation
- Survey oversamples, training weights



Prediction

Predicting Future College Graduation from Childhood Data

Graduation Prediction | Feature Exploration | Model Performance

Family Status (2001)

Wealth

50000

0 250000

Income

50000

0 250000

Living Environment

Large Metro

Food Security

High

Lives With:

Two Parents

Test Scores

Math Percentile

50

0 100

Reading Percentile

50

0 100

Demographics

Age

28

22 35

Race

White

We predict that this student has a

24%

chance of receiving a bachelor's degree.

Future Work

Time Analysis

Data are inherently time-based

Set fixed age, drop feature

Fairness

Community ethics

Black-box algorithms

Survey Weights

Train and test set selection

Mathematical rigor

Thank you!

Questions?

Credits

- Presentation template from slidesgo.com
- Thanks to all the wonderful folks at Metis!