

Predicting American Salary

Jake Cho

Background

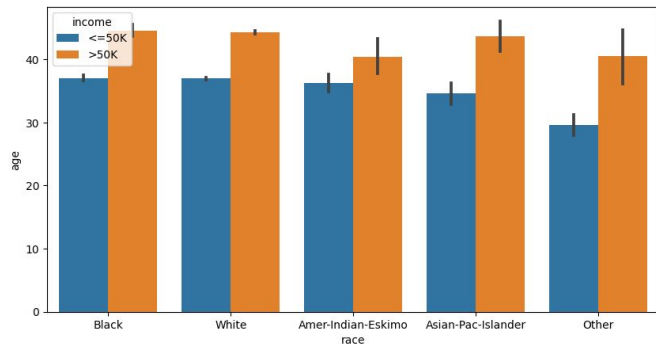
Stakeholder: Business targeting people above 50k/need to see distribution of people making more than \$50k

Business Target: people with salary > \$50k

Data Information

- Variables and prediction
 - Age: Older age increases likelihood of making more than \$50k
 - Educational years: More school years increase likelihood of making more than \$50k
 - Race: Certain race has higher likelihood of making more than \$50k
 - Gender: Certain sex has higher likelihood of making more than \$50k
 - Hours per week: More hours increase likelihood of making more than \$50k
- Result
 - Income: Greater or less than \$50k

Figure 1

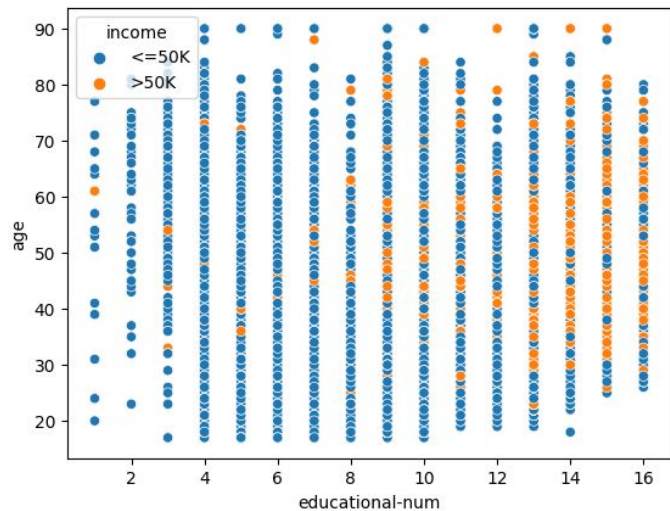


Interesting trends to notice

Average age is higher for ones who make more than \$50k

For every race, there are more people who make more than \$50k than those who don't

Figure 2



Interesting trends to notice

People with longer education have higher tendency to make more than \$50k

There is no clear trend on age

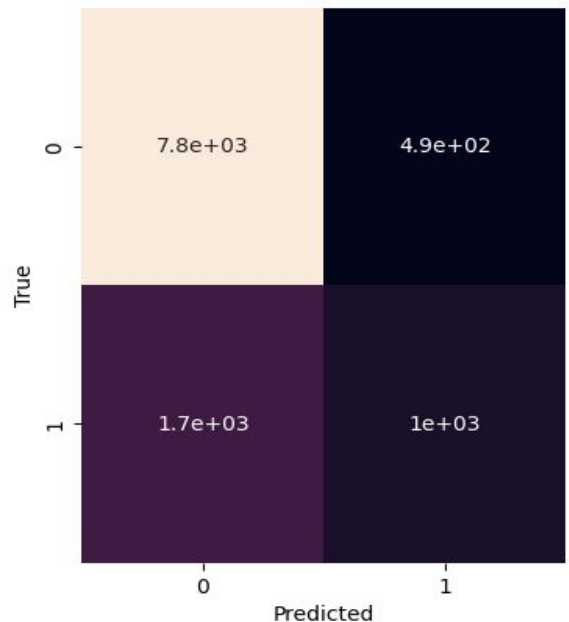
Models

Final: Decision Tree w/ GridSearchCV

False positive: top right, predict \$50k, but subject does not

False negative: bottom left, predict sub \$50k, but subject makes more than \$50k

Impact? Depends



Test Report:		precision	recall	f1-score	support
0	0.82	0.94	0.88	8260	
1	0.67	0.37	0.48	2688	
accuracy			0.80	10948	
macro avg	0.75	0.66	0.68	10948	
weighted avg	0.78	0.80	0.78	10948	

Recommendation

The model works, but it can definitely improve

Improvements

- More variables
- Numerical data for income

