May 2023

Adult Income Analysis

Predicting Income Levels Based on Personal Features

Hannah Ploutz

Understanding Our Stakeholder

Our stakeholder is the marketing department for a firm and aims to understand the income level of adults based on their personal features

This analysis will assist in designing effective marketing strategies

About the Data

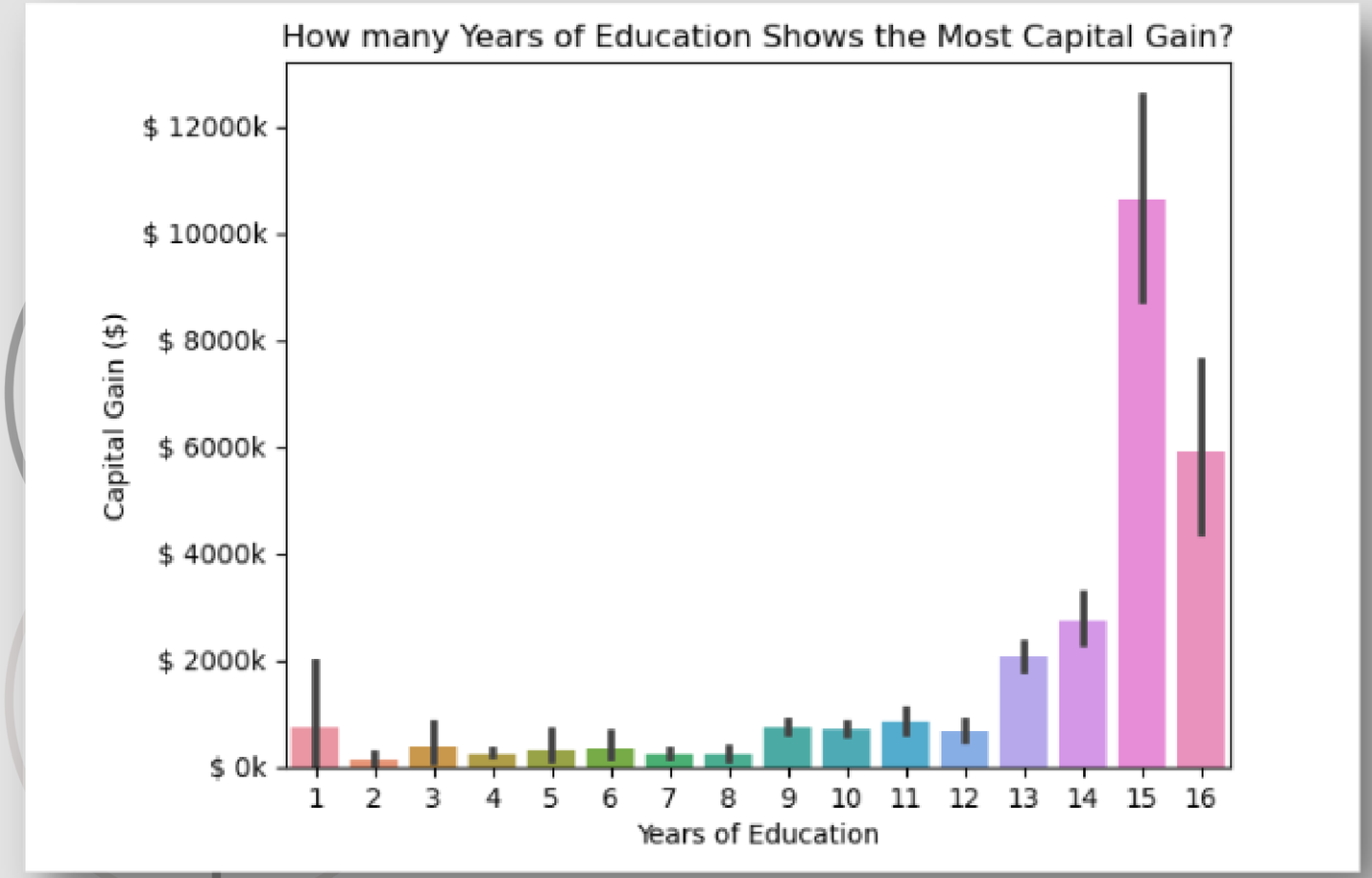
The data is comprised of 48,842 adults with features including education, marital status, race, and gender

The target variable is income, categorized as >50k and <=50k

Key Finding #1: Education and Capital Gain

A correlation exists between years of education and capital gain

Increased education levels tend to lead to higher capital gain, dropping off at 16 years of education



Key Finding #2: Gender and Income Level

Gender also correlates with income level, with males typically earning more than females

The disparity in income distribution between genders is significant

Our Predictive Model: Ada Boosted Model

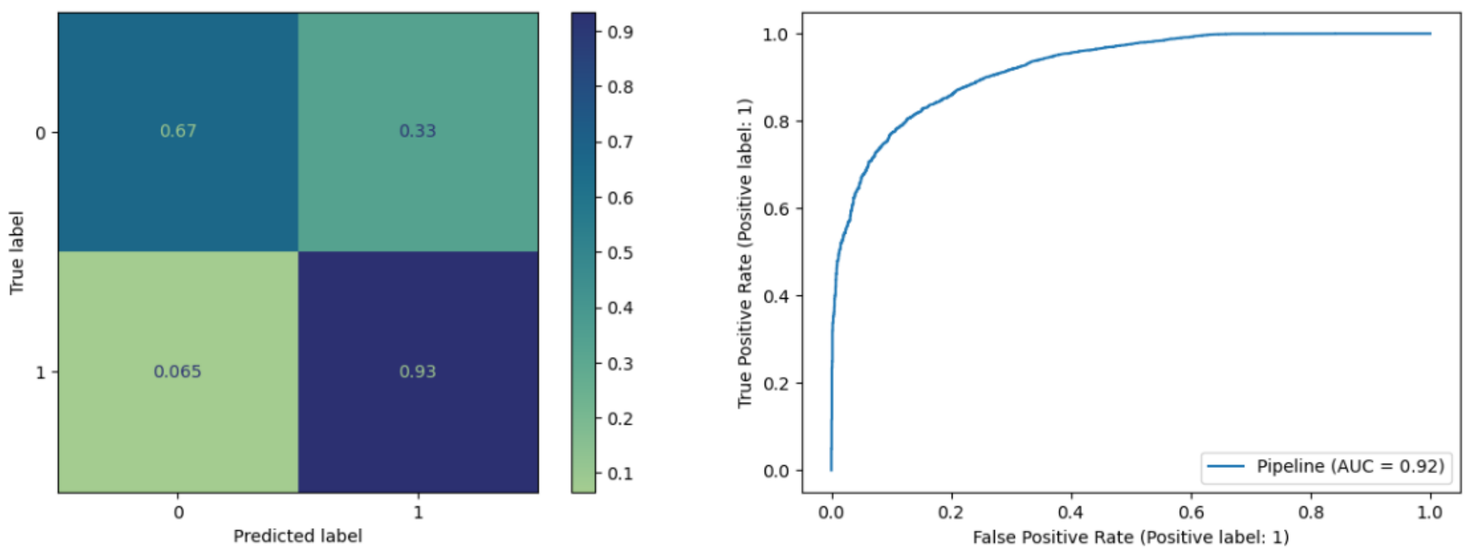
Our chosen predictive model is an Ada Boosted Model with Tuned Hyperparameters

Compared to other models such as Random Forest and KNN, this model yielded the highest metrics

```
------------------------------------------------------------
[i] CLASSIFICATION REPORT FOR: Best Ada Boost Test Data
------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.79      0.67      0.72      2250
           1       0.88      0.93      0.91      6113

    accuracy                           0.86      8363
   macro avg       0.84      0.80      0.82      8363
weighted avg       0.86      0.86      0.86      8363
```



Understanding Our Model's Performance

The model is 86% accurate, performing best for predicting incomes <=50k

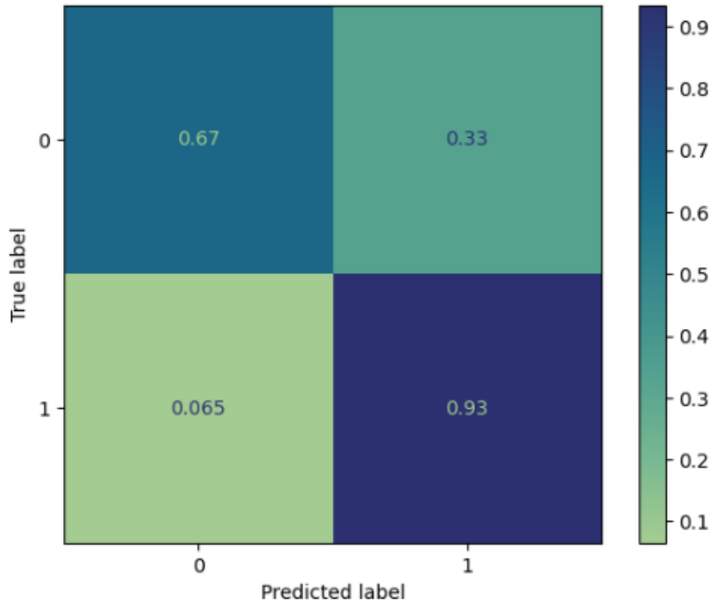
The model has a 93% success rate for predicting <=50k incomes and a 67% success rate for predicting >50k incomes

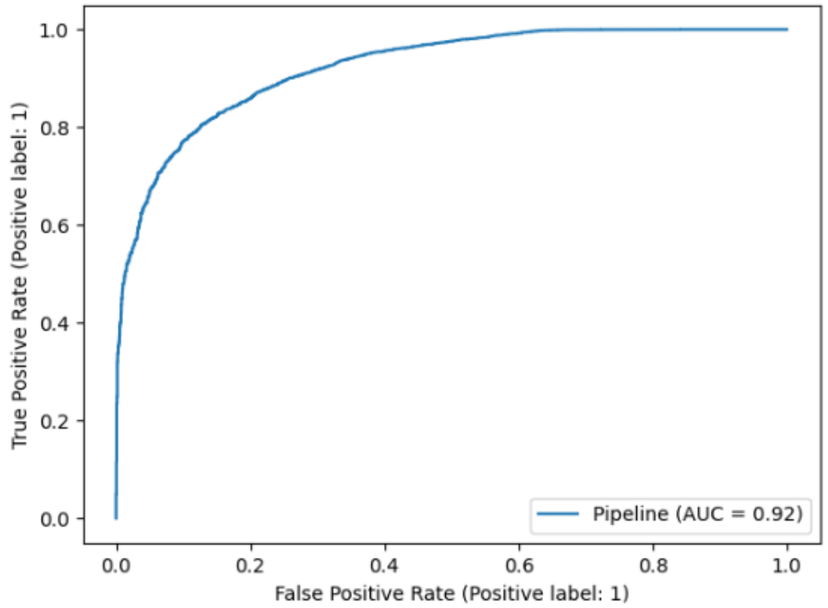
---

[i] CLASSIFICATION REPORT FOR: Best Ada Boost Test Data

---

	precision	recall	f1-score	support
0	0.79	0.67	0.72	2250
1	0.88	0.93	0.91	6113
accuracy			0.86	8363
macro avg	0.84	0.80	0.82	8363
weighted avg	0.86	0.86	0.86	8363





Recommendations and Limitations

Recommendation

Use this model for real-world analysis, especially for adults earning under 50k

Limitation

The model's data is primarily from the US and consists predominantly of white individuals

May 2023

CONCLUSION