Predicting Wages

Jenny James, Lydia Kajeckas, Nikhil Lonberg, & Tony Lucci

Agenda

1. Problem Statement
2. Exploratory Data Analysis
3. Modeling
4. Proposal

Our predatory lending fund is running out of vulnerable individuals to prey on. We need to determine which individuals would likely default on a payday loan. Using the 1994 census data, let's build a model to identify individuals with an income of less than $50,000 a year.

To save time and money, we will make a simple model using no more than 10 features.

Feature Engineering

New features created:

- Higher education: more than High School

- Is USA: from the United States

- Top 7 occupations: Executive Manager, Professional Specialty, Protective Service, Tech Support, Sales, Craft Repair, Transport Moving
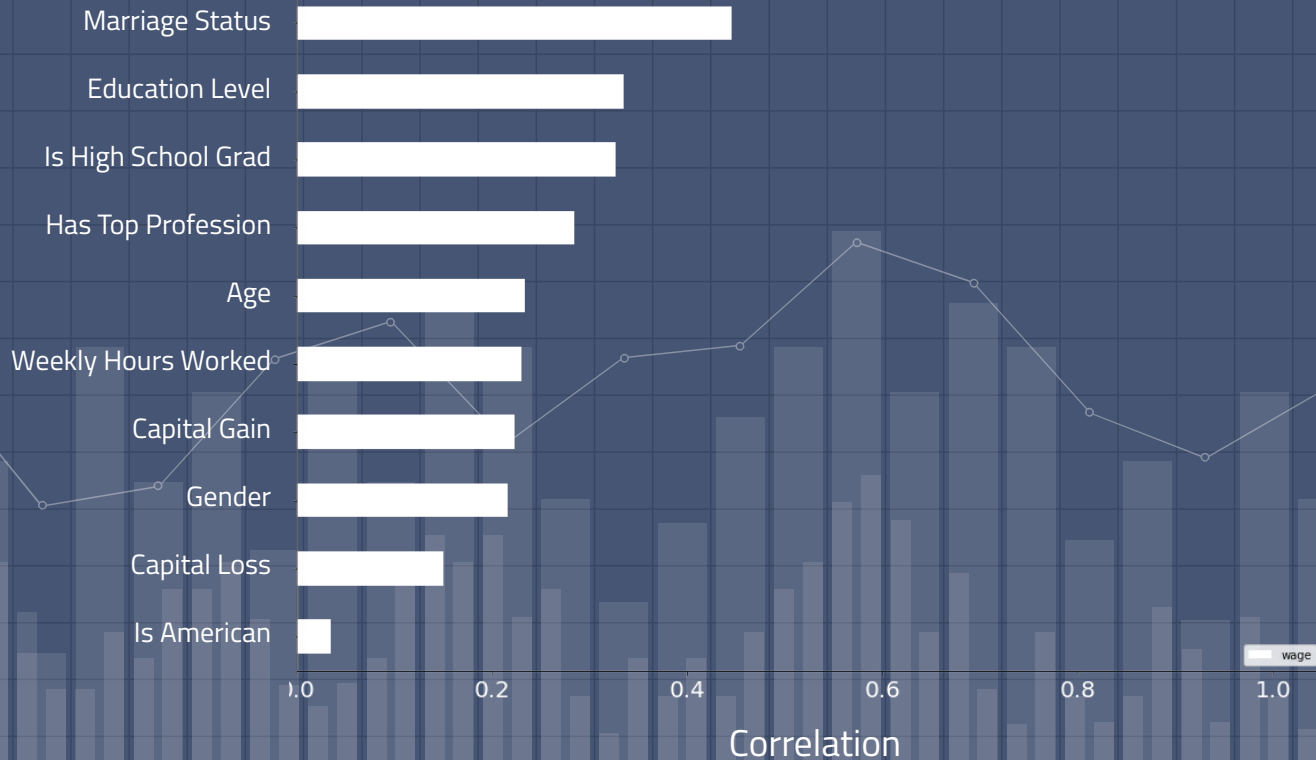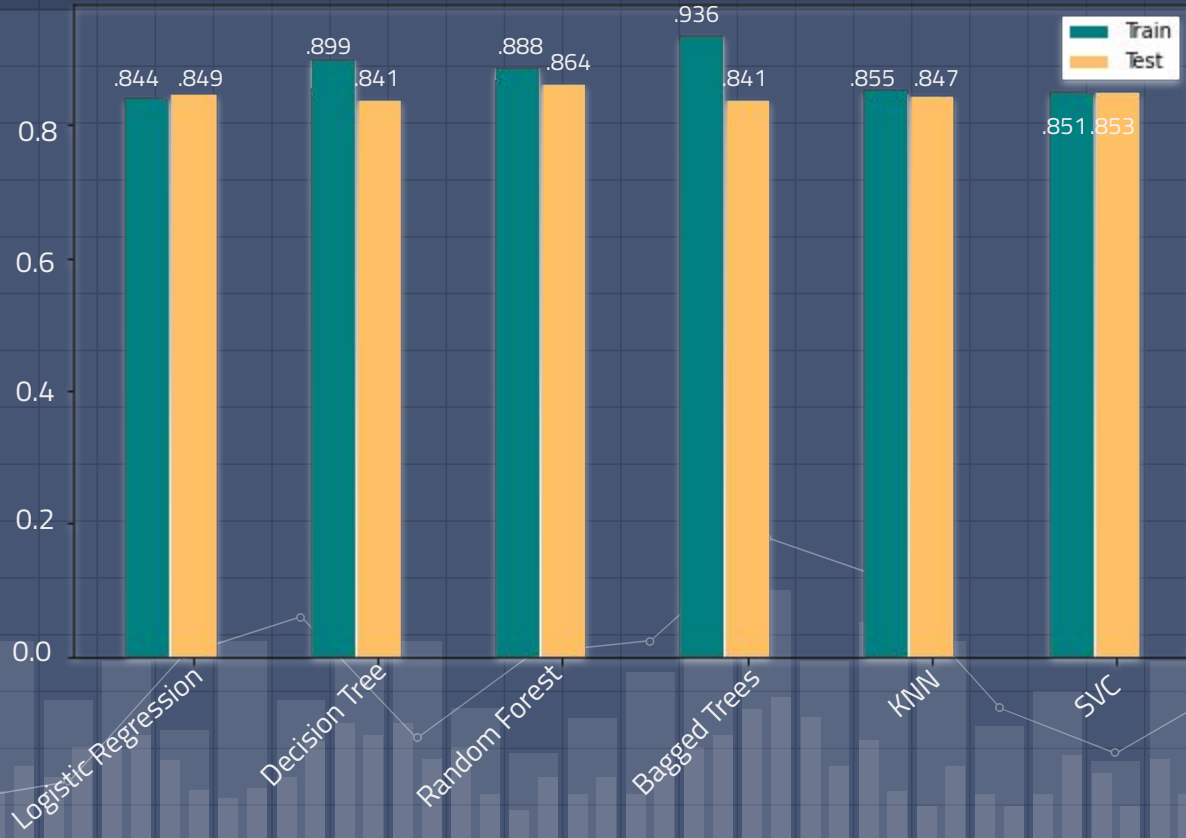
- Gender dummies

Marriage Status
Education Level
Is High School Grad
Has Top Profession
Age
Weekly Hours Worked
Capital Gain
Gender
Capital Loss
Is American

0.0   0.2   0.4   0.6   0.8   1.0

Correlation

wage

Proposal

Machine Learning can help us do evil.

Random Forest Algorithm generalizes to new data best.

Simple is better. Our model predicts income class on testing data with 86% accuracy using only 10 features.

Sensitivity: 63%
Specificity: 94%
Precision: 77%

SOURCE

https://fee.org/articles/without-vulture-capitalists-our-economy-would-rot/

https://en.wikipedia.org/wiki/Vulture_capitalist

https://en.wikipedia.org/wiki/Predatory_lending

https://en.wikipedia.org/wiki/Payday_loan

END