Matt Frei

Executive Summary

An analysis of census data reveals important associations between income and other personal information. The simple model presented as Figure 1 correctly classifies 84% of US residents as high-income (greater than $50,000 annually) or low-income using only three variables. The model reveals that married persons with a spouse present are almost 8 times more likely to be high-income than others. Graduation from college and the presence of significant capital gains revenue are other important indicators. Almost two-thirds of married college graduates (72%) earn at least $50,000 per year. These factors are meaningful even in a logistic regression that holds constant race, age, sex, immigration status, weekly hours worked, occupation, and other factors.

The logistic regression analysis also provides insight into additional important factors. For example, African Americans have 65% lower odds of being high-income than Caucasians, even when accounting for the other factors discussed here. Similarly, women have about 13% lower odds than men and immigrants' odds are 30% lower than for those native to the US. The relationship between age and income is somewhat more complex, however. Those between 50 and 64 have the highest probability of earning at least $50,000 per year while those in the oldest and youngest groups (less than 30 or older than 65) are the least likely to be high-income.

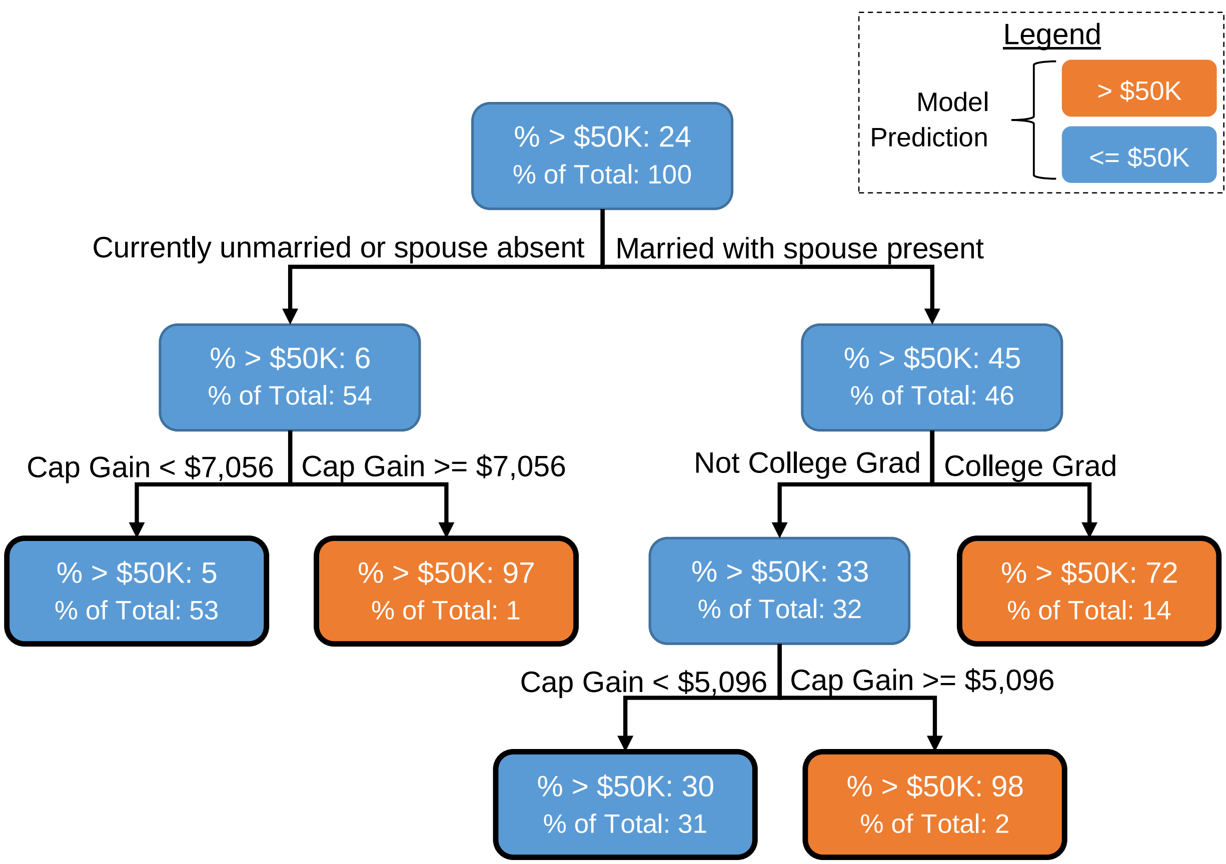

Figure 1: Decision Tree Income Classification Model