Learn About Random Forest in Python With Data From the Adult Census Income Dataset (1996)
Learn About Random Forest in Python With Data From the Adult Census Income Dataset (1996)

About This Dataset

Data source citation


Full title of originating dataset

Adult Census Income

Data author(s) and affiliations

Ronny Kohavi (Data Mining and Visualization, Silicon Graphics)
Barry Becker (Data Mining and Visualization, Silicon Graphics)

Dataset source website address

https://www.kaggle.com/uciml/adult-census-income

First publication date

2016

Data Universe

Civilian, noninstitutionalized individuals 16 years or older from a nationally representative sample in the
Weighting

The weights on the Current Population Survey (CPS) files are controlled to independent estimates of the civilian noninstitutional population of the US. These are prepared monthly by Population Division at the Census Bureau. Three sets of controls are used. These are:

1. A single-cell estimate of the population 16+ for each state.
2. Controls for Hispanic origin by age and sex.
3. Controls by race, age, and sex.

All three sets of controls are used in the weighting program and “rake” through them 6 times so that by the end, we come back to all the controls we used. The term estimate refers to population totals derived from CPS by creating “weighted tallies” of any specified socioeconomic characteristics of the population. People with similar demographic characteristics should have similar weights. There is one important caveat to remember about this statement: Since the CPS sample is actually a collection of 51 state samples, each with its own probability of selection, the statement only applies within state.

Data collection dates

1994

Unit of analysis

Individual

Location covered by data

United States

Other sources


List of variables
income
Annual Income Class

workclass
Work class

age
Age

education
Education

education.num
Education (numeric)

marital.status
Marital status

occupation
Occupation

relationship
Status in household

race
Race

sex
Gender

capital.gain
Capital gain

capital.loss
Capital loss

hours.per.week
Weekly working hours

native.country

Native country