Classification: Adult_50K

Introduction

In this assignment, you will use machine learning techniques in order to predict whether a person makes over 50K a year. Your prediction will be based on features of different nature (demographic, financial, etc.)

This data was extracted from the census bureau database found at

http://www.census.gov/ftp/pub/DES/www/welcome.html

Target Class:

The target attribute is binary: 1- if the person makes more that 50K US dollars in a year or , 0- the client did not makes more that 50K US dollars in a year.

Attributes Information:

id	Attribute	Type	Description/Values
1	age	numerical	Age of the potential client.
2	workclass	categorical	Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
3	fnlwgt	numerical	Census weight.
4	education	categorical	Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
5	education-num	numerical	Years of education
6	marital-status	categorical	Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
7	occupation	categorical	Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
8	relationship	categorical	Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
9	race	categorical	Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
10	sex	categorical	Female, Male.
11	capital-gain		Capital gains.
11	capital-loss	numerical	Capital losses.

12	hours-per-week	numerical	Hours per week of work.
13	native-country	categorical	United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

Reference: Ron Kohavi, "Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996