Adult Income Data
http://mlr.cs.umass.edu/ml/datasets/Adult

Summary

Predict whether income exceeds $50K/year based on census data. Also known as "Census Income" dataset. Extraction was done by Barry Becker from the 1994 Census database. Adult income helps to determine the growth of the country/region. In this project, I will explore the key attributes / factors that could affect the adult income (like higher education, age, area etc.). I will do the classification, regression and prediction to determine whether the income of adult is greater than 50K or not.

Source of the data

The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms

Dataset Donors: Ronny Kohavi and Barry Becker

Dataset URL: http://mlr.cs.umass.edu/ml/datasets/Adult

48842 instances, mix of continuous and discrete (train=32561, test=16281)

Research Questions
1) How many adults have income less than 50K and greater than 50K
2) Which gender makes more income
3) Individuals with higher education vs individuals with basis education
    a. Higher education means high income?
4) What factors/attributes should be considered to predict the income

Technique
- Loading, cleaning and exploring
- Classification
- Regression testing
- Prediction (real data vs test data)

Tools
R Language and RStudio Version 1.1.419 will be used

Shabbir Yousuf Ali (syali@ryerson.ca)   **Student ID** 039081070