Caleb Pudvar

Final Project Proposal

**US Census**

**Income Analysis**

For my final project, I desire to analyze something very prominent in the news across the country: annual income. Using data from nearly 50,000 instances of eligible adults taken from the 1994 census, I plan to analyze salary patterns. I found a strong source for this data provided from the Machine Learning Repository. This set of data provides me with fourteen different attributes. These attributes include, but are not limited to:

* Age
* Sex
* Education
* Marital status
* Native country

I hope for the end results to estimate, with a degree of certainty, whether or not a given individual will make an annual income greater than $50,000 annually. To visualize the results, I first plan to run the data through a simple cluster analysis using python. This will merely be a simple test to see if anything surprising stands out. The main analysis will be CART. Using this method of analysis, I will determine which features will provide an adult the means to make my goal income. All data will be processed using python to feed into valid .csv files which can properly accept the incoming data. I may decide to account for inflation, and extrapolate the annual income to the value of today’s dollar as well.

Source: <http://archive.ics.uci.edu/ml/datasets/Census+Income>