**About Dataset**

US Adult Census data relating income to social factors such as Age, Education, race etc.

The Us Adult income dataset was extracted by Barry Becker from the 1994 US Census Database. The data set consists of anonymous information such as occupation, age, native country, race, capital gain, capital loss, education, work class and more.  
Each row is labelled as either having a salary greater than ">50K" or "<=50K".

This Data set is split into two CSV files, named adult-training.txt and adult-test.txt.

The goal here is to train a binary classifier on the training dataset to predict the column income\_bracket which has two possible values ">50K" and "<=50K" and evaluate the accuracy of the classifier with the test dataset.

Note that the dataset is made up of categorical and continuous features. It also contains missing values  
The categorical columns are: workclass, education, marital\_status, occupation, relationship, race, gender, native\_country

The continuous columns are: age, education\_num, capital\_gain, capital\_loss, hours\_per\_week

This Dataset was obtained from the UCI repository, it can be found on

<https://archive.ics.uci.edu/ml/datasets/census+income>,  
<http://mlr.cs.umass.edu/ml/machine-learning-databases/adult/>

USAGE  
This dataset is well suited to developing and testing wide linear classifiers, deep neutral network classifiers and a combination of both. For more info on Combined Deep and Wide Model classifiers, refer to the Research Paper by Google <https://arxiv.org/abs/1606.07792>

Refer to this kernel for sample usage : <https://www.kaggle.com/johnolafenwa/wage-prediction>

Complete Tutorial is available from <http://johnolafenwa.blogspot.com.ng/2017/07/machine-learning-tutorial-1-wage.html?m=1>