**Assignment**

In this assignment students need to predict whether a person makes over

50K per year or not from classic adult dataset using XGBoost. The

description of the dataset is as follows:

**Data Set Information:**

Extraction was done by Barry Becker from the 1994 Census

database. A set of reasonably clean records was extracted using the

following conditions: ((AAGE>16) && (AGI>100) &&

(AFNLWGT>1)&& (HRSWK>0))

Attribute Information:

Listing of attributes: >50K, <=50K.

age: continuous.

workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov,

Local-gov, State-gov, Without-pay, Never-worked.

fnlwgt: continuous.

education: Bachelors, Some-college, 11th, HS-grad, Prof-school,

Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th,

Doctorate, 5th-6th, Preschool.

education-num: continuous.

marital-status: Married-civ-spouse, Divorced, Never-married,

Separated, Widowed, Married-spouse-absent, Married-AF-spouse.

occupation: 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.

relationship: Wife, Own-child, Husband, Not-in-family, Other-relative,

Unmarried.

race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.

sex: Female, Male.

capital-gain: continuous.

capital-loss: continuous.

hours-per-week: continuous.

native-country: 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.

**Following is the code to load required libraries and data:**

import numpy as np

import pandas as pd

train\_set = pd.read\_csv('http://archive.ics.uci.edu/ml/machine-

learning-databases/adult/adult.dat a', header = None)

test\_set = pd.read\_csv('http://archive.ics.uci.edu/ml/machine-learning-

databases/adult/adult.test' , skiprows = 1, header = None)

col\_labels = ['age', 'workclass', 'fnlwgt', 'education', 'education\_num',

'marital\_status', 'occupation','relationship', 'race', 'sex', capital\_gain',

'capital\_loss', 'hours\_per\_week', 'native\_country', 'wage\_class']

train\_set.columns = col\_labels

test\_set.columns = col\_labels

**Task:** Deploy this assignment in any cloud platform.(Try to look for free cloud platform)

**Assignment:** Submit assignment’s deployable link only.