**Project**

**Task**

Using the provided dataset (each row represents one individual), you need to build a model that will predict whether they have a high or low income level (variable: **Inc**).

**Requirement**

1. Use the train file to create an appropriate model using any of the techniques that will predict a High Income Level.
2. Once you have a final model you are satisfied with, you will apply the model to predict outcomes for the observations in the test file. That is, based on the predictor variables in the test file you will predict if each observation has a High Income Level. Once you have added a ‘Predicted’ label to the test dataframe, write the dataset to a file.

**Files**

You have been provided two files (test and train) each containing the same predictor variables. The train file contains one extra variable, Inc, which indicates whether the subject has Income >

$50K.

PROG8430-23W-Final\_train.txt

PROG8430-23W-Final\_test.txt

**Data Dictionary**

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Index | Unique identifier |
| Inc | Binary outcome for income >$50K or <$50K |
| Capital | Capital Gains |
| Hours | Hours worked in a typical week. |
| Age | Age in years |
| Industry | Industry in which the subject works |
| Education | Education Level of subject |
| Marital | Marital Status of the subject |
| Occupation | Occupation type of the subject |