

Exploring Classification Algorithms on a Census Income Dataset:

Code Readme

Yuwen Yu

Alexander Castro

Brendan McShane

Ismail Mustafa

CISC 5790: Data Mining

Table of Contents

1 Readme

2

1 Readme

The file provided, *preprocessing_feature_selection_model_build.ipynb*, should be ran using Jupyter Notebook.

For the loading of the code, set two file path variables within the code to point to the unprocessed train and test data sets. The necessary lines to edit in the code are as outlined below.

```
filepath_train = ''
filepath_test = ''

df = pd.read_csv(filepath_train,sep=' ',
*',names=columns, na_values='?')

df_test = pd.read_csv(filepath_test,sep=' ',
*',names=columns, na_values='?')
```

These changes alone will allow for the code to run.