Submitted By:Naneeta Talreja

Submitted to: Sir Gulam Murtaza

CMS-Id: 021-19-0030

Data mining project

# Dataset Description:

Name: Adult data set

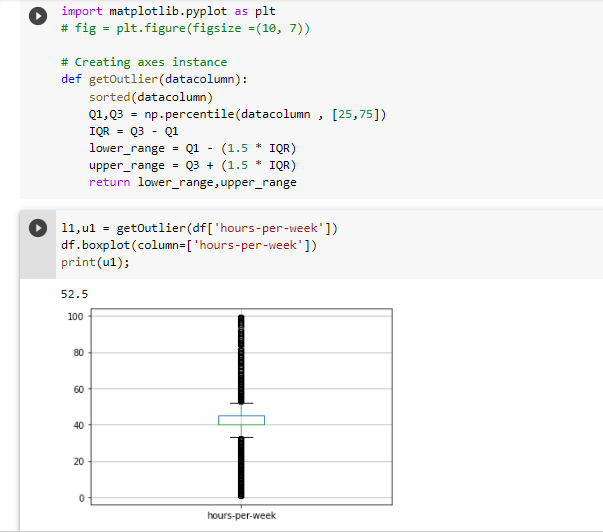
Number of Attributes: 14

Number of instances: 48842

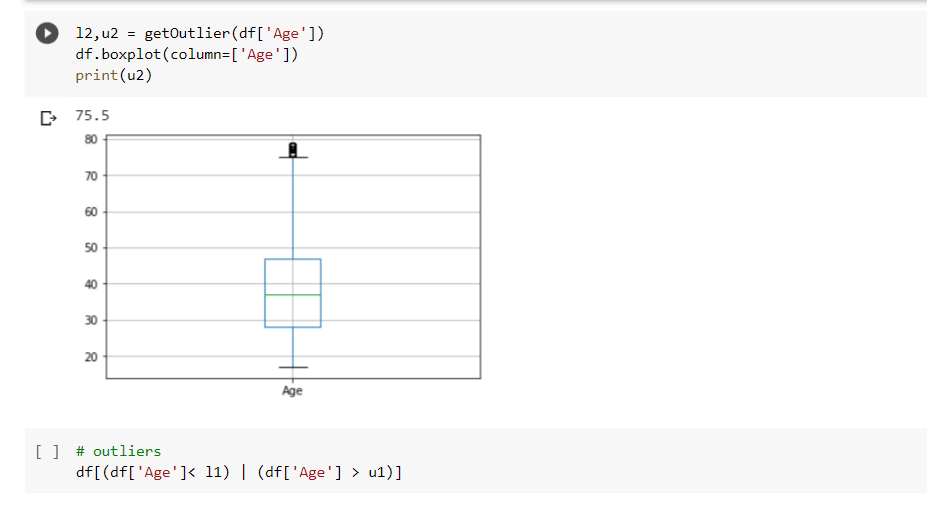
Task: Classification

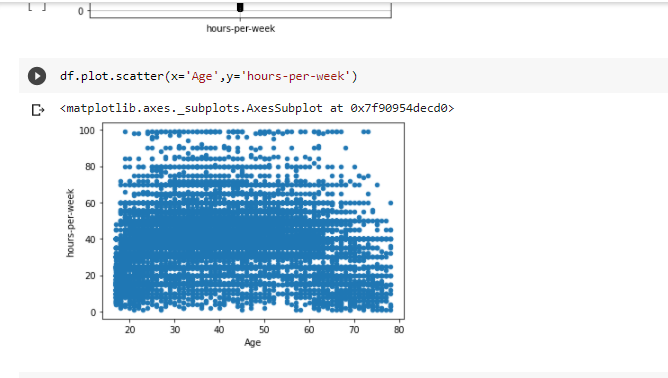
# Visualization of data:

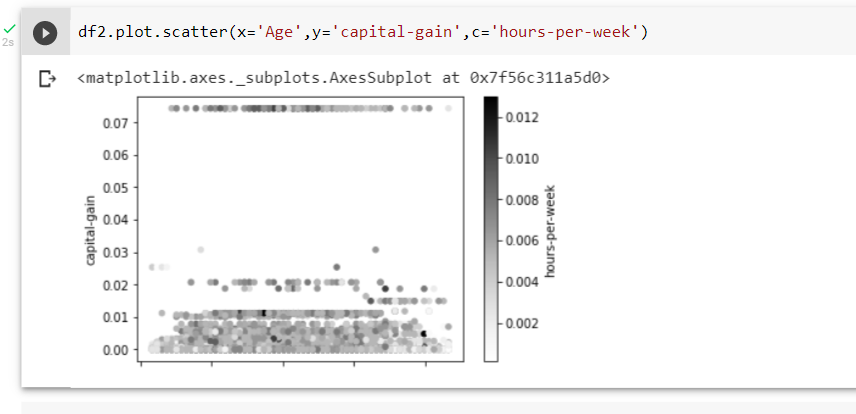
1. Visualization of attribute “hours per week” using box-plot:



1. Visualization of attribute “Age” using box-plot:

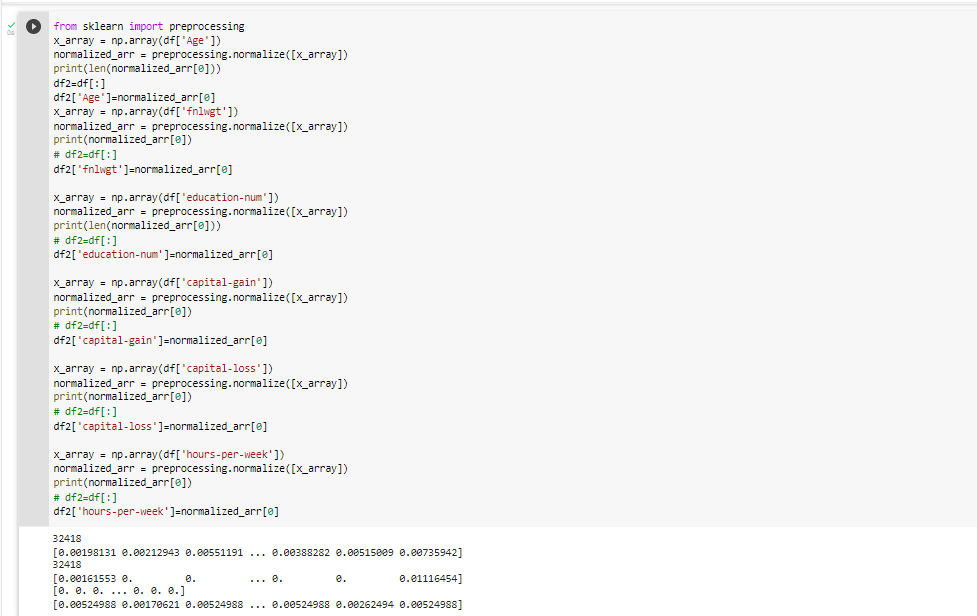


1. Visualization of age and hours per week attributes using scatter plot.
2. Visualization of age hours-per-week and capital-gain after preprocessing.

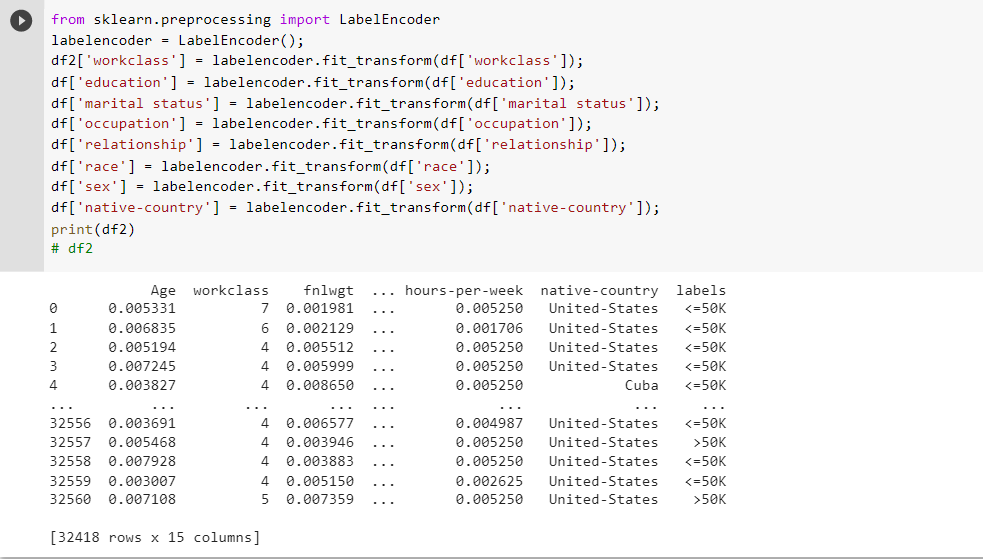


# Preprocessing:

Normalizing the integer columns

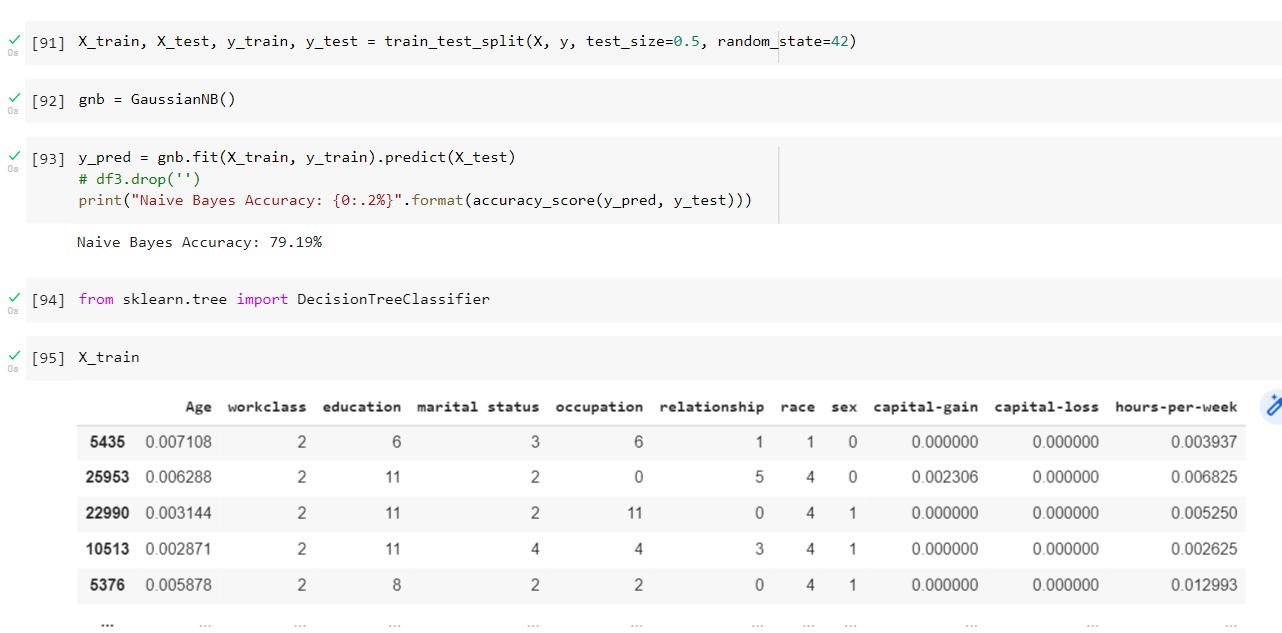


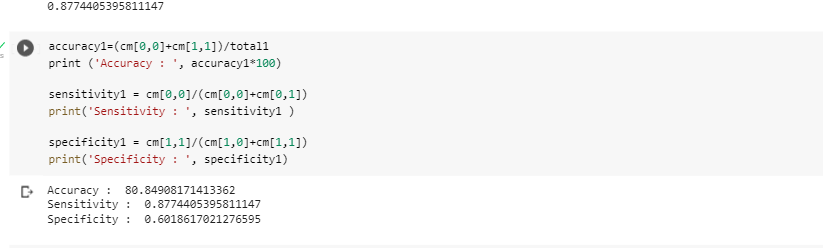
Label encoding the string columns



# Classification:

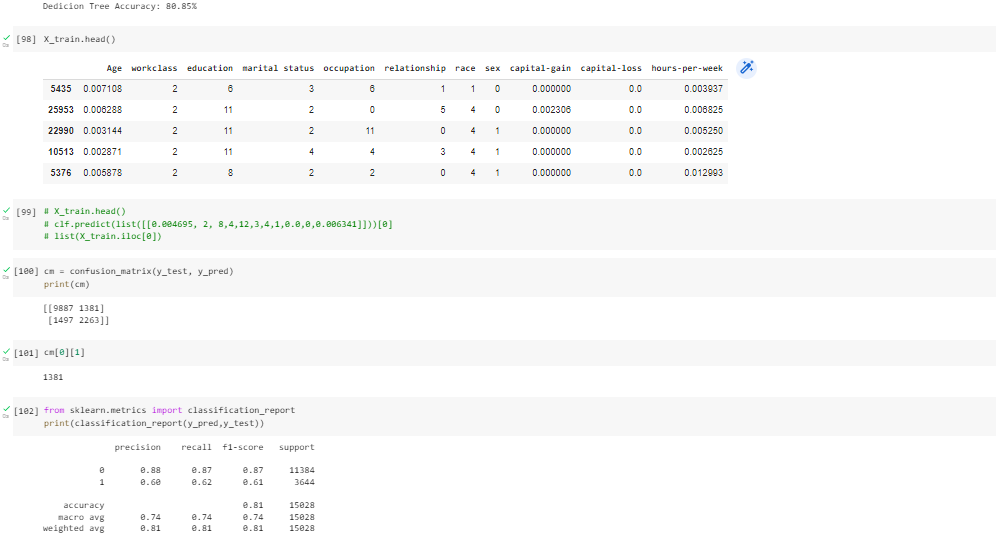
Splitting of data into training and testing and classification using GaussianNB and decisionTreeClassifier.





# Feature extraction:

Removing highly correlated Features



# Performance evaluation:

Confusion Matrix

