**Logistic Regression with scikit-learn**

**Preksha Koirala Dr. Mugizi Robert Rwebangira  
Sophomore @02765387 Professor, Machine Learning  
Howard University Howard University**

## **Logistic regression in Python with the**[**scikit-learn module**](http://scikit-learn.org/)**.** **Dataset**

The dataset I chose is of Parole Violation at   
<https://github.com/chicheongweng/MITx-15.071x-The-Analytics-Edge/blob/master/data/parole.csv>

**Description of Variables:**sex: Male 1, Female 0

Race: (1,2,3) Not necessary for our prediction

Age: Integer

State: (1,2,3)Not necessary for our prediction

Max.sentence: integer

Multiple.offenses: binary

Crime: integer

Violator: True(1), False(0)

**Problem Statement**

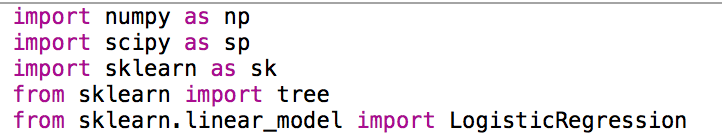
Let’s treat this as a classification problem by focusing on three variables sex, multiple.offenses and violater. Since they are in binary form, it will be easier to use logistic regression function on those variables.

**Libraries Used:**

Scipy

Numpy

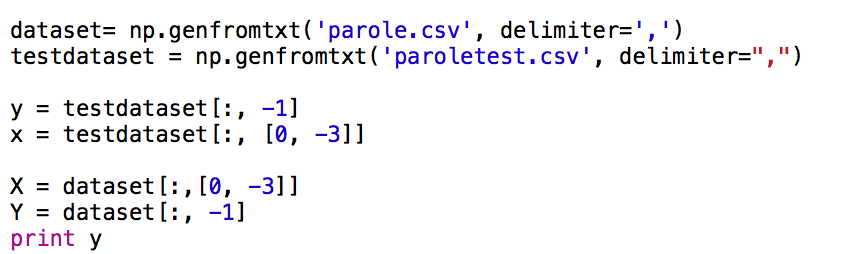
Sklearn



**Process:**

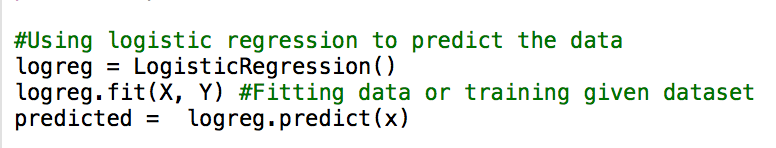
**1) Read file using numpy**

**2) Categorized X and y with the variables that were chosen to be used**

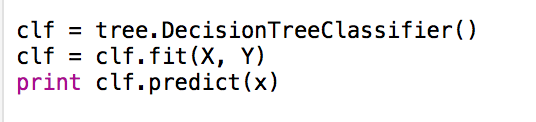


**3) Instantiated LogisticRegression class available in Sklearn.linear\_model**

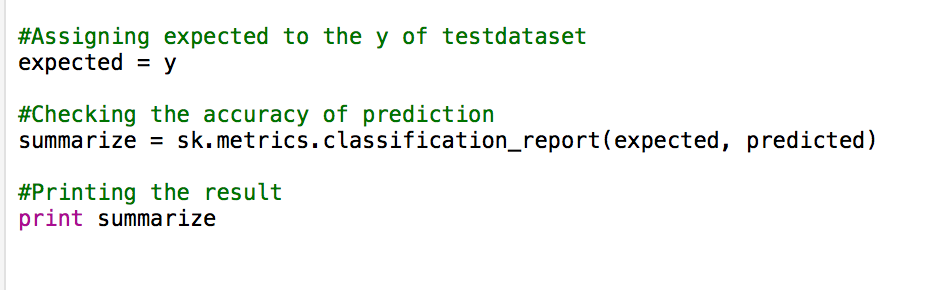
**4) Called predict function from LogisticRegression on X**



**5) Also used decision Tree for predicting data**

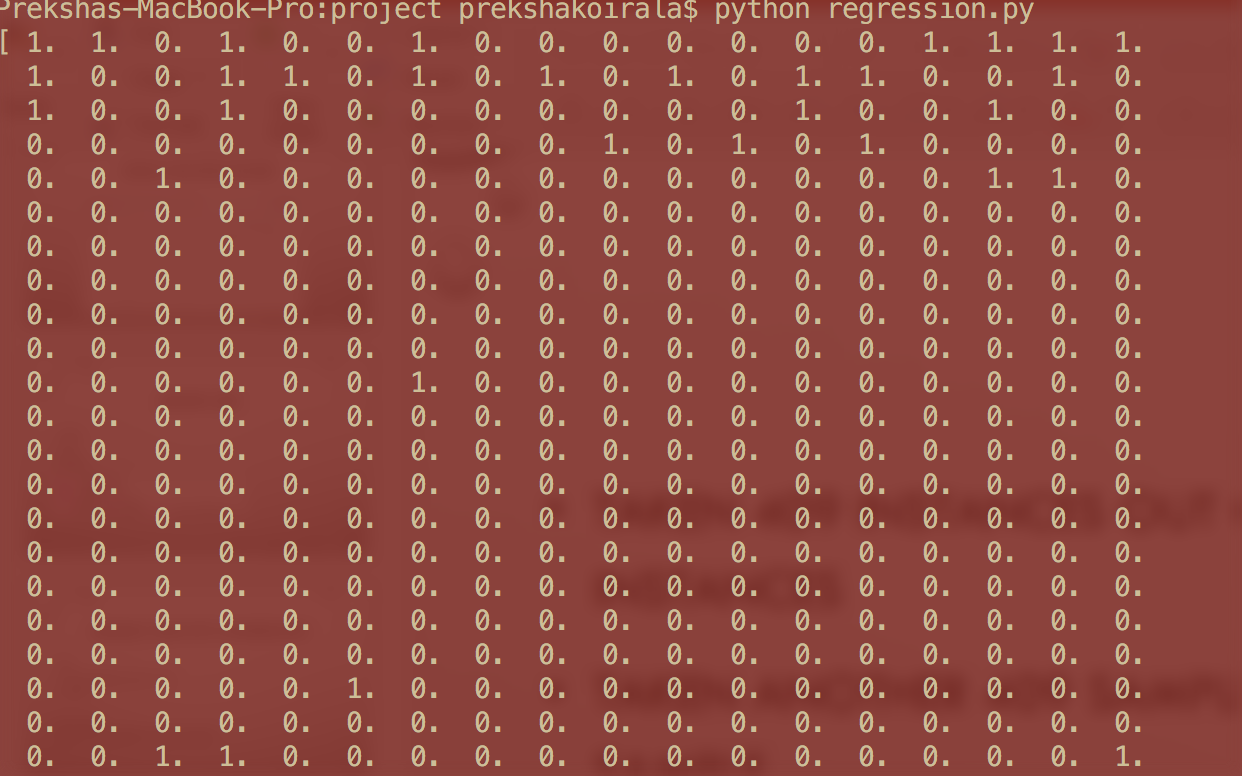


**6) Calculated the accuracy of the prediction using metrics.classifictionreport() from sklearn**

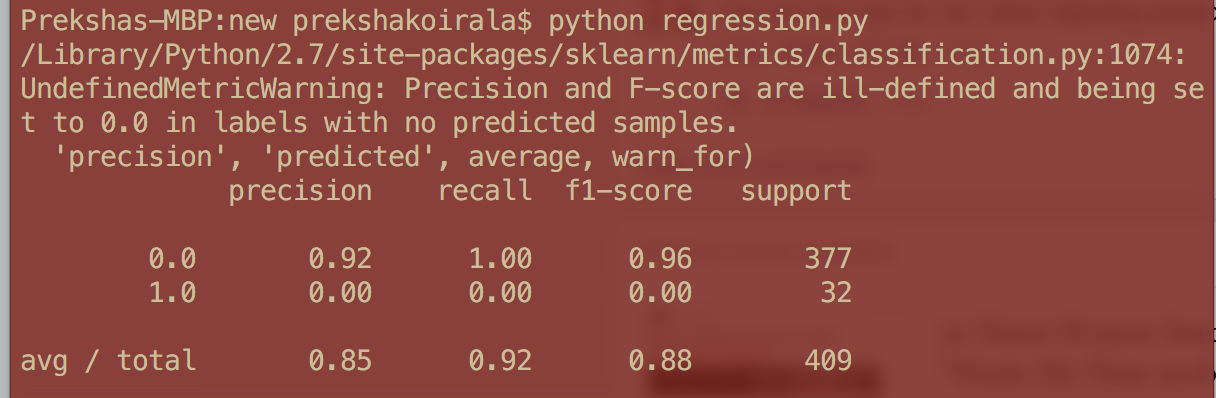


**7)Ran python file**

**Predicted data:**



Result:85% precision rate for 409 samples and 409 test cases.



**Thank you**

**Machine Learning Final Project**

**Logistic Regression on Parole Violation**