Joshua Hoshiko

CS 3120

Dr. Jiang

Homework 2 Report

**Program Overview:**

This program is designed use the SKLearn Logistic Regression on a Diabetes data set

1. The program accepts a CSV data file and creates a data-frame
2. The program then splits the data into testing and training data (40/60 split) using the train\_test\_split function from model\_selection in SKLearn
3. The program then uses the data to create a model using logistic regression by using the training X and Y slices.
4. The program then tests the model while generating a confusion matrix and ROC curve.

**Program Details:**

1. The 5 feature used were ‘pregnant’, ‘age’, ‘bp’, ‘insulin’, and ‘skin’.

**Program Output:**



