**Assignment #10**

**Objective**

Become familiar with the basics of Machine Learning.

**Due**

Before 9am on March 19th. Commit your completed Jupyter notebooks to your GitHub repository. And when completed let Dr. Ficklin know via Slack that it is available. Please indicate the amount of time the assignment took to complete.

**Tasks**

Task A

Complete the online tutorial at <https://machinelearningmasTaktery.com/machine-learning-in-python-step-by-step/>. You can skip the sections titled **1.1. Install SciPy Libraries** (as we have already done this), and **1.2 Start Python and Check Versions**.

Before proceeding, you will need to install the sklearn package. You can do that using this command:

pip install sklearn

or

pip3 install sklearn (on Ubuntu 16.04 linux)

Begin the tutorial at the section titled **2. Load the Data.** You simply need to read the tutorial, take time to understand it. Resist the urge to just cut-and-paste to get through it quickly. There are no exercises to complete, simply repeat the sample code provided. Upload a Jupyter notebook containing the code into an Assignment10 folder on your GitHub repository.

**Why do we have to do from pandas.plotting import scatter matrix?**

**How does the command Dataset.describe()) know what columns to use (ie not use class)**

Task B

Kaggle (<https://www.kaggle.com/>) is an online community for learning, practicing and performing machine learning. It provides training datasets, real-world datasets and online competitions for machine learning. For beginners, they offer the **Titanic: Machine Learning from Disaster** tutorial found here: <https://www.kaggle.com/c/titanic>. Please read the introduction to the problem on that page. You may remember that we have already looked at this Titanic dataset in a previous assignment.

To help solidify your understanding of how machine learning can be applied, you will complete this Titanic tutorial. Fortunately, solutions have already been supplied! Follow the tutorial solution at this link: <https://www.kaggle.com/startupsci/titanic-data-science-solutions>.

Before proceeding, you will need to install the seborn package. You can do that using this command:

pip install seborn

or

pip3 install seborn (on Ubuntu 16.04 linux)

Please skip over the top bolded section that discusses a Speedml implementation. Begin at the section titled **Workflow Stages**. The datasets have already been provided to you on Slack. You do not need to register for an account on Kaggle to complete the assignment. Again, there are no exercises to complete. Simply repeat the sample code provided into Jupyter notebook and take time to learn what is going on. Upload a Jupyter notebook containing the code into an Assignment10 folder on your GitHub repository.

If you spend more than 7-9 hours total on this assignment, please stop and indicate to Dr. Ficklin how far you were able to get.