**9/1/2018**

Read several articles about machine learning at the Machine Learning Mastery website:

<https://machinelearningmastery.com/>

<https://machinelearningmastery.com/how-to-define-your-machine-learning-problem/>

<https://machinelearningmastery.com/process-for-working-through-machine-learning-problems/>

<https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>

<https://machinelearningmastery.com/why-you-should-be-spot-checking-algorithms-on-your-machine-learning-problems/>

**9/8/2018**

Came across Weka as the tool to experiment with machine learning concepts without any programming:

<https://machinelearningmastery.com/how-to-run-your-first-classifier-in-weka/>

<https://www.cs.waikato.ac.nz/ml/weka/downloading.html>

**9/22/2018**

Learned about scikit-learn as a popular tool to run machine learning algorithms:

<https://machinelearningmastery.com/compare-machine-learning-algorithms-python-scikit-learn/>

<https://machinelearningmastery.com/setup-python-environment-machine-learning-deep-learning-anaconda/>

<https://machinelearningmastery.com/make-predictions-scikit-learn/>

<https://scikit-learn.org/stable/install.html>

**9/29/2018**

Read about the PMLB research paper that compared machine learning algorithms:

<https://machinelearningmastery.com/start-with-gradient-boosting/>

<https://arxiv.org/abs/1708.05070>

<https://github.com/EpistasisLab/penn-ml-benchmarks>

**10/6/2018**

Read about the XGBoost classifier as the most popular algorithm for tabular datasets:

<https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>

<https://machinelearningmastery.com/develop-first-xgboost-model-python-scikit-learn/>

<https://www.ibm.com/developerworks/community/blogs/jfp/entry/Installing_XGBoost_on_Mac_OSX?lang=en>

**10/13/2018**

Noticed that the PMLB code added XGBoost evaluation support even though it wasn’t listed in their research paper. I decided to evaluate my dataset with XGBoost in addition to Random Forest and Gradient Boosting classifiers.

**10/20/2018**

Found the UCI machine learning website that offers datasets in several areas:

https://machinelearningmastery.com/practice-machine-learning-with-small-in-memory-datasets-from-the-uci-machine-learning-repository/

<http://archive.ics.uci.edu/ml/datasets.html>

**10/27/2018**

Searched for a heart related dataset at UCI as I got interested in bioinformatics after I talked to my cardiologist, Nafiz Kiciman MD, during my heart exam:

<http://archive.ics.uci.edu/ml/datasets/Echocardiogram>

<http://archive.ics.uci.edu/ml/datasets/Cardiotocography>

<http://archive.ics.uci.edu/ml/datasets/Heart+Disease>

<http://archive.ics.uci.edu/ml/datasets/SPECT+Heart>

<http://archive.ics.uci.edu/ml/datasets/SPECTF+Heart>

<http://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>

**11/10/2018**

Decided on my OCSEF project: A web application that can be run on any computer or mobile device, which will let a cardiologist enter a patient’s heart data and predict whether the patient is healthy or sick.

<http://archive.ics.uci.edu/ml/datasets/Heart+Disease>

**11/17/2018**

I called Dr. Kiciman for a private meeting to go over the heart disease data. He immediately noticed some attributes that looked irrelevant, and I took a note of them:

age, sex, chol, fbs, restecg, trestbps

**11/17/2018**

Selected the attributes using Weka. Used several different methods to find the best set of attributes, and Weka consistently dropped the number of attributes from 13 to 7. The irrelevant attributes were the same as the ones recommended by Dr. Kiciman. Appreciated the great expert advice.

**11/24/2018**

Compared the performance of Gradient Tree Boosting, XGBoost, and Random Forest by running the PMLB tools, and selected XGBoost as it had the highest accuracy and the fastest speed.

**12/1/2018**

Wrote a web page in HTML and Javascript as a user interface for the medical personnel to key in patient data and predict results.

**12/29/2018**

Used the PMLB source code in Python to create a machine learning model from the UCI heart data.

**1/12/2018**

Wrote a web service in Python Flask to let the web page communicate with the machine learning model.

**1/26/2019**

Decided to use weebly as a free website hosting company for my web page:

<https://www.weebly.com/features/free-web-hosting>

**2/2/2019**

Published my web page on weebly:

<http://heartdx.weebly.com/>

**2/9/2019**

Dad helped me deploy the server-side Python program and the web service onto a public machine.

**2/9/2019**

Started writing the report for my project:

<https://www.ocsef.org/images/resources/Writing_Reports-OCSEF.pdf>

**2/16/2019**

Finished my report.

**2/23/2019**

Started working on my poster board:

<https://www.ocsef.org/images/resources/Science_Fair_Displays-OCSEF.pdf>

**3/9/2019**

Finished my poster board.

**3/16/2019**

Prepared for my presentation and interview with the judges.