December 4, 2019
COMP 520: Artificial Intelligence Name: **Daniel Babalola**

Progress Report 1

1. Project Summary:
My project is to carry out a **Letter Recognition** task using **k-Nearest Neighbor**. The learning algorithm is implemented in **Python**. My data set consists of **20000** characters from the English alphabet, gotten from the **UCI machine learning repository**. In this project, I planned to examine:

a. How well the k-NN performs with this learning task
b. How the error and performance varies with different values of k
c. How the error varies with different sample sizes of training data
d. How **boosting** could be used to reduce the number of training data required to train and still maintain a low error value

2. Current Status(work completed):
So far, I've been able to perform training successfully using the kNN model with accuracy averaging **95%** on the training data.

I've also been able to run several instances of my model over various values of k. It turns out that as k typically grows, the error grows as well. I settled for a k-value of **3** nearest neighbours as this gave very good consistency.

I also have been able to train several instances with varying values of sample data. Just like in theory, as the size of n grew, the accuracy did as well. Results were most stable after about **10000** examples.

3. Experiments in Final Report:
In the final report, I plan to conclude my experiments on the boosting process. Much progress has already been made in the implementation of the algorithm. I had to do this from ground up as the existing boosting package doesn't fit well with the k-NN model.

The completion of the boosting algorithm would help show how an accuracy of 95% or more can still be achieved with as little training examples as possible, hence reducing classification time and improving overall performance.