Note: Because we have not covered deep learning in class, much of the work I did for this deliverable was just to practice implementing the models we learned in class. Once we cover CNNs, I will likely revise and/or redo the steps in this deliverable.

Problem Statement:

My project is to classify paintings by artist. Given a painting, I would like my program to predict with about 70% accuracy which artists in a specified set painted the piece of art.

Data Preprocessing:

I chose to work with the data I proposed in the last deliverable. The data I have is very skewed. I have close to 100 paintings for some artists and only 50 for others, so I decided to choose the 10 artists that I had the most data for. And to further simplify my dataset I just took 200 paintings for each of the 10 artists. Because of the model I was using I needed all the images to be the same size, so resized them and made them significantly smaller. I also converted all of the images to black and white because I knew I would need to flatten my matrices to one dimensional arrays, and so I needed to reduce the number of dimensions of the matrices.

Machine Learning Model:

None of the models we have learned in class really apply to my projects, so I decided to use the PCA model mostly for the sake of practice. I had originally planned to use a CNN, but because I am still unfamiliar with the model, I did not feel comfortable implementing it.

Preliminary Results:

The PCA model produced no significant results. The explained variance for all ten components was negligibly low.

Next Steps:

The PCA model is simply not the right fit for this problem. It forces me to reduce my data too much. In particular having to turn my images into gray-scale is a problem because color is a helpful indicator of artist. I think my next step will be to spend some time learning about CNNs so that I can try to implement one and get some significant results.