Group 2 Project 4 Design Document

Goals:

Long Term Goal: Creating a model that can classify different letters in images to help with converting physical lettering into data

Short Term:

- Creating a model with 80 percent accuracy
- Being able to highlight the letters in the image

Roles:

- Data Analyzer(Alexander Aybar) - Corrects the data so it can be properly used in the model
- Presentation Designer(Talaya Sherdon) - Works on the presentation and makes sure that everything follows what we are trying to show
- Calculations(Tyrell Green) -Calculates the precision, recall and f1 - scores from the model to make sure the model is the best it can be
- Model Designer(Ishma'il Scott) - Trains the model and works close with calculations to make sure the model is the best it can be

Process:

Part 1:

1. We receive the data
2. Upload the data into the colab
3. Read and analyze
4. Convert the txt. file into a data set
5. Make the last column into the target column
6. Then show the image as a 16 x 8 array and show it

Part 2:

1. Create a multinomial Naïve Bayes model
2. Use all the training data to train the model with a smoothing factor of 1
3. Calculate the confusion matrix and display

Part 3:

1. Reformat the confusion matrix to better display it
2. Display the metrics
3. Create a heatmap

Part 4

1. Get the ROC for each letter
2. Create the graph for the roc curve

Problem Space:

Finding out a way to teach a machine to classify any letter is a challenge we had to face. This can be used for various different things such as license identification, text to speech, and even image to text. However, we need to do this as accurately and precisely as possible. In order to achieve this we will train a model that will be able to identify letters and be even more efficient than any previous models