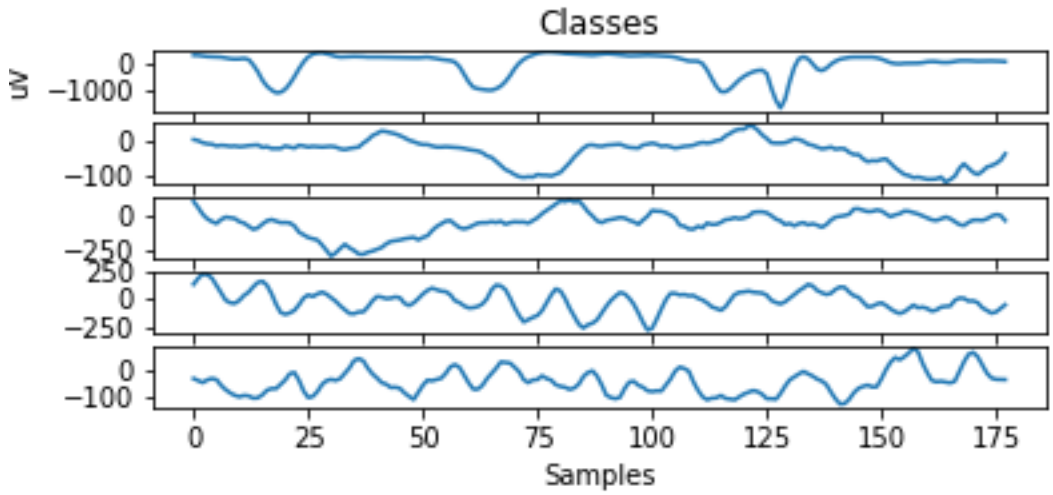
Solim LeGris
November 13, 2020

MAIS 202 - PROJECT DELIVERABLE 3

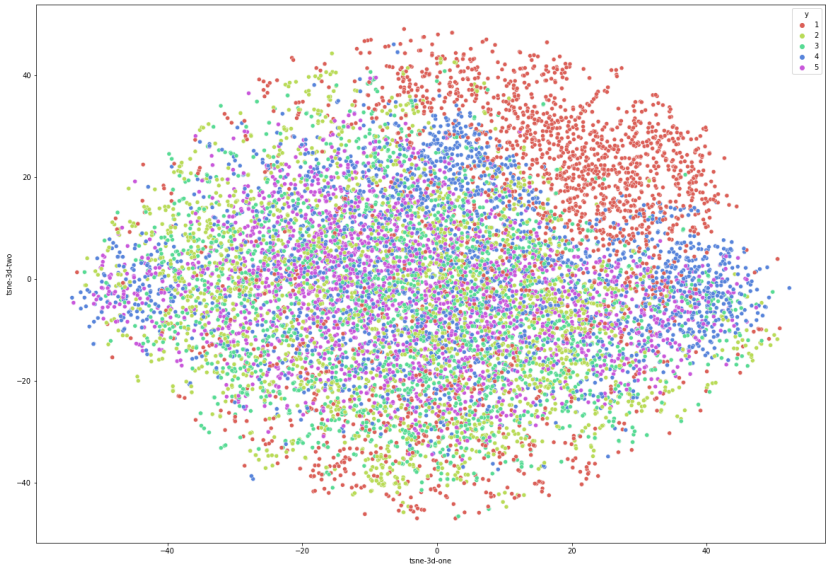
Further Data Analysis

As mentioned in my previous deliverable, I was getting extremely good results on both classifiers that I trained so I attempted to decipher my data furthermore.

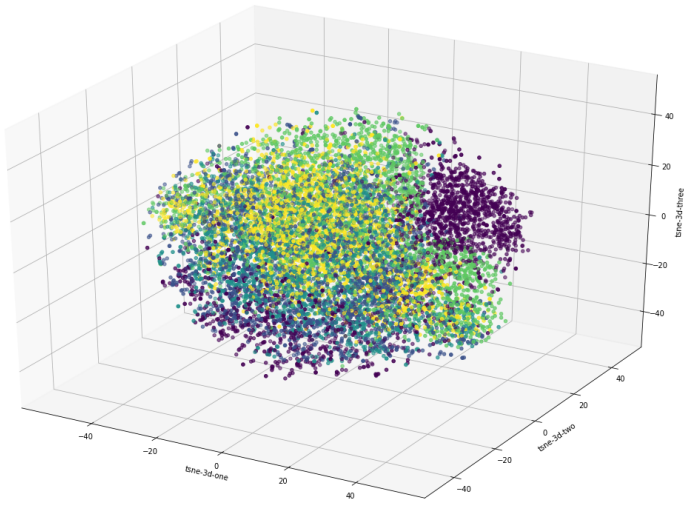
I plotted the 5 classes' EEG time-series to visualize the averaged electrical activity of the patients over the 1s recordings for each condition.



I also applied dimensionality reduction using PCA to my dataset (for visualization) and used t-distributed Stochastic Neighbor Embedding to visualize my data in 2D and 3D. I plotted the graphs for the data with the 5 original labels and with the 2 target labels as well.

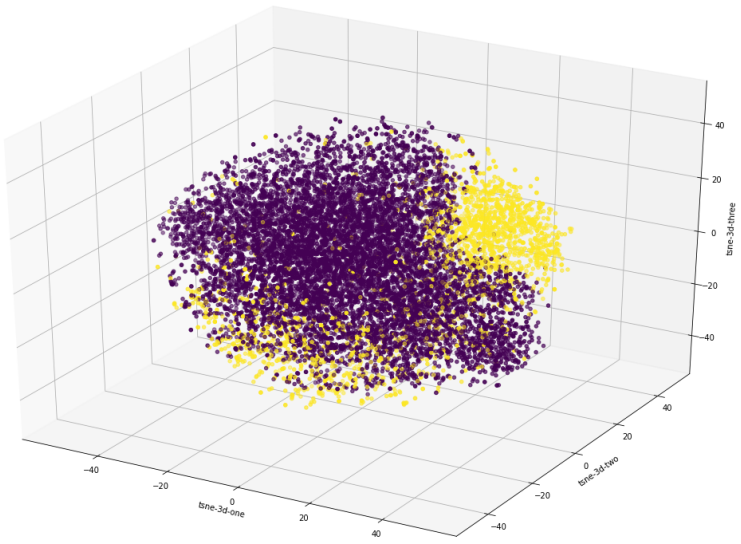


2D Scatter Plot with Multiclass Coloring



3D Scatter Plot with Mutliclass Coloring

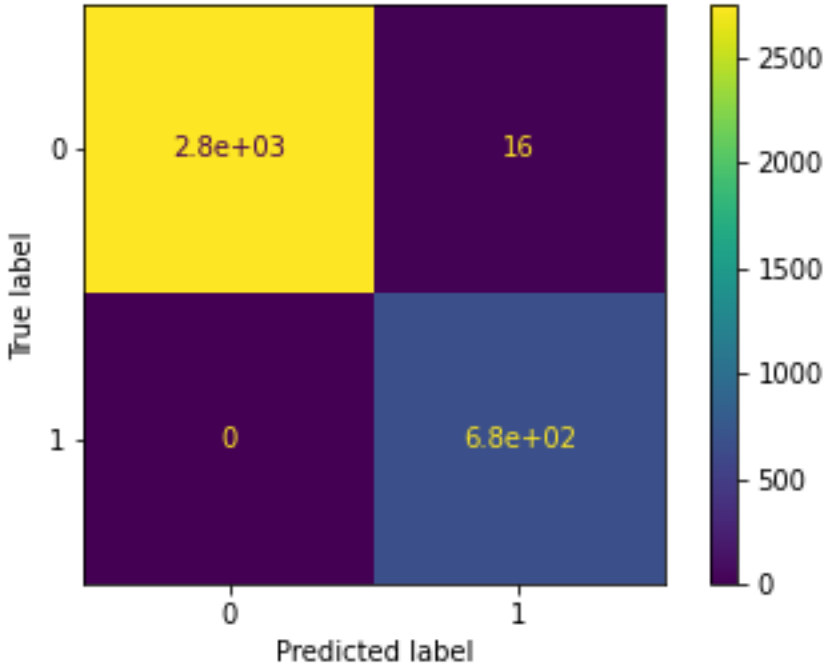
2D Scatter Plot with Binary Coloring



3D Scatter Plot with Binary Coloring

Final Training Results

Given that I was getting very accurate results, I reduced the size of the training set to 70%to account for potential overfitting. I also standardized my data and introduced a Gaussian noise factor with $\mu = 0$ and $\sigma = 0.5$ to my data. After training a Support Vector Machine using that noisy data, I obtained an accuracy of 0.9953, a log loss of 0.1601 and the following confusion matrix. The data in this preprocessed form is very robust and yields accurate results.



Solim LeGris
November 13, 2020

Final Demonstration Proposal

For my final demonstration, I will be constructing a landing web page using Flask as a framework for my backend development and React as a framework for my frontend development. I want to build a webpage that demonstrates the usefulness of my model in the medical setting. Given that people do not generally have EEG data at hand, I will allow for a random generator button that will pick out a random EEG datapoint from the dataset I used to train the model and plot the time series for the user to see. This datapoint will serve as input for the model and will yield a prediction. Furthermore, I will allow for users to upload EEG time series in a specific format and try predictions on it. Lastly, I also want to add a noise introduction feature (in the form of a slider) that will demonstrate how robust my model is to the user and will allow the user to test different variations of a given input.