Data Visualisation final- docmuntation:

Models I used:

Knn, GradientBoostedTrees, RandomForest, LogisticRegrassion, k\_means-clustering

Avaluation methods:

Accuracy score, confusion matrix, roc-curve, recall and precision scores , f1-score

I used PrincipleComponantAnalysis & standard scaling

A little about how I conducted each task:

I used different libraries including pandas to convert the cats and dogs pictures into an array of 2500

features, each representing the 'grayscale' value of a unique pixel of the picture reduced to an 50x50 array.

Both in the Fmnist and CatsVSDogs notebooks I used k\_means-clustering to cluster different

Square subparts of the picture. Than I used a graph of the number of features as a function

of the explained variance from the original data to figure out what value should I use as an input

For the PCA algorithm. I compressed the data with it, applied the different ML algorithms,

,compared between them with cross validation and the evaluation methods above.

Lastly I tried different ensamble and optimization techniques to improve upon them.

In the notebook about the neurological experiment, I defined several functions to prosses the data into a coherent

Pandas's DataFrame and another for splitting the data into train\test parts without ruining the continuity of the data.

I used graphs and pandas method to visualizes the data and even build a new feature.

After experimenting with different models

and methods I tried to condense the data such that each 5 consecutive instances would be as one.

Eventually I imported the validation data used scaling & pca and checked the results