Data Science Project

The first part of this project is about three classification notebooks.
The first notebook is about the improvement of the classification notebook from the last semester by using the Ensemble Learning methods and PCA (Principal Component Analysis).
The second notebook is about multiclass classification of the FMNIST data set, which was arranged as a csv file, in which every row represented an image by giving us the values of each pixel in an array.
The third notebook is about binary classification of the DOGS VS CATS data set, we worked with pictures that we made into a numpy array.

The Ensemble Learning methods I used were:

- Voting Classifier
- AdaBoost Classifier
- XGBoost
- Stacking Classifier
- Bagging

PCA (Principal Component Analysis):

- In order to create a compact model, we would like to reduce the dimensions by using the PCA method from sklearn and train again the models over the reduced train set. By using PCA we make the models compact but with a high score, mostly we will see a slightly decrease of the accuracy score.

The second part of this project is about the Harnessing Machine Learning for interpersonal physical alignment project.
While loading the data to the notebook I added a labels column by using the file's name, when the Alone file was added I merged the data with the Right-Hand data to get a balanced data set. Then separated the left hand from the right hand and merged them by their Time column (they both were taken at the same time). At last, I recreated the data set columns by adding r_ to the right hand's columns, and l_ to the left hand's columns.
Then after loading each person's data set, I merged their all three data sets and finally I merged all the person's data sets into one data set.
I did it all on the training files and the validation files, so I got two different data sets one for the test and one for the train.

After I have all the data sets I need, I separated the test into the test labels and the rest of the data, did the same to the train set, then started working on the train set.
I split the train set into valid-train data sets, then normalized the train, the valid and the test data sets. I have put aside the test set to use it in the end, and started visualizing the train set, then finally started training models on the data. I used the same models I used in the other three notebooks.
At the end of the project, I ran the trained models over the test set.