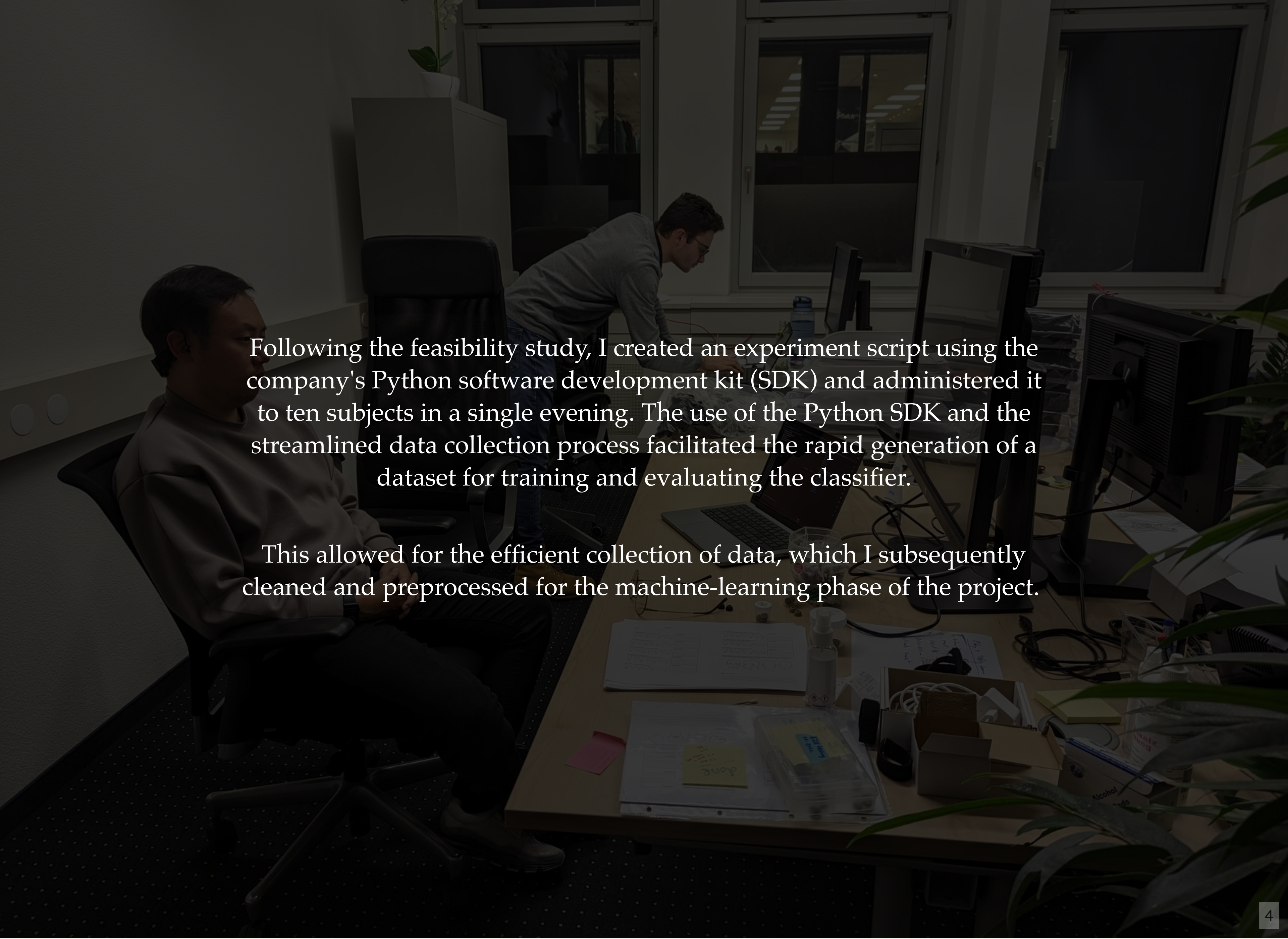
Capstone Project

Daniel Burger — Dec. 2022

SIT Academy Zürich

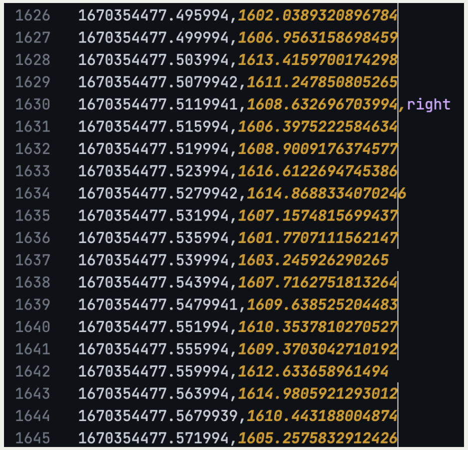
The project involved creating **a classifier for detecting eye winks** from electroencephalographic (EEG) signals. I used in-ear EEG earbuds from IDUN Technologies to collect the necessary data. The classifier was trained on a dataset derived from these recordings.

The project's initial phase involved **conducting a feasibility study** to assess the presence of patterns in the processed and transformed neural signals. I was pleased to discover that there were indeed discernible patterns, indicating the potential viability of developing a classifier to identify them. This finding provided the foundation for further exploration and analysis.

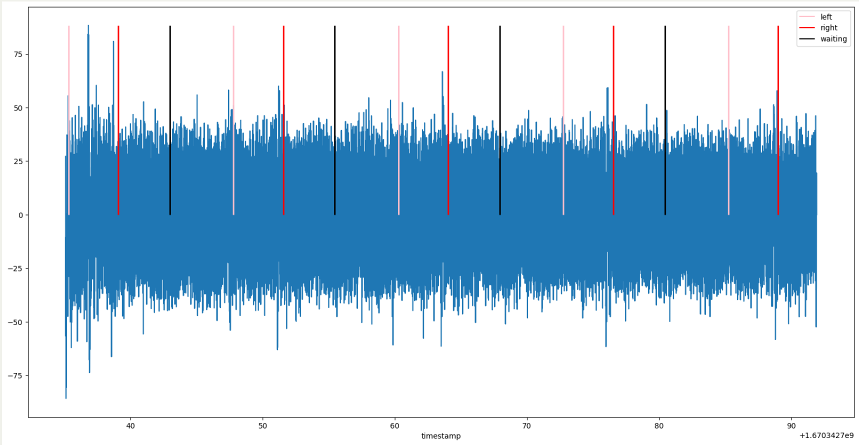


Following the feasibility study, I created an experiment script using the company's Python software development kit (SDK) and administered it to ten subjects in a single evening. The use of the Python SDK and the streamlined data collection process facilitated the rapid generation of a dataset for training and evaluating the classifier.

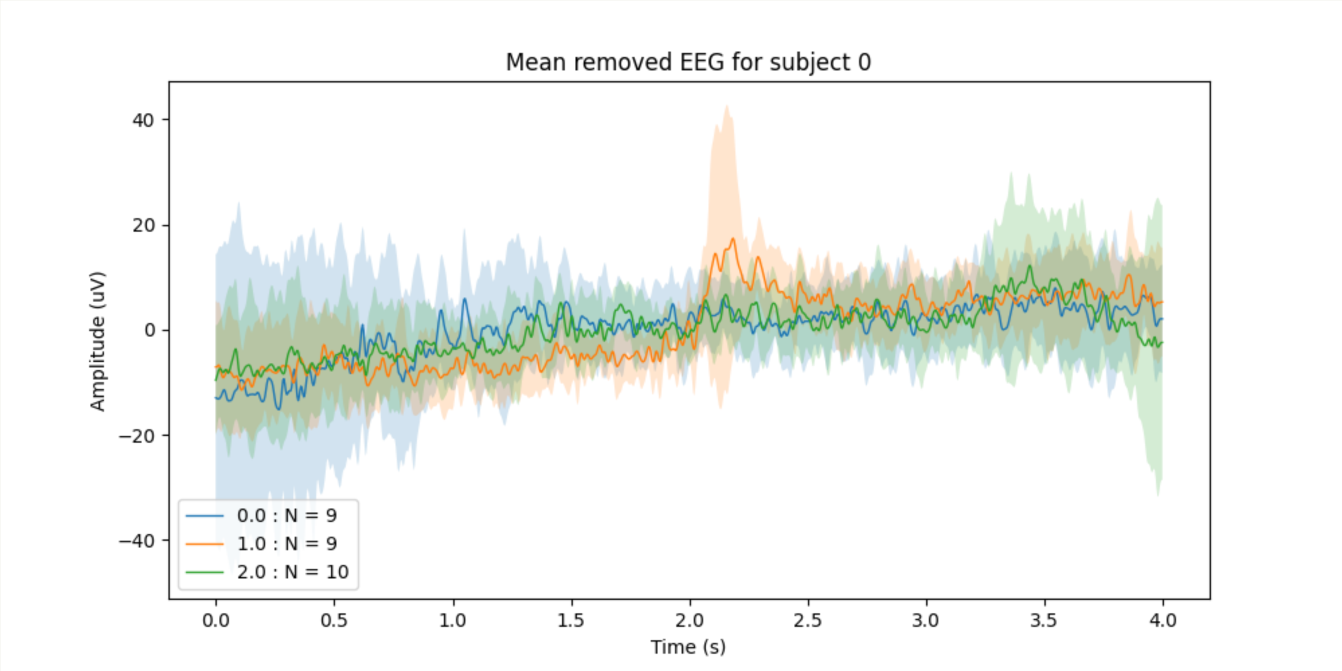
This allowed for the efficient collection of data, which I subsequently cleaned and preprocessed for the machine-learning phase of the project.



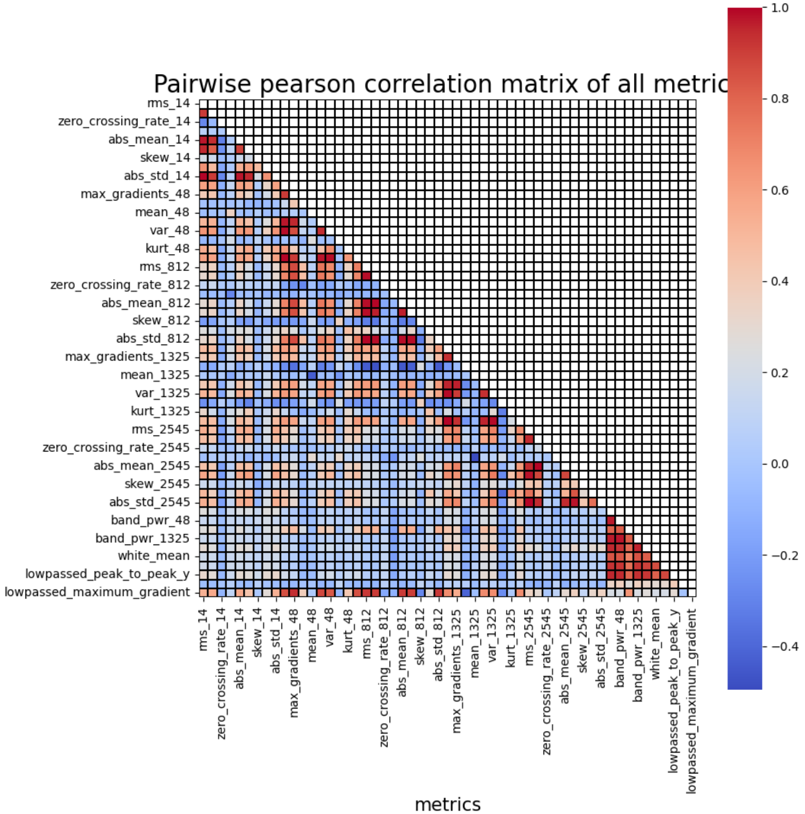
```
1670354477.495994,1602.0389320896784
1670354477.499994,1606.9563158698459
1670354477.503994,1613.4159700174298
1670354477.5079942,1611.247850805265
1670354477.5119941,1608.632696703994,right
1670354477.515994,1606.3975222584634
1670354477.519994,1608.9009176374577
1670354477.523994,1616.6122694745386
1670354477.5279942,1614.8688334070246
1670354477.531994,1607.1574815699437
1670354477.535994,1601.7707111562147
1670354477.539994,1603.245926290265
1670354477.543994,1607.7162751813264
1670354477.5479941,1609.638525204483
1670354477.551994,1610.3537810270527
1670354477.555994,1609.3703042710192
1670354477.559994,1612.633658961494
1670354477.563994,1614.9805921293012
1670354477.5679939,1610.443188004874
1670354477.571994,1605.2575832912426
```



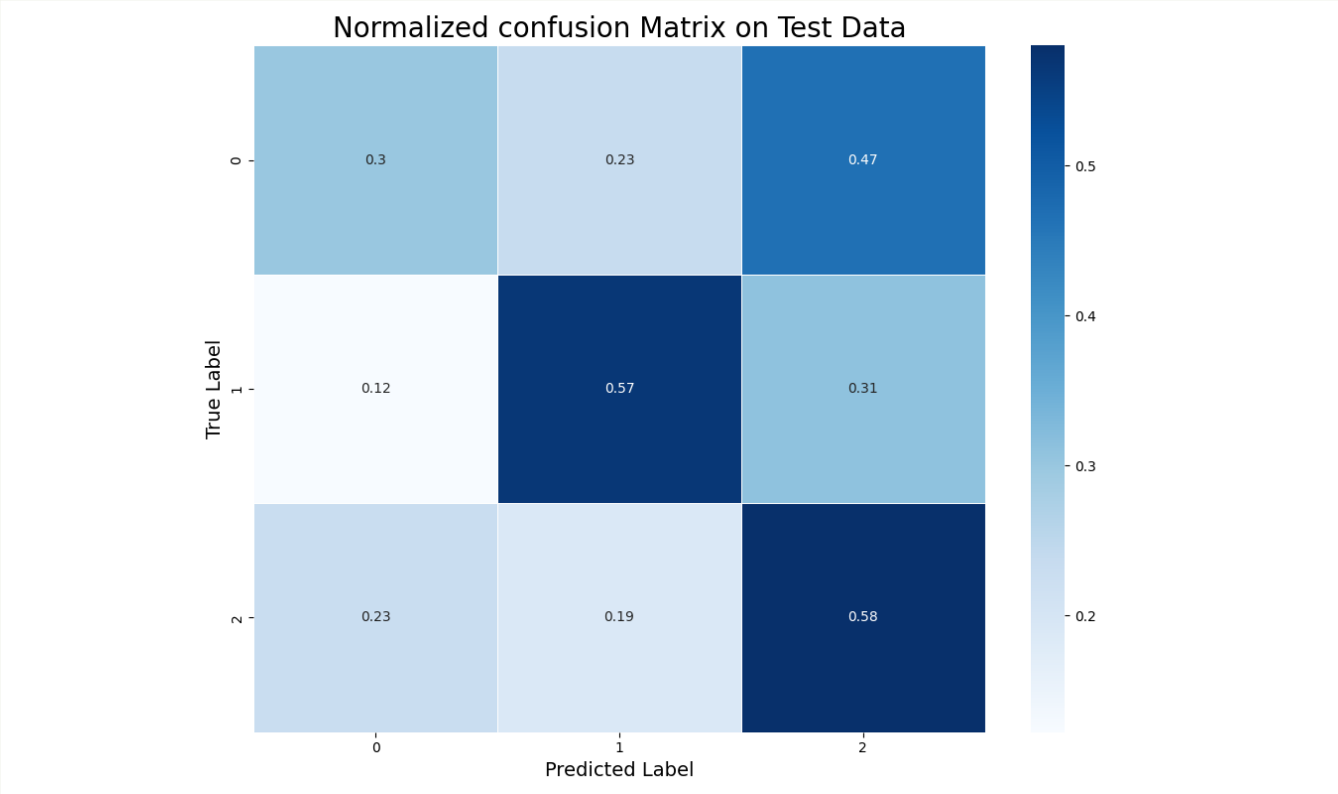
The outcome of the data collection process was the generation of multiple merged CSV files **containing the timestamps of the markers for the features** and the raw EEG signal, as illustrated in the left figure above. The right figure above displays the bandpass-filtered EEG signal with the markers superimposed, as specified in the experimental script. These CSV files provided the input for the subsequent preprocessing and machine learning phases of the project.



I epoched the data around the three different types of markers and stored the processed and filtered (including bandpass filtering and denoising) epochs in the repository. The figure above shows an example of the mean EEG signal from all subjects in the first epoch. **This preprocessing step facilitated the generation of a dataset suitable for training and evaluating the classifier.**



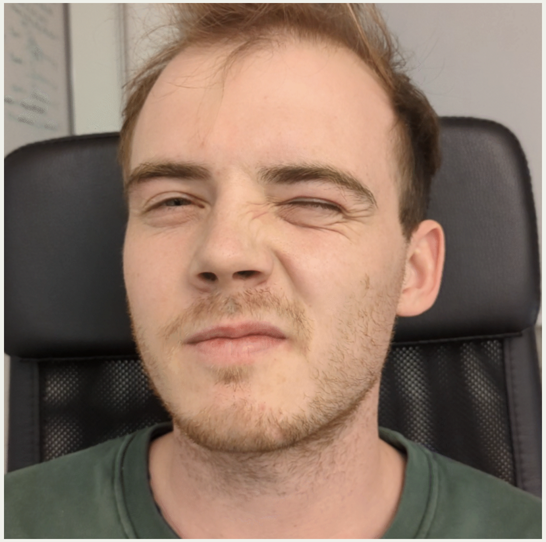
I employed **MLflow on an Azure compute cluster for automated machine learning** (AutoML) to train and evaluate a range of machine learning models on the preprocessed data. This approach allowed for the efficient and systematic exploration of multiple model architectures and hyperparameter settings, facilitating the identification of the most promising models for further analysis and refinement.



Although I was satisfied with the results of the automated machine learning process, I encountered a few challenges during the course of the project. For instance, the **collected data was somewhat noisy,** which presented difficulties in training accurate models. Additionally, there is potential for further improvement, as indicated by the observed limitations in the performance of the models generated by the AutoML process.

I am continuing to work on this project and am eager to learn and grow as a machine learning engineer. I am excited to see where this project leads and am confident that I will be able to make further progress in the development of the classifier.

*“ One of the key challenges of machine learning with BCI is **collecting high-quality data from neural signals.** Careful planning and execution of the data collection process is crucial for obtaining relevant and accurate data for training and evaluating BCI models.*

Additionally, **each individual's unique physiology** presents a further challenge in identifying patterns in neural signals. As illustrated in the figures above, individuals may exhibit significant variations in the muscles and movements involved in winking their eyes. This diversity of physiological factors complicates detecting consistent patterns in the corresponding neural signals.

Next steps

I plan to generate a larger dataset and devote more time to model selection to **avoid relying too heavily on automated methods** for this task in the future. This will enable a more thorough exploration of potential model architectures and hyperparameter settings, allowing for the identification of the most promising models for further development and evaluation.

Thanks for reading

GitHub repository with all data and code

https://github.com/danburonline/eye-wink