Data engineering applied to battery data

The BFH Center for Energy Storage (ESReC) is researching and developing solutions for applications in energy and mobility. They develop and test battery management systems and batteries for various applications. To achieve this, they have to create models to estimate the SoH (State of Health) and SoC (State of Charge) of a battery. This process involves a lot of testing in a laboratory. Machine learning approaches for generating such models could help with this process.

Goals

An industrial partner of the ESReC has recorded and stored operating data of their e-bikes over several years. The recorded data contains log data of the batteries operating conditions. The goal of this thesis was to apply different machine learning algorithms to the data for predicting the SoH of a battery. For this purpose, the log data had to be preprocessed and cleaned. To enable the reproducibility of the results, a pipeline was to be built.

Workflow

With the workflow shown below multiple machine learning algorithms were tested and the results saved in the webapp. The first line displays the pre-processing, by which a usage profile of each battery was generated. Lines two and three show our automated machine learning pipeline. To efficiently train the various machine learning algorithms, a shared NVIDIA DGX Station from the BFH was used. A usage and install manual was created to be able to recreate our work. The manual also covers the setup process on the docker container of the NVIDIA DGX Station.

Results

To evaluate the trained models, different plots were generated which show how different operating conditions affect the SoH of the battery. Additionally, the accuracy of the models was measured on a validation set. The algorithms were tested for two different battery types. The resulting models could be used for a better estimation of battery life.

Summary

We created an efficient way to train machine learning models on battery data for predicting the SoH. The pre-processing enables various different data origins to be used. With the usage profile of the batteries, the original powered object (e.g., E-Bikes or E-Scooters) does not matter. The results of a machine learning workflow run are easily accessible displayed in a webapp.
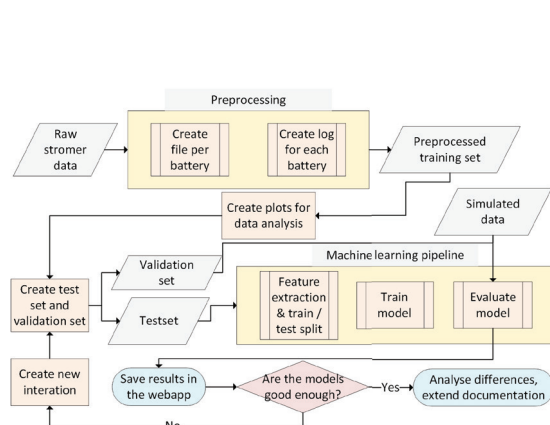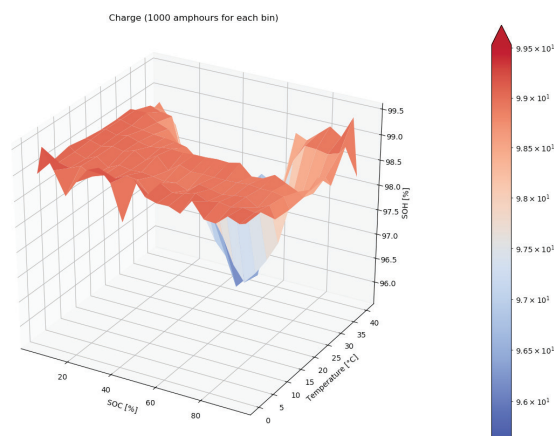Following projects by the ESReC could use our research and environment as baseline.

Christian Haldi

Nicola Florin Michaelis

Machine learning workflow

Charge (1000 amphours for each bin)

Charge wear per SoC and temperature bin