**SCHEDULE 1:**

**THE PROJECT**

**PROJECT TITLE**: **Building smart power grids**

**Description of the Project**

Toumetis is a small scale-up company based in Bristol developing software with cutting edge machine learning for engineering applications. Our flagship platform; Cascadence, has found success in the power utility and renewable energy generation industries and has been nominated for two E&T innovation awards. Our stakeholders use the platform for preventative maintenance to avoid power outages as well as for public and environmental safety, such as wildfire prevention.

The problem is we have a huge amount of data, meaning it is very diff icult for our subject matter experts to comprehend what the data means at this scale. Therefore we would like an intelligent visualisation tool to help these engineers understand this data. We hope that you can build this for us!

The idea we have is to generate synthetic data consisting of voltage and current waveforms with disturbances using numerical methods. Then use this synthetic data to train a deep neural network for identifying disturbances types for use on real data (as we lack labels for the real data) collected from monitors installed on power grids. This real data will feed into a visualisation tool for further investigation by experts and may incorporate dimensionality reduction, unsupervised clustering or other machine learning methods to help with visualisation. Our data scientists may find some of the processed data useful too, so an API to pull data from this app is another option.

This is an ambitious project but one with the option to use a huge amount of cutting edge machine learning/AI methods, which we can provide some guidance in, and package this up in a usable web-app or similar piece of software. Incorporating machine learning into software is greatly in demand so we feel students will get a lot out of this project and we look forward to seeing what they come up with.

black: finished

red: not started yet

blue: in progress

**Deliverables**

There are two main components:

1. A tool to classify power quality events according to IEEE standards [2].
   1. The CNN should be trained on synthetic waveforms.
   2. Synthetic waveforms should be generated according to IEEE power disturbances definitions [2]
   3. [Stretch] All IEEE disturbances listed in [1] should be used
   4. [Stretch] Further disturbances, such as combinations of IEEE power disturbances or additional categories not included in [1] could be used.
   5. As an input, the tool will be expected to take in sets of power quality (PQ) events.
      1. this could be in the form of pointing to a directory where PQ files are stored,
      2. a list of paths to the PQ csv files
      3. or a single file containing multiple events
   6. The csv format will be 10 cycles long at 256 samples per cycle, with 6 columns [Ia, Vab, Ib,Vbc,Ic,Vca] (the phases) + index (time since start),
   7. [Stretch] - The service should be flexible to the length of this file changing to accommodate larger PQ events
   8. The tool will then utilise a CNN model as described in [1] to classify the models according to IEEE waveform classifications [2] for each channel of the PQ csv file
   9. Performance of the CNN must be comparable to that reported in [1]
   10. [Stretch] Explore changes to the model by making changes to the hyperparmeters, training sets, architecture, feature processing or other method(s) to improve performance of the CNN
   11. [Stretch] The tool should be easily usable by data scientists or other technically capable researchers, such as via a CLI tool or similar.
   12. The output is expected to be a csv with the probability of a given IEEE classification [2] for each phase/channel of the input csv file, (e.g. 6 IEEE classifications \* 6 channels for the csv, total 36 columns)
2. A self-contained local web/browser-based application to view classifications from the classification tool/CNN model and assist subject matter experts (SMEs) to understand and label the event data.
   1. The app will receive the features from the CNN output for the probability that a waveform should fall into a particular IEEE classification [2] and the original waveform data.
   2. It will then use dimensionality reduction to reduce the CNN features down to 2 dimensions for a scatter plot.
   3. [Stretch] The user should be able to adjust the hyperparameters for the dimensionality reduction algorithm manually.
   4. [Stretch] The user should be able to select from more than one dimensionality reduction algorithm
   5. The user should also be able to choose an unsupervised algorithm (e.g. clustering, K-nearest neighbours) to group the data within the scatter plot.
   6. [Stretch] The service should allow for the hyperparameters of the unsupervised algorithm to be user modified.
   7. [Stretch] The user should be able to choose from more than one unsupervised algorithm.
   8. The user should be able to select a point(s) within the scatter plot and see the corresponding waveform
   9. The user should be able to see this point and compare its waveforms alongside similar events within the local cluster/group.
   10. The user should then be able to group and label sets of waveforms based on patterns they may observe from the output of (b.4)
   11. Any grouped/labelled waveform sets should be able to be exported from the service in the form of a csv which contains: the user-defined name of the group, the waveforms or event ids and paths to the csv that make up the group

[1] Cai K, Hu T, Cao W, Li G. Classifying Power Quality Disturbances Based on Phase Space Reconstruction and a Convolutional Neural Network. *Applied Sciences*. 2019; 9(18):3681. <https://doi.org/10.3390/app9183681>

[2] IEEE Recommended Practice for Monitoring Electric Power Quality; IEEE Standard 1159–2009; IEEE: Piscataway, NI, USA, 2009.