Replication

The code is split into 3 parts:

- Linear baseline.ipynb
- Shallow neural network.ipynb
- Statistical Hypothesis test.ipynb

The first two files can be run either way round. There are two methods depending on how many iterations you want to run the experiment for: it takes ~13hrs to retrain the shallow neural network 30 times for all 9 workloads like in the report, but if you just want to confirm the results reported in Figures 2 and 5, then 1 iteration should be enough (~30 mins). You would obviously need to run for multiple iterations to complete the Mann-Whitney U test seen in Figure 4, which can be found in **Statistical Hypothesis test.ipynb**.

- **If you want to run the models once**:
  1. Run all cells in order in **Linear baseline.ipynb** and **Shallow neural network.ipynb**.
  2. **Read the printed outputs** of each notebook (just the final cell for **Shallow neural network.ipynb**).
  3. They should list the data set being trained on, followed by the results (MAPE, MAE and RMSE on the test set) on the trained model. These should be very similar to the results in Figures 2 or 5 in the report (there is some stochastic bias so they may not be exact). The **Shallow neural network.ipynb** will also display a graph after training to show the training loss against test loss (should match the relevant graph in Figure 3).

- **If you want to run the models for multiple iterations**:
  1. Change the repetitions variable in the first two notebooks to the number you want (**line 36** for **Linear baseline.pynb** and **last cell, line 21** for **Shallow neural network.pynb**). The results in the report are over 30 iterations, but as mentioned, that takes a long time (~13hrs).
  2. Run all cells in order in Linear baseline.ipynb and Shallow neural network.ipynb.
  3. Instead of just reading outputs, it would be highly recommended to use the "Collated test analysis" spreadsheet in the Data Analysis folder. Each notebook writes its results to this folder, and this spreadsheet collates all the csv files to produce the tables you find in Figures 2 and 5.
  4. Run **Statistical Hypothesis test.ipynb** and read the printed outputs, they should closely match Figure 4.