

Stage 1

Load in data:

- REFIT
- UCID

Data cleaning:

- Account for missing days
- Account for incomplete days

Stage 2

Dimensionality reduction:

- SP: $n^{96} \rightarrow n^{20}$
- t-SNE/UMAP: $n^{20} \rightarrow n^2$

Apply clustering algorithm:

- HDBSCAN

Extract daily load profiles:

- Evaluate the resulting clusters
- Manual inspection of results

Stage 3

Feature engineering:

- Temporal data
- Weather data (Solcast)

Feature selection:

- Correlation/causation
- Statistical significance

Data preprocessing:

- Data smoothing
- Identify and remove outliers

Train classification and regression tree

Trained classification and regression tree

Stage 4

Train-validation-test split:

- 70:15:15 ratio
- Per cluster

Train forecasting models:

- CNN-LSTM hybrid network

Optimize model:

- Cross-validation
- Hyperparameter tuning

Trained forecasting model

New sample

