NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
School of Computer Science and Engineering
College of Engineering

Time Series Classification (TSC)

Development of Feature-based Time Series Classification Tools for Multivariate Time-series Data

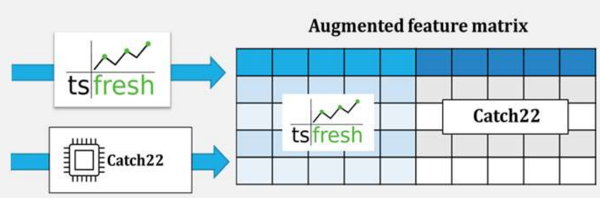
Student: Serene Quek Kai Zhen **Supervisor**: Assoc Prof A S Madhukumar

This project developed a pipeline to execute Feature-based Time series classification (TSC) to predict the label of a time series. The pipeline is made up of 3 parts, Feature Engineering (FE), Feature Selection (FS) and predictive modelling.

In Feature-based TSC, the time series input is converted from the time domain into a feature set of its characteristics. The characteristics are extracted using 2 state of the art FE methods: TSFRESH and Catch22.

The datasets created through FE are combined into one dataset to create a comprehensive augmented feature matrix (AFM).

ID	Time (T)	Value
1	1	1.98
1	...	...
1	15	2.41
4	1	1.19
4	...	...
4	15	0.05



As FE generate many features, FS identifies the most relevant features from the AFM through the following methods:

(1) Laplacian Score ranks the features based on its ability to preserve locality.

Rank	Feature
1	Feature_19
2	Feature_2
3	Feature_31
4	Feature_N
…	…
N	Feature_75

(2) TSFRESH feature selection determines if a feature is relevant with Hypothesis Testing.

Relevant	Not Relevant
Feature_34	Feature_21
Feature_76	Feature_3
Feature_81	Feature_11
…	…

Finally, we make use of the generated and selected dataset for predictive modelling with CatBoost which uses a series of gradient boosted decision trees to perform the classification task.

CatBoost