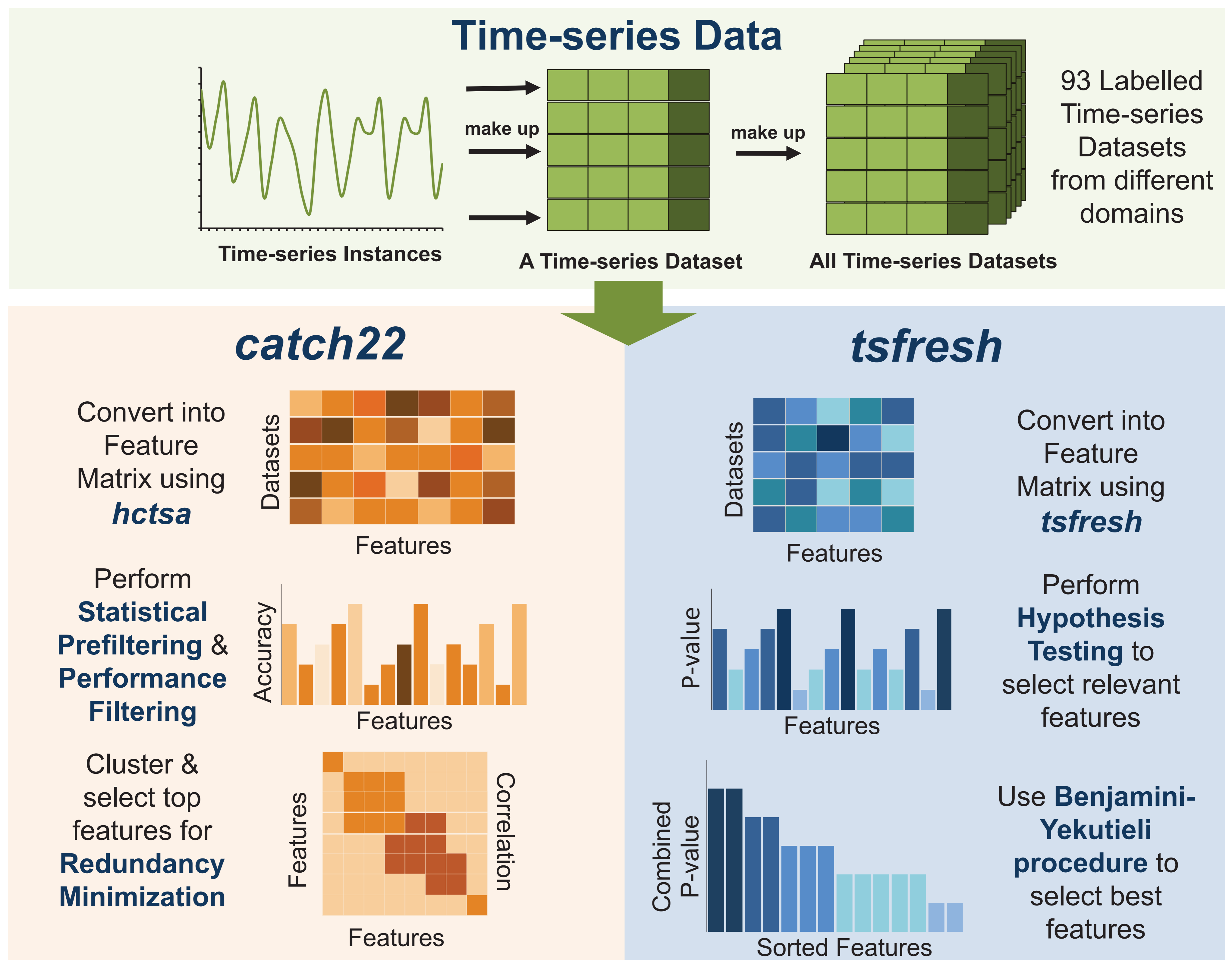# Analysis on the Impact of Feature Reduction on Time-Series Data

Student: Ernest Tan Yong En   **Supervisor:** Assoc Prof A S Madhukumar

Comparison of catch22 and tsfresh Feature Extraction Pipelines

Automated Feature Engineering is an emerging field that significantly reduces the time, effort, and domain knowledge needed to perform time-series classification. In this project, we explore 2 cutting edge methods shown below: catch22 and tsfresh

Time-series Data

Time-series Instances   make up   A Time-series Dataset   make up   All Time-series Datasets

93 Labelled Time-series Datasets from different domains

catch22

Convert into Feature Matrix using **hctsa**

Datasets / Features

Perform **Statistical Prefiltering** & **Performance Filtering**

Accuracy / Features

Cluster & select top features for **Redundancy Minimization**

Features / Features / Correlation

tsfresh

Convert into Feature Matrix using **tsfresh**

Datasets / Features

Perform **Hypothesis Testing** to select relevant features

P-value / Features

Use **Benjamini-Yekutieli procedure** to select best features

Combined P-value / Sorted Features

Results

The resulting feature sets from both pipelines were evaluated on the time-series data to compare their performance. The results showed that *catch22* seems to perform better than *tsfresh* in terms of **F-score** but exhibits similar performance on **class-balanced accuracy**. However, catch22 requires significantly longer computation time to generate the feature set.