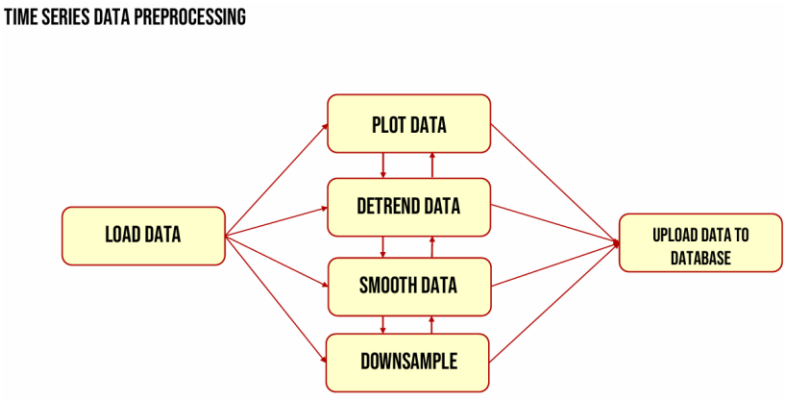
Report on Time series Data preprocessing Tools in Cogniforge

The code developed in this section includes tools as part of Cogniforge to preprocess Time Series Data before using them for other analysis. This document delineates the features that are developed in this Winter Semester 2024/2025 and provides an overview of their current basic functionalities.



1. **Load Data tab**

Key features: Load data from Furthr database, preview, data selection and choosing sample for subsequent analyses.

User workflow:

- User selects data source from Furthr Database
- Preview of the first 10 rows is shown
- Users can choose between using the entire dataset or only a range of rows
- Chosen data is processed, and a successful message is displayed to the user

After loading the data, users can move freely back and forth from any of the analysis tabs, i.e. users can choose to apply only the preprocessing steps in whatever order they see fit.

2. **Plot Data:**

Key features:

- Plot several columns to plot at once
- Normalize data: enable comparison data with different scales in one plot.
- Interactive zoom and time slider: enable user to specifically explore specific data points.
- Collapsible Widget for ‘Analysis status’: keep track of analysis performed on the data

User Workflow

- Pick which columns you want to visualize
- Option to normalize data to better compare different measurements

3. Detrend Data

Key features:

Trend Detection

- **Statistical Analysis**: Uses R-squared values from simple linear regression, p-values, and unit root test (in this code: ADF) to detect trends
- **Automatic Recommendations**: Suggests appropriate detrending methods based on data characteristics.
- **Collapsible Visual Inspection**: Provides a so-called “before and after” of the original and detrended data and an expandable widget**.**

Detrending Methods:

- **Linear Detrending**: Removes linear trends using polynomial fitting
- **Moving Average Detrending**: Removes complex non-linear trends using rolling windows
- **Parameter Customization**: Allows manual adjustment of detrending parameters.

User Workflow

- **Select Columns**: User chooses which columns to analyze for trends
- **Review Recommendations**: System detects trends and recommends detrending methods
- **Customize Parameters**: User can adjust parameters like window size for moving averages
- **Visualize Results:** User explicitly clicks to view interactive plots in a collapsible panel, showing original data, trend line, and detrended data
- **Track Changes**: System records all detrending operations in analysis history
- **Revert if Needed**: User can revert to the original data if the detrending results are not satisfactory

4. Smooth Data

Key Features

Noise Detection

- **Automated Noise Analysis**: Uses multiple statistical indicators to determine if smoothing is needed: Checks rolling standard deviation, outlier ratios, and local changes.

Smoothing Methods

- **Exponential Smoothing**: Applies weighted averaging where recent observations receive more weight.
- **Parameter Recommendations**: The code computes optimal smoothing parameter based on volatility and outlier prevalence.

User Experience

- **Column-by-Column Analysis**: Analyzes each selected column individually with dedicated tabs
- **Collapsible Visual Inspection**: Provides before-and-after visualizations in expandable panels that users can open as needed
- **Smoothing Statistics**: Offers quantitative metrics on noise reduction effectiveness
- **Analysis History**: Tracks all applied smoothing operations

User Workflow

- **Select Columns**: User chooses which columns to analyze for noise
- **Review Recommendations**: System detects noise and recommends smoothing parameters
- **Customize Parameters**: User can adjust smoothing strength (alpha) using a slider
- **Visualize Results**: User explicitly clicks to view interactive plots in a collapsible panel, showing original and smoothed data. Result of each variable is displayed in a separate tab.
- **Apply Smoothing**: User can apply smoothing to selected columns
- **Track Changes**: System records all smoothing operations in analysis history
- **Revert if Needed:** User can revert to the original data if the results are not satisfactory

5. **Downsample Data**

Key Features

- **Downsample using LTTB** : LTTB divides data into buckets and picks points that best represent the overall shape
- **Sample Size Control**: User can specify the desired number of data points

User Workflow

- **Configure Parameters**: User sets the desired sample size (number of points to retain)
  - **Collapsible Visual Inspection**: Provides a so-called “before and after” of the original and detrended data and an expandable widget**.**
  - **Apply Downsampling**: User can apply smoothing to selected columns
  - **Track Changes**: System records all downsample operations in analysis history
  - **Revert if Needed:** User can revert to the original data if the results are not satisfactory

6. **Upload Data**

Key Features

Upload Options

- **Choose Dataset Versions:** Choice between uploading original data or fully processed data with all analyses
- **Location Selection**: Interactive selection of project, group, and experiment for upload
- **Upload Progress** Tracking: Visual feedback during the upload process
- **Status Reporting**: Clear success/failure feedback with detailed error messages

User Workflow

- **Select Upload Location**: User navigates through project, group, and experiment selection
- **Choose Dataset Version**: User selects between original or processed data
- **Review Information**: System displays row count and available columns (for double check)
- **Monitor Progress**: System provides visual feedback during the upload process
- **Receive Confirmation**: User receives success message or detailed error information