Efficient AI with Rust Lab
Rapid Time Series Datasets Library
RWTH Aachen University

Group 1

Marius Kaufmann[1] Amir Ali Aali[2] Kilian Fin Braun[1]

[1]Masters of Computer Science
[2]Masters of Data Science

14th Jun, 2025

RWTH AACHEN UNIVERSITY

Marius's Part

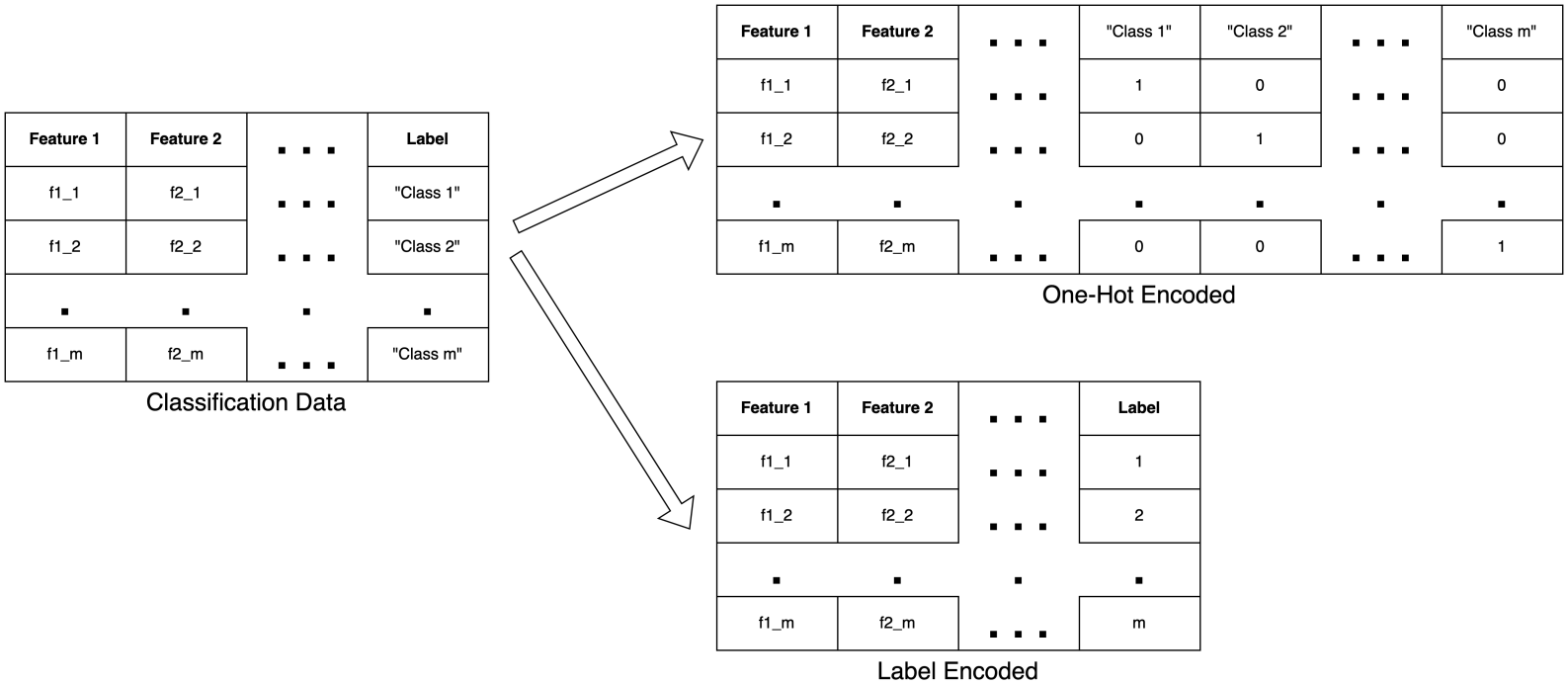
Data Abstraction I

In time series datasets, we often have to deal with mainly two types of data:

- **Forecasting Data:**
  - Contains only floating point values
  - Used for predicting future values
  - Example: Stock prices, weather data
- **Classification Data:**
  - Contains a mix of floating point and categorical values
  - Used for classifying time series data into categories
  - Example: Medical data, sensor data

Data Abstraction II

We require categorical columns to be provided as either one-hot or label-encoded values.
This enables us to save both datasets in a unified way, which is a table of floating point values.



Data Point Representation

For our current implementation, we defined a function *.get(index)* that returns a data point at the given index.
In each of the two dataset types, we have a different representation of the data point.

- **Forecasting Data Point:**
  - **ID:** A unique identifier for the data point
  - **Past:** A vector of floating point values representing past observations
  - **Future:** A vector of floating point values representing future observations
- **Classification Data Point:**
  - **ID:** A unique identifier for the data point
  - **Features:** A vector of floating point values representing the features of the data point
  - **Label:** A vector of floating point values representing the label of the data point

Splitting Strategies

As one of the main features of our library, we provide different splitting strategies for the datasets.

- **Random Split:**
  - Randomly splits the dataset into training and test sets
  - Can be used only for classification data
- **Temporal Split:**
  - Splits the dataset ordered by time
  - Can be used for both forecasting and classification data

Kilian's Part