**RAL\_TCM Dataset Readme**

1. **Overview of the Dataset**

The dataset, named RAL\_TCM (Tandem Cold-rolling Mill), is a large-scale benchmark dataset designed for multivariate time series forecasting in the steel industry, particularly targeting complex process operations in process industries. The RAL\_TCM dataset was collected from the tandem cold rolling production line of a steel company in Jiangsu Province, China, through Level-2 process control systems and PDA systems. It encompasses almost all critical process parameters involved in the tandem cold rolling production process.

1. **The Composition of the Dataset**

* **Data Source**: The tandem cold rolling production site of strip steel at a steel enterprise in Jiangsu Province, China.
* **Collection System:** Recorded through various sensors, instruments, and automatic control systems at the rolling production site, as well as data acquisition PDA systems.
* **Data content**:
  + **- Number of strip coils: 5**
  + **- Time steps**： A total of 20,299 time steps with millisecond time granularity
  + **- Number of** **process variables:** Each data point consists of 170 process variables
* **- Target predicted variables (8)**：
  + - Rolling forces for Stands 1 to 5:

RFC\_STD#1

RFC\_STD#2

RFC\_STD#3

RFC\_STD#4

RFC\_STD#5

* + - Thickness of outlet strips for Stands 1, 4 and 5:

THK\_STD#1,

THK\_STD#4,

THK\_STD#5

1. **Characteristics and split of the dataset**

* **High-time resolution**: Millisecond data acquisition provides high temporal resolution data for high-precision prediction tasks.
* **Multivariate Temporal Sequencing**: Multivariate time-series data with 170 process variables per time step captures multiple complex relationships in the production process and provides sufficient information for modelling.
* **Large-scale data**: The data volume of 20,299 time steps provides rich samples for model training and validation。
* **Split of the dataset**： The recommended split ratio of training, validation, and test sets is 7:1:2.

**4. Dataset applications**

* **Application Fields**： The dataset can be used for research on topics such as machine learning, time-series forecasting models, process optimization, production quality control in artificial intelligence, etc.
* **Forecasting tasks**： The main task is to predict key process parameters (e.g. rolling force and strip thickness) that contribute to the productivity and product quality of the strip cold rolling process.

**5. 数据集获取与使用**

* **Github Open Source URL**： RAL\_TCM Dataset | Github(地址).
* **Copyright Statement**： This dataset and the associated work are provided by the State Key Laboratory of Rolling and Automation (RAL), Northeastern University, Shenyang 110819, Liaoning, PR China. The dataset is utilized exclusively for scientific research purposes. If used, proper citation in the literature is required. Please cite it as follows: **论文引用格式**

**6. Potential application areas for the dataset**

Engineers in the steel industry can use the dataset for precise control and optimization of production processes, while data scientists and AI experts can use it to develop and validate new time-series forecasting algorithms to drive process industry intelligence. In conjunction with the dataset, the performance of different models can be objectively evaluated, providing data to support further process improvements. At the same time, it provides the basis for creating more informative sensor layouts and improving equipment regulation strategies.