Machine Learning-Based Estimation of Monthly GDP

This repository is created to replicate the results from the study "*Machine Learning-Based Estimation of Monthly GDP.*"

The repository consists of three main directories:

1. `data/`
2. `model/`
3. `figures/`

Each directory contains subfolders for the countries analyzed in the study: **China**, **Germany**, **the UK**, and **the US**.

1. `data/` directory

This folder includes all source data and preprocessing scripts.

- **`collecting_data.ipynb`**
  This notebook collects macroeconomic indicators via public APIs, including FRED and Yahoo Finance.
  To use the FRED API, you must register for a (free) API key at https://fred.stlouisfed.org/.
- **`merging_data.ipynb`**
  This notebook merges and cleans the collected data into a unified file format:
  `master_<country>.csv`, which is used in model training.
  While you can collect the data yourself, we also provide pre-processed datasets for each country in this directory.

2. `model/` directory

This folder contains modeling scripts organized by country.

Each country folder contains implementations of the following models:

- **MLP (Multi-Layer Perceptron)**
- **LSTM (Long Short-Term Memory)**
- **XGBoost**
- **Elastic Net**

Training and evaluation results are automatically saved as `.csv` files. These results are later used to generate tables and figures for the paper.

3. `figures/` directory

This folder creates final figures and tables used in the paper, based on model outputs from the `model/` directory.

Before running the scripts here, you need to manually compile a country-level summary result file and save it as a `.csv`.
Reference `.csv` files used in the paper are included and can serve as examples.

Contact

For questions regarding this project, please contact:

Yonggeun Jung
✉ yonggeun.jung@ufl.edu