The codebase is organized into the following components:

1. **Libraries** – Importing the necessary Python libraries required for execution.
2. **Data Import** – Loading input files located in the “Initial indexes data” folder.
3. **Auxiliary Functions** – Supporting functions required for data processing and model evaluation:
   * **create\_dataset** – Converts data into the appropriate format for prediction tasks.
   * **calculate\_metrics** – Computes statistical performance metrics (MSE, MAE, R²).
   * **generate\_signal** – Helper function for generating trading signals (buy, sell, hold).
   * **calculate\_trades** – Implements a trading strategy based on generated signals.
   * **calculate\_return\_and\_drawdown** – Calculates portfolio returns and maximum drawdown.
   * **financials** – Performs the final financial calculations, producing all relevant performance metrics.
   * **prepare\_data** – Prepares the dataset for use in machine learning algorithms.
4. **Models** – Implementation of predictive models used in the study:
   * LSTM
   * GRU
   * DMLP
   * XGBoost
   * Transformer
5. **Training and Ensemble Preparation** – For each index and model, optimal hyperparameters were identified using data from 2020–2024. The resulting training and test datasets were then consolidated for financial analysis. These processed datasets are stored in the “Final indexes predictions” folder.
6. **Data Reading** – Scripts to load the processed files from the “Final indexes predictions” folder.
7. **Financial Data Preparation** – Adds any missing columns required for financial calculations if files were loaded from disk.
8. **Ensemble and Financial Analysis** – Functions related to the ensemble approach and financial evaluation:
   * **Saving final results** – Option to save outputs from the financial analysis.
   * **Loading final results** – Loads results from final\_results\_df.xlsx for review.
   * **Scatter plot of R² vs. Total Return** – Visualization of predictive accuracy against profitability.
   * **Heatmaps** – Comparative visualizations for R² values and total returns.