GDP Prediction and Analysis Project

This repository contains all the resources and code used for analyzing and predicting GDP trends for the USA and Japan, along with the corresponding visualizations, workflows, and results.

Project Structure

The repository is organized into the following folders:

1. Preprocessing Codes

Contains all the Python scripts (in `.ipynb` format) used for data preprocessing, including: - Cleaning the data - Feature engineering - Splitting the datasets into training and testing subsets - Normalization and scaling

2. CSV Files

Includes all the datasets used for: - Model training and testing - Generating visualizations

3. Images

This folder contains: - Final visualizations of GDP trends and comparisons - Tables summarizing key metrics - Workflow diagrams explaining the methodology

4. Rough Code

Contains experimental Jupyter Notebook files (`.ipynb`) with exploratory analyses and tests conducted during model development.

5. Demo PPT

Includes the presentation used for showcasing the project and summarizing its objectives, methodologies, results, and conclusions.

Getting Started

Prerequisites

Ensure the following libraries and tools are installed: - Python 3.x - Required Python packages: `pandas`, `numpy`, `matplotlib`, `seaborn`, `scikit-learn`, `statsmodels`

Usage

Preprocessing

1. Navigate to the `Preprocessing Codes` folder and open the relevant Jupyter Notebooks to preprocess the data.

Model Training and Testing

1. Use the CSV files in the `CSV Files` folder for model training and testing.
2. Experiment with different models and parameters using the notebooks in the `Rough Code` folder.

Visualizations   Run the visualization scripts or open the provided notebooks to generate charts and tables

Demo   Open the `Demo PPT` file to review the presentation and results.

Outputs

The project outputs include: - Predictions of GDP trends for the USA and Japan from 2024 to 2029. - Comparisons of model performance metrics (e.g., MAPE, accuracy). - Visualization of actual vs. predicted GDP values. - Workflow diagram outlining the visual analytics pipeline.

Key Highlights

- **Accurate Models**: Models such as Linear Regression, Polynomial Regression, ARIMA, and Random Forest were evaluated.
- **Comparative Analysis**: USA and Japan GDPs were compared across multiple metrics.
- **Visual Analytics Workflow**: A systematic approach was adopted for preprocessing, modeling, and visualization.

License

This project is licensed under the MIT License.

Acknowledgments

Special thanks to the contributors and reviewers who provided feedback and helped refine this project.