Predicting GDP per Capita Growth in Emerging Market Economies Using Machine Learning

Group 6: Matthew Kwon, Piper Lickteig, Yewen Liu, Jesse Schmolze, Liufeng Yang

Research Question

Can machine learning accurately predict future GDP per capita growth in emerging markets using historical economic, demographic, and governance data?

Data Sources

We have three files in our dataset folder: `Global_Data.csv`, `Demographics.csv`, and `wgidataset_excel.zip`.

- `Global_Data.csv` (1960–2023): Includes macroeconomic indicators such as trade openness, inflation, FDI, investment ratio, etc.
- `Demographics.csv`: Contains population growth, median age, education, and labor force statistics.
- `wgidataset_excel.zip` (1996–2024): Extracted from Worldwide Governance Indicators (WGI) dataset. Includes six variables: voice & accountability (va), political stability (pv), government effectiveness (ge), regulatory quality (rq), rule of law (rl), and control of corruption (cc).

Variables

Target (Y): GDP per capita growth rate.
Predictors (X): Governance variables (va, pv, ge, rq, rl, cc), demographic data, and macroeconomic variables.

Methods

We plan to use four machine learning models to predict GDP per capita growth:

1. Linear Regression
2. Lasso Regression
3. Random Forest Regressor
4. K-Nearest Neighbors (KNN)

We will evaluate performance using Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2.

Python Code Example

```
import pandas as pd

# Load datasets
macro = pd.read_csv("Global_Data.csv")
demo  = pd.read_csv("Demographics.csv")
wgi   = pd.read_excel("wgidataset_excel.xlsx")

# Rename and merge
wgi = wgi.rename(columns={"countryname": "Country", "year": "Year"})
df = pd.merge(macro, demo, on=["Country", "Year"], how="inner")
df = pd.merge(df, wgi, on=["Country", "Year"], how="inner")

print(df.shape)
```

Notes

All variables are derived from World Bank and WGI datasets. Ensure consistent variable naming and documentation for reproducibility.