



# Machine Learning for Predicting GDP Growth

---

ENRICO FIORI FERRAZ

# IMPORTANCE OF PREDICTING GDP GROWTH

- In economic planning and policy-making, GDP growth is one of the most important aspects for a country development;
- Predicting GDP growth is challenging, considering its dependence on a lot of factors, which can be difficult to accurately define;
- The scope of this project is to investigate how government debt and others factors can be used to predict GDP growth.

# ABOUT THE DATASET

- Main Source: World Bank
- Time Frame: From 1980 to 2020
- Number of countries: 196
- Features: 17 (all numerical)
- GDP Growth is the target!

## Problems in Dataset:

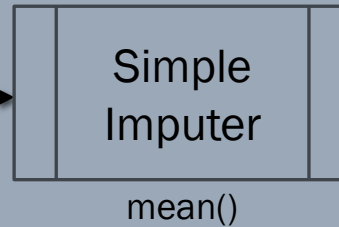
Almost half of the columns have **more than 20%** null or NaN values

# PREPROCESSING AND DATA CLEANING

## Raw data

Total investment (% of GDP)	1720
Gross national savings (% of GDP)	1744
General government revenue (% of GDP)	2315
General government total expenditure (% of GDP)	2368
General government net lending/borrowing (% of GDP)	2381
General government primary net lending/borrowing (% of GDP)	2747
General government net debt (% of GDP)	5789
General government gross debt (% of GDP)	3056
Current account balance (% of GDP)	1149
GDP GROWTH	0
Exports of goods and services (% of GDP)	861
Inflation, consumer prices (annual %)	861
Imports of goods and services (% of GDP)	861
Expense (% of GDP)	861
Domestic credit to private sector (% of GDP)	861
Research and development expenditure (% of GDP)	861
Tax revenue (% of GDP)	861

dtype: int64



## Processed data

Total investment	0
Gross national savings	0
General government revenue	0
General government total expenditure	0
General government net lending/borrowing	0
General government primary net lending/borrowing	0
General government net debt	0
General government gross debt	0
Current account balance	0
GDP GROWTH	0
Exports of goods and services (% of GDP)	0
Inflation, consumer prices (annual %)	0
Imports of goods and services (% of GDP)	0
Expense (% of GDP)	0
Domestic credit to private sector (% of GDP)	0
Research and development expenditure (% of GDP)	0
Tax revenue (% of GDP)	0

## MODELS TRAINED

### Extra Trees Regressor

- Good for GDP growth forecasts as it balances between randomness and precision in economic predictions

### Random Forest Regressor

- Suitable for GDP growth prediction as it handles complex relationships and interactions among the data.

### K Neighbors Regressor

- Effective for GDP growth prediction due to its reliance on nearby observations for forecasting.

## PERFORMANCE METRICS

MODEL	MAE	MSE	R-SQUARED	N. ESTIMATORS	N. NEIGHBORS
Extra Trees	2.5060	19.1982	0.1983	300	-
Random Forest	2.5740	20.4767	0.1449	300	-
K-Neighbors	2.7167	218187	0.0888	-	7

**MAE** is the best metric for this problem because the idea is to know how much the predictions on average deviate from real values of GDP Growth

## CONCLUSION AND NEXT STEPS

- Find more data to accurately fulfill the gaps in the original dataset;
- Even the best model (Extra Trees) trained could not properly predict the GDP growth with the features provided;
- For the problem the model is tackling, ~2.5% of error is a lot;
- Get more data of different indicators, not only government and trade data.

