Project Proposal: Economic Impact Analysis Using Machine Learning Techniques

Project Overview

The objective of this project is to analyze the economic impact of certain key factors on a country's GDP growth over time. Utilizing the provided dataset, which encompasses data from various countries across multiple years, this study aims to examine how variables such as population, investment rates, and other economic indicators influence the GDP of these nations.

Research Question

The main research question is: "How do key economic factors affect the GDP growth of different countries over time, and can machine learning models provide accurate predictions and insights into these relationships?"

Data Description

The dataset contains historical economic data for multiple countries. Key variables include Year, Country, Population, Real GDP measures, Investment rates

Other economic indicators

There are 59 columns in total, indicating a rich set of variables for analysis. The dataset spans from the year 1870 onwards, providing a substantial historical perspective.

Methodology

Machine Learning Task: This project will primarily involve a regression analysis, given the continuous nature of the GDP variable.

Techniques: We plan to employ various machine learning techniques covered in class, such as linear regression, decision trees, and random forests. Regularization methods (like LASSO or Ridge Regression) might be used to handle the high number of predictors and avoid overfitting.

Data Processing: The dataset will be preprocessed to handle missing values, normalize data, and encode categorical variables.

Computational Tools

Language: R (primary) and Python (optional for those comfortable with it).

Libraries: In R, we'll use tidyverse for data manipulation, ggplot2 for data visualization, and packages like caret and randomForest for machine learning. In Python, pandas, scikit-learn, and matplotlib will be our primary libraries.

This proposal aligns with the project guidelines, focusing on a relevant economic question and employing the machine learning techniques discussed in the course. The analysis will be documented in a paper not exceeding 20 pages, submitted in R Markdown format. ​​