Stock Prices and Analyzed Relation with Kconomy (SPARK)

Team Members: Shekhar Sharma, Royyam Saqib, Mina Gong, Liwei Huang |
Github Repository

Introduction

Economic growth of a country is typically measured by its Gross Domestic Product (GDP) per capita, which provides insight into the average economic well-being of a country's citizens. A rising GDP per capita signals improving economic conditions, attracting multinational firms seeking profitable investment opportunities. Such investments further contribute to economic expansion, creating a positive cycle that drives sustained growth.

GDP per capita is influenced by various economic and socio-economic factors, including Foreign Direct Investment (FDI), Gross Capital Formation, and even less intuitive indicators like mobile cellular subscribers and life expectancy.

Among these, corporate investment plays an important role. Since the performance and investment behavior of these companies are often influenced by the broader economic environment, stock market activity can serve as a reflection of their economic sentiment and trend. As a result, stock performance has the potential to be an indicator of a country's overall economic conditions.

Project Goal

This project aims to develop a predictive model for GDP per capita (US$) in G20 countries and identify key economic indicators that influence it. Specifically, we evaluate the role of stock market indicators in predicting economic conditions. By doing so, we seek to uncover meaningful correlations between stock market indicators and a country's GDP per capita. Establishing such links could provide valuable predictive insights for both governments and corporations, enabling more informed decision-making in areas such as investment strategy, policy planning, and economic forecasting.

Stakeholders and KPIs

- Multinational firms with investments in the respective nations
- Government agencies/Banks
- GDP per capita
- Model accuracy metrics (R^2, RMSE)
- Feature importance scores (economic indicators significantly impacting GDP per capita)
- Stock market performance indicators

Data Acquisition and Description

We collected datasets from the World Bank Open Data Platform, which provides a collection of time series data with a wide range of global indicators from 1960 to 2023. The dataset

encompasses economic, social, environmental, and demographic metrics. We focused exclusively on data from G20 countries, as our analysis centers on their GDP per capita. This scope provides more meaningful insights into global development trends and socio-economic analysis.

Exploratory Data Analysis and Feature Engineering

Since GDP per capita is more likely to be influenced by the values of indicators from recent years rather than current-year values, we converted all features into lag-1 variables (i.e., values from the previous year) for inclusion in the GDP per capita prediction model. After filtering out features with more than 30% missing data, we selected a range of variables likely to impact GDP per capita based on domain knowledge.

Modeling and Analysis

First, we developed a baseline predictive model for GDP per capita using the selected features, excluding the ‘Stock traded total value’ variable.

We employed two types of regression models: (1) Linear Regression with Ridge and LASSO regularization, and (2) a Random Forest Regressor as a non-linear alternative. Ridge regression is well-suited for handling multicollinearity by shrinking coefficients, while LASSO not only addresses multicollinearity but also performs feature selection by driving some coefficients to zero. Random Forest, a tree-based ensemble method, is robust to overfitting and can effectively capture complex, non-linear relationships between predictors and the target variable.

After establishing the baseline, we incorporated the ‘Stock traded total value’ feature into the model to evaluate its additional predictive value. We then compared the model performance to assess the impact of including stock market activity as an economic indicator.

Results

All baseline models without the ‘Stock traded total value’ feature showed relatively low performance. In particular, the Random Forest model had a cross-validation mean R^2 score of 0.44 on the training set and 0.48 on the test set, with corresponding mean RMSE values of 11,518 (training) and 12,657 (testing).

After incorporating the stock market indicator (‘Total Value of Stock Trading’) into the Random Forest Regressor, the model’s performance significantly improved. The cross-validation mean R^2 score on the training set increased to 0.79, with a reduced RMSE of 7,804. Following hyperparameter tuning, the model achieved an R^2 of 0.89 and an RMSE of 5,355 on the test set. This indicate

The finalized Random Forest model also provided feature importance scores, highlighting the most influential predictors of GDP per capita. These included the one-year lag of ‘Stock traded

total value', 'School enrollment (secondary)', 'Life expectancy at birth', 'Mobile cellular subscriptions', and 'Electric power consumption'. Further details and visualizations can be found on the project's GitHub page.

Conclusion and Future Directions

The project successfully demonstrated that stock market activity, specifically the "Stock traded total value," is a significant predictor of GDP per capita. The Random Forest Regressor, when enhanced with this indicator, showed a substantial improvement in predictive power (R^2 of 0.89 and RMSE of 5,355 on the test set), highlighting its crucial role in economic forecasting. This strong correlation provides valuable insights for both governments and multinational firms in making more informed decisions regarding investment strategies, policy planning, and economic forecasting. We conclude that the project successfully demonstrate the stock market activity as an important predictor for the GDP growth, the project can further be extended in the following possible future directions :

- **Expand scope to include more countries/regions:** While the G20 countries provide meaningful insights, extending the analysis to a broader range of countries or specific economic indicators could reveal more diverse patterns and enhance the model's generalizability.

- **Incorporate additional financial indicators:** Explore the impact of other stock market indicators such as market capitalization, trading volume of specific sectors, or volatility indices. Additionally, integrate other financial metrics like interest rates, inflation, and bond yields to build a more comprehensive predictive model.

- **Explore advanced modeling techniques:** Investigate the use of time series models (e.g., ARIMA, LSTM, Prophet) that are specifically designed to capture temporal dependencies in data, potentially improving the accuracy of GDP per capita predictions.