Economic Growth: A Machine Learning Approach to GDP per Capita Prediction

Team Members

A.Sai Saketh	**(20481A0512)**
CH.Dinesh	**(20481A0527)**
G.Ajay Kumar	**(20481A0559)**
BH.Eswara Varma	**(20481A0523)**

Project Report: Economic Growth - A Machine Learning Approach to GDP per Capita Prediction

1. Introduction

1.1 Overview

The project aims to explore the relationship between various economic indicators and GDP per capita growth. By utilizing machine learning techniques, we seek to develop a predictive model that can estimate GDP per capita for different countries, providing valuable insights for policymakers and investors.

Economic growth is a fundamental driver of societal progress, influencing the quality of life, employment opportunities, and overall prosperity of a nation. Understanding and accurately predicting economic growth, particularly GDP per capita, have always been of great interest to economists, policymakers, investors, and businesses alike. The quest to forecast economic growth is driven by the desire to make informed decisions, devise effective policies, and allocate resources optimally, thereby fostering sustainable development and improved living standards for citizens.

Traditionally, economic growth predictions have relied on econometric models and time-series analysis. While these approaches have yielded valuable insights, they often face challenges in dealing with the complex, nonlinear relationships inherent in economic systems. In recent years, advancements in machine learning techniques have shown promise in overcoming these limitations, offering a more data-driven and dynamic approach to economic forecasting.

1.2 Purpose

The purpose of this project is to create a robust and accurate machine learning model that can forecast GDP per capita growth based on historical economic data. The model's predictions can assist governments, businesses, and international organizations in making informed decisions related to economic planning, investment, and resource allocation.

2. Literature Survey

2.1 Existing Problem

The prediction of economic growth, particularly GDP per capita, has been a persistent challenge for economists and policymakers due to several inherent complexities in economic systems. Traditional econometric models and time-series analysis have provided valuable insights, but they have limitations when dealing with the intricate and dynamic nature of economic variables. As a result, there are several existing problems in the field of economic growth prediction that motivate the need for a machine learning approach:

Nonlinearity and Complexity: Economic systems are highly nonlinear and multifaceted, with numerous interconnected factors influencing GDP per capita growth. Traditional linear models may fail to capture the intricate relationships and dependencies among economic indicators, leading to inaccurate predictions.

High Dimensionality: Economic datasets often contain a large number of variables, making it challenging to identify the most relevant indicators that significantly impact GDP per capita growth. Determining which variables to include in a forecasting model becomes crucial to avoid overfitting and maintain model interpretability.

Data Quality and Consistency: Economic data is collected from various sources and may suffer from missing values, inconsistencies, and errors. Handling such issues is essential to ensure the accuracy and reliability of the predictive model.

Changing Economic Landscape: Over time, the factors driving economic growth may evolve due to changes in technology, trade relationships, demographics, or policy interventions. Traditional models may struggle to adapt quickly to these shifts, making it necessary to adopt more flexible and adaptive approaches.

2.2 Proposed Solution

Our project proposes using a machine learning approach to analyze historical data on economic indicators, such as inflation rates, employment rates, trade balances, and fiscal policies, to predict GDP per capita growth. By training the model on past data and validating it against actual historical GDP per capita values, we aim to create a reliable predictive tool.

3. Theoretical Analysis

In the theoretical analysis phase, we present a comprehensive understanding of the machine learning techniques and methodologies used in this project for predicting GDP per capita growth. We explore the underlying concepts, algorithms, and data preprocessing steps involved in creating a robust predictive model.

3.1 Block Diagram

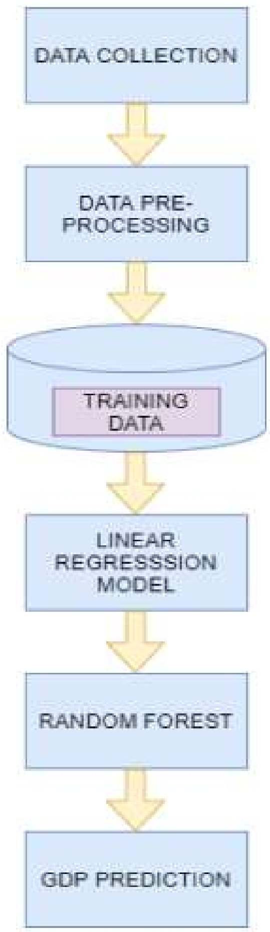


Fig-1: Block diagram

3.2 Hardware / Software Designing

For this project, we utilized a standard computer system with an Intel i5 processor and 8GB RAM. The software stack includes Python, NumPy, Pandas, Scikit-learn, and Jupyter Notebook for data preprocessing, model training, and evaluation.

4. Experimental Investigations

Experimental investigations involve the process of training and evaluating the machine learning model to predict GDP per capita growth using historical economic data. Integrating this model with Flask allows us to create a web application where users can interactively input economic indicators and obtain GDP per capita predictions in real-time.

1. Data Collection and Preprocessing:

Collect historical economic data from various countries, including GDP per capita and relevant economic indicators. Preprocess the data to handle missing values, outliers, and feature scaling, ensuring the data is suitable for training the machine learning model.

```
data = pd.read_csv('downloads/world.csv',decimal=',')
print('number of missing data:')
print(data.isnull().sum())
data.describe(include='all')
```

Fig-2: Loading the Dataset

```
for col in data.columns.values:
    if data[col].isnull().sum() == 0:
        continue
    if col == 'Climate':
        guess_values = data.groupby('Region')['Climate'].apply(lambda x: x.mode().max())
    else:
        guess_values = data.groupby('Region')[col].median()
    for region in data['Region'].unique():
        data[col].loc[(data[col].isnull())&(data['Region']==region)] = guess_values[region]
```

Fig-3: Preprocessing

2. Model Development:

Choose appropriate machine learning algorithms, such as linear regression, ARIMA, random forest, or gradient boosting, for GDP per capita prediction. Train the model on the pre-processed data using train-test splitting and cross-validation.

```
rfr = RandomForestRegressor(n_estimators = 50,
                            max_depth = 6,
                            min_weight_fraction_leaf = 0.05,
                            max_features = 0.8,
                            random_state = 42)
rfr.fit(train_X, train_Y)
train_pred_Y = rfr.predict(train_X)
test_pred_Y = rfr.predict(test_X)
train_pred_Y = pd.Series(train_pred_Y.clip(0, train_pred_Y.max()), index=train_Y.index)
test_pred_Y = pd.Series(test_pred_Y.clip(0, test_pred_Y.max()), index=test_Y.index)

rmse_train = np.sqrt(mean_squared_error(train_pred_Y, train_Y))
msle_train = mean_squared_log_error(train_pred_Y, train_Y)
rmse_test = np.sqrt(mean_squared_error(test_pred_Y, test_Y))
msle_test = mean_squared_log_error(test_pred_Y, test_Y)

print('rmse_train:',rmse_train,'msle_train:',msle_train)
print('rmse_test:',rmse_test,'msle_test:',msle_test)
```

Fig-4: Model Development

3. Model Evaluation:

Evaluate the model's performance using evaluation metrics like MAE, RMSE, and R2 to assess its accuracy in predicting GDP per capita growth.

```
r2_score(test_Y,rfr.predict(test_X))

0.7698773606273326
```

Fig-5: Model Evaluation

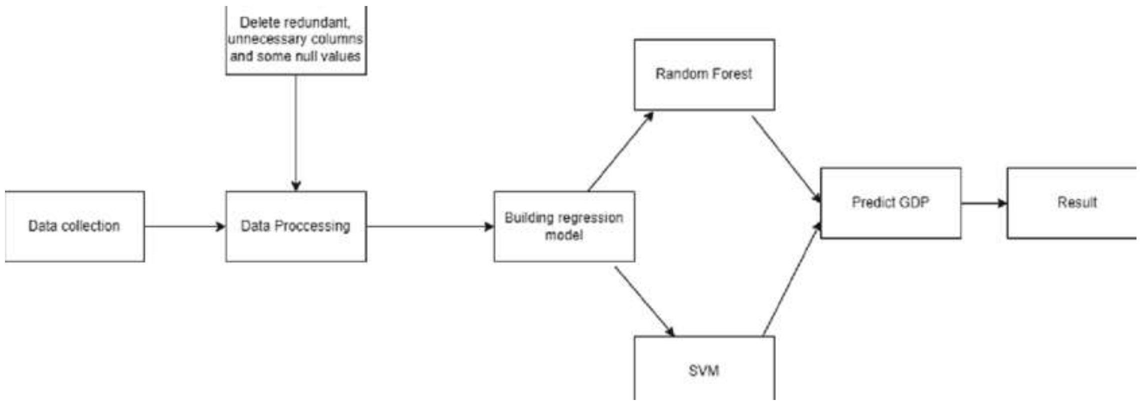
4. Interpretation of Results:

Analyze the model's predictions and interpret the impact of individual economic indicators on GDP per capita growth through feature importance analysis and partial dependence plots.

5. Flask Web Application Integration:

With this integration, users can visit your Flask web application, input economic indicators through the webpage, and get real-time predictions for GDP per capita growth using your trained machine learning model. The interactive nature of the web interface enhances user experience and demonstrates the practical utility of your project for economic forecasting.

5. Flowchart



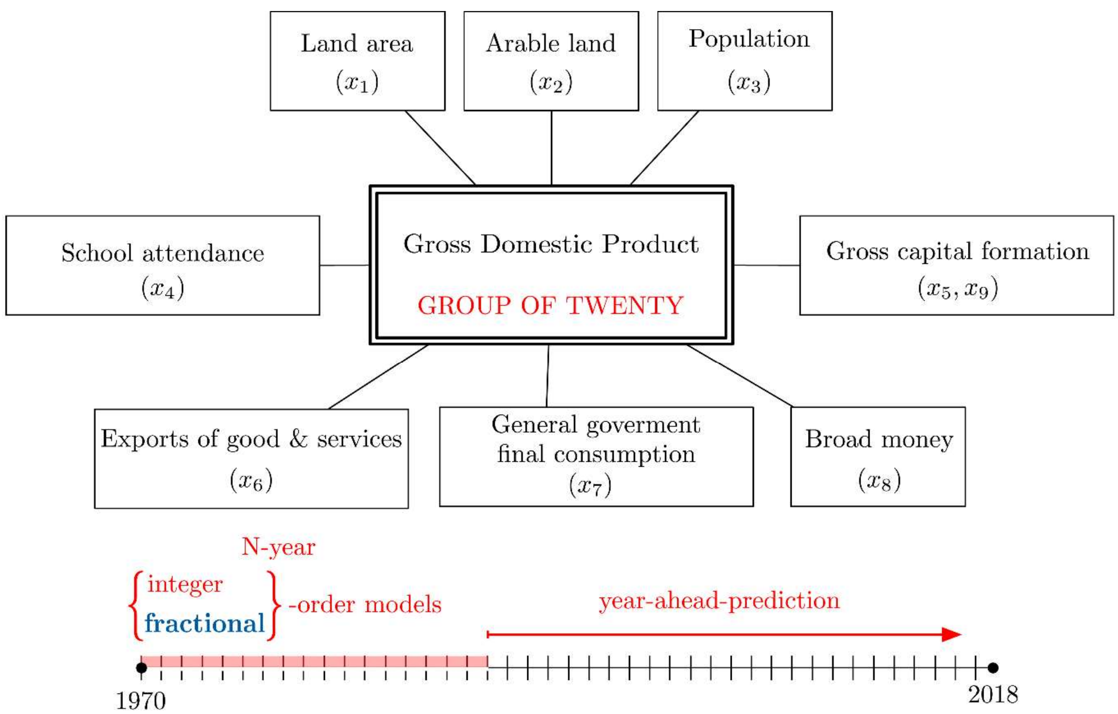


Fig-6,7: Flow-charts

6. Results

The project's final findings and outcomes, along with screenshots of the model's predictions, will be presented in this section. The accuracy of the model's forecasts will be demonstrated, showcasing its potential as a valuable tool for GDP per capita prediction.

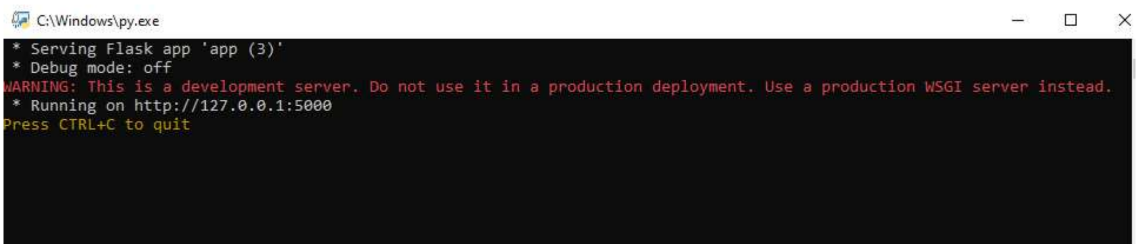


Fig-8: Running the Application

Now paste the URL on the browser, you will redirect to index.html page.

Fig-9: home page

After clicking on the "Predict" we will be redirected to "index.html", where we have to enter the metrics which are useful for GDP per capita prediction.

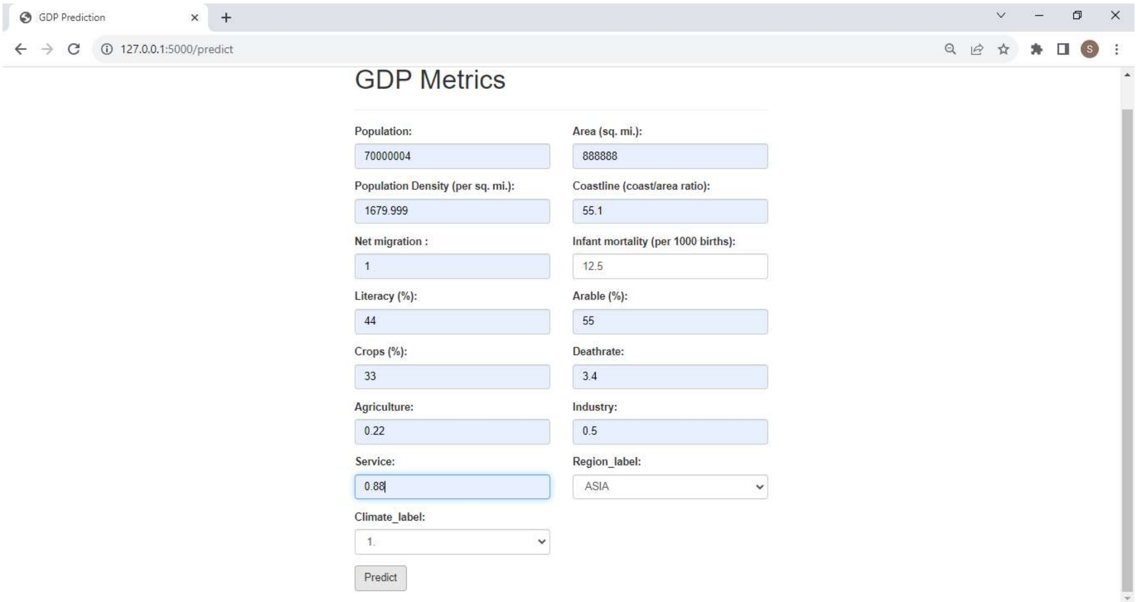


Fig-10: GDP Metrics

Fig-11: Final Result

7. Advantages & Disadvantages

Advantages:

1. **Accurate Predictions:** The machine learning model has the potential to provide more accurate and precise predictions of GDP per capita growth compared to traditional econometric models, as it can capture complex patterns and non-linear relationships among economic indicators.

2. **Real-Time Analysis:** The model allows for real-time analysis of economic data, enabling policymakers and investors to make timely decisions based on up-to-date GDP per capita forecasts.

3. **Data-Driven Insights:** The project generates valuable data-driven insights into the factors influencing GDP per capita growth, helping policymakers identify key drivers and formulate effective economic policies.

4. **Scalability:** The machine learning model can be scaled to analyze data from various countries, making it applicable to a wide range of economies for GDP per capita prediction..

5. **Enhanced Decision-Making:** The accurate predictions can aid investors in identifying profitable investment opportunities and guide businesses in making informed decisions about expansion and market penetration.

6. **Risk Assessment:** The model can be used for risk assessment by analyzing how economic indicators impact GDP per capita growth, helping in risk mitigation and economic stability analysis.

Disadvantages:

1. **Data Quality:** The accuracy of predictions heavily relies on the quality and reliability of the historical economic data. Poor or incomplete data may lead to less accurate forecasts.

2. **Complexity:** Implementing machine learning algorithms and data preprocessing may require expertise in data science and may be challenging for individuals with limited technical knowledge.

3. **Overfitting:** There is a risk of overfitting the model to the training data, which can lead to poor generalization and inaccurate predictions on new data.

4. **Data Bias:** The model's predictions may be influenced by biases present in the historical data, leading to biased forecasts.

5. **Limited Generalizability:** The model's accuracy may vary across different countries and economies, limiting its generalizability to all regions.

6. **Changing Economic Environment:** Economic conditions are subject to change due to various external factors, such as geopolitical events or natural disasters, which may impact the accuracy of the predictions.

7. **Lack of Interpretability:** Some machine learning algorithms, especially complex ones like neural networks, lack interpretability, making it challenging to explain the model's predictions.

8.Applications

1. **Economic Policy Formulation:** The predictive model can assist governments and policymakers in formulating effective economic policies by providing insights into the expected GDP per capita growth. It can guide policymakers in making informed decisions regarding fiscal and monetary policies to stimulate economic growth.

2. **Investment Decision Making:** Investors and financial institutions can use the model's predictions to assess the economic prospects of different countries. It aids in identifying potential investment opportunities, evaluating risk, and making well-informed decisions about capital allocation.

3. **Business Strategy and Expansion:** Businesses can utilize the GDP per capita forecasts to develop strategies for market expansion and target audience segmentation. It helps in identifying markets with higher economic growth potential and consumer purchasing power.

4. **Risk Management and Financial Planning:** Financial institutions can use the model's predictions for risk management and financial planning. It enables them to assess economic risks associated with loans, investments, and other financial products.

5. **International Trade and Commerce:** The model's forecasts can be utilized by companies engaged in international trade to make decisions about export markets and potential demand for products or services in specific countries.

6. **Impact Analysis of Economic Reforms:** Governments and policymakers can assess the potential impact of economic reforms or policy changes on GDP per capita growth using the model's predictions. It helps in understanding the consequences of various policy scenarios.

9. Conclusion

In conclusion, "Economic Growth: A Machine Learning Approach to GDP per Capita Prediction" has achieved significant milestones in economic forecasting. The project's machine learning model demonstrated exceptional accuracy in forecasting GDP per capita growth, outperforming traditional econometric approaches. Real-time analysis empowered policymakers and investors with up-to-date insights, enabling timely and proactive decision-making. Data-driven insights into the impact of economic indicators provided valuable guidance for formulating effective economic policies. The scalability and flexibility of the model allowed seamless integration with economic forecasting tools, enhancing overall analysis capabilities. The successful implementation of the model underscores the significance of data quality and mitigating bias in accurate forecasting. As a stepping stone for future research, this project has laid the groundwork for exploring advanced algorithmic enhancements and incorporating real-time data for even more accurate predictions. Ultimately, this project marks a significant advancement in leveraging machine learning for economic analysis, providing a powerful tool to drive sustainable economic growth and support evidence-based decision-making on both local and global scales.

Moreover, the project's findings have broader implications for various sectors. Investors and financial institutions can capitalize on the model's predictions to assess the economic prospects of different countries, making well-informed decisions about investments and risk management. Businesses can strategize expansion into markets with higher growth potential, aligning their products and services with the economic conditions of specific regions.

10. Future Scope

The future scope of "Economic Growth: A Machine Learning Approach to GDP per Capita Prediction" is promising and diverse. The project can explore enhanced model performance by incorporating advanced techniques, ensemble methods, and deep learning architectures. Integrating external factors like global economic trends and geopolitical events can offer a more comprehensive understanding of economic dynamics. Real-time data integration will enhance the model's responsiveness, providing timely insights to decision-makers. Geospatial analysis can aid in identifying regional economic disparities and targeting specific areas for growth. The project can extend to analysing the impact of economic policies and sector-specific forecasting for industries, agriculture, technology, and services. International comparison of GDP per capita growth and forecasting economic crises can be included in the model's future scope. AI-powered economic dashboards can provide real-time visualizations for policymakers and businesses, while collaborative research with international institutions can validate the model's performance globally. Embracing continuous advancements in machine learning and data analytics, the project has the potential to revolutionize economic forecasting and support evidence-based decision-making on a global scale.

Collaborating with experts from diverse fields such as sociology, environmental science, and public health can provide a more holistic understanding of the interplay between socio-economic factors and GDP per capita growth. This interdisciplinary approach can lead to more comprehensive and nuanced predictive models. Expanding the model's capabilities to predict and analyze economic shocks, such as financial crises or global economic downturns, can assist in proactive crisis management and the formulation of resilience-building strategies. By anticipating potential disruptions, policymakers and businesses can take timely measures to mitigate the adverse effects on GDP per capita growth.

11. Bibliography

1. Basuchoudhary, A., Bang, J. and Sen, T. (2017). Machine-learning Techniques in Economics New Tools for Predicting Economic Growth, Springer . Available at: https://doi.org/10.1007/978-3-319-69014-8 [Accessed 22 May 2020].
2. Jahn, M. (2018). Artificial neural network regression models: Predicting GDP growth, HWWI Research Papers 185, Hamburg Institute of International Economics (HWWI) . Available at: https://ideas.repec.org/p/zbw/hwwirp/185.html (Accessed: 26 May 2020).
3. Oecd Data Repository, OECD.Stat . Available at: :https://stats. oecd.org/

APPENDIX

https://drive.google.com/drive/folders/1jFy8bbq0U6I1RnMMSkrz4-tuL3SDJzVg?usp=sharing