Capstone Project

Predicting GDP Using Socioeconomic Indicators

PRESENTED BY SEHBA KHAN

STUDENT NAME: SEHBA KHAN

COLLEGE NAME: TAGORE MAHAVIDYALAYA

DEPARTMENT: EDUNET FOUNDATION

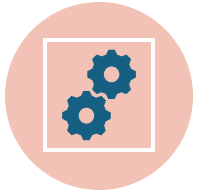
EMAIL ID: SEHBAKHAN43@GMAIL.COM

AICTE STUDENT ID: AINSI_108458

OUTLINE

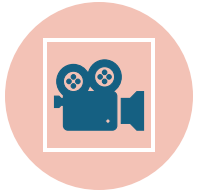
Problem Statement

Proposed Solution



System Development Approach

Algorithm & Deployment



Result

Conclusion

Future Scope

References

PROBLEM STATEMENT

🧩 Problem Statement

- Many countries struggle with assessing their future GDP due to complex dependencies on social, environmental, and economic factors.
- Traditional forecasting lacks real-time insights and fails to leverage multi-factor datasets.
- There is a need for a data-driven model that can understand and quantify the impact of key indicators on GDP.

PROPOSED SOLUTION

The proposed system aims to address the challenge of accurately predicting a country's GDP based on various socioeconomic, environmental, and demographic indicators. This involves leveraging historical data and applying machine learning techniques to model complex relationships among features. The solution is structured into the following components:

1. Data Collection

- Gather historical data from reliable sources like the World Bank, including:
- Use multi-year datasets to capture trends and variations across time and geography.

2. Data Preprocessing

- Encoding categorical variables (e.g., Country).
- Standardizing/normalizing features using scaling techniques.
- Perform **feature engineering** to derive meaningful insights (e.g., GDP per capita).

3. Machine Learning Modeling

- Implement and compare regression algorithms including:
  - Linear Regression, Random Forest Regressor, Gradient Boosting Regressor
- Tune hyperparameters using cross-validation to improve accuracy and generalization.
- Evaluate models based on:
  - Mean Absolute Error (MAE)
  - Root Mean Squared Error (RMSE)
  - R^2 Score

4. Deployment

- Build an interactive web application using **Streamlit** to:
  - Accept user input (country code, population, life expectancy, etc.).
  - Predict GDP in real-time using the trained model.
- Use **Pickle** to serialize and load the trained model and scaler.

5. Evaluation

- Evaluate the best model (Gradient Boosting) achieving:
  - **R^2 Score:** 0.878
  - **MAE:** 0.256
  - **RMSE:** ~0.84
- Visualize predictions vs actuals for validation.

SYSTEM APPROACH

⚙️ System Development Approach

- **Technology Stack:**
  - Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)
  - Streamlit for deployment
  - Pickle for model serialization
- **Steps:**
  - Data cleaning and preprocessing (missing values, encoding, scaling)
  - Exploratory Data Analysis (EDA)
  - Model training, tuning, and evaluation
  - Deployment using a web interface

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  Gradient Boosting Regressor was selected for its ability to capture nonlinear relationships and outperform simpler models in regression tasks.
- **Data Input:**
  Model inputs include Country, Year, Population, Life Expectancy, Unemployment Rate, CO_2 Emissions, and Access to Electricity.
- **Training Process:**
  Data was scaled and split into training/testing sets. Hyperparameters were tuned using cross-validation to minimize MAE and RMSE, and improve R^2.
- **Prediction Process:**
  User inputs are preprocessed and passed through the trained model to predict GDP. The app delivers real-time results based on socioeconomic indicators.
- **Deployment:**
  The model was deployed using Streamlit, allowing users to input data and receive GDP predictions through an interactive web interface.

RESULT

📊 The result of Model:

- **MAE:** 0.256
- **RMSE:** ~0.84
- **R^2 Score:** 0.878
- Best performance from **Gradient Boosting** after tuning
- Developed an efficient GDP forecasting system that outperformed baseline models.
- Enabled interpretation of feature importance, identifying key socioeconomic drivers.
- Delivered an end-to-end solution accessible through a user-friendly web interface.

CONCLUSION

📜 **The Conclusion of Model is:**

- The model successfully predicts GDP with high accuracy using socioeconomic data.
- Gradient Boosting proved most reliable in capturing complex nonlinear relationships.
- Streamlit deployment provides an accessible interface for predictions.

FUTURE SCOPE

- Expand model to include real-time data sources (e.g., live economic feeds).
- Apply model to regional/state-level data for localized GDP predictions.
- Integrate external economic shocks (wars, pandemics) for robust forecasting.

REFERENCES

📚 **References**

GDP Prediction Model:

GitHub Link: GDP-Prediction-Model

Thank you

sehbakhan43@gmail.com
github.com/sehbakhan