`

University of Birmingham

College of Engineering and Physical Sciences

School of Computer Science

Project Report on

Predicting Future GDP Growth by Analyzing Key Economic Indicators

Submitted by

Sara Kale, Suyash Yadav, Suchit Pathak, Shalin Sam

MSc Data Science group project

Under the Guidance of

Qamar Natsheh, Jianqiao Mao, Huiping Chen

Acknowledgment

We extend our heartfelt gratitude to all those who contributed to the completion of this research project on "Predicting Future GDP Growth by Analyzing Key Economic Indicators."

First and foremost, we would like to express our sincere appreciation to our Module Lead, Qamar Natsheh, for their invaluable guidance and support throughout this endeavor. Their expertise and encouragement were instrumental in shaping our research approach and refining our analysis.

We are deeply indebted to our main mentor and the backbone of this project, Jianqiao Mao. Their unwavering dedication, insightful feedback, and tireless efforts significantly enriched our research journey and paved the way for its successful completion.

We would also like to extend our gratitude to Huiping Chen for their assistance and valuable input, which contributed to the enhancement of our project.

Furthermore, we wish to acknowledge the support and understanding of our families and loved ones, whose encouragement and patience sustained us during the demanding phases of this research endeavor.

Lastly, we express our gratitude to all others who provided assistance, feedback, and encouragement along the way. Each contribution, whether significant or modest, has been pivotal in bringing this project to fruition.

TABLE OF CONTENTS

TABLE OF FIGURES

ABSTRACT
In urban profiling areas, GDP is a crucial indicator of the health of the national and global economy. It is common practice to use the real GDP growth rate as a gauge of economic health. This paper presents a comprehensive approach for forecasting GDP Growth by analyzing the key economic factors across four major economies: The United Kingdom, The United States, Japan, and China. With data spanning from 1994 to 2022, the study aims to understand the complex relationships between economic indicators and GDP fluctuations. By employing interpretable machine learning methods, namely ARIMA and LSTM, the project seeks to provide transparent and understandable models, facilitating informed decision-making for various policymakers, economists, and businesses. By examining historical data and trends, the project aims to find potential shifts in these relationships and their implications on future economic forecasts. Additionally, the project outcomes involve determining the economic indicators that have a notable influence on GDP growth in both the long term and the short term. Root Mean Squared Error (RMSE) is used as an evaluation metric for demonstrating the performance of the proposed model. The results of this paper indicate that Government Expenditure is considered as the most significant factor for forecasting the GDP across the studied economies. The experimental results demonstrated that the ARIMA model outperforms the LSTM Model by achieving the highest accuracy of 91.878% for UK. This work highlights the effectiveness of our machine learning models in improving the precision of macroeconomic forecasts and recommends their wider application in economic research.

KEYWORDS
Gross Domestic Product, Machine Learning, Economic Performance, Decision Making, Prediction Model

1. INTRODUCTION

In the contemporary realm of global economics, accurately forecasting Gross Domestic Product (GDP) growth is a paramount objective essential for crafting effective economic policies and strategies. The intricate interaction among diverse economic indicators and GDP trends profoundly shapes the trajectory of national economic prosperity, exerting significant influence on individuals' livelihoods and businesses' operations worldwide. Consequently, comprehending the dynamics of this relationship has become a fundamental pursuit for policymakers, economists, and stakeholders across various sectors. As official data are frequently unavailable for at least a quarter after being released, predicting real GDP growth is a complex process.

Gross Domestic Product (GDP) stands as a pivotal measure reflecting the economic health and vitality of a nation. Serving as a barometer of economic performance, GDP holds immense significance for policymakers, investors, and citizens alike. It offers insights into the overall growth trajectory, employment trends, and standard of living within a nation. Governments utilize GDP data to formulate fiscal and monetary policies, aiming to stimulate growth, curb inflation, and address unemployment. Investors rely on GDP figures to assess market potential and allocate resources efficiently.

One of the most significant measures of macro-economy, Gross Domestic Product (GDP) is crucial for gauging a country's economic progress. Future macro-economic goals and economic regulatory strategies will be heavily impacted by this indicator (Götz & Knetsch, 2019). As demonstrated by (Plakandaras et al., 2015) in the instance of predicting US home values, various machine learning techniques actually frequently outperform conventional econometric models. In the past (Biau and D'Elia, 2010) employed a Random Forest Model to estimate the euro area's GDP data and discovered that the machine learning model could generate predictions that were more accurate than those made by a conventional autoregressive model. (Emsia and Coskuner, 2016) forecasted Turkey's GDP growth using support vector regression. (Shaobo, 2021) recently coupled the ARIMA Model with a back propagation (BP) Neural Network to estimate the GDP's non-linear residual. In order to predict Nigeria GDP, (Oden et al., 2020) and (Yua et al., 2020) used Ordinary Least Squares (OLS) on a dataset of variable indicators. Using a reduced Vector Autoregressive (VAR) approach, (Patrick & Sebastian, 2009) forecasted the GDP growth of Baltic States, Estonia, Latvia, and Lithuania.

This study endeavors to undertake a comprehensive exploration of the intricate nexus between key economic indicators and GDP across four prominent economies: the United Kingdom, the United States, Japan, and China. Covering a substantial time span from 1994 to 2022, this research aims to uncover the subtle variations in how economic indicators impact GDP within the distinctive socio-economic contexts of each country. The significance of this research extends beyond its potential to elucidate the drivers of GDP growth. It also lies in its ability to illuminate the evolving dynamics of economic performance amidst the backdrop of globalization, technological advancements, and shifting geopolitical landscapes. By examining a diverse array of economic indicators drawn from reputable sources such as The World Bank and OECD websites, this study seeks to offer a comprehensive perspective on the factors underpinning GDP fluctuations. Thus, it provides a sturdy basis for evidence-based policy formulation and strategic planning. Furthermore, the selection of the United Kingdom, the United States, Japan, and China as focal points for analysis underscores the strategic importance of these economies on the global stage. With their varied economic structures, differing levels of development, and significant roles in shaping international trade and finance, these countries present compelling case studies for exploring the differential impacts of economic indicators on GDP growth. By delving into these intricacies, this research aims to equip stakeholders with actionable insights to navigate economic challenges adaptively and seize emerging opportunities in an increasingly interconnected world. In essence, this study endeavors to contribute to the ongoing discourse surrounding economic development and resilience by unpacking the intricate relationships between economic indicators and GDP growth across diverse national contexts. By elucidating these dynamics, the research strives to empower stakeholders with the knowledge and tools necessary to navigate the complexities of the global economy, foster inclusive growth, and advance the collective pursuit of sustainable prosperity.

1.1. BACKGROUND RESEARCH

GDP encapsulates the total market value of all goods and services produced within a country's borders over a specific period, typically annually or quarterly. Beyond economics, GDP influences social welfare programs, infrastructure development, and resource allocation, shaping the fabric of society. However, it's crucial to recognize that GDP alone may not capture the complete picture of a nation's well-being, as it overlooks factors like income distribution, environmental sustainability, and quality of life. Nonetheless, as a widely acknowledged metric, GDP remains indispensable in understanding and navigating the complexities of modern economies.

Economic indicators play a vital role in assessing the health and performance of an economy. The Consumer Price Index (CPI) measures the average change in prices paid by consumers over time, providing crucial insights into inflationary pressures within an economy. Exchange rates determine the relative price of one currency against another, impacting international trade, investment, and economic competitiveness. Fluctuations in exchange rates can affect export competitiveness, import costs, and ultimately, inflation and GDP growth. Exports of merchandise represent tangible goods produced domestically and sold internationally, contributing significantly to economic output and trade balances. Imported merchandise, on the other hand, encompasses goods brought into a country for consumption or production, influencing domestic supply chains and consumer choices while affecting trade deficits and currency demand. Industrial production drives economic growth, job creation, and technological advancement, reflecting manufacturing, mining, and utility activities. Inflation, the general increase in prices of goods and services over time, impacts purchasing power and economic stability, affecting interest rates, investment decisions, and consumer and business confidence. Tourism plays a crucial role in boosting economic activity, job creation, and infrastructure development, contributing to GDP growth, regional development, and cultural exchange. The unemployment rate measures the proportion of the labor force actively seeking employment but without a job, impacting consumer spending, government budgets, and overall economic productivity. Foreign Direct Investment (FDI) represents investments made by non-residents into the economy, fostering growth, technology transfer, and market integration, influencing employment, productivity, and export competitiveness. Life expectancy reflects overall health, quality of life, and demographic trends, impacting healthcare expenditures, retirement planning, and social welfare policies. Government expenditure constitutes spending on goods, services, and investments to support economic activity and public welfare, influencing aggregate demand, social programs, and fiscal sustainability.

1.2. MOTIVATION

In embarking on this research endeavor, our motivation is deeply rooted in the imperative to decipher the intricate relationship between economic indicators and Gross Domestic Product (GDP) dynamics. Amidst the complexity of today's global economy, the ability to accurately forecast GDP growth holds immense significance for policymakers, economists, and analysts alike. Recognizing the challenges posed by traditional forecasting methods and the pressing need for advanced modeling techniques, our group is driven by a collective commitment to shed light on the multifaceted interplay of economic variables. Through meticulous analysis and innovative methodologies, we aim to offer actionable insights that can inform strategic decision-making and policy formulation. Ultimately, our research endeavors to empower stakeholders with the knowledge and tools needed to navigate the complexities of the economic landscape with precision and foresight.

1.3. COMMON MODELLING APPROACHES FOR GDP FORECASTING

a. Linear Regression

Linear Regression is a fundamental statistical technique used to model the relationship between a dependent variable (GDP) and one or more independent variables (economic indicators). The model assumes a linear relationship between the predictors and the target variable, expressed by the equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon$$

Here, Y represents the GDP, X_1, X_2,…Xn is the independent variables, β_0, β_1,… βn are the coefficients to be estimated, and ε is the error term. The coefficients are determined through the method of least squares, which minimizes the sum of squared residuals.

This technique offers several advantages, including simplicity, interpretability, and ease of implementation. It provides insights into the relationship between the predictors and the target variable, allowing analysts to identify potentially significant factors influencing GDP growth. Additionally, linear regression can handle both numerical and categorical predictors, making it versatile for various types of economic data.

However, linear regression has limitations, particularly in capturing nonlinear relationships and complex interactions among variables. It assumes a linear relationship between predictors and the target variable, which may not hold in real-world scenarios with intricate economic dynamics. Moreover, linear regression is sensitive to outliers and multi-collinearity, which can affect the model's performance and interpretability.

b) Ensemble Methods: Random Forest

Random Forest is a powerful ensemble learning technique that combines multiple decision trees to improve predictive accuracy and robustness. The basic idea behind Random Forest is to train a large number of decision trees on different subsets of the training data and aggregate their predictions to obtain the final output.

Each decision tree in the Random Forest is trained independently on a random sample of the training data, with a random subset of features considered at each split. This randomness helps to decorate the trees and reduce overfitting, leading to more accurate and generalizable predictions.

During Prediction, the output of Random Forest is obtained by averaging or voting across all individual trees' predictions. This aggregation process helps to mitigate the bias and variance associated with individual trees, resulting in a more stable and reliable model.

Random Forest offers several advantages such as:

Robustness to overfitting: By aggregating predictions from multiple trees, RF reduces the risk of overfitting and improves generalization performance.

Feature Importance: Random Forest provides a measure of feature importance, allowing analysts to identify the most influential predictors for GDP forecasting.

However, even this method has certain limitations:

Computational Complexity: Training and evaluating a large number of decision trees can be computationally expensive, especially for large datasets.

Lack of Interpretability: While random forests can identify important features, the underlying decision-making process of individual trees is less interpretable compared to linear models.

c) Ensemble Methods: XGBoost

XGBoost is another ensemble learning algorithm that has gained popularity for its superior performance in various machine learning tasks, including regression and classification. Like the random forest, XGBoost is based on the principle of ensemble learning but employs a different approach known as the gradient approach.

In XGBoost, decision trees are built sequentially, with each new tree trained to correct the errors made by the previous ones. Unlike Random Forest, which builds trees independently, XGBoost learns from the mistakes of previous trees and focuses on the instances that are difficult to predict. This iterative process allows XGBoost to improve its predictive performance and capture complex patterns. One key feature of XGBoost is its regularization techniques, which help to prevent overfitting and enhance generalization performance. These techniques include learning rate, and column subsampling, which controls the complexity of the model and reduces the risk of overfitting.

However, it also has some limitations,
Hyperparameter Tuning: XGBoost requires careful tuning to achieve optimal performance, which can be time-consuming and computationally expensive.

Black Box Nature: While XGBoost delivers excellent predictive accuracy, the internal workings of the model are less interpretable compared to simpler techniques like linear regression.

2. RELATED WORKS

Forecasting GDP is a critical aspect of macroeconomic analysis and policymaking, enabling businesses and researchers to anticipate economic trends and plan accordingly. In recent years, machine learning algorithms and traditional econometric models have emerged as powerful tools for GDP forecasting. Understanding the strengths and limitations of different forecasting approaches is crucial for developing robust models and improving economic decision-making.

The paper (Yoon, J. 2020) presents a methodological approach using machine learning models, specifically gradient boosting, and random forest, to forecast real GDP growth in Japan. The study's primary objective is to improve upon benchmark forecasts provided by the IMF and Bank of Japan. By employing cross-validation techniques to optimize hyperparameters, the authors demonstrate that their models outperform the benchmark forecasts in terms of accuracy. One of the strengths of this paper lies in its rigorous methodology, which includes a thorough evaluation of model performance using mean absolute percentage error and root mean squared error. Additionally, the study's focus on real GDP growth in Japan adds relevance and specificity to its findings, allowing for insights that are tailored to the Japanese economic context. However, while the study demonstrates improved forecast accuracy, it also has certain limitations. For instance, the use of machine learning models may prioritize predictive accuracy over interpretability, potentially limiting the insights gained from the forecasting process. Additionally, the study's reliance on historical data may pose challenges in predicting economic trends influenced by unprecedented events, such as global pandemics or financial crises.

Another approach by (Saadah, S. and Wibowo et al., 2020) explores the prediction of GDP in Indonesia using deep learning algorithms, specifically LSTM and RNN architectures. The study focuses on understanding the stability of Indonesia's financial condition by forecasting GDP growth, particularly during the COVID-19 pandemic. One notable strength of this paper is the use of deep learning architectures which allows for the modelling of temporal dependencies and non-linear relationships, which may be challenging to capture using traditional econometric methods. Despite its contributions, this paper exhibits certain limitations. The reliance on deep learning algorithms may introduce challenges related to model interpretability. Deep learning models often operate as 'Black boxes', making it difficult to understand the underlying factors driving predictions. The paper by (Krishna M et al., 2020) focuses on predicting Indian GDP using ML algorithms such as Linear regression and Polynomial Regression techniques. While these models are valuable tools for capturing relationships between independent and dependent variables, they may struggle to account for the complex dynamics and non-linear patterns present in the data. As a result, accuracy may be limited, particularly during periods of economic volatility or structural change.

Our approach addresses limitations observed in existing techniques by adopting a methodology that considers both linear and non-linear relationships, and diverse predictors and incorporates PFFRA and feature importance analysis, which enables us to identify the most influential predictors and enhance model interpretability.

3. METHODOLOGY

3.1 DATA

3.1.1 RETRIEVING THE DATA

The initial phase of our methodology involved gathering economic data from reputable sources such as the Organization for Economic Co-operation and Development (OECD) and The World Bank. These organizations provide comprehensive datasets covering a wide range of economic indicators across various countries, making them ideal sources for our research. The datasets were accessed through their respective websites, where we obtained both Excel and CSV files containing the required information. Each dataset comprised multiple columns representing different economic indicators and rows corresponding to periods. The datasets were organized in a tabular format, facilitating easy access and manipulation for further analysis.

Our data retrieval process was twofold, focusing on both the selection of relevant countries and the collection of key economic indicators. We identified a diverse set of countries such as the USA, UK, China, and Japan representing different regions and levels of economic development to ensure the generalizability and applicability of our findings. These countries were selected based on criteria such as economic significance, geographical representation, and data availability. Simultaneously, we curated a list of key economic indicators that are widely recognized as influential in assessing a country's economic health and predicting GDP growth. These indicators encompassed various aspects of the economy, influencing inflation, trade, market conditions, Government spending, and Health care, among others. By selecting a diverse set of indicators, we aim to capture the multifaceted nature of economic activity and its drivers across different countries.

3.1.2 DATA PREPARATION

Upon retrieving the data, our next step was to prepare it for analysis. This involved several preprocessing tasks aimed at ensuring data quality and integrity. One crucial aspect was handling missing values, which we addressed by identifying and assessing the extent of missingness in the dataset. Based on the proportion of missing values and their impact on the analysis, we decided on an appropriate strategy for the imputation of missing data to prevent bias in our results. Linear Interpolation is a straightforward and widely used technique for estimating missing values based on the observed values before and after the missing points. The method assumes a linear relationship between consecutive data points and fills in the missing values with estimates derived from this linear interpolation. This approach is particularly useful for time-series data, where the temporal continuity of observations makes linear interpolation a suitable and intuitive imputation method. By leveraging this technique as part of our data preparation pipeline, we ensured the completeness and reliability of our dataset, laying the groundwork for robust exploratory analyses.

Additionally, outlier detection was performed to identify any extreme values in key economic indicators. In our case, dealing with real-world economic data, we recognized that outliers may represent genuine, albeit unusual, phenomena within the economic landscape. Therefore, instead of outright removal, we retained them as we aimed to capture the full spectrum of economic behavior. One of the primary challenges in working with multiple datasets sourced from different sources is the lack of uniformity in data formats and structures. To address this issue, we meticulously examined the data schemas, variable names, and data types across all datasets. We then applied standardized conventions such as renaming variables and aligning data types to ensure consistency and compatibility across the board.

3.2 RATIONALE TOWARDS EXPLORING THE DATA

3.2.1 EXPLORATORY DATA ANALYSIS (EDA)

EDA plays a crucial role in understanding the characteristics of our dataset and identifying patterns or relationships among variables. To begin, we computed basic descriptive statistics for each economic indicator and GDP, including measures of central tendency. This provided us with valuable insights into the data's distributional properties and variability. Furthermore, data visualization techniques such as Time-Series plots and heatmaps were employed to explore the relationships between features and GDP. This helps to uncover patterns that may not be apparent from numerical summaries alone. Given the temporal nature of our data, we conducted a comprehensive analysis of temporal patterns over time. Time Series plots and line graphs allow us to visualize the evolution of key variables across different periods, revealing trends and cyclic patterns inherent in the data. This helps us to discern long-term trends, identify abrupt changes, and assess the impact of external factors on economic indicators.

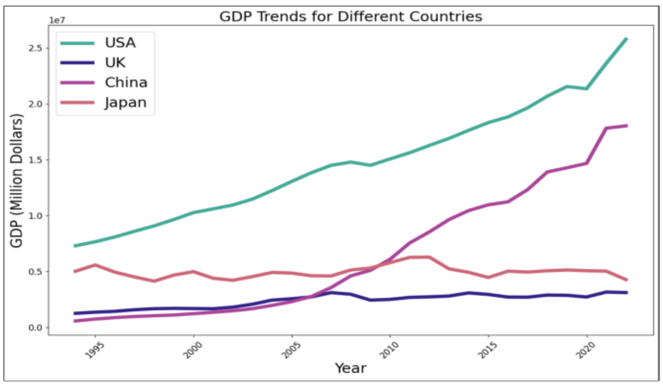

Fig 1. GDP Trend across Various Economies

3.2.2 FEATURE ENGINEERING

This process involves selecting and transforming features from the existing dataset to improve model performance. In our study, feature selection was guided by various considerations, including correlation analysis, domain knowledge, data availability and assessment of multicollinearity using techniques such as implementing heatmaps, VIF Analysis. Correlation analysis helped us to identify features with significant correlations with GDP, enabling us to prioritize relevant features for further analysis. Features such as Core CPI, Exchange Rate, Inflation and Environmental Factors were deemed unsuitable for inclusion due to data inconsistency, and high multicollinearity with other predictors. An important technique for assessing multicollinearity among predictor variables is Variance Inflation Factor (VIF) analysis. Multi-collinearity arises when predictor variables in a regression model are highly correlated, leading to inflated standard errors of regression coefficients and reduced reliability of parameter estimates. By examining the VIF Values of predictors, we can determine the degree of independence among variables and prioritize those with lower VIF values for inclusion in the model.

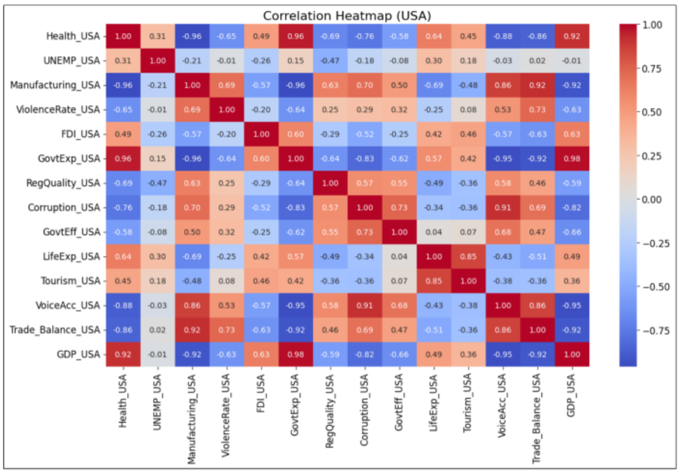

Fig 2. Correlation Heatmap

4. DATA MODELLING

4.1 LSTM MODEL

4.1.1 THEORY

Long Short-Term Memory (LSTM) networks are a type of recurrent Neural Network (RNN) architecture, designed to address the vanishing gradient problem in traditional RNNs. They are particularly effective in capturing long-term dependencies and sequential patterns in time series data, making them well-suited for tasks such as Time Series forecasting, natural language processing, and speech recognition.

General Architecture

The LSTM Architecture consists of multiple memory cells and gating mechanisms that regulate the flow of information through the network. Key components include:

Memory cells: LSTM networks contain memory cells that maintain a memory state over time, allowing them to remember information from earlier steps. These cells are equipped with self-loop connections, enabling them to retain information for long durations.

Forget Gate: The forget gate determines which new information from the previous time step to discard or forget. It takes as input the previous hidden state and current input, and outputs a forget gate activation vector that modulates the memory cell's content.

Input Gate: This gate decides which new information to store in the memory cell. It consists of a sigmoid activation function and a tanh activation function, which controls the flow of input information into the memory cell.

Output Gate: The Output Gate controls the flow of information from the memory cell to the output of the LSTM unit. It selects relevant information from the memory cell to pass on to the subsequent layers or time steps.

In the context of GDP Prediction, LSTM networks excel at capturing the complex temporal dependencies and nonlinear relationships present in economic indicators over time. By learning from historical data patterns, LSTM Models can effectively forecast future GDP values based on sequential patterns observed in the input features. The ability of LSTM Networks to retain long-term memory enables them to capture subtle trends and seasonality in the data, making them a powerful tool for time series forecasting tasks like GDP prediction.

4.1.2 IMPLEMENTATION

In our LSTM implementation for GDP Prediction, we utilize the Keras library to construct a deep learning model consisting of sequential layers of LSTM units. Each layer processes input sequences over time, allowing the model to capture long-term dependencies in the data. The architecture includes multiple LSTM layers stacked on top of each other, with each layer capable of learning complex patterns from the input data. To prepare the data for training, we first convert the DataFrame containing economic indicators and GDP Values into NumPy arrays. These arrays serve as input and target data for the LSTM model, respectively. We then apply min-max scaling to normalize the input features to a similar range, ensuring that all features contribute equally to the model's learning process. This normalization step is crucial for preventing any single feature from dominating the training process. Next, we define the LSTM Model architecture using the sequential API provided by Keras.

Algorithm:

Iterating over Years: The implementation begins by iterating over the years starting from the year 2000 up to the year 2019. This iterative process allows for the sequential training of the LSTM Model on the Historical Data. The window size parameter is set to 20. This window size ensures that the model has sufficient historical data to learn meaningful patterns and relationships.

Model Training: The LSTM Model is trained exclusively on the training set, using historical data of all economic indicators from the designated training years. During this training phase, the model learns temporal patterns inherent in the GDP time series data, enabling it to make accurate predictions based on past observations.

Forecasting GDP for Test Years: Following training, the LSTM Model is deployed to forecast GDP values for the test years (2020 to 2022). For each test year, the LSTM model utilizes the corresponding economic indicators to generate a forecasted GDP value. These forecasted values represent the model's predictions for future GDP trends based on the learned patterns and relationships extracted during training.

Hyperparameter	Value
Number of Layers	1
Number of Neurons	100
Fully Connected Layers	Yes
Batch Size	64
Learning Rate	Default (0.001)
Number of Epochs	300
Optimizer	Adam
Time Steps	20
Activation Function	Linear (default)

Fig 3. Hyperparameters in LSTM

After defining the model architecture, we compile the model using the Adam Optimizer, a variant of stochastic gradient descent and specify the loss function as mean squared error (MSE). The optimizer adjusts the model's weights and biases iteratively during training to minimize the loss function, optimizing the model's performance.

During training, we feed batches of input sequences and corresponding target GDP values to the model, iteratively updating its parameters to minimize the loss. We train the model for a fixed number of epochs, monitoring its performance on a separate validation set to prevent overfitting. By adjusting the hyperparameters such as batch size, number of epochs, and learning rate, we can fine-tune the model's training process to achieve optimal performance.

4.2 ARIMA MODEL

4.2.1 THEORY

The Autoregressive Integrated Moving Average (ARIMA) model is a widely used time series forecasting method that combines autoregression, differencing, moving average components to capture the temporal dependencies and patterns present in the time series data. ARIMA models are particularly effective for modelling and predicting time series data with stationary or near-stationary properties.

Components:

1. Autoregressive (AR) Component

The autoregressive component models the relationship between an observation and several lagged observations (i.e., its past values). It captures the linear relationship between the current observation and its previous observations.

Mathematically, an AR(p) process of order p is represented as:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

Where:

- $\mathbf{Y_t}$ is value of the time series at time t,
- **c** is a constant
- $\phi_1, \phi_2, \dots, \phi_p$ are the autoregressive parameters
- ϵ_t is the white noise.

2. Integrated (I) Component

The Integrated component accounts for differencing, which helps to transform the original non-stationary time series into a stationary series. Differencing involves computing the difference between consecutive observations to remove trends or seasonality. The order of differencing, denoted by d, represents the number of times differencing is applied to achieve stationarity.

Reason for Differencing:

- **Stationarity:** Many time series models, including ARIMA, assume stationarity, where the statistical properties of the data remain constant over time. Stationarity simplifies the modelling process and allows for more reliable forecasts.
- **Trend Removal:** Time series data often exhibits long-term trends, which can obscure the underlying patterns. Differencing removes the trend component by computing the differences between consecutive observations, thereby making the data stationary.
- **Stabilization of Variance:** Differencing can also help stabilize the variance of the time series data, making it easier to model.

3. Moving Average (MA) Component

The moving average component models the relationship between an observation and a linear combination of past residuals. It captures the impact of past shocks or disturbances on the current observation.

Mathematically, an MA(q) process of order q is represented as:

$$Y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Where:

- $\mathbf{Y_t}$ is value of the time series at time t,
- **c** is a constant
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive parameters
- $\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q}$ are the error terms at different lags

4.2.2 IMPLEMENTATION

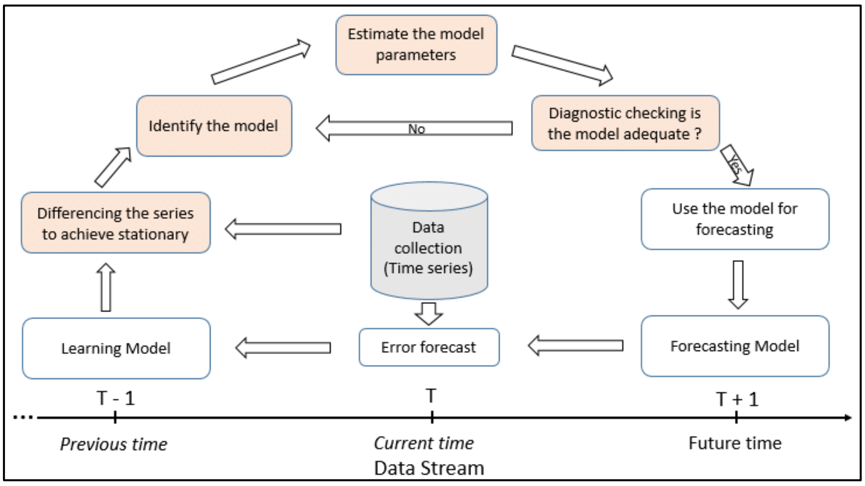

Fig 4. ARIMA Model Training Process

In our project, the ARIMA Model is employed to forecast future GDP growth by leveraging key economic indicators.

Model Training

The Autoregressive parameter (p) captures the linear relationship between an observation and its lagged values. In our project, we selected a value of 20 for p based on the requirement to consider the past 20 years of historical data for accurate forecasting. The differencing parameter is set to 1 to perform first-order differencing. The moving average parameter is set to 0, indicating that there is no moving average component in the model.

Model Fitting

Once the parameters are selected, the model is fitted to the training data using the selected parameters. This process aims to minimize the sum of squared errors between the observed and the predicted values.

Model Testing

Using the trained ARIMA model, forecasts for GDP growth are generated for each year in the test period (2020-2022). During the testing phase, the forecasted GDP values are compared to the actual GDP values for the corresponding years.

4.3 PERFORMANCE METRICS

In our project, we utilized two main performance metrics to evaluate the accuracy of our predictive models:

Root Mean Squared Error (RMSE)

RMSE is a widely adopted metric for assessing the predictive accuracy of regression models, including time series forecasting models such as LSTM and ARIMA. It quantifies the average magnitude of the errors between the predicted values and the actual values.

Mathematically, RMSE is calculated as the square root of the average of the squared differences between the predicted and actual values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Here n represents the number of observations, y_i represents the actual GDP values, $\hat{y}i$ the predicted GDP values. A lower RMSE indicates better model performance, suggesting smaller errors between the predicted and actual values.

Accuracy

Accuracy is a key metric in time series forecasting, often calculated as a percentage based on the RMSE. In our project, the accuracy metric measures the percentage of improvement achieved by the model over the mean actual GDP values. It is computed using the formula:

$$Accuracy = 100 \times \left(1 - \frac{Mean\ Actual\ GDP}{RMSE}\right)$$

Here, the mean actual GDP represents the average of the actual GDP values across the dataset. The accuracy metric provides a relative assessment of the model's performance compared to a simple baseline model that predicts the mean GDP value. A higher accuracy percentage indicates superior predictive performance.

	Country	LSTM RMSE	LSTM Accuracy (%)	ARIMA RMSE	ARIMA Accuracy (%)
0	China	3.255896e+06	80.638659	2.242993e+06	86.661
1	Japan	7.456920e+05	84.383000	5.883982e+05	87.677
2	UK	4.990669e+05	83.256000	2.420672e+05	91.878
3	USA	1.944027e+06	91.746000	3.538478e+06	84.977

Figure5. Accuracy Comparison (LSTM v/s ARIMA)

5. EXPERIMENTAL RESULTS – MODEL INTERPRETABILITY

5.1. PERMUTATION-BASED FEATURE IMPORTANCE ANALYSIS

We conducted a comprehensive analysis using permutation-based feature importance to gain insights into the key determinants of GDP growth and enhance the interpretability of our forecasting models. This technique allowed us to quantify the relative importance of predictor variables and identify critical drivers influencing model accuracy.

We implemented this technique within the framework of the Random Forest Regressor, a powerful ensemble learning algorithm widely used for regression tasks. We trained a Random Forest regressor using our GDP dataset of 4 countries, comprising various predictor variables and evaluated the initial accuracy of the model. Next, we permuted the values of each predictor individually while keeping other variables constant. This involved randomly shuffling the values of a particular feature across all samples in the dataset. After permuting the feature values, we re-evaluated the trained model on the permuted dataset and calculated the change in prediction accuracy. This change in accuracy serves as a measure of the feature's importance, a significant decrease in accuracy indicates high feature importance, while a minor change suggests lower importance. The number of decision trees in the ensemble (n_estimators) were set to 100 to balance the model complexity and computational efficiency. We limited the maximum depth of each tree to 10 to prevent overfitting. Based on the analysis, we can conclude that out of all the economic indicators, GovtExp emerged as the most significant feature for predicting GDP Growth. This implies that variation in this feature has significant impact on the economic performance, as reflected in GDP.

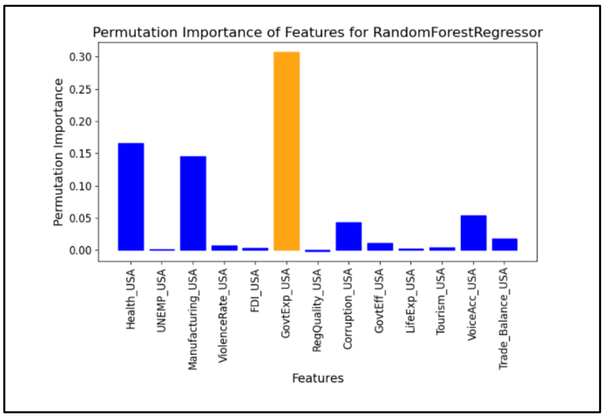

Fig 6. Permutation Based Feature Importance - USA

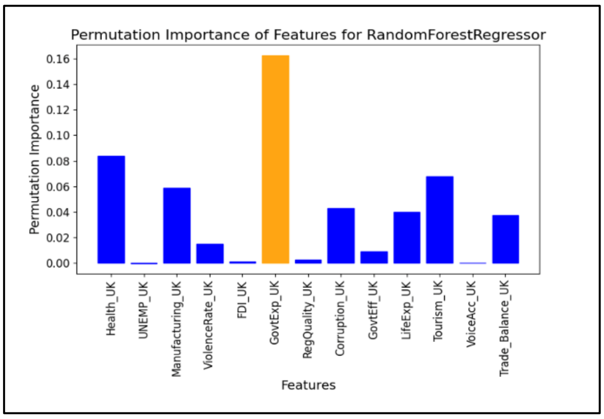

Fig 7. Permutation Based Feature Importance - UK

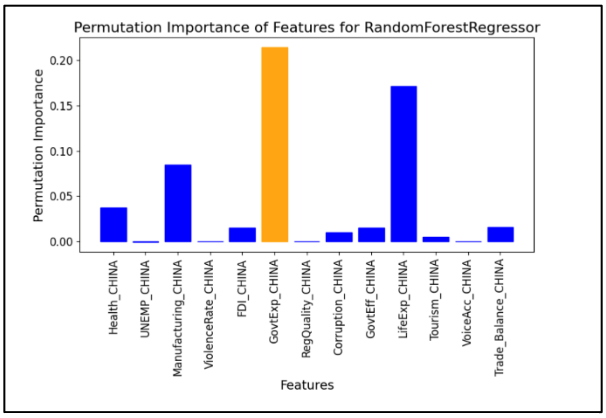

Fig 8. Permutation Based Feature Importance - China

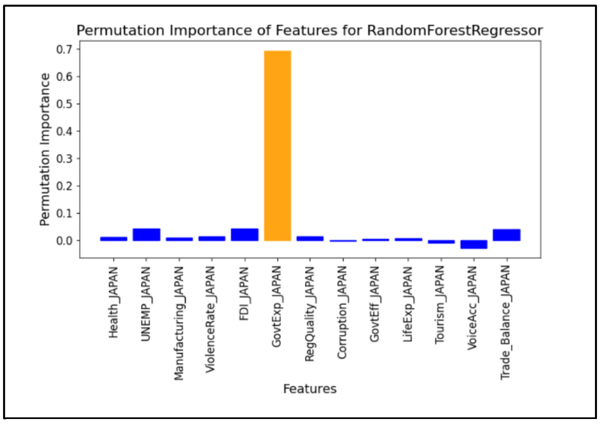

Fig 9. Permutation Based Feature Importance - Japan

5.2. PFFRA ANALYSIS

PFFRA is a data analysis technique used to identify and extract most influential factors driving the variability in the dataset. In addition to identifying the most influential factors driving economic GDP growth, PFFRA also helps in analyzing the trend associated with these factors over time. By examining the temporal patterns and trends in the factor loadings or weights assigned to each variable, we can gain valuable insights into how the relative importance of different economic factors may change over time and across different economic conditions. We implemented PFFRA using XGBoost Regressor with the following hyperparameters: Max_depth = 5, Learning rate = 0.1, n_estimators = 100. The dataset was split into training and test sets. Then PFFRA was performed using the 'permutation_importance' function from the 'eli5' package. This function calculated the importance of each feature by permuting its values and measuring the change in the model performance. PFFRA allows us to detect shifts in the underlying structures of the data and how these changes impact GDP growth. For example, we can identify periods where certain indicators become more or less important in driving economic growth.

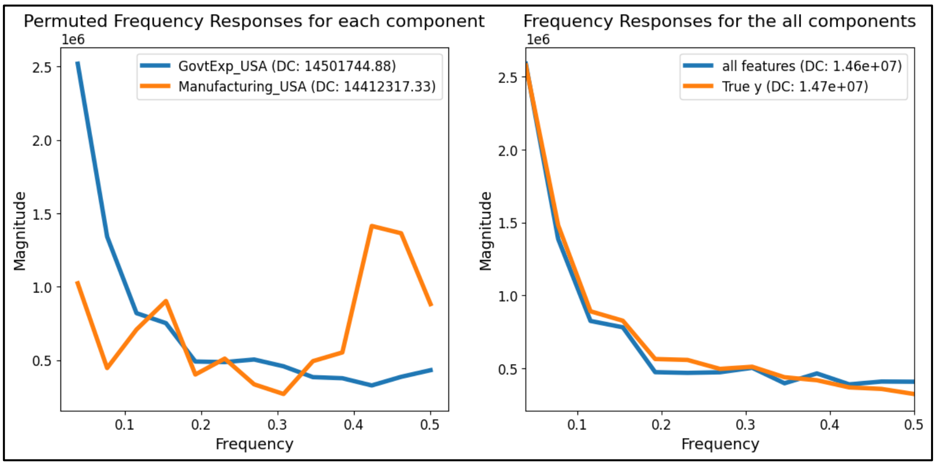

Fig 10. PFFRA Analysis (USA)

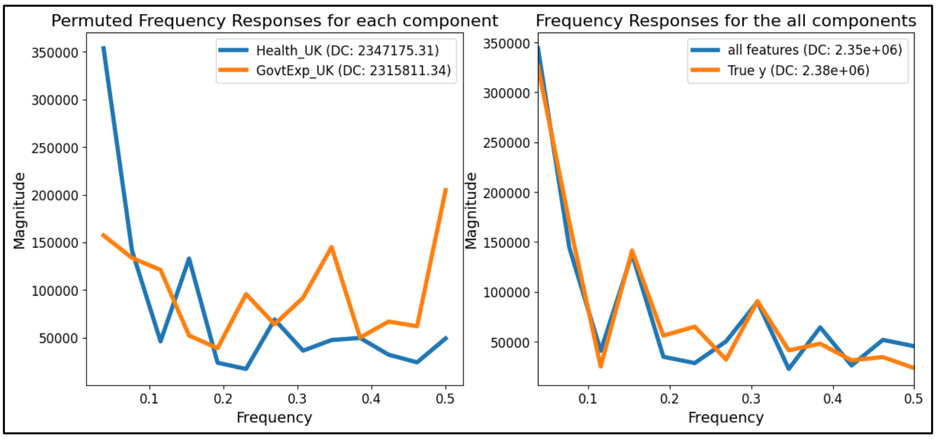

Fig 11. PFFRA Analysis (UK)

6. DISCUSSION

Our analysis compared the performance of LSTM and ARIMA in forecasting GDP growth for China, Japan, the UK, and USA. We evaluated the models based on their RMSE and accuracy Percentages. Results revealed variations in performance across different countries and modes. In China, Japan, and UK, our analysis revealed that ARIMA outperformed LSTM in terms of both RMSE and Accuracy percentages. This finding contrasts with the performance observed in the USA, where LSTM showed better performance.

One of the potential reasons behind the superior performance of ARIMA is the relatively stable and linear nature of GDP time series data in these 3 regions. ARIMA relies on Linear regression to capture the relationships between lagged observations and the target variable. Additionally, the simplicity and interpretability of ARIMA contributed to its better performance in these countries. Another factor to consider is the amount of available data for training the model. ARIMA typically requires a small amount of training data compared to LSTM, which makes it more suitable for datasets with limited historical observations.

Permutation-based feature importance analysis suggests that GovtExp is the feature with the highest importance value across all 4 countries. This suggests that government spending plays a critical role in influencing the GDP growth of these countries. In the USA, we also see Manufacturing and Health Expenditure emerge as top features. This reflects the importance of manufacturing (key driver of employment and economic activity) and healthcare (Investment in population health) sectors in driving the GDP growth of the USA. The prominence of LifeExp underscores the role of demographic factors in shaping the economic growth of China. Manufacturing remains a crucial sector contributing to China's GDP. In the UK, Healthcare and Tourism emerge as top features that highlight the significance of public health outcomes and foreign exchange earnings/employment generation. Features with near-zero importance values do not have a significant impact on GDP growth in the context of the specific model and data analyzed. Further investigation and sensitivity analysis may be warranted to validate the robustness of these findings.

The PFFRA (Permutation Feature Frequency Response Analysis) reveals interesting insights into the frequency domain dynamics of key features impacting GDP growth across different countries. For instance, in the USA, GovtExp exhibits high magnitude at lower frequencies, indicating its significance on the long-term economic trends. Conversely, Manufacturing shows high magnitude at higher frequencies, suggesting its impact on short-term fluctuations in GDP growth. Similar patterns can be seen across significant features from the UK which highlights the complex interplay between various economic factors and their temporal dynamics. It is essential to note that these observations provide only a preliminary understanding of the frequency domain dynamics and further research is needed to comprehensively capture the trends of all features across different countries.

7. LIMITATIONS

While our approach to the project has shown promising results, it is essential to acknowledge several limitations that may affect the generalization of our findings. One of the primary limitations is the availability of data. Utilizing annual data may introduce limitations due to its low granularity. Higher frequency data such as quarterly or monthly figures may provide more detailed insights into economic dynamics and improve the accuracy of the models. The complexity of models such as LSTM, used in our approach may pose challenges in terms of computational resources. These models require tuning numerous hyperparameters and may be prone to overfitting, especially when dealing with small datasets. Our models may not account for external factors or exogenous variables such as geopolitical events, policy changes, or natural disasters. Neglecting these factors may result in incomplete or biased forecasts reducing the robustness of our model. While we used commonly employed metrics like RMSE and accuracy to evaluate model performance, these metrics may not fully capture the predictive uncertainty of our forecasts.

7.1. FUTURE WORK

An avenue of future research involves integrating higher-frequency economic data into the forecasting framework. This would enable more accurate predictions by capturing short-term economic fluctuations. Techniques such as probabilistic forecasting and Bayesian Methods could be employed to generate probability distributions of future GDP Values, rather than single-point forecasts. This would provide decision-makers with valuable insights into the range of possible outcomes and associated confidence intervals. We also plan to extend the application of PFFRA to assess the importance of all features in our dataset comprehensively. Future efforts could focus on integrating alternative data sources such as social media data and online search trends into the forecasting framework. These sources could offer insights into various aspects such as consumer behavior and supply chain dynamics which may not be captured by traditional economic indicators.

8. CONCLUSION

In conclusion, our project represents a significant step forward in the field of GDP forecasting by leveraging advanced machine-learning techniques and innovative methodologies for model interpretability. Through the implementation of models such as LSTM and ARIMA, we have demonstrated the efficacy of different approaches to improve the accuracy and reliability of GDP forecasts. Additionally, we leveraged innovative approaches such as PFFRA to enhance model interpretability and identify key drivers of economic growth across 4 different countries. As we continue to refine and expand our analytical techniques, we aim to provide policymakers and stakeholders with robust tools for navigating the complexities of economic forecasting in an increasingly dynamic global landscape.

9. CONTRIBUTIONS

9.1. GROUP WORK

As a group, we went through various critical aspects in the research process and demonstrated good teamwork and coordination along the way. First, we brainstormed and developed sub-questions that aimed to capture the complexity of the underlying dynamics for the GDP forecasting problem. Collectively exploiting our various expertise and diverse perspectives, we fine-tuned our research focus in such a way that our research considers a comprehensive study on the economic features influencing the GDP across multiple countries. We collectively focused on finding and downloading relevant datasets from different countries for a comprehensive analysis capturing the nuances of different economic contexts. Each one of our group members was responsible for extracting relevant data for their assigned country and integrating everything together to make the data diverse. We also collectively did the EDA and created insightful visualizations and statistical summaries that uncovered trends and patterns within the data across the respective economies.

Additionally, we also helped each other out during the model selection and evaluation phase, running through the various forecasting algorithms and evaluation techniques to identify the best approaches to use in our project. Our group collaborated on interpreting the model's results, understanding the relevance of the performance metrics, and figuring out ways to fine-tune the hyperparameters used to improve the model generalizability. By collectively analyzing the results, we came to a much deeper understanding of the underlying factors driving GDP prediction across multiple countries. Finally, we collaborated on the development of our project presentation, Poster, and Final report. Through brainstorming and iterative refinement, we synthesized our findings into clear visual narratives. This ensured that every contribution from the team members was essential in forming various sections and maintaining the effectiveness of our presentation materials. Through our collective efforts, we successfully navigated through various challenges and produced a comprehensive analysis which enriches our understanding of GDP forecasting dynamics.

9.2. INDIVIDUAL CONTRIBUTION

9.2.1 SARA KALE

Throughout this project, I carefully designed core research questions to guide our investigation into the impact of the United Kingdom's Gross Domestic Product (GDP) on economic performance. I conducted thorough data analysis, employing detailed plotting techniques to uncover underlying trends. By the third week, I presented our progress in a clear and structured manner. The integration of various datasets was pivotal to ensure the robustness of our analysis, with a particular focus on investigating multicollinearity to validate regression assumptions.
Concentrating on linear regression methods due to their relevance, I applied an array of predictive models and systematically evaluated their efficacy. By discerning the most reliable model, I critically assessed its performance, engaging in nuanced discussions on hyperparameters and optimization techniques to refine accuracy. Analyzing the results, I drew insightful conclusions and crafted a visually engaging poster for the final presentation. Additionally, I developed a comprehensive final presentation for the midterm demonstration, encapsulating our research journey and its outcomes.

This meticulous approach not only facilitated a deeper comprehension of the relationship between GDP and economic performance but also underscored the significance of rigorous research methodologies. By adhering to systematic procedures and rigorous analysis, we were able to extract valuable insights that contribute to advancing our understanding of economic dynamics.

9.2.2 SUYASH YADAV

In this thesis, I have made several significant contributions to the field of economic forecasting. Initially, I suggested and conducted preliminary research on Tourism as a potential factor to explore within our Project. Following collaborative discussion, the team opted to shift its focus primarily towards predicting GDP. I established a solid foundation for our research by formulating precise research questions and sub-questions. This was followed by a meticulous process of acquiring and integrating a comprehensive dataset for GDP prediction in the USA. I conducted a basic Exploratory Data Analysis (EDA), which included plotting to understand the underlying patterns and trends within the data to uncover the economic trends of USA. To ensure a systematic approach, I created a presentation detailing our progress by the third week, which facilitated a clear and organized progression of our research. A critical aspect of our work involved addressing the correlations between various features and conducting a multicollinearity analysis, which is crucial for the reliability of any predictive model.

Furthermore, I applied a range of models from the domain of Ensemble Learning, including the Random Forest and XGBoost Regressors, to the dataset. Through rigorous testing and evaluation, I identified the best-performing model. I also delved into evaluation techniques and model tuning, discussing, and implementing various hyperparameters to enhance the model's predictive power. The culmination of my efforts was the analysis of the results, which provided insightful conclusions and potential implications for economic forecasting. Additionally, I synthesized my findings into a poster for the final presentation and prepared a comprehensive presentation for the midterm demonstration, both of which were instrumental in communicating our research to a broader audience.

9.2.3 SUCHIT PATHAK

My contributions to the project for predicting GDP were extensive and integral to its success. Initially, I suggested and conducted preliminary research on Inflation as a potential factor to explore within our Project. However, after collaborative deliberation, the team decided to pivot towards focusing primarily on GDP prediction. Despite this shift, my initial exploration into inflation dynamics provided valuable insights into the broader economic landscape, informing our understanding of the factors influencing GDP Trends. Furthermore, I actively participated in the data acquisition efforts, specifically sourcing relevant datasets related to the economic factors influencing the GDP of Japan.

In collaboration with my teammate Shalin, I undertook the critical task of handling missing values in the dataset by applying techniques such as Linear Interpolation/Mean Imputation to maintain the integrity and reliability of the data. I contributed to the normalization of the data, ensuring that the dataset was properly scaled for modeling purposes. This step helped in minimizing the bias in the data. Furthermore, I played a significant role in conducting the EDA Process which involved generating basic line plots for uncovering the economic trends in Japan. Moreover, my expertise in the time series analysis and forecasting proved invaluable during the model selection phase. Leveraging my understanding of the time series dynamics, I reviewed and implemented the ARIMA Model. Moreover, I played a significant role in troubleshooting issues related to ARIMA/LSTM Modelling, assisting in addressing challenges such as Model convergence problems and accuracy inflation issues. My contribution was significant in interpreting the model's results, employing advanced techniques such as PFFRA and Permutation Based Feature Importance to generate meaningful insights about the short term/long term trends and understanding the underlying drivers of the GDP.

9.2.4. SHALIN SAM

In this project, my contributions were integral to several key aspects of our research and analysis. Firstly, I played a significant role in framing our research question, which focused on identifying the economic indicators with a substantial impact on GDP forecasting. This involved extensive research across various countries to understand the key variables affecting GDP dynamics. Additionally, I took the lead in identifying and procuring the appropriate datasets essential for our GDP prediction models, ensuring the reliability and relevance of our data sources.

Furthermore, I undertook the critical task of data processing and analysis. This included merging datasets, handling missing values through sophisticated techniques like linear interpolation, and implementing normalization to standardize the data for modelling. I also conducted thorough statistical analysis, particularly focusing on feature analysis and correlation assessment to refine our dataset. Through this process, I identified and eliminated highly correlated features to enhance the quality and efficiency of our predictive models. Additionally, I was pivotal in selecting and implementing two advanced machine learning models through rigorous tuning to achieve superior accuracy. Moreover, I contributed significantly to the interpretability of our models, conducting PFFRA tests to assess the long and short-term impacts of individual features on GDP, which provided valuable insights into economic dynamics within our chosen countries. Overall, my multifaceted contributions were instrumental in shaping the research direction, data preparation, modelling strategies, and interpretative analyses crucial to the success of our project.

10. REFERENCES

[1] Okaro, C.S.O. (2020) 'Causal Relationship between Financial Structure and Economic Growth in Contemporary African Economy: A Case Study of Nigeria from 1990-2018,' Awka [Preprint]. https://www.academia.edu/86050426/Causal_Relationship_between_Financial_Structure_and_Economic_Growth_in_Contemporary_African_Economy_A_Case_Study_of_Nigeria_from_1990_2018.

[2] Emsia, E. and Coşkuner, Ç. (2015) 'Economic growth prediction using optimized support vector machines,' *Computational Economics*, 48(3), pp. 453–462. https://doi.org/10.1007/s10614-015-9528-1.

[3] Biau, O. and D´Elia, A. (2010) Euro Area GDP Forecast using large Survey Dataset - a Random forest approach. https://econpapers.repec.org/paper/ekd002596/259600029.htm.

[4] Saadah, S. and Wibowo, M.S. (2020) 'Prediction of gross domestic product (GDP) in Indonesia using deep learning algorithm,' 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) [Preprint]. https://doi.org/10.1109/isriti51436.2020.9315519.

[5] Pilström, P. and Pohl, S. (2009a) Forecasting GDP growth : The case of the Baltic States. https://lnu.diva-portal.org/smash/record.jsf?pid=diva2%3A229044&dswid=-8974.

[6] Yoon, J. (2020) 'Forecasting of real GDP growth using machine learning models: gradient boosting and random Forest approach,' Computational Economics, 57(1), pp. 247–265. https://doi.org/10.1007/s10614-020-10054-w.

[7] Plakandaras, V. et al. (2015) 'Forecasting the U.S. real house price index,' Economic Modelling, 45, pp. 259–267. https://doi.org/10.1016/j.econmod.2014.10.050.

[8] Götz, T. and Knetsch, T.A. (2019) 'Google data in bridge equation models for German GDP,' International Journal of Forecasting, 35(1), pp. 45–66. https://doi.org/10.1016/j.ijforecast.2018.08.001.

[9] Agu, S.C. et al. (2022) 'Predicting gross domestic product to macroeconomic indicators,' Intelligent Systems With Applications, 14, p. 200082. https://doi.org/10.1016/j.iswa.2022.200082.

[10] Predicting Indian GDP with Machine Learning: A Comparison of Regression Models (2023). https://ieeexplore.ieee.org/document/10113035.

[11] Lu, S. (2021) 'Research on GDP Forecast analysis combining BP Neural Network and ARIMA model,' Computational Intelligence and Neuroscience, 2021, pp. 1–10. https://doi.org/10.1155/2021/1026978.