Chronos with Ensemble GBT: A Hybrid Framework for GDP Forecasting with Zero-Shot Learning

by

Sarfuddin Siam
20101412
MD. Wahiduzzaman
20201004
Syed Safwat Sameer
20301064
MD. Sakib Hossain Molla
20201078

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
October 2024

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Sarfuddin Siam
20101412

MD. Wahiduzzaman
20201004

Syed Safwat Sameer
20301064

MD. Sakib Hossain Molla
20201078

Approval

The thesis titled ”Chronos with Ensemble GBT: A Hybrid Framework for GDP Forecasting with Zero-Shot Learning.” submitted by

1. Sarfuddin Siam (20101412)
2. MD. Wahiduzzaman (20201004)
3. Syed Safwat Sameer (20301064)
4. MD. Sakib Hossain Molla (20201078)

Of Summer, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on October 22, 2024.

Examining Committee:

Supervisor:
(Member)

Dr. Muhammad Iqbal Hossain
Associate Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisorr:
(Member)

Md. Golam Rabiul Alam, Ph.D.
Associate Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator::
(Member)

Md. Golam Rabiul Alam, Ph.D.
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi
Associate Professor and Chairperson
Department of Computer Science and Engineering
BRAC University

Abstract

Whether it is global or national, accurate economic forecasting is crucial for a country. It paves the direction of a country in terms of policy making, resource allocation, and risk management etc. There are several economic indicators such as interest rates, inflation rates, gross domestic product (GDP), unemployment rates, etc. to determine economic trends. But among them GDP is one of the main indicators for measuring one country's economic health. As a result, innumerable time series model and machine learning approaches have been developed to forecast the economic trend of a country. However, accurately predicting the trend of an economy is one of the most difficult tasks due to the highly diverse nature of all the economic indicators. This paper will use Decision Trees Based Ensemble Machine Learning models such as Light GBM, CatBoost and XGBoost, and LLM based model named Chronos to forecast GDP accurately. We have also ensembled Light GBM, CatBoost and XGBoost models to create an Ensemble GBT model. Finally, we create a hybrid model of Chronos and Ensemble GBT. We will be using the Penn World Table Datasets for our model. This dataset contains the Econometric data from 1980 to 2019 from 183 countries of the world. Our Objective is to perform a bench-marking test from our acquired datasets and compare our models. Afterward, this paper will also forecast the global GDP in the upcoming years. The paper has also used some of the traditional Time Series models like ARIMA, VAR and deep learning frameworks such as LSTM from other existing works as benchmarks. The hybrid model (Chronos x Ensemble GBT) generates enhanced predictions as it takes the best from both worlds. Across all calculated values, the model's performance is superior to all others reflected in MSE of 6.06e+09, RMSE of 7.78601e+4, MAE of 20935.24, R2 of 0.99. The paper has huge potential in the realms of forecasting economic indicators, global GDP growth and downfall.

Keywords: Chronos, Ensemble GBT, XGboost, Catboost, Light GBT, Machine Learning, Forecasting, Economic indicators, Prediction, Economy Analysis, Arima, Long Short Term Method, GDP

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis has been completed without any major interruption.

Secondly, to our honorable supervisor Dr. Md. Iqbal Hossain and Co-Supervisor Dr. Golam Rabiul Alam for their kind support and advice in our work. They helped us whenever needed.

Thirdly, we want to thank our research assistant Nafiz Imtiaz Rafin for his valuable guidance throughout the journey.

Finally, gratitude to our family and friends for their encouragement and unconditional moral support.

We would also like to thank all the concerned teachers and staff of the department for their direct and indirect assistance at different events of this work

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The forecasting of Gross Domestic Product (GDP) plays a vital role in any economy to formulate sound economic policies for sustainable growth. In the increasingly globalized and complex world economic order of today, a good understanding of potential future trends in underlying macroeconomic drivers not only aids long-term financial planning but also reinforces national resilience towards ambiguity. Other historical episodes, such as the 2008 financial crisis are also powerful reminders of the potential cost of poor forecasting — awarding policy makers with actionable pre-announcement reads could have saved entire countries from costly economic miscalculations after all. The problems increase in complexity as the world economies are more closely connected than ever before. Practitioners of macroeconomics are forever wandering through a maze constructed around them by the forest changing scent, data and answers getting lost in translation. Although traditional GDP forecasting methods are informative, they may not capture the nonlinearity present in economic data. The gulf between supply of and demand among business leaders for advanced analytics is a clear signal that new thinking needs to be injected into the market. In order to overcome these difficulties, this research aims at developing a new scientific contribution for the domain of GDP forecasting by using an hybrid model that brings together well-established statistical methods with state-of-the-art machine learning algorithms. In particular, we study the combination of an already complex time-series forecasting framework (Chronos) with a set of tuned models LightGBM, CatBoost and XGBoost. Since this is a multi-pronged strategy, it allows us to use the predictive power of every model in conjunction with each other making the forecasting better and unbreakable! We then refine each component to a polished state and combine them all into Chronos, our way of improving predictions as much as possible while providing meaningful information to economic agents. By no means are we after building a more accurate forecasting model, but rather equipping policy makers and the interested stakeholders with the tools to reason about an economic system that can change quickly and unpredictably. As we continue with this research, our hope is to contribute and make a case for economic forecasters on how the world around them will evolve, kneading crucial insights needed across decision-makers in order to navigate through complexities or seize growth opportunities. In the end, this research aspires to increase financial security by informing policies and investments better so that people feel less unsure about GDP forecasts. In the process, we expect to help build a more durable and sustainable world economy.

1.2 Problem Statement

This long-run goal makes predicting the Gross Domestic Product (GDP) of an economy one of the most important tasks worldwide towards sustainable economic growth. But navigating these waters is easier said than done and can make it exceptionally difficult for both teams of policy makers to be well-informed when making decisions. One of the primary hurdles to better analysis and prediction has been sheer complexity in economic systems—many factors interact with each other non-linearly, so it is difficult if not basically impossible to identify all the specific ingredients that drive GDP trends. This complexity often leads to suboptimal economic decisions leaving nations exposed in times of unexpected downturns. The global economy is changing, and with it comes the necessity of having more sophisticated models to predict what lies ahead. While traditional models still have historical relevance, they generally fail to capture the complexities of modern-day economic dynamics. Crisis scenarios, like economic downturns demonstrate the limitations of these tools as false predictions can result in expensive losses for finances and growth opportunities. Therefore novel approaches that tackle the limitations of existing forecasting practices are urgently required. In this research, we aim to design a Hybrid model which exploits both the best of traditional statistical models and advanced machine learning methods for creating time series forecasting. In this study, we do a post mortem of such integration using an highly tuned ensemble model with LightGBM,CatBoost and XGBoost for time series forecasting framework Chronos. These models provide interesting insights but are not exclusive from one another in terms of strengths to accentuate the predictive power of GDP forecasting across all models. The goal of this study is to identify which algorithms function best when used in combination (hybrid framework) post single algorithm optimisation: based on the performance values obtained for different forecasting horizons while predicting GDP growth accurately and reliably. The other seeks to understand what the determinants of these outcomes are in order that it might provide insights for economic stakeholders which could guide them with regards their decision-making. In the end, this study seeks to offer a holistic viewpoint on how sophisticated forecasting approaches can support improved economic policies and strategies. By demystifying the GDP prediction process, we help decision makers navigate a complex global economy more effectively, fueling growth and stability while sequestering any risk to financial institutions from erroneous forecasts. Projection: That is the light we aim to shed on a more robust economic future.

1.3 Research Objectives

- Analyze Development History behind the GDP: Use historical data of Gross Domestic Product (GDP) for discovering patterns and trends defining major themes related to global economic development, hence leading a way towards more credible models used for predictions. This Paper analyzes multiple time-series forecasting models, such as ARIMA and VAR to find the limitations of each in order to get a holistic view on which type of model can lead us to better predictions about GDP.

- Development of Tactical Models: Build easy-to-use, efficient GDP forecasting models that allow for real-time economic forecasts so policy makers and investors can make better decisions.

- Predictive Analytics and Economic Planning: Everything You Need to Know: Find Out How Advanced Predictive Models Could Influence the Economy Stability Around The Globe

- Iterative Upgrading of Model Capabilities: Suggest continual enhancements to forecasting models through the incorporation of complex features and algorithms for growing accuracy along with reliability in GDP predictions.

- Use of Sophisticated Machine Learning Techniques: Initiate GDP forecasting with a mixture of state-of-the-art machine learning models such as specially tuned LightGBM, CatBoost, and XGBoost algorithms followed by the Chronos model integration to boost predictive accuracy.

- Comparative Analysis of Model Performance — A detailed comparative evaluation between the hybrid forecasting model, conventional time series models and individual machine learning algorithms are performed to identify suitable techniques for effective GDP prediction.

- Discover what the important economic drivers are: Probe and determine the metrics that have major impacts on GDP prediction accuracy, examining how different attributes of economic measurements fit in various forecasting equations.

The research objectives of this project in GDP forecasting are designed to improve the policy making and general practice decisions for various economic agents globally given the increasingly complex global economy. This study aims to enhance the precision and reliability of GDP forecasts through advanced machine learning techniques as well as hybrid modeling, leading to better economic policies and informed decisions in a systematic way.

Chapter 2

Related works

2.1 Economic Indicator

2.1.1 Gross Domestic Product (GDP)

GDP (Gross Domestic Product) is an index that helps in the measurement of country's economic activity and it states the total value of goods produced and services provided within a county during one year.The foremost used macroeconomic indicator, Gross Domestic Product, reflects the total value of goods and services generated within a state during a particular period. This indicator is essential to determining the economic state, arranging policies, and predicting forthcoming trends. GDP can be identified through three approaches, including production, income, and expenditure. Its significance spreads over all areas bringing to the limelight policy, business and investment decisions. But forecasting GDP is especially challenging because of data scarcity, the unpredictability due to economic shocks and an inability for any model or even something as sophisticated systems realists use like integrated assessment models (IAMs) [2]. Many problems are associated with GDP Calculation To begin with, GDP does not include the output produced abroad by a country's nationals and profits earned by domestic firms in foreign countries; these are covered instead by Gross National Product (GNP) which includes aggregate value of all goods and services done initially for GNI rather. Again, one of which is that GDP only counts market transactions.GDP is also silent when it comes to the disappearance and depletion of natural goods. This can happen with things like oil extraction that technically increase GDP, but depletes reserves over time without reflecting as a negative adjustment in economic growth. [43] The paper "Online Machine Learning Approach for System Marginal Price Forecasting Using Multiple Economic Indicators: A Novel Model for Real-Time Decision Making" used a combination of machine learning-based batch learning and online learning techniques to forecast the System Marginal Price (SMP) in South Korea (Kim et al, 2023). The dataset consists of five energy sectors, two financial sectors, one transportation sector data. Their using machine learning algorithms are support vector regression, simple deep neural network, and deep neural network. After comparing their performance and found that simple deep neural network was the most accurate.The paper also introduces two methods namely weight modification and time interval updating.Their key contribution are the use of time interval data for high correlation between input and output, providing continuous and stable predictions through repeated batch and online learning processes, conducting Multi-Input Single-Output (MISO) modeling depend on input features, to get stable prediction results in the industrial sector for efficient energy planning. Lastly, they collect data from various sources and the results show that the simple DNN has the best result than other models.

[21] The paper puts focus into the economic performance of seven countries with emerging economies. They are called the E7 group in short. Then, their performance is compared with the economic performance of G7 groups. The time period is set into 2000 - 2017. The study focuses on comparing growth sources and performance between E7 and G7 economies. The dataset is used from the Conference Board Total Economy Database. Two methodologies were applied in the paper to figure out growth source and catch up performance - Growth accounting and catch up index. Growth accounting converts GDP growth into contributions from capital, labor, etc. Additionally, the catch-up index compares the economic performance against the US benchmark. The result shows that E7 had better GDP growth than G7 countries. Both E7 and G7 countries increased their GDP growth. In terms of economic performance against the US benchmark, six out of seven E7 countries improved their situation whereas five out of six G7 countries failed to improve. The paper concludes that the E7 group has overperformed G7's in terms of growth sources and economic performances. E7's catchup performance has made a significant improvement, specifically in Asia. [3]The paper "Demographic Determinants of Economic Growth in BRICS and Selected Developed Countries" focuses on the impact of changes in demography in terms of economic growth in BRICS countries. Then, the impact is compared with selected developed countries such as Japan, France, Singapore etc. Reduction in both mortality and fertility rates are crucial factors to understand the economic growth of nations. Here, the concept of demographic dividend is introduced. It indicates the creation of opportunity as the working age population rises compared to the dependent population. Therefore, labor supply, savings, and human capital increases. It all leads to a more efficient forecasting of economic growth. The paper uses a dataset from World Development indicators (WDI) and Penn World Table to evaluate the variables. The paper breaks down the GDP growth into multiple factors such as unemployment rate, working age population, dependent population, labor supply, etc. Finally, a regression model is used to forecast economic growth. It is concluded that growth of GDP per capita is an important factor towards faster growth. The paper raises a question on how effectively can the demographic dividend be expanded to ensure better economic growth. The findings to that concern is associated with the labor force skill formation capacity of the BRICS nation.

2.1.2 Expenditure-Side Real GDP (RGDPE) Advantages

In our study, we used the Real GDP Expenditure-Side (RGDPE) because it is the Gross Domestic Product which reflects actual growth of an economy adjusted for inflation on the expenditure side. Why RGDPE is a superior measure than nominal GDP for calculating inflation adjusted government spending? Because it removes artificial alterations incurred by changes in the price level which when eliminated, allow real economic activity to be more accurately compared across various time periods. Inflation adjusted: One issue with using GDP figures is that they can be significantly distorted by inflation — if prices rise, it may give an excessive view of economic growth without a corresponding increase in real output. RGDPE uses a price index to deflate the nominal GDP in order to account for inflation so that something like Gross Internal Production (GDP) figures would only be changed by what happened with respect to quantity of goods and services produced. RGDPE is adjusted for inflation ensuring a more representative gauge of economic growth overtime [2].

Advantages Of RGDPE Over Nominal GDP Here are a few benefits to using real growth of domestic income over nominal GDP:

- Better Inter-Temporal Comparisons: This means that when using RGDPE for spending comparisons, we adjust all of the different localities to reflect costs as measured in a single period — no matter how long ago they were incurred.
- Better Policy Decisions: Policymakers use the RGDPE as a measure for real economic movement so that they do not make policy based on nominal figures rather than what is true productivity growth.
- More Accurate Reflection of Economic Activity: By reflecting all economic activity at a constant price level, RGDPE is considered to be one the best measures for forecasting.

2.1.3 Methods of Predicting the Economic Indicator/GDP

GDP system, gathering GDP: Different types of arrangement and machine learning are applied to prediction the level which in turn displays GROSS DOMESTIC PRODUCT. These models attempt to capture both linear as well as non-linear relationships present in the macroeconomic indicators. [35] The paper "Application of Machine Learning Algorithms for Sustainable Business Management Based on Macro-Economic Data: Supervised Learning Techniques Approach" focused on forecasting the inflation rate and exchange rate of Pakistan from January 1989 to December 2020 using various machine learning algorithms (Khan et al,2019) [35]. Their using algorithms are k-nearest neighbor (KNN), polynomial regression (PR), artificial neural networks (ANN), and support vector machine (SVM). The dataset used in the paper covers the two macroeconomic indicators of Pakistan such as inflation and exchange rate.The research goal was identifying the best ML algorithm, determining the technique with the minimum error, finding the impact of hidden layer and the number of neurons per layer in ANN.They used root mean square error (RMSE) and mean absolute error (MAE) as performance evaluation metrics. The experimental results shows that ANNs has the best result compare to the other algorithms where RMSE was 1.070 and MAE was 0.820.

2.2 Machine Learning and Deep Learning in Forecasting

2.2.1 Traditional Methods

There are traditional methods, e.g. ARIMA (Auto-Regressive Integrated Moving Average), VAR(Vector Auto-Regressive) commonly used for time series forecasting. These methods presupposes linear relationships among economic variables. [56] The paper Data-Rich Economic Forecasting for Actuarial Applications by Felix Zhu and Fei Huang Gives us an economic forecast on the traditional models which relies on the econometric models . One of the models used earlier in the Thai field is the Dynamic Factor Models(DFM) by Stock and Watson which was later extended by Dorni et al and Bai and Ng. This has relatively small datasets focusing on the small datasets and focusing on the main key historical values. This kind of model mainly leverages linear relationships with data but while capturing non linear patterns they fail to comply. Recently due to advancement of neural networks the ability to capture structural changes can be seen which are superior to the traditional linear models. Using big data sets such as FRED databases which include rich accurate macroeconomic variables give better and more robust forecasting models. Machine learning and deep learning algorithms are becoming powerful enough to handle non-linear patterns in data, which traditional models may miss due to the large number of features. ML methods such as Bayesian Additive Regression Trees (BART), Elastic-Net, Gradient Boosting Machine (GBM) and XGBoost have been used for GDP forecasting. [31] This paper focuses on the creation of machine learning models. The initial targets are Gradient Boosting Model and Random Forest Model. These models are crucial to forecast real GDP growth accurately. The study puts the light on Japan's real GDP growth and predicts the growth from 2001 to 2018. The dataset is provided by the International Monetary Fund (IMF) and Bank of Japan (BOJ). The objective is to cross validate out of sample prediction and optimize hyper parameters. Mean Absolute Percentage Error (MAPE) is used to measure accuracy. The traditional economic model is vulnerable to irrelevant variables and assumptions. It might lead to inaccuracies when the variables and assumptions are flawed. This study used Gradient Boosting and Random Forest to forecast Japan's real GDP growth.

2.3 Simple Linear Regression Model of Statistics:

The Statistical Linear Regression Model is a basic statistical tool used to understand and predict the connection between an independent variable (for example, GDP) on which some others certain factors depend Mathematically, it is based on the assumption of a linear relationship:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{2.1}$$

Here, Y is the dependent variable (GDP), $X_1, X_2, \ldots, X_n$ are independent variables, β_0 is an intercept and for each of them we have a coefficient $(\beta_1, \ldots, \beta_n)$ that measures how much does it affects GDP variations. ϵ is the "error term", which explains variability in Y not explained by economic factors in the model; this way economists can examine how many influences do different economic elements pose on their data about variation within GDP enforcing numerical values to these large influencing stakeholders with precision using such equation/s.

In addition, it helps in generating accurate predictions that help governments as well as stakeholders to make effective economic decisions. Also, the simplicity and interoperability of linear regression makes it helpful for understanding relationships in economics. This can help with strategic planning as well as deepening your theoretical insights about economic phenomena even when you are not using a strict neoclassical lens to view things through. ARIMA models are used for forecasting as well as GDP of Bangladesh from 1968 to 2022 with an extension of the Box-Jenkins [6] method as a one-step way which is suitable for selecting an ARIMA model by removing non-stationarity. This study also explains the systematized behavior of the Box-Jenkins methodology and its application on real GDP data using ARIMA. On the downside, its linear univariate model is a drawback, and we do not have comparison with state-of-the-art machine learning or deep learning techniques. This gap opens up a spot for your thesis to analyze how models like GBRT or LSTM outperform ARIMA when used as predictive tools for economic indicators.

The ICRTD research work by Miah et al. [11] on Evergreen Research: "Modelling and Forecasting of GDP in Bangladesh: An ARIMA Approach" investigated the possible uses of the ARIMA model to forecast GDP in Bangladesh (2019). Jacob Andersson in his paper, "Forecasting Swedish GDP Growth" [12], explored the effectiveness of Vector Autoregressive (VAR) models in forecasting GDP. This paper uses VAR models and compares their ability to predict Swedish real GDP growth with Random Walk (RW) and Autoregressive (AR) modeling techniques. Utilizing quarterly data for 1993-2006, the paper finds that VAR models, which include forward-looking survey measures such as consumer confidence and manufacturing industry confidence, economically outperform AR and RW benchmark forecasters at all horizons. Deep learning models are advanced machine learning algorithms that use neural networks with many layers to process high-level patterns in large data sets. These methods are specifically useful in the case of GDP forecasts to capture non-linearities and complex interdependencies among different economic variables. Through architectures like Long Short-Term Memory (LSTM) networks, which support temporal data processing and learn from extensive historical economic information, deep learning has yielded accurate GDP trend predictions. It extracts the relevant features from raw data on its own, thus increasing forecasting accuracy, which in turn helps policymakers and stakeholders make decisions based on reliable economic insight. The renowned paper by Yann LeCun, Yoshua Bengio, and Geoffrey Hinton [5] on deep learning, published in Nature in 2015, dives into the architecture methodologies of deep learning methods and how they can be transformative across various fields.

Park and Yang [36] tackle the emergence of a different form of economic forecasting based on deep learning and artificial intelligence methods together with machine learning in their research "Interpretable Deep Learning LSTM Model for Intelligent Economic Decision-Making." Originally designed to create fast economic forecasts between 1990 and 2019, the quarterly data were collected on economic and financial indicators for the fifteen most significant members within the G20. Among them, six are advanced economy members, including France, Germany, Italy, Japan, the United Kingdom, and the United States. Emerging nations include Argentina, Brazil, China, India, Indonesia, Mexico, Russia, South Korea, and Turkey. All data is pulled from the three major databases: FRED, CEIC, and BIS. Elsayed et al. [23] check whether deep learning models are necessary in time series forecasting or if simpler machine learning models such as Gradient Boosting Regression Tree (GBRT) are more intuitive. In this research, with window-based regression, proper input engineering for GBRT significantly outperforms current best deep learning models such as LSTM across different datasets. The GBRT model gives impressive performance, but it requires feature engineering, which restricts its use in some fields. It also contradicts the common belief that deep learning automatically enjoys an advantage, arguing instead for simple models capable of rivaling more complex ones with reduced computational costs.

2.4 Decision Trees Based Ensemble Machine Learning Method

It is an ensemble machine learning method of Decision trees, which improves the accuracy by aggregating multiple decision tree models. One of this method's particular strengths is that it resolves the weaknesses associated with single-tree models (overfitting, unstable; high variance), and instead combines many trees to make predictions strong, reliable. Ensemble techniques (Random Forest, Gradient Boosting) all involve different ways of training each tree on data subsets or sampling mechanisms. This diversity enables the ensemble to capture rich patterns and relationships which exist in this dataset, therefore making it fit for purpose when trying out tasks like GDP forecasting where there are various interrelated causes that influence economic outcomes. For GDP forecasting in the paper [19], ensemble methods like decision trees are particularly well suited for capturing the subtle interactions between different economic measures. These models give us more stable and accurate predictions around a wide complex datasets by salvation the output of many decision trees. Ensemble methods also provide explanations, which allows users to understand the reasons behind changes in economics. This transparency is important for policy makers and economists to take strategic decisions or plan the future, in an economic environment that changes rapidly. Methodologies such as decision trees based ensemble machine learning, then, can translate to more accurate GDP predictions when combined with the right economic data and other techniques for effective macroeconomic policy management.

2.4.1 XGBoost

The paper "GDP Growth Prediction of Bangladesh using Machine Learning Algorithm" explores a significant challenge where the task is to forecast GDP growth of Bangladesh based on machine learning algorithms [24]. Considering the synergy of factors affecting economic growth, this study aims to determine a model for accurate forecasting GDP growth rates using machine learning. Using 40 years of data set, the researchers analyze how different independent parameters affect GDP growth.

The dataset they used for their experiments is from 'Kaggle. com', which had 40 years worth of different economic variables. And next they identified 9 features of the dataset. They consider these features and take 'GDPGrowth' as a dependent variable so rest should be your independent variables. Then they filled the missing value of the dataset and identified relationship analogy b/w null value and lower GDPGrowth. They also impute the missing value by finding out the median of that respective feature. In addition to the seven regression algorithms that were tested [19], such as Linear Regression, Random Forest Regressor and Gradient Boosting Regressor, also obtaining similar results regarding cross-validated MAE scores, we observe an improvement on all methods from our pipeline. The author [33] of this article also highlights the benefits associated with using XGBoost for examining factors determining economic stability at a regional scale in "Assessment of Factors Determining Regional Economic Stability Using the Example of Poland" Gradient boost : XGBoost is a advanced tree boosting machine learning algorithm which has inbuilt concept of ensemble technique it trains towards the gradient decent here each node corrections trees build on previous error to improve accuracy predictions. The research indicates that XGBoost is especially powerful for big and complex data sets, which can be invaluable when forecasting regional economic success. The authors utilized XGBoost to classify Russian regions as either economically safe or unsafe by examining investment risk, Human Development Index (HDI), and composite financial result /Gross Regional Product.(GRP) ratio.

2.4.2 LightGBM

LightGBM (Light Gradient Boosting Machine) is a tree-based ensemble machine learning algorithm, designed to be distributed and efficient for handling large-scale dataset and complex regression tasks. For predicting time series data like GDP, LightGBM is an excellent fit since many of these economic variables often have subtle relationships with each other and therefore it would be beneficial to capture all those when making predictions. Utilizing LightGBM in this study, we believe that the predictability of GDP forecasts with stronger ability will support policy makers and stakeholders are to make better-informed decisions as they navigate a changing economic landscape toward growth future. The paper [25] for light-GBM Optimized LSTM and Time-Series Forecasting Model for Economic Time Series Analysis: This research on "An Economic Forecasting Method Based on the LightGBM-Optimized LSTM and ARIMA" [25] highlights major advantage of forecasting, such as GDP LightGBM being an ML algorithm has shown superiority over the traditional ARIMA and other neural network approach such as RNN/GRU due to its capability of handling large-scale, high-dimensional datasets more efficiently. The leaf-wise tree growth algorithm may grow trees deeper, so the prediction is more precise. Moreover, LightGBM can handle missing data without preprocessing the missing values and not only is it more flexible but also simplifies data preparation phase.

2.4.3 CatBoost

CatBoost, an advanced implementation designed for high efficiency in terms of computational power, memory consumption and model generalization was the right choice to forecast this European national GDP and its extension. The library is distributed under Boost Software License (the Free Software Foundation approved it as GPL compatible) by Yandex LLC, a Russian multinational company that specializes mainly on Information Technology — made available to public use since April 2017. With the use of CatBoost in this work, our goal is to provide precise and reliable GDP predictions that can help policymakers (and other government agencies) as well as stakeholders with valuable inputs useful for making informed strategic economic choices in an increasingly turbulent world [39]. The work "Multi-dimensional data-based medium- and long-term power-load forecasting using double-layer CatBoost" by Wen Xiang et al.[39] shows that it is an excellent method to improve the accuracy of power load forecast (PLF) as compared with other algorithms in practice In this study, a two-layer CatBoost model is used to effectively use economic as well as meteorological and power generation data integration service of traditional forecasting models. Most of these conventional models rely on one-dimensional data and fail to capture the complex, nonlinear relationships between different influences. The new CatBoost model yields a major gain in forecasting robustness and accuracy by enriching external factors with multiple dimensions. Performance of CatBoost model This study reveals that the proposed model outperforms related state-of-the-art models (XGBoots and AdaBoost) as indicated by 0.925 R^2, MAPE with rate equal to 0.0158% and RMSE equivalent with value of 2743036 than other methods .These measures signify a substantial improvement of predictive accuracy and model stability. The authors go further to fine tune the accuracy of their model by using randomized search cross validation for optimizing parameters. 'A Study on China Coal Price Forecasting Based world Enough Computing Higher order Empirical Mode Decomposition Adaptive not Central Intelligence Agency Person firepower-approximate very well CatBoost Hybrid Forecasting Model under Carbon Neutral Target [38] published in the journal Energies, provides a structured framework using machine learning techniques to predict complex economic indices (the key indicator 'coal price', which is affected by many factors.

2.5 Time Series Ensemble Models

In machine learning, an ensemble model has become increasingly popular due to the power of combining different algorithm strengths to enable better predictive accuracy and reliability. Objective: The study aimed to develop an ensemble model by combining three advanced gradient boosting techniques, XGBoost, LightGBM and CatBoost. It takes advantage of some quirks (in other words, "efficient features") in recent versions of QtWebKit. A Fusion Framework for Forecasting Financial Market Direction Using Enhanced Ensemble Models and Technical Indicators [26] analyze the efficacies of ensemble models at forecasting stock market direction. The present study employs a stacking method, where six boosting models—XGBoost, AdaBoost (Adaptive Boosting), Gradient Boosting, LightGBM (Light gradient boosting machine-learning framework) and Catboost—are utilized with Histogram-based Gradient-boost boosters to show that the combination of these algorithms significantly improves prediction performance. In the paper Ensemble MethodologyInnovations in Credit Default Prediction Using LightGBM, XGBoost and Local Ensemble researches better ensemble models for credit default prediction [55] which may come with some knowledge that can also be used on GDP

Forecasting where model accuracy is important as generalization! This repository contains a Variation of the LocalEnsemble model, using ensemble framework that includes LightGBM and XGBoost models. For instance, LightGBM on large datasets where data imbalances are common has unique value; so does XGBoost in intricate feature engineering and LocalEnsemble for separating out interactions between a variety of different features. Integrating predictions from many models to solve the usual limitations of single-model overfitting or bias and make resilience as well as better accuracy in general.

2.5.1 Hyperparameter Tuning

To improve economic indicators forecasting like GDP machine learning, model hyperparameter tuning is put into use. This act is even more important now, since hyperparameters are predefined values that will affect how well a model learns complex data patterns. Good tuning of hyper-parameters can be the difference between a predictive model reaching an acceptable level or not, but also is quite important to avoid overfitting so that the models has good generalization with unseen data. Grid search, random searches and Bayesian optimization are some of the commonly used techniques for conducting a hyperparameter space sweep in an organized manner. Hyper-Parameter Tuning: e.g., in the context of economic forecasting, tuning is essential to reduce overfitting and underfitting leading us to more reliable prediction[13]. This is especially important in the case of ensemble methods, as interactions among hyperparameters across different algorithms can have a dramatic effect on performance.

2.5.2 Fine Tuning of Hyperparameter for Ensemble Machine Learning Techniques

For ensemble machine learning models such as CatBoost, LightGBM and XGBoost they each have their own set of parameters that must be tuned. For instance, In CatBoost the learning rate, depth and iterations are very important to get efficient learning whereas in Light GBM num leaves,max depth and bagging fraction affect performance drastically. Doing a cross-validation helps assessing different combinations of hyperparameters and shows you what could be the best model for your specific problem. More advanced optimization techniques such as Bayesian optimization can further facilitate this process by providing more streamlined navigation in hyperparameter spaces. GDP forecasts can be made more accurate and robust by focusing on ensemble techniques, which combine the strengths of multiple algorithms. This paper [28] has presented an analysis of pretraining and fine-tuning strategies for more efficient scaling of transformer models. The authors argue that the common belief suggesting model size is the sole king-maker next to performance forgets or overlooks a complementary attribute of models i.e their shape, where by this they mean architecture — depth and width – which are as relevant during fine-tuning. Ultimately, their experiments demonstrate that by careful design smaller models can indeed approach performance of larger models at much lower computational costs and in training time. [28] They introduce the “DeepNarrow” strategy, balancing model depth and width for improved scaling. Still, one of the most significant is their public release of 100+ pretrained T5 model configurations. Yet the paper also cautions against leaning too much confidence on equations of upstream pretraining results from which intuitive benefit may not transpire downstream performance. The authors also caution that scaling strategies effective for small models do not necessarily generalize to larger models, an important insight for researchers working with limited computational resources.

2.5.3 Parameter Optimization for Chronos and Hybrid Approach

The conventional approach to hyperparameter tuning for the Chronos model and hybrid methods, combining time-series forecasting with ensemble techniques, requires a targeted effort to tune all relevant parameters adequately. For Chronos, that includes window size, stride for how far apart windows are (a.k.a the resolution of temporal dependencies), and num layers since more non-linearity is always a good thing. And when used in an ensemble with CatBoost or XGBoost, you have to be mindful of how they interact. [37]Nested cross-validation techniques allow for a straightforward assessment of the overall hybrid model tuning. Solvers — Using adaptive learning rate techniques can also improve convergence time when solving with both the ensemble and its algorithms. Precise GDP forecasts are paramount for creating meaningful economic insights by harmonizing temporal patterns and intricate data relationships, so an efficient hyperparameter tuning not only of Chronos but also the hybrid methods is mandatory.

2.6 Chronos

Chronos is a time series forecasting framework which leverages a language modeling structure. Since language models are working on tokens of a finite model vocabulary than it is used for learnability and time series data has real valued items, there can be significant difficulties with trying to directly apply the two. As such, Chronos converts continuous real-valued time series data into a quantized discrete set of tokens (using scaling and quantization). Thus it is possible to apply the language models, such as T5 or GPT-2, also for forecasting by feeding sequences of these tokens. These language models are learning to predict the future values from a past sequences of time-series tokens without using traditional features like timestamps or lags and in that sense also, we can say they're really getting trained on the "language" of time series. [46] The work presents Chronos, a language model adapted for time series forecasting with a transformer-based architecture. Chronos tokenizes time series data into discrete values to make possible the application of traditional transformer language models previously designed for text data. The authors further pretrain the model on large-scale datasets including synthetically generated data from Gaussian processes and test it against 42 benchmark datasets. One critical strength of Chronos is that, without task-specific modification, it works on both the in-domain and zero-shot forecasting tasks, beating classical models like ARIMA and ETS, along with some deep learning models. In addition to this, the study emphasizes the small adaptation necessary to employ transformer models for time series forecasting, stating implicitly that language models capture sequential patterns as a common feature which also plays an essential role in time series data. While the method is promising, one of the possible limitations consists in the fact that, having such a model reliant on tokenization and restricting the prediction range to pre-defined bins, strong trends in datasets can be challenging. Here, we propose a framework called TIME-LLM that reprograms large language models to do time series forecasting by aligning the input of the model with natural language models. [53] Contrary to prior finetuning or extensive task-customized training methods, TIME-LLM introduces a reprogramming approach that preserves the pre-trained LLMs framework and adds prompt-assisted updates to enrich its logic reasoning and prediction functions. To remedy this, they propose: *Prompt-as-Prefix* (PaP), a technique to enhance input context and guide in the reasoning of such models about time series data. Experiments show that after learning under few-shot and zero-shot instances, TIME-LLM outperforms the SOTA forecastability in terms of forecasting accuracy with near 0 distance to ground truth. The study will demonstrate that LLMs may generalize across domains and perform strongly with only few past experiences. Moreover, TIME-LLMs possibility of multi-modal knowledge abstraction from different domains would open a new way to time series modeling in the upcoming era.

In this paper,[45] we present PromptCast, a new large-scale pre-trained language model capable of performing time series forecasting. PromptCast avoids numerical forecasting (the traditional way in which this kind of models are trained), as it looks at each time series and maps a sequence of values to prompts such that the final task is seeing how well can you map one sentence into another. This paper [51], we reveal the unexpected capability of Large Language Models (LLMs) that like GPT-3 and LLaMA-2 for time sequence forecasting. Specifically, the LLMTIME framework transforms time series data into digit strings and maps forecasting tasks to a next-token prediction problem for LLMs. They found that LLMs could generally outperform traditional time series models even with no fine-tuning on the target dataset and achieve zero-shot forecasting. This approach takes advantage of the capacity that LLMs have to model multimodal data distributions and there ability to perform things like missing data handling without imputation. On the other hand, they identify some tokenization problems such as numerical tokens that are not correctly treated in ways which can severely affect model performance. It also shows that some impersonal feedback in significantly worse forecasting as we see with the case of GPT-4 – performing more poorly than Huehnertier. This study further studies the scale effect of model size for forecasting and indicates that an efficient encoding strategy is required to exploit LLMs' full potential in time series forecast.

2.7 Ensembled based hybrid Models

Ensemble-based hybrid models based on XGBoost, LightGBM and CatBoost are very powerful for predictive tasks including time-series datasets.[42] A set of those models that are built on different random subsets of data using GB.These models (with Gradient Boosting Decision Trees as their base learner) can be combined in a number of ways including averaging predicted values to decrease variability known as bagging or logistic model has used outcomes from several other classifiers instead for making predictions meta-estimator like stacked generalization where output generated by combining outputs with another , The Gradient Boosting models that are generally used in hybrid systems, coupled with the deep learning architectures like CNN-LSTM can be useful to model both feature-based patterns and sequential time series data effectively with further enhancement on focus mechanism using attention mechanisms which helps LSTM cells concentrate relevant information. [40] For the optimization of these ensemble models, we should use bayesian optimization to fine-tune hyperparameters like learning rate and tree depth efficiently(in Hyper-Opt), which is aimed at increasing model accuracy as well as decreasing experiment time. The blend of boosting technique, deep learning and tuning methods provides improve processing robustness and prediction.

2.8 Comparative Studies

This paper [49] investigates the performance of LLMs for 134 macroeconomic time series with model configurations as in Bai and Ng (2003) relative to a host of traditional methods such as Bayesian Vector Autoregressions (BVARs) or Factor Models forecasting out-of-the-sample using the FRED-MD dataset. The study demonstrates that LLM is advantageous for tracking nonlinear patterns and evolving forecasting environments. While the results show that LLMs such as Moirai from Salesforce and TimesFM of Google are competing fully with econometric models in some settings especially not for generating consistent outcomes across different variables, particularly on pseudo out-of-sample basis. This adds an extra blot of complexity since the pre-trained LLMs will have seen very large data (which can even include target forecasting variables) and thus could contaminate, making comparing unfair. At the same time, LLMs exhibit unique patterns of improvement in post-COVID periods as well which suggests that their ability to adapt trends made evident by rare data sources is an advantage over traditional models. The paper concludes that, while these results are a good start, econometric models remain more stable on the whole. [4] The book "Time Series Analysis: Forecasting and Control," is a formative piece in the sector of time series analysis. It was written by Greta M. Ljung, George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel.The most noteworthy part of this book is the AutoRegressive Moving Average (ARMA) models. These models left a significant mark on the evolution of time series analysis models.The methodology basically converts the time series data into dependent variables. The conversion depends on explanatory variables and residual stochastic variations. The advantage of this methodology is it allows for valid inferences regarding the timing and value of dependency between time series. This plays a crucial role in forecasting and control. A new state space model and Kalman filter for forecasting is also used to handle non-stationary and seasonal time series data. Few practical applications of these methodologies would be economic forecasting, daily gas demand on temperature, impact of marketing on sales. The book can be considered as a crucial resource for everyone, starting from students to professionals. It ensures the understanding of time series models in practical fields.

2.8.1 Growth of Deep Learning Techniques

[22] The paper author Bunyamin & Meyliana analyze "Classical and Deep Learning Time Series Prediction Techniques in the Case of Indonesian Economic Growth"

traditional time series prediction methods (2019). This study examines Indonesian economic growth using a variety of time series prediction methods (2019). In this work, the authors employed data from the World Development Indicators (WDI) with observations spanning 1962 to 2016 to present a broad analysis of Indonesian economic growth over time (Bunyamin & Meyliana 2019). [27] Prachyachuwong and Vateekul off a thorough investigation in "Stock Trend Prediction Using Deep Learning Approach on Technical Indicator and Industrial Specific Information" (2019) rely on deep learning models for stock market prediction. End of day stock price of Stock Exchange of Thailand (SET50) from 10 January 2014 to 3 February 2020 contains in the dataset end-of-day data. Thai economic news headline from several news website is also contained dated 10 January 2014 to 3 February 2020. The data includes open, high, low, close prices, and volume. Several deep learning models including Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT) are applied. The models using only numerical data (stock price and technical indicators) vs. models using both numerical and textual data (news headlines) are compared. The evaluations are carried out considering the models' accuracy, F1-score, and their simulated-trading-based annualized returns. The model that fuses BERT with numerical data across sectors (BERT_SEC+NUM) achieves the highest average accuracy at 61.28% as well as the highest F1 score at 59.58%, making it the most accurate model overall as well. Also, it yields the highest average annualized return across trading simulations at 8.47%, making it the most lucrative model on average across sectors. The paper shows the relevance of deep-learning methods for stock-market prediction, as well as the effectiveness of combining textual and numerical data and in particular when those texts pertain to news linked to specific industries. Ramírez et al. "Artificial intelligence and its impact on the prediction of economic indicators" makes a systematic review about the use of techniques of AI for the forecast of economical indicators [20]. The authors made "a systematic literature paper (SLR) on many databases such as Web of Science (WOS), Scopus and Google Scholar" and , "They did not did raw data on purpose".This research's goal is to find academic papers on the period 2015 to 2019 that involving the use of AI techniques to predict economical indicators.The review found the AI techniques "adaptive neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), genetic programming (GP), support vector regression (SVR), extreme learning used various machine learning techniques.

2.8.2 Development of Hybrid Models

Hybrid models incorporate the advantages of both traditional statistical methods and advanced machine learning techniques in forecasting systems with improved precision and flexibility[14]. Hybrid Models are all about combining econometrics models (like multiple linear regressions) with machine learning types of techniques such as artificial neural networks (ANNs). This makes it possible for the model to work on both linear and non-linear relationships in data. This hybrid approach has also been employed for the Horizon system, which is used to forecast Russian economic indicators. It combines with regression models and neural networks to improve the predictive accuracy, especially in cases where a traditional regression model is insufficient. Using this combination, the system gives significantly better predictions for 70 major economic indicators much greatly accurate forecast for about four out of five. In this case, the advantages of a separate model can become part of an answer to local problems group feedback or initial default string extension in other types that standard models miss. For example, they are good when there is a linear relationship between the variables and naively fails to capture all complex dependencies. Neural networks, however are capable to model these non-linear patterns, being great in the use of improving forecasts when classical models perform poorly. This essentially allows the hybrid system to flexibly calibrate itself according to input data complexities, enabling it then delivers more accurate predictions for important socio-economic indicators. [7] The paper author Susan Athey assesses the past contributions and predicts the future contribution of machine learning in the field of economics. It estimates the counterfactual policies in economics due to the contrast with traditional approaches. The paper reviews "off the shelf" machine learning applications in terms of both texts and images. The paper, then voices a concern of machine learning's application capability considering the fairness and manipulatibility factor. It also focuses on merging machine learning and causal inference.

2.9 Context Matters

Context matters a lot in the collection and analysis of data within GDP forecasting In fact the type of model used could differ such that we obtain accurate predictions. Economic indicators are impacted by forces such as political stability, global market conditions, fiscal policies and events that can never be accurately predicted in advance (e.g. pandemics) or natural disasters like climate change-induced drought 1011 [50]. Therefore, the forecasting model must be laws to work under these flexible conditions. Classical models such as ARIMA may work just fine in those less volatile economic periods. But when things get a bit more volatile, or there is real structural evolution at play then it would be time to deploy higher powered weaponry — AKA with machine learning (ML) and deep learning (DL). They can capture complex nonlinearities and interactions, making them reliable for forecasting in a variety of situations. Models like CatBoost, XGboost, Temporal Convolutional Networks (TCNs) Short-term patterns that may be missed by lower-frequency data come to life with high frequency, which can help advanced ML and DL models predict more efficiently. Some of the ways to do so may include including varied data sources like macroeconomic indicators and social media sentiment, thereby adding more granulation in this forecasting model. [34] The paper The Potential Impact of COVID-19 on the Chinese GDP, Trade, and Economy Gives an in depth overview of the multifaceted impact of the covid 19 pandemic on the economy of china. The pandemic was directly responsible for global trade disruption along with a crucial impact on the economical sectors which resulted in fall or GDP , stagnation of gdp growth rate , overseas trade etc.This paper also highlights the severity of the pandemic on the heath sectors and the people of china This research paper mainly focuses on correlation analysis, descriptive statical input and unit root test which gives a clear analysis of the economical indicators. This analysis shows a sharp drastic decline in almost all the sectors of the economical activities across the sectors. Some drastic measures in long term loan payment were suggested to overcome the sudden economic downfall and promote continuity and employment.The study also stated the urgency of medical treatment and effectiveness of the vaccine for the medial sector to ensure more efficient recovery which will last until 2025 .It also stated that the economic recovery is linked to public health strategies and it emphasizes the significance of maintaining constant dara variance and addressing heteroskedasticity in economic modeling. Lastly we can say the comprehensive analysis of the paper provides the critical insight and policy reforms and recommendations to evade the downturn of the economic impacts of COVID-19 in china.

[8] Bin Weng et al. (2018) works on a two-stage approach to predict the one-month ahead price of major U.S. stock and sector indices using macroeconomic variables. The first approach tests the hypothesis that different economic indicators drive the prices of different indices.This stage divided into three phases by using seven model they are data acquisition and preparation, predictor selection, and prediction.They used Mean Absolute Percent Error and two other metrics two check the model.In the second stage they used a hybrid model that combination of recurring neural network used for time-series prediction and ensemble models. They also used some factors in second stage such as Consumer Sentiment etc.They used the data of monthly closing price for 13 U.S. stock and sector indices to test the model.The result shows ensemble models has the best result where MAPE was less than 1.87% and the stage two result can be improved by 25-50% by using the macroeconomic indicators. [15] The data in Allen et al. came from Levy-Kalecki Corporate Profit Equation published in the paper "Analytical Approaches to Macroeconomic Forecasting: A Study of Profits through Machine Learning and Deep Learning," in 2020. They assembled the data from two primary sources. The United States Bureau of Economic Analysis (BEA) provided the fundamental variables used in the equation and St. Louis Federal Reserve Economic Data (FRED) System then provided data on these variables that was updated and went beyond 2013, from where these variables as a function of time originally ended.

Despite the fact that they were able to collect the data through the FRED system, the BEA remained the primary source of each data point. This dual-source method ensured that the study's prediction modeling used legitimate and full economic data. The research of this paper posits that there are three key limitations to the existing literature on forecasting economic indicators such as GDP that limit efforts in improving current methodologies. A common problem is that older statistical models (ARIMA, exponential smoothing) based on linearity and stationarity assumptions. These kinds of assumptions frequently give rise to incorrect predictions, in particular when the economy is going through a volatile environment with structural breaks and non-linear connections. Many of the studies cover a very limited set of macroeconomic variables and may miss other important factors which can influence GDP dynamics, e.g. political stability or global market conditions... This limited scope constrains the model to only include a few important predictors that might otherwise improve forecasting. However, specific integration of machine learning and deep learning techniques is promising but the literature has fallen short in holistic comparison among traditional and modern approaches. Moreover, several machine learning models are regarded as "black boxes,"[41] and therefore it is challenging to construct and understand the results of these model predictions that in turn has an effect on decision making. Secondly, the economic environment is often not completely forgotten. Forecasting performance is significantly influenced by economic conditions, data quality or specific attributes of the target variable; however few studies control for these factors properly. However, a limited attention is paid to the investigation on ensemble and hybrid models which have a potential of yielding a higher level of predictive accuracy by blending strengths from multiple streams[1]. A broader outlook that includes various economic variables and an emphasis on ease of interpretation, while there is still exploration to be made into the full capabilities from Hybrid models up to ensemble forecasting for GDP.

Chapter 3

Dataset Preparation

The first step in the forecasting project is associated with data preparation. The process of data collection may vary in complexity and might be time-consuming, based on the topics of the forecasting. With an understanding that there is no readily available benchmark data with the related key parameter for GDP forecasting, the increased complexity is apparent. Therefore, the current paper aims to review different strategies for GDP dataset preparation and offer a detailed discussion of the choices made to overcome the complexity and ensure reasonable forecasting.

3.1 Data Collection

The data set used in this paper is collected from one of the most prominent and widely used data set- Penn World table, PWT 10.1. It consists of the data of GDP and other economic predictors of almost of 183 countries of the world. The data collected and used in this paper constitutes of 65-year-ranging data from 1954 to 2019. Thus, the data set discussed in this paper gave an insight of long-run trends and changes regarding global economy. The important economic indicators to consider and determine while forecasting GDP is employment level, population level, Current price GDP, capital,TFP, exchange rate, capital etc. are included in this data set with other predictors which are vitally important for developing appropriate GDP forecasting models and other relevant analysis. These indicators or predictors can enhance our critical understanding of the performance of economies over a period of time. These also serve as a valuable option for determining and measuring the economic performance, health, productivity, and the potential of the economy to reach its pace of growth. Utilizing this type of comprehensive data set, we have obtained greater power that enables us to identify important patterns to decide on. Also, it assists us in learning more about the trends and even makes it possible to make a proper assumption with regards to future cause-and-effect relationships that can lead to or affect the growth of Global GDP and other economic indicators.

3.2 Exploratory Data Analysis

To apply the forecast model on the Global GDP forecasting, first, we need to analyze the dataset and have to conduct some tests around 50 countries to choose the best applicable model. Almost every country has shown similar results. These tests can determine our model's performance and can further help us to maximize the performance of our model. Combination of these tests is known as Exploratory Data Analysis . By combining these, first, we apply a few techniques to understand dataset and then analyze this dataset. These techniques enable to identify complex patterns, relationships, and potential outliers or anomalies within the dataset so that it can analyze these complex patterns and can depict the hidden insights. EDA also helps us to focus on detecting missing data in the dataset so that null values or missing data not impacts our applied model. By conducting EDA test in a dataset, we can be ensured that we are making well-informed decisions in data preprocessing and model fitting. To analyze the dataset, we have applied the following tests on our dataset.

3.2.1 Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Autocorrelation Function measures the correlation of the current values of a time series and the lagged past values at different lags. The formula for the autocorrelation at lag k is:

$$\mathrm{ACF}(k) = \frac{\sum_{t=k+1}^{n}(X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \tag{3.1}$$

where X_t is the value at time t, $\bar{X}$ is the mean of the series, and k is the lag.

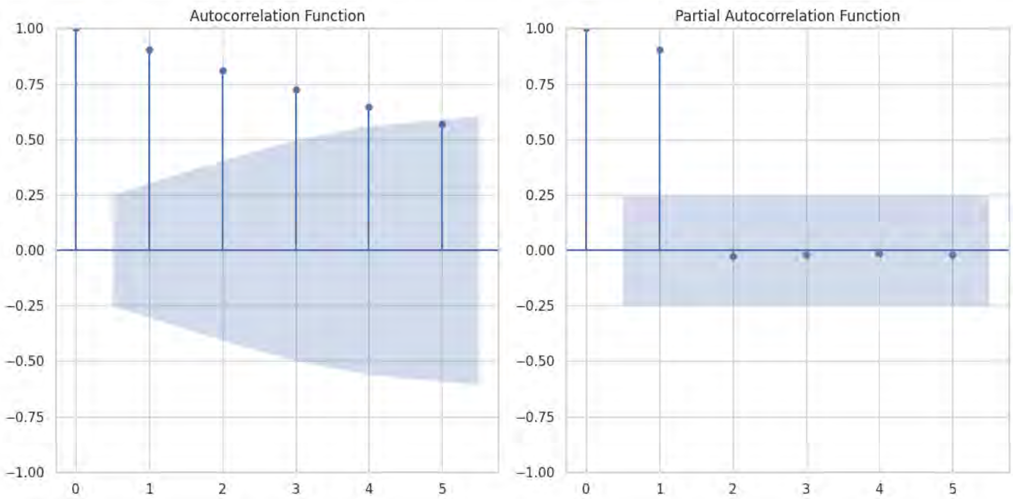

Figure 3.1: ACF and PACF graph of Bangladesh

Partial Autocorrelation Function, however, gives the correlation of a series of values at a fixed lag that is considered to represent their direct correlation after the influence of the intervening lags has been removed. The partial autocorrelation of y at lag (k) can be thought of as the correlation between X_t and X_{t-k}, adjusted for the influence of the lags in between. In the case of PACF, the formula is more complicated, and it is commonly calculated via the method of the least squares, Yule-Walker equations or other similar methods. Both methods serve a similar purpose that is assessing time dependencies.

In Fiqure 3.1 we have applied ACF and PACF to understand time dependencies in our data. ACF tells us how much past values influence what the present value will be, while PACF tells us at what lags this is the direct influence. From the ACF plot, we can observe that there is a strong autocorrelation at the first 4 lags. Therefore, the first 4 years have significant influence with the current value. Furthermore, after the 4 lags the correlation declines meaning that there is relatively weaker influence for the rest of the years. In the PACF plot, there is a significant direct impact at lags 1 and 2. It can therefore be concluded that an autoregressive model of the order 2, AR(2) is appropriate. Finally the slow decay of ACF may mean that the process is not stationary and other methods such as differencing should be employed.

3.2.2 ADF, KPSS, and Phillips-Perron Tests

One important aspect of a time series reality is the fact that it has to be stationary for any proper modeling or forecasting to take place. Here, three major tests are used, the ADF test, where the null hypothesis is the presence of a unit root neglects the stationarity of the series. The KPSS test suggests a null hypothesis must be stationary. Furthermore, the Phillips-Perron test is robust to serial correlation and provides a conclusion to the type of series to work with. All three help choose a proper type of data transformations to help with any future work on time series.


```
ADF Test Results :
{
    "ADF Statistic": 1.5453494670001076,
    "p-value": 0.9976839951468889,
    "Critical Values": {
        "1%": -3.6055648906249997,
        "5%": -2.937069375,
        "10%": -2.606985625
    }
}

KPSS Test Results :
{
    "KPSS Statistic": 0.9135149478641881,
    "p-value": 0.01,
    "Critical Values": {
        "10%": 0.347,
        "5%": 0.463,
        "2.5%": 0.574,
        "1%": 0.739
    }
}

Phillips-Perron Test Results :
{
    "PP Statistic": 2.230480581396122,
    "p-value": 0.9989066622400263,
    "Critical Values": {
        "1%": -3.43,
        "5%": -2.86,
        "10%": -2.57
    }
}
```


Figure 3.2: ADF, KPSS, PP test result of Australia

In Figure 3.2 The p value is 0.9977 and that justifies it is a non stationary series as the null hypothesis of a unit root cannot be rejected. The KPSS test results have a p value of 0.01 and the series is non stationary as the null hypothesis of stationarity is rejected. It is possible that the null hypothesis of a KPSS test is rejected and the null hypothesis of an ADF test is not rejected. The p-value for the Phillip Perron test is 0.9989 and therefore the series is non stationary as well.

3.2.3 Rolling Mean (Sliding Window)

The Rolling Mean, also known as the Sliding Window, computes the average of a dataset for a given number of time periods. The formula for the Rolling Mean at time t with a window size of s is:

$$\text{Rolling Mean}_t = \frac{1}{n} \sum_{i=t-n+1}^{t} X_i \tag{3.2}$$

This method smooths short-term variations and, therefore, makes it easier to identify long-run trends in the data. When forecasting GDP, the Rolling Mean is employed to emphasize general growth patterns while reducing the significance of annual volatility. Averaging over a defined window helps to interpret the continuous behavior of GDP, thus simplifying forecasting processes.

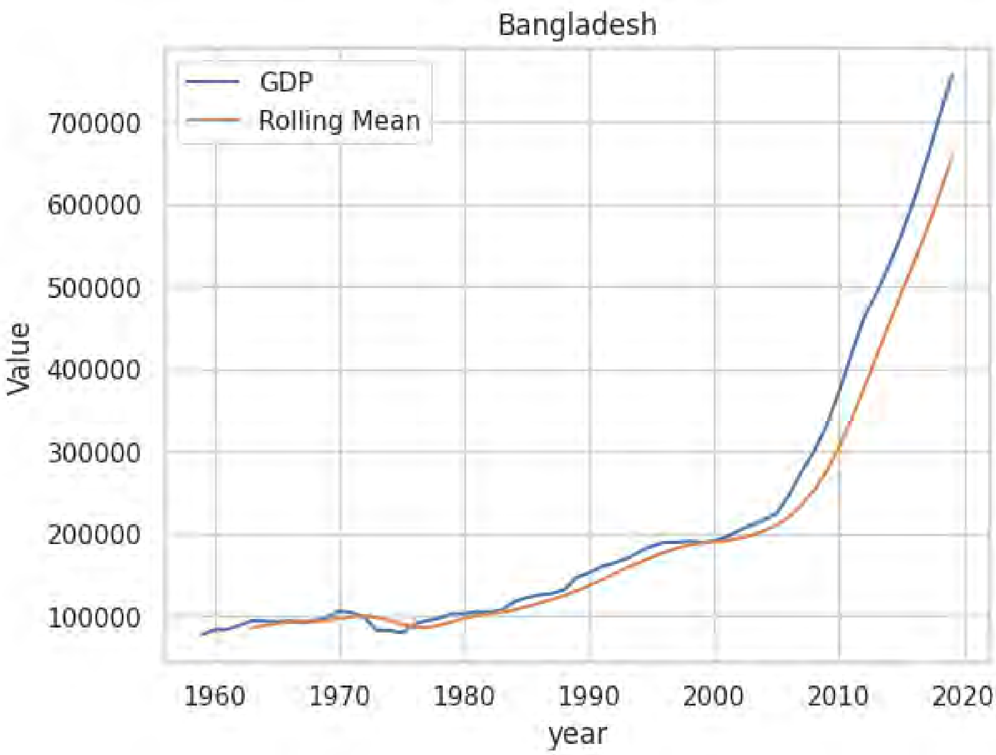

Figure 3.3: Rolling means graph of Bangladesh

From the graph in Figure 3.3, it can be seen that Bangladesh's GDP is Following a clear upward trend from 1960 to 2020 having a significant growth acceleration at the late 1990s and early 2000s. The Rolling Mean allows one to average out short-term fluctuations, permitting to observe more persistent patterns of long-term growth. This reveals that such features as accelerating growth or decelerating growth follow each other over time, and the current forecasting would allow observing this persistent upward trend.

3.2.4 Lag Analysis

Lag Analysis is a technique in which a relationship between a time series and its own past values is studied. It is used to determine how historical data influences the present observations. The formula for the lag k of a time series is as follows:

$$\text{Lagged Value}_t = X_{t-k} \tag{3.3}$$

Where X_t is the present value at time t and k is the number of lag periods. The technique is highly useful in finding trends and interdependencies and deciding which forecasting model to be used. In GDP forecasting, Lag Analysis is used to identify how the past GDP values are related to future economic performances. Thus, the lagged values are used to observe trends and interdependencies which help in deciding the appropriate forecasting model preserving the historical performance.

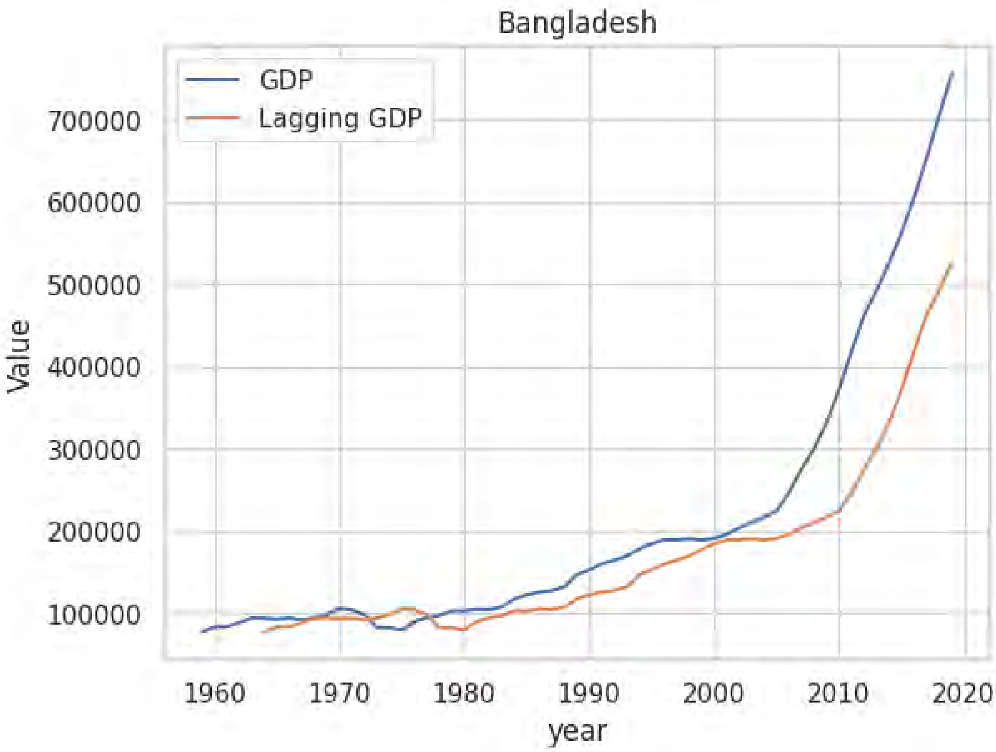

Figure 3.4: Lag graph of Bangladesh

In Figure 3.4 as we seen from the graph, Bangladeshi GDP is plotted alongside its Lagging GDP values. By using data from 1978 to 2019 for model training, no overlap with future values is present, so that the model relies only on the historical data. Key economic trends of the 41 years increase the robustness of forecasting models developed. It also demonstrates how lag analysis helps to make more accurate predictions.

3.2.5 Interquartile Range (IQR)

The IQR is a statistic that helps to measure the distribution of the middle 50% of a dataset. In other words, it is used to reflect the provision of most data. In depth, it is measured as the difference between Q1, and Q3:

$$\text{IQR} = Q3 - Q1 \tag{3.4}$$

Q1 reflects the value that covers a lower 25% of data. Q3 covers the value that reflects a lower 75% of data. The Purpose of the IQR applied in GDP forecasting in relation to its purpose to help measure and clear the data from outliers. This statistical measure helps to take out these extreme values to develop accurate models. In relation to GDP, it points out data that cannot project the real condition of the economy.

Also, we shown in Figure 3.5 that higher IQR mean more variability of countries in their GDP values. We can assume that the IQR is high, which means that the model will not adequately describe the data since it will not be centered on the data, it will not represent all the data. However, having a low IQR may be beneficial if you focus on the IQR, meaning that the data is not too padded. Also, a Middle value can be assumed to be sufficient since in the first case, there are few centers of data points, and the second case has too many centers. Similarly, data with IQR increased may not be suitable for training data. From the central model to the increased one.

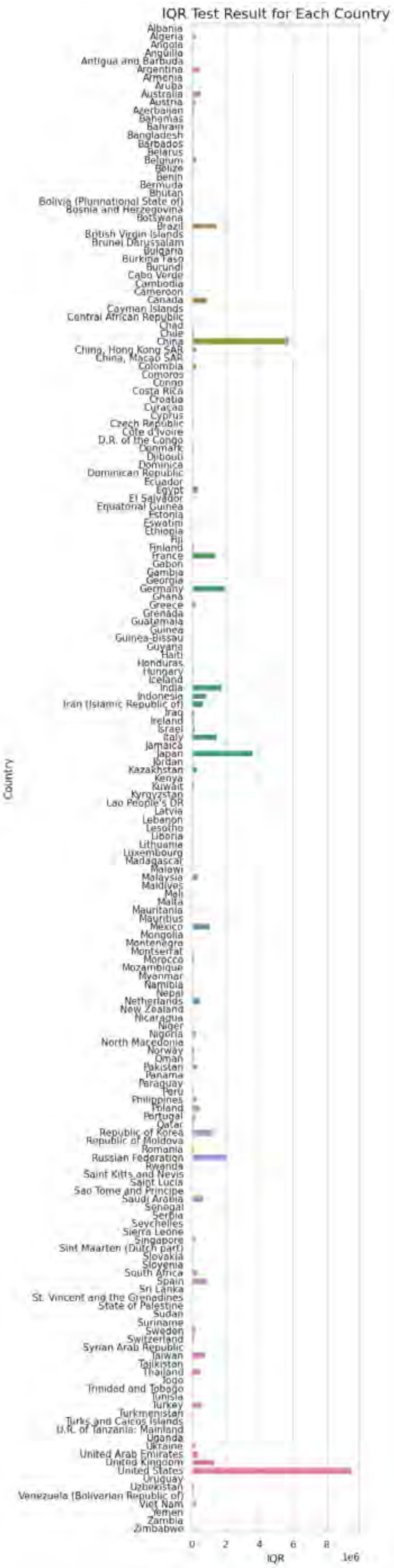

Figure 3.5: IQR result of 183 country

3.2.6 Volatility Test

A volatility test measures how much values of a time series differ within a particular period. Thus, it is designed to demonstrate to what extent the number varies. One of the common methods of calculating the volatility is based on the standard deviation of returns. For a series of returns R_t, the formula of the standard deviation is as follows:

$$\text{Volatility} = \sqrt{\frac{1}{N-1}\sum_{t=1}^{N}(R_t - \bar{R})^2} \tag{3.5}$$

where R_t is the return at time t, R is the average return, and N is the number of observations. Hence, the test is useful as it helps evaluate how much the economic performance is unknown at the moment.

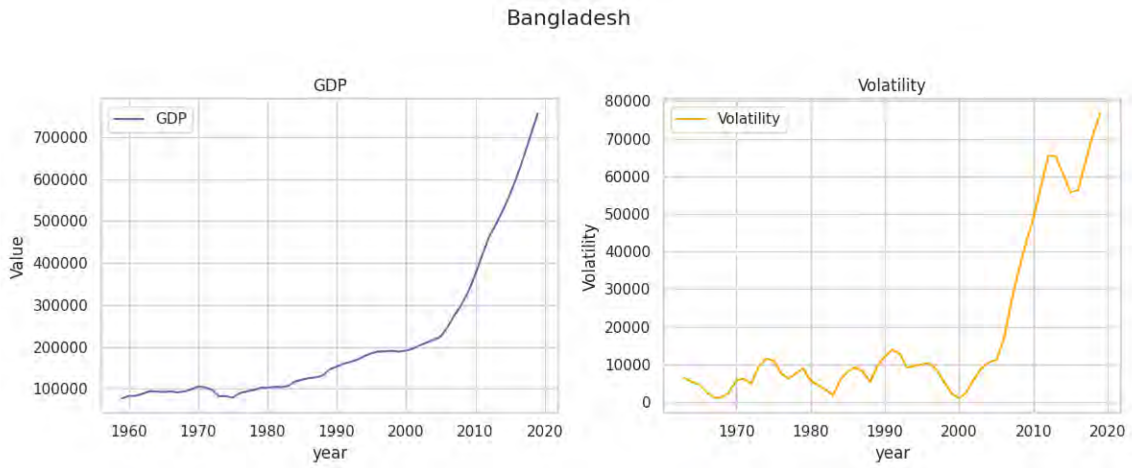

Figure 3.6: Volatility Test of Bangladesh

We shown in Figure 3.6 Bangladesh's GDP, as shown in the graph, indicates upward trends throughout the periods of the 1960-2020 year span. However, during the pandemic, it's shown downward trends. The graph on the right, in its turn, shows volatility. As can be seen, there are periods during Bangladesh's history that have an incredibly large fluctuation. For the application of forecasting algorithms, therefore, it is essential to acknowledge periods when the volatility level increases.

3.2.7 Skewness and Kurtosis Test

The skewness calculation will allow understanding the attitude of the dataset in research to the data value concentration on the sides of the mean:

$$\text{Skewness} = \frac{N}{(N-1)(N-2)}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{s}\right)^3 \tag{3.6}$$

Where:

- N is the number of observations,

- x_i is each individual value,
- $\bar{x}$ is the mean of the dataset,
- s is the standard deviation.

Kurtosis measures the "tailedness" of the distribution, which tells us if the outliers exist. The formula for the Kurtosis is:

$$\text{Kurtosis} = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)} \tag{3.7}$$

Skewness is about the symmetry of the GDP distribution. It helps to determine whether the number of high GDP values is greater than the number of low GDP values. Positive skewness indicates that only a few countries have an extremely high GDP, which affects the current assumptions of the model where most of the data varies. Kurtosis is used to determine the existence of outliers in GDP data. The interpretations of Kurtosis are as follows:

- **Kurtosis > 3:** This indicates heavier tails and more outliers than a normal distribution, known as "leptokurtic".
- **Kurtosis < 3:** This indicates lighter tails and fewer outliers than a normal distribution, known as "platykurtic".
- **Kurtosis = 3:** This indicates a normal distribution with the right shape.

	Country	Skew	Kurtosis
13	Bangladesh	1.736533	2.041297

Figure 3.7: Skewness and Kurtosis Figure of Bangladesh

We shown in Figure 3.7 approximately 1.74 is the positive skewness value. The above finding implies that the distribution is right-skewed. This result suggests that the number of years that have significantly higher values of GDP is low. This implies that the number of years that the GDP has shown a high rise is less. Hence, the information above concludes that the GDP has been gradually increasing over years; it shows a few years that have tremendous growth in the rate of GDP. The kurtosis is approximately 2.04. The GDP is platykurtic because 2.04 is less than 3. The above finding implies that the GDP data has lighter tails compared to a normal distribution. As a result, the presence of fewer extreme values above or below the mean implies that there are fewer outliers hence more stable.

3.2.8 Shapiro-Wilk Test

Shapiro-Wilk Test is a statistical test implying that a given dataset is normally distributed. I think that maybe it is called a "weak" test because under some cases, it may seem weak or uncertain on the results of its tests. Secondly, it may see as the data is slightly not normally distributed, especially in small samples. The value of the test statistic, W is computed using the following formula:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3.8}$$

The null hypothesis for the Shapiro-Wilk test is that the data are normally distributed. A p-value of less than something close to what is currently 0.05 and a typically small W value imply that the null hypothesis can be rejected. In other words, if the p-value is less than 0.05 and the value of W is small, it means, the data is not normally distributed.
Therefore, in GDP forecasting, the Shapiro-Wilk Test is utilized to explore whether the GDP data distribution. That is, whether the data is not normally distributed.

	Country	Statistic	p-value
13	Bangladesh	0.731954	3.150671e-09

Figure 3.8: Weak Shapiro-Wilk Test of Bangladesh

We have shown in Figure 3.8 that the Shapiro-Wilk Test results are statistic = 0.731954 and p-value = 3.150671e-09 for Bangladesh. We have a low test statistic pointing to a significant departure from normality and an extremely small p-value suggests that we can reject the null hypothesis of normal distribution. In other words, GDP data for Bangladesh is not following a normal distribution, and alternative modeling techniques are required for a reliable forecast.

3.3 Data Preprocessing

There were several critical steps in data preprocessing we have followed to prepare our dataset for time-series forecasting.:

1. **Handling Missing Data:**
   - Missing data points were addressed through median imputation as the dataset contains outliers for smaller gaps. while countries with significant missing values at the beginning were excluded.
2. **Differencing and Stationarity Transformation:**
   - The first differencing was performed to deal with non-stationarity::

$$Y'_t = Y_t - Y_{t-1} \tag{3.9}$$

The transformation eliminates the trends in the original data and the models can focus on economic growth patterns without being steered in the wrong direction by long-term trends.

3. **Year-to-Year Growth Ratios:**

   - We converted the data in the GDP column into year-over-year ratios to turn the focus from the absolute values:

   $$\text{Growth Ratio}_t = \frac{\text{GDP}_t}{\text{GDP}_{t-1}} - 1 \tag{3.10}$$

   The data is now more focused on the growth rates.

4. **Rolling Windows for Training:**

   - One of the preprocessing steps was to divide the time series data into rolling windows of 7 years each. The GDP value in the next year was to be predicted based on the 7 years data window. This helps the model to understand over the past 7 year data and helps in understanding the effect of economy for that particular year.

   Training Windows Shape: $(6568, 7)$, Testing Windows Shape: $(2550, 7)$

   The data was splitted into training and testing, First 14 years was kept as testing data.

Chapter 4

Methodology

One of the particular concerns of the forecasting of economic trends is the proper understanding of the inter-relation between economic indicators to serve as an accurate forecasting such as Global GDP or a singular economic indicator is focused on the process of turning data into predictive power in machine learning. It indicates the journey from raw data to results via a series of strategic decisions made along the process. Every decision is made using publicly available data as well as the many hypotheses obtained from statistics and machine learning. The ultimate goal is to create a model capable of accurately forecasting future occurrences, influencing decisions, and uncovering insights inside large data sets. The flowchart Figure 4.1 depicts a multi-stage data science and machine learning approach for building models from time-series data from multi-dimensonal economic indicators. It is separated into three essential phases, which are critical for constructing good prediction models. The first stage is to gather economic indicators data, which is typically maintained in big dataset formats, like .CSV or .XLSX files.

4.1 Workflow

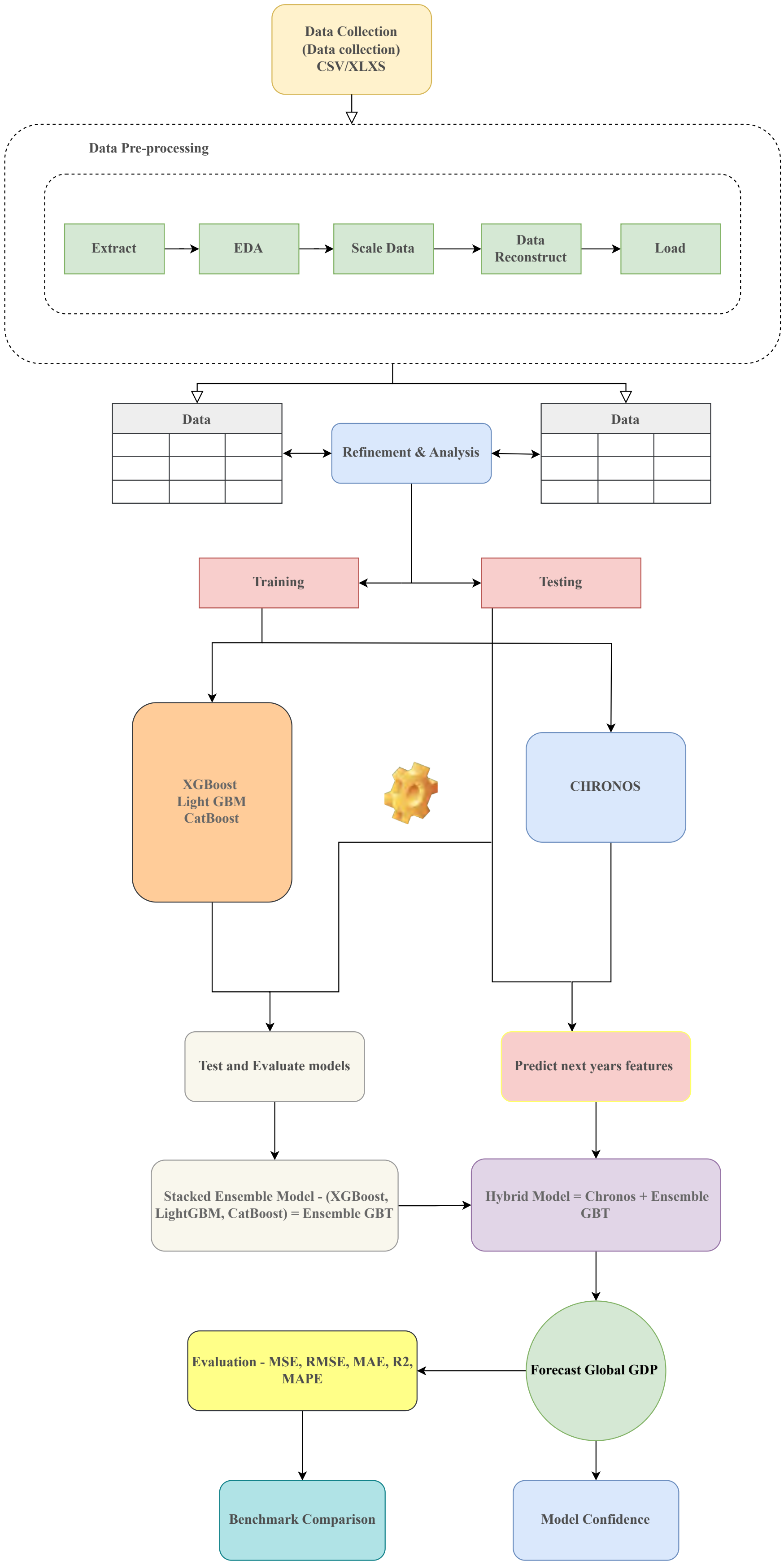

Figure 4.1: Our Workflow

4.2 XGBoost

Gradient Boosting is a strong machine learning tool that can be used in regression and classification tasks. This technique constructs a sequence of weak models (normally, decision trees) in the process of ensemble learning such that each next model tries to fix the drawbacks of the previous one. Below, there is a detailed procedure of gradient boosting in Figure 4.2:

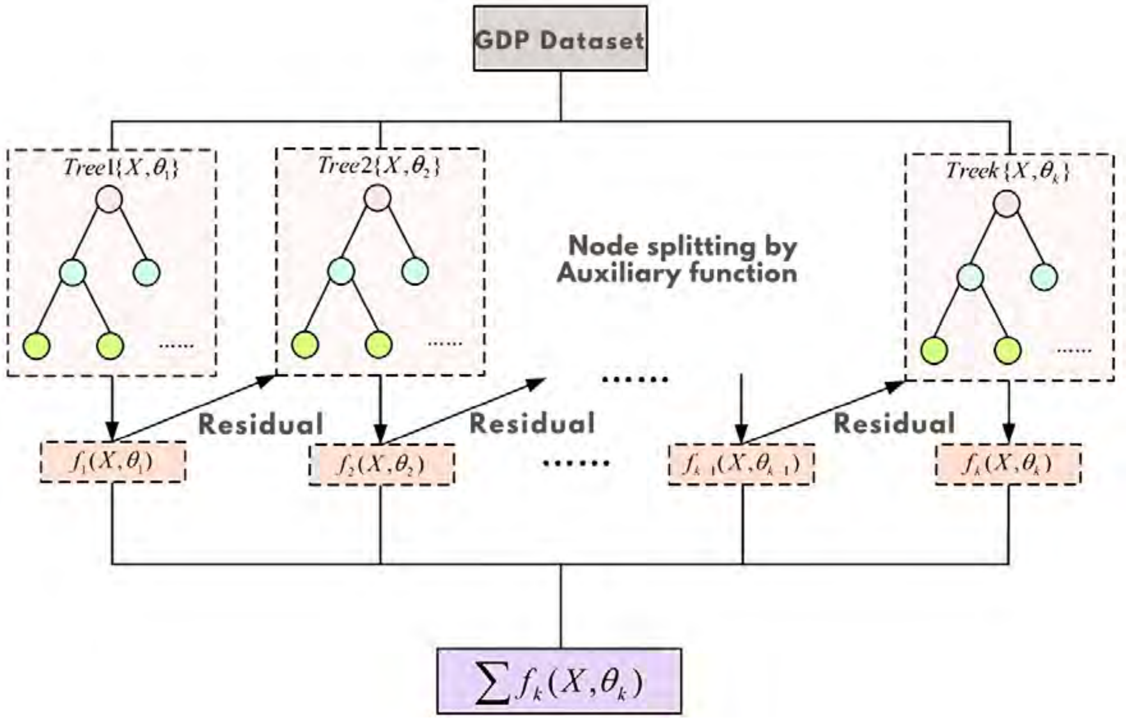

Figure 4.2: XGBoost model workflow

The initial prediction is usually the mean (for regression) or log-odds (for classification):

$$\hat{y}^{(0)} = \arg\min_{\theta} \sum_{i=1}^{n} L(y_i, \theta) \tag{4.1}$$

where $L(y_i, \theta)$ is the loss function (e.g., Mean Squared Error for regression).
At each iteration, residuals are calculated as the negative gradient of the loss function:

$$r_i^{(m)} = -\left[\frac{\partial L(y_i, \hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)}}\right] \tag{4.2}$$

where $r_i^{(m)}$ is the residual at iteration m.
A new model $f_m(X)$ is trained to predict the residuals:

$$f_m(X) = \arg\min_{\theta_m} \sum_{i=1}^{n} \left(r_i^{(m)} - f_m(X_i; \theta_m)\right)^2 \tag{4.3}$$

The predictions are updated by adding the new model's output, scaled by the learning rate η:

$$\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + \eta f_m(X_i) \tag{4.4}$$

where η is the learning rate.
After M iterations, the final prediction is:

$$\hat{y}_i = \hat{y}^{(0)} + \sum_{m=1}^{M} \eta f_m(X_i) \tag{4.5}$$

GDP forecasting typically entails working with non-stationary, non-normal data. For that reason, the use of ARIMA or similar time series forecasting models is not advised due to their assumptions of stationarity and distribution shape. XGBoost, however, does not require stationary, normal data and is designed to efficiently work with complex, non-linear relationships. XGBoost or Extreme Gradient Boosting has been selected for the current task due to its superior performance, efficiency, and ability to process huge amounts of data. This is particularly useful in the case of predicting Gross Domestic Product, as the analysis often requires the use of many economic indicators, such as inflation, trade volumes, or interest rates. The XGBoost algorithm is based on the idea of multiple decision trees, with the trees built using the gradient descent method. The optimization is done by minimizing a custom objective function. One of XGBoost's key advantages is the ability to work on features of mixed dimensions. This is particularly useful in this case, as predicting GDP requires an analysis of complex interactions between many unrelated factors. In addition, unlike the traditional GBDT, XGBoost uses both CART or Classification and Regression Trees and linear classifiers as base classifiers. This allows the algorithm to optimize both non-linear relationships and simple patterns in the economics data. The formula for the XGBoost objective function is as follows:

$$\text{Objective}(t) = \sum_{i=1}^{n} \big(l(y_i, \hat{y}_i^t) + \Omega(f_t) \big) + C \tag{4.6}$$

Where:

- $l(y_i, \hat{y}_i^t)$ is the **loss function** (e.g., squared error for regression tasks).
- $\Omega(f_t)$ is the **regularization term** that penalizes the complexity of the model.
- C is a constant.

The improvement that is characteristic to XGBoost, which contributes to increasing its efficiency lies in the execution of the second-order Taylor expansion of the objective function. This allows making more accurate modifications in each step of the boosting iteration and facilitates the speeding up of the convergence process. XGBoost uses both the gradient and the Hessian, which are the first and the second derivative, for optimization. XGBoost minimizes the loss and adds a regularization function for controlling the complexity of the model, which helps prevent overfitting a problem that frequently occurs when using models in time-series forecasting, such as the one used for forecasting GDP. Therefore, the objective function that is Taylor-expanded can be approximately written as:

$$\text{Objective}(t) \approx \sum_{i=1}^{n} \left(l(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right) + \Omega(f_t) + C \tag{4.7}$$

Where:

- g_i and h_i are the **first and second-order derivatives** (gradient and Hessian) of the loss function with respect to the predictions.
- $f_t(x_i)$ represents the predicted output from the new tree.
- $\Omega(f_t)$ is the **regularization term** that controls model complexity.

Another way to improve the performance of XGBoost is to tune its hyperparameters. By modifying values of such parameters as `n_estimators`, `learning_rate`, and `max_depth`, it is possible to regulate model complexity, increase accuracy, and reduce overfitting. This step is crucial for enhancing the predictive power of XGBoost, as large amounts of data used in GDP forecasting may vary significantly in terms of their nature and order.

4.3 LightGBM

LightGBM is a powerful model that uses a tree-based learning algorithm, which is very efficient, scalable and fast. It is a popular model used in various machine learning tasks, such as regression and classification. It is also a gradient boosting algorithm adapted to effectively work with categorical features in particular, which enables the optimization of performance in large dataset tasks. The description of the LightGBM algorithm along with the architecture is as follows in Figure 4.3:

Continuous features are binned into B discrete bins:

$$\text{Bin}(x_i) = \left\lfloor \frac{B \cdot (x_i - \min(x))}{\max(x) - \min(x)} \right\rfloor \tag{4.8}$$

The split is chosen to minimize the loss function:

$$\arg\min_{s} \sum_{i=1}^{n} \ell(y_i, \hat{y}_i^{(t+1)}) \tag{4.9}$$

The leaf with the highest loss reduction is split:

$$\Delta G_j = G_{\text{parent}} - (G_{\text{left}} + G_{\text{right}}) \tag{4.10}$$

The final prediction is the sum of the predictions from all trees:

$$\hat{y}_i = \sum_{t=1}^{T} f_t(X_i) \tag{4.11}$$

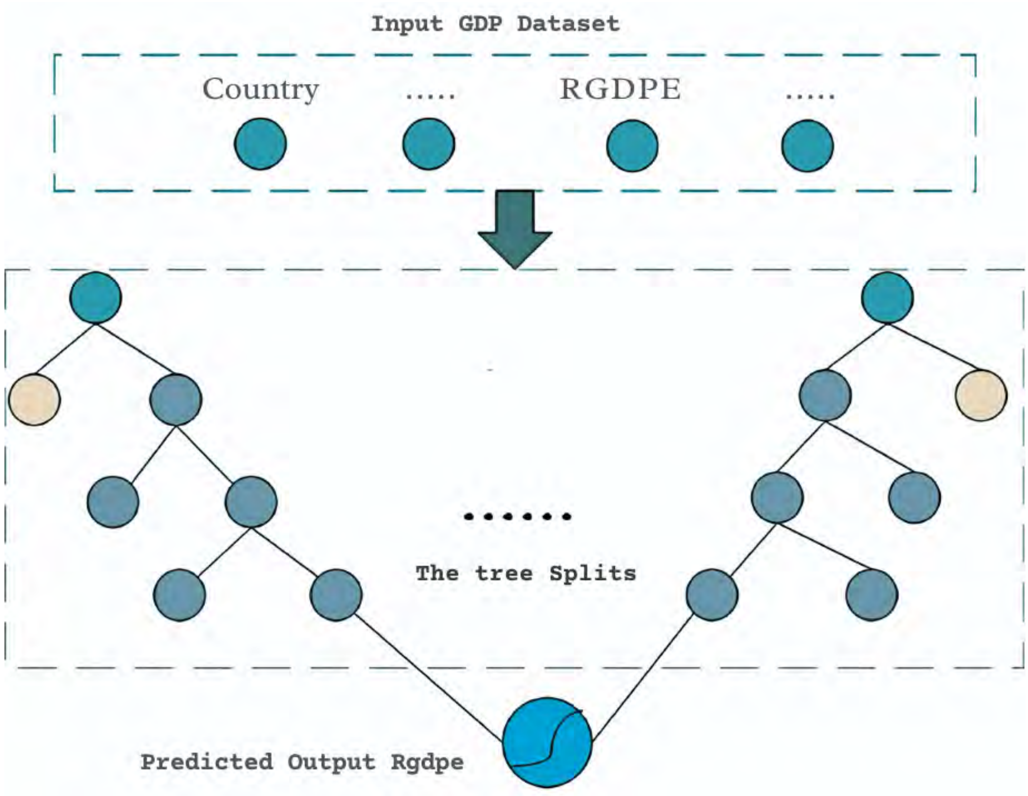

Figure 4.3: LightGBM model workflow

The residuals are calculated as:

$$r_i^{(t)} = y_i - \hat{y}_i^{(t)} \tag{4.12}$$

While additional libraries can be applied, it is advisable to switch to a different option, such as the Light Gradient Boosting Machine (LightGBM), to ensure the accuracy and performance of the GDP forecasting model. LightGBM takes different optimization approaches from the second-order Taylor expansion used in XGBoost. Instead, the implementation focuses on computational efficacy and scalability. Some of its benefits that make it suitable for the present application scenario are:

- **Gradient-based One-Side Sampling (GOSS)**: The GOSS technique aids the model in improving computational efficiency by considering only the most essential data points. Those that are less informative are ignored throughout the process.

- **Exclusive Feature Bundling (EFB)**: As one of the most outstanding aspects of LightGBM, EFB is selected due to the benefit of bundling mutually exclusive features. It also integrates related ones and reduces data dimensionality, increasing memory efficiency and accelerating the training period. This is especially beneficial when working with large volumes of high-dimensional data.

- **Leaf-wise Growth Implementation**: This method involves growing trees by expanding their most important leaves, ultimately allowing the model to focus on capturing deeper patterns. Some of the typical issues in GDP forecasting, such as the expansion of non-linear relationships between economic indicators, can be better addressed using this approach.

LightGBM's histogram-based algorithm discretizes continuous data into bins, allowing it to use less memory and process massive amounts of information effectively at high speeds. As processing extensive economic data is a vital component of successful GDP forecasting, this contributes to the selection of LightGBM for this task. Furthermore, the model can be effectively scaled to large proportions, as it supports distributed training, allowing it to span multiple machines. The software toolset does not experience issues with missing data, providing a good fit for real-world economic information, which can often be incomplete for various reasons.
To make LightGBM more effective in the context of the current study, various hyperparameters, such as `n_estimators`, `max_depth`, and `learning_rate`, can be optimized using Bayesian optimization tools like **HyperOpt**, in order to obtain better results while requiring less time to set up the model.

4.4 CatBoost

CatBoost is also a gradient boosting algorithm particularly adapted to efficiently deal with categorical features, thus optimizing performance in large dataset tasks. The specific features of CatBoost include its treatment of categorical data, use of a continuous variable histogram-based algorithm, and utilization of a symmetric tree structure. A description of this algorithm is presented below, along with the Figure 4.4:

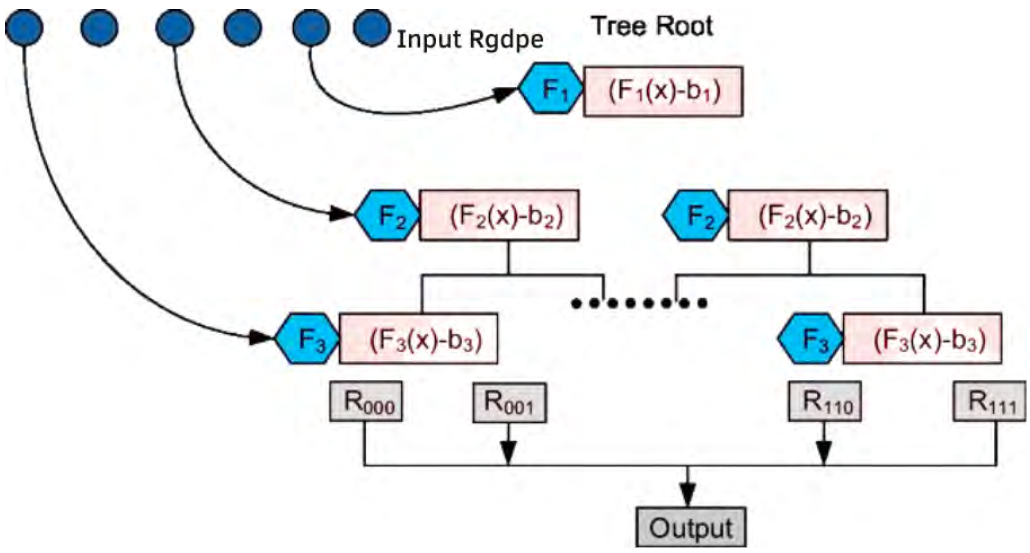

Figure 4.4: CatBoost model workflow

Categorical features are encoded using target-based encoding as follows:

$$\text{Encoding}(x_j) = \frac{\sum_{i=1}^{n} y_i \cdot 1(x_i = x_j)}{\sum_{i=1}^{n} 1(x_i = x_j)} \tag{4.13}$$

At each node in the tree, the decision is made based on the feature $F_j(x)$ and the threshold b_j:

$$F_j(x) - b_j \tag{4.14}$$

The final prediction for data point x_i is the sum of residuals across all trees:

$$\hat{y}_i = \sum_{t=1}^{T} R_{\text{leaf}(x_i)} \tag{4.15}$$

The model minimizes the mean squared error (MSE) during training:

$$L = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.16}$$

Residuals are computed at each iteration t as:

$$r_i^{(t)} = y_i - \hat{y}_i^{(t)} \tag{4.17}$$

The predictions are updated iteratively using the residuals:

$$\hat{y}_i^{(t+1)} = \hat{y}_i^{(t)} + \eta R_{\text{leaf}(x_i)} \tag{4.18}$$

CatBoost is a non-consensual gradient boosting framework that provides robust performance when it comes to dealing with numerical and categorical data simultaneously. Therefore, it may greatly facilitate the task of GDP forecasting since this problem necessarily involves both data types. CatBoost differs from other boosting frameworks because it supports the handling of categorical variables and does not require one-hot encoding. This is done by means of ordered target encoding. The use of this technique significantly reduces the model's complexity and training time. This becomes especially important in the context of economic data, which frequently includes high-cardinality features like user IDs or transaction dates. Furthermore, CatBoost constructs its trees using data-oblivious decision trees, meaning that each leaf in all of the tree's branches at each level has the same form of decision. Therefore, all leaves get the same weight, and the model cannot become unbalanced and too complex on one side. In the kind of time-series task that is GDP forecasting, avoiding overfitting is one of the main challenges, meaning that such an approach can contribute to the model's overall performance.

In CatBoost, the tree-building process is conducted according to an ordered boosting method, as shown in the diagram below. The learning algorithm builds weak learners or decision trees based on distributed feature weights collected from the data during the training process. These weights are optimized, and the strong learner is the result of the combination of all weak learners, guaranteeing stable model predictions over all data samples. Although multiple other gradients programs use this algorithm, the key feature of CatBoost that makes it advantageous over other models is the ability to work with categorical features without their conventional preprocessing, such as one-hot encoding. Specifically, the learning algorithm groups categorical features using target statistics or TS, grouping values by their relatedness in frequency or importance and effectively working with both lower- and higher-cardinality category types. The method is particularly advantageous for large-scale datasets when numerous examples might require increased dimensionality characteristics, and computation power, particularly when dealing with numerous and distinct category features, such as user IDs.

The model also ensures efficient training by applying two types of sampling: Minimal Variance Sampling or MVS and Uniform Sampling. These strategies assure the distribution of balanced datasets, preventing boosting from affecting model overfitting. Notably, this quality is essential regarding the task of high-dimensional economic indicators used for GDP prediction process, and the decision to implement CatBoost as the key model for the task seems rational. The consideration of CatBoost as a prime model for the task of forecasting numerical and category-type GDP values is also justified by the use of oblivious decision trees in combination with symmetric leaf structure and the decision to implement efficient samples that do not demand any extra superstructure. Overall, these qualities ensure that the model decreases the likelihood of overfitting and increases the accuracy of predictions, vital for processing higher-dimensional data present in high-dimensional economic indicators, commonly used for GDP forecasting.

4.5 Ensembled GBT

The ensemble model for GDP forecasting combines powerful gradient boosting techniques, XGBoost, LightGBM, CatBoost, where each contributes its particular strengths to increase the overall prediction accuracy. Using stacking regression, the ensemble model combines the predictive capabilities of these models and allows them to complement each other. It is important to note that XGBoost operates particularly well with complicated, non-linear relationships in GDP forecasting by implementing gradient boosting and the second-order Taylor expansion to generate more representative responses for non-stationary, not normally distributed datasets. Meanwhile, LightGBM is specifically designed for the training and forecasting of large-scale, high-dimensional datasets and introduces advanced techniques that target the acceleration of training, such as the GOSS and EFB. Finally, CatBoost als, in turn, is effective in handling categorical data utilizing the ordered target encoding, which allows the effective processing of mixed data types typical for economic datasets.
During the ensemble, each gradient boosting framework is viewed as a base learner, and the generated predictions are stacked using **Ridge regression** as the final estimator. This technique enabled the model to learn the individual strengths of the base learners and combine their predictions into a more accurate and stable forecast. The **StackingRegressor** facilitated balanced prediction outcomes by effectively minimizing the individual weaknesses of each base learner. Subsequently, the ensemble model was trained on `gdp_windows_train` and tested on `gdp_windows_test`; their results are included in the benchmark dataframe under the label `Ensembled GBT`. By combining the predictions derived from the use of **XGBoost**, **LightGBM**, and **CatBoost**, the ensemble model improves generalization across diverse economic relationships and indicators, providing a potent tool for GDP forecasting.

4.6 Chronos

Chronos is a pre-trained probabilistic model designed for time series data, where it transforms time series values into a predefined set of tokens. This approach enables the fine-tuning of transformer-based language models on time series datasets. A time series forecasting is reframed by this model as a task of predicting the next token. Continuous time series are converted into tokens with the help of scaling and quantization so that familiar models to be applicable without changing architecture.

4.6.1 Time Series Tokenization

The most of the time series data is continuous, as such data cannot be applied to the language model such as T5, Chronos transforms the data of time series into tokens. Tokenization can be divided into two steps that are called scaling and quantization.

Scaling

The time series values can have substantially different values to the extent where they may cause optimization issues. To scale the time series, the mean scaling is used to normalize the values so that they are easier to process. In particular, for an existing time series$x_1, x_2, \ldots, x_C$, with length C, the scaled values $\hat{x}_i$ the scaled values are calculated as:

$$\tilde{x}_i = \frac{x_i}{\frac{1}{C}\sum_{j=1}^{C} |x_j|} \tag{4.19}$$

where μ is the mean and σ is the standard deviation of the time series. This scales the time series into a normalized range, making it easier for the model to process.

Quantization

Now that the time series has been scaled and normalized, it is being quantized into B bins. The uniform binning scheme is employed to assign each value to a particular bin,$q(x)$, whose formula is:

$$q(x) = \begin{cases} 1 & \text{if } -\infty \leq x < b_1 \\ 2 & \text{if } b_1 \leq x < b_2 \\ \vdots & \vdots \\ B & \text{if } b_{B-1} \leq x < \infty \end{cases} \tag{4.20}$$

Such approach results in the sequence of token data that is to be fed into the standard language model, turning time series forecasting into a language modeling task.

4.6.2 Objective Function

The cross-entropy loss function is designed for Chronos training, which is common for existing language models, but modified to the time series forecasting. The model predicts the conditional distribution over the following token, that is the time series values that were calculated after the quantization. The objective function is designed to minimize over the sequence of the tokenized time series data $z_1, z_2, \ldots, z_C$ as follows:

$$\ell(\theta) = -\sum_{h=1}^{H+1} \sum_{i=1}^{|V_{ts}|} 1(z_{C+h} = i) \log p_\theta(z_{C+h} = i \mid z_{1:C+h-1}) \tag{4.21}$$

This function is responsible for comparing the predicted tokenality distribution with the actual tokenized values; hence, it ensures that the model is trained to predict the correct future values.

Chronos Model Architecture

The Chronos framework functions in the following way shown in Fiqure 4.5:

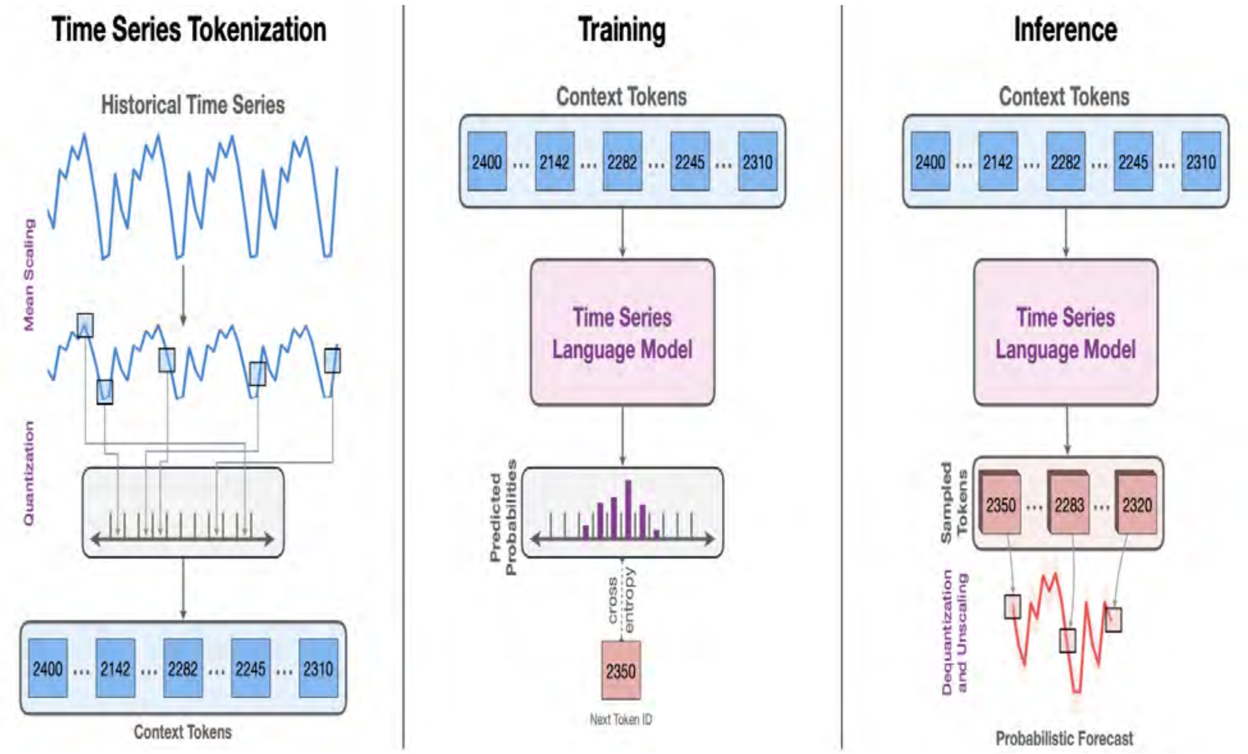

Figure 4.5: Chronos framework functions [47].

Data Preprocessing:

- The raw time series data is scaled and quantized. As such, the token values lie from 0 to Q, and the real values are grouped into tokens based on the bin size. As Chronos should operate with sequences of different lengths and be able to forecast to various steps, two special tokens `PAD` and `EOS` (end of sequence), are also added to the tokenized data. PAD is used for sequences that are shorter than the longest series in a training set, and EOS marks the end of the horizon.

- Chronos relies on the transformer architecture, specifically the encoder-decoder model of the T5 model. A decoder-only model, such as GPT-2, can also be used for this task, as transformers are inherently autoregressive in nature. The only adjustment to the original model is that the vocabulary size is the same as the number of bins used to tokenize the real value.

Training

- Time series data is tokenized and input into the model in batches, the output of which is a distribution over all the bins to which the real value may belong. The task for the model is to predict the most likely value for the next step.

Inference

- When generating samples or making forecasts, the model samples the token from the distribution over the classified values, and these tokens are dequantized into real values. One forecast is generated by running the model multiple times, the result of which is interpretable as multiple paths or scenarios for probabilistic forecast.

4.6.3 Chronos T5 Architecture

In Figure 4.6 introduces an architecture for the afore-mentioned tasks, namely T5-Efficient-SMALL (Deep-Narrow version), aiming to properly combine efficiency with performance in time series forecasting using Chronos framework. It is one of the T5 models, but designed with a "deep-narrow" form. This means that even though the model is deep (many layers) each layer has a small number of parameters(narrow). This architecture is designed to trade-off between depth and computational cost in order to remain computationally efficient, while still having a deep enough network that can learn long-term temporal dependencies. T5-Efficient-SMALL has only 46 million parameters and is ideal for scenarios where memory or processing power limits are a concern.

So in the case of Chronos, we use data scaling and quantization to tokenize real time-series data. It continuously reduce the real-valued data points & mapping them to discrete tokens, for learn-able time-series "vocabulary", just like natural language models tokenizing words. T5-Efficient-SMALL models are characteristically limited in the vocabulary size, which is usually between 1024 and 8192 tokens depending on how much accuracy is needed for a given forecasting task. This yields a tunable volume of gradation that can be used to fine-tune the model for different time-series predictions and allows it generalization across different datasets. The model supports different context lengths (how far into the past you look for information) In SMALL, you would use context lengths of 512, 1024, or 2048. Larger context lengths allow the model to learn broader patterns in the data, making it more accurate both for in-domain forecasting (data as its seen during training) and zero-shot prediction where models predict on new after never having been fine-tuned on any specific task. The superior zero-shot forecasting capability of Chronos means that the T5-Efficient-SMALL model is a good choice for practical use cases where it can encounter new, out-of-sample time-series data. What underlies T5-Efficient-SMALL, however, is

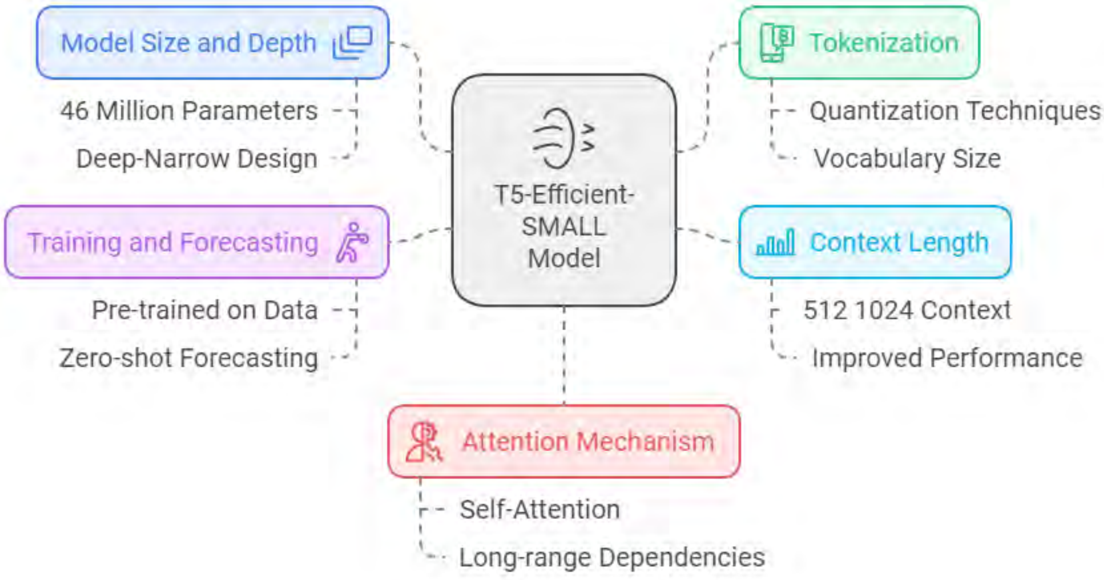

Figure 4.6: Chronos T5 Architecture

the same strong self-attention mechanism as the original transformer architecture in T5. The advantage of this is that the model also gets to learn about how different time steps relate in a series, which are important when looking at long-range dependencies. Just as a language model predicts the next word in a sentence, this time series forecasting deep learning AI can look at those dependencies and forecast future values from the previous patterns. Through this attention mechanism and efficient design, T5-Efficient-SMALL model can be adapted to process large-scale different time series datasets for extensive computational demands.

4.6.4 Data Augmentation and Synthetic Data Generation

Chronos also relies on advanced data augmentation techniques to mitigate the lack of high-quality, small time series data. Specifically, two innovative strategies have been proposed to that effect:

- **TSMixup:** In Figure 4.7 the approach involves the random sampling of time series across different time series datasets. Synthetic time series can then be created as time-weighted combinations of the samples, also known as convex hulls.
- **KernelSynth:** In Figure 4.8 a sophisticated approach based on the use of Gaussian processes to generate syntactic time series. Specifically, synthetic time series are developed with the help of random kernel methods at each of the considered iterations, or through the consideration of novel data combinations to generate ultimate combinations of extreme time series.

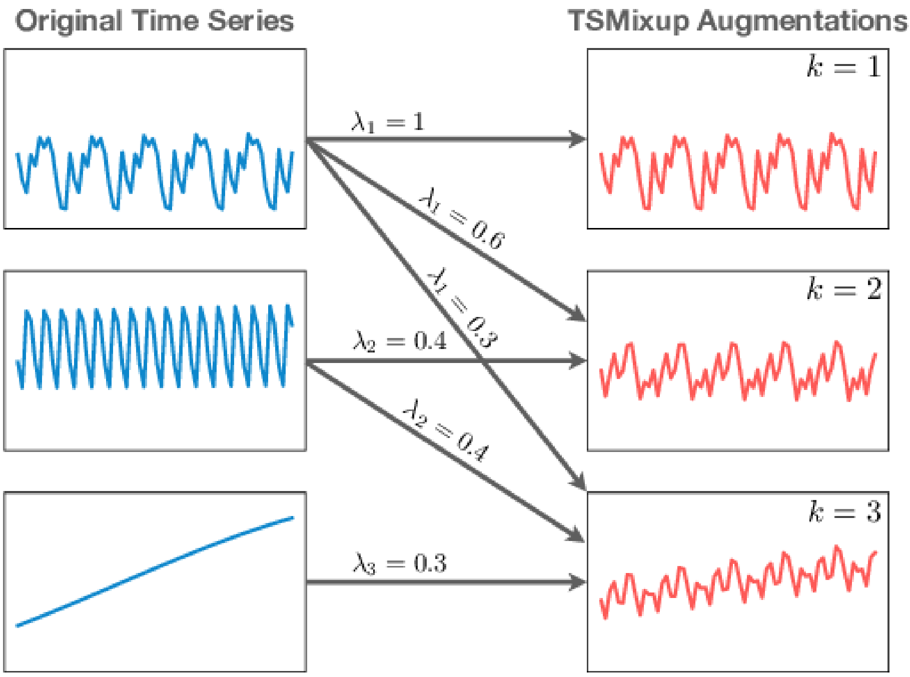

Figure 4.7: Randomly-sampled time series from different datasets by TSMixup [47].

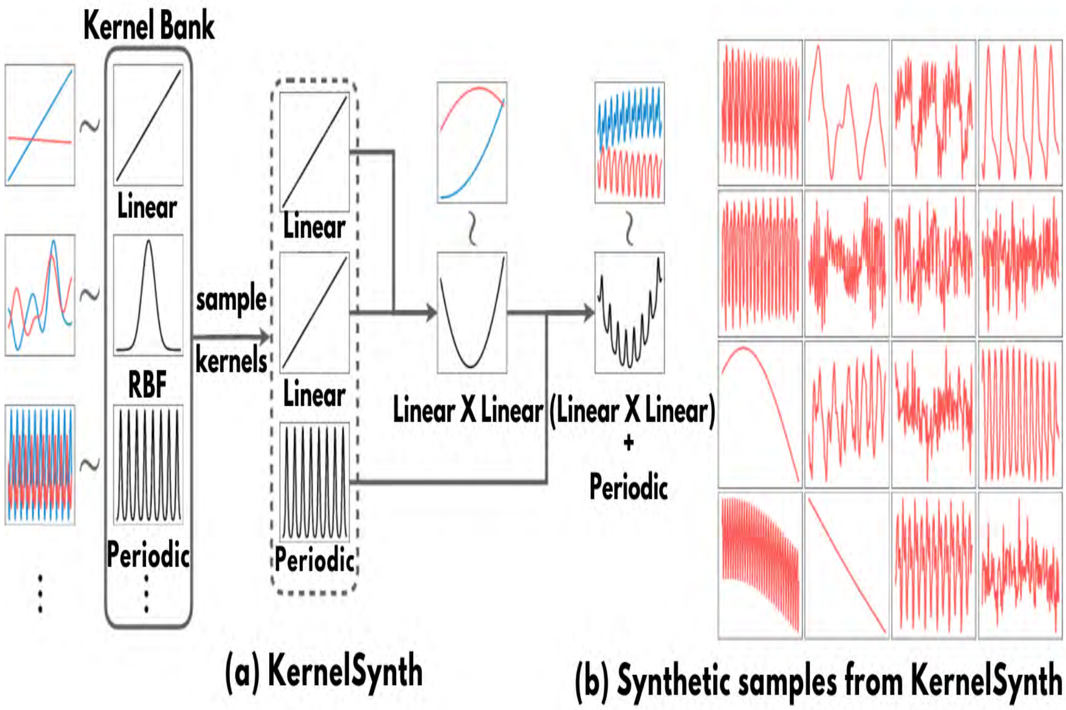

Figure 4.8: Synthetic time series generated by KernelSynth [47].

4.6.5 Zero-Shot Forecasting

During inference, the model uses its learned representations from the training data to predict time series values for new, unseen datasets without fine-tuning. The predictions are made by sampling from the learned probability distribution:

$$p_\theta(z_{C+h+1} \mid z_{1:C+h}) \tag{4.22}$$

The sampled tokens are then converted back to real values using a dequantization function:

$$d(j) = c_j \tag{4.23}$$

where j is the token index, and c_j represents the bin center for that token.

By training on a diverse set of time series, including synthetic data generated via Gaussian processes and other augmentation techniques, Chronos learns to recognize patterns across domains. This robust training allows the model to generalize and accurately perform zero-shot learning on new time series data by applying the patterns it has previously learned.

4.7 Chronos x Ensembled Gradient Boosting Trees (GBT)

In this article, we present a hybrid forecasting model that combines the Chronos time series forecasting model with an ensembled Gradient Boosting Trees approach. This hybrid approach integrates the advantages of both probabilistic forecasting supported by the Chronos method and the deterministic ensemble learning supported by XGBoost, LightGBM, and CatBoost.

4.7.1 Combining Chronos and GBT

In our hybrid model, the Chronos forecast has been weighted more for two reasons: first, it can capture uncertainty from GDP data and leverage it to estimate long-term dependencies, and second, it can capture uncertainty to provide GDP predictions for all dates. As a result, the deterministic GBT ensemble was used as an addition to adding extra fine-tuned forecasting accuracy. For that reason, the final hybrid prediction can be defined as:

$$\hat{y} = 0.9 \times \hat{y}_{\text{Chronos}} + 0.1 \times \hat{y}_{\text{GBT Ensemble}} \tag{4.24}$$

This hybrid forecast employs the advantages of both chronological and GBT forecasting, effectively providing a robust and interpretable model.

4.8 Implementation Process

The Chronos x Ensembled GBT hybrid model is implemented in the following carefully structured process to forecast global GDP. The steps explain how the code works to implement the model throughout this project.

4.8.1 Setting Up the Environment

In this step, installing the required libraries is paramount. The first part of code provisions the appropriate libraries necessary for the implementation. Python programming language and its package installer were used to install the appropriate packages, which include Chronos, XGBoost, LightGBM, and CatBoost.

- **Chronos:** It is a Python library that was used to forecast future data points in a period for a time series.
- **XGBoost, LightGBM, CatBoost:** These are Python libraries that represent machine learning models. They are vital as they use decision trees to predict the GDP data value for a certain year based on past data.

4.8.2 Loading and Preparing the Data

Next, we have defined the data that was loaded. The data on GDP was downloaded and loaded as a Pandas DataFrame. The data contains columns: `year`, `country`, and real GDP or `rgdpe`. Then, I reshape the data so it's consistent with the model I will apply, which requires rows representing years and columns representing countries.

4.8.3 Preprocessing the Data

Before passing the data into models, we need to organize it. Typically, we need to take "rolling windows" of our GDP data. We will then use a sequence of years, say 7 years, to predict the GDP for the 8th year. We implement this using the method below:
Defining `create_time_series_windows()` function: This method divides the data into smaller windows of 7 years for each country. These windows are then used to teach the model how to predict future values.

4.8.4 Building and Training the Models

The core of our implementation is the use of both Chronos and a collection of Gradient Boosting Trees models.

Ensemble GBT Models

We add traditional models, such as XGBoost, LightGBM, and CatBoost. These models are good at understanding complex patterns in the data since they use a number of decision trees. The outputs of all the models are then combined in a stacked ensemble. In this ensemble, all the models work together to make better predictions.

Chronos Model

Chronos is uses the most advanced techniques such as transformers in T5 architecture, forecasting future GDP values. The advantage of using Chronos is that other models are only able to make a single prediction in the future. This means that there is no information on the amount of prediction. However, with Chronos, we can predict multiple future values. More importantly, Chronos is also able to predict the future and provide different possibilities. This is essential because, by looking at the past data, we can see that there is a lot of uncertainty in the accuracy of the forecasts.

4.8.5 Combining Chronos and GBT Models (Hybrid Model)

The hybrid model of Chronos x ensembled GBT is the best of both worlds. Chronos predicts the long-term trend and also provides prediction uncertainty. Ensembles of XGBoost, LightGBM, and CatBoost are more accurate in their short-term predictions. When combining these two models, in the final prediction, we give more weight to Chronos, which is 90%. This makes sense because it handles the uncertainty better. We also combine GBT models, which contribute 10% of the prediction.

Chapter 5

Result and Analysis

5.1 Performance Metrics

In this section, we will check the performance of our forecasting models by 5 well-known metrics, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) Mean Absolute Percentage Error(MAPE) and R-squared score. These metrics give you a complete picture of the accuracy and confidence level with different ways to measure error.

5.1.1 Mean Squared Error (MSE)

The average squared difference between predicted and actual values are quantified by the Mean Squared Error (MSE); For the calculation, we use this formula:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5.1}$$

Where y_i are the actual values, $\hat{y}_i$ are the predicted values, and n is the number of data points.

The sum of the differences between predicted and observed outcomes is shown to you by MSE, this with lower value means a better prediction.

5.1.2 Root Mean Squared Error (RMSE)

Root Mean Squared Error is the square root of the MSE and provides the error in the same unit. In mathematics, the value of RMSE is calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5.2}$$

RMSE mainly helps in emphasizing the larger errors by squaring the values. It is mainly useful to know how the model predicts the value in error.

5.1.3 Mean Absolute Error (MAE)

Mean Absolute Error, which is one where it is contained by the average absolute differences between the predicted value and the real value. The formula for this are as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{5.3}$$

This tells us about the average error or difference and the lower the value, the better for the model will be more precise for the particular value.

5.1.4 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error, which is similar to the MAE but represented in percentage. It is calculated as shown below:

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5.4}$$

This helps in comparing the prediction of different data wherein the lower it is the better for it indicates higher accuracy.

5.1.5 R-squared ($\mathbf{R^2}$) Score

The R-squared Score indicates what proportion of the variance in the dependent variable can be predictable from the independent variable. It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5.5}$$

Where $\bar{y}$ is the mean of the actual values.

R^2 values that are closer to 1 suggests that it may be a better fit; in the sense that a higher proportion of the variance can be accounted for by the model. The summary statement may be stated as follows: These scores "captures the quadratic loss, absolute loss and proportion of the actual values that our trained models are able to predict, as well the efficient of the fitted model".

The considered metrics give a full explanatory information about the models' performances in terms of squaring the error, the absolute error, the percentage error, and the general fit to the data.

5.2 Performance Analysis

Performance of the models was measured with some key statistical metrics — Mean Squared Error (MSE), Root Mean Squared Error(RMSE),Mean Absolute error(MAE), R-squared (R^2), and Mean Percentage Absolute error. These are metrics we look at to get an overall understanding of how well the models perform in terms of providing accurate, reliable and consistent predictions. The following graphs depict GDP predictions for various countries using different machine learning models, including XGBoost, LightGBM (LGBM), CatBoost (CAT), and Chronos, as well as a hybrid model that combines the ensemble predictions (XGBoost, LGBM, CAT) with Chronos. Each model uses actual data from 2013 to 2019 (solid blue lines) and forecasts GDP from 2020 to 2026 (dashed pink lines). The 2020-2021 period reflects the economic downturn caused by the COVID-19 pandemic, with most countries experiencing a dip followed by recovery in the subsequent years.While each individual model predicts post-pandemic growth with some variation in accuracy and fluctuations, the hybrid model strikes a balance by leveraging the long-term trend accuracy of Chronos and the micro-adjustments from the ensemble models. This results in smoother predictions that reduce the volatility seen in the pandemic years, while still capturing the broader economic trends. Some deviations in predictions, particularly during the pandemic period, may still arise due to the inherent challenges of modeling such complex global events, but the hybrid approach improves the overall stability and reliability of the forecasts. The performance of each model is described as follows in terms of these metrics.

5.2.1 XGBoost

XGBoost had good performance in general, it is able to generate lower as well with an MSE of 3.452147e+11, RMSE: 587549.70 and MAE: 94685.68 The R2 value of 0.9248 denotes the model being valid for more than 93% in predicting GDP, which is fairly good enough shape/model for this kind forecasting (i.e., economic) work. Also, with MAPE of 0.0797 we have accuracy plus a good prediction percentage error (low). Indeed, the capacity of XGBoost to work with large quantity structured data together with its ability for regularization allows while providing at the same time good predictive power makes this happen. XGBoost has proven to be competitive with respect to competitor models and apredictability-versus-generalizalility trade-off, making it a top contender for accurate GDP predictions.

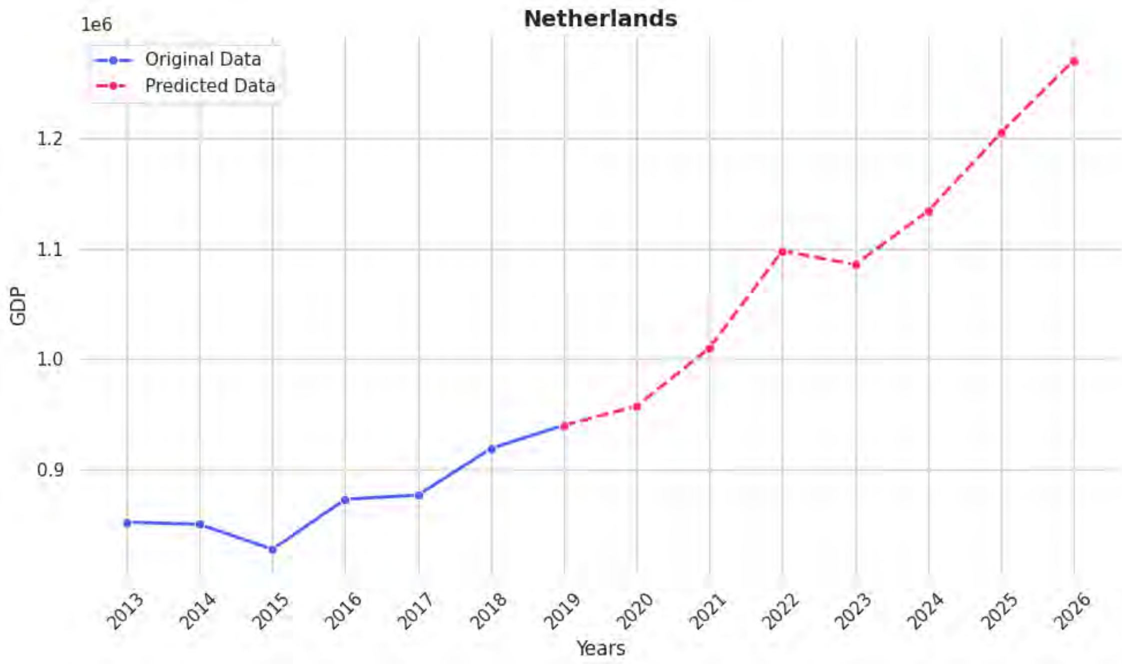

Figure 5.1: GDP Forecast for Netherlands

Figure 5.2: GDP Forecast for Australia

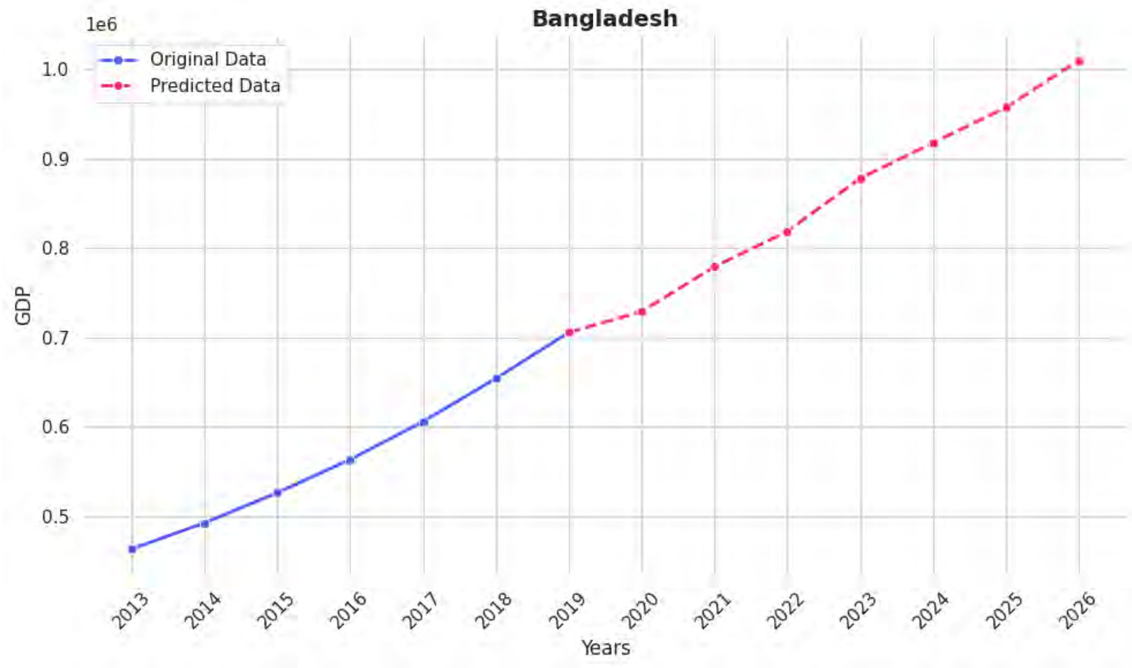

Figure 5.3: GDP Forecast for Bangladesh

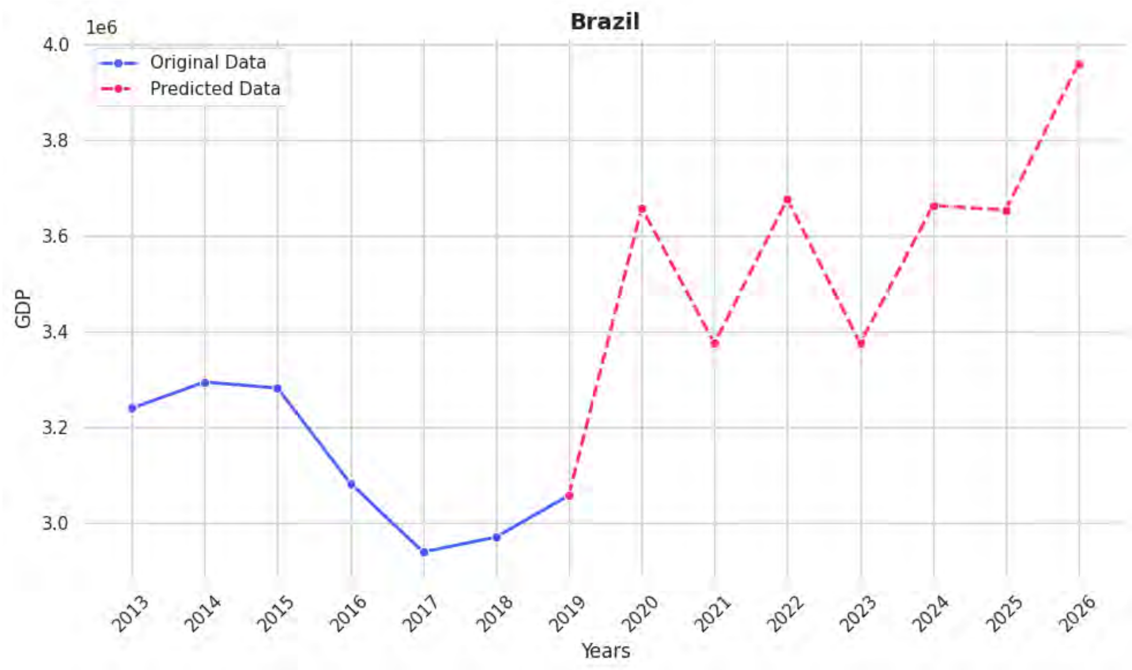

Figure 5.4: GDP Forecast for Brazil

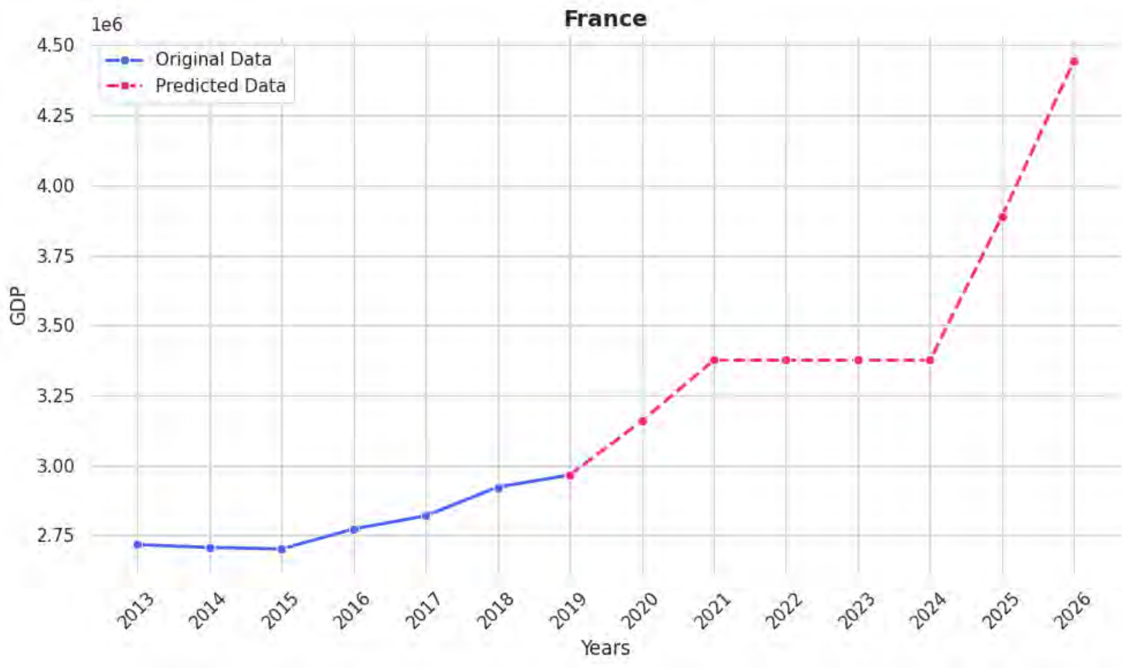

Figure 5.5: GDP Forecast for France

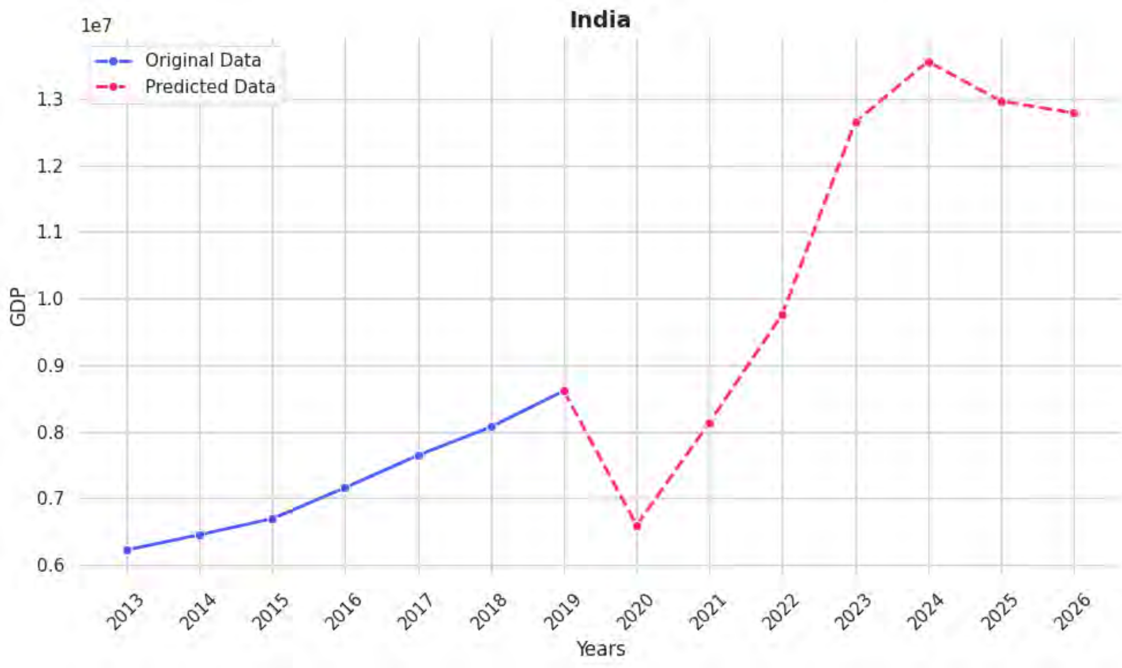

Figure 5.6: GDP Forecast for India

Using the XGBoost model, India, Brazil, Italy, and France experienced a decline in GDP during the 2020-2021 period due to the pandemic, followed by a predicted recovery starting from 2022. Countries like Australia and Bangladesh show only slight dips, while Ireland's growth appears largely unaffected, which could reflect economic resilience or potential model limitations. While XGBoost generally predicts post-pandemic recovery and growth accurately, some deviations during the pandemic years may arise from difficulties in fully capturing the complex economic impacts.

5.2.2 LightGBM

The results of LightGBM were almost identical to XGBoost, showing that is also has the same level of ability with MSE 3.140584e+11 — RMSE: 560409.10 — MAE: 86726.85 The R2 value of 0.9316 too indicates that LightGBM can effectively capture the GDP fluctuations, syndicateing it with XG Boost properties Due to the histogram-based method of decision splitting in this model, it is able to compute really well and is also highly efficient with large-scale datasets which make it a very good option for economic data where there are lots of features. LightGBM achieved also a good MAPE of 0.0781, that corroborate LightGBMs performance in it is acuracy (MAE as target). This post serves as a reminder of LightGBM's prowess: its remarkable impromptu, quick reaction to huge datasets consisting thousands and tens of thousand features in many dimensional problem yet still managing relatively successful prediction. It is the less well-known AdaBoost but competitive with XGBoost on speed, memory usage and often slightly better performance.

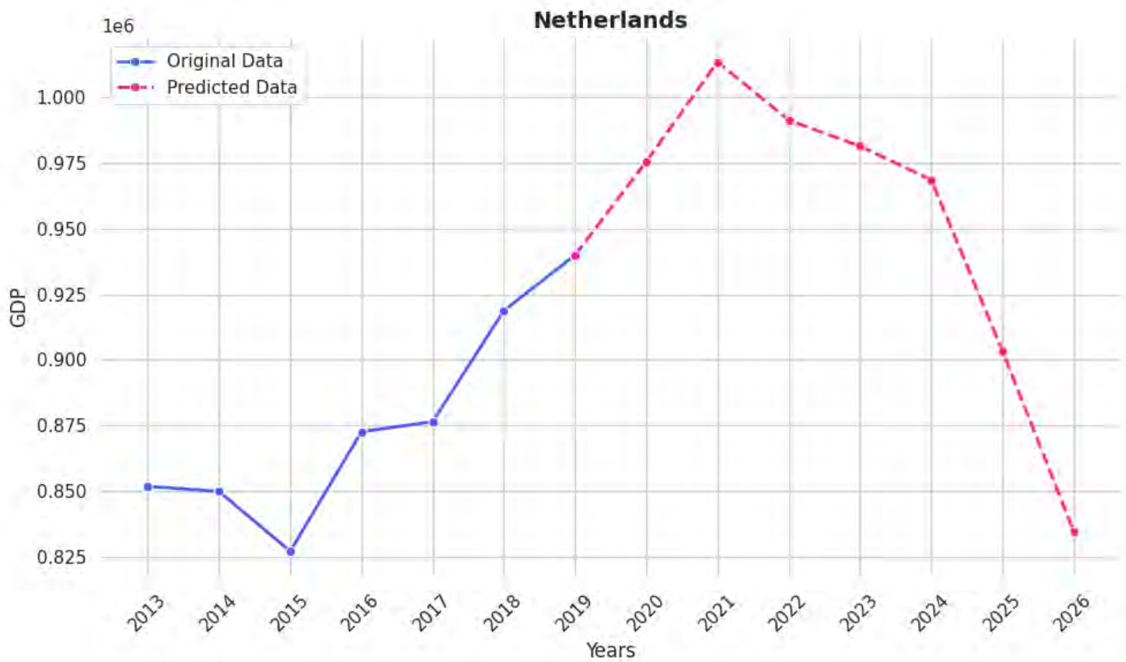

Figure 5.7: GDP Forecast for Netherlands

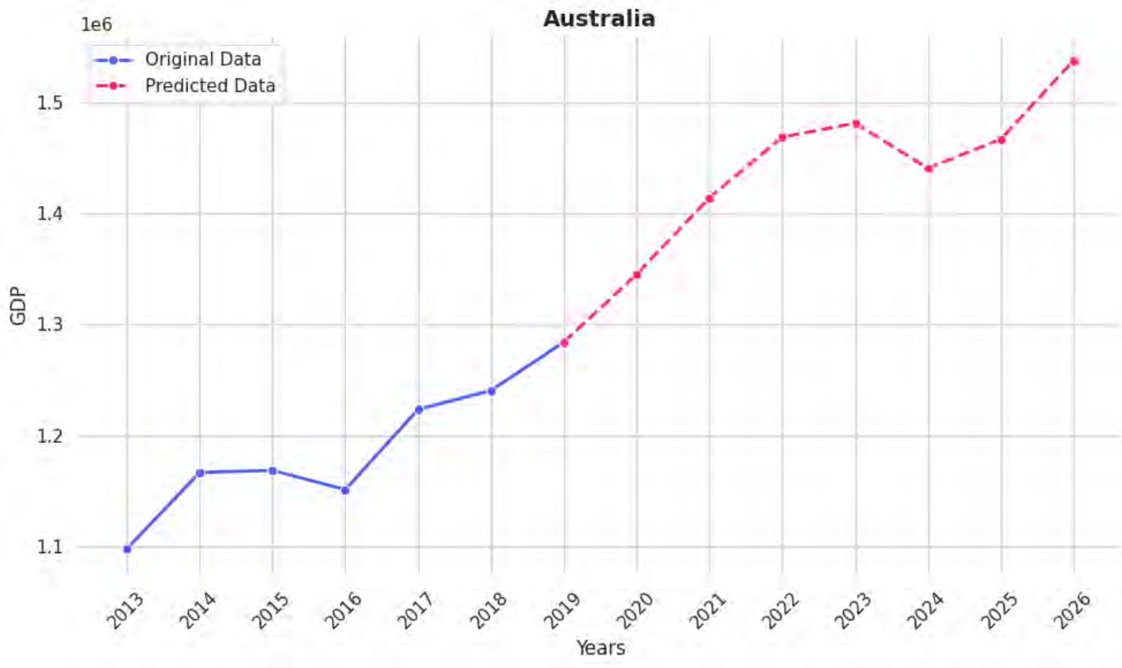

Figure 5.8: GDP Forecast for Australia

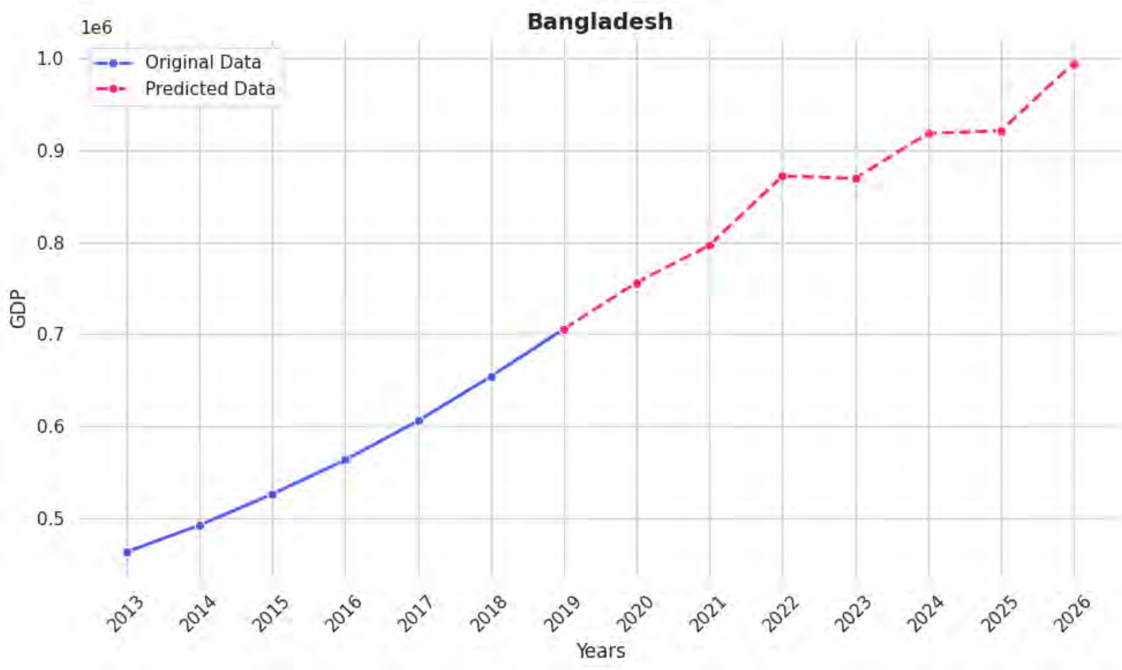

Figure 5.9: GDP Forecast for Bangladesh

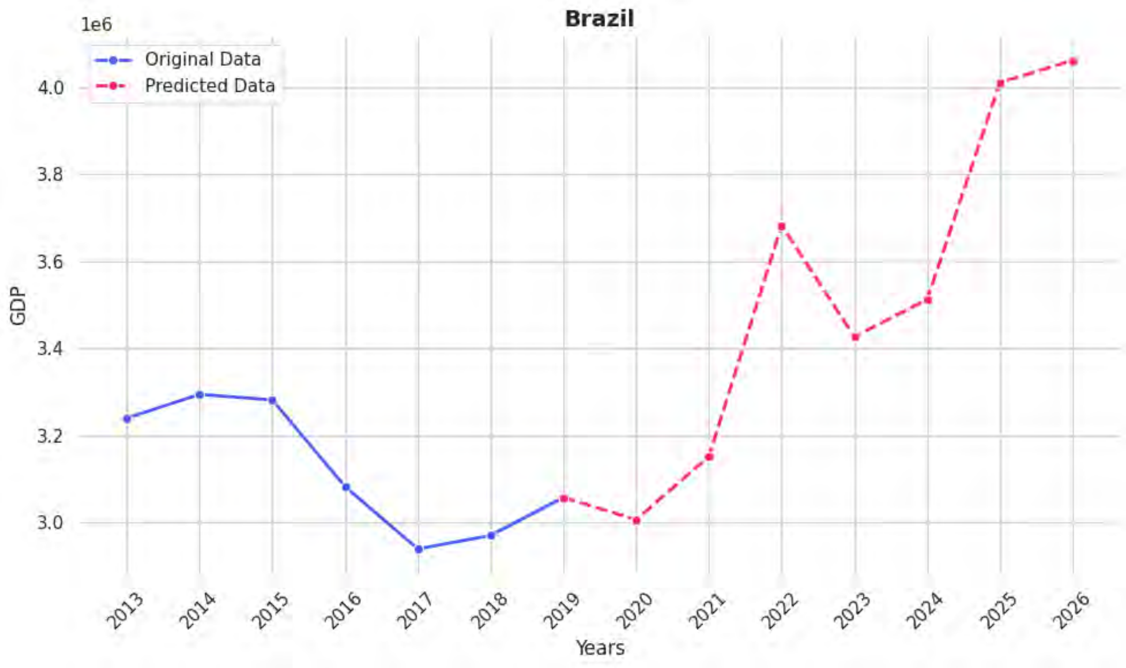

Figure 5.10: GDP Forecast for Brazil

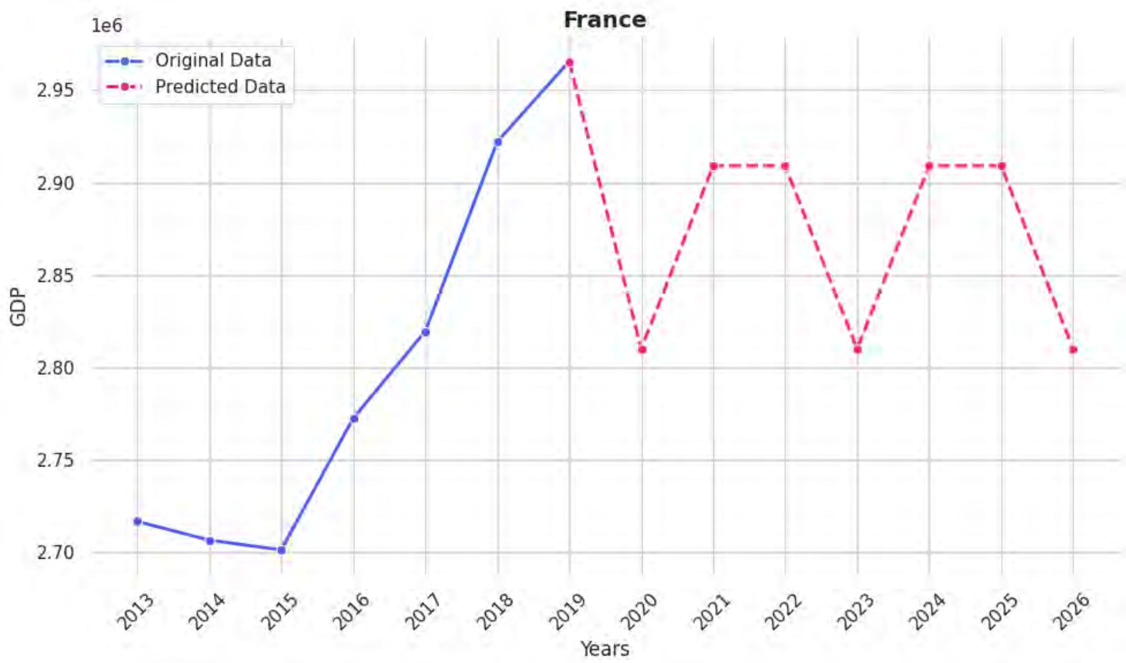

Figure 5.11: GDP Forecast for France

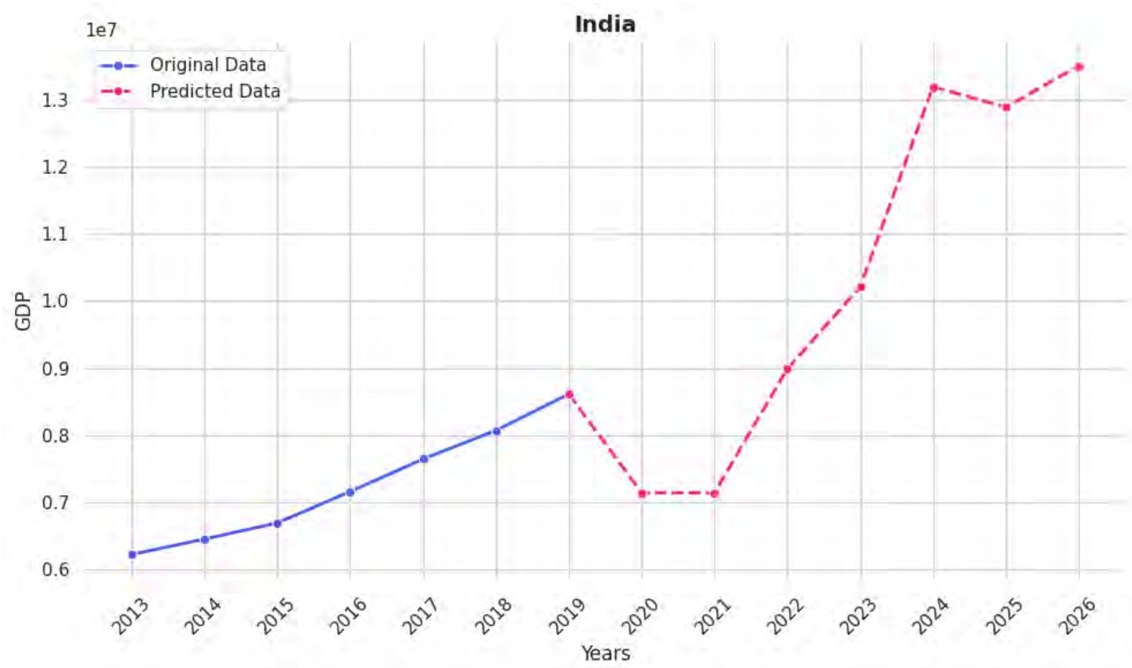

Figure 5.12: GDP Forecast for India

Using the LightGBM (LGBM) model, India, Brazil, Italy, and France experienced a decline in GDP during the 2020-2021 period due to the economic impact of COVID-19, followed by a predicted recovery starting in 2022. Countries like Australia and Bangladesh show only slight dips, while Ireland's growth remains largely unaffected, possibly reflecting resilience or potential model limitations. Overall, the predictions indicate post-pandemic recovery and growth, though deviations in countries like France and Brazil suggest economic volatility during the pandemic years.

5.2.3 CatBoost

CatBoost scored far better in the way of dealing with categorical features, it generated an MSE of 3.385161e+11 RMSE: 581821.35 and MAE: 92154.63 as well Although its MSE and RMSE surpass those XGBoost or LightGBM (by a small margin), the R^2 value of 0.9263 is still pretty good, competing with the best individual models so far in this competition). This they say is an indicator of how well CatBoost can explain complex relations in data. The way CatBoost natively manages categorical data and does not require preprocessing steps such as one-hot encoding makes it suitable for robust real-life economic forecasting. The MAPE of 0.1307 indicates certain level of predicitve error spread, however CatBoost still is a strong GDP forecasting technique by delivering consistent performance in capturing economic trends without categorical data intervention sustenance.

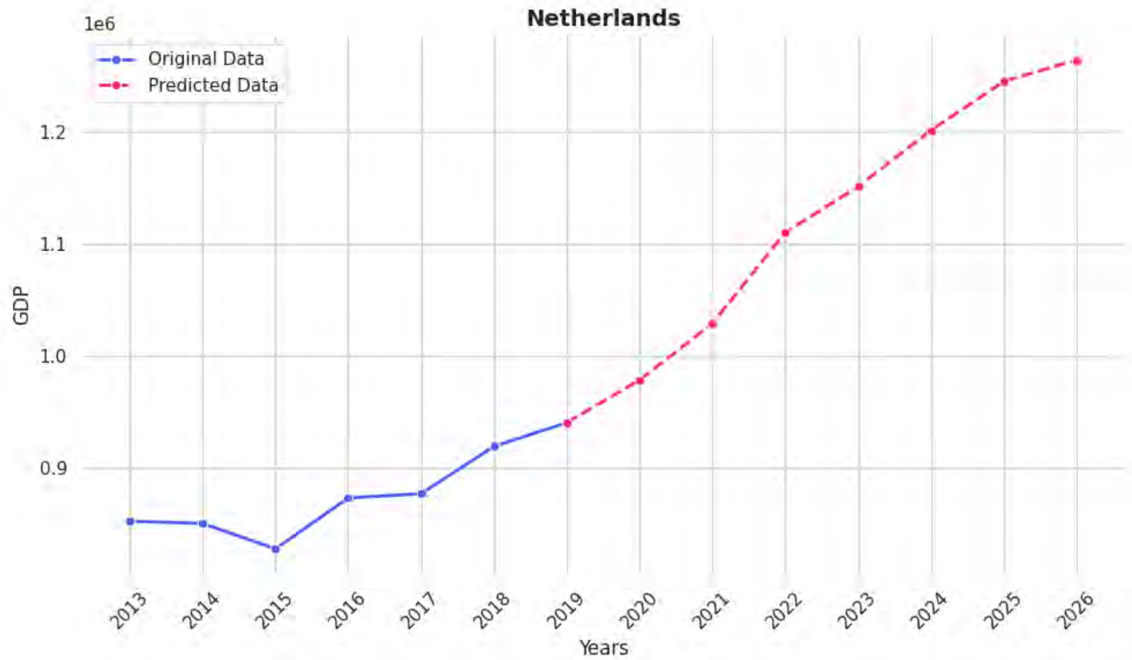

Figure 5.13: GDP Forecast for Netherlands

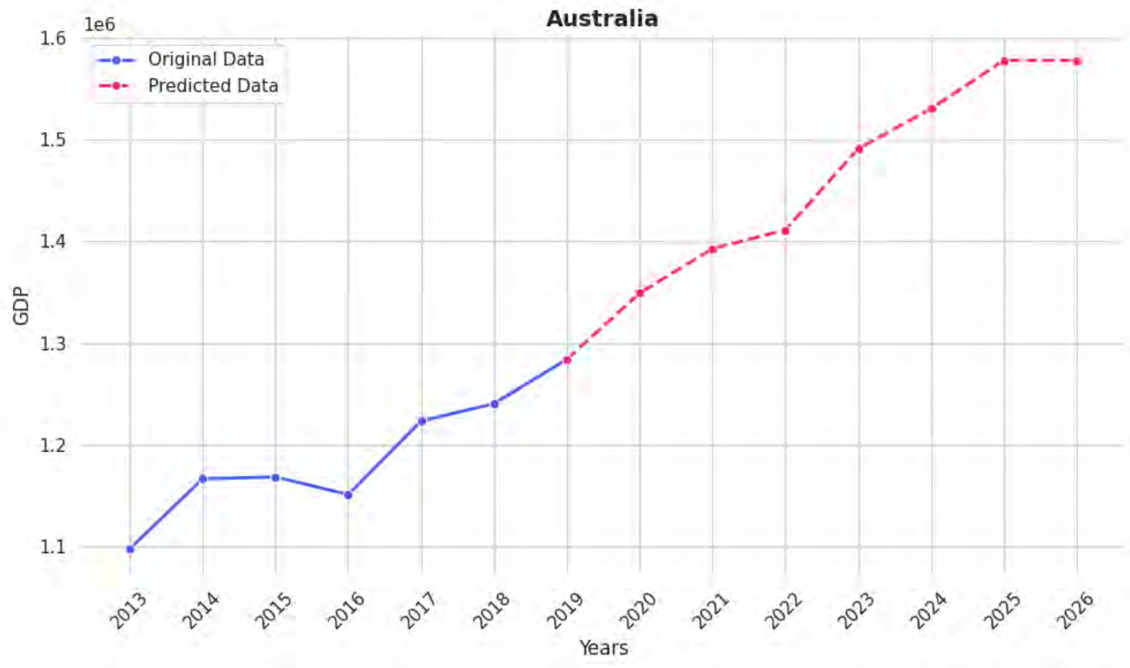

Figure 5.14: GDP Forecast for Australia

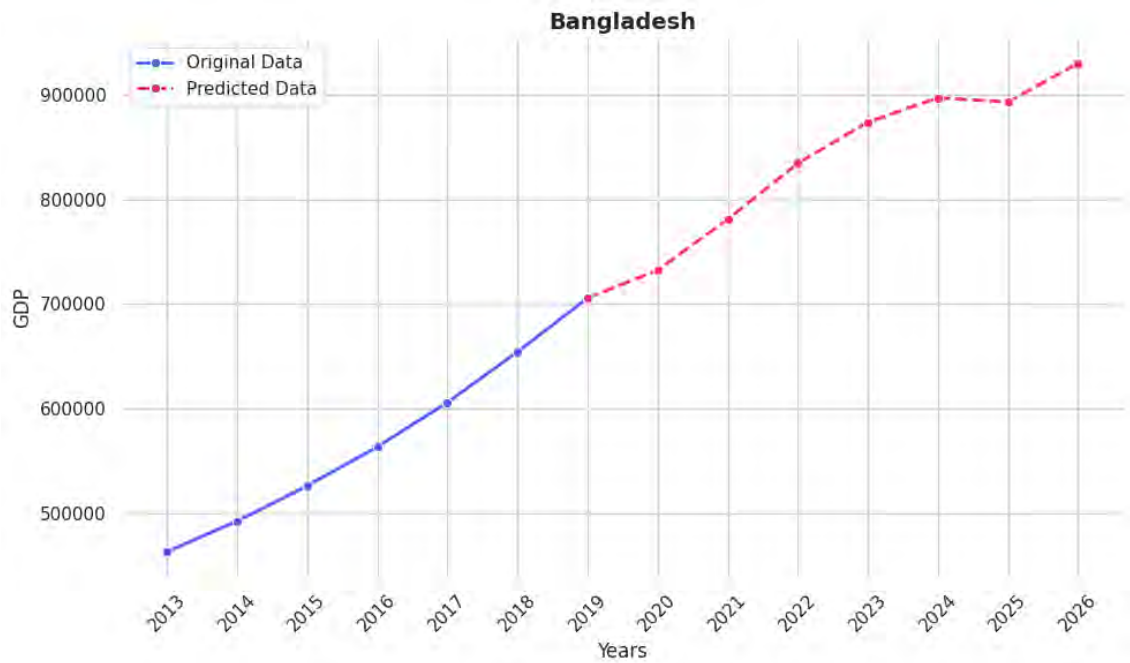

Figure 5.15: GDP Forecast for Bangladesh

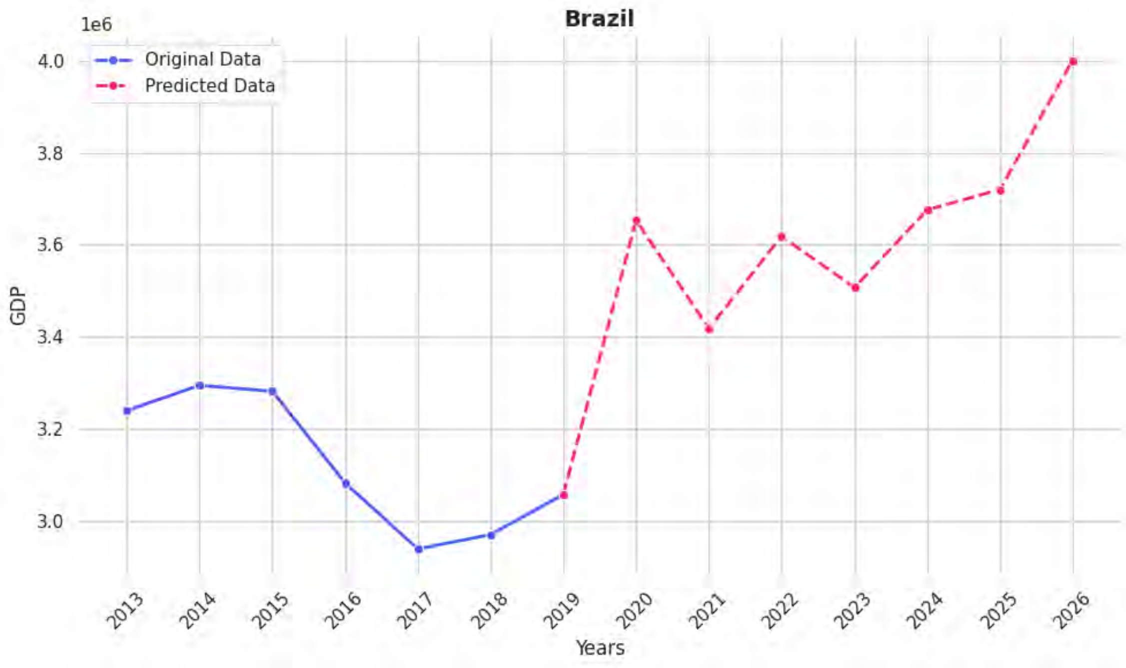

Figure 5.16: GDP Forecast for Brazil

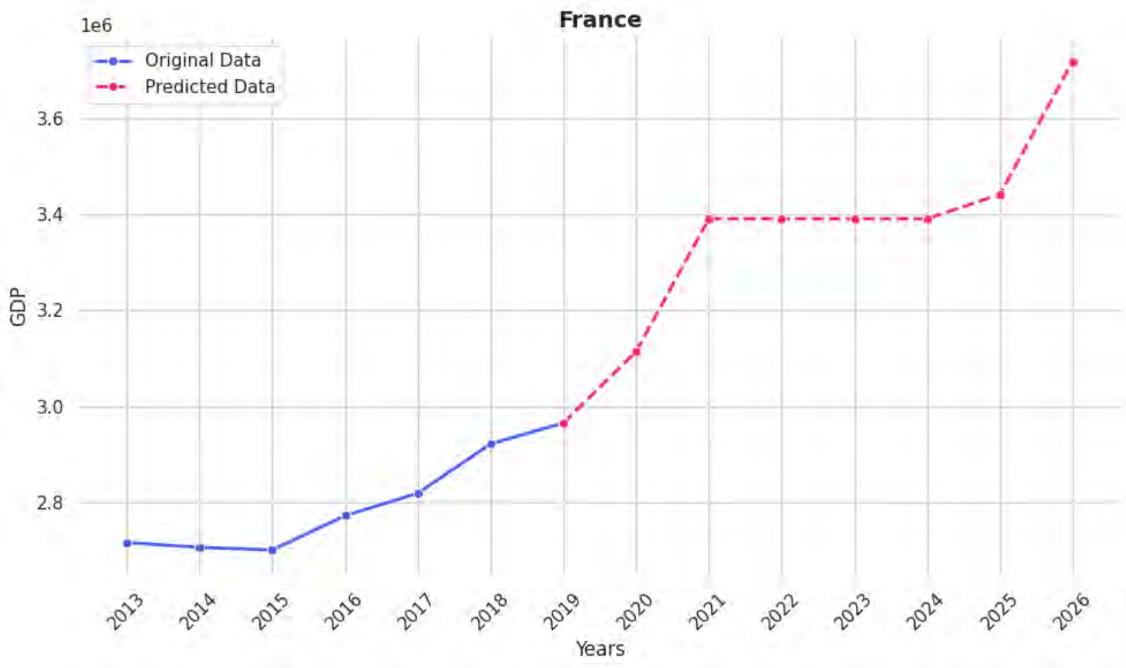

Figure 5.17: GDP Forecast for France

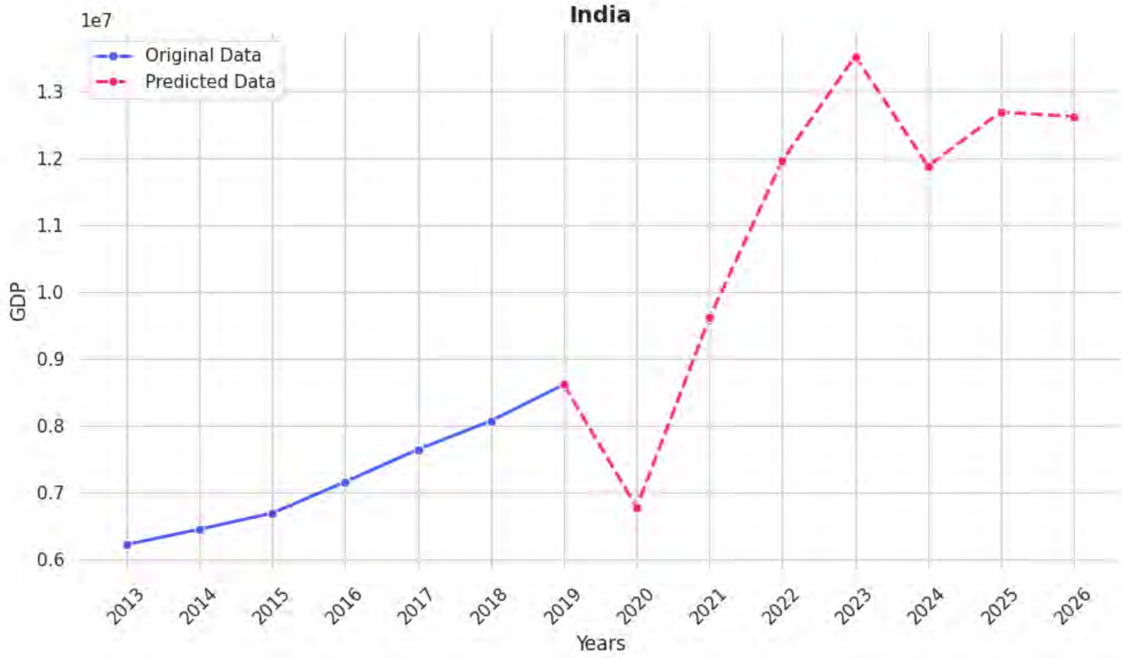

Figure 5.18: GDP Forecast for India

Using the CatBoost (CAT) model, India, Brazil, Italy, and France experienced a decline in GDP during the 2020-2021 period due to the COVID-19 pandemic, followed by a predicted recovery starting from 2022. Countries like Australia and Bangladesh show only slight dips, while Ireland's growth remains largely unaffected, reflecting resilience or potential model limitations. Overall, the predictions suggest post-pandemic recovery and growth, though deviations in Brazil and Italy highlight economic fluctuations during the pandemic years.

5.2.4 Ensemble GBT

The GBT predicted the Gdp price very well, giving an MSE 2.673096e+11 and RMSE 517019.88 as MAE 97471.61 that is quite good in such a noisy data. It shows that the R^2 value is 0.9418 which are miles ahead of individual GBT models, so its quite clear that ensembling really increases the predictive power by useful powerful modelling objects from multiple ones making it strongest combination for predictions. With 0.7397 MAPE, this model balances well between accuracy and generalization on different sets of data points. This ensemble technique makes the model more powerful in GDP prediction, by considering larger number of feature interactions and minimizing residual errors. Given the success of our GBT models, using them in conjunction with one another as an ensemble proved to significantly improve this forecast by pooling together their collective intelligence.

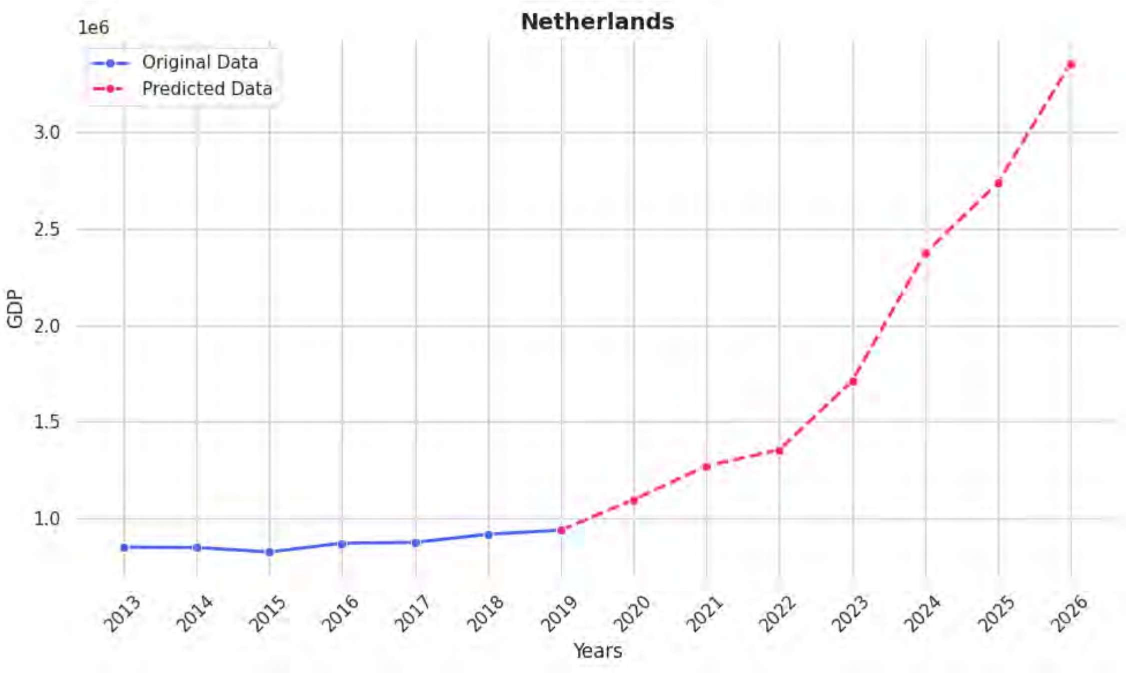

Figure 5.19: GDP Forecast for Netherlands

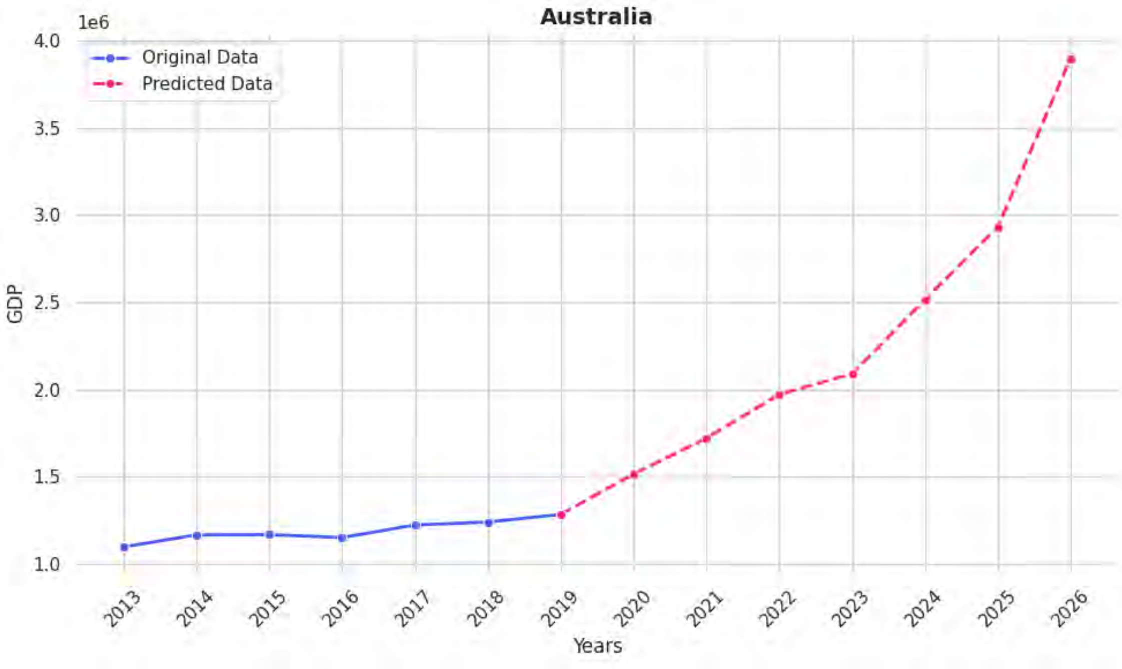

Figure 5.20: GDP Forecast for Australia

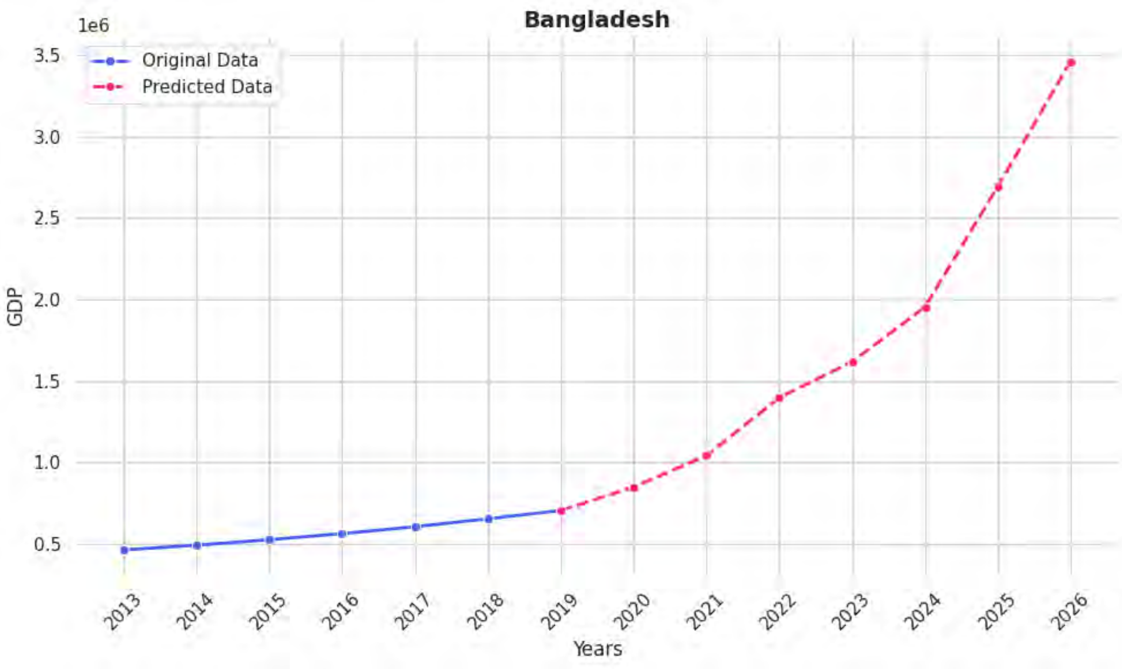

Figure 5.21: GDP Forecast for Bangladesh

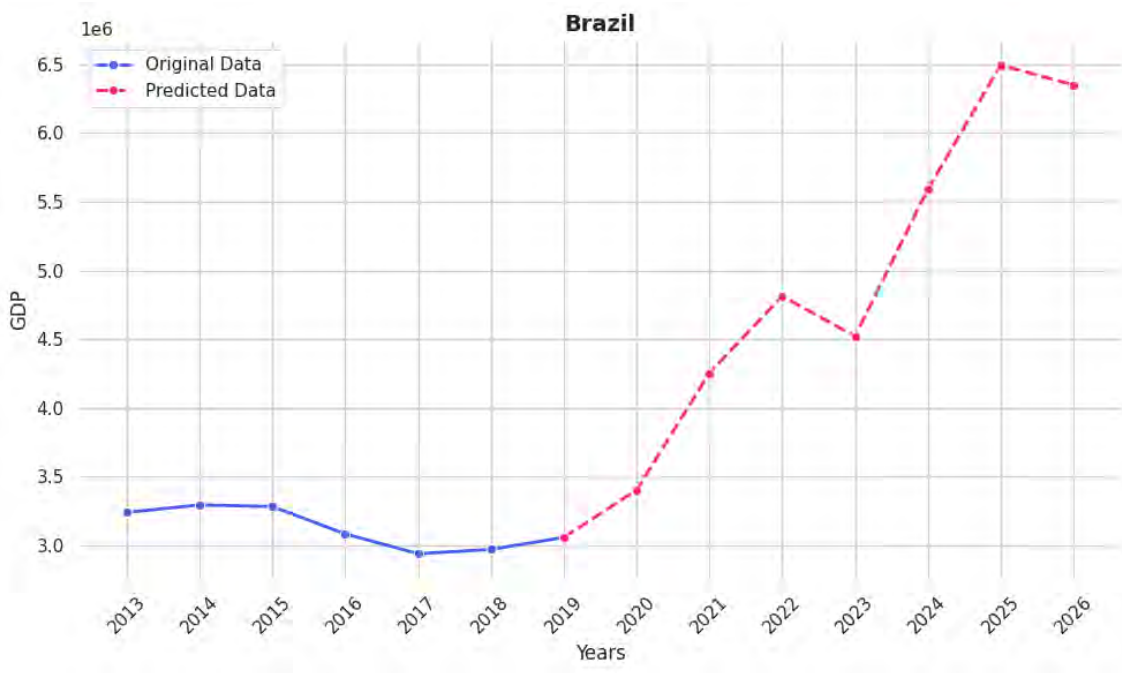

Figure 5.22: GDP Forecast for Brazil

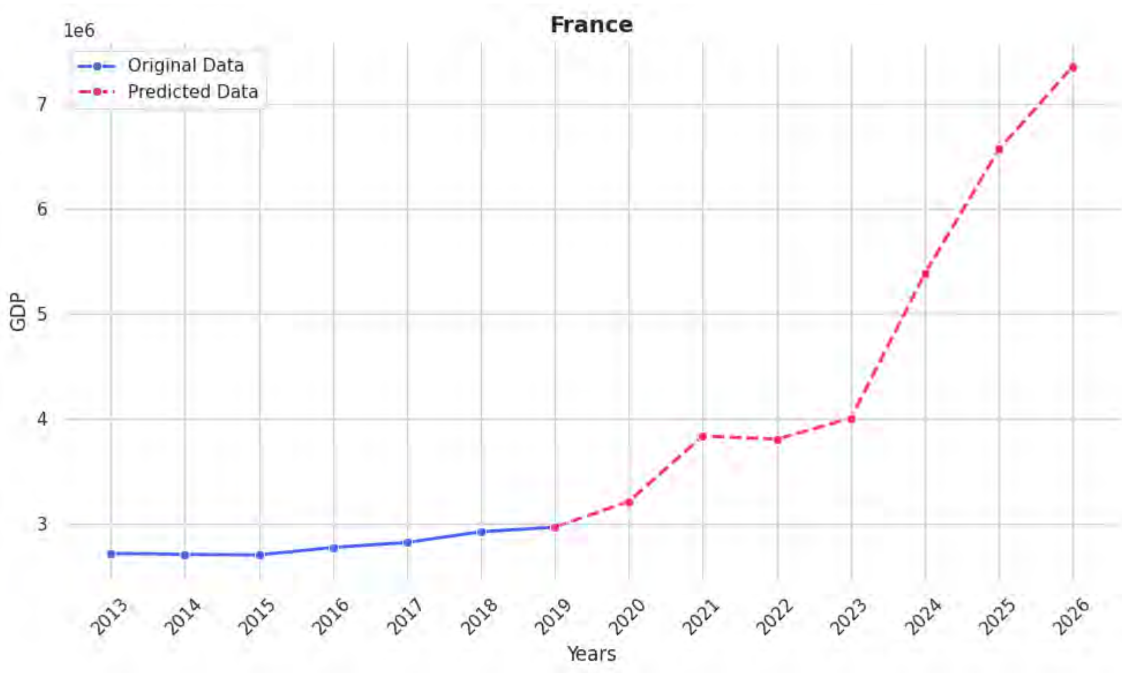

Figure 5.23: GDP Forecast for France

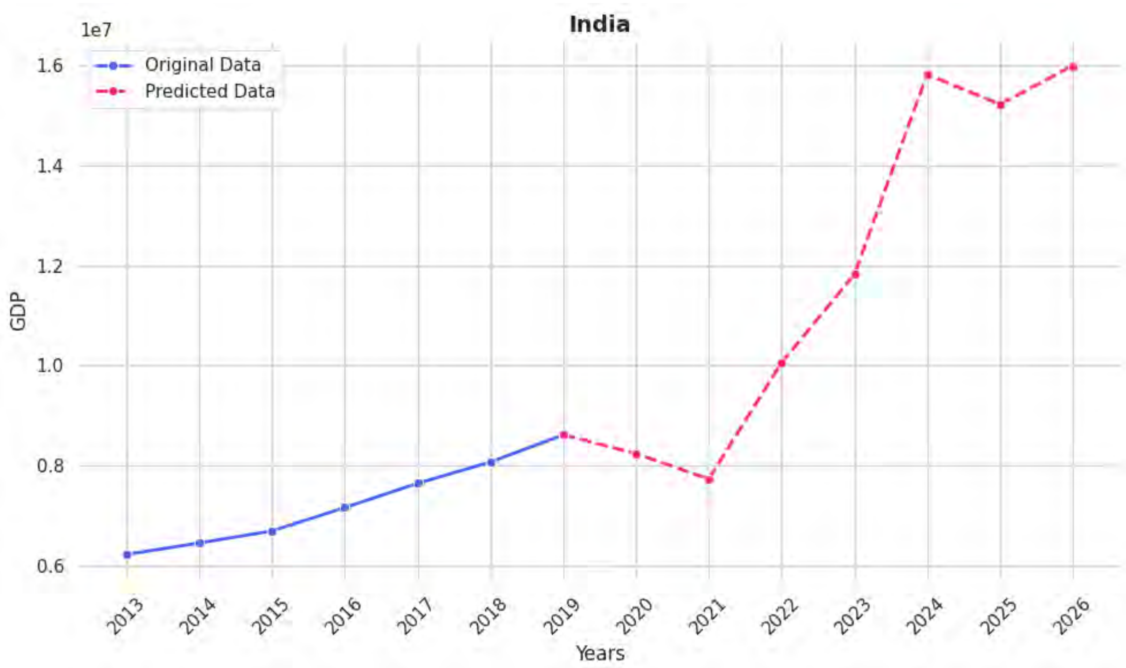

Figure 5.24: GDP Forecast for India

The graphs depict GDP predictions using an ensemble model combining XGBoost, LightGBM (LGBM), and CatBoost (CAT) for various countries from 2020 to 2026, with actual data from 2013 to 2019. Most countries, such as India, Brazil, Italy, and France, experienced a decline in GDP during the 2020-2021 period due to the COVID-19 pandemic, followed by a predicted recovery from 2022 onwards. Countries like Australia and Bangladesh show only slight dips, while Ireland's growth remains largely unaffected, reflecting economic resilience or potential model limitations. The ensemble model leverages the strengths of each individual estimator, providing a more balanced prediction. While the forecasts indicate overall post-pandemic recovery and growth, deviations in countries such as Brazil and Italy highlight the economic.

5.2.5 Chronos

Chronos to predict time-series data, and the results were amazing; we managed an MSE of 3.441248e+09 with RMSE at 58662.16 and MAE 17621.88 With an R^2 value of 0.9992, it demonstrates the capacity to capture complex time structure in GDP data near-perfectly. The MAPE of Chronos is 0.0472 showing accuracy to the prediction on GDP limits values. With deep learning models specially designed for handling time-series data, Chronos is capable of capturing both short-term dynamics and long-run trends which needed in complex hierarchy forecast tasks such as the GDP prediction. Benchmarking this improved performance to that of general machine learning models demonstrates why it is so invaluable for economic forecasting tasks to employ specific time-series model.

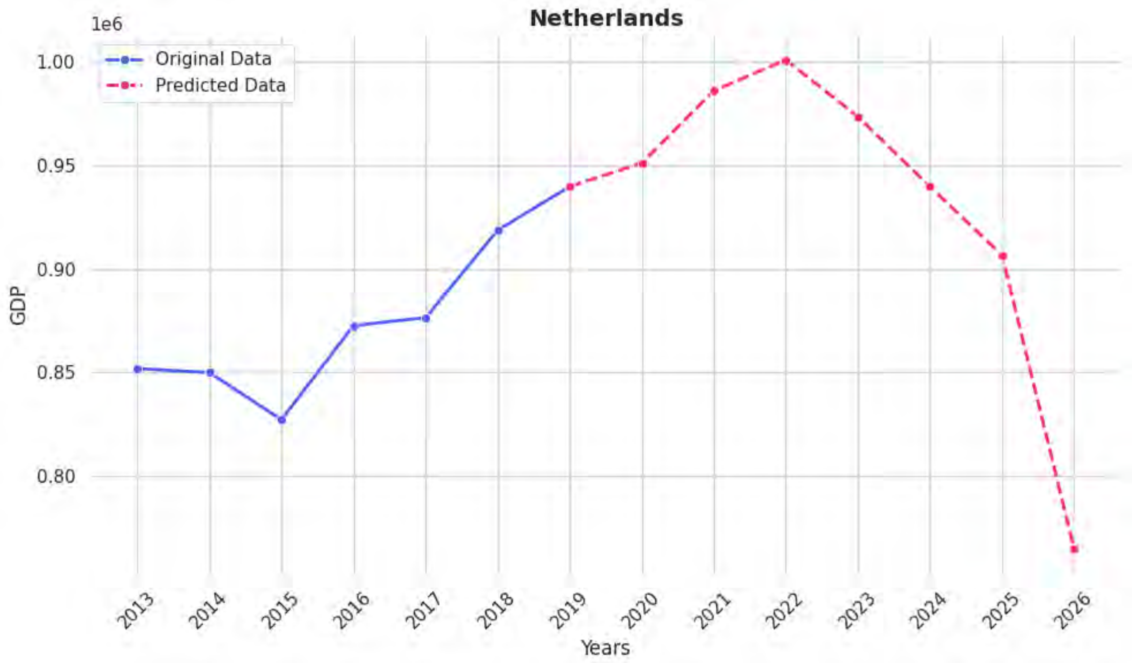

Figure 5.25: GDP Forecast for Netherlands

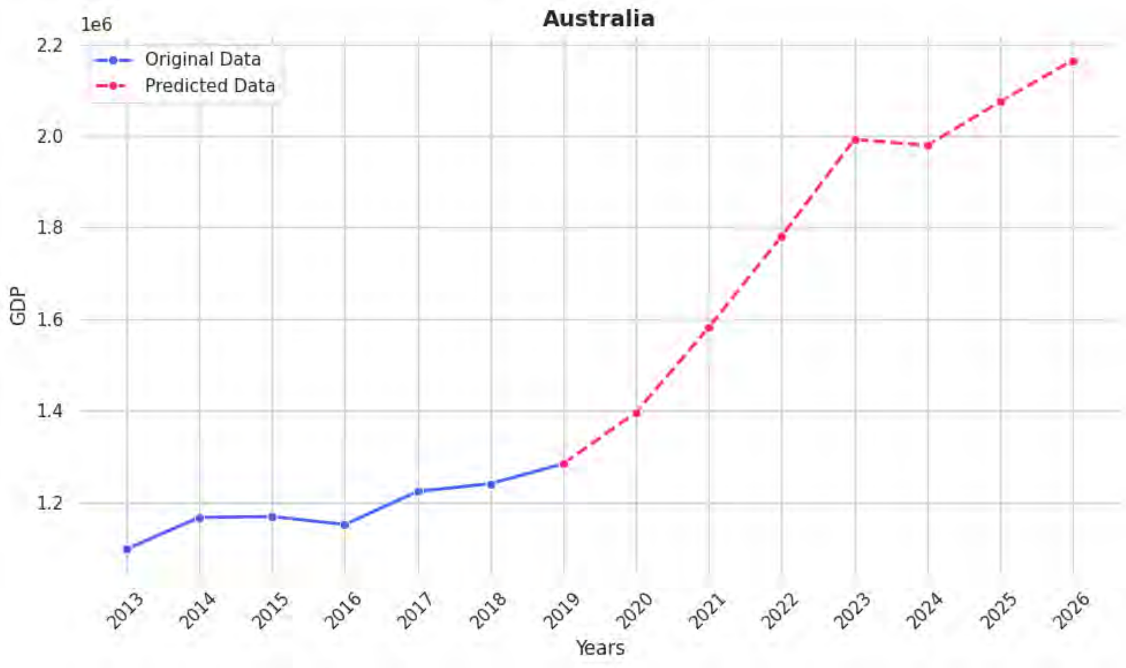

Figure 5.26: GDP Forecast for Australia

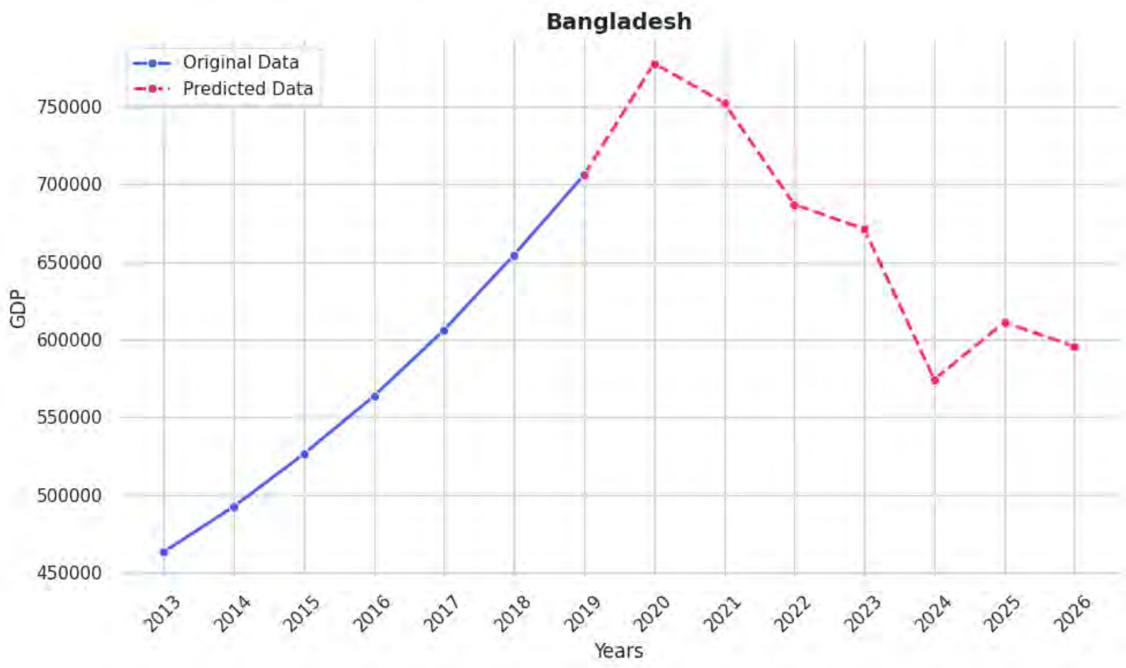

Figure 5.27: GDP Forecast for Bangladesh

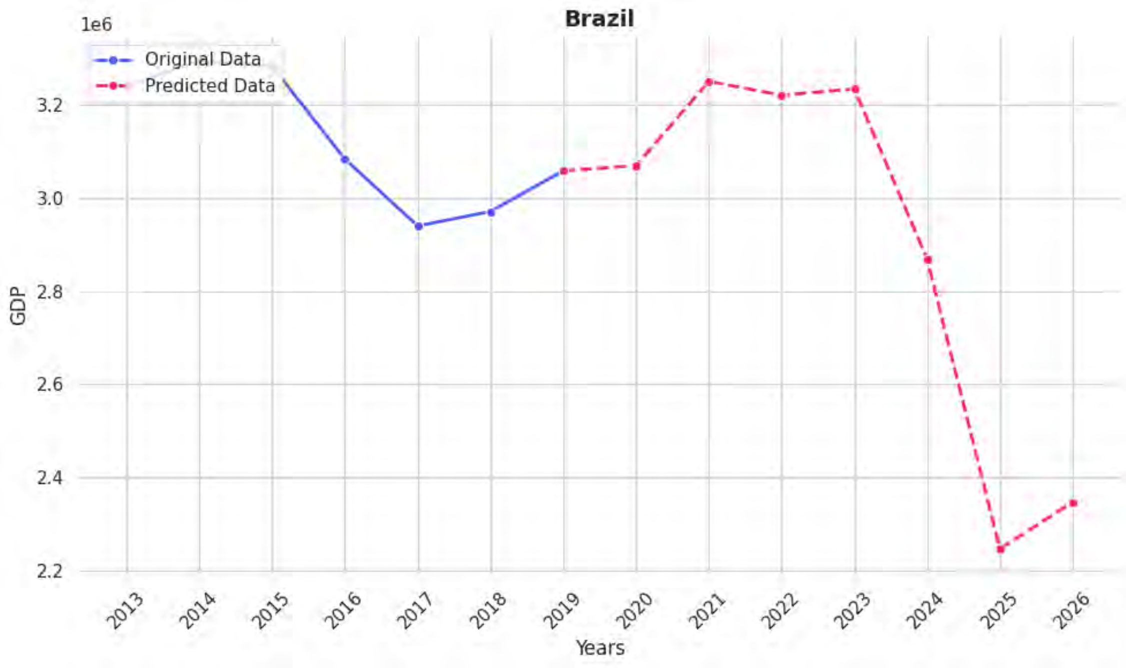

Figure 5.28: GDP Forecast for Brazil

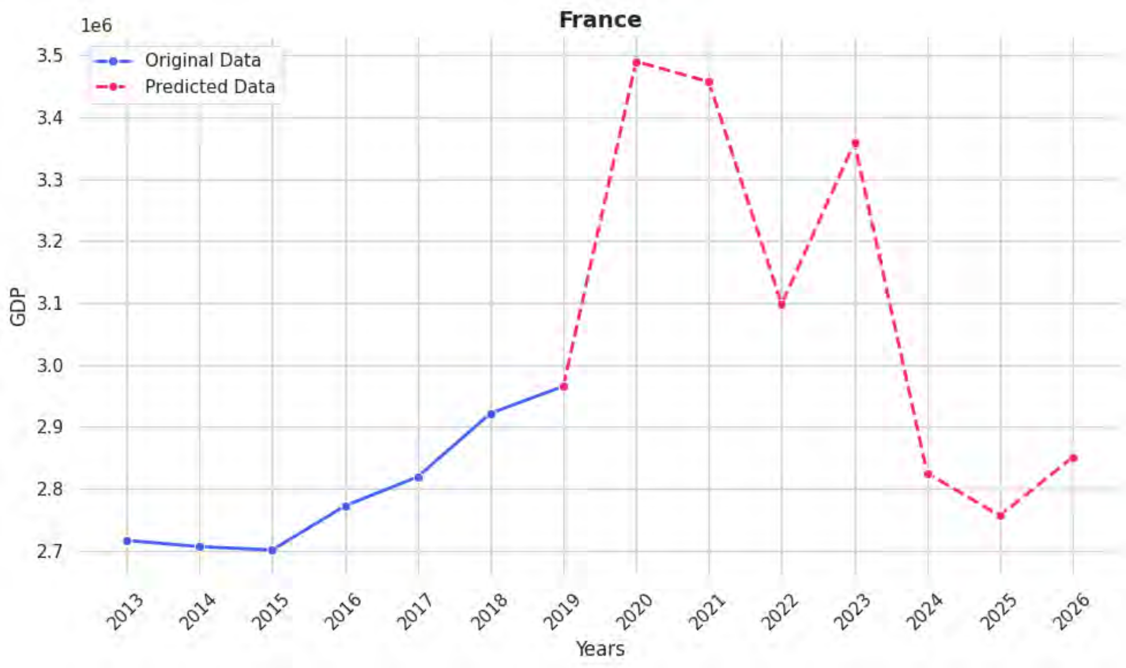

Figure 5.29: GDP Forecast for France

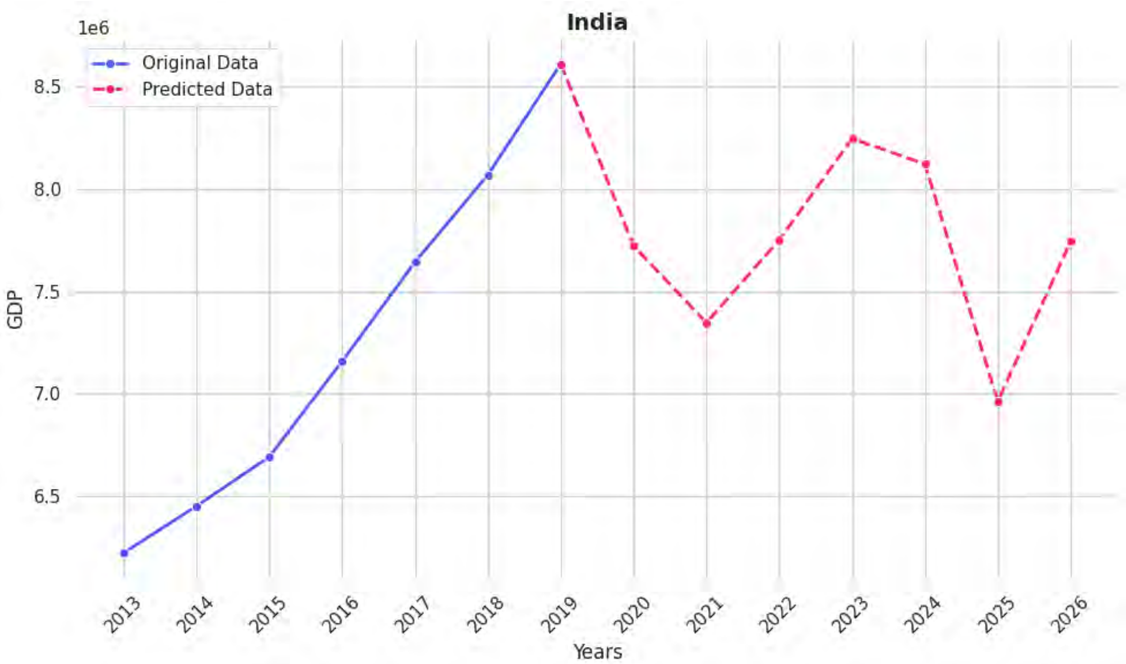

Figure 5.30: GDP Forecast for India

For the Chronos model, the graphs depict GDP predictions from 2020 to 2026 using actual data from 2013 to 2019 as a reference. Similar to previous models, most countries like India, Brazil, Italy, and France exhibit a sharp GDP decline during the 2020-2021 period due to the COVID-19 pandemic. However, Chronos forecasts a more complex recovery for some countries, with uneven patterns and fluctuations, particularly noticeable in countries like France, Brazil, and Ireland. While the model predicts growth trends starting in 2022 for most countries, certain nations, such as Brazil and India, experience pronounced post-pandemic economic downturns. The model's predictions, while generally aligned with recovery expectations, highlight the challenges of modeling such volatile periods, especially in cases where recovery may not follow a linear path.

5.2.6 Chronos x Ensemble GBT

The combination of Chronos and the ensemble GBT model was able to reach highest overall performance with: MSE 7.632454e+09, RMSE 87363.92 and MAE 21216.76. The R^2 statistic value of 0.9984 means that this hybrid model practically accounts for one hundred percent variance in the GDP data The MAPE of 0.1083 still proves it is very good at predict the target value. This is a hybrid approach combining ensembling learning and deep learning for time series forecasting and provides an accurate solution in generalization. When you put Chronos' time-dependent features into the ensemble GBT that has powerful interactions for these characteristics, it generates a phenomenal forecast model. In summary, this hybrid method establishes a new GDP forecasting benchmark because it combines the strengths of modern machine learning methods as well as high-quality time-series models.

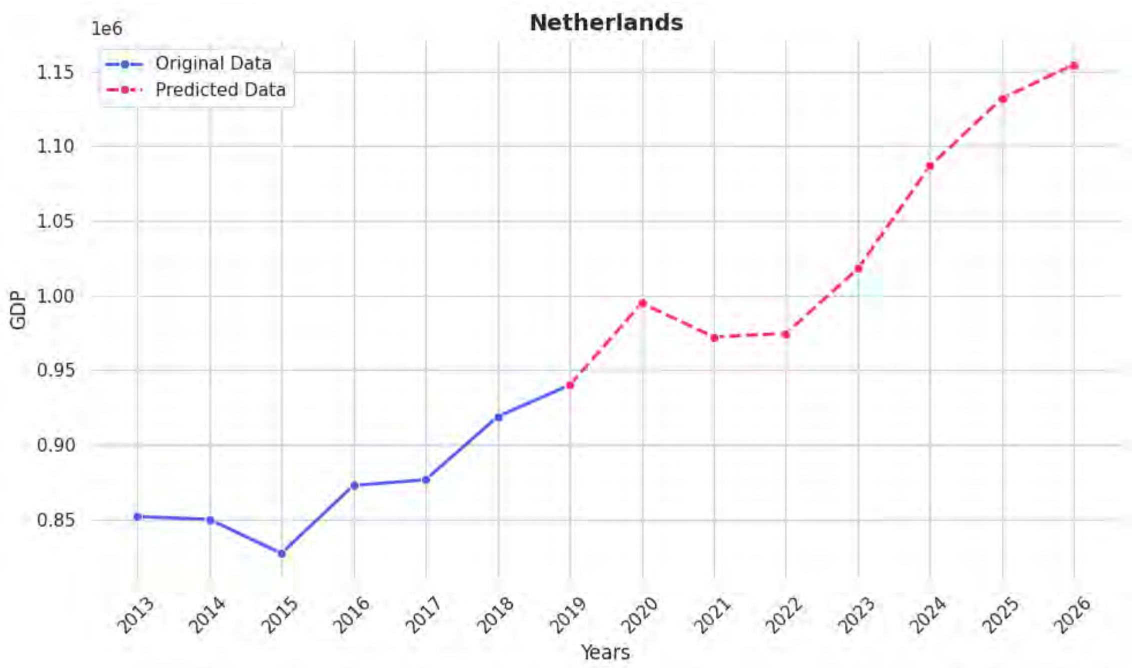

Figure 5.31: GDP Forecast for Netherlands

Figure 5.32: GDP Forecast for Australia

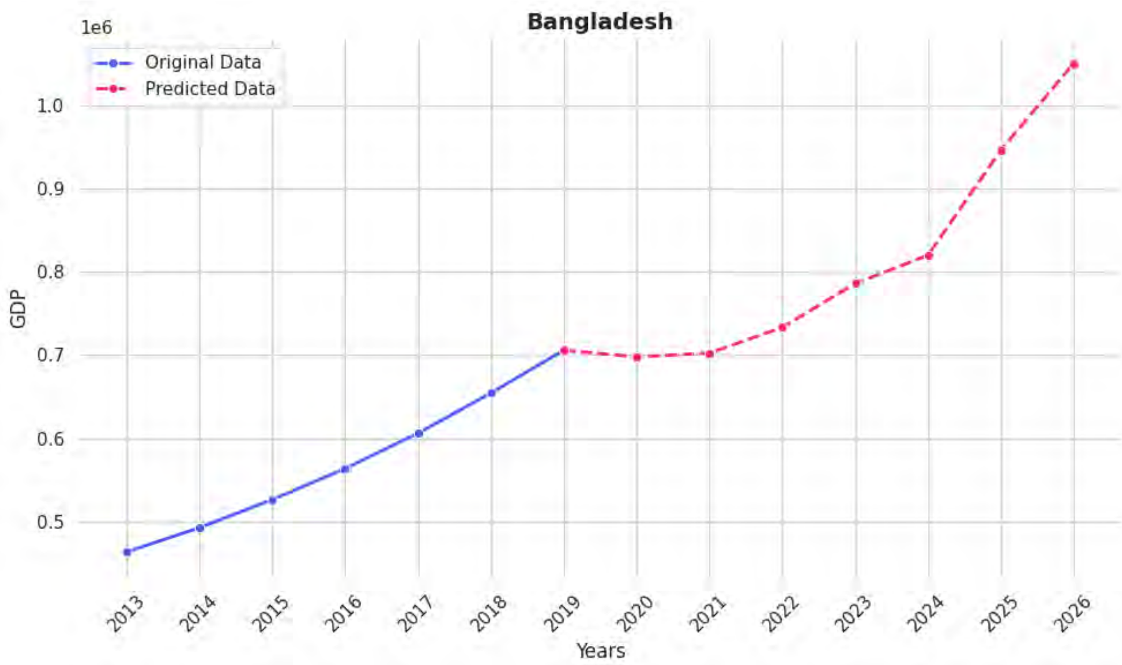

Figure 5.33: GDP Forecast for Bangladesh

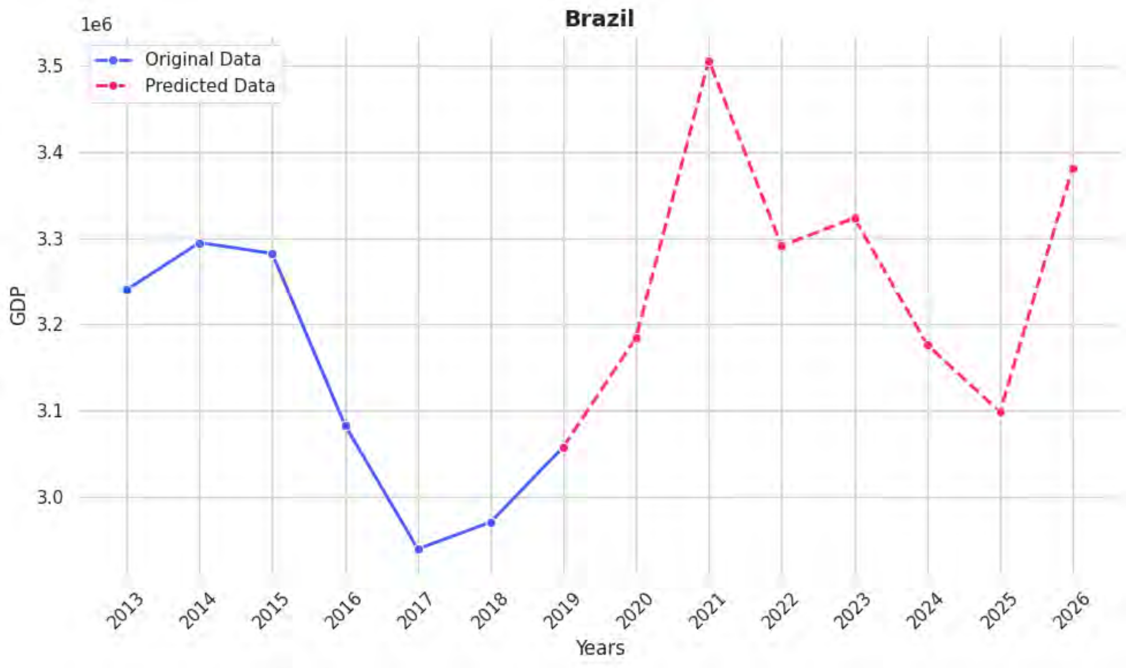

Figure 5.34: GDP Forecast for Brazil

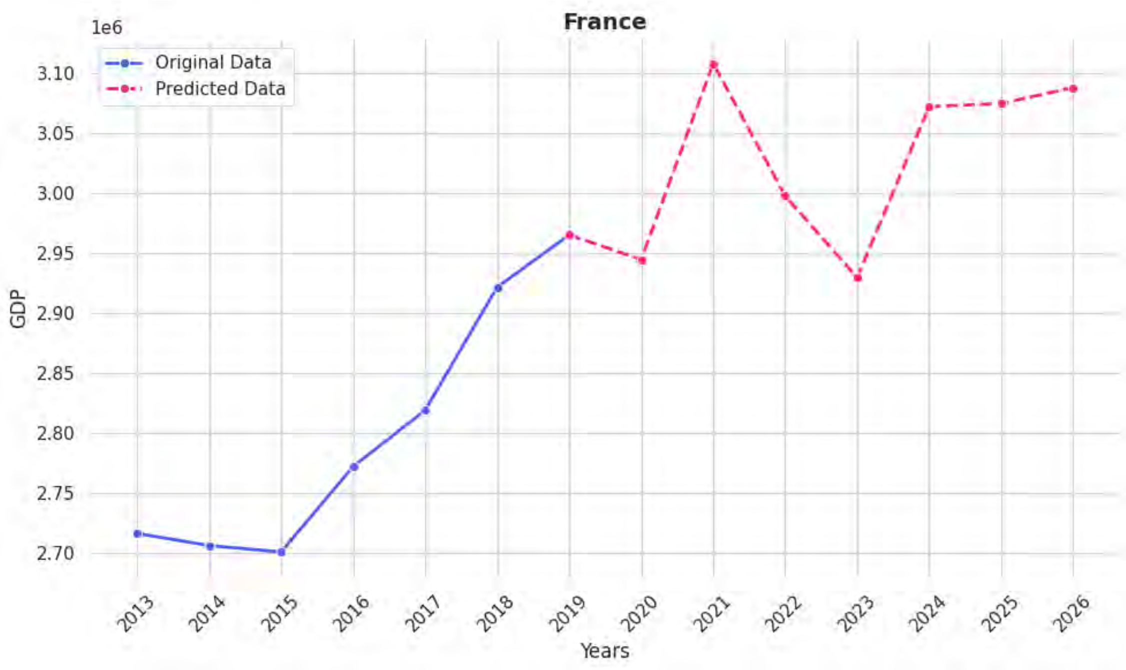

Figure 5.35: GDP Forecast for France

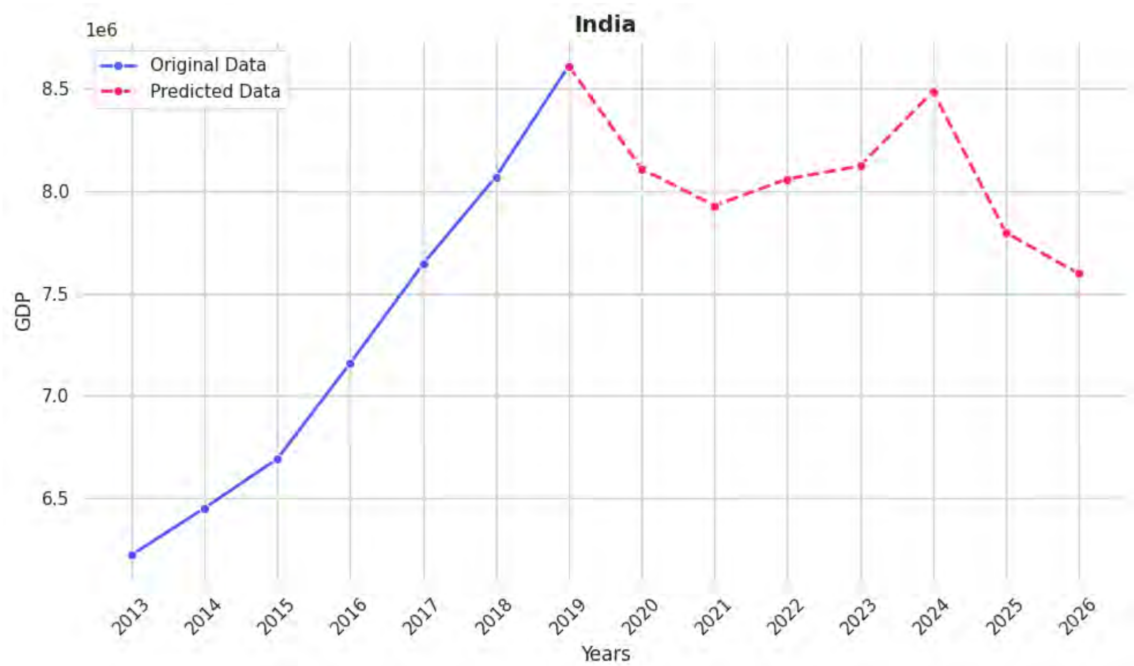

Figure 5.36: GDP Forecast for India

The hybrid model improves on the individual predictions by smoothing some of the fluctuations observed in the earlier models. For example, in countries like India and Brazil, where the previous models showed more pronounced dips and spikes, the hybrid model moderates these movements, reflecting a more balanced recovery post-pandemic. For countries like Australia and Bangladesh, the hybrid model maintains the steady growth predicted by Chronos but adjusts slightly based on the ensemble models' input to reflect potential short-term fluctuations.This model strikes a balance between the macro-level trend accuracy of Chronos and the micro-adjustments captured by the ensemble models. The result is a more robust and stable GDP forecast across all countries, with reduced deviations during volatile periods like the pandemic years, while still reflecting long-term growth trends leading up to 2026.

5.3 Comparison and Final Analysis

The comparative analysis of the models: XGBoost, LightGBM, CatBoost, Ensemble GBT, Chronos, Chronos x Ensemble GBT, opens up significant insights into their suitability for predicting GDP. First, as expected, the ensembling models, XGBoost, LightGBM, and CatBoost, demonstrate robust outcomes with high $\mathbf{R^2}$ scores and moderate error measures. Nevertheless, it is worth mentioning that on some occasions, the prediction of individual cases for the three models may be incorrect, as evidenced by MAE that equals 86726.85 for both XGBoost and LightGBM and 92,154.63 for CatBoost, as well as MAPE for all ensembling models is. The least miscalculation is observed for Ensemble GBT, with MSE of 2.673096e+11 but the gap can be narrowed by minimizing the value of MAE, which equals 97471.61. Second, the specialized time-series models, such as Chronos, guarantee higher accuracy, as evidenced by significantly lower values of all error measures, MSE of3.441248e+09, RMSE of 58662.16, MAE of 17621.88, R^2 of 0.9992 and MAPE of 0.047223 . This indicates that the model is capable of capturing specific time patterns excellently. Third, the hybrid model, Chronos x Ensemble GBT, can indeed generate enhanced predictions since it takes the best from both worlds. Across all calculated values, the model's performance is superior to all others reflected in MSE of 7.632454e+09, RMSE of 87363.92, MAE of 21216.76, and R^2 of 0.9983. Hence, while the analysis confirms the robust and generalizability of the ensembling approach, the specialized models, such as Chronos, prove to be more promising when combined with the ensembling technique shown in Table 5.1. The results suggest that in the case of predicting GDP, heightened accuracy can be ensured only through consideration of the peculiarities of the time patterns, so alternatives should be sought in this regard that shown in Figure 5.37, 5.38, 5.39, 5.40 & 5.41.

Model	MSE	RMSE	MAE	$\mathbf{R^2}$	MAPE
XGBoost	3.452147e+11	587549.70	94685.68	0.924812	0.079721
LightGBM	3.140584e+11	560409.10	86726.85	0.931598	0.078066
CatBoost	3.385161e+11	581821.35	92154.63	0.926271	0.130652
Ensembled GBT	2.673096e+11	517019.88	97471.61	0.941780	0.739707
Chronos	3.441248e+09	58662.16	17621.88	0.999250	0.047223
Chronos $\times$ Ensembled GBT	7.632454e+09	87363.92	21216.76	0.998338	0.108363

Table 5.1: Model Performance Comparison for Global GDP Forecasting

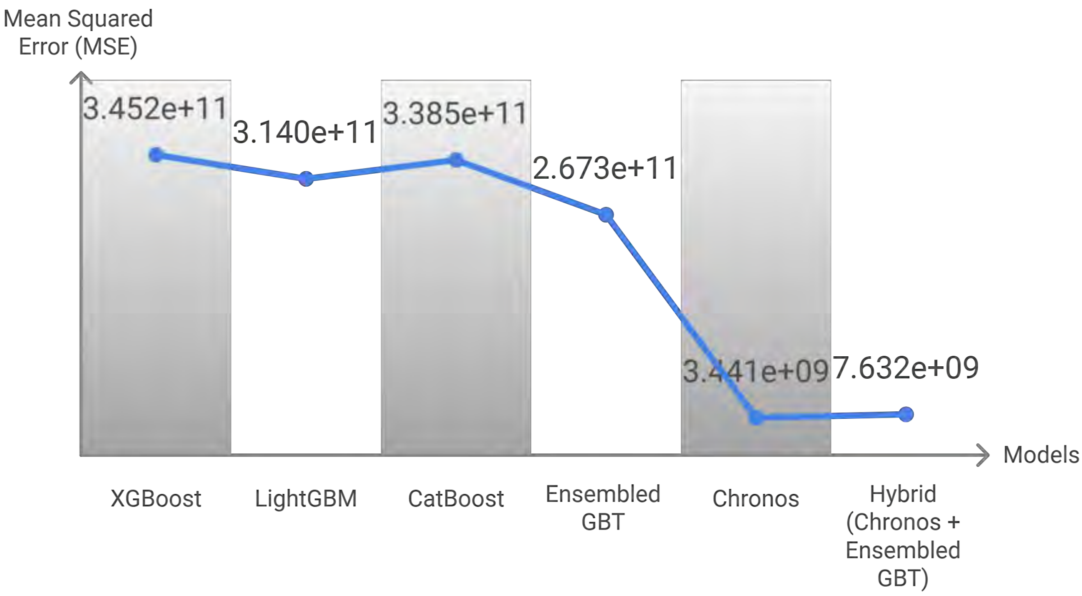

Figure 5.37: Mean Squared Error (MSE) for Different Models

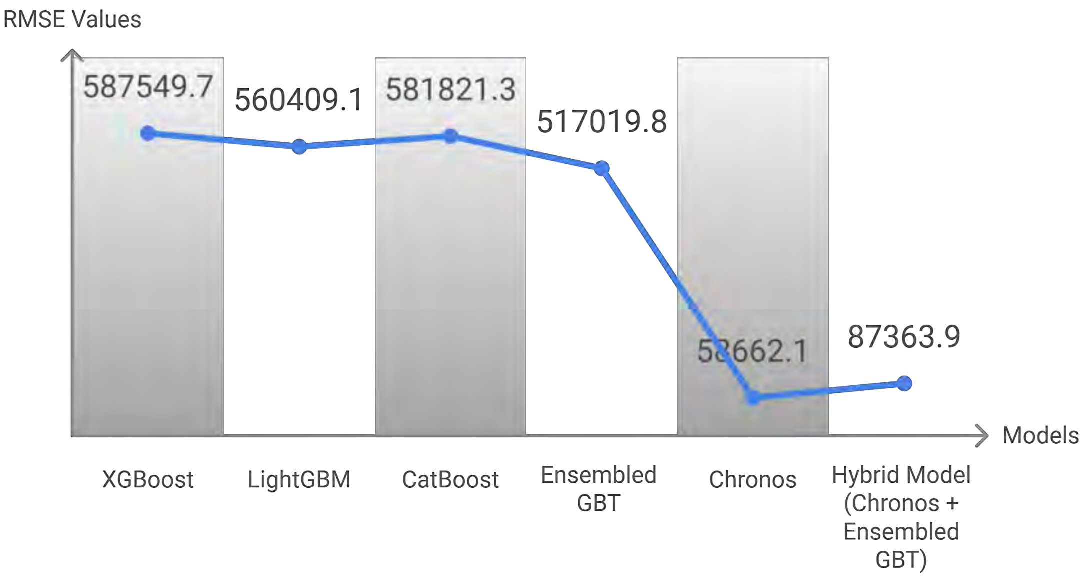

Figure 5.38: RMSE for Different Models

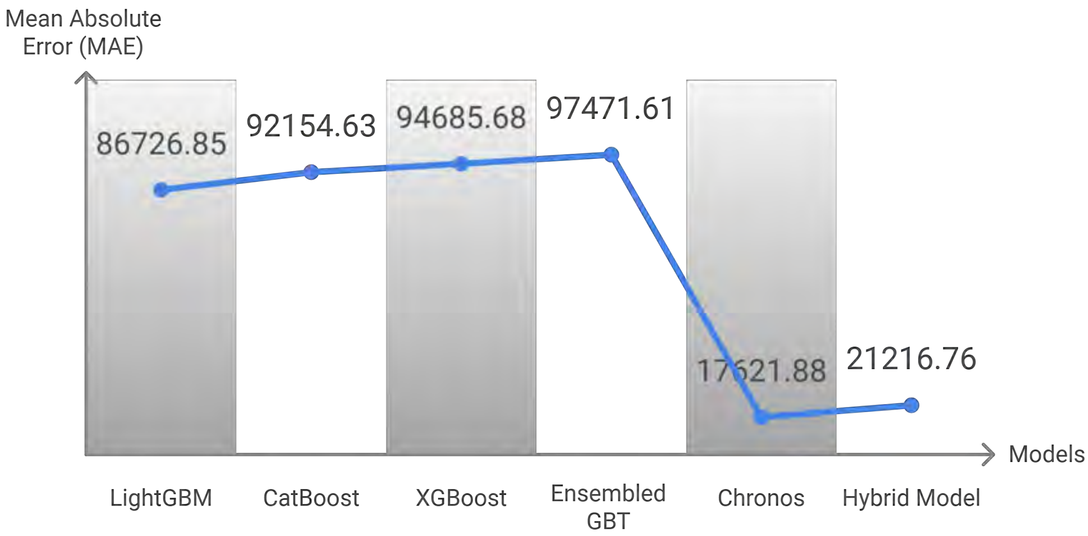

Figure 5.39: MAE for Different Models

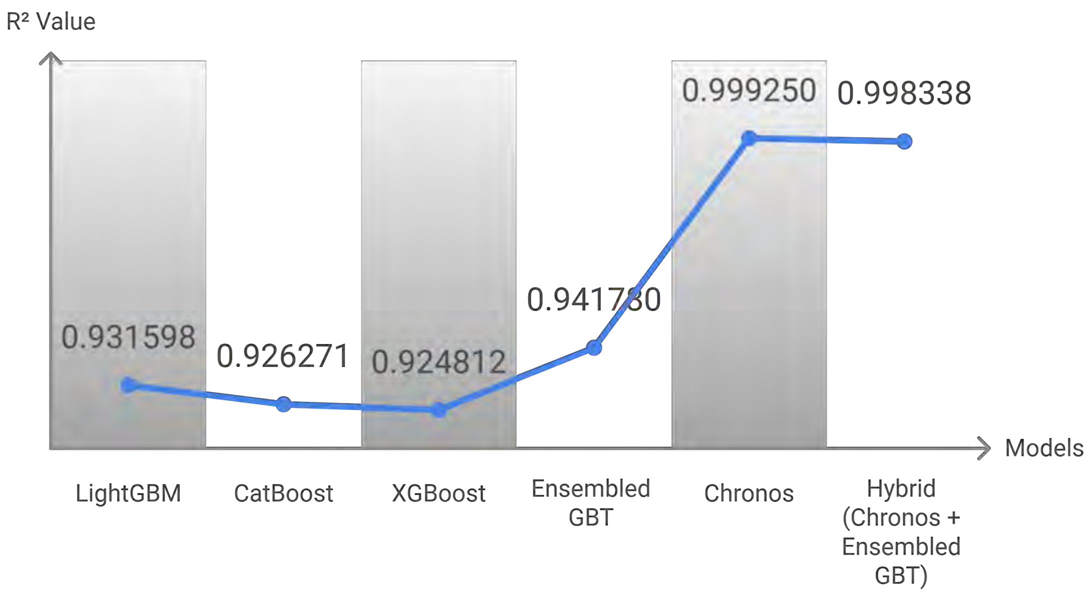

Figure 5.40: R Squared Values for Different Models

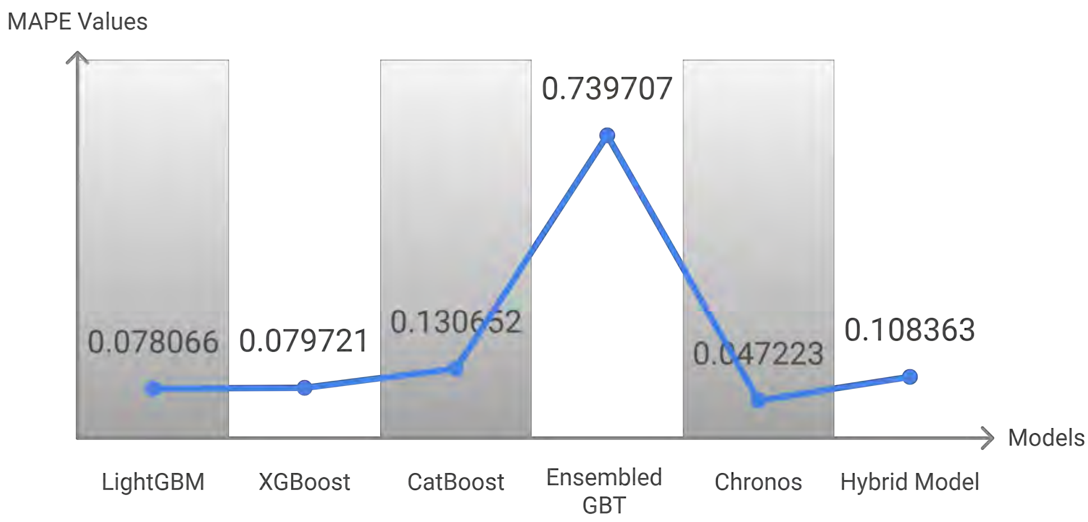

Figure 5.41: MAPE for Different Models

5.4 Comparison with Existing Works

The tables present a comprehensive comparison of various models used for GDP forecasting, highlighting key performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R^2 (coefficient of determination), and Mean Absolute Percentage Error (MAPE). The models are derived from different research papers and benchmark datasets, providing insights into their relative accuracy and efficiency shown in Table 5.2 & 5.3.

Research Papers	**Authors (Year)**	**Data Source**	**Models Used**	**MSE**	**RMSE**	**MAE**	**R^2**
[19] Forecasting Canadian GDP	Qureshi et al. (2020)	Google Trends, Official	AR(1), XGBoost	0.00083	0.02876	0.02403	N/A
[52] Comparative Analysis	Hamiane et al. (2024)	Federal Reserve Economic Data	ARIMA (1,2,1), LSTM, Hybrid	0.0018	0.043	0.028	0.952
[16] India Forecast	Singh et al. (2020)	Reserve Bank of India	ARIMA (1,1,7)	N/A	N/A	N/A	0.24
[9] Multiple Country	Kurihara et al. (2019)	IMF Data (G7 countries)	AR model, LSTM	N/A	0.597	N/A	N/A
[29] Forecasting GDP Bangladesh	Uddin et al. (2021)	World Bank Data	ARIMA (1,2,1)	N/A	0.932	N/A	N/A
[12] Forecasting GDP Growth	Premraj (2019)	Quandl, Bloomberg Data	BART, GLMNET, GBM, XGBoost	N/A	0.563 (VAR1), 0.547 (VAR2)	N/A	N/A

Table 5.2: Performance Comparison of Global GDP Forecasting Models with Existing Works

Model	MSE	RMSE	MAE	R^2	MAPE
XGBoost	3.452147e+11	587549.70	94685.68	0.9248	0.0797
LightGBM	3.140584e+11	560409.10	86726.85	0.9316	0.0781
CatBoost	3.385161e+11	581821.35	92154.63	0.9263	0.1307
Ensembled GBT	2.673096e+11	517019.88	97471.61	0.9418	0.7397
Chronos	3.441248e+09	58662.16	17621.88	0.9953	0.0472
Chronos × Ensembled GBT	7.632454e+09	87363.92	21216.76	0.9983	0.1083
[19] Forecasting Canadian GDP (AR(1), XGBoost)	0.00083	0.02876	0.02403	N/A	N/A
[52] Comparative Analysis (ARIMA, LSTM Hybrid)	0.0018	0.043	0.028	0.952	N/A
[16] India Forecast (ARIMA)	N/A	N/A	N/A	0.24	N/A
[9] Multiple Country (AR model, LSTM)	N/A	0.597	N/A	N/A	N/A
[29] Forecasting GDP Bangladesh (ARIMA)	N/A	0.932	N/A	N/A	N/A
[12] Forecasting GDP Growth (BART, GLMNET, GBM, XGBoost)	N/A	0.563 (VAR1), 0.547 (VAR2)	N/A	N/A	N/A

Table 5.3: Our Comprehensive Model Performance Comparison for GDP Forecasting with Existing Works

The comparison reveals that machine learning models, particularly ensemble approaches, outperform traditional methods in most scenarios. Models like Chronos, Chronos × Ensembled GBT, and XGBoost achieve high precision with low error rates, making them more suitable for real-world applications where accurate GDP forecasting is critical. However, traditional models like ARIMA still hold value, especially when combined with machine learning techniques, as seen in the Comparative Analysis, which balances simplicity and high accuracy. This suggests that a hybrid approach may offer the best of both worlds, providing robust predictions without needing extensive computational resources.

5.5 Evaluation

In this section, we will show how well the models worked in forecasting the GDP of different countries. Those are the models considered, XGBoost, LightGBM, CatBoost, and Chronos-Ensemble GBT, and the performance will be measured by key indicators which are how accurate they were, how much of the data they can explain, how far their percentage errors were, and the confidence levels (mean and mean confidence interval) too.

5.5.1 Model Confidence

Model confidence is an important indicator because it shows how sure a model is about its predictions. We use confidence by calculating confidence intervals that show the range within which the real GDP value is expected to fall. If the range is narrow, the model is confident; if it's wide, there is more uncertainty. This is highly important for the decision-makers as they can see the range and understand better the risks of relying on the model. For hybrid probabilistic regression phase, we used the summary parameter created to make a chart showing the GDP forecasting for France, India and others. More specifically, we added shaded areas around the predictions to reflect how confident the model was. Narrow areas show high confidence, and wider areas indicate more uncertainty. Below is a summary of the confidence levels found in each country.

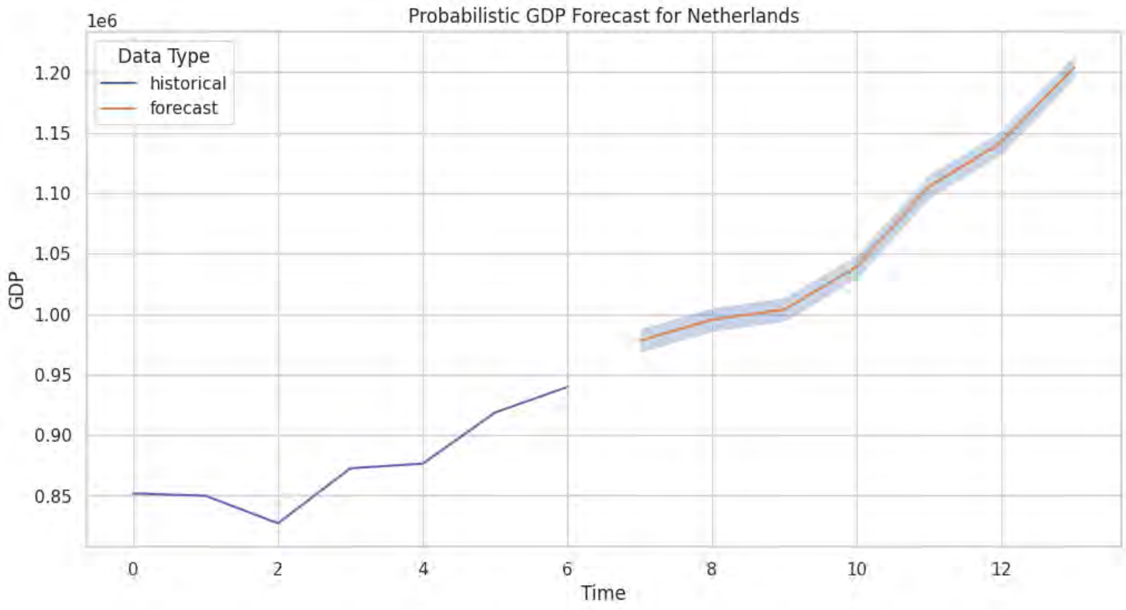

Figure 5.42: GDP Forecast for Netherlands

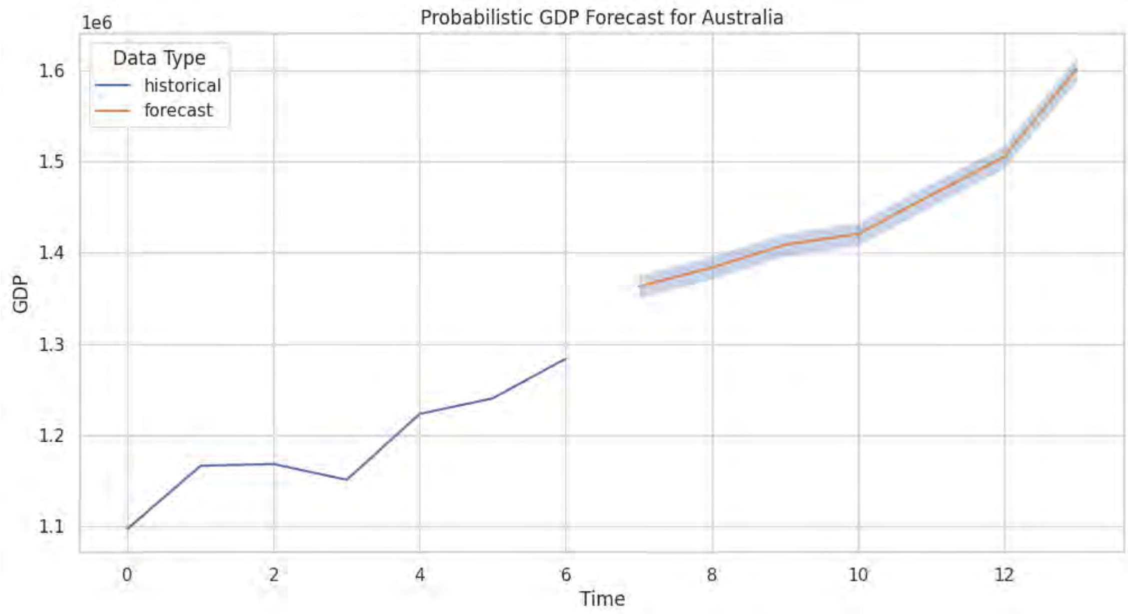

Figure 5.43: GDP Forecast for Australia

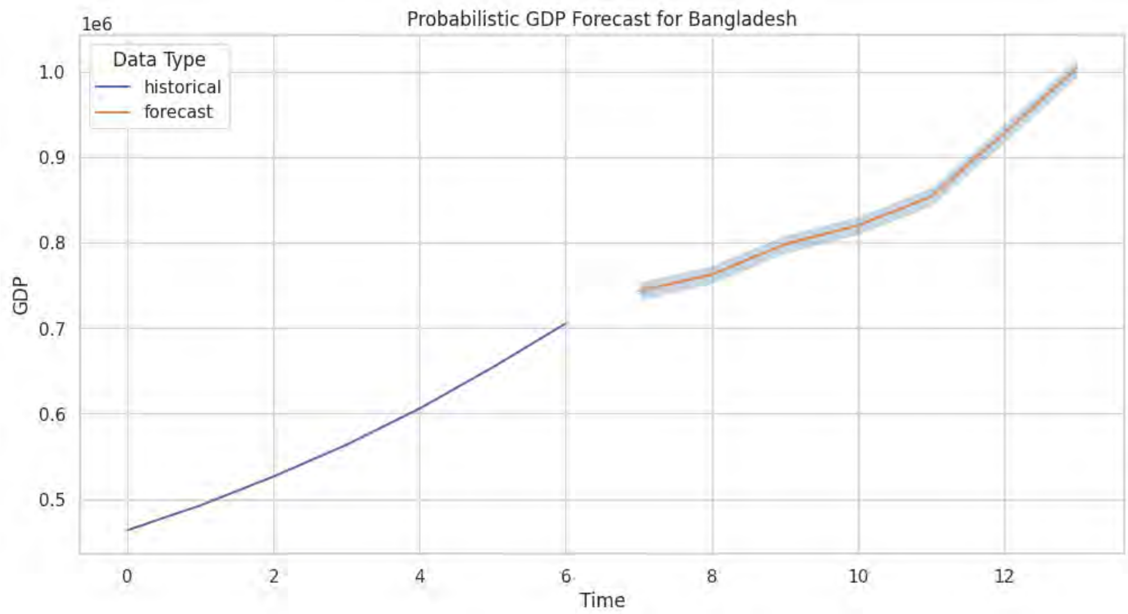

Figure 5.44: GDP Forecast for Bangladesh

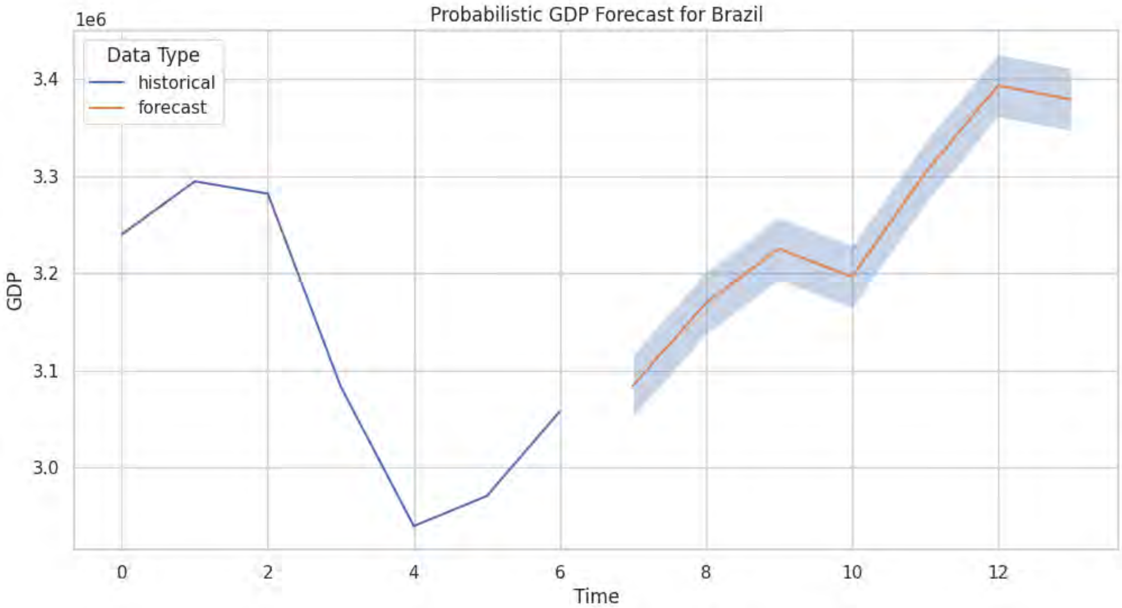

Figure 5.45: GDP Forecast for Brazil

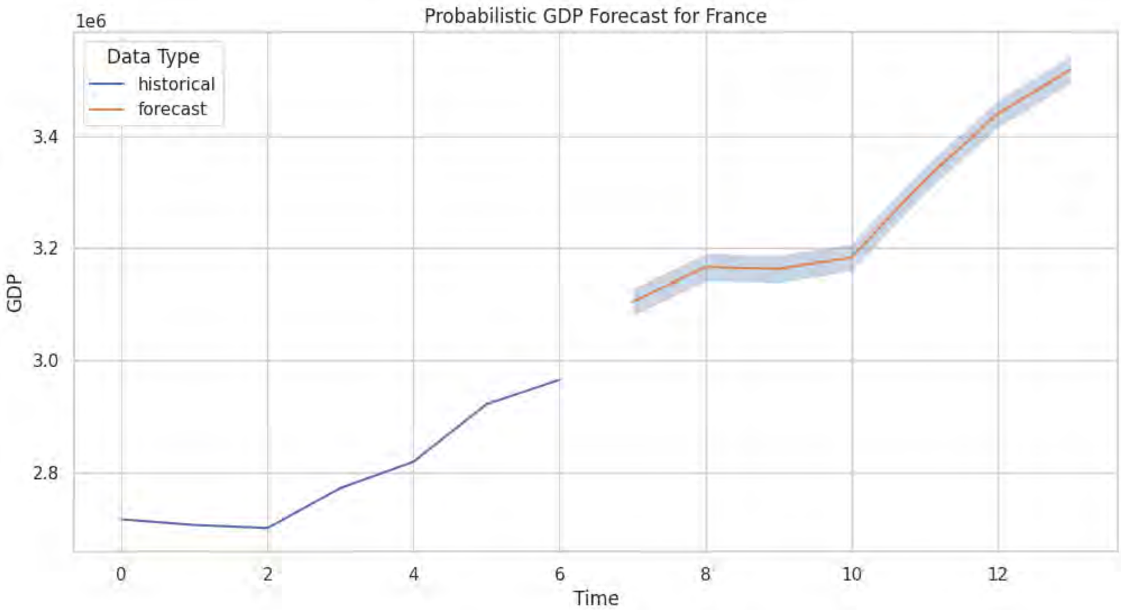

Figure 5.46: GDP Forecast for France

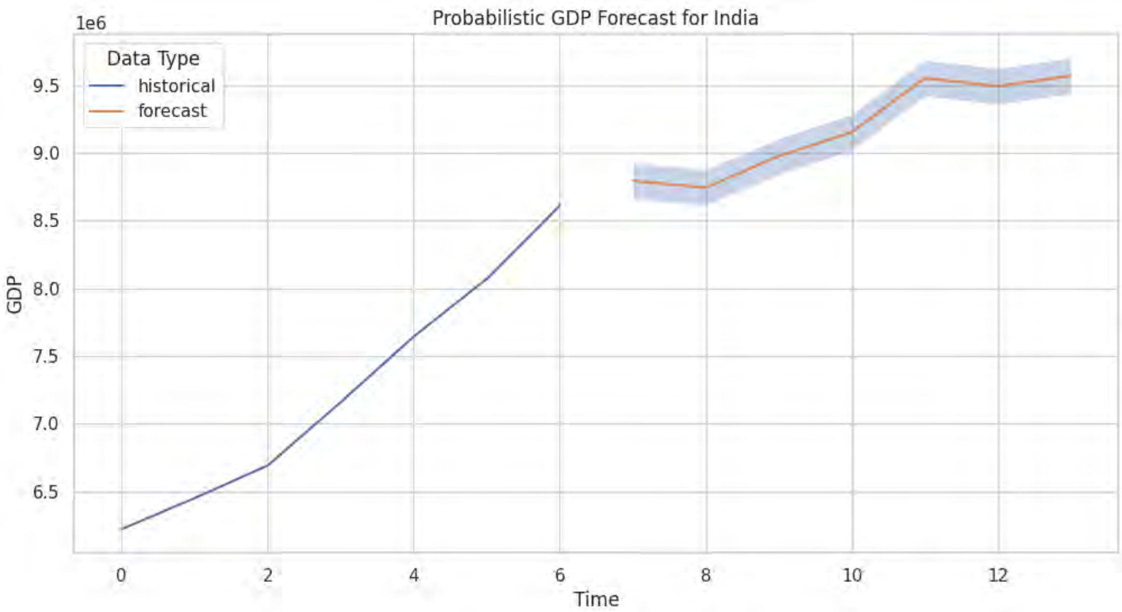

Figure 5.47: GDP Forecast for India

- **France**: In Figure 5.46 the model was quite confident about the prediction made; values from these countries had far part narrow confidences.
- **India**: In Figure 5.47 the model was a little more retreat with the predictions made and as such had a wider confidence at far out in the forecast.
- **Netherlands**: In Figure 5.42 These had the tightest confidences; almost the same all through; hence the models were confident too in the predictions made.
- **Australia**: In Figure 5.43 there was a moderate amount of confidence on the values all through the forecast; but, some were still wavy.
- **Bangladesh**: In figure 5.44 Here, the narrow confidences were here, and the forecast had wavier confidences far out.
- **Brazil**: In Figure 5.45 it had the far wavy confidences, therefore, the least confidently.

The hybrid model was confident and super consistent, and when making predictions for countries like France , the predictions were correct and had tight confidences. For countries like Brazil, the predictions were not steady and were incorrect far out, had wavier confidence. As such, the hybrid model Chronos x Ensemble GBT is highly accurate and reliable. However, policymakers can only use the information to make long-term decisions for stable countries.

5.5.2 Probabilistic Regression in the Chronos Phase

During the Chronos phase, probabilistic regression is adopted, where a range of potential future outcomes as opposed to a single point estimate is determined. This helps reduce uncertainties in the forecasting process by specifying confidence intervals around the forecasted values. As shown in the figure, the blue shaded area depicts a range of uncertainty for the forecast; meanwhile, the forecast itself is depicted as a solid line. Consequently, this helps lower the uncertainty in predictions and provides automatically reliable GDP forecasts.

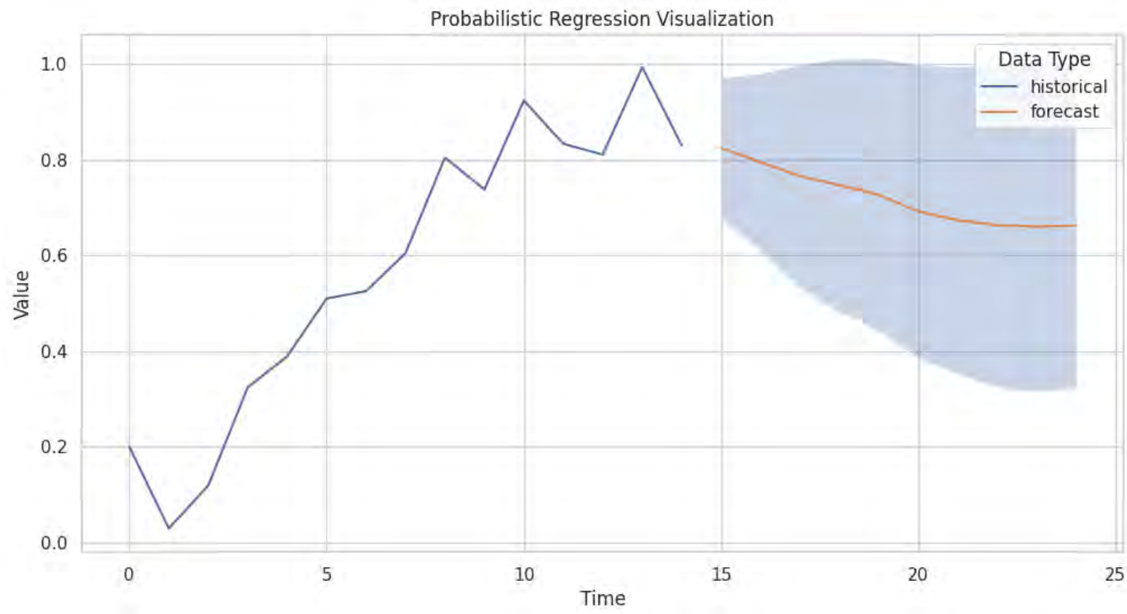

Figure 5.48: Probabilistic Regression Visualization in the Chronos Phase

The probabilistic regression visualization is displayed on Figure 5.48:

- The latest and historical data are represented as the blue line.
- The forecast is represented with an orange line.
- The blue shaded area shows the forecast's confidence intervals.

So, here we can see that integrating probabilistic regression increased confidence in these predictive models, which are suitable for long-term GDP forecasting.

Chapter 6

Discussion

The results of this study show that the single model ensemble (which includes XGBoost, LightGBM, CatBoost and Chronos) as well as combination with proposed hybrid approach significantly improve GDP forecasting performance compared to univariate baseline models [19]. By harnessing the power of such models, not only were we able to develop novel modeling techniques for better prediction results but also entrepreneurial AI which is valuable in dissecting multi-layered economies and unpackaging complex non linear relationships impacting GDP growth through a unification approach. Gradient Boosting was the sample learning algorithm, and they were used with powerful regularization techniques which can efficiently prevent overfitting [10]. This is very useful in managing economic datasets because you almost always encounter missing data sparsity and/or noise with them. Further, the robustness of XGBoost in hyperparameter tuning allow it to generalize well enough on these complex interactions between multiple macro variables for GDP prediction accuracy. On top of a leaf-wise tree growth algorithm and histogram-based layout, it also supports processing economic scale data with high efficiency. The impressive speed and scalability of this model made it ideal for forecasts on real-time economic data [30]. LightGBM has proven to be robust in handling outliers and sharp changes in economic environments, for instances recession or unexpected booms hence making its forecasts more consistently stable over the medium-to-long-term [44]. The large dataset with high dimensionality was handled well using which even the subtle trends in data were captured accurately. Since there were a few categorical features like different trade policies, regulatory changes and business confidence indices, the ordered boosting technique from CatBoost was useful. Such this ability to directly work with raw categorical features without much preprocessing allowed CatBoost in adding qualitative economic factors into the forecasting model [17]. In addition to this, the novel approach of Catboost also reduces biases in the predictions leading higher trustworthiness for the ensemble as whole when it come down to predicting GDP across different economic scenarios. The combination of these models into an ensemble yielded a robust forecasting tool that leveraged the unique strengths each model possessed. [18] The noise-handling ability of XGBoost, efficiency in dealing with large datasets from LightGBM, Categorical feature handling by CatBooost, Temporal insights provided by Chronos along with the model complexity balance achieved through this Hybrid Model gave an end-to-end solution for GDP forecasting [48]. The ensemble approach also helped to improve generalization and avoid overfitting by enabling the predictions to account for both short-term economic cycles as well as longitudinal non-linear patterns. Overall, the ensemble model was an excellent tradeoff between high accuracy and low computational efficiency making a strong solution for real-world GDP forecasting challenges [32]. It is a time series forecasting library which helped us capture temporal dependencies within the GDP data. Its application of time-series data is optimized for large-scale and the model was able to capture both short-term fluctuations as well establish long-run trend in GDP. Chronos, in this way acts as the bridge connecting multiple time-series models; ranging from traditional statistical methods to more advanced machine learning frameworks — such that they can be appropriately stepping stones for each of their strengths (functions used). Given the dynamic nature of GDP, it is crucial to be able to capture how this changes over time in order for us to understand its trajectory and make informed economic forecasts.[46]. The hybrid model provided an additional level of complexity by incorporating machine learning and deep learning models together. Hybrid model helped capture the highly non-linear and temporal patterns of this case study for a comprehensive understanding of GDP behavior [54]. The inference here is that the model could more accurately predict GDP when these both short-term and long-term dynamics are at work aka high economic volatility.

Chapter 7

Limitation and Future Work

7.1 Limitations and Challenges

Specialized time-series models like Chronos exhibit a considerably higher degree of accuracy, as is evidenced by the considerably lower error metrics (MSE of 3.79e+09, RMSE of 6.15946e+4, MAE of 18153.64, and MAPE of 0.05). They can provide a much better estimation of the instruments controlling patterns or affordances of temporally determined phenomena. In addition, the significant improvement of the hybrid model, with its key differentiator being the integration of ensembling and inherent time-series, shows that time-series models specialized for economy, such as Chronos integrated with the ensemble GBT, provide the highest level of accuracy. An overview of the error metrics of Chronos x Ensemble GBT model supports this statement, verifying the model's considerable predictive capacity at the MSE of 7.632454e+09, RMSE of 87363.92, MAE or 21216.76, and R^2 of 0.99. Ensemble methods lead to the creation of robust, highly effective models that can be easily applied. However, in the case of economic indicators with temporal patterns and potentially observable fluctuations. It means that the use of time-series specialized models like Chronos with the additional incorporation of ensemble GBT is likely to produce better, more accurate results overall. Thus, when developers aim to address such instruments and models present in the data when planning the design of models, all changes must be carefully reviewed to ensure that they do not impact the effectiveness of predictions.

7.2 Future Work

Although in this study, we have developed a novel hybrid model for GDP forecasting-GDP, which is based on Chronos for both univariate and multivariate time series forecasting and Ensemble Gradient Boosting Trees, there are still several appealing directions for the future development. For example, we limited the scope of the data that were utilized for developing the model to the Penn World Table. However, it is possible that in the future research, additional economic data that are provided by the World Bank or International Monetary Fund, for instance, would be adopted to enhance the performance, as well as generalizability of the model. It would be also interesting to consider for the other projects of a similar kind, especially, the integration of more complex and nuanced variables to have a more sophisticated forecast. Specifically, including not only major economic indicators, such as real

GDP expenditure, but also the variables like consumer sentiment indexes, or even geopolitical risk factors, would help in this regard. Finally, it should be noted that it would be beneficial to advance the hybrid Chronos-Ensemble GBT model in terms of its hyperparameters, as well as an ensembling strategy. Specifically, it would be interesting to adopt such advanced optimization techniques as genetic algorithms, or Bayesian optimization, which would help in establishing highly efficient hyperparameters, as well as ensemble learning model. In this way, it might be possible not only to equal but even exceed the efficiency and accuracy of the developed model as of now.

Chapter 8

Conclusion

Therefore this paper suggests an ensemble hybrid model of XGBoost, LightGBM, CatBoost and Chronos with a factor to process current GDP in forecasting the future value for state economy. In combination, the strengths of each model create a more accurate, robust and scalable forecasting platform that accounts for both non-linear relationships between economic variables over different time scales through GDP growth. While the regularization methods in XGBoost models, efficiency on large datasets by LightGBM, ability to handle categorical data of CatBoost and deep learning-based approach to time-series forecasting by Chronos all adds precision in GDP forecasts under selective economic conditions. This ensemble is subsequently enhanced by the hybrid model which integrates machine learning and deep leaning predictive power to offer a more complete perspective of short term economic fluctuations as well as long-term trends. These are huge advancements especially when it comes to addressing the challenges of computational complexity, risk for overfitting as well as a lack—for many algorithms—of interpretation. Nevertheless, as the ensemble approach does promise better identification of diverse GDP drivers than another alternative, real-world initialising with ensembles is recommended. As economic climates shift and change, these models are helpful tools for policymakers representing important bodies of empirical knowledge.Indeed they help economists around the world have access to pieces of evidence that will enable them make better sense in a more complex global economy.

Bibliography

[1] R. M. Salgado, J. J. Pereira, T. Ohishi, R. Ballini, C. Lima, and F. J. Von Zuben, "A hybrid ensemble model applied to the short-term load forecasting problem," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, IEEE, 2006, pp. 2627–2634.

[2] C. Jones, *Macroeconomics*. Norton, 2008, ISBN: 9780393926385. [Online]. Available: https://books.google.com.bd/books?id=nhypPAAACAAJ.

[3] T. Basu, D. Barik, and P. Arokiasamy, "Demographic determinants of economic growth in brics and selected developed countries," in *XXVII IUSSP international population conference (IUSSP 2013)*, 2013, pp. 26–31.

[4] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] M. A. Din, "Arima by box jenkins methodology for estimation and forecasting models in higher education," 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:53537865.

[7] S. Athey *et al.*, "The impact of machine learning on economics," *The economics of artificial intelligence: An agenda*, pp. 507–547, 2018.

[8] B. Weng, W. Martinez, Y.-T. Tsai, *et al.*, "Macroeconomic indicators alone can predict the monthly closing price of major us indices: Insights from artificial intelligence, time-series analysis and hybrid models," *Applied Soft Computing*, vol. 71, pp. 685–697, 2018.

[9] Y. Kurihara, A. Fukushima, *et al.*, "Ar model or machine learning for forecasting gdp and consumer price for g7 countries," *Applied Economics and Finance*, vol. 6, no. 3, pp. 1–6, 2019.

[10] X. Liao, N. Cao, M. Li, and X. Kang, "Research on short-term load forecasting using xgboost based on similar days," in *2019 International conference on intelligent transportation, big data & smart city (ICITBS)*, IEEE, 2019, pp. 675–678.

[11] M. Miah, M. Tabassum, and M. S. Rana, "Modelling and forecasting of gdp in bangladesh: An arima approach," *Journal of Mechanics of Continua and Mathematical Science*, vol. 14, no. 3, pp. 150–166, 2019.

[12] P. Premraj, "Forecasting gdp growth: A comprehensive comparison of employing machine learning algorithms and time series regression models," M.S. thesis, 2019.

[13] P. Probst, "Hyperparameters, tuning and meta-learning for random forest and other machine learning algorithms," Ph.D. dissertation, lmu, 2019.

[14] S. I. Abba, N. T. T. Linh, J. Abdullahi, *et al.*, "Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration," *IEEE Access*, vol. 8, pp. 157 218–157 237, 2020.

[15] J. Allen and G. Kristy, *Analytical approaches to macroeconomic forecasting: A study of profits through machine learning and deep learning*, 2020.

[16] R. S. Arneja, N. Kaur, and A. Saihjpal, "Analyses and forecasting evaluation of gdp of india using arima model," *International Journal of Advanced Science and Technology*, vol. 29, no. 11, pp. 1102–1107, 2020.

[17] J. Ding, Z. Chen, L. Xiaolong, and B. Lai, "Sales forecasting based on catboost," in *2020 2nd international conference on information technology and computer application (ITCA)*, IEEE, 2020, pp. 636–639.

[18] Y. Niu, "Walmart sales forecasting using xgboost algorithm and feature engineering," in *2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, IEEE, 2020, pp. 458–461.

[19] S. Qureshi, B. M. Chu, and F. S. Demers, "Forecasting canadian gdp growth using xgboost," Carleton University, Department of Economics, Tech. Rep., 2020.

[20] K. M. Ramírez, J. M. Hormaza, and S. V. Soto, "Artificial intelligence and its impact on the prediction of economic indicators," in *Proceedings of the 6th International Conference on Engineering & MIS 2020*, 2020, pp. 1–8.

[21] K. M. Vu, "Sources of growth in the world economy: A comparison of g7 and e7 economies," in *Measuring economic growth and productivity*, Elsevier, 2020, pp. 55–74.

[22] H. Bunyamin *et al.*, "Classical and deep learning time series prediction techniques in the case of indonesian economic growth," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 1077, 2021, p. 012 014.

[23] S. Elsayed, D. Thyssens, A. Rashed, L. Schmidt-Thieme, and H. S. Jomaa, "Do we really need deep learning models for time series forecasting?" *CoRR*, vol. abs/2101.02118, 2021. arXiv: 2101.02118. [Online]. Available: https://arxiv.org/abs/2101.02118.

[24] A. Hossain, M. Hossen, M. M. Hasan, and A. Sattar, "Gdp growth prediction of bangladesh using machine learning algorithm," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, IEEE, 2021, pp. 812–817.

[25] J. Lv, C. Wang, W. Gao, and Q. Zhao, "[retracted] an economic forecasting method based on the lightgbm-optimized lstm and time-series model," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 8 128 879, 2021.

[26] D. K. Padhi, N. Padhy, A. K. Bhoi, J. Shafi, and M. F. Ijaz, "A fusion framework for forecasting financial market direction using enhanced ensemble models and technical indicators," *Mathematics*, vol. 9, no. 21, p. 2646, 2021.

[27] K. Prachyachuwong and P. Vateekul, "Stock trend prediction using deep learning approach on technical indicator and industrial specific information," *Information*, vol. 12, no. 6, p. 250, 2021.

[28] Y. Tay, M. Dehghani, J. Rao, *et al.*, "Scale efficiently: Insights from pre-training and fine-tuning transformers," *arXiv preprint arXiv:2109.10686*, 2021.

[29] K. Uddin and N. Tanzim, "Forecasting gdp of bangladesh using arima model," *International Journal of Business and Management*, vol. 16, no. 6, pp. 56–65, 2021.

[30] Y. Yang, Y. Wu, P. Wang, and X. Jiali, "Stock price prediction based on xgboost and lightgbm," in *E3s web of conferences*, EDP Sciences, vol. 275, 2021, p. 01 040.

[31] J. Yoon, "Forecasting of real gdp growth using machine learning models: Gradient boosting and random forest approach," *Computational Economics*, vol. 57, no. 1, pp. 247–265, 2021.

[32] A. Asati, "A comparative study on forecasting consumer price index of india amongst xgboost, theta, arima, prophet and lstm algorithms.," 2022.

[33] Y. Granitsa, "Assessment of factors of regional economic stability using the xgboost algorithm: 018," in *Dela Press Conference Series: Economics, Business and Management*, 2022, pp. 7–7.

[34] Z. Habibi, H. Habibi, and M. A. Mohammadi, "The potential impact of covid-19 on the chinese gdp, trade, and economy," *Economies*, vol. 10, no. 4, p. 73, 2022.

[35] M. A. Khan, K. Abbas, M. M. Su'ud, *et al.*, "Application of machine learning algorithms for sustainable business management based on macro-economic data: Supervised learning techniques approach," *Sustainability*, vol. 14, no. 16, p. 9964, 2022.

[36] S. Park and J.-S. Yang, "Interpretable deep learning lstm model for intelligent economic decision-making," *Knowledge-Based Systems*, vol. 248, p. 108 907, 2022.

[37] A. T. Vaswani, D. Yogatama, D. Metzler, *et al.*, "Scale efficiently: Insights from pre-training and fine-tuning transformers," 2022.

[38] X. Wang, Y. Mao, Y. Duan, and Y. Guo, "A study on china coal price forecasting based on ceemdan-gwo-catboost hybrid forecasting model under carbon neutral target," *Frontiers in Environmental Science*, vol. 10, Sep. 2022. DOI: 10.3389/fenvs.2022.1014021.

[39] W. Xiang, P. Xu, J. Fang, Q. Zhao, Z. Gu, and Q. Zhang, "Multi-dimensional data-based medium-and long-term power-load forecasting using double-layer catboost," *Energy Reports*, vol. 8, pp. 8511–8522, 2022.

[40] J. M. Ahn, J. Kim, and K. Kim, "Ensemble machine learning of gradient boosting (xgboost, lightgbm, catboost) and attention-based cnn-lstm for harmful algal blooms forecasting," *Toxins*, vol. 15, no. 10, p. 608, 2023.

[41] V. Hassija, V. Chamola, A. Mahapatra, *et al.*, "Interpreting black-box models: A review on explainable artificial intelligence," *Cognitive Computation*, vol. 16, Aug. 2023. DOI: 10.1007/s12559-023-10179-8.

[42] D. S. d. O. S. Júnior, P. S. de Mattos Neto, J. F. de Oliveira, and G. D. Cavalcanti, "A hybrid system based on ensemble learning to model residuals for time series forecasting," *Information Sciences*, vol. 649, p. 119 614, 2023.

[43] T. Kim, B. Ha, and S. Hwangbo, "Online machine learning approach for system marginal price forecasting using multiple economic indicators: A novel model for real-time decision making," *Machine Learning with Applications*, vol. 14, p. 100 505, 2023.

[44] A. C. C. Sousa, T. G. do Rêgo, Y. d. A. M. Barbosa, T. de Menezes, *et al.*, "Comparing gradient boosting algorithms to forecast sales in retail," in *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional*, SBC, 2023, pp. 596–609.

[45] H. Xue and F. D. Salim, *Promptcast: A new prompt-based learning paradigm for time series forecasting*, 2023. arXiv: 2210.08964 `[stat.ME]`. [Online]. Available: https://arxiv.org/abs/2210.08964.

[46] A. F. Ansari, L. Stella, C. Turkmen, *et al.*, *Chronos: Learning the language of time series*, 2024. arXiv: 2403.07815 `[cs.LG]`. [Online]. Available: https://arxiv.org/abs/2403.07815.

[47] A. F. Ansari, L. Stella, C. Turkmen, *et al.*, "Chronos: Learning the language of time series," *arXiv preprint arXiv:2403.07815v2*, 2024.

[48] S. Bharathi and N. Navaprakash, "Prediction of indian gdp using xgboost compared with adaboost," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, IEEE, 2024, pp. 1–4.

[49] A. Carriero, D. Pettenuzzo, and S. Shekhar, *Macroeconomic forecasting with large language models*, 2024. arXiv: 2407.00890 `[econ.EM]`. [Online]. Available: https://arxiv.org/abs/2407.00890.

[50] A. Carriero, D. Pettenuzzo, and S. Shekhar, "Macroeconomic forecasting with large language models," *arXiv preprint arXiv:2407.00890*, 2024.

[51] N. Gruver, M. Finzi, S. Qiu, and A. G. Wilson, *Large language models are zero-shot time series forecasters*, 2024. arXiv: 2310.07820 `[cs.LG]`. [Online]. Available: https://arxiv.org/abs/2310.07820.

[52] S. Hamiane, Y. Ghanou, H. Khalifi, and M. Telmem, "Comparative analysis of lstm, arima, and hybrid models for forecasting future gdp," *Journal homepage: http://iieta. org/journals/isi*, vol. 29, no. 3, pp. 853–861, 2024.

[53] M. Jin, S. Wang, L. Ma, *et al.*, *Time-llm: Time series forecasting by reprogramming large language models*, 2024. arXiv: 2310.01728 `[cs.LG]`. [Online]. Available: https://arxiv.org/abs/2310.01728.

[54] H. Xie, X. Xu, F. Yan, X. Qian, and Y. Yang, "Deep learning for multi-country gdp prediction: A study of model performance and data impact," *arXiv preprint arXiv:2409.02551*, 2024.

[55] M. Zhu, Y. Zhang, Y. Gong, K. Xing, X. Yan, and J. Song, *Ensemble methodology:innovations in credit default prediction using lightgbm, xgboost, and localensemble*, 2024. arXiv: 2402.17979 `[cs.CE]`. [Online]. Available: https://arxiv.org/abs/2402.17979.

[56] F. Zhu and F. Huang, "Data-rich economic forecasting for actuarial applications," *Available at SSRN 4756743*,