Forecasting the Sector wise Gross Domestic Product of India

MSc Research Project
Data Analytics

Ramanan Loganathan
Student ID: 21202702

School of Computing
National College of Ireland

Supervisor: Furqan Rustam

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ramanan Loganathan
Student ID:	21202702
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Furqan Rustam
Submission Due Date:	14/08/2023
Project Title:	Forecasting the Sector wise Gross Domestic Product of India
Word Count:	8933
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	18th September 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Forecasting the Sector wise Gross Domestic Product of India

Ramanan Loganathan
21202702

Abstract

This analysis delves into the significance of temporal sequence prediction techniques for forecasting India's Gross Domestic Product (GDP), a pivotal economic metric. The dataset encompasses historical GDP values across eight distinct sectors and GDP at factor cost, presenting a comprehensive overview of the economy. The primary objective is to employ and assess diverse forecasting approaches to accurately predict GDP trends. The initial phase involves meticulous data preprocessing, including managing missing values, scaling data, and partitioning it into training and testing subsets. Four distinct methodologies, specifically SARIMA (Seasonal Autoregressive Integrated 'Moving Average), ARIMA (Autoregressive Integrated Moving Average), RNN (Recurrent Neural Network), and LSTM (Long Short-Term Memory), were utilized. The 'statsmodels' library is utilized for SARIMA and ARIMA, while 'keras' facilitates the implementation of RNN and LSTM models. The training process involves constructing SARIMA and ARIMA models on the training set, followed by forecasting GDP values for the test set. Performance metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE) evaluate the precision of the predictions. For RNN and LSTM, a sequence-based training approach is adopted. The findings of this examination underscore the efficacy of the models in predicting GDP values across sectors. Notably, the LSTM model demonstrates superior precision, boasting the lowest RMSE among the models. The implications of this analysis are substantial, offering valuable insights for decision-makers in various economic sectors. The robustness of these forecasting models assists in generating well-informed projections, thereby contributing to strategic deliberations and policy formulation. Essentially, this research illuminates the pivotal role of accurate GDP prediction and advocates for the strategic integration of suitable forecasting models based on data attributes. The outcomes pave the path for future advancements in economic forecasting, highlighting the ever-increasing significance of data-driven insights in shaping economic paths.

1 Introduction

A significant measure that captures a country's overall economic performance is its gross domestic product (GDP) *International Monetary Fund)* (2023) . India possesses one of the largest and most rapidly expanding economies globally, projecting its GDP is essential for a range of individuals, including policymakers, firms, stakeholders, and the wider community. Accurate GDP forecasts aid in the development of sensible decisions and efficient economic strategies, as well as useful insights into the nation's economic health *International Monetary Fund)* (2023). The following factors such as Economic Planning and Policy Formulation, Fiscal Management, Investment Decisions, Business Strategy, Employment and Labor Market makes predicting India's GDP essential *Economic Times)* (2023).

As policymakers rely on GDP predictions to evaluate the current health of the economy, identify growth trends, and make strategic decisions to promote economic stability and growth, accurate GDP forecasts are essential for economic planning and policy formulation (Mohsin et al.; 2022). Fiscal management heavily relies on GDP forecasts, to preserve fiscal deficit, governments must synchronize their spending plans and income forecasts based on the anticipated GDP growth. International and domestic investors both largely rely on GDP projections to assess prospective investment possibilities*investopedia)* (2023). Economic growth forecasts are a vital source of data for risk analysis and strategic investment choices. Understanding the future economic picture is essential for firms operating in India to establish successful business plans for resource allocation, product development, and business expansion. The labour market and employment prospects are affected by accurate GDP estimates. Estimating job possibilities and the number of workers needed in various industries requires an understanding of GDP patterns. Predicting the sector-wise GDP is also as important like predicting overall GDP because sector-specific policies, diversified economy, investment allocation, identifying growth driver and risk management can be addressed to make better policies.

Different economic sectors contribute differently to the expansion of the overall GDP. Policymakers can develop sector-specific policies to solve issues and take advantage of development possibilities by projecting GDP at the sector level. India's economy is varied, with numerous industries including manufacturing, services, agriculture, and more. Predictions made at the sector level give a thorough understanding of the dynamics of the economy and the potential for growth in particular industries. Investors can identify high-growth industries and allocate resources accordingly with the help of sector-specific GDP estimates. By reducing risks and optimizing investment portfolios, this information is helpful. Forecasting GDP on a sectoral basis helps in identifying the main forces driving economic expansion. It enables firms and politicians to prioritize strategic projects and concentrate on the industries with the most potential. The ability to estimate risk for specific industries is aided by precise GDP predictions made by sector. This enables businesses to prepare for economic downturns in particular industries and create backup plans. So, to address these problems machine learning algorithms can be used to forecast the sector wise GDP of India.

The goal of the research is to evaluate how well different machine learning algorithms estimate the GDP of various sectors in India using a wide range of economic data. By training algorithms like Random Forest, LSTM, ARIMA, and SARIMA on historical GDP data broken down by sector and relevant economic indicators like industrial production, inflation rates, trade balances, and more, this study intends to examine the predicting powers of these algorithms. The study seeks to accomplish two objectives through this investigation: first, to establish which machine learning algorithm provides the most accurate and trustworthy GDP projections; and second, to identify the major economic factors that significantly affect GDP fluctuations per sector. The goal of the study is to shed light on the potential of machine learning models as macroeconomic analytical predictors and to advance knowledge of the intricate connections between economic indicators and sector-specific GDP performance in the context of India's economy.

After referring to a dozen of previous works related to this problem, the niche of this research is to predict the sector-wise GDP of India, analyzing its debts and GDP growth and existing policies and its impact which will help in modeling new policies that help to improve GDP growth and repayment of debts. This study uses SARIMA, ARIMA, RNN, and LSTM, four advanced time series forecasting models, to forecast India's GDP by sector. The analysis attempts to provide useful insights into economic patterns, educate policy choices, and enable stakeholders to make knowledgeable judgements when navigating the changing Indian economy.

- Forecasting the sector-wise GDP of India using machine learning models
- Finding which machine learning algorithm can forecast accurate results

2 Related Work

The literature evaluation prioritizes prior research, and the publication (Li et al.; 2022) explores how high-precision GDP forecast technology enables the development of sustainable regional resource usage and the recommendation of economic management policies. In order to create a precise GDP forecast model, this study developed a novel multi-predictor ensemble decision framework based on deep reinforcement learning. Then, using the characteristics of GRU, TCN, and DBN as the main predictors, three GDP forecasting models are trained. The DQN algorithm then effectively evaluates how well these three neural networks may be combined to construct an ensemble model on different GDP datasets. The ensemble weight coefficients of these three neural networks were adaptively optimized using the DQN algorithm to get the final GDP forecast results. This paper's citations make it easier to approach regional GDP data and giving better understanding how to apply machine learning algorithms to sector wise data.

In this work (Hossain et al.; 2021), the forecasting of GDP growth utilizing additional variables—such as GDP per capita, inflation rate, public debt, total investment, remittances, and unemployment rate—was the main topic of discussion. A machine learning method is used to estimate GDP Growth Rate after learning the complex correlations between GDP Growth Rate and other characteristics. This could help everyone better understand economics and help economists validate their economic projections. As a result, it is straightforward to pinpoint potential tactics for accelerating the targeted GDP development. This paper can also help to illustrate our eco-social future scenario, help to set economic goals for Bangladesh, and help to pinpoint the variables that directly affect GDP growth in Bangladesh the most and those that have the least direct influence and are responsible for stifling it. Since India's population and demography are comparable, using the information in this research can help create a machine learning model that is stable.

Further study on other papers that helps in policy making 2.1 and machine Learning in GDP prediction 2.2.

2.1 Policy Making

The study on "One Belt and One Road" (Han and Li; 2018), which develops a multivariate regression model to define the important variables impacting GDP per capita and provide a more precise projection model, is referenced in order to further determine the impact of new policies. According to research, the "One Belt and One Road" initiative has had a considerable and rapid increase in the value of Xinjiang's GDP per capita. This shows that the "One Belt and One Road" policy may help to support Xinjiang's economic development. A possibility for Xinjiang to prosper is provided by the "One Belt and One Road" plan. Long-term, the Silk Road Economic Belt will be the main focus of China's economic growth. Xinjiang is one of the top five provinces in northwest China in terms of economic output, resource endowment, geographic advantage, and rate of development. Xinjiang is the most significant political position on the silk route economic belt since it is located in the frontier and center of the region.

In order to make decisions in real time, policy makers frequently rely on imprecise information about the status of the economy, as was covered in the study (Richardson et al.; 2018). This report claims that many significant statistics are constantly changed and provided slowly from a further distance. Because nowcasting models are increasingly in demand as instruments for lowering some of these uncertainties, forecasters at various central banks and other organizations have used them extensively (Giannone et al. 2008, Banbura et al. 2013, Jansen et al. 2016, Bloor 2009). Due to advances in computing power, machine learning (ML) approaches have recently been proposed as potential alternatives to the time-series regression models that central banks typically use to forecast significant macroeconomic indicators. When there are more potential regressors than there are observations, ML models are especially skilled at handling large datasets. ML techniques to get precise nowcasts of the real gross domestic product growth in New Zealand for the current quarter. We use multiple vintages of historical GDP data and multiple vintages of a large features set, which includes over 550 domestic and international variables, to evaluate the real-time performance of these algorithms across the 2009Q1-2018Q1 timeframe. The forecasts generated by these algorithms are then compared to a benchmark's forecasting accuracy, which was determined using naive autoregressive analysis and other data-rich methodologies like a factor model, a Bayesian VAR (BVAR), and a number of statistical models used by the RBNZ. This paper is the first to assess the relative nowcast performance of several ML approaches using real-time data. This study looked at how effectively various ML algorithms produce accurate forecasts of New Zealand's real GDP growth for the quarter in question. The evaluation of the real-time performance of these algorithms over the 2009Q1–2018Q1 period utilizing several eras of historical GDP data and multiple vintages of a substantial feature set, which contains over 550 domestic and foreign variables. The forecasts generated by these algorithms are then compared to the forecasting accuracy of a benchmark naïve autoregressive method and other data-rich techniques such as a factor model, a Bayesian VAR (BVAR), and a group of statistical models used by the RBNZ. This study is unique in that it compares how well various ML algorithms perform when used to make nowcasts using real-time data. The majority of ML models, according to the results, produce point nowcasts that are superior to the traditional AR benchmark. Other high-performing models, including as support vector machines, Lasso, and neural networks, can reduce average nowcast errors by around 16–28% compared to the AR benchmark. This investigation adds to a growing corpus of research examining how well machine learning (ML) models forecast data when compared to more traditional time-series techniques.

The research conducted (Aprigliano et al.; 2019) found that payment data mirrors economic activity. We examine numerous aggregates of flows in the Italian payment system and other indicators commonly employed in macroeconomic forecasting and find that they still contain some additional information. LASSO selects the payments together with other conventional business cycle indicators (such industrial production and company surveys), for both GDP and household expenditure, starting with a big database of short-term monthly variables. Additionally, in an out-of-sample forecasting application using retail payment flows as a forecasting input, a mixed frequency factor model outperforms the one based exclusively on conventional short-term indications. In addition to GDP, results are shown for consumption, investments, and value added in the service sector. It is found that the weights attached to PS are comparable to those of some of the most significant short-term indicators typically used to track economic activity after estimating the weights suggested in Koopman and Harvey (2003) in order to separate the contributions of the observable variables to the forecasts of GDP. our study showed that, compared to other comparable short-term indicators, the mixed-frequency feature of our model boosts forecast accuracy during the quarter.

2.2 Machine Learning in GDP Prediction

A machine learning framework for growth prediction was produced by this study Bang et al. (2015). serve as a sort of manual for using machine learning to solve economic problems in addition to establishing this framework. Though machine learning as a concept is not new, advancements in computing technology and a greater comprehension of how it may be used to solve economic problems have made it a new tool for economists (Varian, 2014). The study's challenges and questions were stated in part 2, and the suggested approaches were offered in section 3. The data and some of its issues are described in Section 4's information. The results are reported in Sections 5 and 6, together with some concluding remarks on the possible policy implications and future research initiatives. There was discussion of factors that both directly and indirectly affected GDP projection, including geography, political institutions, religion, market distortions and performance, investment and its composition, dependence on primary products, trade, market orientation, and colonial history. The market value of all goods and services produced by the economy during the measurement period is captured by the gross domestic product (GDP), which also includes private inventories, paid-in construction expenses, individual consumption, government purchases, and the balance of international trade (exports minus imports). The spending approach, the production approach, and the income approach are the three traditional approaches for estimating GDP that are used in this work(Abonazel and Abd-Elftah; 2019). According to Ning et al. (2010), the GDP issue has become the most important among macroeconomic variables, and GDP data is acknowledged as a key indicator for assessing macroeconomic health in general and national economic progress. It is frequently regarded as the most accurate indicator of the state of the economy. It serves as a crucial foundation for the government's initiatives and plans for economic development. In this work, time series analysis statistical techniques are used to estimate and forecast Egypt's GDP. The four phases of the Box-Jenkins technique are applied in order to develop an acceptable ARIMA model for the Egyptian GDP, which is used to forecast the Egyptian GDP for the subsequent ten years (from 2017 to 2026).

This article's (Jena et al.; 2021) main contribution is the development of an ANN model that can forecast GDP for eight major economies one quarter in advance. This model captures the nonlinearities included in the quarterly time-series data and provides accurate forecasts. These countries' health has suffered greatly as a result of the ongoing COVID-19 pandemic. Various forms of lockdown have been implemented by governments to contain the outbreak since infection and mortality rates have become concerning. As a result, their economies have suffered, which has caused a number of industries to close their doors and a rise in unemployment rates. If policymakers have a comprehensive grasp of the future economic prediction, they will be better prepared to take the proper response in such a situation. For instance, there are varying opinions in the US Congress over the amount of aid required to bolster the economy. The ANN model given in this study correctly predicted the GDP values since the MAPE is less than 2% in each of the country examples.

Although more frequent updates and data broken down by geographic unit are frequently required for policy decisions, this paper (Ortega-Bastida et al.; 2021) covers how official economic indicators are computed and published yearly at the national level by National Offices of Statistics around the world. The article presents a methodology that combines historical GDP data from Twitter with historical GDP data using a multimodal technique based on a recurrent neural network in order to estimate GDP for regions or any other geographic unit at any frequency. The approach is evaluated for predicting quarterly GDP for four regions of Spain using a sizable dataset. When compared to standard and cutting-edge approaches, the results show improved forecasting accuracy. The paper also thoroughly evaluates several designs and embedding strategies, tests the method's resilience, and assesses hyperparameters. Additionally, the research assesses the feasibility of the provided method to predict the effects of COVID-19 on local economies using tweets from the first two quarters of 2020.

In this work (Schumacher and Breitung; 2008), factor models were employed to anticipate German GDP using mixed-frequency real-time data with various statistical publication lags. According to Stock and Watson (2002a), missing observations have been handled using the EM approach and principle components decomposition. The EM method provides us with monthly estimates of the variables and also makes numerous recommendations for how to produce iterative and direct forecasts for a low-frequency variable like GDP based on high-frequency variables. The capability of the factor model for short-term forecasting was evaluated using recursive forecast comparison. The empirical application makes use of a medium-sized real-time dataset for post-unification Germany, which has about fifty monthly and quarterly time series. The accuracy of forecasting is not significantly affected by data alterations, according to our empirical findings. The mixed-frequency factor model outperforms factor models based on balanced data by a small margin. More large variances are visible when real-time factor forecasts are compared to straightforward benchmark models. A study of forecast mistakes over time demonstrates that although the differences across the factor model estimations are quite small, they do exist. The overall moderate forecast performance of the factor models indicates that German GDP forecasting in real time is a difficult task. However, this outcome is in line with earlier studies on the recent decline in forecast accuracy.

LSTM have displayed outstanding potential in time-series estimation, especially when dealing with lengthy sequences with time-related connections. This Study (Zhang et al.; 2022) employed Lester to forecast Consumer Price Index (CPI) in Indonesia, highlighting its ability to capture intricate inflation designs. It expanded LSTM's usage to optimum hedging strategies, showcasing its flexibility in intricate financial examinations. Despite these individual contributions, the incorporation of LSTM-HMM for Gross Domestic Product (GDP) projection remains an under explored domain and the method presented in the present analysis, utilizing Lester to anticipate real-time CPI fluctuation statuses and subsequently forecasting GDP fluctuation statuses, introduces an innovative technique that connects the gap between profound learning and statistical modeling. Machine learning, in particular RNNs, are being used in economic forecasting to show off their potential for capturing intricate temporal correlations and managing structural discontinuities. RNNs have proven to be adept in modelling long-term dependencies in sequences, which makes them a good choice for capturing the complex dynamics of economic variables. This is seen from the study's (Longo et al.; 2022) focus on the contribution of RNNs to improving forecast accuracy in times of crisis, such as the 2008–2009 financial crisis and the most recent COVID–19 recession. The use of integrated gradients as an interpretive tool is a fascinating aspect of the current investigation. This strategy clarifies the relative weights of various model components, enabling a clearer comprehension of how particular indicators affect estimates of GDP growth. The model's transparency is increased by the inclusion of interpretable machine learning techniques, making it potentially useful for policymakers trying to understand the causes of economic fluctuations brought on by various shocks.

3 Methodology

The proposed methodology starts by scrapping the raw data from different comma separated files and making it into different sector wise file by python functions as part of data preprocessing step. Further dividing the data into test and train to apply the different machine learning model to predict the sector wise GDP of India is the high level architecture of this study as shown in 1.

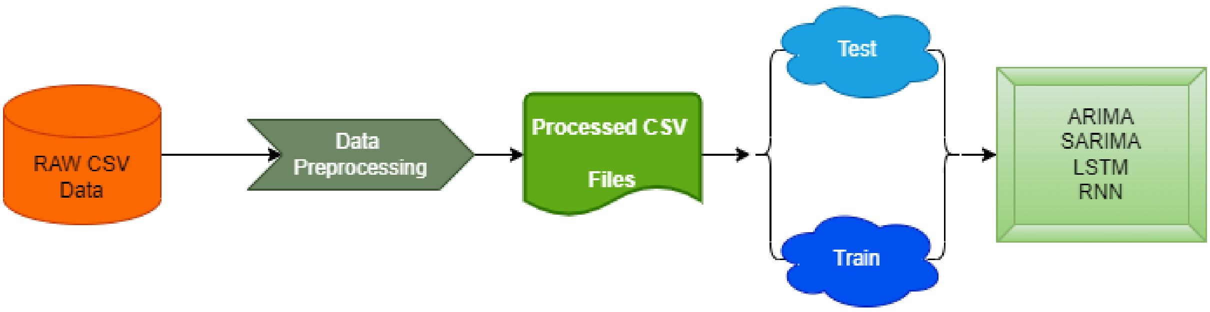

Figure 1: Architecture of this Study

3.1 Data Description

As part of initial analysis, the idea was to take data from world bank and NITI AAYOG website but, after implementing those data in a machine learning model the results are not appropriate and unreliable. All these data were yearly outcomes and starting from 1970's hardly 45 years of data was only available so, the machine learning models were not able to process the data properly which resulted in the shortcomings in the output. Later the quarterly data from the Ministry of Statistics and Implementation Programme had the quarterly GDP data of different sectors. This dataset had Gross Domestic Product (GDP) figures for various industries across different quarters for a span of 15 years. The industry column specifies the particular industry or sector to which the GDP values pertain as shown in the figure 2. Each industry represents a distinct sector of the economy, such as agriculture, mining, manufacturing, trade, construction, financing, and community services. The Q1, Q2, Q3 and Q4 are the Quarterly GDP values for each industry, representing economic output for each quarter of the year (Q1: January-March, Q2: April-June, Q3: July-September, Q4: October-December). The dataset comprises GDP values for multiple industries, offering insights into the contribution of each sector to the overall economy and for each industry, the respective annual growth was also provided. Researchers and policymakers can analyze the GDP data to understand the relative importance of different industries in the economy, track growth rates, and identify sectors that may require targeted interventions. This data can be utilized to develop predictive models for GDP growth, enabling informed forecasts about the economic trajectory of specific industries.

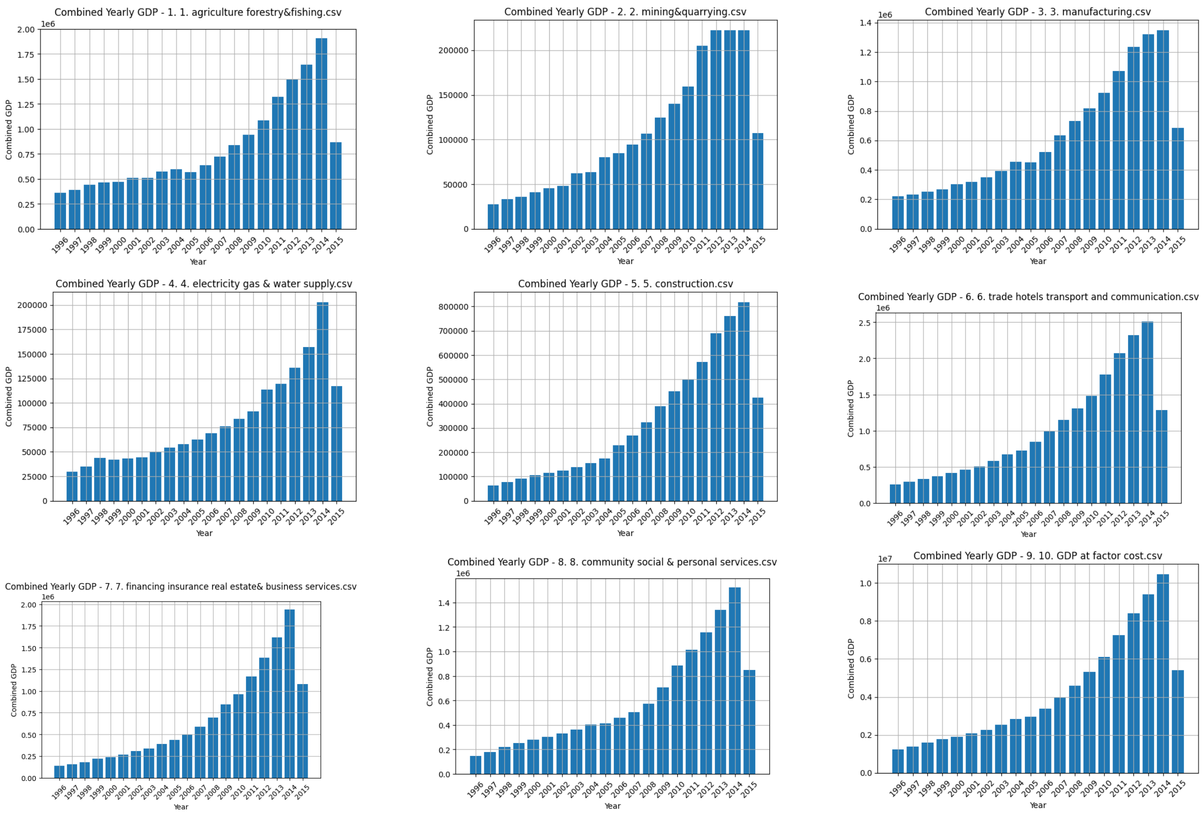

Figure 2: GDP data of different sectors

3.2 Data Preprocessing

The initial data preprocessing aimed at converting an initial raw GDP dataset into a structured and more manageable format that can be fed to the machine learning models. The preprocessing steps involved the transformation of the dataset into separate industry-specific CSV files, each document the time-series data of GDP values for distinct quarters of respective years. The process starts by designating input and output file paths from google cloud. Later on, the code iterates through the dataset's rows, corresponding to different industries and extracts relavant data, excluding annual figures, and for each industry, establishes a new dataframe. This newly built dataframe, denoted as 'industry_df,' is devised to house of processed data that featuring 'Time' and 'GDP' columns and through a precise iteration across quarters, the code computes corresponding dates and refines GDP values. These processed data entries were appended to the industry-specific dataframe. After finalizing the processing of all quarters for a given industry, the code persists the precisely curated information by saving the 'industry_df' dataframe as a distinct CSV file. This resultant file, appropriately named with the industry's index and name which is situated within the designated output directory. By doing this data preprocessing step, the code fundamentally transforms the raw GDP dataset into a coherent structure, primed for subsequent analytical exploration, modeling endeavors, and in-depth interpretation of GDP trends across various sectors of the economy.

3.3 Future Forecasting Models

Future forecasting models such as ARIMA, SARIMA, LSTM and RNN has been used to forecast the sector wise GDP of India. Along with these four discussed models other forecasting models such Exponential Smoothing, LASSO and Random Forest Regressor were implemented to check if they provide better forcasted values than the discussed models.

3.4 ARIMA

As ARIMA is the well-known model for time series forecast, So the very first approach is to try ARIMA in this project. The process begins with data preparation, where historical GDP data is loaded from a processed CSV file and the time information is formatted as datetime and set as the index of the dataset. Next, the data is split into training and testing sets, in which approximately 80% was used for training set and the remaining 20% for testing. The core of the methodology lies in the ARIMA model training and forecasting so, the model is constructed and trained using the training data, with specific parameters (1, 1, 1) representing the autoregressive, differencing, and moving average components. Later, the trained model is deployed to generate forecasts for the test data, covering the number of periods equivalent to the test set. Next the confidence intervals are calculated to provide a measure of uncertainty around the forecasted values which was achieved by utilizing the 'conf_int' method, which computes 95% confidence intervals for the forecasted GDP values. These confidence intervals gave an indication of the range within which the actual values are likely to fall. For a comprehensive evaluation, the methodology computes error metrics to assess the accuracy of the forecast that too specifically, the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were calculated, quantifying the deviation between the forecasted and actual GDP values. The results are then presented, starting with the sector name and parallelly the forecasted GDP values, confidence intervals, differences between actual and forecasted values, and the percentage difference are displayed for the test set.

$$ARIMA \tag{1}$$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

3.5 SARIMA

The SARIMA GDP forecasting model requires a systematic process for accurate predictions so as part of intial step, the processed data is loaded from the specified file path and organized, with the 'Time' column converted to datetime format and set as the index. This data is then partitioned into a training set, constituting 70% of the data, and a test set for evaluating the model's performance. After that, the SARIMA model is configured by determining the appropriate order (p, d, q) to account for autoregressive, differencing, and moving average components, along with a seasonal order (P, D, Q, S) to capture seasonal patterns. The model is trained on the training data to estimate parameters. Forecasting for the test set commences and parallely with the trained SARIMA model generating predictions for the GDP values. The difference between actual and predicted GDP values is calculated, and the percentage difference is calculated to assess forecast accuracy. The results are presented in a different dataframe showcasing forecasted and actual values, differences, and percentage differences. Apart from that, the performance evaluation metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE) are calculated to quantify predictive precision. The final step involves displaying the sector name alongside calculated error metrics, facilitating a clear understanding of the model's forecasting capabilities.

$$SARIMA \tag{2}$$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t + \Phi_1 y_{t-s} + \Theta_1 \varepsilon_{t-s}$$

3.6 LSTM

After applying the two statistical models ARIMA and SARIMA, the next ideas was to implement the deep learning model. The process begins with data preparation, involving the loading of preprocessed GDP data from a CSV file and the time column is formatted as datetime and set as the index of the dataset. To facilitate training and testing, the data is partitioned into training and testing sets, typically with a split ratio of 80:20, where 80% of the data is utilized for training the model and the remaining 20% for assessing its predictive performance. The core of the methodology lies in the construction and training of the LSTM model which involves designing a neural network architecture comprising LSTM layers, which are particularly suited for capturing sequential dependencies in time series data. Hyperparameters such as the number of LSTM units, the number of layers, and the batch size are configured to optimize the model's predictive ability and also the LSTM model is trained using the training data, with the aim of learning patterns and relationships within the historical GDP values. During training, the model adjusts its internal parameters iteratively to minimize the difference between predicted and actual GDP values and after training the model is ready for forecasting. The test data, which represents unseen future GDP values, is fed into the trained LSTM model and then the model processes the sequential data and generates predictions for the corresponding GDP values. These predictions are compared with the actual GDP values to evaluate the model's accuracy and performance. And like the previuos statistical models evaluation metrics like Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are commonly calculated to assess the deviation between the predicted and actual GDP values were used.

$$
\begin{aligned}
&LSTM &(3)\\
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)\\
\tilde{C}_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)\\
C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)\\
h_t &= o_t \odot \tanh(C_t)
\end{aligned}
$$

3.7 RNN

Apart from statistical and deep-learning models, the Recurrent Neural Networks(RNN) is also an another machine learning technique that can be used to forecast to future. Like the three previous methodologies this process also begins by preprocessing the GDP data, transforming the time series into a suitable format, and splitting it into training and test sets. The training data is normalized to enhance model performance so that the data get splitted into Sequential input-output pairs are created from the training data, ensuring the model captures temporal dependencies. An RNN model, consisting of an LSTM layer followed by a dense layer, is constructed and trained on the prepared training sequences. The model's performance is evaluated on the test set by predicting future GDP values and the predicted values are then inverse-transformed to obtain real GDP values. Key evaluation metrics such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) are calculated to test the model's accuracy and further presents the forecasted GDP values alongside actual values in tabular form, facilitating comparison.

$$
\begin{aligned}
&RNN &(4)\\
h_t &= \tanh(W_{hx}x_t + W_{hh}h_{t-1} + b_h)
\end{aligned}
$$

3.8 Evaluation

The application of RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and MSE (Mean Squared Error) evaluations in the context of ARIMA, SARIMA, RNN, and LSTM time series models serves as a crucial framework for assessing the performance and accuracy of these forecasting techniques. RMSE, a widely used metric, measures the average magnitude of prediction errors, offering insights into the model's overall forecasting precision, which can be used in the case of ARIMA and SARIMA models, which leverage historical data patterns, RMSE helps quantify the deviation between predicted and actual GDP values. In the RNN and LSTM models, RMSE helps in measuring the effectiveness of the neural networks in capturing complecated temporal relationship within the data. Through offering an easy-to-understand indicator of prediction accuracy, MAE complements RMSE and this helps for better visualization of prediction quality across all four models and its an important measure for determining how closely forecasts match actual GDP statistics. MSE gives a better insights into the level of forecasting error variability in the ARIMA and SARIMA models, which is highly capable of handling seasonality and trends. In the case of RNN and LSTM, MSE helps in evaluating the model's ability to reduce prediction discrepancies, which excel at capturing complicated temporal patterns.

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

4 Implementation

The initial idea was to execute the code in Jupyter notebook using python as it offers good visualizations and easy to use later on the idea was changed to use Google colab instead of Jupyter notebook for the following reasons. As google colab is a cloud based python notebook so that the code can saved in cloud and it can be retrived anywhere at any point of time and also all the latest python packages were availabe in it. Apart from this it provides free access to Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), which are powerful hardware accelerators that significantly speed up machine learning computations.

The implementation starts by importing necessary libraries, including pandas for data handling, numpy for numerical computations, matplotlib for visualization, and statsmodels for building the ARIMA model. The dataset is loaded from the processed CSV file, where the time column is converted to datetime format, and the index is set to the time column and the data is split into training and testing sets, with 80% used for training and the remaining 20% for testing. In turn this split allows us to train the model on historical data and evaluate its performance on unseen data and the 'order' variable is defined as (1, 1, 1), representing the order of the ARIMA model (p, d, q) on which the values are chosen based on domain knowledge or model tuning. The ARIMA model is built using the 'ARIMA' class from statsmodels where the training data is used to fit the model, and the 'order' variable is passed as an argument. After fitting the trained data, the model is ready to make predictions that are made on the test data using the 'get_forecast' method of the fitted model. The number of steps in the forecast is determined by the length of the test data and the confidence intervals are calculated using the 'conf_int' method to provide a range of possible values for the predictions. A confidence level of 95% is commonly used, but this can be adjusted as needed. The actual and predicted GDP values, as well as the confidence intervals, are extracted from the forecast object and these values are later used to calculate the Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) using respective functions from scikit-learn. The actual and predicted GDP values, along with the confidence intervals, are plotted using Matplotlib for visualization where the x-axis represents time, and the y-axis represents GDP values as shown in fig(). These plots allow to visually assess how well the model's predictions align with the actual data and the accuracy of these predictions.

After importing necessary libraries, including pandas for data handling, numpy for numerical computations, and statsmodels for building the SARIMA model, the dataset is loaded from the provided CSV file, and the time column is converted to datetime format. Post that the index is set to the time column, like the ARIMA model. Then the data is divided into training and testing sets, with approximately 70% of the data used for training and the remaining 30% for testing, here the training set is reduced because with these data its giving an reliable predictions. As per the traditional process the SARIMA model is constructed using the SARIMAX class from statsmodels and the order and seasonal_order parameters are set based on model tuning. These parameters define the autoregressive, differencing, moving average, and seasonal components of the model. The model is fitted to the training data using the fit method, which estimates the model parameters based on the provided data. As the model is trained and predictions are generated for the test set using the get_forecast method of the fitted model. The number of itteration in the forecast is respective to the length of the test data which is 30%. To capture the uncertainty around the predictions the confidence intervals are calculated using conf_int and confidence level of 95% is commonly used wich indicates that the actual values are likely to fall within the calculated intervals. The forecasted GDP values, along with their corresponding confidence intervals, are extracted from the forecast object and these values are then utilized to calculate performance metrics such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) using functions available in scikit-learn. To visualize the model's predictions, both the actual and forecasted GDP values are plotted over time using Matplotlib. The plot also includes shaded regions representing the confidence intervals, which help quantify the uncertainty of the forecasts.

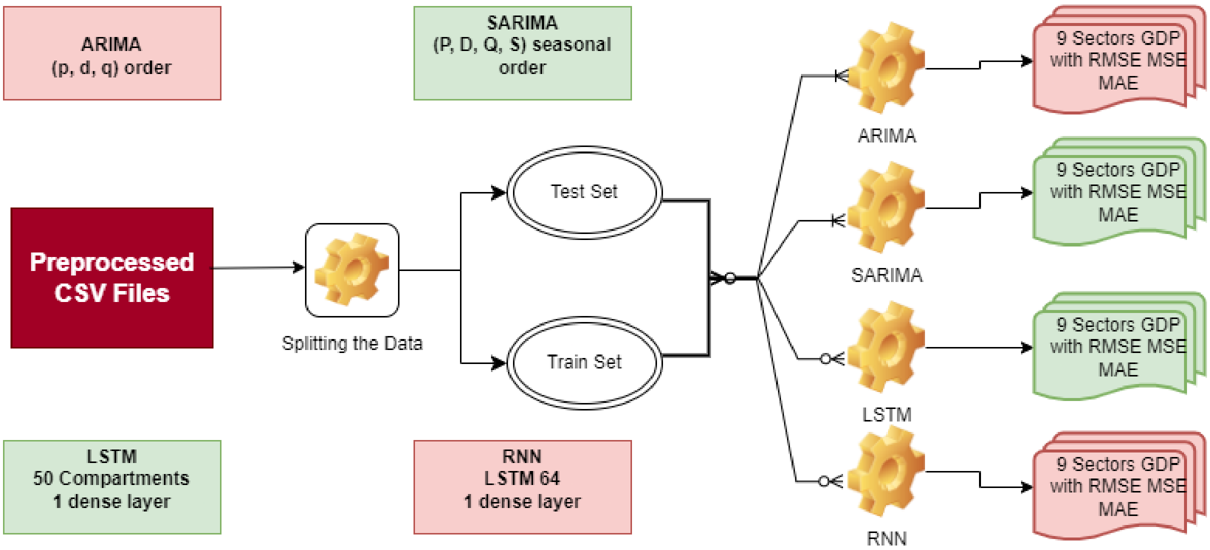

Figure 3: Implementation of ARIMA, SARIMA, LSTM & RNN Models

Like the other two models the initial steps are followed for LSTM as well. To enhance the scaling process and model stability during training and to ensure that the LSTM captures underlying patterns more effectively the MinMaxScaler from scikit-learn is employed to transform the GDP values into a range between 0 and 1. Next, the dataset is divided into training and testing sets, following an approximate split of 70% for training and 30% for testing and the training data is further prepared by creating input sequences (X) and their corresponding target GDP values (y) in which each input sequence is designed with a window size of 5 representing five quarter), provides historical context for the model to learn from. The LSTM model architecture is constructed using Keras, a high-level neural network API and it comprises of two LSTM layers, each with 50 units, to capture sequential patterns in the data where the second LSTM layer returns only the output for the last time step. A dense output layer with a single unit predicts the next GDP value. To adapt the learning rate during training, the model is compiled with the mean squared error loss function and the Adam optimizer. The model is trained using the prepared training data and during training, the model iteratively adjusts its internal weights to minimize the difference between predicted and actual GDP values, the training is performed over a specified number of epochs, where one epoch represents a complete pass through the training dataset. Similar to training data preparation, input sequences (X) are created for the test data followed by training and the trained LSTM model then generates predicted GDP values for the corresponding input sequences. Finally, the predicted and actual GDP values are plotted using Matplotlib for visualization, the time (quarters) is represented on the x-axis, while GDP values are shown on the y-axis by which the model's ability to capture trends and patterns in the data becomes visually apparent.

In addition to the initial procedures, TensorFlow and Keras are used for constructing the RNN model. To design a facilitated creation of input-output pairs suitable for training an RNN create_sequences function is used which accepts two arguments one is a dataset ('data') and another one sequence length ('seq_length'). This function systematically extracts subsequences from the dataset, each of length 'seq_length', along with the corresponding target value that follows each subsequence and essentially constructs the temporal context necessary for sequential data modeling. After that the key hyperparameters are set for the RNN model. The 'seq_length' parameter determines the length of input sequences, the 'n_epochs' parameter specifies the number of training epochs, and the 'batch_size' parameter defines the number of samples utilized in each training iteration. A loop that iterates through pairs of 'file_path' and 'sector' using the 'zip' function was created and for each pair, a CSV file is read into a Pandas DataFrame. The dataset is then partitioned into training and testing subsets with a split ratio of 70-30. As part of scaling process, the training data is normalized using the MinMaxScaler from scikit-learn which brings all GDP values into a range between 0 and 1 and aids in stabilizing the model training. The RNN model is built using Keras' Sequential API which consists of an LSTM layer with 64 units, a specialized type of RNN designed for sequential data. The input shape is defined as '(seq_length, 1)', signifying the shape of each input sequence and additionally, a fully connected dense output layer with a single unit is appended to predict the upcoming GDP value. Widely used optimization algorithm such as mean squared error (MSE) loss function and the Adam optimizer is used for compilation of the model. Using scaler technique both the predicted and actual normalized GDP values are inverse transformed to interpret predictions in terms of actual GDP values. To evaluate the performance of the model metrics like Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) are computed. The figure 3 explains the implementation of these four models.

5 Evaluation

Evaluation metrics such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) were used to evaluate the accuracy and reliability of the implemented models Rustam et al. (2020). The following reasons prove the necessity of these metrics: Quantifying Prediction Accuracy, Comparing Models, Identifying Overfitting or Underfitting, Tuning Model Parameters and Communicating Results.

Sector	RMSE	MAE	MSE
Agriculture, Forestry & Fishing	**171,033.76**	**149,518.86**	**29,252,547,848.03**
Mining & Quarrying	**9,967.10**	**8,489.84**	**99,343,120.89**
Manufacturing	**41,481.33**	**39,073.49**	**1,720,700,977.10**
Electricity, Gas & Water Supply	**12,547.17**	**9,706.56**	**157,431,419.45**
Construction	**36,706.85**	**34,252.45**	**1,347,392,988.43**
Trade, Hotels, Transport & Communication	**157,052.60**	**142,753.41**	**24,665,518,195.24**
Finance, Insurance, Real Estate & Business Services	**91,615.74**	**75,623.69**	**8,393,443,484.60**
Community, Social & Personal Services	**104,947.31**	**88,336.35**	**11,013,937,374.04**
GDP at Factor Cost	**531,571.08**	**490,692.46**	**282,567,808,179.67**

Table 1: **ARIMA Evaluation Results**

5.1 ARIMA Evaluation

The evaluation results of the ARIMA model[Table] for GDP prediction across various sectors reveals the insights to its forecasted value performance and the analysis hold within the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE) metrics. Sectors like "Mining & Quarrying" and "Electricity, Gas & Water Supply" shows good prediction accuracy when compared to other sectors, as its indicates lower RMSE and MAE values, implying that the model's forecasts align closely with actual data for these sectors. In contrast, sectors like "Trade, Hotels, Transport & Communication" and "GDP at Factor Cost" demonstrate larger prediction errors, suggesting a relatively larger divergence between the model's projections and the actual values. Adding to this, the graphical representation of the predicted and actual GDP values further prove that these findings are reliable. Overall, the RMSE, MAE, and MSE metrics, along with the graphical analysis, provide a complete understanding of the ARIMA model's strengths and limitations for different sectors, and also it paves for the next model building to get improved results.

5.2 SARIMA Evaluation

The evaluation outcomes of the SARIMA model for GDP prediction, performed across various sectors, shows a valuable insights into its forecasting efficiency and the evaluation is based on key metrics such as the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). As the results show sector-wise performance trends, the sectors like "Mining & Quarrying" and "Construction" demonstrates a notably lower RMSE and MAE, which shows the higher precision between the model's predictions and actual data. Opposing that, sectors like "Trade, Hotels, Transport & Communication" gives a relatively higher prediction errors, which was proven by the larger RMSE and MAE values, implying that the model's projections exhibit greater divergence from actual values in these sectors. The graphical visualization further highlights these findings by presenting a visual comparison between the predicted and actual GDP values. The Graphical representation in [] confirms the numeric results, helping to

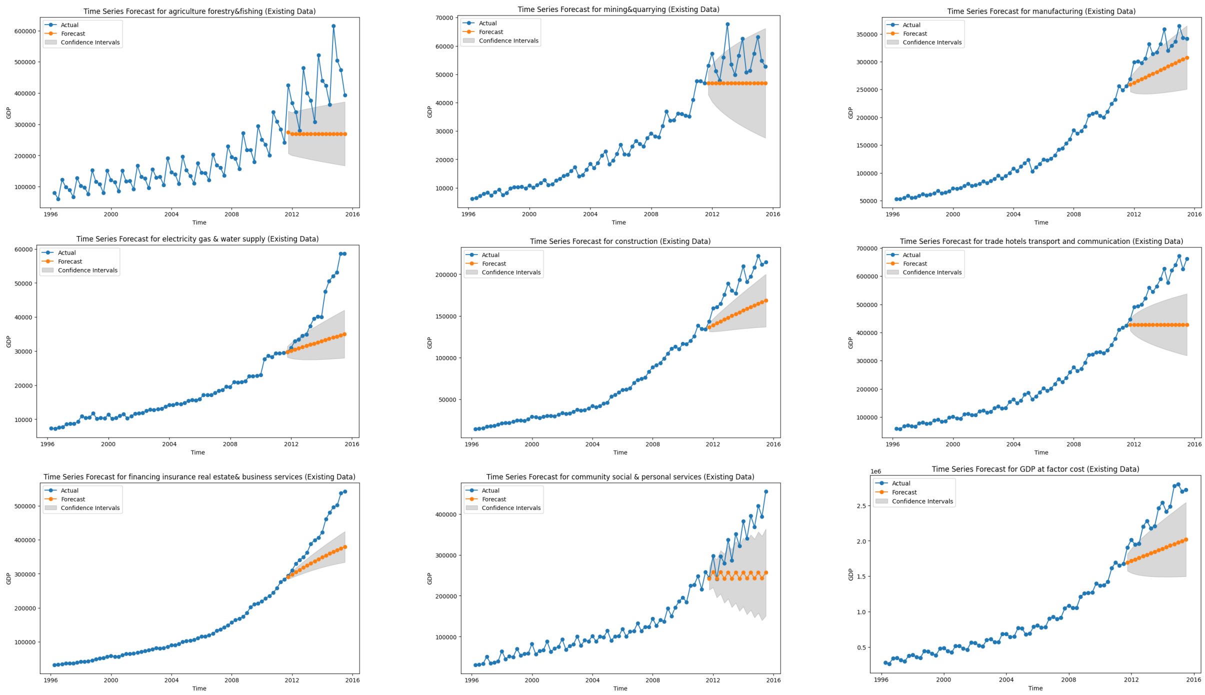

Figure 4: Sector wise graphical representation of ARIMA model's forecast

collectively assess the SARIMA model's predictive capabilities. Overall with the help of [] and [] says that the implemented SARIMA model is reliable and forecast the values effectively than the ARIMA statistic model.

Sector	RMSE	MAE	MSE
Agriculture, Forestry & Fishing	76767.28	60574.85	5893214657.54
Mining & Quarrying	5447.48	4287.72	29675084.05
Manufacturing	20587.15	16416.63	423830754.15
Electricity, Gas & Water Supply	11185.55	8471.28	125116436.59
Construction	8353.77	7230.68	69785461.68
Trade, Hotels, Transport & Communication	23731.00	18176.82	563160167.32
Finance, Insurance, Real Estate & Business Services	24818.66	18446.19	615965738.00
Community, Social & Personal Services	59736.08	50150.79	3568398679.00
GDP at Factor Cost	131198.44	107882.34	17213030693.30

Table 2: **SARIMA Evaluation Results**

5.3 LSTM Evaluation

After the significant results from SARIMA, the expirement was carried in LSTM as well to achieve better GDP forecast. Same like previous models evaluation metrics including the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared

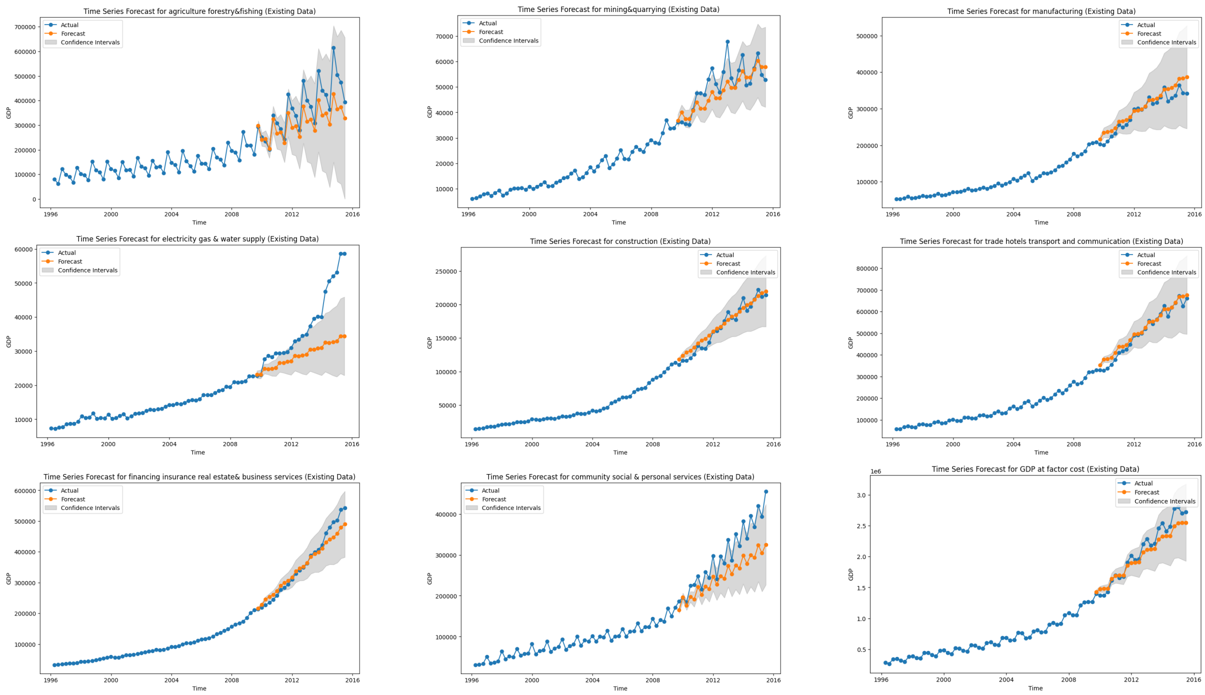

Figure 5: Sector wise graphical representation of SARIMA model's forecast

Error (MSE) were calculated [] which in turn offer a comprehensive assessment of the model's predictive accuracy. The outcomes of this analysis across the sectors reveal surprising trends. For example, the "Electricity, Gas & Water Supply" has more precise predictions in this domain sector when compared to other sectors as it has relatively lower RMSE and MAE values. On the other hand, the "Manufacturing" and "Trade, Hotels, Transport & Communication" sectors manifest comparatively larger RMSE and MAE which shows that there was a greater variance between the model's projections and actual observations. Along with this the graphical representations [] of visual comparison between predicted and actual GDP values proves the same. The visual analysis aligns with the quantitative evaluation further helps in understanding the sectors that perform good prediction and other sectors that might require data cleansing for improved results. These further pushes to implement the model in other machine learning technique to achieve better results.

5.4 RNN Evaluation

Though the above three significant results but at the sametime the results are not significant for all the sectors. The evaluation is carried out by essential metrics like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE), which collectively offers a strong assessment of the model's forecasting accuracy. Investigating sector-specific results[] reveals that the predicted results are in a intresting trends. With the lower RMSE and MAE values it can be concluded that the "Electricity, Gas & Water Supply" sector's prediction provides an trustable predictions where as the "Manufacturing" and "Trade, Hotels, Transport & Communication" sectors stands out from other sectors prediction as it has highest RMSE and MAE values. Visual analysis [], with graphs providing a direct comparison of expected and actual GDP levels across sectors, strengthens the interpretive potential of these measures and this visual review

Figure 6: Sector wise graphical representation of LSTM model's forecast

Sector	RMSE	MAE	MSE
Agriculture, Forestry & Fishing	40961.27	33616.78	1677825289.04
Mining & Quarrying	5758.88	4966.68	33164711.77
Manufacturing	14341.92	12706.33	205690708.68
Electricity, Gas & Water Supply	3921.82	2701.25	15380705.79
Construction	8967.86	7381.15	80422449.36
Trade, Hotels, Transport & Communication	34712.57	27785.76	1204962485.00
Finance, Insurance, Real Estate & Business Services	21218.04	18903.31	450205208.72
Community, Social & Personal Services	34744.70	30556.67	1207194446.69
GDP at Factor Cost	260151.78	215122.42	67678948691.69

Table 3: **LSTM Evaluation Results**

confirms the quantitative assessment, emphasizing potential areas for model improvement and highlighting positives. A thorough evaluation of the forecasting effectiveness of the RNN model is provided by the combination of RMSE, MAE, MSE metrics and the graphical representations.

Figure 7: Sector wise graphical representation of RNN model's forecast

5.5 Discussion

Among the multiple models implemented for GDP prediction, the SARIMA model became the most reliable model 8 as it gives the reliable forecast for most of the sectors, showcasing its efficiency in capturing complex economic trends of different sector GDP. Except for the "Electricity Gas & Water supply" sector the forecasted value is very close to the actual value and the difference in percentage is also less for all the other sectors. Followed by the LSTM model which gives a reliable forecast for six sectors, RNN gives only two and ARIMA gives for only one sector. Apart from these model's other time series models such as Exponential Smoothing, LASSO and Random Forest Regressor were also implemented but these models did not give the expected results, so it was not included in the report. The Exponential Smoothing model, known for its simplicity, efficiency, and ability to capture underlying trends, faced difficulties when applied to sector wise GDP prediction. Considering the dynamic and frequently non-stationary character of economic data, it may be less accurate in predicting trends with irregular trends because of its assumption of stationary nature. Moreover, the model's reliance on previous observations and exponentially decreasing weights might struggle to fit in with the sudden shifts or structural changes in the data. Similarly, the Random Forest Regressor, which is considered as one of the most powerful machine learning techniques, also encountered challenges like overfitting because the data might need little tuning. The Lasso Regression model outperformed might be because of the insufficient subset of features.

Sector	RMSE	MAE	MSE
Agriculture, Forestry & Fishing	69335.56	59653.04	4807419726.38
Mining & Quarrying	10371.88	8656.46	107575990.65
Manufacturing	75662.92	66777.14	5724877753.83
Electricity, Gas & Water Supply	1869.75	1520.01	3495950.12
Construction	44261.40	39641.67	1959071846.07
Trade, Hotels, Transport & Communication	116358.54	99962.31	13539309562.32
Finance, Insurance, Real Estate & Business Services	68948.21	62820.95	4753856184.11
Community, Social & Personal Services	24827.55	20703.16	616407029.40
GDP at Factor Cost	347486.74	294986.38	120747034748.26

Table 4: **RNN Evaluation Results**

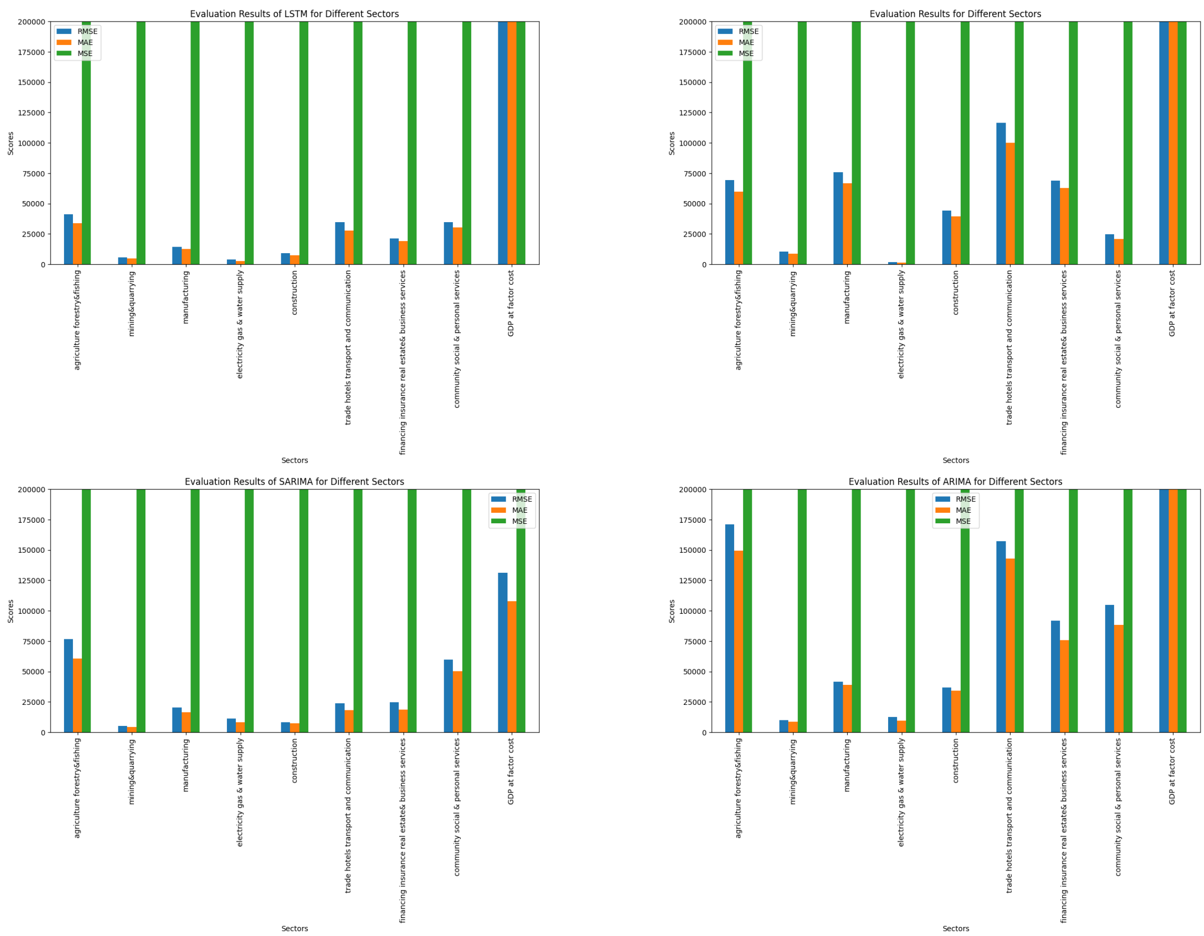

Figure 8: Models performance comparison

6 Conclusion and Future Work

In the study of GDP prediction, multiple forecasting models were implemented and thoroughly examined. Among them, the SARIMA model stands out as reliable due to minimal percentage differences and effectiveness across sectors. The SARIMA model had more accuracy because forecasted values closely aligned with actual values and had very less differences in percentage across sectors. Following SARIMA, the LSTM model showcases adaptiveness in complex economic data prediction. Although less reliable, the RNN model offers credible forecasts for two sectors, and ARIMA performs well for one. Exponential Smoothing, LASSO, and Random Forest Regressor were tested but excluded. Exponential Smoothing cannot able to handle the irregular stationarity in economic data, impacting accuracy. Random Forest encountered overfitting issues, while LASSO's inefficiency with fewer subsets led to unexpected results.

- SARIMA and LSTM models gives the best sector wise GDP forecasting results
- RNN and ARIMA models are providing the expected results.

In conclusion, as mentioned above the SARIMA model is the most efficient model because of its reliability, which gives accurate and consistent GDP predictions across different sectors of the Indian economy. As this study was undertaken to forecast the sector-wise GDP using machine learning techniques, the main aim was to build a suitable model to forecast the GDP values SRIMA model was concluded as the best performing model which can help the different sector investors either from inside the country or foreign investors which in turn provides employment opportunities for millions of youths in INDIA and it further helps to boost the emerging economy. Accurate forecasts can help in policy decisions that were made for specific sectors, helping in targeted interruption and promoting economic stability. Also helps businesses by benefiting from sector-specific planning, aligning resources with forecasted GDP trends to improve the sector's operations and manage risks. Apart from this long-term planning and transparent forecasting improves India's demand for foreign investment, maintaining economic progress and technological advancement. All the above study's forecast contributes significantly to India's economic growth, attracting investment, guiding policy, and ensuring sustainable development. As SARIMA and LSTM give the most reliable results among the other models, a combined model of both can implement in future to forecast the sector-wise GDP of India. Models such as Exponential Smoothing and LASSO may be implemented with Standardization, Stationarity, and other required preprocessing steps.

References

Abonazel, M. R. and Abd-Elftah, A. I. (2019). Forecasting egyptian gdp using arima models, *Reports on Economics and Finance* **5**(1): 35–47.

Aprigliano, V., Ardizzi, G. and Monteforte, L. (2019). Using payment system data to forecast economic activity, *60th issue (October 2019) of the International Journal of Central Banking* .

Bang, J. T., Sen, T. and Basuchoudhary, A. (2015). New tools for predicting economic growth using machine learning: a guide for theory and policy, *Southern Economic Association Conference*.

Economic Times) (2023). https://economictimes.indiatimes.com/news/economy/policy/one-factor-that-can-put-india-ahead-of-the-us-as-the-2nd-biggest-economy/articleshow/101668658.cms?from=mdr.

Han, X. and Li, G. (2018). The development trend and prediction model of xinjiang gdp in the context of "one belt and one road", *2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 468–471.

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

International Monetary Fund) (2023). https://www.imf.org/en/Publications/fandd/issues/Series/Back-to-Basics/gross-domestic-product-GDP.

investopedia) (2023). https://www.investopedia.com/articles/investing/121213/gdp-and-its-importance.asp#:~:text=GDP%20enables%20policymakers%20and%20central,and%20tax%20and%20spending%20plans.

Jena, P. R., Majhi, R., Kalli, R., Managi, S. and Majhi, B. (2021). Impact of covid-19 on gdp of major economies: Application of the artificial neural network forecaster, *Economic Analysis and Policy* **69**: 324–339.
URL: *https://www.sciencedirect.com/science/article/pii/S0313592620304604*

Li, Q., Yu, C. and Yan, G. (2022). A new multipredictor ensemble decision framework based on deep reinforcement learning for regional gdp prediction, *IEEE Access* **10**: 45266–45279.

Longo, L., Riccaboni, M. and Rungi, A. (2022). A neural network ensemble approach for gdp forecasting, *Journal of Economic Dynamics and Control* **134**: 104278.
URL: *https://www.sciencedirect.com/science/article/pii/S016518892100213X*

Mohsin, M., Naseem, S., Sarfraz, M. and Azam, T. (2022). Assessing the effects of fuel energy consumption, foreign direct investment and gdp on co2 emission: New data science evidence from europe & central asia, *Fuel* **314**: 123098.

Ortega-Bastida, J., Gallego, A. J., Rico-Juan, J. R. and Albarrán, P. (2021). A multimodal approach for regional gdp prediction using social media activity and historical information, *Applied Soft Computing* **111**: 107693.
URL: *https://www.sciencedirect.com/science/article/pii/S1568494621006141*

Richardson, A., Mulder, T. et al. (2018). Nowcasting new zealand gdp using machine learning algorithms.

Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B.-W., Aslam, W. and Choi, G. S. (2020). Covid-19 future forecasting using supervised machine learning models, *IEEE Access* **8**: 101489–101499.

Schumacher, C. and Breitung, J. (2008). Real-time forecasting of german gdp based on a large factor model with monthly and quarterly data, *International Journal of Forecasting* **24**(3): 386–398.
URL: *https://www.sciencedirect.com/science/article/pii/S0169207008000393*

Zhang, J., Wen, J. and Yang, Z. (2022). China's gdp forecasting using long short term memory recurrent neural network and hidden markov model, *Plos one* **17**(6): e0269529.