A

PROJECT REPORT

(PS04CAST53)

ON

"Forecasting GSDP of Maharashtra state by using time series analysis "

DEPATMENT OF STATISTICS

SARDAR PATEL UNIVERSITY

VALLABH VIDYANAGAR, ANAND-388120

PROJECT GUIDE

Dr. JYOTI M. DIVECHA

Dr. KHIMIYA TINANI

Dr. AGNIVA DAS

Mr. TANAJI MANE

Miss. RUPALI KHARAT

SUBMITTED BY

Mr. LAKHAN DEVIDAS JADHAV

M.Sc. APPLIED STATISTICS (SEMESTER IV)

BATCH 2024-2025

CERTIFICATE

This is to certify that Mr. Lakhan D. Jadhav student of Master of Science in Applied Statistics, Roll No. 05 has satisfactorily completed her Project work on **"Forecasting GSDP of Maharashtra state by using time series analysis "** for M.Sc. (Applied Statistics) semester IV during the Dec 2024 – May 2025.

Place: Vallabh Vidyanagar

Date:

Dr. Jyoti M. Divecha	**Dr. Khimiya Tinani**	**Dr. Agniva Das**
(Head of Department)	**(Project Guide)**	**(Project Guide)**
&		
(Project Guide)		

Mr. Tanaji Mane	**Miss. Rupali Kharat**
(Project Guide)	**(Project Guide)**

ACKNOWLEDGEMENT

This year has been an extremely informative journey for, my friend and me. We would like to extend our gratitude to **Dr. Jyoti M. Divecha** (Head of department), **Dr. Khimiya Tinani, Dr. Agniva Das, Mr. Tanaji mane and Miss. Rupali Kharat** for entrusting upon me these invaluable projects. The journey of the study at the department and the projects gave me immense insight the world of analytics. Documentation is heart of project, so we take opportunity to express my heartfelt thanks to all my dear friends who support and encourage my project partner and me to complete our documentation successful. This project has been the outcome of ideas of combination of suggestions and contribution of many people. Our project is dedicated to all the people whom we met, took guidance interviewed and something from them. At this occasion, we want to grab this opportunity to acknowledge our sincere thanks to all of them while submitting.

Lakhan Devidas Jadhav (Applied Statistics 05)

Place: V.V Nagar, Anand

Date:

PREFACE

It is great opportunity for me to have MASTER OF APPLIED STATISTICS IN SARDAR PATEL UNIVERSITY, VALLABH VIDYANAGAR. In the accomplishment of this degree, I am doing a project report on **"Forecasting GSDP of Maharashtra state by using time series analysis "**. Subject to the limitation of time efforts and resources, every possible attempt has been made to study the problem deeply. The whole project is measured through the secondary data, the data further analyzed and interpreted and the result was obtained.

A practical knowledge in a student’s life is very important. It helps a student to know the real-life situation and problems of life. Theoretical knowledge is very much needed but practical knowledge is equally important. The practical knowledge to student is given in a form of project.

This project provides an opportunity and platform to know the current situation and the behavior of environment.

For the preparation of this project, I feel deep sense of gratitude to all faculty members, staff members of the respective organization and all other persons who helped me to prepare such project report.

TABLE OF CONTENTS

Abstract

This study focuses on forecasting the Gross State Domestic Product (GSDP) of Maharashtra using Time series analysis**.** Maharashtra, India's largest state by economic output has undergone significant economic transformations, with major contributions from manufacturing, information technology (IT), finance and real estate**.** The study assesses sectoral contributions, evaluates the impact of COVID-19 and develops accurate forecasting models**.**

The research utilizes annual GSDP data from 2000 to 2023, sourced from RBI and Maharashtra government reports. Exploratory Data Analysis (EDA) identifies trends and stationarity using the Augmented Dickey-Fuller (ADF) test.

Two forecasting models are applied:

1. ARIMA (1,2,1) – Selected based on ACF/PACF plots and AIC/BIC criteria, achieving an accuracy of 89.45%.
2. The Long Short-Term Memory (LSTM) a deep learning model designed to capture complex temporal dependencies, achieves 96.31% accuracy for real GDP forecasting and 93.72% accuracy for nominal GDP forecasting**.**

Model evaluation is performed using Mean Absolute Percentage Error (MAPE). The results indicate that the LSTM model significantly outperforms ARIMA, making it a more effective tool for economic forecasting. The study also reveals that Maharashtra economy has shown resilience, with a steady recovery post-pandemic. The findings provide valuable insights for policymakers and economists in planning future economic strategies.

Introduction

Introduction to Topic:

"Gross State Domestic Product is a monetary measure of the market value of all the final goods and services produced in a period of time, often annually."

GSDP is an aggregate measure of production equal to the sum of the gross values added of all resident and institutional units engaged in production (Plus any taxes, and minus any subsidies, on products not included in the value of their outputs).

An **IMF (International Monetary Fund)** publication states that " **GSDP** measures the monetary value of final goods and services those purchased by the final user produced within a state in a given period of time (such as a quarter or a year)."

"GSDP is considered the most powerful Statistical indicator of a state's economic development and progress."

How to Determine GSDP

There are three primary methods by which GSDP can be determined. All, when correctly calculated, should yield the same figure. These three approaches are often termed the expenditure approach, the output (or production) approach, and the income approach.

1. **GSDP Based on Spending**

The expenditure approach or spending approach, which is the most common method, calculates the monies spent by the different groups that participate in the economy. For instance, consumers spend money to buy various goods and services, and businesses spend money as they invest in their business activities (buying machinery, for instance). Governments also spend money. All these activities contribute to the GSDP of a country. In addition, some of the goods and services that an economy makes are exported overseas, their **net exports.** And some of the products and services that are consumed within the country are imports from overseas. The GSDP calculation also accounts for spending on exports and imports.

This approach essentially measures the total sum of everything used in developing a finished product for sale. To return to the example of the ship, the finished ship's contribution to a nation's GSDP would here be measured by the total costs of materials and services that went into the ship's construction. This approach assumes a relatively fixed value of the completed ship relative to the value of these materials and services in calculating value added.

A country's gross domestic product can be calculated using the following formula:

GSDP=C+G+I+NX

- **C** is equal to all private consumption or consumer spending in a nation economy.
- **G** is the sum of government spending.
- **I** is the sum of all the country's investment, including business capital expenditures.
- **NX** is the nation total net exports, calculated as total exports minus total imports (**NX = Exports - Imports**).

2. GSDP Based on Production

The production approach is something like the reverse of the expenditure approach. Instead of exclusively measuring input costs that feed economic activity, the production approach estimates the total value of economic output and deducts costs of intermediate goods that are consumed in the process, like those of materials and services. Whereas the expenditure approach projects forward beyond intermediate costs, the production approach looks backward from the vantage of a state of completed economic activity.

3. GSDP Based on Income

Considering that the other side of the spending coin is income, and since what you spend is somebody else's income, another approach to calculating GSDP something of an intermediary between the two aforementioned approaches is based on a tally of the national income. Income earned by all the factors of production in an economy includes the wages paid to labour, the rent earned by land, the return on capital in the form of interest, as well as an entrepreneur's profits. An entrepreneur's profits could be invested in his own business or it could be an investment in any outside business.

All this constitutes national income, which is used both as an indicator of implied productivity and implied expenditure.

In addition, the income approach factors in some adjustments for some items that don't show up in these payments made to factors of production. In addition, depreciation, which is a reserve that businesses set aside to account for the replacement of equipment that tends to wear down with use, is also added to the national income.

Impact of the Balance of Trade on GSDP

The **balance of trade** is a key factor influencing a state's **Gross State Domestic Product (GSDP)**. GSDP increases when the total value of goods and services that state-based producers sell to other states or countries exceeds the total value of goods and services imported into the state, creating a **trade surplus**. Conversely, if the state imports more goods and services than it exports resulting in a **trade deficit** its GSDP growth may be negatively impacted. A favourable balance of trade strengthens the state's economy, fosters industrial growth, and enhances overall economic development.

Nominal vs. Real GSDP

Since Gross State Domestic Product (GSDP) is based on the monetary value of a state's economic output, it is influenced by inflationary pressures. Over time, prices in an economy generally rise, affecting the GSDP figures. Therefore, analyzing a state unadjusted GSDP alone makes it difficult to determine whether the growth is due to an actual increase in production or merely a result of rising prices.

To address this, economists adjust for inflation to derive the real GSDP instead of the nominal GSDP, which does not account for inflation or deflation. By adjusting the output for inflation to reflect price levels from a reference year, known as the base year, economists can isolate the effect of inflation and make meaningful year-over-year comparisons. This helps assess whether the state's economy is experiencing genuine growth.

Real GSDP is calculated using a GSDP price deflator, which measures the change in prices between the current year and the base year. Typically, nominal GSDP is higher than real GSDP since inflation is usually positive. Real GSDP accounts for price level changes, reducing distortions in output figures over time.

A significant gap between a state's real and nominal GSDP indicates strong inflationary pressures (if nominal GSDP is higher) or deflationary trends (if real GSDP is higher). Nominal GSDP is useful for comparing different quarters within the same year, whereas real GSDP is essential for analyzing long-term trends across multiple years, as it removes the effects of inflation and focuses on actual economic output.

Overall, real GSDP serves as a more reliable indicator for evaluating a state long-term economic performance.

Adjustment for Inflation in GSDP

GSDP figures reported for economic analysis are typically adjusted for inflation. For instance, if a state's nominal GSDP growth is calculated at 6% over the previous year but inflation during the same period is 2%, the real GSDP growth would be reported as 4%, reflecting the net economic expansion after accounting for price changes.

Why Does Inflation Increase with GSDP Growth?

An increase in unadjusted GSDP indicates that a state economy has experienced one of five possible scenarios:

1. **Produced more at the same prices.**
2. **Produced the same amount at higher prices.**
3. **Produced more at higher prices.**
4. **Produced significantly more at lower prices.**
5. **Produced less at much higher prices.**

Scenario 1:

If production increases while prices remain stable, it suggests that businesses are expanding output to meet rising demand. This leads to job creation, reducing unemployment and increasing disposable income. As wages rise, consumer spending grows, further driving demand. Eventually, this results in higher GSDP growth, often accompanied by inflation.

Scenario 2:

If production remains constant but prices rise, inflation occurs despite no increase in consumer demand. This can happen when businesses face higher input costs such as rising fuel or raw material prices forcing them to increase prices. In this case, both GSDP and inflation rise, not due to increased economic activity but because of cost-driven inflation.

Scenario 3:

If both production and prices rise, it indicates strong demand alongside supply constraints. To meet demand, businesses hire more workers, increasing wages and consumer spending. However, if supply cannot keep pace with demand, prices rise rapidly, leading to an unsustainable economic environment where both GSDP and inflation grow at an uncontrollable rate.

Scenario 4:

A scenario where production increases significantly while prices decline is rare in modern economies. Such a deflationary growth environment would indicate extreme efficiency gains or supply surpluses, which are uncommon over long periods.

Scenario 5:

This scenario resembles stagflation, similar to what the U.S. experienced in the 1970s. Here, GSDP grows at a sluggish pace while inflation remains high. Unemployment stays elevated as production struggles to keep up with economic needs, leading to economic stagnation.

By understanding these scenarios, policymakers and economists can better analyze a state's economic trends and implement strategies to maintain stable and sustainable GSDP growth.

Why GSDP Fluctuates

GSDP fluctuates due to the economic cycle. When a state's economy is booming and GSDP is rising, inflationary pressures build up as labour and production capacity approach full utilization. In response, monetary authorities or policymakers may implement tighter fiscal or monetary policies to prevent overheating and control inflation.

As interest rates rise or government spending is curtailed, businesses and consumers reduce their expenditures, slowing the economy. Declining demand may lead businesses to cut jobs, further impacting consumer confidence and spending. To counter this downturn, policymakers ease monetary or fiscal policies to stimulate economic growth and employment, setting the stage for the next expansion phase. This cycle continues over time.

Consumer confidence plays a crucial role in a state economic growth. High confidence leads to increased spending, driving GSDP growth, while low confidence signals uncertainty, reducing spending and slowing economic activity.

Business investment is another key driver of GSDP, as it enhances productive capacity and generates employment. During periods of weak consumer spending and reduced private investment such as after an economic slowdown government expenditure becomes a vital factor in stabilizing GSDP and fostering economic recovery.

Objectives

1. To assess the contribution of the primary, secondary and tertiary sectors to gross state domestic product (GSDP) growth.
2. To forecasting Maharashtra gross state domestic product (GSDP).
3. To evaluate the impact of the COVID-19 pandemic on Maharashtra gross state domestic product (GSDP).

Data collection

Research Methodology

Research methodology is an essential aspect of any research or investigation. It enables the investigator to look at the problem in a systematic, meaningful and orderly way.

Secondary Data:

Secondary data is one type of quantitative data that has already been collected by someone else for a different purpose to yours. Secondary data can be used in different ways:

- You can simply report the data in its original format. If so, then it is most likely that the place for this data will be in your main introduction or literature review as support or evidence for your argument.
- You can do something with the data. If you use it (analyse it or re-interpret it) for a different purpose to the original, then the most likely place would be in the 'Analysis of findings' section of your dissertation.

There are many sources of data and most people tend to underestimate the number of sources and the amount of data within each of these sources. Sources can be classified as:

- **Paper-based sources:** Books, Journals, Periodicals, Abstracts, Indexes, Directories, Research reports, Conference papers, Market reports, Annual reports, Internal records of organizations, Newspapers, and Magazines.
- **Electronic sources:** Online databases, Internet, videos, and broadcasts.
- **Government Websites** – RBI and Maharashtra government reports.

Tools of Analysis:

In order to analyze the objective of this study, various **visualization tools** such as **graphs, diagrams, and charts** were used to interpret the trends in GSDP data. **Python** was used for time series analysis, leveraging libraries like **NumPy, Pandas, Matplotlib, Seaborn, Statsmodels, and TensorFlow**. Through these tools, key insights were derived, and a conclusion was drawn based on statistical forecasting models such as **ARIMA and LSTM**.

Software Used

- **MS Excel** – Used for initial data collection, pre-processing, and basic visualization.

- **Python** – Used for time series analysis, forecasting, and visualization. The following Python libraries were utilized:
  - **Pandas & NumPy** – For data manipulation and pre-processing.
  - **Matplotlib & Seaborn** – For graphical representation of trends and patterns.
  - **Statsmodels** – For statistical analysis and ARIMA modelling.
  - **TensorFlow & Keras** – For building and training the LSTM deep learning model.

Methodology

I. Time Series Analysis

Time series analysis is a statistical technique used to analyse and interpret sequential data points collected over time. This method of data analysis provides insights into the underlying patterns, trends, and behaviours of a given dataset with a different perspective than other statistical analysis. By making observations at equally spaced intervals, data professionals can make informed decisions based on historical trends, accurate predictions for better future outcomes, and impactful improvements in the areas that matter most to the organization.

❑ Augmented Dickey-Fuller test

The Augmented Dickey-Fuller (ADF) test is a statistical test used to determine if a time series data is stationary or not, by testing the null hypothesis that a unit root is present, meaning the data is non-stationary.

- **Null Hypothesis:** The time series data contains a unit root (non-stationary).
- **Alternative Hypothesis:** The time series data is stationary or trend-stationary.

- **Stationarity:** A time series is considered stationary if its statistical properties (mean, variance, and autocorrelation) do not change over time.
- If the p-value is less than 0.05, reject the null hypothesis, indicating the series is stationary.

❑ Autoregressive Process

In a multiple regression model, we forecast the variable of interest using a linear combination of predictors. In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable.

The term Autoregression indicates that it is a regression of the variable against itself.

An autoregressive process is one where the current values of a variable y depend upon only the values that the variable took in previous periods plus an error term.

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t$$

Where,

δ: the mean of y,

u_t: an uncorrelated random error term with zero mean and constant variance σ^2, then we say that Y_t follows a first–order autoregressive, or AR (1), stochastic process.

Here the value of Y at time t depends on its value in the previous time period and a random term; the Y values are expressed as deviations from their mean value. In other words, this model says that the forecast value of Y at time t is simply some proportion (= α_1) of its value at time (t − 1) plus a random shock or disturbance at time t, again the Y values are expressed around their mean values. But if we consider this model,

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + u_t$$

Then we say that Y_t follows a second-order autoregressive, or AR(2), process. That is, the value of Y at time t depends on its value in the previous two time periods, the Y values being expressed around their mean value δ.

In general, we can have

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \cdots + \alpha_p(Y_{t-p} - \delta) + u_t$$

in which case Y_t is a p^{th} order autoregressive, or AR(p), process. Notice that in all the preceding models only the current and previous Y values are involved; there are no other regressors.

❑ A Moving Average (MA) Process

As we have seen, the idea behind AR process is to feed past data back into the current value of the process. This induces correlation between the past and present. The effect is to have at least some correlation at all lags. Sometimes data show correlation at only short lags, for example, only at 1 or only at lags 1 and 2. AR processes do not behave this way and will not fit such data well. In such situation, a useful alternative to an AR model is a moving average(MA) model. A process Y_t is a moving average process if Y_t can be expressed as weighted average (moving average) of the past values of the white noise process u_t, rather than of past values Y_t itself as happens in an AR process.

The MA (1) (moving average of order 1) process is

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1}$$

where μ is a constant and u, as before, is the white noise stochastic error term. Here Y at time t is equal to a constant plus a moving average of the current and past error terms. Thus, in the present case, we say that Y follows a first-order moving average, or an MA (1), process.

But if Y follows the expression

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2}$$

then it is an MA (2) process. More generally,

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \cdots + \beta_q u_{t-q}$$

is an MA(q) process. In short, a moving average process is simply a linear combination of white noise error terms.

❑ Autoregressive Integrated Moving Average (ARIMA) Process

ARIMA models are an extension of the ARMA model, which combines autoregressive (AR) and moving average (MA) components. The "I" in ARIMA stands for "integrated," which refers to the differencing of the time series to make it stationary.

An ARIMA model is defined by three parameters: p, d, and q, which correspond to the order of the autoregressive component, the order of differencing, and the order of the moving average component, respectively. The notation for an ARIMA model is ARIMA (p, d, q).

where:

p: - This parameter represents the order of the autoregressive (AR) component of the model. The AR component models the relationship between the current value of the time series and its past values. A higher value of p indicates a stronger dependence on past values and a more complex model.

d: - This parameter represents the order of differencing required to make the time series stationary. Differencing involves subtracting each value from its previous value, which can remove trends and seasonality. A higher value of d indicates that more differencing is required to make the series stationary.

q: - This parameter represents the order of the moving average (MA) component of the model. The MA component models the relationship between the current value of the time series and past error terms. A higher value of q indicates a stronger dependence on past error terms and a more complex mode

❑ The Box–Jenkins (BJ) Methodology

1. Identification

That is, find out the appropriate values of p, d, and q. We will show shortly how the correlogram and partial correlogram aid in this task.

2. Estimation

Having identified the appropriate p and q values, the next stage is to estimate the parameters of the autoregressive and moving average terms included in the model.

3. Diagnostic checking

Having chosen a particular ARIMA model, and having estimated its parameters, we next see whether the chosen model fits the data reasonably well, for it is possible that another ARIMA model might do the job as well. This is why Box–Jenkins ARIMA modelling is more an art than a science; considerable skill is required to choose the right ARIMA model.

One simple test of the chosen model is to see if the residuals estimated from this model, are white noise; if they are, we can accept the particular fit; if not, we must start over. Thus, the **BJ methodology is an iterative process**.

4. Forecasting

Select the number of time periods to forecast. The first step is to decide on the number of time periods for which you want to generate forecast.

❑ Diagnostic checking

Residuals should be white noise: The residuals should exhibit no significant patterns or trends, and their distribution should be approximately normal with constant variance. This can be checked visually using a time plot and a histogram of the residuals, as well as a normal probability plot and a Boxplot. Additionally, statistical tests such as the Ljung-Box test for autocorrelation and the Shapiro Wilk test for normality can be used to check the adequacy of the model.

ACF and PACF: The ACF and PACF of the residuals should be examined to ensure that there is no significant correlation remaining in the residuals. Any remaining correlation indicates that the model may be mis-specified or inadequately fitted.

Model adequacy testing: Checking if the selected model is adequate and meets the assumptions by conducting tests such as the Ljung-Box test, Q-Q plot, and residual autocorrelation function (ACF) plot.

Diagnostic checking is a crucial step in the Box-Jenkins methodology, as it ensures that the chosen ARIMA model is a good fit for the time series data and can be used for forecasting and decision making.

- **Ljung Box Test**

H_0: - Autocorrelation is not Present.
H_1: - Autocorrelation is Present.
Test statistic:

$$Q = n(n+2)\ \Sigma_{k=1} \frac{\hat{\rho}_k^2}{n-k}$$

If P-value <0.05, thus we reject H_0 at 5% level of significance. Thus, there is autocorrelation present in the residuals.

II. Long Short-Term Memory (LSTM): -

LSTM is a type of recurrent neural network (RNN) that effectively captures long-term dependencies in sequential data. It is particularly suitable for time-series forecasting due to its memory cell structure, which mitigates vanishing gradient issues.

1. Data Pre-processing

Let X = $\{x_1, x_2, \ldots \ldots \ldots, x_t\}$ be the GSDP time series, where t is the total number of observations. The pre-processing steps include:

- **Normalization:**

  $x' = \frac{x - X_{min}}{X_{max} - X_{min}}$ This ensures all values lie between 0 and 1 for stable training.

- **Supervised Learning Transformation:** The dataset is converted into overlapping input-output sequences: $Y_t = (X_{t-n}, X_{t-n+1}, \ldots \ldots, X_t)$ where n is the lookback window.
- **Train-Test Split:** The data is divided into training (80%) and testing (20%) sets.

2. Model Architecture

The LSTM model follows a structured layer architecture:

- **Input Layer:** Accepts a sequence of n past observations $(X_{t-n}, \ldots \ldots \ldots X_{t+1})$.
- **LSTM Layer:** Contains memory cells with the following equations:
  - Forget gate: $f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f)$
  - Input gate: $i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i)$
  - Candidate cell state: $\widetilde{C_t} = ReLU(W_c . [h_{t-1}, x_t] + b_c)$
  - Cell state update: $C_t = f_t . C_{t-1} + i_t . \widetilde{C_t}$

- Output gate: $O_t = \sigma(W_O.[h_{t-1}, x_t] + b_O)$
- Hidden state: $h_t = O_t.ReLU(C_t)$

- **Weight and Bias Update:** During backpropagation through time (BPTT), weights and biases are updated using the gradients of the loss function with respect to the parameters:
  - Weight Update :- $W_{new} = W_{old} - \eta.\frac{\partial L}{\partial W_{old}}$
  - Bias Update :- $B_{new} = B_{old} - \eta.\frac{\partial L}{\partial B_{old}}$

  where η is the learning rate, and L is the loss function
- **Dropout Layer:** Prevents overfitting by randomly deactivating neurons.
- **Dense Layer:** Outputs the predicted GSDP value.

The model is optimized using **Adam optimizer** and trained with **Mean Squared Error (MSE)** loss function: MSE=$\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$

3. Model Training and Hyperparameter Tuning

The model was trained with:

- **Batch size:** (16) The number of samples processed before the model updates its weights.
- **Epochs:** (100) The number of times the model goes through the entire dataset during training. (One epoch consists of one complete cycle of forward propagation and backward propagation across the entire dataset)
- **Hyperparameter tuning:** Grid search was used to optimize:
  - Number of LSTM units u
  - Learning rate η
  - Dropout probability p

4. Evaluation Metrics

The model's accuracy was assessed using:

- **Mean Absolute Error (MAE):**

MAE=$\frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$

- **Root Mean Squared Error (RMSE):**

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

- **Mean Absolute Percentage Error (MAPE):**

MAPE=$\frac{100}{N}\sum_{i=1}^{N}|\frac{y_i-\hat{y}_i}{y_i}|$

5. Forecasting

The trained LSTM model was used to predict future GSDP values: $\hat{Y}_{t+1}$=$f(X_{t,}X_{t-1,\ldots\ldots}x_{t-n+1,})$.

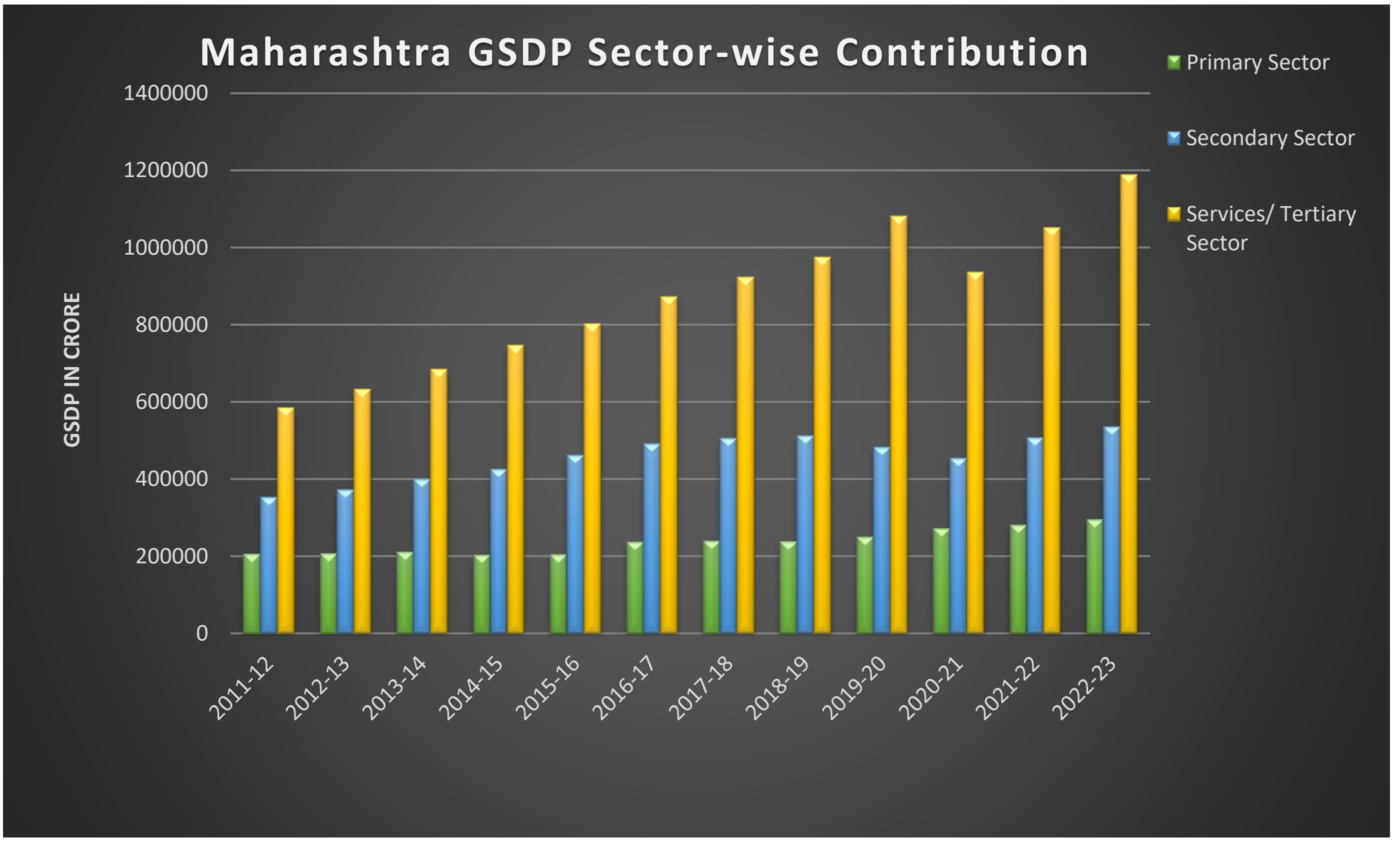
Data overview

Sr. no	Variable	Description	Data Type
1	Year	The year period for which the economic data is recorded. (2000-01, 2022-23)	DateTime
2	Primary Sector	Represents the value of economic activities related to natural resources. This includes agriculture, forestry, fishing, and mining. (Values in Crore)	Quantitative
3	Secondary Sector	Represents the value of industrial and manufacturing activities, including construction and utilities. (Values in Crore)	Quantitative
4	Services/ Tertiary Sector	Represents the value of activities related to services such as trade, transportation, banking, healthcare, education, public administration, and other services. (Values in Crore)	Quantitative
5	Gross State Domestic Product (GDDP)	The total monetary value of all goods and services produced within the district in a given year. (Values in Crore)	Quantitative
6	Net State Domestic Product (NSDP)	The total monetary value of all goods and services produced within the district after deducting depreciation of capital assets. It represents the net value added. (Values in Crore)	Quantitative

Statistical Analysis

Exploratory Data Analysis (EDA)

- Maharashtra Real GSDP Sector-wise Contribution



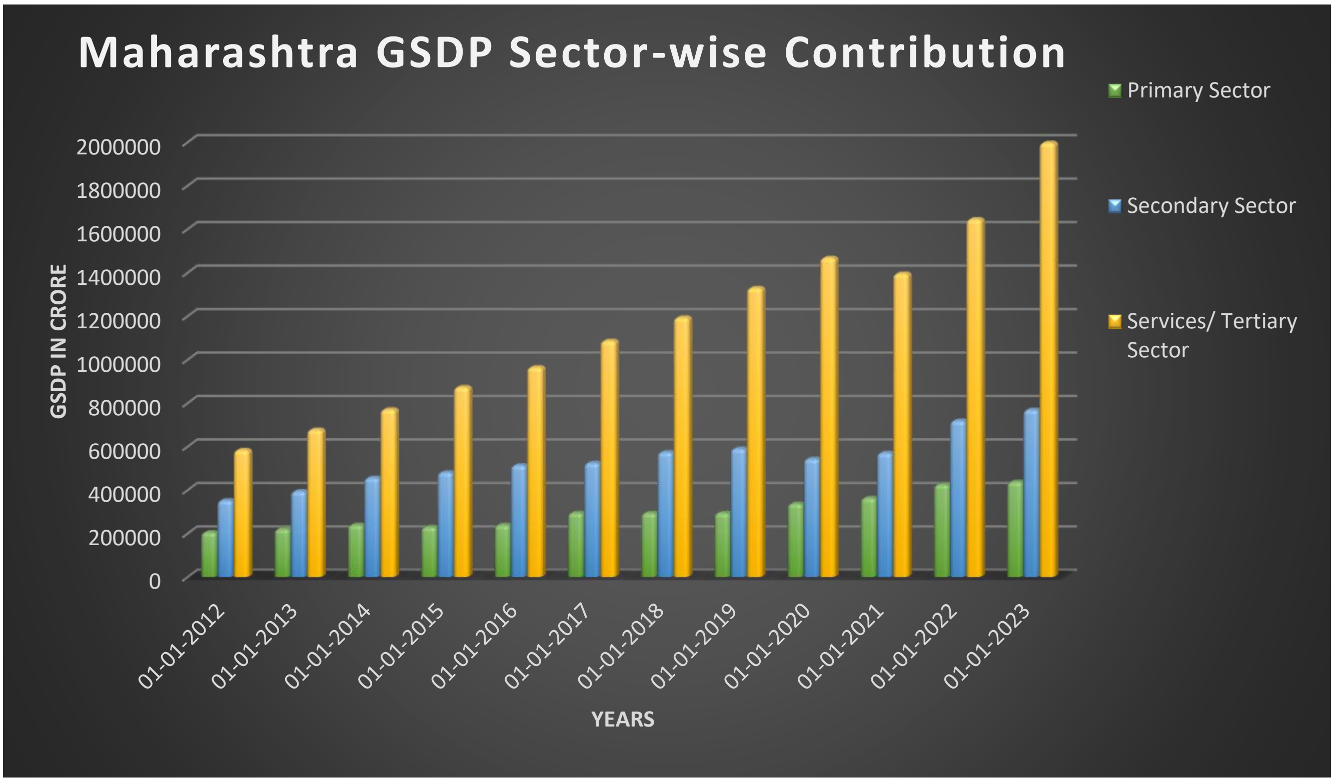
Interpretation: -

- The services sector is the biggest contributor to Maharashtra economy and plays a major role in its growth.
- Agriculture has the smallest contribution to Maharashtra economy but is slowly growing over time.
- In 2021, the economy slowed down because of the covid-19 pandemic, but it has gradually recovered and grown stronger since then.
- Maharashtra economy is strong and continues to grow after the pandemic.

Conclusion: -

Maharashtra economy is strong and has grown steadily over the years. After the pandemic all three sectors bounced back with the services sector leading the recovery. For the future balanced growth in all sectors will help the economy keep growing.

❑ Maharashtra Nominal GSDP Sector-wise Contribution



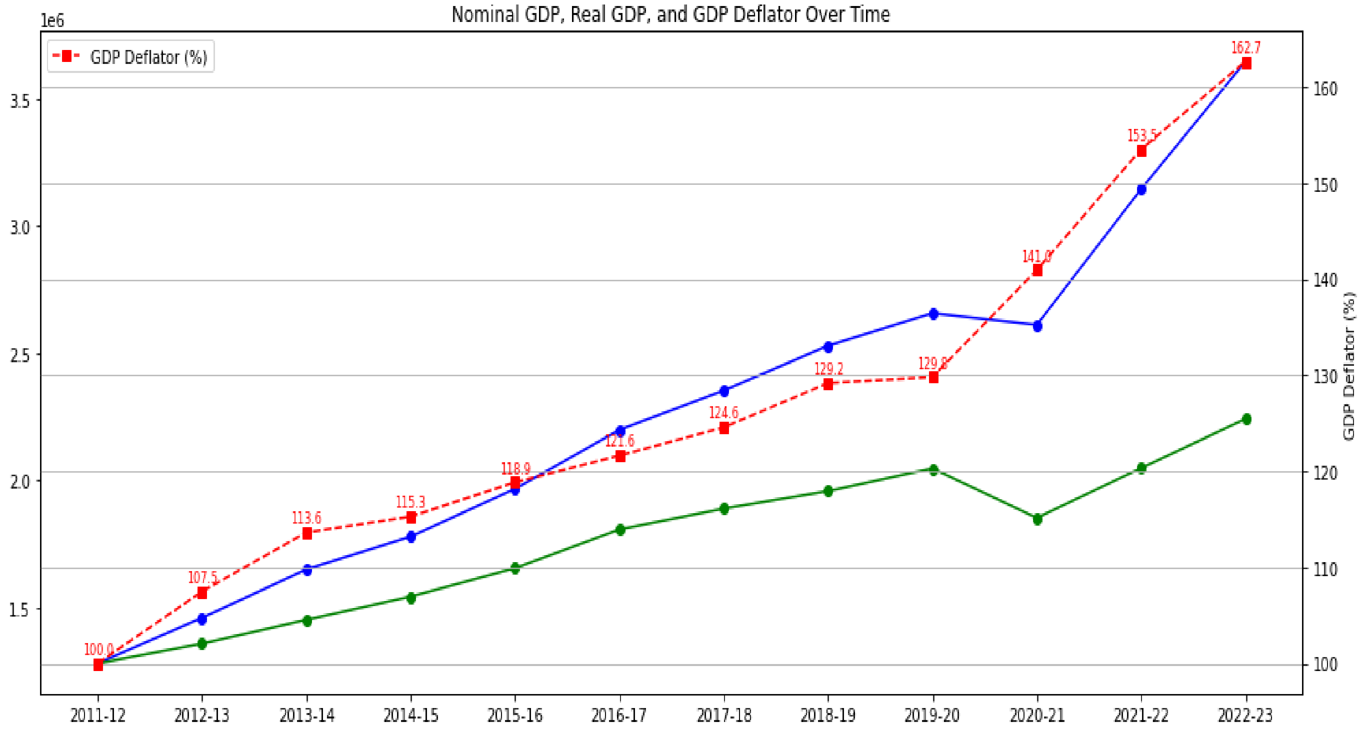
Interpretation: -

- The Services/Tertiary sector is the largest contributor, showing strong growth in trade, IT, banking, healthcare and education.
- The Secondary sector has grown steadily, supporting manufacturing, construction, and infrastructure.
- The Primary sector has the smallest share but is gradually improving in agriculture, forestry and fishing.
- A dip in 2020-21 due to COVID-19 was followed by a strong recovery from 2021-22 onwards, led by the Services sector.

Conclusion: -

The graph shows steady economic growth with contributions from all three sectors. Post-pandemic, Maharashtra economy is recovering well, led by the Services sector. A balanced approach across sectors will support long-term development.

❑ GSDP Deflator and Economic Trends Over Time



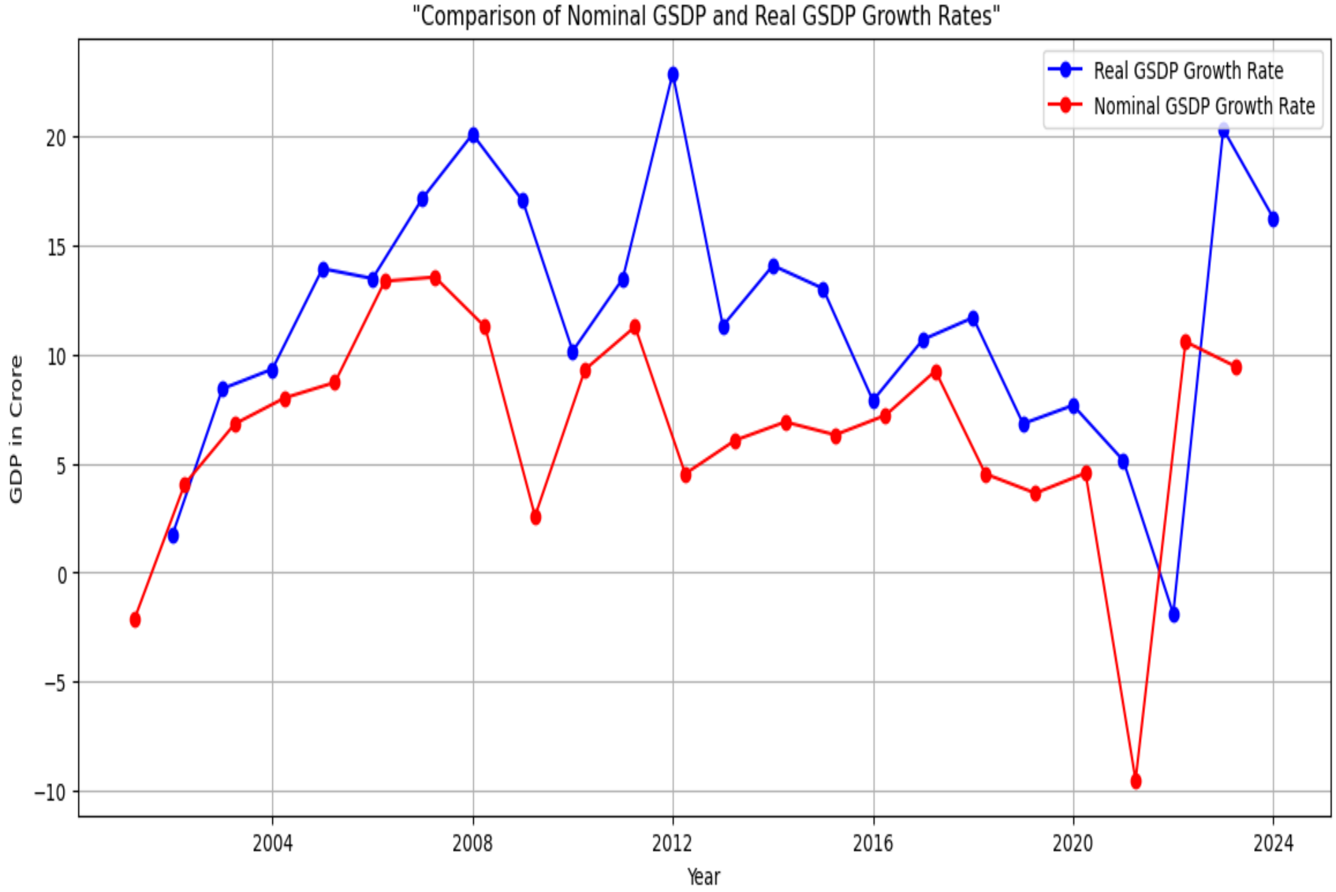
Interpretation: -

- ❑ The gap between the red and green (Real GSDP growth), blue (Nominal GSDP growth) lines is getting bigger, which means prices have been rising over time (inflation).
- ❑ The red line (GSDP deflator) has gone up sharply in recent years, showing that inflation has increased even more.
- ❑ The green line (Real economic growth) is growing steadily but more slowly because inflation is affecting it.

Conclusion: -

The graph shows that the GSDP Deflator is increasing, which means prices are rising (inflation). The blue line shows steady economic growth, while the green line has ups and downs, suggesting that some sectors are not growing as fast. This highlights the need for measures to control inflation and support weaker sectors to ensure stable and balanced economic growth.

❑ Comparison of Nominal GSDP and Real GSDP Growth Rates



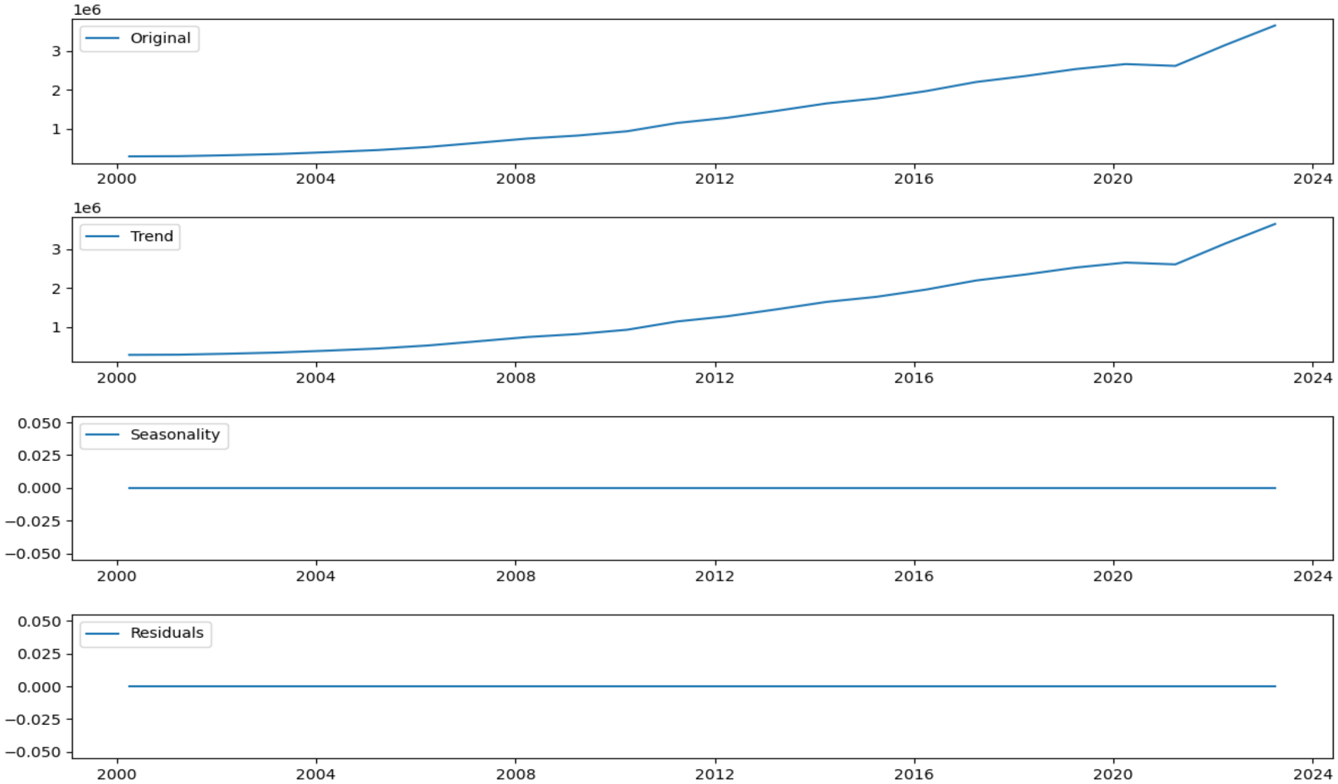
Interpretation: -

- ❑ After 2021, the economy started recovering and showing signs of strength.
- ❑ The difference between nominal and real GSDP growth helps us understand how much prices (inflation) are affecting the economy.

Conclusion: -

The graph shows that both real and nominal GSDP growth rates went up and down over time. There was a big drop around 2020, likely due to economic problems, but after that, the economy recovered well.

Time Series Decomposition of GSDP



Interpretation: -

- **Original Data:** The first plot shows the actual time series, which exhibits an overall increasing trend, indicating long-term growth.
- **Trend Component**: The second plot highlights the trend, confirming a consistent upward movement, except for a slight dip around 2020-2021, likely due to the COVID-19 pandemic.
- **Seasonality Component**: The third plot appears flat, indicating that there is no significant seasonal pattern in the data. This suggests that the variations in the time series are not driven by seasonal factors.
- **Residuals Component:** The fourth plot is also flat, implying that there are no significant irregular variations or unexplained fluctuations after removing the trend and seasonality.

Conclusion: -

The data shows a steady increase over time without regular ups and downs. There is a small drop around 2020, likely due to the pandemic. Overall, the trend suggests strong and stable economic growth.

Identifying the order of d using Augmented Dickey-Fuller test

- ADF Test Hypothesis: -

  Null Hypothesis (H_o): Data is non-Stationary

  Alternative Hypothesis(H_1): Data is Stationary

- ADF Test on Original Data

  - ADF Statistic: -0.8390368
  - p-value: 0.8073855

Conclusion: - Since the p-value (0.8073855) is greater than 0.05, we fail to reject the null hypothesis (H_o) indicating that the data is non-stationary.

- ADF Test on 1st Differenced

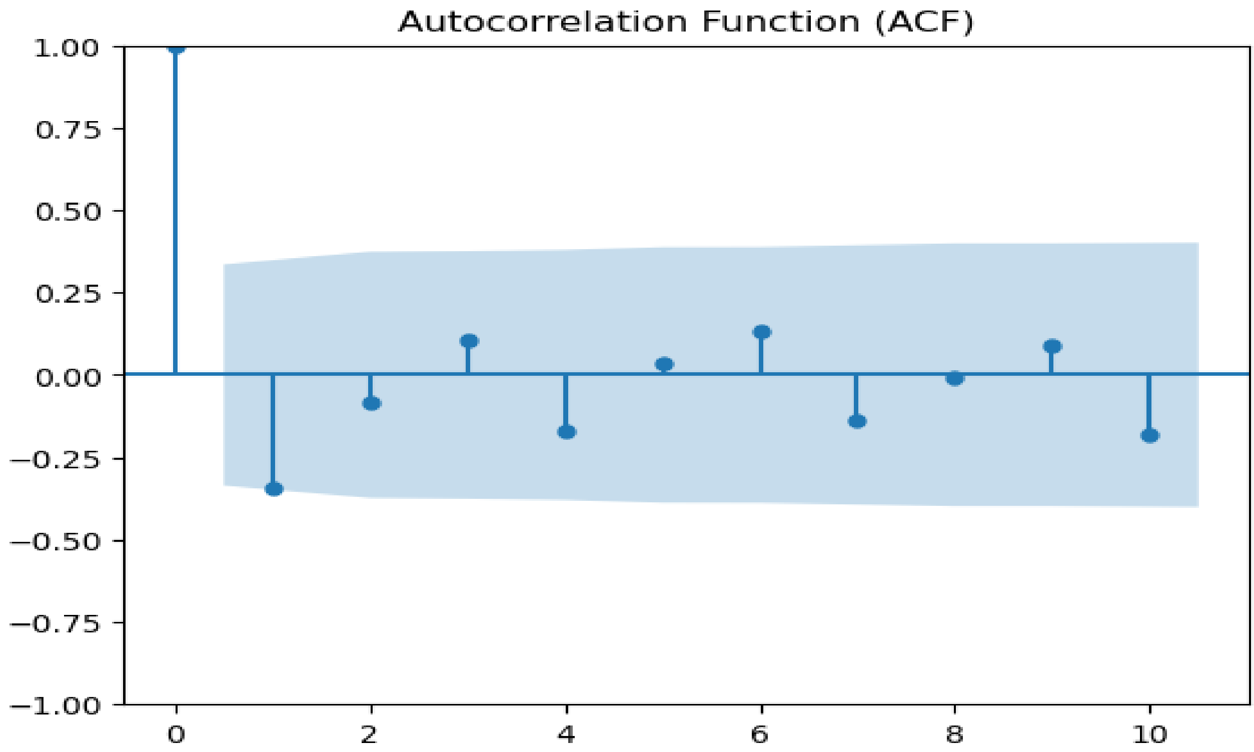
  - ADF Statistic: (-2.0016656)
  - p-value: (0.285843236)

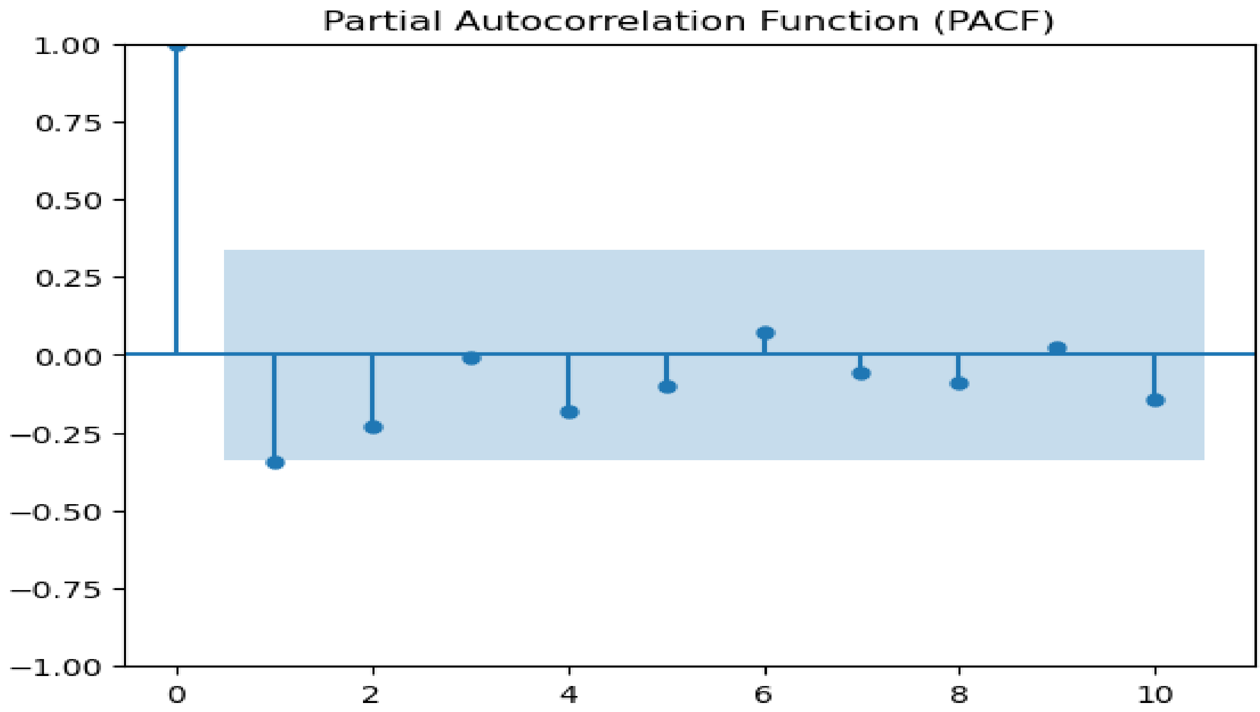
Conclusion: - Since the p-value (0.285843236) is greater than 0.05, we fail to reject the null hypothesis (H_o) indicating that the data is non-stationary.

- ADF Test on 2nd Differenced

  - ADF Statistic: (-3.483161)
  - p-value: (0.0084302)

Conclusion: - Since the p-value (0.0084302) is less than 0.05, we reject the null hypothesis (H_o), indicating that the data becomes stationary after the second difference.





ACF: There is a significant spike at lag 1, q = 1

PACF: Significant spikes at lags 1 , p = 1

ARIMA (p, d, q) = (1, 2, 1)

Parameters Estimation of Model

Model=ARIMA (1,2,1)

	coefficient
ar.L1	0.2497
ma.L1	-0.7613
AIC	585.151

Interpretation: -

- Autoregressive Term (AR (1) = 0.2497):
  - ($0.2497GDP_{t-1}$) This means that the GSDP at time t is influenced by 24.97% of the GSDP from the previous time step (t−1).
  - A positive coefficient indicates that if the previous GSDP was high, the current GSDP is also likely to be high, and vice versa.
- Moving Average Term (MA (1) = -0.7613):
  - ($-0.7613u_{t-1}$) The coefficient −0.7613 determines how strongly the past error u_{t-1} influences the current GSDP.
  - This coefficient shows how past forecasting errors (u_{t-1}) affect the current GSDP.
- Error Term (u_t):- Represents the random or unpredictable part of GSDP at time t which cannot be explained by past values or errors.

$$\boldsymbol{GSDP_t = 0.2497GSDP_{t-1} - 0.7613u_{t-1} + u_t}$$

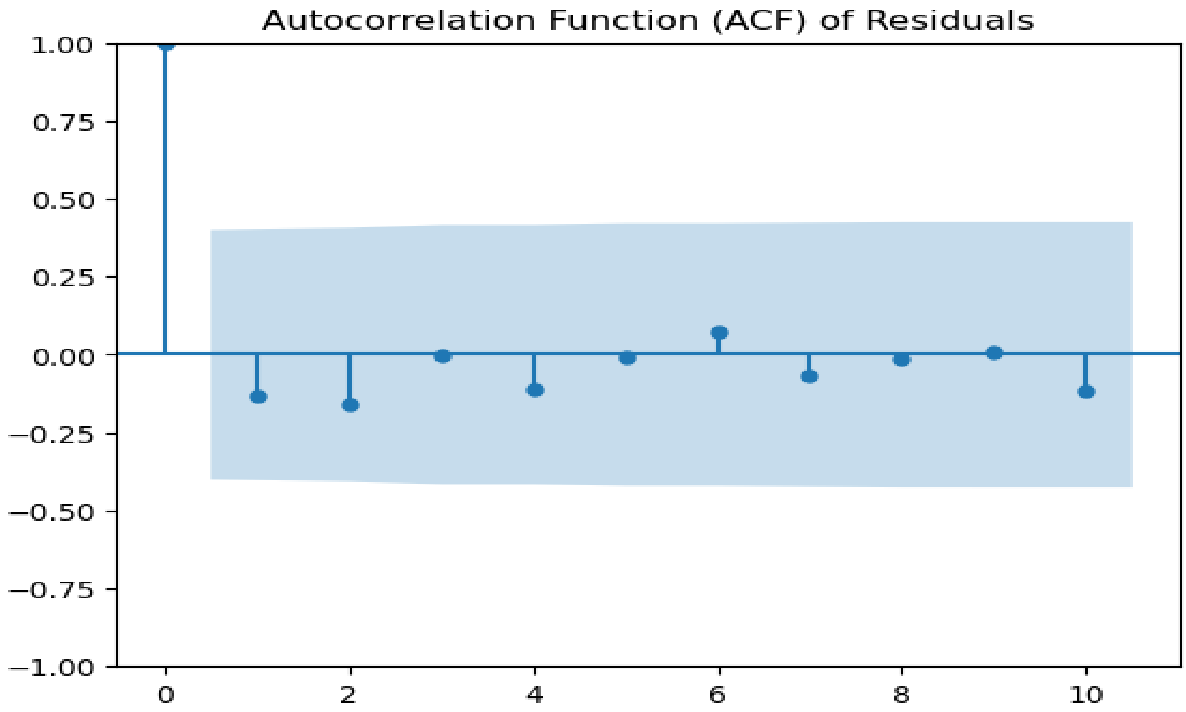
Interpretation: -

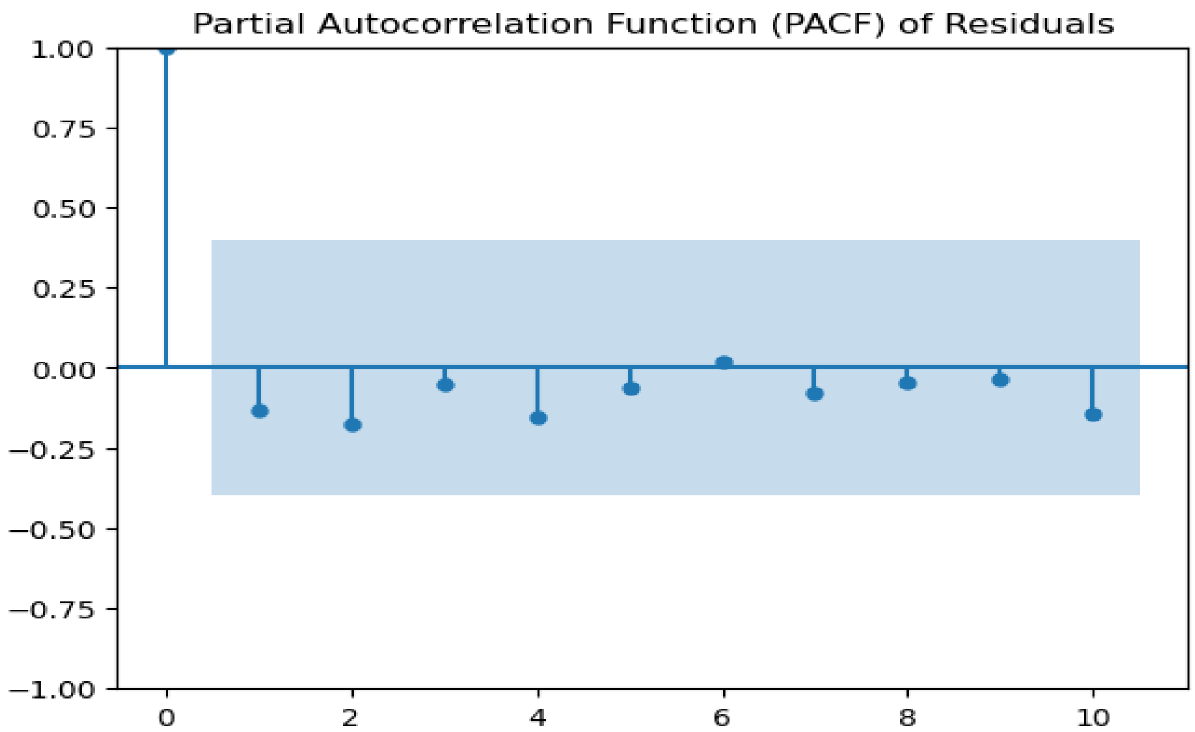
Maharashtra GSDP is influenced by its past values (24.97%) and previous forecasting errors (-0.7613). The negative moving average coefficient suggests that past prediction mistakes lead to corrections in future estimates. This model accounts for past trends and forecasting errors to improve future GSDP predictions.

Conclusion: -

The AR value (0.2497) shows that past data has a small effect on future predictions, while the MA value (-0.7613) helps correct past errors. The model performs well based on its AIC score (585.151), but some improvements could make it more accurate.

Statistical Validation of Model Assumptions





Interpretation: - No lag is significantly different from zero therefore there is no autocorrelation is present.

Checking stationarity of Residual

Null Hypothesis (H_0): The residuals are non-stationary.
Alternative Hypothesis (H_1): The residuals are stationary.

- ADF Statistic: (-3.5066749)
- p-value: (0.00781731)
- Conclusion:-

Since, p-value (0.00781731) < 0.05**,** we reject the null hypothesis (H_0) and conclude that the residuals are stationary**.**

Performing Ljung-Box test for model diagnostic checking

H_0:No autocorrelation

H_1:Autocorrelation is present

- LB Statistic: 2.570966
- p-value: 0.989801
- Conclusion: p-value > 0.05

Since, p-value (0.989801) is greater than 0.05 therefore No autocorrelation is present in residuals

Shapiro-Wilk Test for Normality

- Null Hypothesis (H_0): The residuals follow a normal distribution.
- Alternative Hypothesis (H_1): The residuals do not follow a normal distribution.

Test Results:

- W-statistic: 0.8309
- p-value: 0.0010

Conclusion:

Since the p-value (0.0010) is less than 0.05, we reject the null hypothesis (H_0). This means the residuals do not follow a normal distribution, indicating potential deviations from normality in the model error terms.

Forecasting of GSDP using ARIMA model

Accuracy of Model: -

RMSE	154842.2578
MAE	86875.61142
MAPE	11.93%

Interpretation: - Around 11.93 % MAPE implies the model is about 88.07% accurate in predicting the next 3 observations.

Conclusion: -

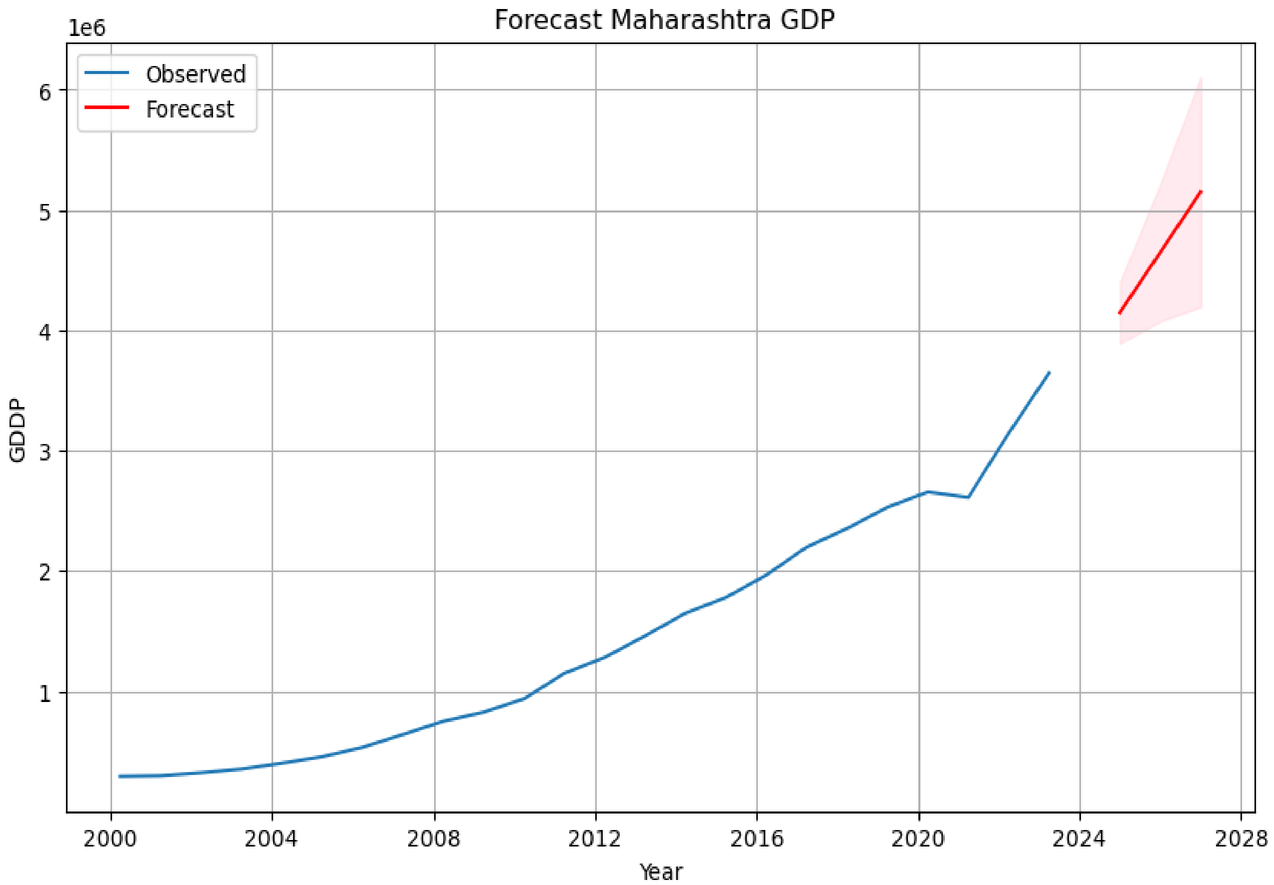
The model predicts Maharashtra future GSDP with **88.07% accuracy**, but there is some error (MAPE: **11.93%**, RMSE: **154842.26**, MAE: **86875.61**). It gives a good estimate, but improving the model with more data or better settings can make predictions even more accurate.

Projected Economic Growth of Maharashtra (2024-2026)

Year	Forecast GDP
31-03-2024	4.15E+06
31-03-2025	4.65E+06
31-03-2026	5.15E+06

Maharashtra economy is expected to grow steadily in the coming years:

- **March 31, 2024**: - ₹4.15 lakh crore
- **March 31, 2025**: - ₹4.65 lakh crore
- **March 31, 2026**: - ₹5.15 lakh crore



Interpretation: -

Maharashtra economy is expected to keep growing after 2024, as shown by the rising red line. This means the GSDP is likely to increase steadily. The light red shaded area shows possible ups and downs, meaning the exact growth might vary. A wider shaded area means there some uncertainty about how fast the economy will grow. Overall, the future looks positive, but there may be some fluctuations along the way.

Conclusion: -

Maharashtra GSDP has grown steadily, with a brief slowdown around 2020 but a strong recovery afterward. The forecast shows continued growth, though some uncertainty remains. Overall, the economic outlook for Maharashtra is positive.

Forecasting of Nominal GSDP using LSTM model

Model Training and Evaluation (Nominal GSDP)

Model1: "sequential"

Layer (type)	Output Shape	Param
lstm_1(LSTM)	(None, 3, 50)	10,400
lstm_2 (LSTM)	(None, 50)	20,200
dense (Dense)	(None, 1)	51

Total params: 30,653

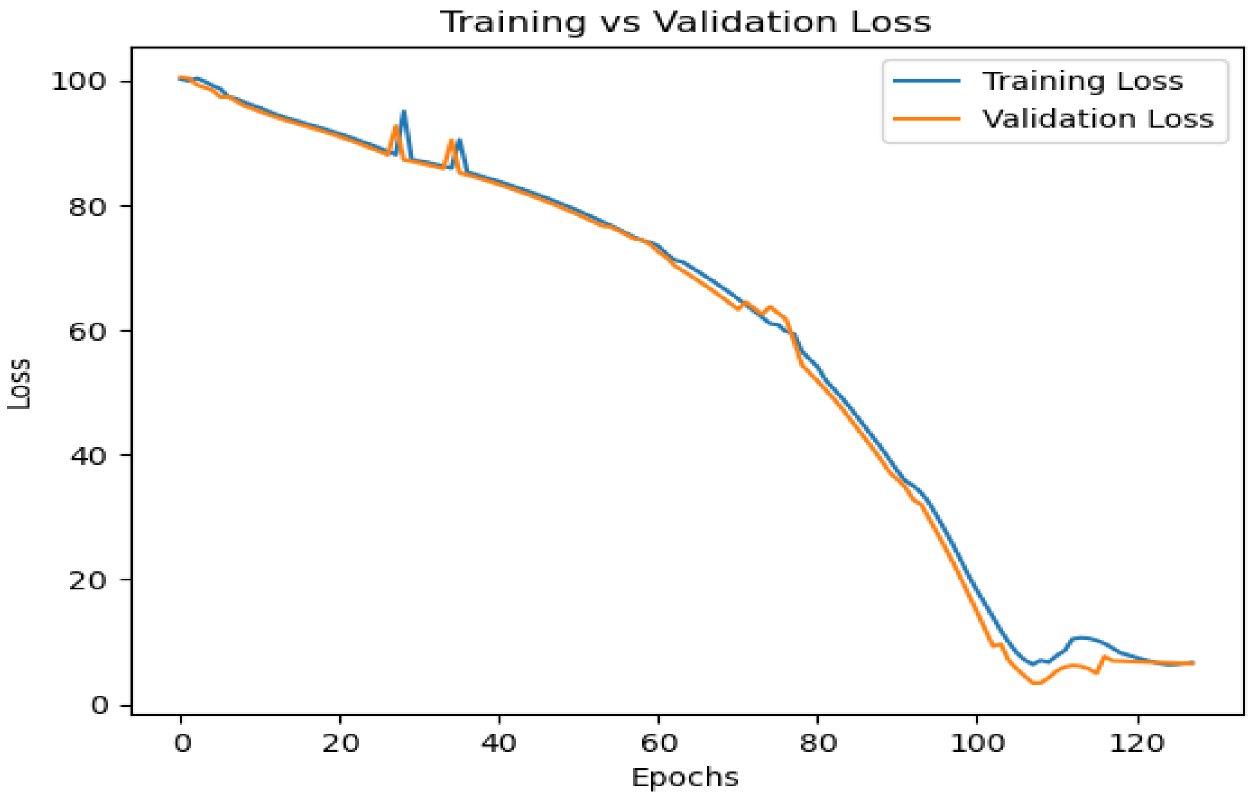
Trainable params: 30,651

Non-trainable params: 0

Optimizer params: 2

Model Architecture: -

This Sequential LSTM Neural Network is designed for time-series forecasting data tasks. It consists of two LSTM layers the first processes input sequences of length 3 with 50 features and has 10,400 trainable parameters, while the second reduces the sequence dimension to a single vector of size 50 with 20,200 parameters. A Dense layer follows, producing a single output value with 51 parameters. In total, the model has 30,653 parameters**,** with 30,651 being trainable**.** The optimizer parameters (2) likely manage learning rate adjustments. This structure helps the model learn sequential dependencies effectively for prediction tasks.



Interpretation: -

- The model is learning well because both training and validation loss are getting smaller over time.
- Since the validation loss is close to the training loss, the model is generalizing well (no overfitting).
- The low loss values at the end mean the model have understood the data patterns and can make good predictions.

Conclusion: -

The model is well-structured for handling sequential data and shows effective training behaviour. The balance between training and validation loss suggests good generalization, meaning it can make accurate predictions on new data.

Forecasting Results

Accuracy of Model: -

MAPE	6.52%
MAE	78687.68
RMSE	90960.33

Interpretation: -

Around 6.52% MAPE implies the model is about 93.48 % accurate in predicting the next 3 observations.

Conclusion: -

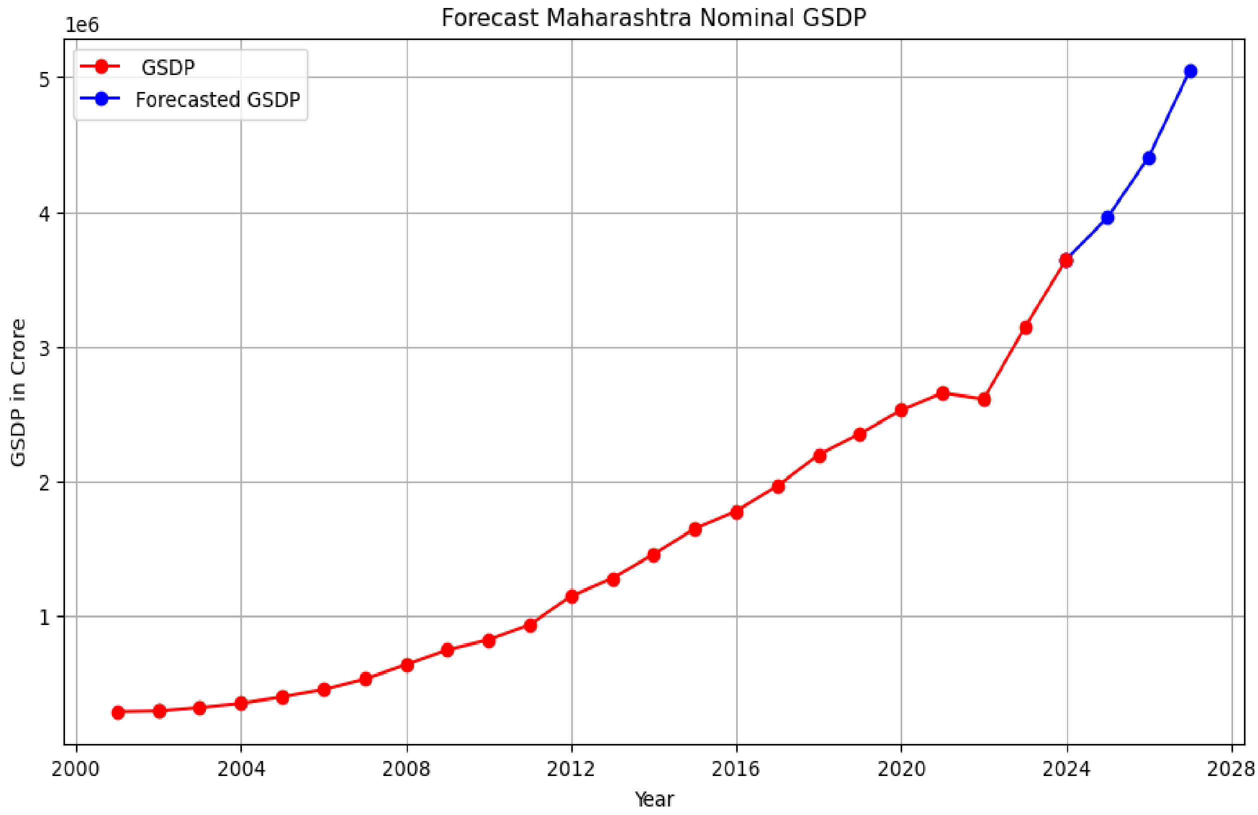
The model makes accurate predictions with errors averaging only 6.52% from the actual values. The average mistake in predictions is around 78,687.68 units and larger errors push the RMSE to 90,960.33. These numbers seem big but their significance depends on the scale of the data.

Projected Economic Growth of Maharashtra (2024-2026)

Year	Forecasted GSDP
31-03-2024	3957839.5
31-03-2025	4407256
31-03-2026	5053009

Maharashtra economy is expected to grow steadily in the coming years:

- **March 31, 2024**: - ₹3.95 lakh crore
- **March 31, 2025**: - ₹4.40 lakh crore
- **March 31, 2026**: - ₹5.05 lakh crore



Interpretation: -

- The forecasted GSDP values for Maharashtra indicate strong economic growth in the coming years.
- Based on the table, the GSDP is projected to rise from 39,57,839.5 crores in 2024 to 50,53,009 crores in 2026 showing a consistent upward trend**.**
- The graph also supports this trend, with historical GSDP (red) showing steady growth and forecasted values (blue) continuing the sharp increase beyond 2024**.**
- The steep rise in GSDP suggests a rapidly expanding economy possibly driven by industrial growth, increased investments and policy support.

Conclusion: -

Maharashtra economy is expected to grow significantly in the coming years. By 2026 its GSDP is likely to cross 50 lakh crores making it one of India's strongest economies. This growth could be driven by industries, investments, and government policies. However, factors like inflation, global economic condition.

Forecasting of Real GSDP using LSTM model

Model Training and Evaluation (Real GSDP)

Model2: "sequential"

Layer (type)	Output Shape	Param
lstm_1(LSTM)	(None, 3, 50)	10,400
lstm_2 (LSTM)	(None, 50)	20,200
dense (Dense)	(None, 1)	51

Total params: 30,653

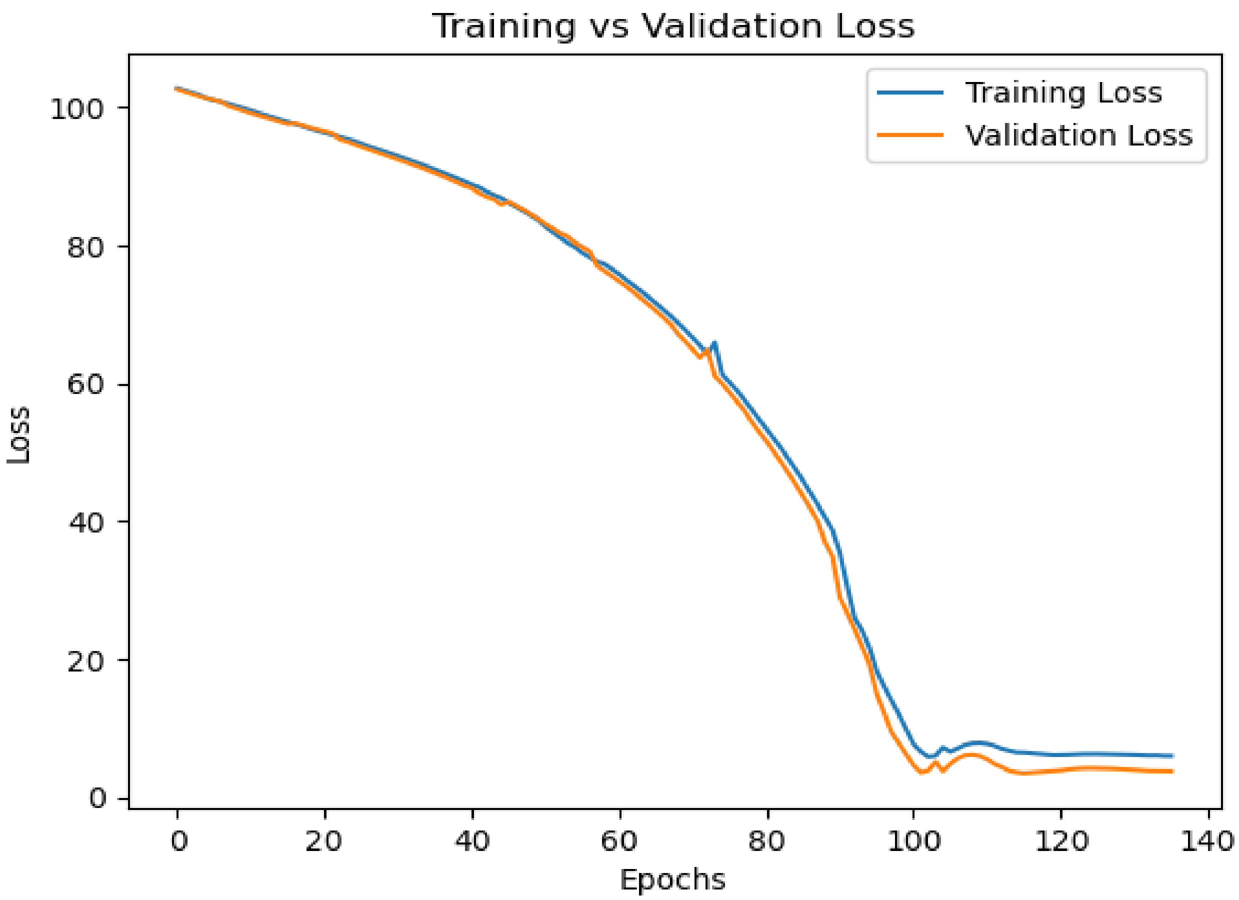
Trainable params: 30,651

Non-trainable params: 0

Optimizer params: 2

Model Architecture: -

This Sequential LSTM Neural Network is designed for time-series forecasting data tasks. It consists of two LSTM layers the first processes input sequences of length 3 with 50 features and has 10,400 trainable parameters, while the second reduces the sequence dimension to a single vector of size 50 with 20,200 parameters. A Dense layer follows, producing a single output value with 51 parameters. In total, the model has 30,653 parameters**,** with 30,651 being trainable**.** The optimizer parameters (2) likely manage learning rate adjustments. This structure helps the model learn sequential dependencies effectively for prediction tasks.



Interpretation: -

- The loss consistently decreases, showing that the model is learning patterns in the data effectively.
- The validation loss follows the training loss closely, meaning the model is not overfitting to training data.
- The model is learning well and generalizing to new data.

Conclusion: -

The model is well-trained and does not show signs of overfitting or underfitting. Both training and validation losses are low and stable, indicating that the model has learned the data patterns well.

Forecasting Results

Accuracy of Model: -

MAPE	3.69%
MAE	41775.73
RMSE	58822.75

Interpretation: -

Around 3.69% MAPE implies the model is about 96.31 % accurate in predicting the next 3 observations

Conclusion: -

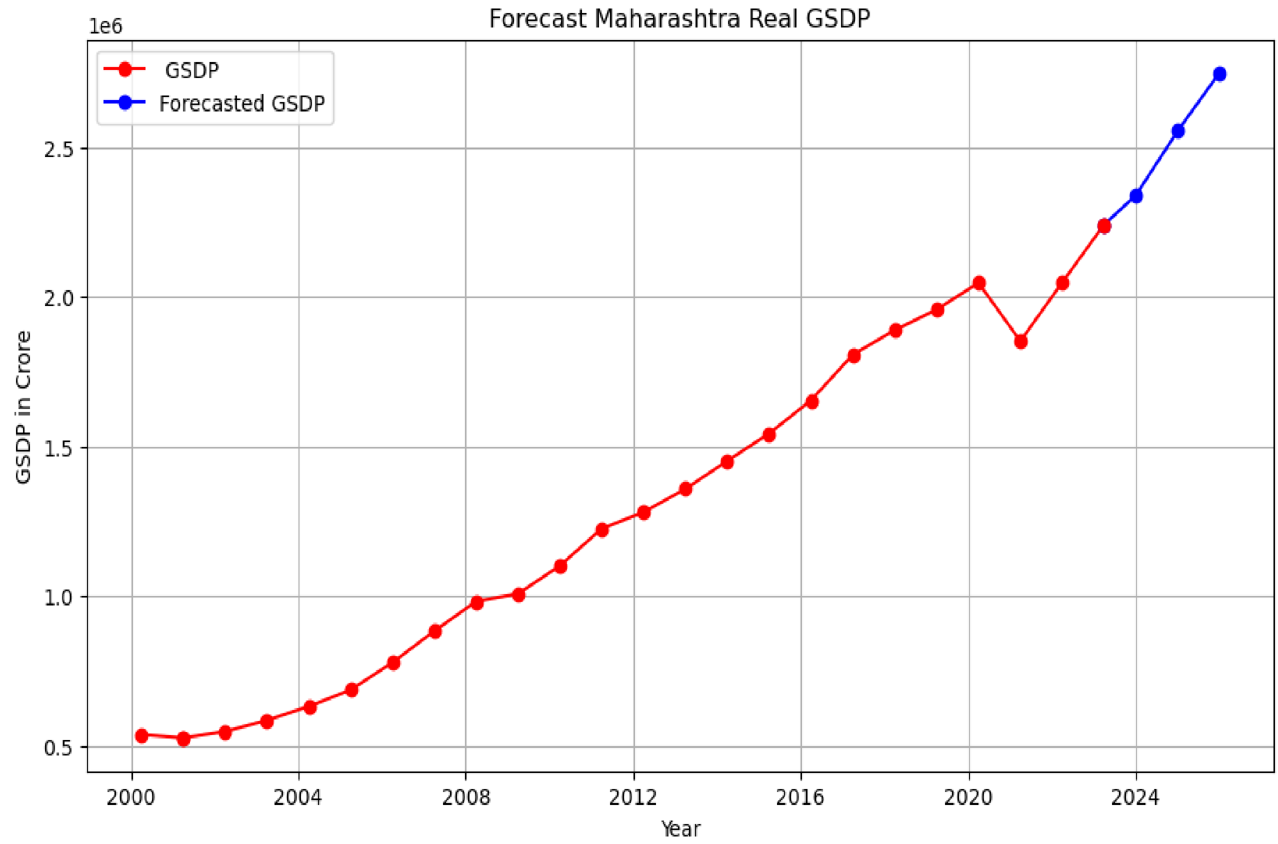
The model makes accurate predictions with errors averaging only 3.69% from the actual values. The average mistake in predictions is around 41775.73 units and larger errors push the RMSE to 58822.75. These numbers seem big but their significance depends on the scale of the data.

Projected Economic Growth of Maharashtra (2024-2026)

Year	Forecasted GSDP
31-03-2024	2338529
31-03-2025	2554263
31-03-2026	2748874

Maharashtra economy is expected to grow steadily in the coming years:

- **March 31, 2024**: - ₹2.24 lakh crore
- **March 31, 2025**: - ₹2.55 lakh crore
- **March 31, 2026**: - ₹2.75 lakh crore



Interpretation: -

- Maharashtra real GSDP showed consistent growth from 2000 to 2019, driven by industrialization and investments.
- A noticeable dip due to the pandemic disrupted economic activities.
- The economy rebounded after 2021, supported by policies and increased investments.
- GSDP is expected to reach ₹25.5 lakh crore by 2025 and ₹27.5 lakh crore by 2026, indicating continued expansion.

Conclusion: -

Maharashtra real GSDP is growing steadily, with a brief dip in 2020 due to COVID-19. A strong recovery followed, and projections show it will exceed ₹27.5 lakh crore by 2026, driven by industrial growth and investments.

Model Training and Evaluation (Impact of COVID19)

Model3: "sequential"

Layer (type)	**Output Shape**	**Param**
lstm_1(LSTM)	(None, 3, 50)	10,400
lstm_2 (LSTM)	(None, 50)	20,200
dense (Dense)	(None, 1)	51

Total params: 30,653

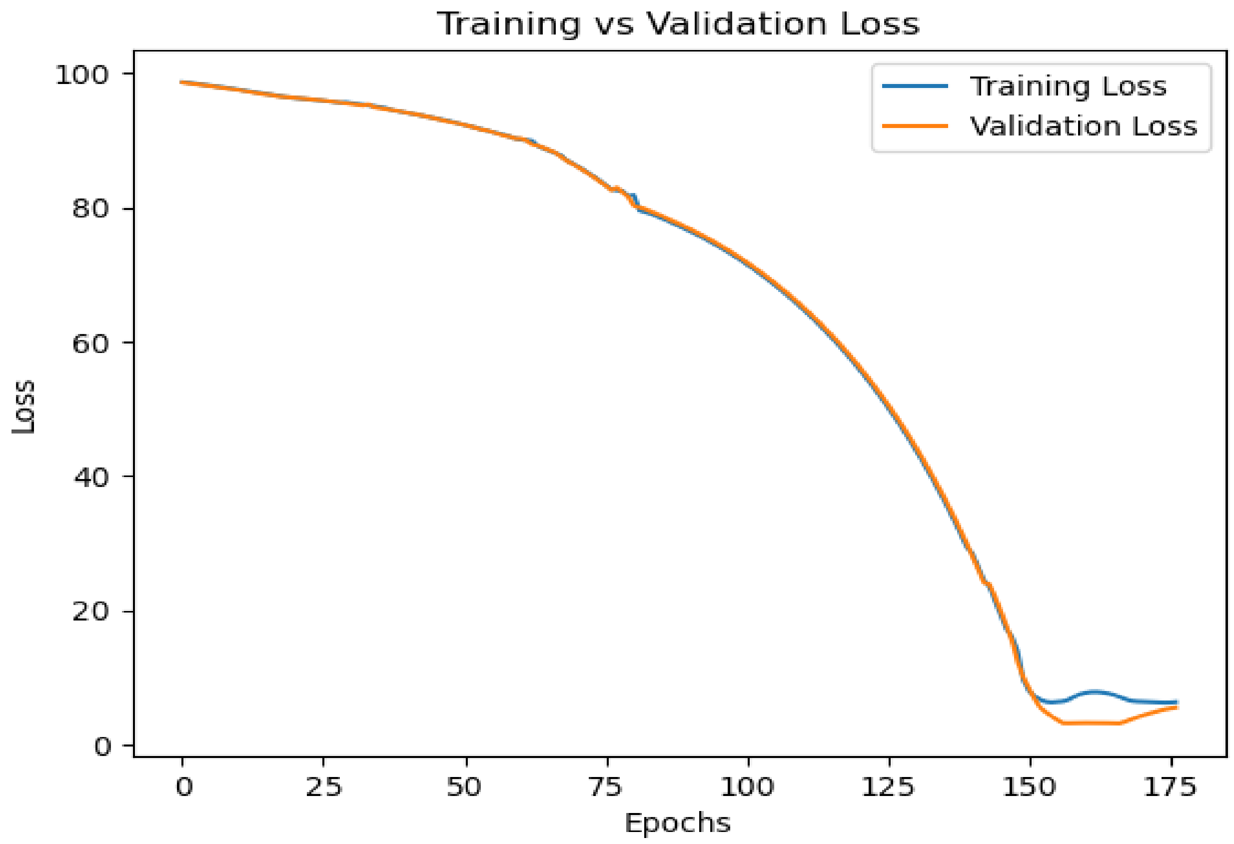
Trainable params: 30,651

Non-trainable params: 0

Optimizer params: 2

Model Architecture: -

This Sequential LSTM Neural Network is designed for time-series forecasting data tasks. It consists of two LSTM layers the first processes input sequences of length 3 with 50 features and has 10,400 trainable parameters, while the second reduces the sequence dimension to a single vector of size 50 with 20,200 parameters. A Dense layer follows, producing a single output value with 51 parameters. In total, the model has 30,653 parameters**,** with 30,651 being trainable**.** The optimizer parameters (2) likely manage learning rate adjustments. This structure helps the model learn sequential dependencies effectively for prediction tasks.



Interpretation: -

- The model is learning well because both training and validation loss are getting smaller over time.
- Since the validation loss is close to the training loss, the model is generalizing well (no overfitting).
- The low loss values at the end mean the model have understood the data patterns and can make good predictions.

Conclusion: -

The training and validation loss decrease consistently, indicating effective learning. Both losses remain close, suggesting no overfitting. The model is well-trained and performs reliably on new data.

Forecasting Results

Accuracy of Model: -

MAPE	6.35%
MAE	56228.28
RMSE	73421.26

Interpretation: -

Around 6.35% MAPE implies the model is about 93.65 % accurate in predicting the next 2 observations.

Conclusion: -

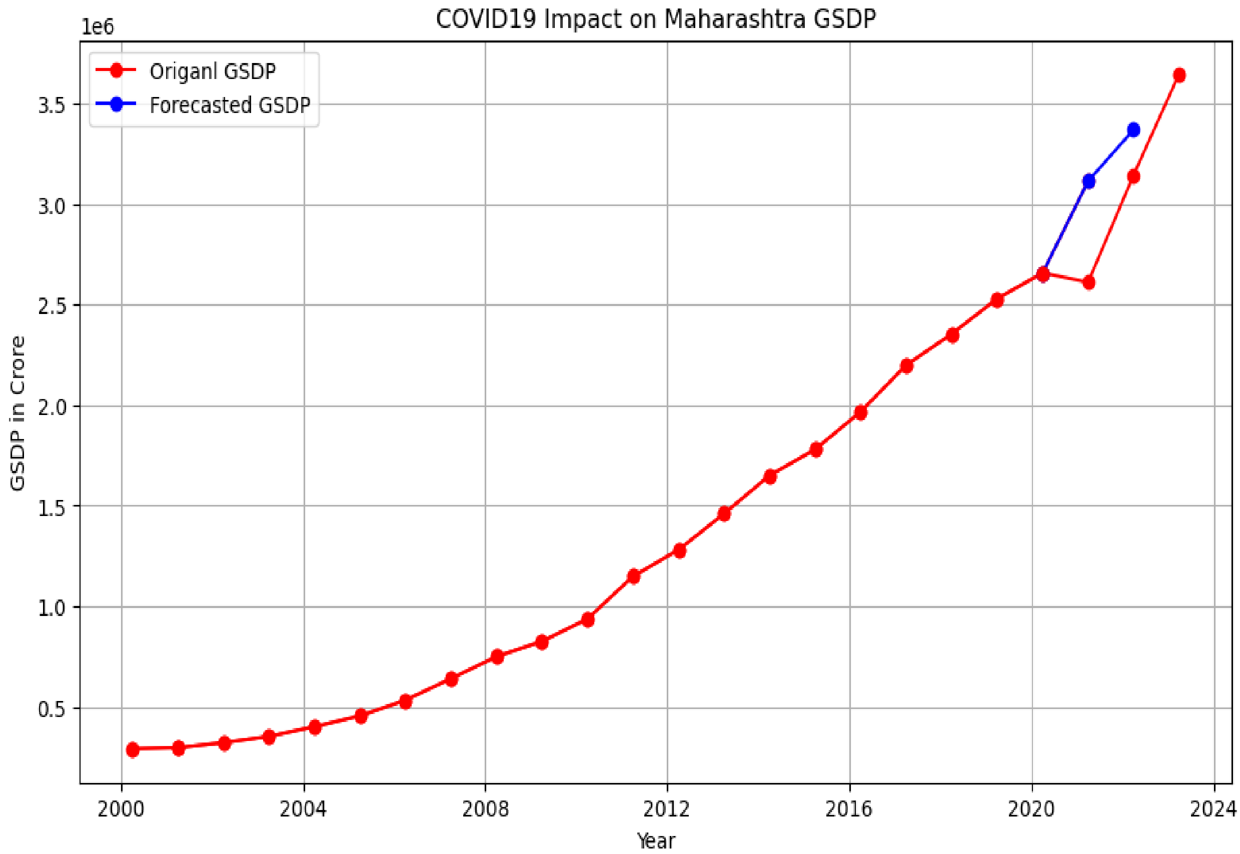
The model makes accurate predictions with errors averaging only 6.35% from the actual values. The average mistake in predictions is around 56228.28 units and larger errors push the RMSE to 73421.26. These numbers seem big but their significance depends on the scale of the data.

Impact of COVID-19: Forecasted vs. Actual GSDP for Maharashtra (2021-22)

Year	Forecast GSDP	GSDP
31-03-2021	3113574.8	2610651
31-03-2022	3368883.8	3144138

Impact of COVID-19 on Maharashtra Nominal GSDP

- **2021 Decline:** Actual GSDP (₹26.10 lakh crore) fell short of the forecasted ₹31.13 lakh crore, reflecting the economic slowdown due to lockdowns and reduced activity.
- **2022 Recovery:** GSDP rebounded to ₹31.44 lakh crore, nearing the forecasted ₹33.69 lakh crore, indicating economic recovery driven by policy support and business revival.



Interpretation: -

- The blue line shows how Maharashtra economy might have grown if COVID-19 had never happened. It moves up smoothly meaning the economy would have kept growing without any interruptions.
- The difference between the blue and red lines highlights the economic loss due to COVID-19.

Conclusion: -

Maharashtra economy grew steadily until 2019 but dropped in 2020 due to COVID-19. It started recovering in 2021 with government support and business reopening. Future predictions show strong growth, crossing ₹35 lakh crore by 2024. This shows Maharashtra economy is bouncing back and has a bright future.

Overall Conclusions

1. Maharashtra economy has grown steadily, with all sectors recovering post-pandemic, led by the services sector. Balanced growth will support future expansion.
2. The economy is recovering well after the pandemic, with contributions from all three sectors. A balanced approach will ensure long-term development.
3. Rising inflation is evident from the increasing GSDP Deflator, highlighting the need for inflation control and support for weaker sectors.
4. Real and nominal GSDP growth rates have fluctuated, with a sharp drop in 2020, but the economy has since recovered well.
5. Maharashtra economy has grown consistently except for a dip in 2020. It has bounced back strongly, and the future looks promising, though some challenges remain.
6. Maharashtra economy is expected to grow significantly, crossing ₹50 lakh crore by 2026, driven by industries, investments, and policies, despite challenges like inflation.
7. Maharashtra real GSDP is steadily growing and is projected to exceed ₹27.5 lakh crore by 2026, supported by industrial expansion and investments.
8. After a dip in 2020 due to COVID-19, Maharashtra economy rebounded in 2021 and is expected to cross ₹35 lakh crore by 2024, showing a strong recovery and bright future.

References

- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill.
- https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/
- OpenAI. (2024). *ChatGPT (March 31 version)*. Retrieved from https://openai.com
- Time Series Analysis and Forecasting using ARIMA models in R YouTube https://youtu.be/zB_0Yxxs0b4

Appendix

GitHub-Link:-

https://github.com/JadhavLakhan/Economic-Growth-Maharashtra