**Forecasting Inflation using Machine Learning Models: A Study on Consumer Price Index**

Aditya Srikant, Mohana Krishnan, Suresh Nepal

Department of Mathematics, School of Advance Sciences

Vellore Institute of Technology, Vellore – 632014, Tamil Nadu, India.

E-mail: suresh.nepal2022@vitstudent.ac.in, aditya.srikant2022@vitstudent.ac.in, [mohanakrishnan.v2022@vitstudent.ac.in](mailto:mohanakrishnan.v2022@vitstudent.ac.in)

***Abstract—For* policymakers and businesses in India, predicting monthly inflation is a crucial challenge. In this study, we assessed the accuracy of machine learning models such as Support Vector Machine Regressor (SVMR), Long Short-Term Memory (LSTM), and Random Forest in predicting the monthly inflation rate based on the Consumer Price Index (CPI) in India. We also compared the performance of traditional time series models such as ARIMA/SARIMA. Data from the time series span the years 2014 through 2022. In comparison to the LSTM model based on the mean sum of errors, we discovered that the ARIMA model and the Random Forest method performed better in terms of accuracy. The lack of sufficient data for training and testing of the LSTM model led to its low performance.is the LSTM model. More specifically, our results imply that conventional time series models like ARIMA and machine learning models like Random Forest may be practical methods for predicting monthly inflation rates in India.**

***Keywords***—***Time series models, Inflation forecasting, ARIMA, Machine learning models, Accuracy, Consumer Price Index (CPI)***

1. INTRODUCTION

The pace at which the prices of goods and services in an economy rise over time is measured by the important macroeconomic indicator known as inflation. For governments and policymakers to create successful macroeconomic policies, accurate inflation forecasting is essential. In order to increase the accuracy of forecasts, machine learning (ML) algorithms have been used in recent years to anticipate inflation and other aspects of the economy. The research being conducted intends to determine how well machine learning algorithms do at prediction of inflation in India. We use secondary data from the World Bank website to get inflation statistics in order to accomplish this purpose. We specifically contrast and assess the performance of neural networks, random forest methods and SVM Regressor models with that of traditional ARIMA/SARIMA models. Our study uses builds on the following studies. Bhanu and Shovon Sengupta[1] discuss the use of automated macroeconomic forecasting in India They show how SARIMA and XGBoost routinely place among the top five predictions in terms of accuracy scores, suggesting the better efficacy of ML approaches in forecasting inflation. By analysing the performance of several machine learning algorithms for inflation forecasting in the Indian setting, our work expands on the conclusions of these two articles. By examining the prediction precision of various models using our methods, we discover that the ARIMA and Support Vector models provide the most precise results.

1. METHODOLOGY

***Data collection:*** For the years 2014 via 2022, monthly inflation data based on the Consumer Price Index (CPI) in India have been collected.

***Data preparation***: To completely remove any blank or false data points, the data gathered was pre-processed and wiped away.

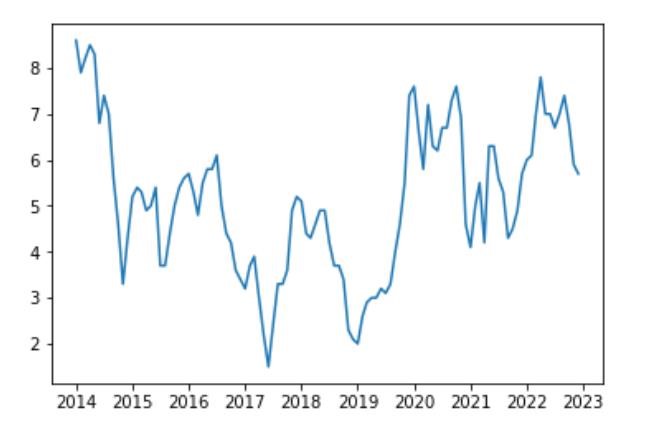
***Data splitting:*** The data was split into a training set and a testing set, with a split ratio of 80:20.

***Model selection:*** Five different models were selected for the study – ARIMA/SARIMA, SVMR, LSTM, and Random Forest.

***Model implementation:*** The selected models were implemented using Python programming language and Jupyter Notebook.

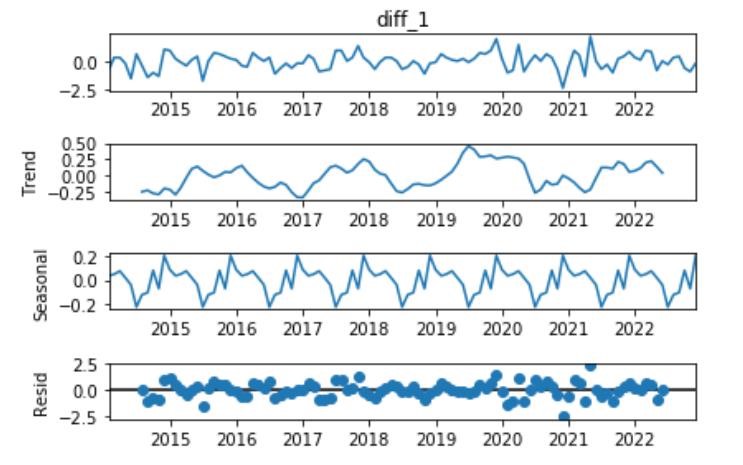
***Model evaluation:*** The testing set was used to assess the trained models. Each model was used to make predictions, and the Mean Sum of Error was committed to compare the models' accuracy. The absolute difference between the actual inflation rate and the projected rate of inflation for each month was taken, and the average of these differences across the testing set was used to figure out the mean total of error.

***Results analysis:*** For the purpose of determining which model performed best in terms of accuracy, the examination results of the models were investigated. The actual time series data plot is presented in **Figure 1.**



**Figure 1:** Plot of the time series data

**Figure 2** shows the decomposition of the time series data into trend, seasonal and residual component after differencing



**Figure 2**: Decomposition of the time series data

# ARIMA/SARIMA

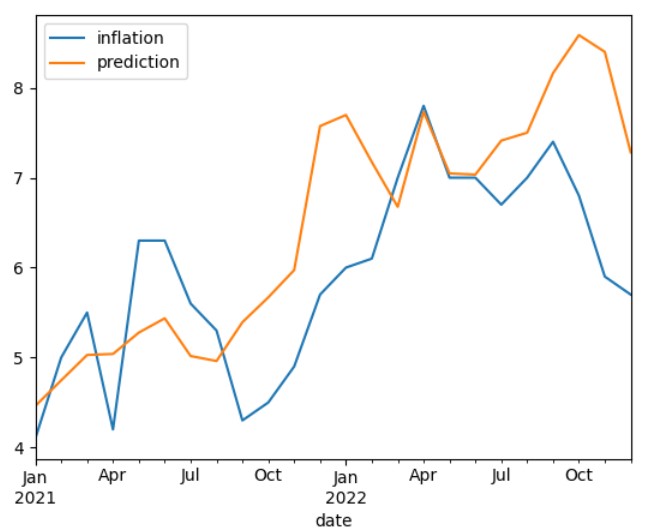
Traditional time series forecasting models like ARIMA (Autoregressive Integrated Moving Average) are used to estimate future values based on data from the past.

Since the ARIMA model only fits stationary data, first order differencing was employed to transform the non-stationary data into stationary data. The MSE (Mean Squared Error) for the ARIMA model was 0.6827.**Figure 3** shows actual data and predicted data using ARIMA model.



**Figure 3:** Actual data and predicted data using ARIMA model

The Seasonal ARIMA model was additionally developed using Auto ARIMA because the time series data contained a seasonal component. The SARIMA model did not outperform the ARIMA model despite the data having a seasonal component. Our SARIMA model's Mean Squared Error (MSE) was 1.1674. **Figure 4** shows actual data and predicted data using SARIMA model.

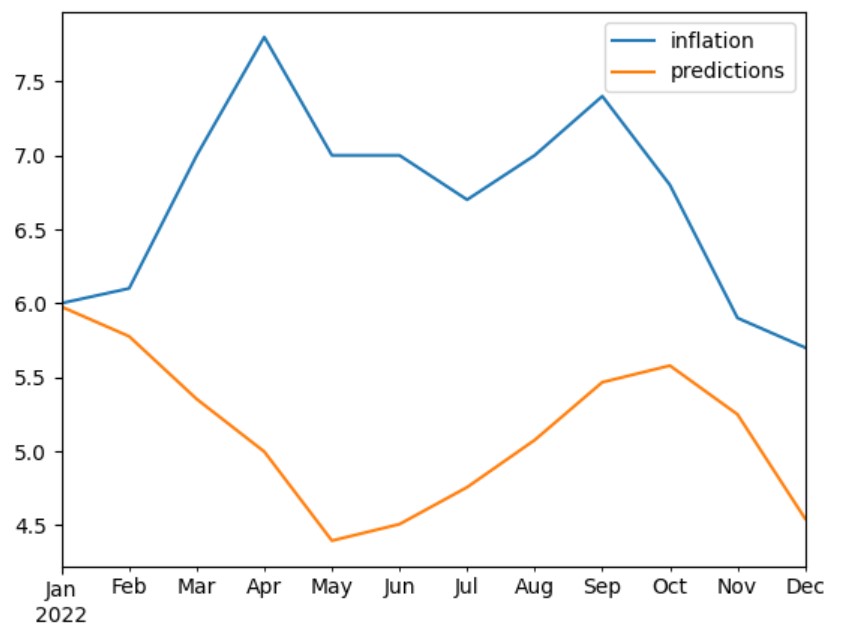


**Figure 4:** Actual data and predicted data using SARIMA model

# LSTM

A powerful machine learning tool for time series analysis, the LSTM (Long Short-Term

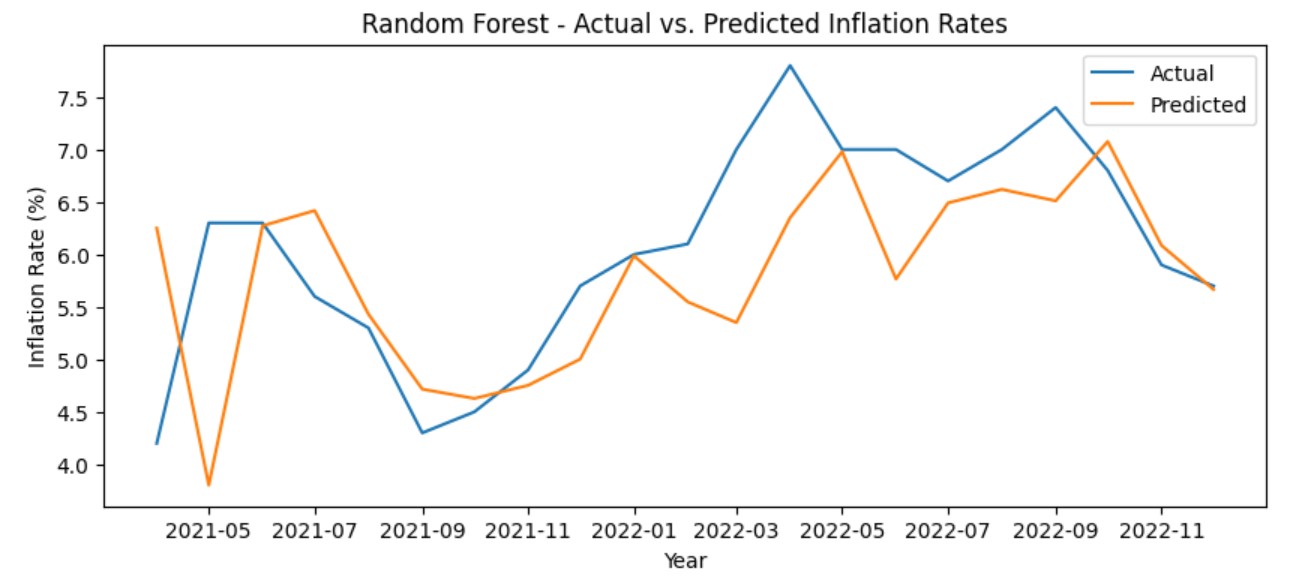
Memory) model is designed to handle data sequences with temporal dependencies. The Mean Squared Error (MSE) for the LSTM model was 3.1820. **Figure 5:** shows actual data and predicted data using LSTM model.



**Figure 5:** Actual data and predicted data using LSTM model

# RANDOM FOREST REGRESSOR

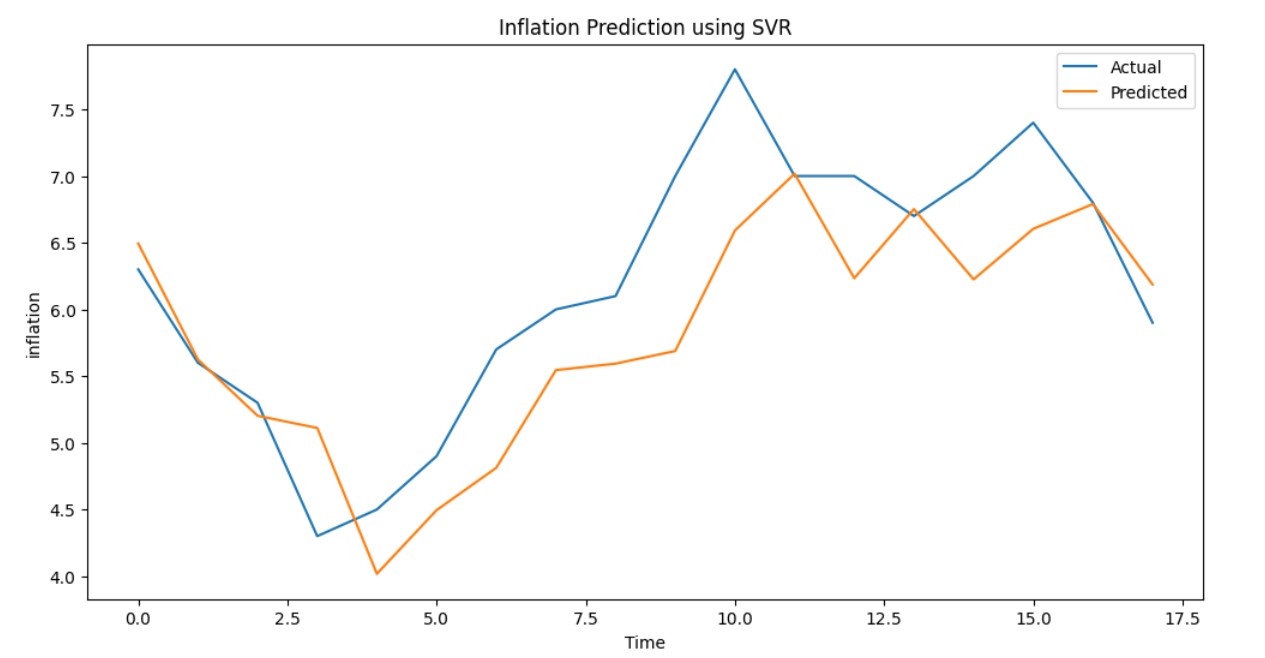
Popular machine learning algorithm Random Forest Regressor is used for regression applications. For the Random Forest Model, Mean Squared Error of 0.9318 was attained. **Figure 6:** shows actual data and predicted data using Random Forest model.



**Figure 6:** Actual data and predicted data using Random Forest model

# SUPPORT VECTOR REGRESSOR

A method based on machine learning titled the Support vector model Vector Regressor may be used to predict time series. A mean squared error of 0.4133 was reported for the Support Vector Model. **Figure 7:** shows actual data and predicted data using Support Vector model.



**Figure 7:** Actual data and predicted data using Support Vector model

1. RESULT AND CONCLUSION:

Using an indicator known as the Consumer Price Index, we explored how well different forecasting methods can predict how much prices will rise each month in India. We compare more recent machine learning-based techniques with more traditional ones, like Random Forest, Long Short-Term Memory, and Support Vector Machine Regressor. On the basis of the study, the SVMR model had the lowest error score (0.4133), making it the most accurate in making judgements. Higher error ratings were achieved by the other models, ARIMA, Random Forest, and LSTM.

Since the time series data for inflation are only reachable monthly, even if we utilise 10 years' worth of data, there will only be 120 records, which is why the LSTM performed poorly.

This study implies that SVMR could be a useful tool for anticipating inflation, while ARIMA/SARIMA are still viable options. In accordance to the study, it is possible to predict future economic trends by using computers to learn and analyse tremendous quantities of information in order.

1. REFERENCE

**[**1]Bhanu Pratap and Shovon Sengupta(2019). Macroeconomic Forecasting in India: Does Machine Learning Hold the Key to Better Forecasts? Working Paper Series, Reserve Bank of India, 2019

[2] Omprakash Yadav, Cynara Gomes, Abhishek Kanojiya, Abhishek Yadav(2019).

International Journal Of Information And Computing Science. ISSN NO: 0972-1347. Volume 6, Issue 5, May 2019

[3] Sven Heeren (2021). Forecasting Inflation using Machine Learning for an Emerging

Economy. Erasmus University of Rotterdam- Bachelor Thesis Econometrics and Operations Research, 2021

[4] Kriti Mahajan & Anand Srinivasan(2018). Inflation Forecasting in Emerging Markets: A

Machine Learning Approach. Issues in Central Banking, Chapter 2, BIS Publications

[5] Gour Sundar Mitra Thakur, Rupak Bhattacharyya, Seema Sarkar Mondal. Artificial Neural

Network Based Model for Forecasting of Inflation in India. Fuzzy Information and

Engineering, March 2016

[6] Gustavo Silva Araujo, Wagner Piazza Gaglianone (2020). Machine learning methods for inflation forecasting in Brazil: new contenders versus classical models

[7] Sune karlsson, Farrukh Javed, JUNE, 2016. Modeling and Forecasting Unemployment Rate In Sweden using various Econometric Measures. Meron Desaling (86/01/30)