IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

INFLATION PREDICTION

V. Vamsy Kartheek, SK. Mehraj Ali, G. Hima Brunda, K. Karthik
Student/Scholar, Student/Scholar, Student/Scholar, Student/Scholar
Computer Science and Engineering (Data Science),
Hyderabad Institute of Technology and Management (HITAM), Gowdavelly(V), Medchal(M), Medchal-Malkajgiri Dist. 501401, Telangana, India

Abstract Inflation occurs when there is a continuous rise in the prices of goods and services within a country. Forecasting the inflation rate can assist government departments and central banks in effectively utilizing their monetary policies to stabilize prices, thereby mitigating the impact of inflation on the market/economy, particularly for low and middle-income groups. The most well-known indicator of inflation is the Consumer Price Index (CPI), which measures the percentage change in the price of a basket of goods and services commonly consumed by households, such as groceries, haircuts, and travel expenses. Another measure of inflation can be done with Wholesale Price Index (WPI). Open-source data is available on the Indian government site MOSPI. We will examine all available datasets and apply data preprocessing methods to prepare the data for model learning. This study establishes an appropriate SARIMA model to forecast the Indian inflation rate for the period from February 2024 to July 2024, using univariate historical data of the country's inflation rates from 2014 to 2024.

IndexTerms **- Data science, Inflation, Data analysis, Data preprocessing, Augmented Dickey fuller test, Consumer Price Index(CPI), Wholesale Price Index(WPI), SARIMA.**

1. INTRODUCTION:

Inflation is a pressing subject in monetary literature, particularly giant for both developing and industrialized countries. Fig 01 explains the continuous rise in the average prices of goods and services throughout an economy, often stemming from a decline within the use of a forex price. Achieving fee stability is paramount, with low inflation fees indicating financial fitness, and high rates signalling instability.

Various elements contribute to inflation, along with fluctuations in meal fees, overseas rate hikes, and shifts in economic interest costs. In India, the Reserve Bank of India (RBI) performs a pivotal position as a valuable bank, wielding an effect on monetary guidelines that impact wage dynamics and commodity fees nationally.

Accurate inflation forecasts are important for guiding economic coverage choices. They function as essential inputs for policymakers, shaping monetary debates and informing macroeconomic techniques. In the realm of financial modelling, forecasting methods vary, with statistical and financial tactics dominating. These techniques leverage ancient facts and multivariate models to anticipate inflation traits.

Machine Learning techniques, inclusive of neural networks, clustering, and time series analysis, have emerged as treasured gear for uncovering styles and insights within inflation records. Their application extends throughout diverse domain names, together with commercial enterprise intelligence, finance, economics, and government policymaking, with potential implications for social sciences and humanities.

Fig 01: Factor that influences Inflation

1.1 TYPES TO MEASURE INFLATION

There are two types to measure inflation
- Consumer price index (CPI)
- Wholesale price index (WPI)

1.1.1 CONSUMER PRICE INDEX(CPI)

The Consumer Price Index (CPI) measures changes in the price level of a basket of goods and services used by households. It is calculated by periodically collecting the prices of selected items, providing a numerical estimate of these price changes at the consumer level. In contrast, the Wholesale Price Index (WPI) tracks price changes at the producer level but does not account for changes in the prices of services.
It involves comparing the overall market price level during a specific period to that of a previous period, known as the base year (typically set to 2012). CPI helps measure inflation, providing insight into how much prices are rising for the average person. Expressed as a percentage, the Consumer Price Index (CPI) helps estimate changes in the cost of living for individuals.

Indian CPI inflation is 5.09% in January 2024

Formula:

$$CPI = \frac{\text{Cost of a Fixed Basket of Goods and Services in the Base}}{\text{YearCost of a Fixed Basket of Goods and Services in the Current Year}} \times 100$$

1.1.2 WHOLESALE PRICE INDEX(WPI)

A Wholesale Price Index (WPI) gauges the fluctuation in the collective price of goods before their sale at retail, encompassing prices set by manufacturers and often wholesalers, particularly beyond India. Typically represented as a percentage shift from the preceding month or year, the WPI serves as an inflation gauge. A Wholesale Price Index (WPI) tracks the overall alteration in producer prices over time. It acts as an inflation metric derived from the prices of goods before they reach end consumers.
Wholesale price indexes are updated monthly to monitor the general pace of change in producer and wholesale prices. Beginning with a base period set at 100, the index is computed based on subsequent alterations in price for the collective output of goods. For instance, if January 2021 serves as the base period, and the aggregate price level escalates by 9.7% in the following year, the WPI for January 2022 would stand at 109.7.
Initially Inflation of India’s is calculated with WPI, which then in 2012 RBI decided WPI doesn't capture every sector's expense, so from 2014 January started implementing CPI with base year as 2012 = 100.

- It helps in calculating inflation based on the prices of services or goods before they reach end consumers.
- Current India’s WPI inflation is 0.27% (Jan2024)

  Formula:

$$\text{Wholesale Price Index} = \frac{\text{Price of Product in Current Year}}{Price\ of\ Product\ in\ Base\ Year} \times 100$$

The calculation of the Wholesale Price Index (WPI) in India initially included around 435 items during the base year 1993-94. However, the updated base year of 2011-12 expanded this to 697 items. Currently, the base year has been revised from 2004-05 to 2011-12 by the Office of Economic Advisor.

1.2 Calculation of Inflation

The inflation rate can be calculated by using the following formula:

$$Inflation\ Rate = \left(\frac{Final\ combined\ CPI(Current\ Month)}{Final\ combined\ CPI(Current\ Month)} - 1\right) \times 100$$

By using this formula, we will get the current inflation rate. India’s retail inflation eased from 5.7% in December 2023 to 5.10% in January 2024.In order to calculate the inflation, rate we need to calculate the CPI.

Formula:

$$CPI = \frac{\text{Cost of a Fixed Basket of Goods and Services in the Base}}{\text{YearCost of a Fixed Basket of Goods and Services in the Current Year}} \times 100$$

$$Annual\ CPI = \frac{Value\ of\ basket\ in\ Prior\ Year}{Vlaue\ of\ basket\ in\ Current\ Year} \times 100$$

Calculating the Consumer Price Index (CPI) for a range of products can be challenging, as each has its price and quantity. A common approach is to use a weighted average to address these variations. This involves multiplying the price of each item by its respective weight and then summing these values. The final step is to divide this sum by the total of all the weights.

Weighted Average:

$$x = \frac{\sum_{i=1}^{n}(x_i * w_i)}{\sum_{i=i}^{n} w_i}$$

2. LITERATURE SURVEY

This study [1] delve into the advancements in machine learning (ML) techniques and the emergence of new datasets for inflation forecasting. Contrary to previous skepticism in the literature, our findings reveal that ML models with an extensive array of covariates consistently outperform traditional benchmarks. Notably, the random forest model emerges as a standout performer, surpassing all other models.

This study [2] states, forward looking forecast of inflation rates are crucial for central banks and governments to stabilize prices and alleviate inflation's impact, especially on low and middle-income groups. Existing literature predominantly utilizes linear models like autoregressive (AR) and vector autoregressive (VAR) models for inflation rate prediction, given its significance in monitoring the macroeconomic landscape.

This study [3] focused where the monetary phenomenon is characterized by a widespread increase in prices, impacting various sectors, including the political arena and economic stability within a nation. The significance of inflation control cannot be overstated, as high and unstable inflation rates can lead to detrimental effects on both the economy and social fabric of society. Predicting the inflation rate emerges as a crucial solution for managing inflationary pressures and fostering economic resilience.

As outlined in reference [4], predicting inflation is crucial for devising strategies and policies aimed at sustaining economic stability. Two main approaches are commonly utilized for this task: the parametric regression method, employing the Autoregressive Integrated Moving Average (ARIMA) model, and the nonparametric regression technique, which relies on the local polynomial estimator.

This study [5] uses the Wholesale Price Index (WPI) as the primary indicator for measuring inflation. Initially, the researchers analyze the data's unit root and stationarity characteristics using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). The results show that the WPI series in India is non-stationary at its level but achieves stationarity upon first differencing. Through the Box-Jenkins methodology, they determine that the ARIMA (1, 1, 0) model best fits the WPI data. The forecasts predict a continuous increase in inflation starting from 2016-17, with confidence intervals indicating a steady annual rise in WPI throughout the forecast period from 2016-17 to 2024-25.

This study [6] shows that the Extreme Gradient Boosting (XGBoost) technique has gained popularity and recognition as a robust machine learning tool. It combines weak learners and employs a regularization method to mitigate overfitting. This study intends to apply the XGBoost method in conjunction with a genetic algorithm to predict inflation across three different forecast horizons.

This study [7] reexamines Stock and Watson's (1999) claim that commodity prices do not enhance traditional Phillips curve-based inflation forecasts. Utilizing Shin et al.'s (2014) methodology, the authors analyse historical data from 1957 to 2017. Their findings suggest that incorporating oil prices into the Phillips curve improves forecast accuracy, contingent upon capturing inherent predictor asymmetries. The study underscores the importance of refining inflation forecasting models to meet the evolving needs of monetary policymakers and analysts.

In economic practice, the annual inflation rate is commonly used, as noted in study [8]. It serves as a lagging one-sided moving average of both the annualized and monthly inflation rates and the Consumer Price Index (CPI), typically trailing about six months behind. The annual inflation rate functions as a smoothing mechanism, aiming to remove the seasonal component from the annualized inflation rate, assuming the presence of all seasonal unit roots. However, in reality, the CPIs and consequently, the annualized inflation rates often lack many of these seasonal unit roots. This disparity results in misleading cycles within the annual inflation rates, making it challenging to assess and interpret their developmental patterns accurately. Consequently, parameter estimates for models of the annual inflation rate are prone to bias and inaccuracy, leading to unreliable short-term forecasts.

The key finding of this study [9] underscores that implementing the auto speed acceleration algorithm improves the effectiveness of particles within the Particle Swarm Optimization (PSO), enabling them to discover superior solutions. These solutions involve enhanced initial weights for the Extreme Learning Machine (ELM), resulting in more accurate forecasting results. To evaluate the precision of the proposed technique, we employ comparison methods including backpropagation, ELM, hybrid ELM with genetic algorithms, and hybrid ELM with particle swarm optimization. Computational experiments indicate that the proposed approach achieves an RMSE of 0.01926, outperforming all comparison methods in terms of performance.

This study [10] developed an ARIMA(p,d,q) model to forecast the Philippines' inflation rate from 2018 to 2022 using historical data from 1960 to 2017. The best model was selected by assigning p, d, and q values based on ACF and PACF plots and unit-root tests, with the lowest AIC and error metrics (RMSE, MAE, MAPE) guiding the choice. ARIMA (1,0,0) emerged as the best model by AIC, predicting a 7.05% inflation rate by the end of 2018 and 8.93% in 2022. Alternatively, based on forecast error criteria, ARIMA (7,0,0) was identified as optimal, forecasting a 4.60% inflation rate for 2018 and 5.21% for the next twelve months.

3. PROBLEM STATEMENT

Inflation refers to the increase in prices over time, affecting the overall cost of living in a country. It diminishes the purchasing power of money, impacting consumers by making goods and services more expensive a general example as in the Fig 03. This decrease in purchasing power particularly affects fixed-income groups and savers. Furthermore, inflation doesn't uniformly affect all goods and services, leading to disparities in relative price increases.

We intend to make a Forecasting model which is trained on historical data to closely predict the inflation rate movements such that RBI monetary policy and other sectors try maintain a stable inflation rate which will avoid make huge impact on commodities.

4. REQUIREMENTS AND INSTALLATION:

a) Hardware Requirements:

- Processor: Handle scanning tasks efficiently using a multi-core processor with sufficient power of processing.
- Memory (RAM): At least 4GB of RAM is the minimum requirement to ensure the smooth running of the vulnerability scanning tool.
- Storage: Sufficient disk quota to keep the project files, dependencies, and any scan outcomes produced during the process
- Internet Connection: The internet connection, which is active, is needed to fetch dependencies, use web applications for scanning, and obtain updates or additional resources during the project's execution.

b) SOFTWARE

- Operating System: Any OS
- Network: Wi-Fi Internet or cellular Network.
- Jupyter Notebook or Google Collab: Used to run our dataset, compute required results, create models execute them on the dataset and obtain the accuracy.
- Packages: TensorFlow, Keras, Seaborn, Pandas, Matplotlib

5. METHODOLOGY:

- Utilization of Open Data Sources: Data available on the Indian government site MOSPI. We will go through all the available data sets and use the data preprocessing methods for preparing data for model learning.
- Forecasting Inflation Rates: Utilize the collected data to generate an approximate forecast for the inflation rate.
- Dashboards and Visualizations: Develop dashboards that allow users to explore data-driven insights and facilitate decision-making.

1. **Data Collection:** The initial phase involves gathering a comprehensive dataset from multiple reliable sources relevant to our forecasting goals. This dataset will undergo intense preprocessing to ensure its quality and readiness for analysis.

2. **Data Preprocessing:** This pivotal step involves cleaning, transforming, and organizing the dataset to ensure quality for analysis. Handling missing values, normalizing data, and encoding categorical variables are among the tasks to be addressed to create a robust dataset for analysis.

3. **Model Exploration:** In this phase, we will use various machine learning models, particularly focusing on Statistical models and deep learning architectures known for their capability to capture complex patterns and nuances in the data. Through a systematic exploration process, we will choose the most suitable models for our forecasting objectives.

4. **Forecasting and Visualization:** Applying the refined models to generate forecasts will be a core aspect of this project. The forecasts will be presented through comprehensive visualizations, aiding in the interpretation and understanding of the projected trends and patterns for better decision making.

6. Model and Architecture

In this section, we will elaborate on the process through which our SARIMA models operate, providing a detailed explanation of their functioning and methodology.

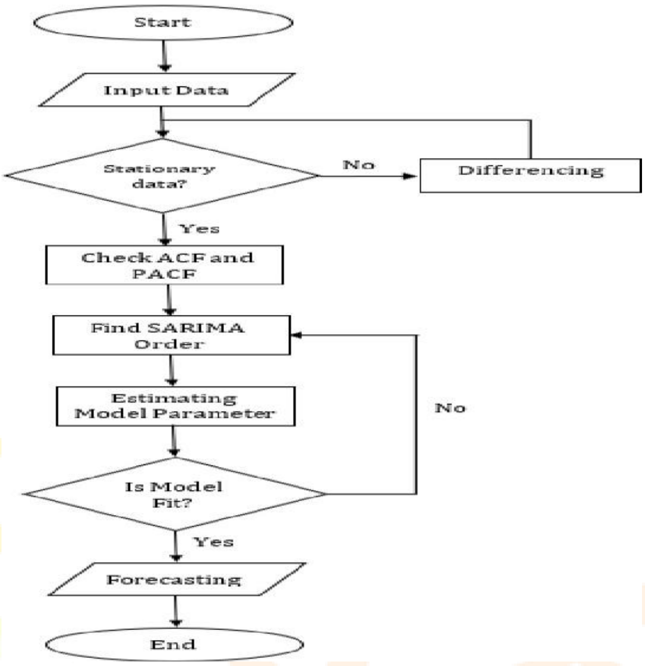


Fig 04: Flow Chart

The process for the inflation prediction model in accordance with Fig 04 is as follows:

Step 1: Input Data: Begin by importing the dataset containing the historical inflation rate data for India from January 2014 to January 2024.Check the format of the dataset and ensure it is in a suitable format for analysis. Create timestamps for each data point if they are not already available. If the date/time column is not in the correct data type, convert it to the appropriate format. Ensure the series is univariate, meaning it contains only one variable (inflation rate) over time. Ensure that there are no missing values on the continuous timestamp format.

Step 2: Stationary Data: Stationarity serves as a fundamental assumption in time series analysis. Assess the stationarity of the series through statistical tests like the Augmented Dickey-Fuller (ADF) test. If the series lacks stationarity, apply transformations such as differencing to stabilize both the mean and variance.

Step 3: Differencing: Before applying differencing, it's essential to check whether the time series exhibits trend or seasonality components. Trends can be linear or nonlinear, and seasonality can be periodic fluctuations occurring at regular intervals. Depending on the trend of data we apply logarithmic transformation (log(x)) or a square root transformation (sqrt(x)) that can help in stabilizing the variance and achieve stationarity.

Step 4: Check ACF & PACF: AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) plots are crucial for determining the parameters of the ARIMA model. ACF plots show the correlation between each observation and its lagged values, while PACF plots show the correlation between observations separated by a specific lag.

Step 5: Find SARIMA Order: Identify the p (autoregressive) and q (moving average) parameters from the ACF and PACF plots. The lag values where the ACF and PACF plots intersect or cross the significance threshold are the values of p, d and q, respectively. The number of times the difference operation is performed to achieve stationarity is the value of d.

Step 6: Estimating Model Parameters: With the processed data and identified ARIMA parameter values (p, d, q). We also identify a perfect fit parameter for SARIMA which exhibits seasonality (P, Q,D,m)

Step 7: Is Model Fit: The Akaike Information Criterion (AIC) was utilized to select the optimal ARIMA(p,d,q) model for forecasting. The model with the lowest AIC value among all iterative combinations is chosen as the best-fit parameter.

Step 8: Forecasting: Once the SARIMA model is trained, use it to predict future values of the inflation rate. Evaluate the performance of the model with MAE, MAPE and RMSE.

7. DATA AND SOURES OF DATA

We are gathering data from the Ministry of Statistics and Programme Implementation (MoSPI) website, which provides comprehensive information on both the Consumer Price Index (CPI) and inflation. The MoSPI platform serves as a reliable source for obtaining separate datasets on these metrics.

Our research work encompasses a thorough exploration of time series analysis, which involves preparing the data, transforming it into suitable formats, and selecting appropriate models for training. We have rigorously trained these models and evaluated their performance to identify the most effective ones for our needs. Furthermore, we've applied various time series forecasting methods, as detailed in relevant research papers, to project future trends.

Data is collected by NSO from 1181 village markets and 1114 urban markets distributed over 310 towns/cities of the country. The MoSPI website consists of both the CPI and inflation data separately. Data set consists of different metrics and categories like Rural, Urban, State, Food Inflation, General index Inflation etc. Annual inflation is considered a combined (rural and urban) General index of CPI.

8. IMPLEMENTATION

8.1 DATA EXPLORATORY ANALYSIS

In the previous phase, we got our hands on the data but it showcases a comprehensive view.

Sl. No.	State Name	Food and beverages			Pan, tobacco and intoxicants			Clothing and footwear			Housing			Fuel and light			Miscellaneous			All Groups		
		R	U	C	R	U	C	R	U	C	R#	U	C$	R	U	C	R	U	C	R	U	C
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
22	Rajasthan	190.4	191.8	190.8	196.1	199.1	196.5	193.9	187.3	191.8		183.6		154.7	170.8	159.1	182.8	175.3	180.0	185.7	184.0	185.1
23	Sikkim	195.2	192.6	194.6	187.3	211.7	192.2	174.2	188.0	177.3		187.9		293.0	146.3	263.3	187.5	171.4	181.8	197.7	184.7	193.5
24	Tamil Nadu	197.8	200.7	199.3	212.1	212.3	212.2	196.5	178.7	187.1		189.9		192.3	197.0	194.8	182.2	181.4	181.7	193.0	191.4	192.1
25	Telangana	202.5	210.6	206.3	207.6	210.9	208.6	217.5	183.3	199.0		191.1		167.2	167.4	167.3	199.3	185.0	191.5	200.8	195.5	197.9
26	Tripura	196.6	199.0	197.1	231.4	187.9	225.5	177.3	189.4	179.7		214.2		349.0	224.2	318.9	190.5	178.8	187.1	206.5	197.3	204.1
27	Uttar Pradesh	192.3	193.2	192.6	187.9	197.8	190.2	189.3	182.9	187.3		190.0		178.5	173.1	176.9	177.3	173.8	176.0	186.7	185.0	186.1
28	Uttarakhand	188.0	197.6	191.0	195.2	203.6	197.6	187.6	214.1	195.4		177.3		171.7	184.9	175.4	170.8	169.0	170.2	181.9	186.2	183.5
29	West Bengal	193.0	206.9	198.0	229.9	222.6	227.5	197.9	182.1	191.2		172.2		196.3	186.5	192.5	187.3	177.9	182.2	193.3	188.9	191.2
30	Andaman & Nicobar Islands	216.6	198.0	208.9	206.4	228.4	212.2	187.8	162.6	176.3		213.1		147.4	164.8	154.7	179.7	170.8	174.7	198.5	187.2	192.8
31	Chandigarh	189.3	186.6	186.8	201.4	227.8	224.1	191.7	177.3	178.4		161.1		146.6	138.6	139.4	186.2	168.2	169.3	184.4	172.0	172.7
32	Dadra & Nagar Haveli	169.8	190.2	182.2	176.8	166.0	174.6	194.2	176.1	182.2		163.4		168.7	166.9	167.8	173.3	186.3	182.3	172.6	181.8	178.7
33	Daman & Diu	178.8	187.3	181.9	210.5	190.6	208.4	204.2	208.1	205.7		165.9		246.2	164.7	216.8	172.1	176.9	174.4	187.9	180.6	184.8
34	Jammu & Kashmir*	185.9	192.3	187.6	215.9	212.6	215.1	219.2	208.2	216.2		206.0		204.6	203.6	204.4	199.1	177.8	191.5	193.9	193.3	193.7
35	Lakshadweep	206.1	179.2	194.1	264.5	250.9	257.2	201.9	177.2	191.4		281.0		263.4	175.2	219.1	157.2	129.0	142.2	194.6	176.4	185.3
36	Puducherry	202.4	201.4	201.7	227.0	209.2	216.3	230.1	199.3	208.7		177.2		205.0	252.5	237.0	176.2	173.8	174.4	196.1	189.2	191.0
99	**All India**	190.2	196.7	192.6	202.9	208.4	204.4	193.7	182.4	189.2		177.9		182.4	175.8	179.9	182.5	174.4	178.6	188.2	184.2	186.3

Notes:

R : Rural.
U : Urban.
C : Combined.
-- : indicates the receipt of price schedules is less than 80% of allocated schedules and therefore indices are not compiled.
: CPI (Rural) for housing is not compiled.
$: CPI (Combined) is not compiled since CPI (Rural) is not compiled.
* : Figures of this row pertain to the prices and weights of the combined Union Territories of Jammu & Kashmir and Ladakh (erstwhile State of Jammu & Kashmir).

Fig 05: Original raw data

We found that the initial dataset wasn't ideal for training purposes Fig 05, so we trimmed it down to include only tri-variables data as shown in Fig 06. During our analysis, we discovered a gap in the data spanning two months in 2020, a period affected by the COVID-19 pandemic. To fill these gaps, we introduced two random low values to approximate the impact of the economic shutdown during that time. As a result, our dataset now consists of just the year, month, and combined inflation rate for urban and rural areas, totaling 121 records. Throughout the dataset, we observed fluctuations in the inflation rate, ranging between 4% and 7%.

Year	Month	Combined
2014	1	8.6
2014	2	7.88
2014	3	8.25
2014	4	8.48
2014	5	8.33
2014	6	6.77
2014	7	7.39
2014	8	7.03
2014	9	5.63
2014	10	4.62
2014	11	3.27
2014	12	4.28
2015	1	5.19
2015	2	5.37
2015	3	5.25
2015	4	4.87
2015	5	5.01
2015	6	5.4
2015	7	3.69
2015	8	3.74
2015	9	4.41

Fig 06: Tri-Variable Data

8.2 SEASONAL DECOMPOSITION

Time series data often reveal underlying trends and repeating seasonal patterns as shown in Fig 07. To analyze these characteristics, we need to visually inspect the data to identify linear trends and consistent seasonal behaviors over time. This stage involves plotting the data to detect any upward or downward trajectories (indicating a trend) and recurring fluctuations at regular intervals (suggesting seasonality). By examining the data visually, we can gain insights into its structure and develop a better understanding of the dynamics driving the observed patterns.

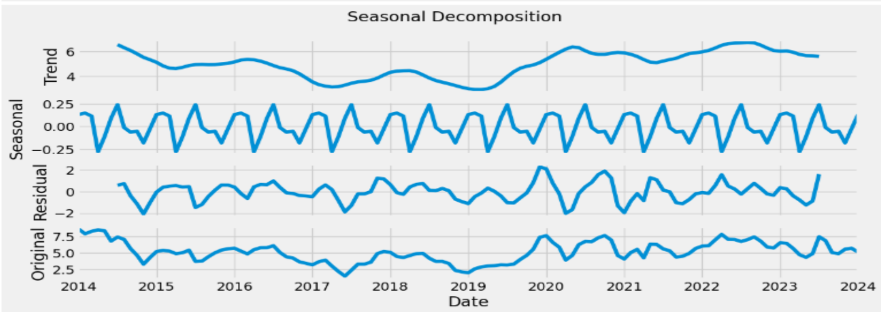


Fig 07: Seasonal Decomposition

8.3 ARIMA (Autoregressive Integrated Moving Average):

ARIMA is a statistical model used to predict future values by leveraging past data points. As an autoregressive model, it incorporates previous observations to forecast what might happen next. This reliance on historical data allows ARIMA to identify and understand patterns or trends, enabling it to make predictions based on the model's understanding of past behaviors. It is mainly composed of three main components which work collectively to achieve the target

1. Autoregressive (AR): The ‘AR’ component of the model has the current value of the series as a linear function of a certain number of past values (lags) of the series itself. It captures the dependence of the current value on its recent history.

2. Integrated (I): The ‘I’ component refers to differencing. If the data has a trend (non-stationary), differencing is applied to remove the trend and make the series stationary.

3. Moving Average (MA): The ‘MA’ component incorporates the error terms (forecasting errors) from past steps to improve the model's accuracy. This essentially takes a weighted average of past forecast errors to adjust future forecasts.

ARIMA Formula:

$$\emptyset(B)(w_t - \mu) = \theta(B)a_t$$

The equation provided in [10] is explained as follows: the time index is represented by t, and the backshift operator as B, ∅ (B) for autoregressive (AR)(p), θ(B) for moving average (MA)(d), and w_t for non-seasonal (I)(d) in the ARIMA (p, d, q) model.

8.4 SARIMA

In our forecasting work, we are confronted with the challenge of having limited data and no variables to rely on other than time. Under these circumstances, among all the available models, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model has proven to deliver the best results.

- Unlike ARIMA (Autoregressive Integrated Moving Average) which focuses on non-seasonal data, SARIMA (Seasonal Autoregressive Integrated Moving Average) explicitly considers seasonal patterns in the data.
- SARIMA is a powerful tool for forecasting time series data with seasonal patterns.

Apart from the parameter of ARIMA (p, d, q), SARIMA also has its own similar parameters which exhibits seasonal characteristics with lag indicators (P, Q, D, m).

SARIMA Formula:

$$y_t(1-B^m)^D(1-\phi_{m,1}B^m-\cdots-\phi_{m,p}B^{pm}=\varepsilon_t(1+\theta_{m,1}B^m+\cdots+\theta_{m,Q}B^{Qm})$$

This SARIMA formula theatrically explained by [26], which can be described and understand by few key points below:

- y_t is the observed time series at time tt.
- B is the backshift operator
- m is the seasonal period.
- D is the degree of seasonal differencing.
- $\phi_{m,i}$ (for i=1, …, P) are the coefficients of the seasonal autoregressive (SAR) terms.
- ε_t is the white noise error term at time t.
- $\phi_{m,j}$ (for j=1, …, Q) are the coefficients of the seasonal moving average (SMA) terms.

8.4.1 PARAMETERS and Its Working

p (AR order):
The number of past values of the series included in the AR model. AR (2), the model considers the effects of the previous two lags (t-1 and t-2) on the current value (t).

d (I order):
The number of times the data is different to achieve stationarity. I(1) indicates that the data has been differentiated once to achieve stationarity.

q (MA order):
The number of past forecast errors included in the MA model. MA (2), the model considers the effects of the previous two error terms (ε_(t-1) and ε_(t-2)) on the current value.

SARIMA also has the same functional parameters but with seasonal influence in it.

P (seasonal AR order): Similar top with the seasonal component.

D (seasonal I order): Similar to d with the seasonal component.

Q (seasonal MA order): Similar to q with the seasonal component.

m (seasonality): The number of periods in a single season (e.g., 12 for monthly data).
8.5 Fighting SARIMA

To tackle this challenge, we are simply iterating in a loop to make the model learn with a specified range of values. In the outcome, we get the best combination of values for the parameter with the lowest AIC (Akaike information criterion) and BIC (Bayes information criterion) value.

$$AIC=-2\log\hat{L}(\theta|data,g)+2K$$
$$BIC=-2\log\hat{L}(\theta|data,g)+\log(n)K$$

The main reason to choose BIC with AIC is that AIC does not take sample size into account.

- L (θ | data, g) is the likelihood function of model parameters given the data and the model g. In short, its likelihood that your model fits your data, A larger value the better.
- K is the number of parameters to be estimated.
- n is the sample size.

It is said that the model with the lowest AIC is better and should be used an example of that is shown in Fig 08.

```
Best order: (1, 1, 2)(0, 1, 2)[12] with AIC = 198.797518 at line 204

(Total time of execution: 1 min 51.87 s)
```

Fig 08: Best ARIMA and SARIMA Parameter

8.6 Check ACF & PACF

8.6.1 ACF Plot:

- The ACF plot displays the relationship between a time series and its lagged versions. Essentially, it quantifies the extent to which the present value of the series is connected to its past values at different time intervals (lags).
- By looking at the pattern of the ACF plot, you can identify:
    - Trend: A significant spike at lag 1 might indicate a trend (positive or negative) in the data.
    - Seasonality: Spikes at regular intervals corresponding to the seasonality period (e.g., every 12 lags for monthly data) indicate seasonal patterns.
    - Autocorrelation: Decaying or oscillating correlations at lags beyond 1 suggest an autoregressive (AR) process, where past values influence future values.

8.6.2 PACF Plot:

- The PACF plot focuses on the correlation between a time series and its lagged versions, but it removes the influence of past lags. This helps isolate the unique contribution of each lag to the current value.
- The interpretation of the x and y-axes is similar to the ACF plot.
- A significant spike at a particular lag in the PACF plot indicates that the lag has a direct impact on the current value, independent of any past lags. Please refer Fig 09 for understanding.

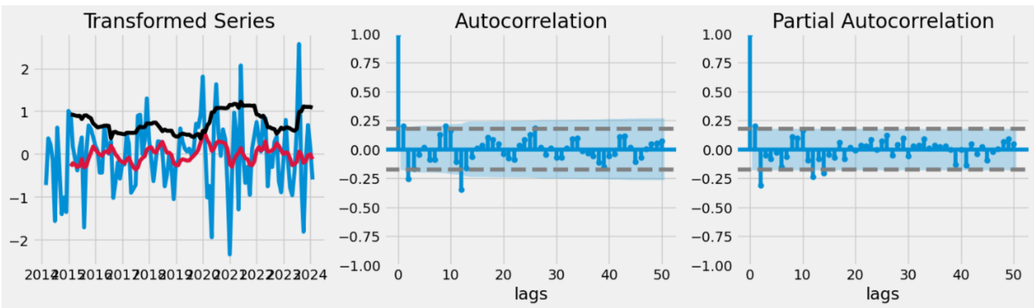


Fig 09: ACF& PACF Plots

8.7 Forecast Evaluation:

Forecast evaluation was done using the various forecast error statistical tools which are modeled as follows

Root Mean Squared Error (RMSE):

$$\mathrm{R.M.S.E.} = \sqrt{\sum_{t=T+1}^{T+h} \frac{(\hat{y}_t - y_t)^2}{h}}$$

Mean Absolute Error (MAE) :

$$M.A.E = \sum_{t=T+1}^{T+h} \frac{|\hat{y}_t - y_t|}{h}$$

Mean Absolute Percentage Error (MAPE):

$$M.A.E = 100 \times \sum_{t=T+1}^{T+h} \frac{\left|\frac{\hat{y}_t - y_t}{y_t}\right|}{h}$$

Akaike information criterion (AIC):

$$AIC = -2\log \hat{L}(\theta|data, g) + 2K$$

Bayes information criterion (BIC):

$$BIC = -2\log \hat{L}(\theta|data, g) + \log(n)K$$

9. TESTING/RESULTS

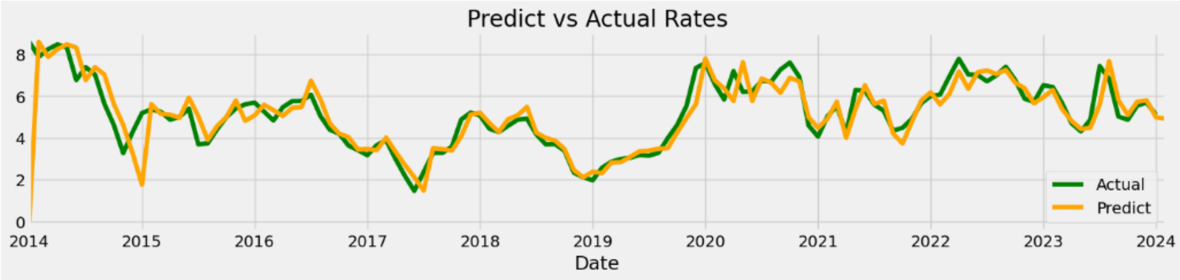
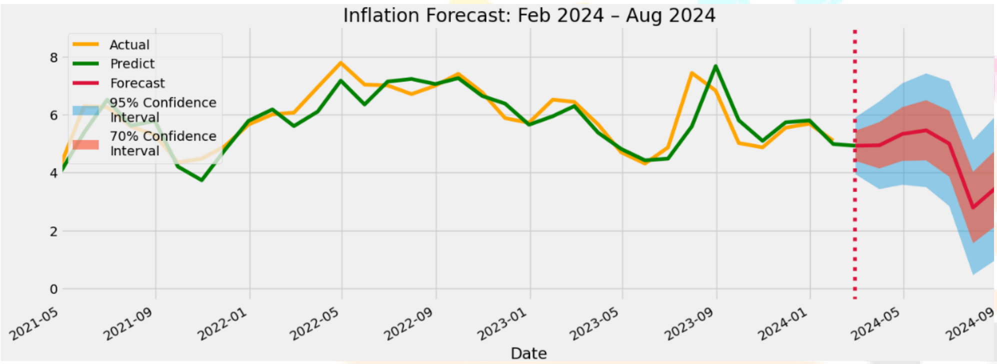


FIG 11: PREDICT VS ACTUAL RATES

The performance of the predictive model displayed in Fig 11 a discernible pattern over time. Initially, between 2014 to 2016, the model's predictions exhibited significant fluctuations from the actual values, indicating challenges in capturing underlying trends. However, between 2018 to 2020, there was a noticeable improvement as the predicted values began aligning more closely with the actual values, suggesting model refinement and better adaptation to the data. Despite this progress, deviations were observed in 2021 to 2023, possibly due to unforeseen events or shifts in data dynamics. Nonetheless, the model's performance rebounded, with predictions gradually converging with actual values thereafter. This pattern underscores the iterative nature of model development, emphasizing the importance of ongoing refinement and adaptation to ensure accurate predictions, especially in the face of evolving circumstances.



	Forecast [%]
2024-02-29	4.9
2024-03-31	4.9
2024-04-30	5.3
2024-05-31	5.5
2024-06-30	5.0
2024-07-31	2.8

Fig 12: Inflation Forecast from Feb 2024 – Aug 2024

From the graph Fig 12 above, based on our predictions using data up to February 2024, the model forecasts that the inflation rate for March 2024 will fall within the range of 4.2% to 6%. This range reflects the uncertainty inherent in predictive modeling and accounts for potential variations or unforeseen factors that could influence inflation levels. The model's prediction suggests a moderate level of inflation for March 2024, indicating a stable but not excessively high or low rate of price increases across the economy.

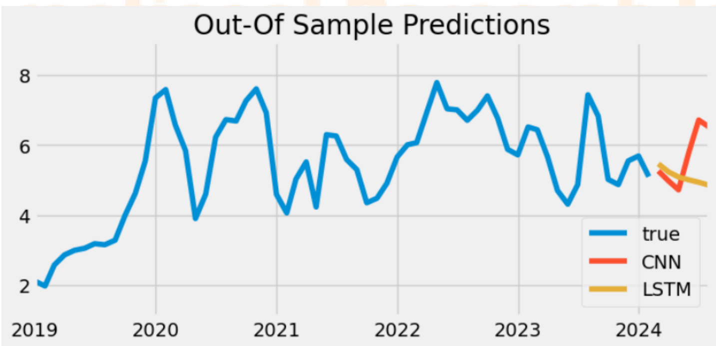


Fig 14: Out - Of Sample Predictions

In the above Fig 14, Testing CNN and LSTM models initially resulted in substantial errors between the actual and predicted values, indicating difficulties in effectively capturing the underlying patterns within the dataset. This could be attributed to factors such as the complexity of the data or limitations in the models' architectures to handle the temporal dependencies inherent in the data. However, upon transitioning to SARIMA, the predictions notably improved, aligning closely with the actual values. SARIMA's success suggests its proficiency in modeling the dataset's temporal dynamics, including seasonal fluctuations and autocorrelation. This highlights the importance of selecting a modeling approach that aligns with the characteristics of the data. Moving forward, leveraging insights from model comparisons can guide further refinements to enhance predictive accuracy and ensure robust performance in future forecasting tasks.

Model Selection Criteria:

- AIC: 158.16
- BIC: 173.36

- HQIC: 164.30

Lower values of AIC, BIC, and HQIC indicate better model fit with a penalty for model complexity. Here's a breakdown:

- AIC: Relatively low compared to BIC, suggesting the model balances goodness-of-fit with complexity reasonably well.
- BIC: Slightly higher than AIC, potentially acceptable depending on the number of alternative models considered. BIC tends to favor simpler models more than AIC.
- HQIC: Falls between AIC and BIC, offering a compromise between goodness-of-fit and complexity.

Error Measures:

- RMSE: 1.01: RMSE measures the average magnitude of the errors between predicted and observed values. It is sensitive to large errors due to squaring the differences. Lower RMSE values indicate better predictive accuracy.
- MAE: 0.51: Similar to RMSE, this suggests the model's predictions are generally within 0.51 units of the actual values on average.
- MAPE: 0.09: The MAPE value of 0.098292 (or approximately 9.83%) indicates the model's predictions are, on average, off by about 9.83% from the actual values. This is a relatively low error rate, implying good predictive performance.

CONCLUSION

Our objective was to construct the most reliable SARIMA model for forecasting India's inflation rate from February 2024 to July 2024. We employed various time series analysis methods, including the AutoCorrelation Function (ACF), Partial AutoCorrelation Function (PACF), stationarity analysis, Akaike Information Criterion (AIC), and forecast evaluation criteria, to identify the most appropriate model for capturing inflation trends.

Our analysis identified the SARIMA (0,1,2) (1,0,2,12) model as the strongest parameters based primarily on the AIC criterion. However, considering forecast evaluation metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), the SARIMA model consistently emerged as the optimal choice. Using the SARIMA model, our forecasts reveal it has down trend in inflation over the next six months, with a predicted inflation rate reaching 2.8 % by July 2024.

Considering both the error measures and model selection criteria, the model demonstrates promising results. The low RMSE, MAE, and MAPE values indicate good forecasting accuracy. Additionally, the AIC and HQIC values suggest that the model achieves a good balance between capturing the data's patterns and avoiding overfitting.

SCOPE OF THE PROJECT

The scope of this project encompasses the development and implementation of a robust forecasting model for predicting inflation trends in India. By leveraging historical data and advanced time series analysis techniques, we aim to provide accurate and timely insights into inflation dynamics.

The project can explore the potential impact of inflation on various sectors and socioeconomic groups, providing valuable information for policymakers and stakeholders to make informed decisions. Model can be improved by introducing model with factor system annotated with inflation with economical variations, for example by feeding model with news articles on real time.

I. Acknowledgment

We would like to thank our guide **Dr M. Rajeshwar** sir for his important suggestions to improve the standard of the paper. We are also grateful to her for helping us review our performance regularly. We would also like to thank the Department of Computer Science Engineering (Data Science), HITAM, Hyderabad.

References

[1] Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2019). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. Journal of Business & Economic Statistics, 1–45.

[2] Algorithms Cheng Yang And Shuhua Guo1,School of Economics, Yunnan University, Kunming 650000, China. Inflation Prediction Method Based on Deep Learning.Computational Intelligence and Neuroscience.Volume 2021

[3] Oktanisa, I., Mahmudy, W. F., & Maski, G. (2020). Inflation Rate Prediction in Indonesia using Optimized Support Vector Regression Model. Journal of Information Technology and Computer Science, 5(1), 104–114

[4] Vita Fibriyani and N Chamidah 2021.Prediction of Inflation Rate in Indonesia Using Local Polynomial Estimator for Time Series Data. Konferensi Nasional Penelitian Matematika dan Pembelajarannya (KNPMP) V, 5th August 2020, Surakarta, Indonesia. Volume 1776

[5] Karthikeyan, Parthasarathy; Manikandan, M.; Mani, N.; Gopalakrishnan, Badri Narayanan.Forecasting Inflation Rate in India.Academic Journal. Finance India, 2021, Vol 35, Issue 4, p1209

[6] Li, YS., Pai, PF. & Lin, YL. Forecasting inflation rates be extreme gradient boosting with the genetic algorithm. J Ambient Intell Human Comput 14, 2211–2220 (2023).

[7] Afees A. Salisu, Kazeem O. Isah. Predicting US inflation: Evidence from a new approach.Center for Econometric and Allied Research (CEAR), University of Ibadan, Nigeria.Volume 71, April 2018, Pages 134-158

[8] Josef Arlt.The problem of annual inflation rate indicator.11 March 2021

[9] Wayan Firdaus Mahmudy1* Adyan Nur Alfiyatin1 Candra Fajri Ananda2 Agus Wahyu Widodo1.Inflation Rate Forecasting Using Extreme Learning Machine and Improved Particle Swarm Optimization.International Journal of Intelligent Engineering and Systems.2021

[10] Allemar Jhone P. Delima, Maria Tavita Q. Lumintac.Application of Time Series Analysis for Philippines' Inflation Prediction.International Journal of Recent Technology and