Impact of Lockdown on India's Index of Industrial Production – Traditional and Deep Learning Statistical Approach

Preethi Patil, Jyothirani S. A., and Haragopal V.V.

***Abstract* — Index of Industrial Production (IIP) data is one of the important economic indicators that track the manufacturing activity of different sectors of an economy. In this paper, an attempt is made to forecast the IIP data using traditional and deep learning statistical approaches. The data from Apr-2012 to Feb-2020 is used for forecasting. The appropriate best model is evaluated by comparing mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The results of the study show that RNN is performing better than the other models i.e ARIMA (Traditional method), FFNN, and LSTM (ANN models). Therefore RNN model is used for forecasting. The forecasted values from Mar-2020 to Jun-2021 are compared with the actual IIP values and resulted in a clear decline in industrial production because of lockdown.**

***Keywords*—ARIMA, FFNN, Index of Industrial Production (IIP), LSTM, RNN.**

I. INTRODUCTION

In India, the Index of Industrial Production (IIP) is an index that captures the growth of various industrial sectors such as Manufacturing, Electricity & Mining. The short-term changes in the volume of Industrial production are measured by this Index [1].

The industrial productions are often influenced by seasonal fluctuations, trend fluctuations as well as calendar & trading day effects, which cover relevant short and long term movements of the time series. The climatic and natural factors may also influence the consumption and production behavior of industries [1].

The Indices of Industrial Production are to systematically grasp activities relating to production, shipment, and inventory in India. Factories in India have manufactured various products, and the Indices of Industrial Production have been prepared as a comprehensive indicator of wide-ranging production activities and are regarded as some of the most important indices among economic indices [2].

The industries sector accounts for a larger share of economic activities in India i.e., around 24.29 % (MOSPI, 17-June-2021) of GDP. The Gross Domestic Product (GDP) on a quarterly basis is compiled using the Gross Value Added (GVA) of the industry sector.

Indices of Industrial Production respond sensitively to economic conditions. Production in industries shows huge fluctuations depending on economic conditions. Characteristically, it shows significant reactions to the economic situation; for example, production is reduced and inventory adjustment takes place when inventory piles up due to an economic depression. On the other hand, inventory is accumulated in the prospect of an expansion of demand when the economic situation improves. Economic fluctuations, such as an inventory cycle, can be understood from the Indices of Industrial Production. On the other hand, tertiary industries such as the service industry do not indicate significant fluctuations, compared with secondary industries such as the manufacturing industry. For this reason, changes in the GDP tend to be generated from the category of mining and manufacturing industries, and movements in the Indices of Industrial Production can indicate the direction of change in the GDP [2].

This data is compiled every month, six weeks after the reference month ends & published by MOSPI. It represents the status of production in the industrial sector for a given period compared to the base period. The base period is assigned with an index level of 100 & the current base year is 2011-12.

The study aims to forecast the Index of industrial production using traditional & deep learning models and Identify the best model for forecasting using the monthly data from April-2012 to Feb-2020.

Published on August 29, 2022.
P. Patil, Department of Statistics, Osmania University, India.
(corresponding e-mail: preethipanchakshari@gmail.com)
Jyothirani S. A., Department of Statistics, Osmania University, India.
(e-mail: jyothi2263@gmail.com)
Haragopal V. V., Department of Mathematics, BITS-Pilani, India.
(e-mail: haragopal_vijjha@yahoo.com).

II. LITERATURE REVIEW

Singh, Salam Shantikumar, L. T. Devi, and T. Deb Roy: "The paper studies the effects of seasonal and trend on IIP in India. Itcomprises general IIP and 26 sub-sectors of industrial productions. "A well-known time series model, ARIMA (p, d, q) model is used to analyze the time-series data. The finding suggests that both seasonal and trend effects are present in IIP. A forecast of the future IIP of India is made by this model after adjusting the effects of short-term and long-term variations" [1].

Çekim, Hatice Öncel: "In this paper, the time series analysis is conducted to the monthly industrial production index of turkey data calculated between 2005 and 2017 and as a result of the analysis, the SARIMA(1,1,1)(3,2,0) $_{12}$ model is determined as the most suitable model for the series. Using this model, the forecast values for the months of 2018 of the index series are calculated"[3].

Tomić, Daniel, and SašaStjepanović : "In this paper, they compare univariate autoregressive integrated moving average (ARIMA) models of the Croatian industrial production and its sub-sectors to evaluate their forecasting features within short and long-term data evolution. The study aims not to forecast industrial production but to analyze the out-of-sample predictive performance of ARIMA models on aggregated and disaggregated levels inside different forecasting horizons. Our results suggest that ARIMA models do perform very well over the whole range of the prediction horizons. It is mainly because univariate models often improve the predictive ability of their single component over short horizons. In that manner, ARIMA modeling could be used at least as a benchmark for more complex forecasting methods in predicting the movements of industrial production in Croatia" [4].

III. OBJECTIVES OF STUDY

- To fit the model using the traditional method-SARIMA.
- To fit the model using the deep learning method-FFNN, RNN & LSTM using **Python code.**
- To find RMSE, MAE, MAPE values for the traditional and deep learning models to compare for identifying the best fit model.
- To forecast the production of IIP for the next 15 months (i.e., March-2020 to June-2021) using the best model and compare with the actual IIP values.
- Estimate the impact of lockdown on manufacturing production in the industrial sector of the economy.

IV. MATERIAL AND METHODS

A. AutoregressiveIntegrated Moving Average (ARIMA)

"The Autoregressive integrated moving average is denoted as ARIMA(p,d,q) where *d* denotes no. of the differences, *p* denotes autoregressive order, and *q* denotes the moving average order

The ARIMA model is given by:

$$z_t = \delta + \emptyset_1 z_{t-1} + \emptyset_2 z_{t-2} + \ldots + \emptyset_p z_{t-p}$$
$$+ a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

where the level of difference is denoted by z_t , the constant is denoted by δ, while φ is an autoregressive operator, a random shock corresponding to time period t is denoted by 'a', and θ denotes moving average operator" [5].

The seasonal ARIMA method (SARIMA) is used when the seasonal variation is exhibited in the time series data. the seasonal autoregressive order is denoted as P, the seasonal moving average order is denoted as Q, seasonal differencing is denoted as D and the multiplicative process will be formed as SARIMA(p,d,q)(P, D, Q)$_s$, where the subscripted letter 's' shows the length of the seasonal period.

1) Step-1-Identification

"The selection of the order of regular differencing (d), seasonal differencing (D), the non-seasonal order of Autoregressive (p), the seasonal order of Autoregressive (P), the non-seasonal order of Moving Average (q), and the non-seasonal order of Autoregressive (Q). The number of orders can be identified by observing the sample autocorrelations function (ACF) and sample partial autocorrelations function (PACF)" [6].

2) Step-2-Estimation

The time-series data is used to estimate the parameters of the model used in Step 1.

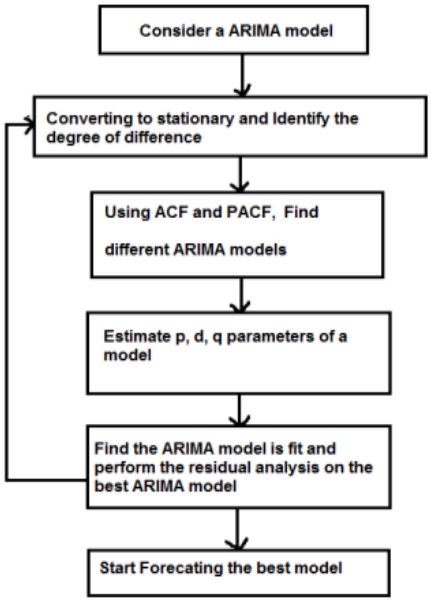

Fig.1. ARIMA Model Flow Chart.

3) *Step-3-Diagnostic Checking*

Residuals of estimated SARIMA model(s) are calculated to see if there is any white noise,pick the best model with well-behaved residuals.

4) *Step-4-Forecasting*

"The above model is used to forecast future values.SARIMA model is widely used because of its capabilities to capture the appropriate trend by examining past data patterns. The Box-Jenkins methodology has the capability of handling stationary and non-stationary time series data in seasonal and non-seasonal elements. It can extract the maximum information from the data using the minimum number of parameters" [6].

B. *Deep Learning Methods*

In recent years Deep Learning Methods are also called Artificial Neural Network (ANN) models have been used widely in many fields. Eg. Biology, Geology, Engineering, Medical, Finance, Physics, etc. From a Statistical perspective, Neural networks have a wide application in classification problems and forecasting the time series data. "An artificial neural network is a massively parallel distributed processor made up of simple processing units which have a natural propensity for storing experiential knowledge and making it available for use" [7]. It resembles the brain in two aspects:

- Knowledge is acquired by the network from its environment through a learning process and the procedure used to perform the learning process are called a learning algorithm
- Interneuron connection strength, known as synaptic weights is used to store the acquired knowledge.

ANN drives its computing power through its massively distributed structure and its ability to learn and generalize. Generalization refers to ANNs ability to produce reasonable outputs for inputs that are not encountered during training. These two features of ANN help the neural network to achieve the following properties:

Non-Linearity: These are the systems where for a given set of inputs the set of outputs is expected. If the relation between input and output is best described by the linear equation then the system is said to be linear but if the output is described not only as of the linear combination of inputs but also on the higher terms of inputs then the system is called non-linear and it is distributed throughout.

Input-Output Mapping: Inputs are provided to the system and in response, the output is received. It is like a learning mechanism where we feed the input as well as the expected output for each input also known as supervised learning. It is possible that initially, the computer model may not be able to provide the expected output for the given input. There may be a difference between excepted and actual output. Therefore the adjustment is done to the "Free Parameters" of the system in such a way that the actual output is close to the expected output for the given set of inputs which creates the mapping between inputs and outputs. This input-output mapping makes the remarkable difference between ANN over other traditional methods.

Adaptivity: ANN has a built-in capacity to adapt free parameters to change according to the surrounding environment where "learning by example" replaces "programming in solving problems.

Evidential Response**:** Apart from the given output it also has the ability to provide a confidence level about its results ie what percentage the neural network is confident about its results?

Fault Tolerance: Its performance only degrades gracefully under adverse operating conditions

These features make the ANN models widely useable throughout all fields. In Statistical application, it can be widely used in stock market prediction, production predictions, economic indicators predictions, etc.

Neural networks architectures: In an Artificial neural network, the basic unit is neurons and there are many neurons that are interconnected to each other forming the network that is why it is called neural networks. There are different types of architectures in neural networks and a few of the most widely used are described below.

Feed-Forward neural networks (FFNN): Feed-forward artificial neural networks are the artificial neural networks with feed-forward topology ie information flows from input to output in only one direction and without back loops. However, there is no limitation on the number of layers, type of activation function used in individual neurons, and the number of connections between individual neurons. Based on the no of layers in the feed-forward neural network we can further classify it into a single layer or multilayer feed-forward network.

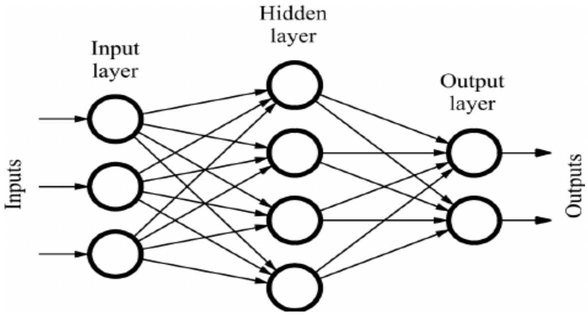

Fig. 2. Feed-Forward Neural Network

Recurrent networks (RNN): It is similar to the feed-forward neural network with no limitation regarding the back loop ie information can be transmitted in both forward and backward directions. The most basic topology of RNN is a fully recurrent artificial neural network where every neuron is connected to every other neuron in all directions. It also allows a self-feedback network.

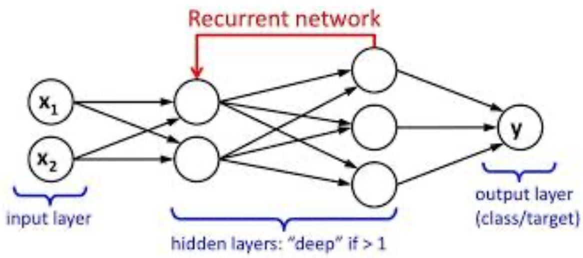

Fig. 3. Recurrent Neural Network

Long short-term memory (LSTM): LSTM is a special kind of RNN that solves the problem of short-term memory. It enables us to learn the long-term patterns in the data. It is designed to overcome the limitation of RNN such as gradient vanishing and exploding, complex training, difficulty in processing very long sequences. LSTM can process not only single data points (such as images) but also entire sequences of data(such as speech or video etc).LSTM has a wide application in music composition, rhythm learning, speech recognition, time series prediction, etc.

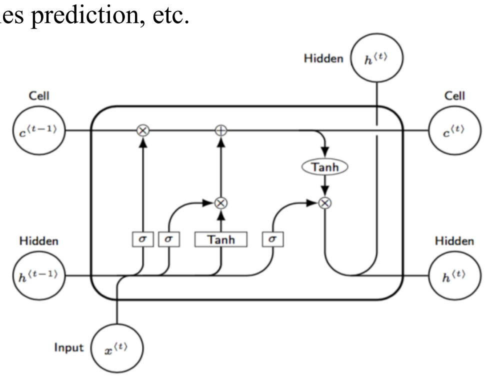

Fig. 4. Long Short-Term Memory

Data Source:The study considers the Index of Industrial Production data, published by the Ministry of Statistics and Programme Implementation (MOSPI).This study examines the above-mentioned objectives at all Indian levels. The period of study is from April-2012 to Feb-2020 (A total of 95 observations). IIP values over months are used for the study. In this section, the results of forecasting are presented using 3 methods. The results are analyzed and compared. The methods are compared using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) which are given below:

$$MAE=1/n \sum |Y_t-F_t|$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_t - F_t}{Y_t}\right| \times 100$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_t - F_t\right)^2}$$

where Y_t is the actual value and F_t is the forecasted value and n is the number of months used as forecasting period.

V. Results and Discussion

A. SARIMA Model

The development of the SARIMA model for a single variable involves identification, estimation, verification, and forecasting. Each of these steps is now explained.

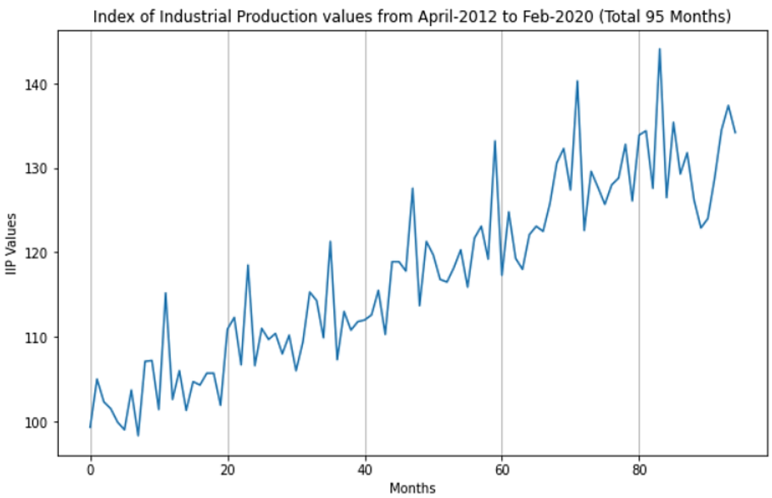

Fig. 5. Index of Industrial Production values from April-2012 to Feb-2020.

```
Result Of Dickey-Fuller Test
pvalue =  0.9051959985230664  if above 0.05, data is not stationary
```

First examined whether the data is stationary or not using Dickey-fuller Test and the results show that the data is not stationary. Then, the data is converted into stationary by using the first or second-order difference of the IIP values. By looking at the trend, it is also observed that seasonal variation is also present in the data. Therefore, the seasonal ARIMA (also called SARIMA) model is used to fit the model. The parameters (p,d,q) & (P,D,Q) of SARIMA are identified using the **auto ARIMA function** in Python & SARIMA$(0,1,1)(1,1,1)_{12}$ is estimated as the best model.

Once the model is identified 75% of the dataset is considered for training and 25% for testing and the model is fitted for both the training & testing data sets.

TABLE I: SARIMA(01,1,1)(1,1,1)12

```
Best model:  ARIMA(0,1,1)(1,1,1)[12]
Total fit time: 20.408 seconds
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                   95
Model:             SARIMAX(0, 1, 1)x(1, 1, 1, 12)   Log Likelihood                -190.067
Date:                            Thu, 13 May 2021   AIC                            388.133
Time:                                    13:34:35   BIC                            397.760
Sample:                                         0   HQIC                           391.998
                                             - 95
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6487      0.074     -8.813      0.000      -0.793      -0.504
ar.S.L12      -0.2425      0.188     -1.288      0.198      -0.611       0.126
ma.S.L12      -0.7260      0.235     -3.092      0.002      -1.186      -0.266
sigma2         5.0858      0.932      5.457      0.000       3.259       6.913
===================================================================================
Ljung-Box (L1) (Q):                   0.20   Jarque-Bera (JB):                 9.79
Prob(Q):                              0.65   Prob(JB):                         0.01
Heteroskedasticity (H):               2.40   Skew:                            -0.62
Prob(H) (two-sided):                  0.03   Kurtosis:                         4.16
===================================================================================
```

Note: Here 'y' as the dependent variable is IIP values.

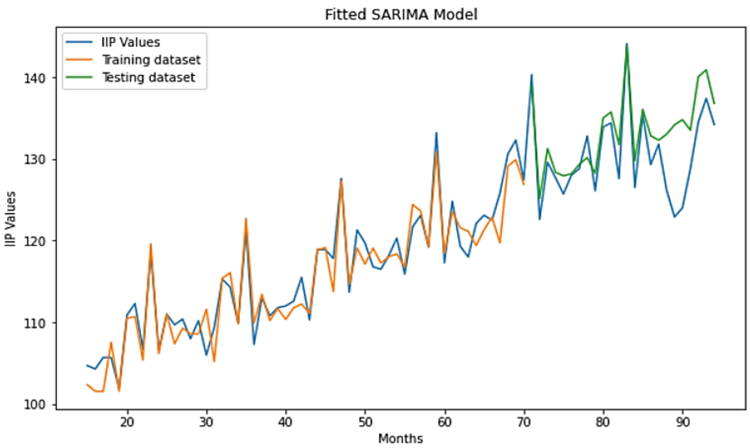

Fig. 6. Fitted SARIMA (01,1,1)(1,1,1)12 model.

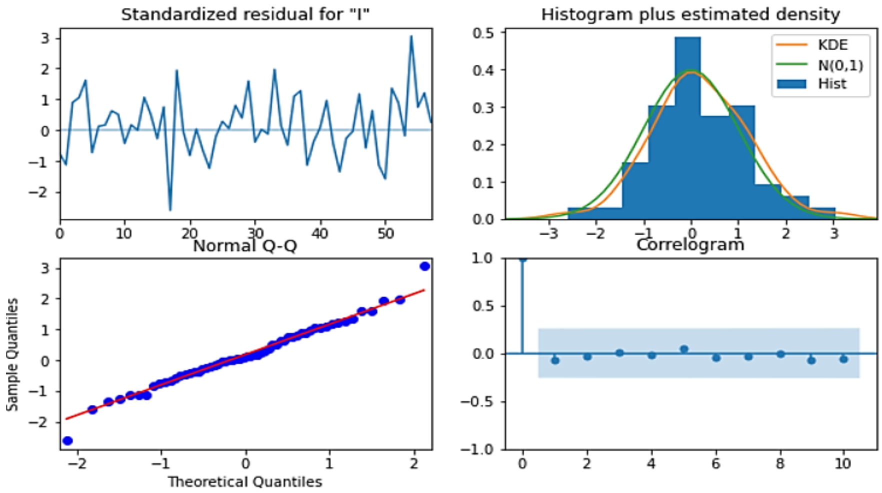

Fig. 7. Plot – Diagnostics.

B. Deep Learning Models

Models are fitted using different specifications and the results show that RMSE, MAE, MAPE is least for 75% of observations. Therefore, 75% of observations are considered as the training data set and 25% of observations as the testing data set for the different models, and the graphs are also plotted for the same data set.

- Activation function: Rectified Linear Unit;

- Validation Generator: TimeseriesGenerator;
- Batch & Sequence Size: 1 & 12.

TABLE II: TEST ACCUARACY

Model Name	Specification	Training Data Set	RMSE	MAE	MAPE
SARIMA	-	75%	13.75	4.14	10.88
FeedForward	Dense Layer-64 Dense Layer-34 Dense Layer-1	75%	1.83	1.49	1.2782
RNN	SimpleRNN-64 Dense Layer-1	75%	0.68	0.55	1.58
LSTM	LSTM(50) LSTM(50) Dense Layer-34 Dense Layer-1	75%	2.13	1.65	7.697
LSTM	LSTM(50) LSTM(50) Dense Layer-1	75%	2.19	1.77	7.477

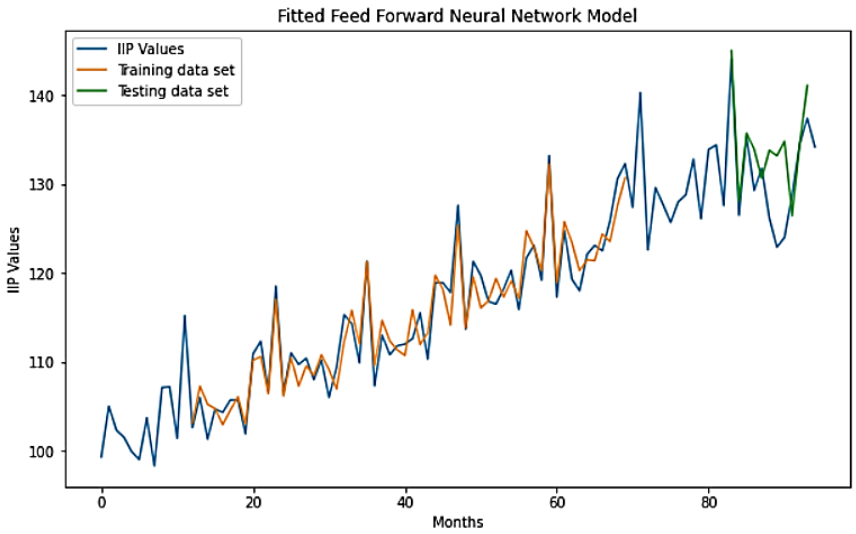

Fig. 8. Fitted Feed Forward Neural Network Model.

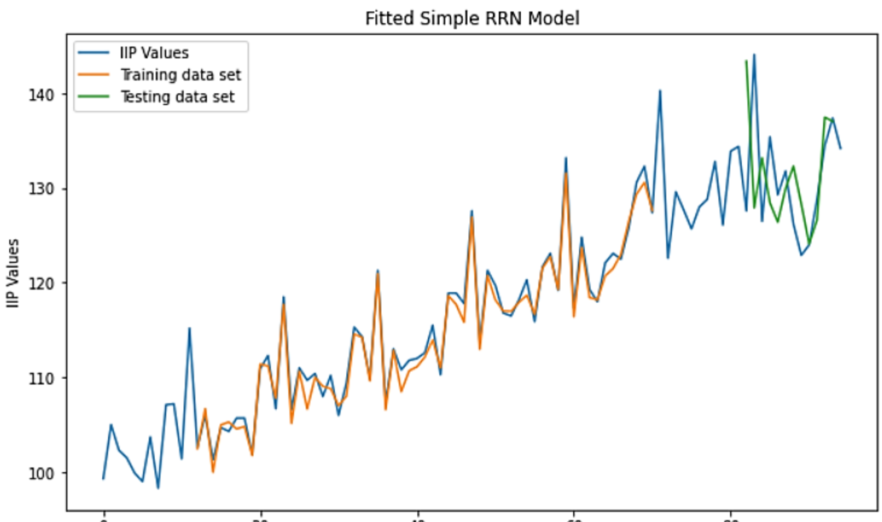

Fig. 9. Fitted RRN Model.

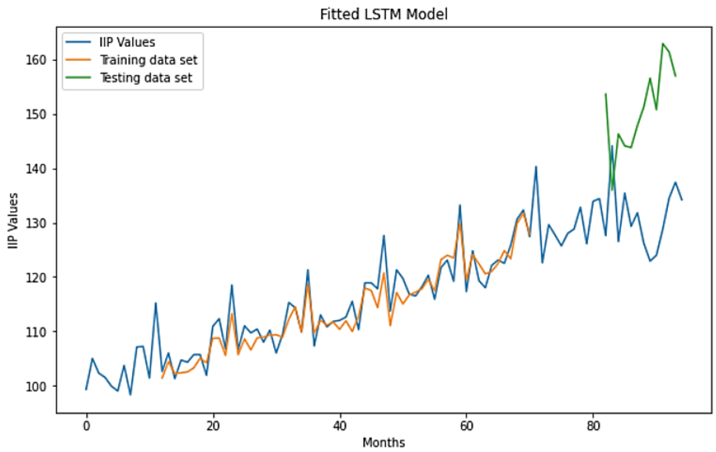

Fig. 10. Fitted LSTM Model.

VI. Conclusion

The test of accuracy (Table II) and the above-plotted graph (Fig.9)shows that the RNN model is performing better than other models like SARIMA, FFNN, LSTM. The forecasted IIP of values from Mar-2020 to Jun-2021 are shown below:

TABLE III:Forecasted IIP Values from Mar-2020 to Jun-2021

Months	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20
IIP Value (Actual)	117.20	54.00	90.20	107.90	117.90	117.20	124.10	129.60
IIP Values(Forecasted)	140.94	132.03	141.17	131.77	131.71	129.92	126.19	127.39

Months	Nov-20	Dec-20	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21
IIP Value (Actual)	126.70	136.60	136.20	129.40	145.60	126.70	116.00	122.60
IIP Values(Forecasted)	126.70	128.51	131.38	132.06	136.24	136.05	133.77	134.57

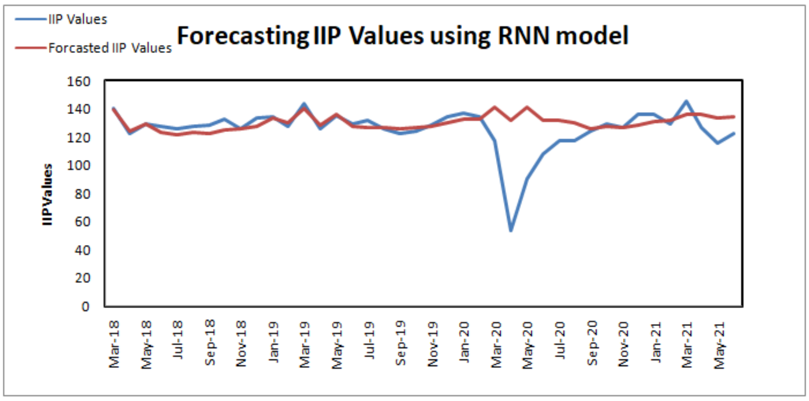

Fig. 11. Forecasting IIP Values using RNN Model.

IIP forecasted values are compared with actual IIP values from Mar-2020 to Jun-2021. The result shows that on average there is around a **36.88%** decrease in IIP Values during the lockdown period i.e.,**Mar-2020 to May-2020** which indicates lockdown has negatively affected the Industrial sector production.

From Fig.11, it is observed there is a declining trend during the lockdown and it took around 6 to 10 months for industrial sector production to reach around to the normal estimated trend.

References

[1] Singh SS, Devi TL, Roy TD. Time series analysis of index of industrial production of India. *IOSR Journal of Mathematics.* 2016; 12(3): 1-7.

[2] Department, Minister's Secretariat, Ministry of Economy, Trade, and Industry. *Economic Analysis Office Research and Statistics.* 2015.

[3] Çekim HÖ. Examination of industry production index in Turkey with time series method. *BalıkesirÜniversitesi Fen BilimleriEnstitüsüDergisi.* 2018; 20(1): 547-54.

[4] Tomić D, Stjepanović S. Forecasting Capacity of ARIMA Models; A Study on Croatian Industrial Production and its Sub-sectors. *Zagreb International Review of Economics & Business.* 2017; 20(1): 81-99.

[5] Permanasari AE, Hidayah I, Bustoni IA. SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence.*International Conference on Information Technology and Electrical Engineering (ICITEE)*. 2013; 203-207.

[6] Shakti SP, Hassan MK, Zhenning Y, Caytiles RD, Iyenger NC. Annual automobile sales prediction using ARIMA model. *International Journal of Hybrid Information Technology*. 2017; 10(6): 13-22.

[7] Mas JF, Flores JJ. The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing.* 2008; 29(3): 617-63.
[8] Wikipedia.org. *Index of industrial production* [Internet]. 2022. Available from: https://en.wikipedia.org/wiki/Index_of_industrial_production.