Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, October : 2023

Modeling and performance of maize production in Telangana using Time series models and ANN

P. Rajeshwar[1] and Dr. K. Sampath kumar[2]

1. Research Scholar, Department of Applied Statistics, Telangana University, Nizamabad, Telangana. Email: palepu108@gmail.com, , Cell: 919505000065.

2. Professor, Department of Applied Statistics, Telangana University, Nizamabad, Telangana. Email: ksampath1@gmail.com, Cell: 919701880796.

Abstract

This paper presents the study of the current situation in Maize production in Telangana using linear and non-linear time series models. The predicted or estimated models such as auto regressive integrated moving average and feed forward neural network models are utilized for forecast the next few years on yearly maize production in Telangana India. The model accuracy are found to be determined using error measures like MAPE (mean absolute percentage error), RMSE (root mean square error) and MAE (mean absolute error) on development and validation data sets. This study, the FFNN (1-2-1) models gives the better model performance than the ARIMA and MAPE provides as 2% in Feed forward neural network.

Keywords: ARIMA, FFNN, RMSE, MAE and MAPE.

1. Introduction:

Telangana is one of the largest maize production state in India. The Madhya Pradesh and Karnataka has used the maize approximately 15% each and 10% Maharashtra and others. After Madhya Pradesh and Karnataka the maize growth states are contributes the more than the 80% of the total maize production area Andra Pradesh has 20.9% highest state productivity. In Telangana, present day scenario the maize cultivating around in 14 lakh acres land generating around16 lakh tones maize. The major growing districts in Telangana are Adilabad, Karimnagar, Warangal, Nizamabad, Mahaboobnagar, Khammam and Medak. In opportunity markets in Telangana exceptional vegetation like Jowar, Chilli and Cotton and beneficial price, additionally to the minimum necessities of irrigation.

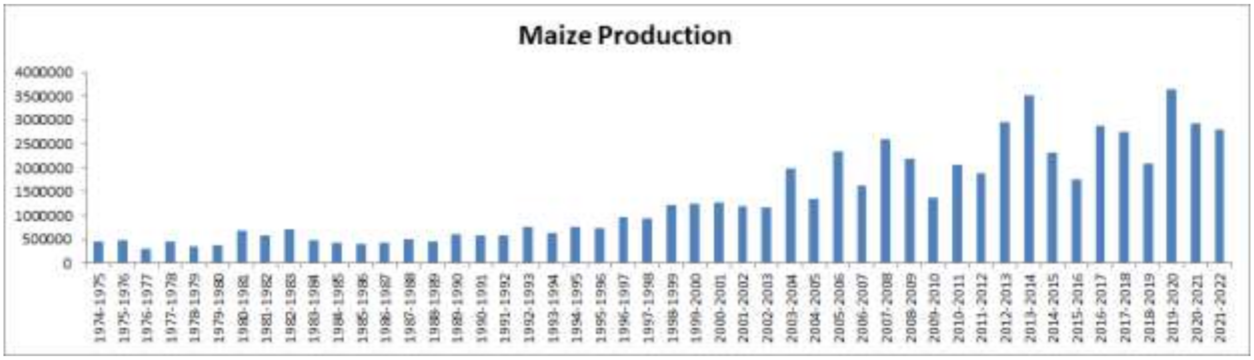

Figure 1.1: Time Series plot for Maize Production In Telangana.

From the above figure, it was observed that the maximum production is showing in 2019-2022 as 36.4 million tones and minimum production is in the year 1976-1977 as 28.9 tonnes due to unpredictable situation was happened in the atmosphere and other several reasons. The average maize production is 13.4 million tonnes. The production is slowly increased at the year 2016-17 to 2018-2019.

2. Data and Methods:

A historical data on Yearly maize production in Telangana is received form the Directorate of Economics and Statistics, Hyderabad, Telangana State. The data comprises the yearly Maize production in tones form the year of 1974-75 to 2021-22. The entire data is divided in to two sets for model development and model validation. The model development is used as train data set and model validation used for test data. The Box – Jenkins and feed forward neural network models are utilized for forecast the future maize production in Telangana. The R software is used for model analysis and for chart and tables, MS-Excel.

2.1 Box-Jenkins methodology:

The Box – Jenkins methodology is used for figuring out the quality model by means of building the auto regressive integrated moving average model (ARIMA) on historical data set. The Box – Jenkins technique consists of several advantages for acquiring the minimal number of parameters in seasonal and non-seasonal. This methodology consists of 4 steps for developing the model including Identification, Estimation, Diagnostic checking and Forecasting. The first step is to check the model identity for model parameters together with p and q by way of the use of the auto correlation and partial auto correlation function plots for the stationary data sets. The auto correlation is dies out for numerous lags and q spikes within the plot then q parameters will appear and partial auto correlation feature dies out for several lags and p spikes inside the plot then this is p parameter. Diagnostic checking is to check the model adequacy by the use of the LJung-Box Q Statistics and check the assumptions with recognize to errors are random. The Ljung Box Q Statistics test is used to test the randomness of the error and also check the parameters importance, if they may be now not giant then take a look at the possible parameters and keep the procedure till to get the parameter significance. In this take a look at, multiple models are examined for the given data sets and identified the best model for predicting the yearly maize production in Telangana, India in line with their error measures together with MAPE, MAE and RMSE.

Therefore, $(1- \Phi_1 B- \Phi_2 B^2 \text{-----} \Phi_p B^p)\ \nabla = (1-\theta_1 B\text{------}\ \theta_q B^q) a_t$

$$\text{Where } \nabla = (1 - B)$$

And d is non-seasonal and D is seasonal components. a_t is a white noise with zero mean and constant variance.

2.2 Artificial Neural Network Model:

The Artificial Neural Network (ANN) models are used on biological neural networks and these networks are linked with numerous organizations of nodes. The following figure is gives an architecture on the structure of the Feed Forward Neural Networks.

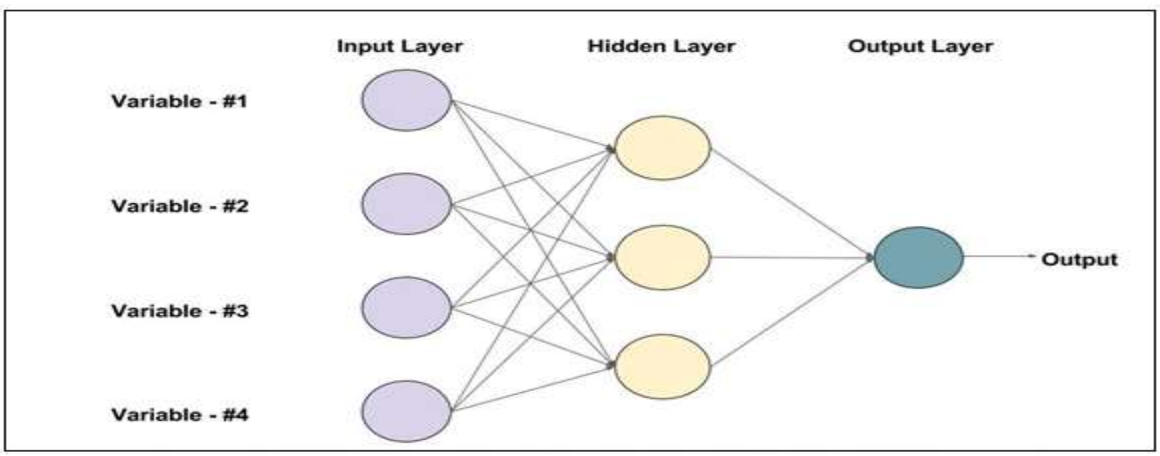

Figure 2.2.1: Feed Forward Neural Network architecture.

An Artificial Neural Network (ANN) is a data-processing worldview inspired by how organic sensory systems, such as the mind, measure data. The unique structure of the data handling system is important to this viewpoint. It is built up of a large number of extremely interconnected preparatory components (neurons) that work together to solve unambiguous difficulties. ANNs, like humans, learn by imitation. Through a learning

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, October : 2023
cycle, an ANN organises a specific application, such as plan acknowledgment, determining, or information order. In natural frameworks, learning remembers modifications for agreement with synaptic attachments that exist between neurons. With their remarkable ability to extract meaning from muddled or ambiguous data, neural networks can be used to separate examples and uncover patterns that are too confusing to be detected by either people or other computer tactics. A trained neural system can be regarded of as a specialist in the classification of data that has been subjected to tests. This master may then be used to make projections given new situations of intrigue and response create a scenario where queries.

The Feed Forward Neural Network model is made up of three layers: input, hidden, and output. The data is transported into the input layer first, then weights based on the pattern of hidden layers are presented, and finally the output layer. The output layer must be one, and there is no specific process for determining the number of layers in the hidden layer. The only process is to evaluate the trail and error approach based on the model's performance. The network information is shown in the table below.

Table 2.2.1: Network Information

Input Layer	Covariates 1	Lag1
	Number of Units[a]	1Normalized
	Rescaling method of covariates	
Hidden Layer	Number of hidden layers	2
	Number of units in the hidden layer 1[a]	2
	Activation function	Hyperbolic Tangent
Output Layer	Dependent variable 1	Stock prices
	Number of units	1
	Rescaling method of scale dependent	Normalized
	Activation function	Identity

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, October : 2023

	Error function	Sum of squares

3. Analysis of data:

The data series of yearly maize production in Telangana is from of 1974-75 to 2021-22 is presents in the following figure. The entire data is divided in to two sets for model development and model validation. The model development is used as train data set (consists of 35 observations) and model validation used for test data (consists of 12 observations) for the selected model.

3.1 ARIMA Model:

In this model, first we find the stationary of the data by using the auto correlation and partial auto correlation functions. The below figures gives an idea of the way the data pattern of yearly maize production is behaving from 1974-75 to 2021-2022.

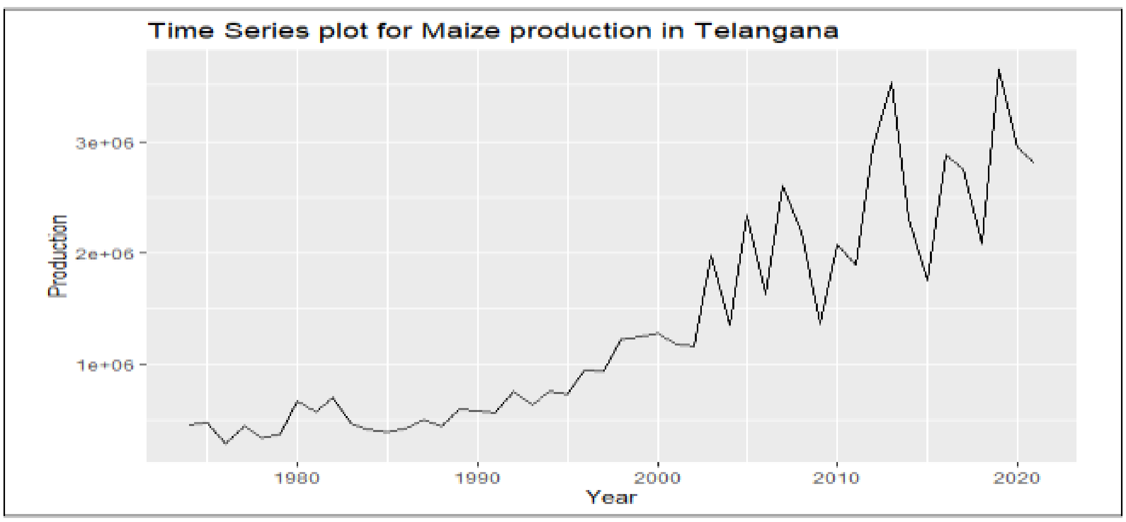

Figure 3.1.1: Time series plot of yearly maize production in Telangana.

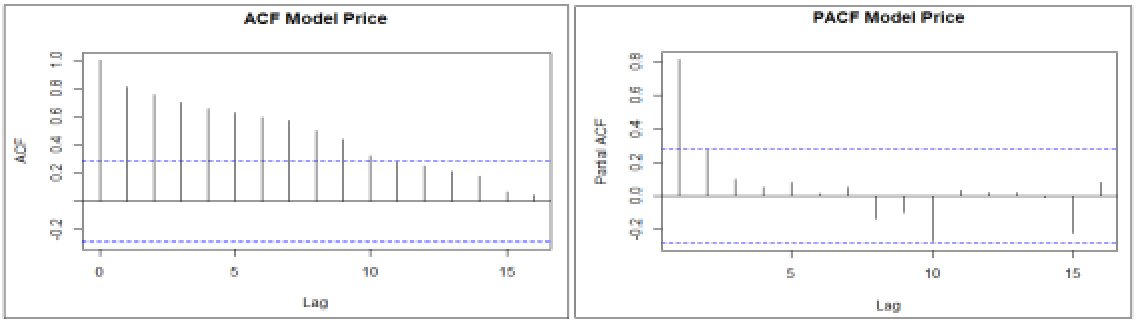

Figure 3.1.2: ACF and PACF plots of yearly maize production in Telangana.

From the above figure 3.1.1 and 3.1.2, shows the data pattern of yearly maize production in Telangana from 1974-1975 to 2021-2022. It was observed that the data consists of multiple fluctuations from one period to another period and it is not constant. Hence the data is non-stationary and apply the transformation and see the pattern of the data including ACF and PACF plots.

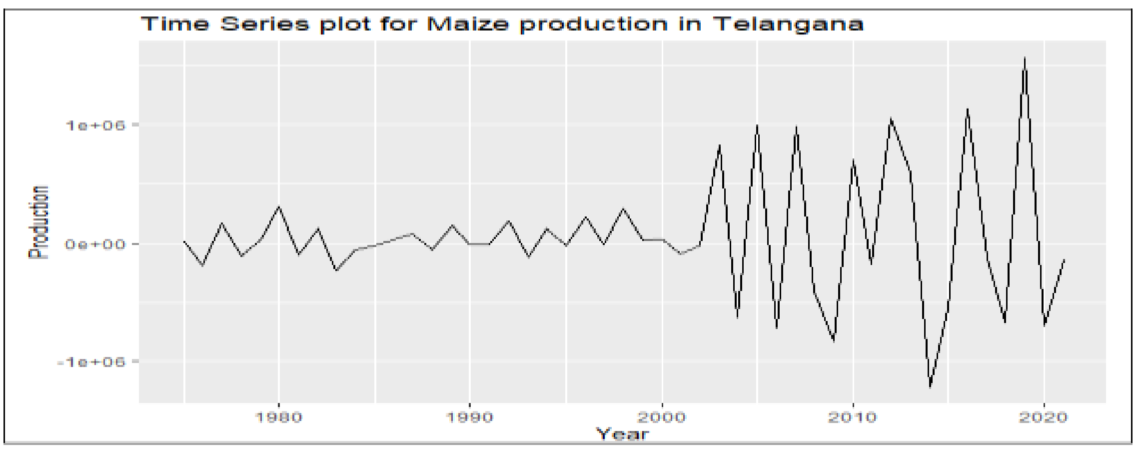

Figure 3.1.3: Transformed time series plots of yearly maize production

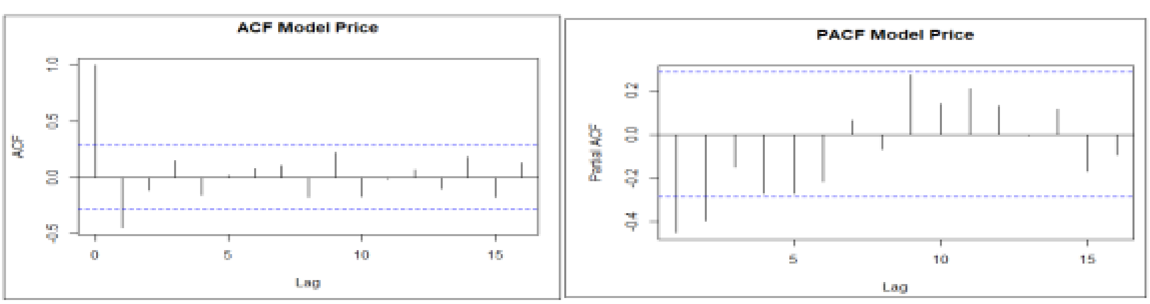

Figure 3.1.4: Transformed ACF and PACF plots of yearly maize production

From the above figures 3.1.3 and 3.1.4, the ACF and PACF dies out with first order difference. Based on the Augmented Dicky – Fuller test the p-value is 0.01, as it is smaller than the significant level (5%). Hence, null hypothesis is rejected and the data is stationary. The below table ADF test results shows the first order difference the data is stationary.

Table 3.1.1: Augmented Dicky-Fuller Test

ADF test	
P-Value	0.01
Lag Order	3
ADF	-5.7589

The parameters of the models are identified based on auto correlation and partial auto correlation function. In the analysis tested several possible models for identifying the best model according with their parameter significance, adequate of model. The adequate of model is used to test the LJung –Box Q Statistics. The following table presents the some tentative models.

Table 3.1.2: Tentative model of ARIMA

ARIMA (p,d,q)			AIC	BIC	Significance of the parameters	L-Jung box	p-value	Adequacy
1	1	0	31.09	34.21	Significant	8.93	0.18	Adequate
0	1	1	33.80	36.91	Significant	12.12	0.06	Adequate
1	1	1	33.07	37.73	Insignificant	8.82	0.12	Adequate
1	1	2	33.64	39.87	Insignificant	6.22	0.18	Adequate
0	1	2	32.85	37.52	Significant	8.00	0.16	Adequate
2	1	0	33.04	37.71	Insignificant	8.73	0.12	Adequate
1	0	2	37.22	45.14	Insignificant	5.63	0.23	Adequate
2	1	1	30.18	36.40	Significant	8.82	0.07	Adequate
2	1	2	29.27	37.05	Significant	4.67	0.20	Adequate

From the above table 3.1.2, it is observed that the ARIMA (2, 1, 1) model is selected for the forecast the future yearly maize production in Telangana. The ARIMA (2, 1, 1) model is chooses based on the parameter significance and adequacy from the above table. The estimated parameters are showed in the following table.

Table 3.1.3: Parameters of the ARIMA (2, 1, 1) Model

Parameters	Estimate	Std. error	z-value	Pr(>|z|)
ar1	0.358	1.093	0.0003	<0.001
ar2	0.637	1.101	0.039	<0.001
ma1	2.000	1.061	3320.5	<0.001

The above table 3.1.3, model ARIMA (2, 1, 1) parameter shows the significant and this is the best model for the forecast the yearly maize production in Telangana. Now the ARIMA (2, 1, 1) model equation is

$$(1- \Phi_1 B- \Phi_2 B^2)(1-B)=(1-\theta_1 B)a_t$$

Now, the ARIMA (2, 1, 1) model is

$$(1- 0.358B- 0.637B^2)\ \nabla = (1-2B)a_t$$

The adequacy of the model is tested based on the Ljung – Box Q Statistics. This test is uses the residuals of the series after building the model. Then the hypothesis of the model is

H_o: Model is adequate and H_1: Model is Inadequate

Table 3.1.4: Ljung-Box Q Statistics

Model	ARIMA(2,1,1)
Statistic	8.82
df	4
p-Value	0.07

From the above tab le 3.1.4, the hypothesis value is more than the p-value as 0.05, and then we accept the null hypothesis and concluded that the selected models are adequate. Hence, the ARIMA (2, 1, 1) model is used for forecasting the future yearly production of maize in Telangana, India. The forecasted yearly maize productions are listed in following table.

Table 3.1.5: Forecasted of yearly maize production using ARIMA(2, 1, 1)

Date	Actual Maize Production	Forecasted Maize Production
2010-2011	2068560	2209759
2011-2012	1892475	1473160
2012-2013	2943717	1966190
2013-2014	3524907	1667805
2014-2015	2308051	1764697
2015-2016	1751074	1806124
2016-2017	2882475	1682184
2017-2018	2752147	1841201
2018-2019	2082991	1686551
2019-2020	3643686	1811359
2020-2021	2942175	1725001
2021-2022	2802498	1770518

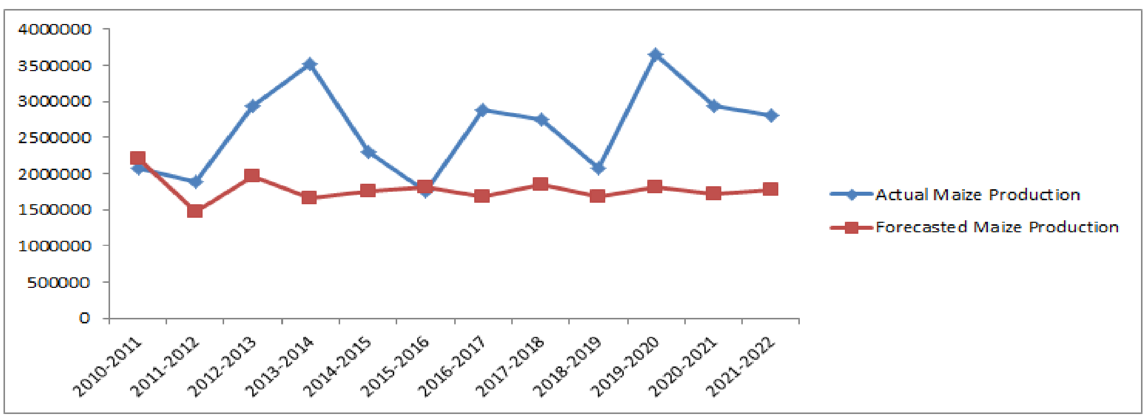

Figure 3.1.5: Forecasted of yearly maize production using ARIMA (2, 1, 1)

The model development is done on training data set and using the test data set validates the model performance. The model performance in test sample is shown in the following table 3.1.6 The model performance is used to consider the test data and compare with the forecasted values by comparing the error measures presented in the below table

Table 3.1.6: Performance of the ARIMA (2, 1, 1) Model

Data	RMSE	MAE	MAPE
Training set	1.23	1.17	2.34
Test set	1.64	1.53	5.99

From the above table 3.6, it was observed that the error measure for ARIMA (2, 1, 1) gives best values as 1.23 and 1.64 million tones for RMSE and 2.34 and 5.99 for MAPE respectively in training and test data sets. The error measures and pattern of the data are very close to test and forecasted values and it was suggested the above level.

3.2 Feed Forward Neural Network Model

The feed forward neural network consists of input neurons which as lag1 in our study. The output layer coming as one and it offers the forecast of the yearly maize production in Telangana in India. Now to find the hidden layer don’t have any specific process for consider the number hidden layers in the model without preparing the forward or backward selection method to determine the hidden layers. In our model, applied hyperbolic tangent function is used for activation function under the back propagation methodology and verified several model and identified the best one based on the minimum error measures such as MAPE, RMSE, MAE. The tentative models are presented in the following table.

Table 3.2.1: Possible Model of FFNN

Number of Layer			Train			Test		
Input	**hidden**	**Output**	**RMSE**	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**
1	2	1	1.31	1.23	2.92	1.41	1.32	3.62
1	3	1	1.27	1.22	2.77	1.50	1.42	5.06
1	4	1	1.31	1.23	2.91	1.44	1.34	3.92
1	5	1	1.28	1.22	2.76	1.45	1.35	4.11

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, October : 2023

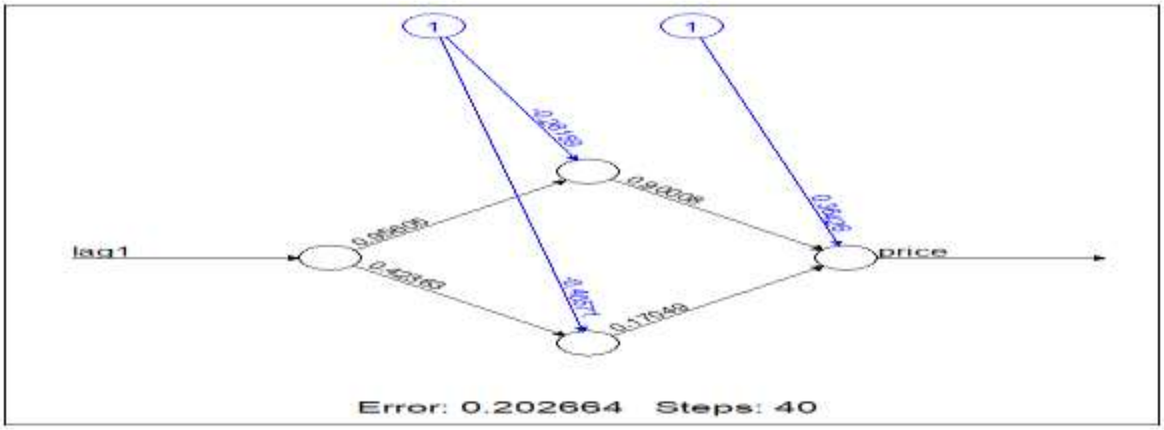

Figure 3.2.1: Feed Forward Neural Network Flow

The above FFNN flow and model development done using the R-software and the parameters of model are listed in the following table 3.2.2

Table 3.2.2: Parameters of the Model (1-2-1) model

Parameter estimates	
error	1.151
reached. threshold	1.006
steps	40
Intercept.to.1layhid1	0.823
lag1.to.1layhid1	1.940
Intercept.to.1layhid2	0.755
lag1.to.1layhid2	1.341
Intercept.to.price	1.287
1layhid1.to.price	1.866
1layhid2.to.price	1.125

The hidden neurons are

H_1 = Tanh[0.823-1.940$\bar{Z}_{t-1}$]

H_2 = Tanh[0.755-1.341$\bar{Z}_{t-1}$]

Where $\bar{Z}_{t-1}$ is rescalled input variables. Then the FFNN 1-2-1 model equation is

O_t = 1.866 H_1 + 1.125 H_2+1.287

The model performance is used to consider the test data and compare with the forecasted values by comparing the error measures presented in the below table

Table 3.2.3: Performance of the model

Data Set	RMSE	MAE	MAPE
Training Set	1.31	1.23	2.92
Test Set	1.41	1.32	3.62

From the above table 3.2.3, it was observed that the error measure for FFNN(1-2-1) gives best values as 1.31 and 1.41 million tones for RMSE and 2.92 and 3.62 for MAPE respectively in training and test data sets. The error measures and pattern of the data are close and near for MAPE values to test and forecasted and it was suggested the above level.

Table: 3.2.4: Forecasted yearly maize production using FFNN (1-2-1) model

Date	Actual PNB Stock Prices	Forecasted PNB Stock prices		Date	Actual PNB Stock Prices	Forecasted PNB Stock prices
01-12-2022	53.25	51.24		16-12-2022	56.75	59.33
02-12-2022	54.00	53.21		19-12-2022	56.05	56.75
05-12-2022	55.15	53.97		20-12-2022	55.55	56.04
06-12-2022	55.45	55.13		21-12-2022	53.15	55.54
07-12-2022	55.45	55.44		22-12-2022	53.75	53.11
08-12-2022	57.70	55.44		23-12-2022	49.70	53.72
09-12-2022	56.25	57.71		26-12-2022	53.50	49.62
12-12-2022	58.10	56.25		27-12-2022	54.20	53.46
13-12-2022	59.90	58.12		28-12-2022	54.90	54.17
14-12-2022	59.25	59.94		29-12-2022	55.40	54.88
15-12-2022	59.30	59.28		30-12-2022	56.45	55.39

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, October : 2023

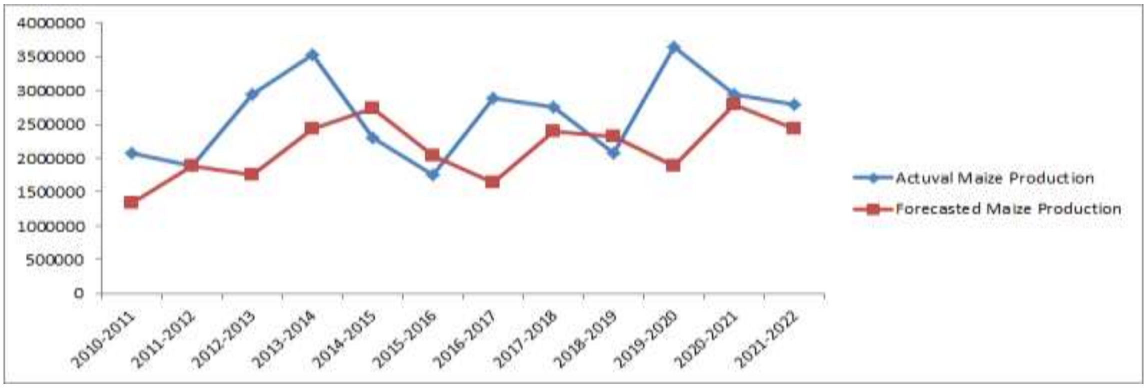

Figure: 3.2.2: Forecasted yearly maize production using FFNN (1-2-1) model

The above figure 3.2.2 shows that the FFNN model provides the best results and graph also gives the better pattern.

4. Comparison of forecasted models for yearly maize production in Telangana.

The comparison between the ARIMA and FFNN models on development and validation data sets, the results provides the significant difference in the error measures. The feed forward neural network model gives the better error measures and flexible for development and validation data sets as verified with the auto regressive integrated moving average model. The performance of the models is listed in the following table and FFNN MAPE values are very close compared to ARIMA model.

Table 4.1: Performance of the ARIMA (2, 1, 1) and FFNN (1-2-1) model

Model	ARIMA			FFNN		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Training set	1.23	1.17	2.34	1.31	1.23	2.92
Test Set	1.64	1.53	5.99	1.41	1.32	3.62

5. Conclusion:

The forecasts of the ARIMA and FFNN models are shown in the following table 5.1 and figure 5.1.

Table 5.1: Forecasts of ARIMA and FFNN models

Date	Actual Maize Production	ARIMA Forecasted Maize Production	FFNN Forecasted Maize Production
2010-2011	2068560	2209759	1340389
2011-2012	1892475	1473160	1880911
2012-2013	2943717	1966190	1754380
2013-2014	3524907	1667805	2434561
2014-2015	2308051	1764697	2745029
2015-2016	1751074	1806124	2044264
2016-2017	2882475	1682184	1648600
2017-2018	2752147	1841201	2399446
2018-2019	2082991	1686551	2323036
2019-2020	3643686	1811359	1891029
2020-2021	2942175	1725001	2803816
2021-2022	2802498	1770518	2433683

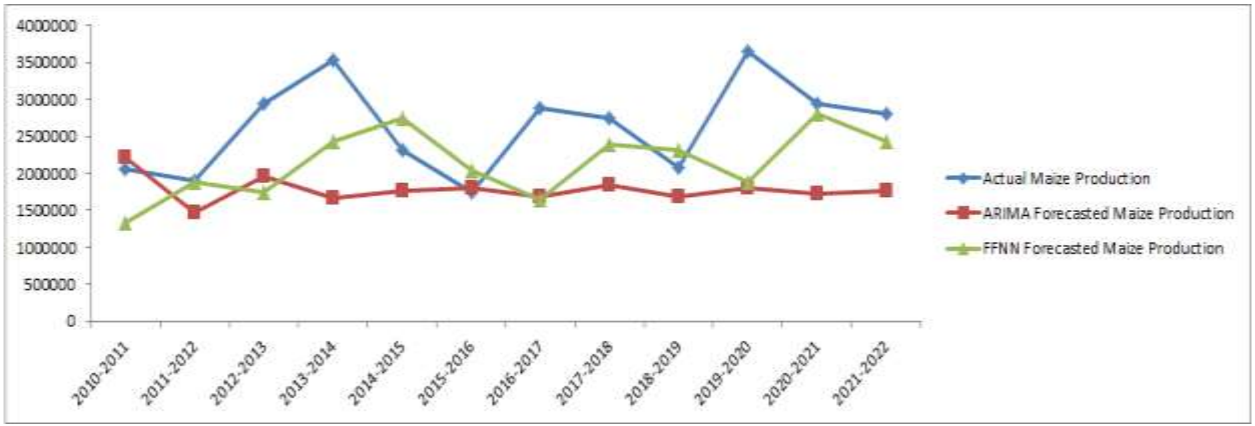

Figure 5.1: Forecasts of ARIMA and FFNN models

From the above study, the FFNN model provides the better outcomes as compared with ARIMA model for fitting and forecasting the yearly maize production in Telangana. Hence, it is concluded that the, FFNN model has been better for forecasting the yearly maize production in Telangana.

References:

1. Dr.M.UdayShankar (2020) et al, FORECASTING OF EGG PRICES IN TELANGANA USING R-SOFTWARE, INDIAN JOURNAL OF APPLIED RESEARCH, Volume -10 | Issue - 3 | March - 2020 | PRINT ISSN No. 2249 - 555X | DOI : 10.36106/ijar.
2. D Srinivas (2022) et al, Forecasting of SBI Stock Prices in India using ARIMA models, JUNI KHYAT JOURNAL, ISSN:2278-4632, Volume -12| | Issue - 08 | August – 2022.
3. Anderson, B. and Ledolder, J.: Statistical Methods for Forecasting, John Wiley & Sons, New York, 1983
4. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31(3) (1986) 309-317.
5. Srinivas Dande, Dr.M.Udayshankar, Dr.M. Raghavender Sharma[2022].: Forecast of SBI stock prices in India Using ARIMA models, JUNI KHYAT JOURNAL ISSN: 2278-4632
6. Banhi Guha and Gautam Bandyopadhyay,Gold Price Forecasting using Arima model, Journal of Advanced Management Science, Vol 4, No.2, March 2016.
7. Raghavender, M. (2009), Forecasting paddy yield in Andhra Pradesh using seasonal Time Series Model, Bulletin of Pure and applied sciences.
8. R.Rama Krishna & Naveen Kumar.B, 2013, Forecasting yield per Hectare of Rice in AP; IJMCAR, ISSN 2249-6955, Volume 3, Issue 1, March 2013, 9-14.
9. Hornik, 1993, Some New result on neural network approximation, Neural Network, 6(1993), pp. 1069-1072.
10. Box, G. E. P. and Pierce, D.: Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models, Journal of the American Statistical Association 65 (1970) 1509-1526.
11. M.A.Khan, 2013, "Forecasting of Gold Prices (Box Jenkins Approach)". 'International Journal of Emerging Technology and Advanced Engineering, Vol.3, PP.662-67.