Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

Accuracy Detection of Maize production in Telangana using time series models and FFNN

P. Rajeshwar Research Scholar, Department of Applied statistics, Telangana University, Nizamabad, Telangana. Email:palepu108@gmail.com,.

Prof. K. Sampath Kumar Professor, Department of Applied statistics, Telangana University, Nizamabad, Telangana. Email: ksampath1@gmail.com ,

Abstract

This work uses linear and non-linear time series models to investigate the present situation of maize production in Telangana. Predicted or estimated models, such as auto regressive integrated moving average and feed forward neural network models, are used to anticipate yearly maize output in Telangana, India, over the following several years. On development and validation data sets, model accuracy was found to be determined by error measures such as MAPE (mean absolute percentage error), RMSE (root mean square error), and MAE (mean absolute error). In this study, the FFNN (1-2-1) models outperform the ARIMA and MAPE models by 2% in the Feed forward neural network.

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

1. Introduction:

Telangana is one of India's major maize-producing states. Madhya Pradesh and Karnataka consumed around 15% maize apiece, with the remaining 10% going to Maharashtra and others. After Madhya Pradesh and Karnataka, the maize growing states provide more than 80% of total maize production area, while Andra Pradesh has the highest state productivity of 20.9%. In Telangana, maize is currently grown on around 14 lakh acres, yielding 16 lakh tonnes per year. Telangana's fastest growing districts include Adilabad, Karimnagar, Warangal, Nizamabad, Mahaboobnagar, Khammam, and Medak. In Telangana's opportunity markets, excellent plants such as Jowar, Chilli, and Cotton are available at a reasonable price, in addition to the bare minimum of irrigation requirements.

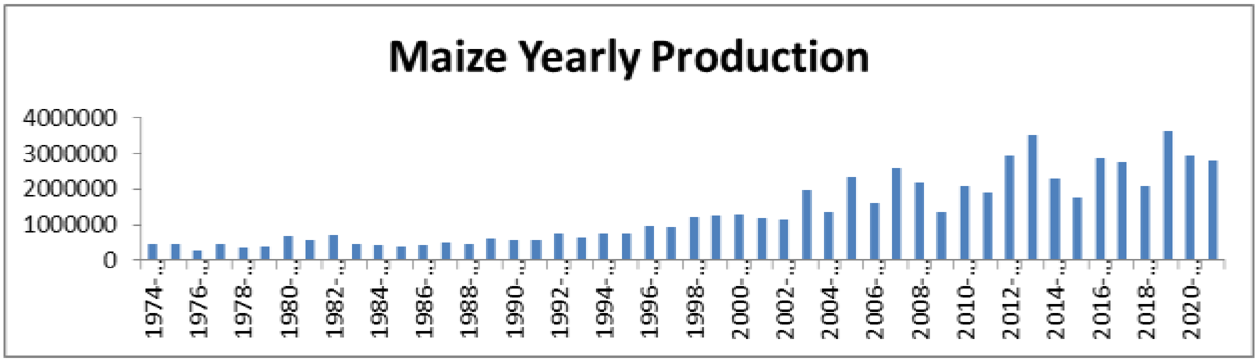

Figure 1.1: Time Series plot for Maize Production in Telangana.

From to the above graph, the highest production is in 2019-2022 at 36.4 million tonnes, while the smallest output is in 1976-1977 at 28.9 tonnes due to an unexpected state in the environment and other factors. Maize output averages 13.4 million tonnes. Production gradually rose from 2016-17 to 2018-2019.

2. Material and Methods:

The Directorate of Economics and Statistics, Hyderabad, Telangana State, has provided historical statistics on annual maize output in Telangana. The data set includes yearly maize output in tonnes from 1974-75 to 2021-22. The data is split into two sets for model development and validation. Model development serves as the trained data set, while model validation serves as the test data set. In Telangana, the Box-Jenkins and feed forward neural network models are used to anticipate future maize output. R software is used for model analysis, as well as charts and tables.

2.1 Box-Jenkins methodology:

The Box - Jenkins approach is used to determine the quality model by creating an auto regressive integrated moving average model (ARIMA) using historical data. The Box - Jenkins approach has various advantages for obtaining the minimum amount of seasonal and non-seasonal characteristics. There is Identification, estimation, diagnostic checking, and forecasting are the four processes in this technique for developing the model. The first stage is to test the model identity for model parameters as well as p and q using auto correlation and partial auto correlation function plots for stationary data sets.

When the auto correlation feature dies for many lags and q spikes inside the plot, q parameters show, and when the partial auto correlation feature dies for several lags and p spikes within the plot, p parameters appear. The purpose of diagnostic checking is to assess the model adequacy using the LJung-Box Q Statistics and to examine the assumptions with the recognition that errors are random. The LJung-Box Q Statistics test is used to evaluate the randomness of the error and also to verify the parameter significance; if they are no longer significant, then look at the alternative parameters and repeat the operation until the parameter significance is obtained. Multiple models are tested for the supplied data sets in this study, and the optimal model for forecasting annual maize output in Telangana, India is determined based on error metrics such as 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:

Artificial Neural Network (ANN) models are utilized on biological neural networks, which are coupled with various node organizations. The architecture of Feed Forward Neural Networks is seen in the diagram below.

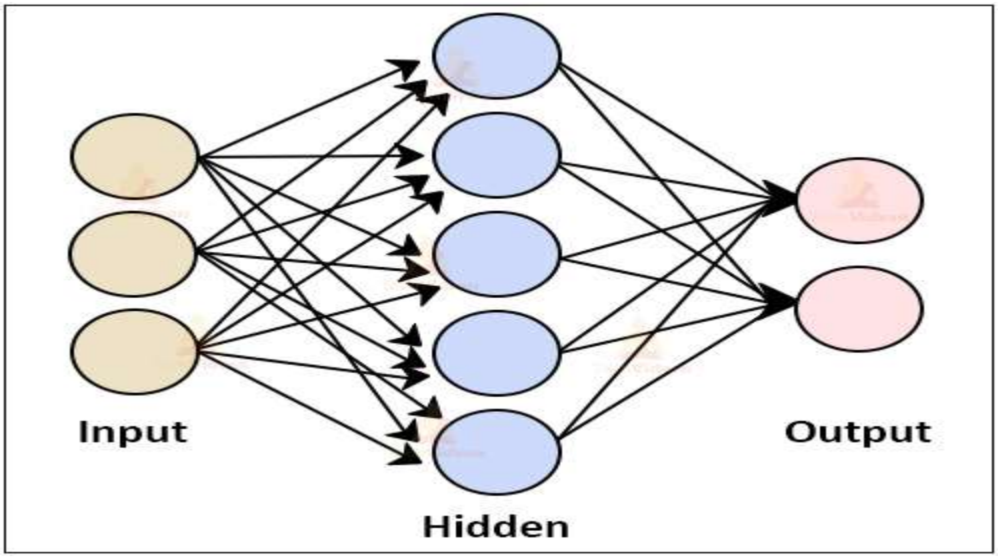

Figure 2.1: Feed Forward Neural Network architecture.

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 precise process for determining the number of layers in the hidden layer. The only step is to evaluate the trial and error approach based on the model's performance. The network information is shown in the below table.

Table 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

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

	Rescaling method of scale dependent	Normalized
	Activation function	Identity
	Error function	Sum of squares

3. Results and Discussion:

The following graph depicts the data series of Telangana's annual maize output from 1974-75 to 2021-22. The data is split into two sets for model development and validation. For the specified model, model development is utilized as the train data set (35 observations) and model validation is used as the test data set (12 observations).

3.1 ARIMA Model:

In this model, first to 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.

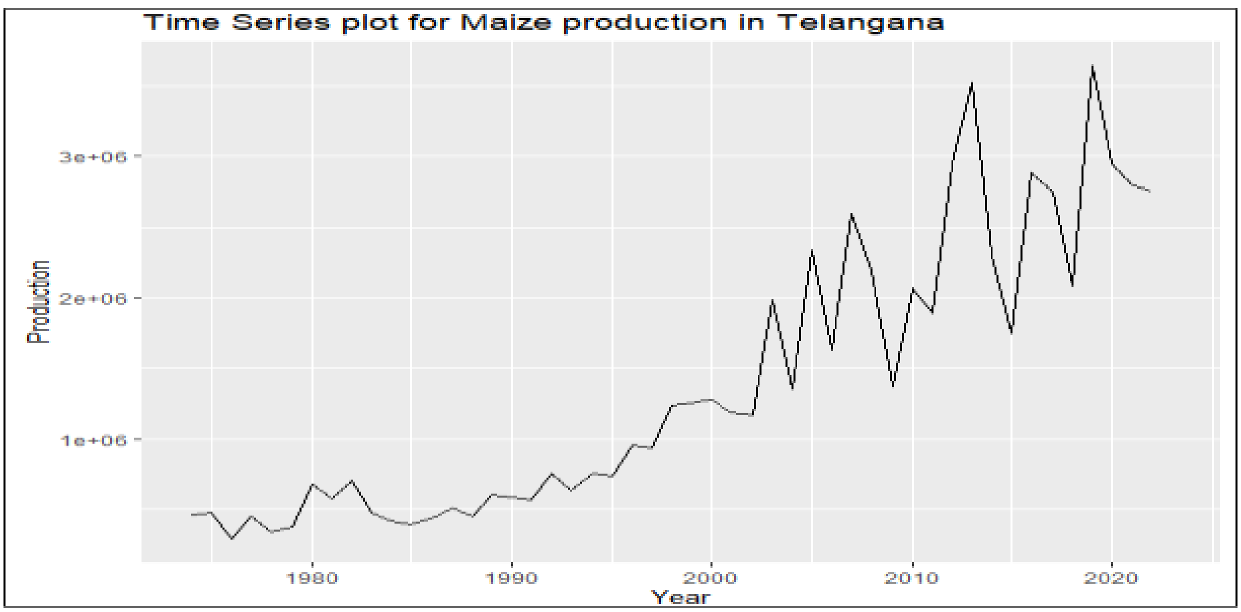

Figure 3.1.1: Time series plot of yearly maize production in Telangana.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

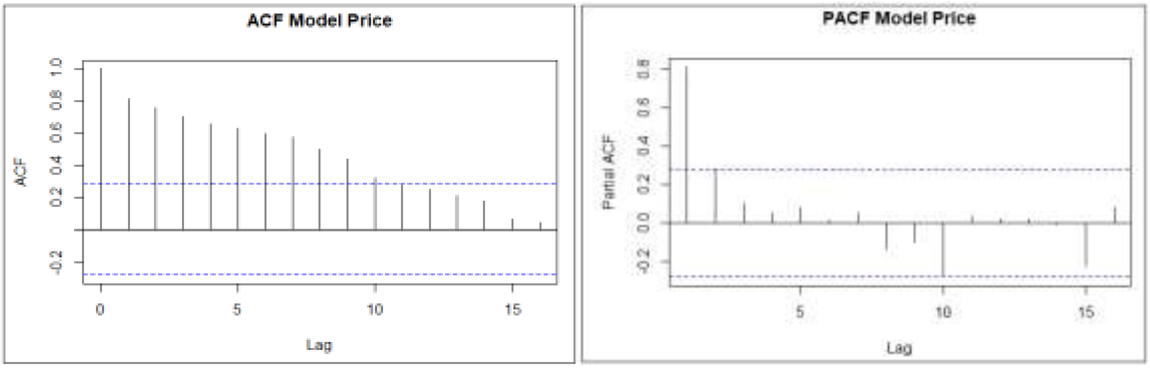

Figure 3.1.2: ACF and PACF plots of yearly maize production in Telangana.

Figures 3.1.1 and 3.1.2 exhibit the data trend of yearly maize production in Telangana from 1974-1975 to 2021-2022. It was discovered that the data consists of many oscillations from one period to the next and is not continuous. As a result, the data is non-stationary, therefore perform the transformation and examine the data pattern, including ACF and PACF plots.

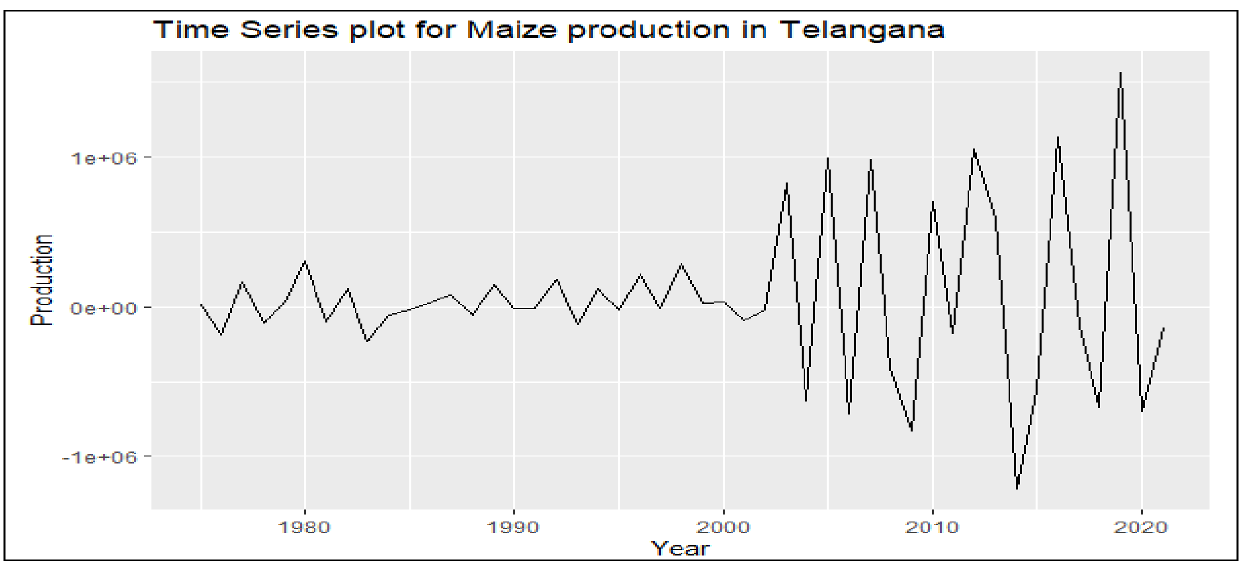

Figure 3.1.3: Transformed time series plots of yearly maize production in Telangana.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

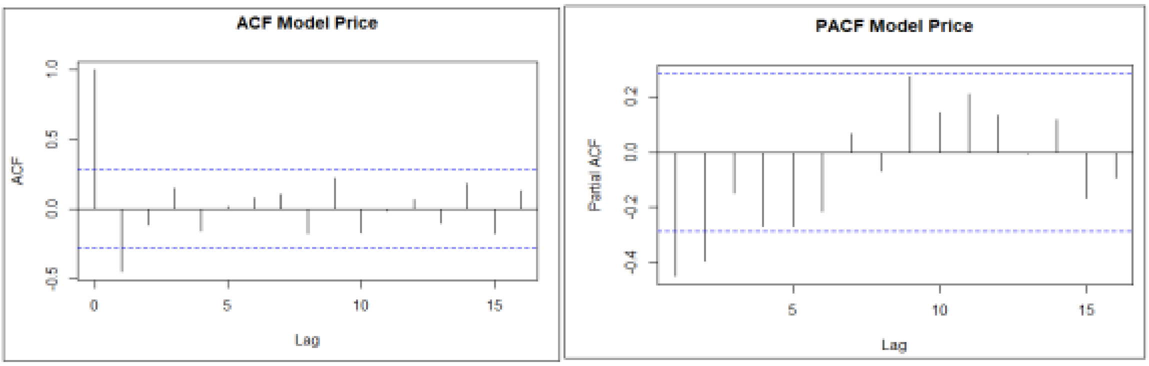

Figure 3.1.4: Transformed ACF and PACF plots of yearly maize production in Telangana.

From to the figures 3.1.3 and 3.1.4 above, the ACF and PACF die out with a first order difference. The p-value for the Augmented Dicky - Fuller test is 0.01, which is less than the significant level. As a result, the null hypothesis is rejected, and the data is stationary. The ADF test results in the table below demonstrate that the data is stationary for the first order difference..

Table 3.1.1: Augmented Dicky-Fuller Test

ADF test	
P-Value	0.01
Lag Order	3
ADF	-5.7

The auto correlation and partial auto correlation functions are used to identify the model parameters. Several alternative models were explored in the study to choose the optimal model based on parameter importance and model appropriateness. To test the LJung -Box Q Statistics, the adequate of model is employed. The table below shows some preliminary models.

Table 3.1.2: Tentative model of ARIMA

ARIMA(p,d,q)			AIC	BIC	Significance parameters	L-Jung box Q	p-value	Adequacy
1	1	0	31.29	34.46	Significant	9.86	0.13	Adequate
0	1	1	34.32	37.48	Significant	13.73	0.03	Inadequate
1	1	1	33.29	38.04	Insignificant	9.85	0.08	Adequate

1	1	2	33.64	39.87	Insignificant	6.22	0.18	Adequate
0	1	2	33.09	37.84	Significant	9.61	0.09	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	29.82	36.15	Significant	8.70	0.07	Adequate
2	1	2	28.72	36.64	Significant	4.58	0.21	Adequate

From the table 3.1.2, the ARIMA (2, 1, 1) model was used to anticipate the future yearly maize output in Telangana. The ARIMA (2, 1, 1) model is chosen from the following table based on parameter importance and appropriateness. The calculated parameters are shown in the table below.

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

Parameters	Estimated	Stnd.error	Z-Value	Pr(>|z|)
ar1	-1.48	0.12	-11.88	<0.001
ar2	-0.64	0.13	-5.00	<0.001
ma1	1.00	0.08	12.29	<0.001

The above table 3.1.3, model ARIMA (2, 1, 1) parameter reveals the significance and this is the best model for forecasting Telangana annual maize output. The ARIMA (2, 1, 1) model equation is now used.

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

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

$$(1+1.48B+0.64B^2)\ \nabla = (1-1B)\ a_t$$

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

The hypothesis value in table 3.1.4 is greater than the p-value of 0.05; we accept the null hypothesis and conclude that the selected models are sufficient. As a result, the ARIMA (2, 1, 1) model is used to anticipate future maize output in Telangana, India. The following table shows the projected annual maize production.

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

Date	Actual Maize Production	Forecasted Maize Production
2010-2011	2068560	1439299
2011-2012	1892475	1887462
2012-2013	2943717	1593773
2013-2014	3524907	1721353
2014-2015	2308051	1711122
2015-2016	1751074	1643644
2016-2017	2882475	1750829
2017-2018	2752147	1636358
2018-2019	2082991	1736792
2019-2020	3643686	1660605
2020-2021	2942175	1708229
2021-2022	2802498	1685883

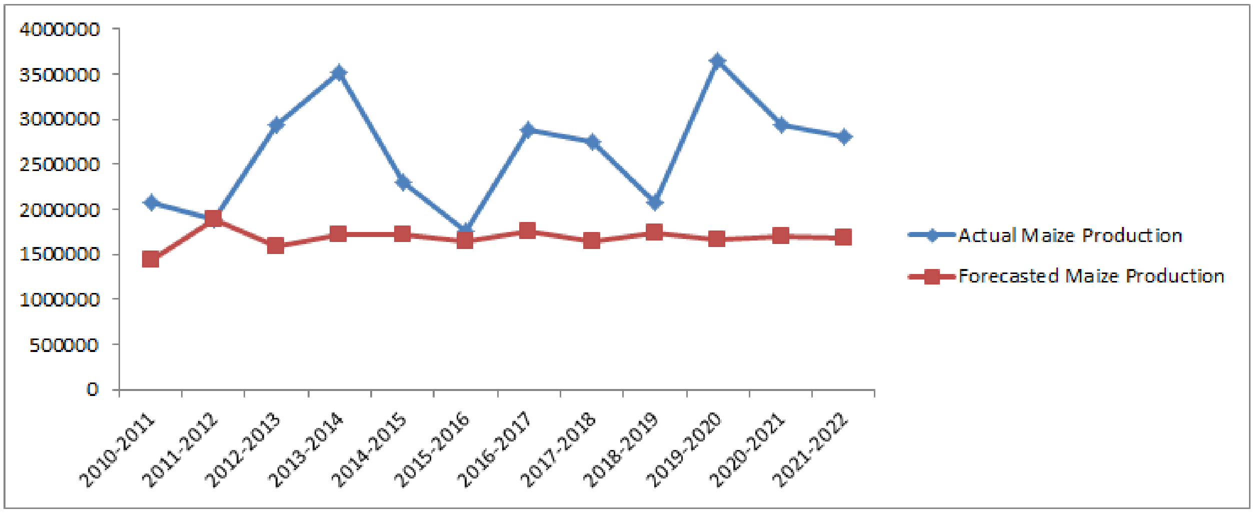

Figure 3.1.5: Forecasted of yearly maize production in Telangana using ARIMA (2, 1, 1)

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	0.31	0.24	1.20
Testing Set	0.76	0.69	3.22

As shown in table 3.1.6, the error measure for ARIMA (2, 1, 1) yields the best values of 0.31 and 0.76 million tonnes for RMSE and MAPE in training and test data sets, respectively. The error measurements and data pattern are extremely near to the test and anticipated values, and the higher level was proposed.

3.2 Feed Forward Neural Network Model

In our study, the feed forward neural network is made up of input neurons known as lag1. The output layer is one and provides a projection of the annual maize production in Telangana, India. Now, there is no special technique for considering the number of hidden layers in the model without developing a forward or backward selection strategy to identify the hidden layers. In our model, the applied hyperbolic tangent function is utilised for activation function under the back propagation approach, and multiple models

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

were validated and the best one was chosen based on lowest error measures such as MAPE, RMSE, and MAE. The preliminary models are shown in the table below.

Table 3.2.1: Possible Model of FFNN

Number of Layer			Error Measures for Train set			Error Measures for Test set		
Input	hidden	Output	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	2	1	0.39	0.30	1.55	0.50	0.40	1.86
1	3	1	0.35	0.29	1.47	0.59	0.50	2.34
1	4	1	0.39	0.30	1.54	0.53	0.42	1.97
1	5	1	0.36	0.29	1.47	0.54	0.44	2.04

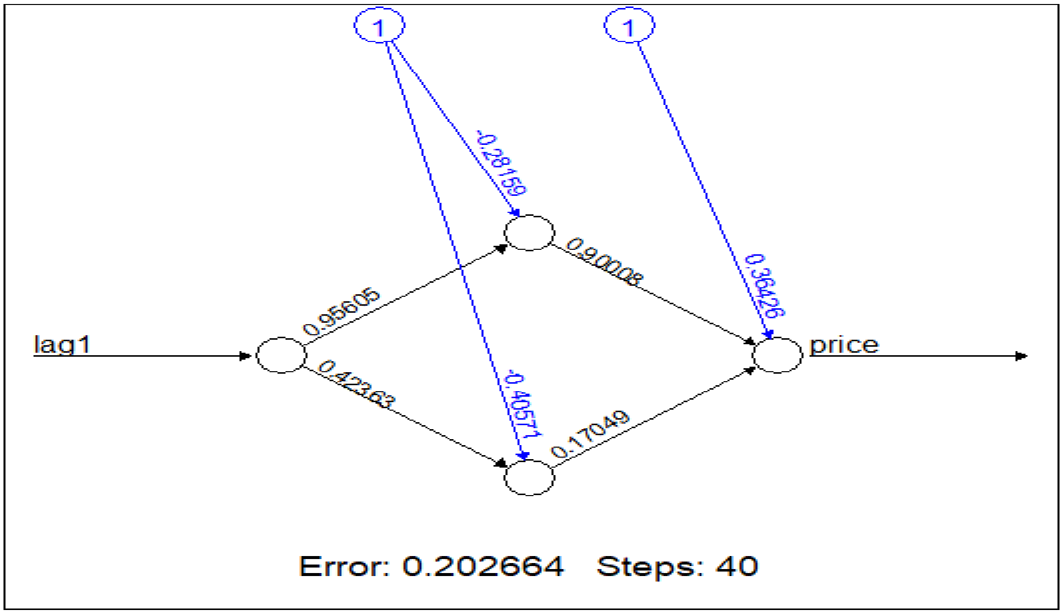

Figure 3.2.1: Feed Forward Neural Network Flow

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	0.20
reached. threshold	0.01
steps	40
Intercept.to.1layhid1	-0.28

lag1.to.1layhid1	0.96
Intercept.to.1layhid2	-0.41
lag1.to.1layhid2	0.42
Intercept.to.price	0.36
1layhid1.to.price	0.90
1layhid2.to.price	0.17

The hidden neurons are

$H_1 = Tanh[-0.28-0.96\bar{Z}_{t-1}]$

$H_2 = Tanh[-0.41-0.42\bar{Z}_{t-1}]$

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

$O_t = 0.90 H_1 + 1.17 H_2 + 0.36$

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
Train Set	0.39	0.30	1.55
Test Set	0.50	0.40	1.86

From table 3.2.3, the error measure for FFNN (1-2-1) yields the best values of 0.39 and 0.50 million tones for RMSE in training and test data sets, respectively. The error measurements and data pattern are close and near for MAPE values to test and forecast, and the above level was indicated.

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

Date	Actual Maize Production	Forecasted Maize Production
2010-2011	2068560	1340389
2011-2012	1892475	1880911
2012-2013	2943717	1754380
2013-2014	3524907	2434561
2014-2015	2308051	2745029
2015-2016	1751074	2044264

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

2016-2017	2882475	1648600
2017-2018	2752147	2399446
2018-2019	2082991	2323036
2019-2020	3643686	1891029
2020-2021	2942175	2803816
2021-2022	2802498	2433683

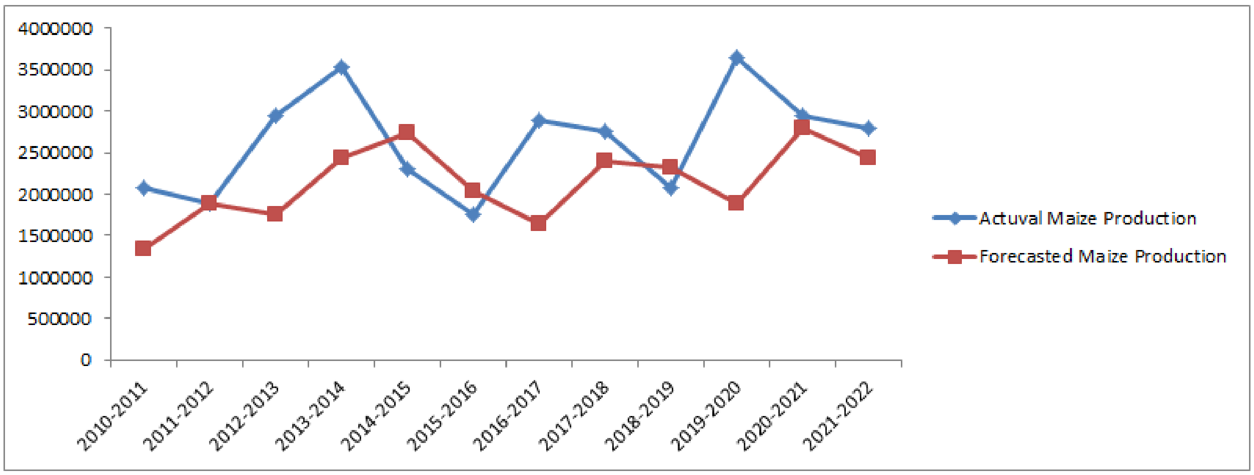

Table: 3.2.2: Forecasted yearly maize production in Telangana using FFNN (1-2-1) model

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	0.31	0.24	1.20	0.39	0.30	1.55
Test Set	0.76	0.69	3.22	0.50	0.40	1.86

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

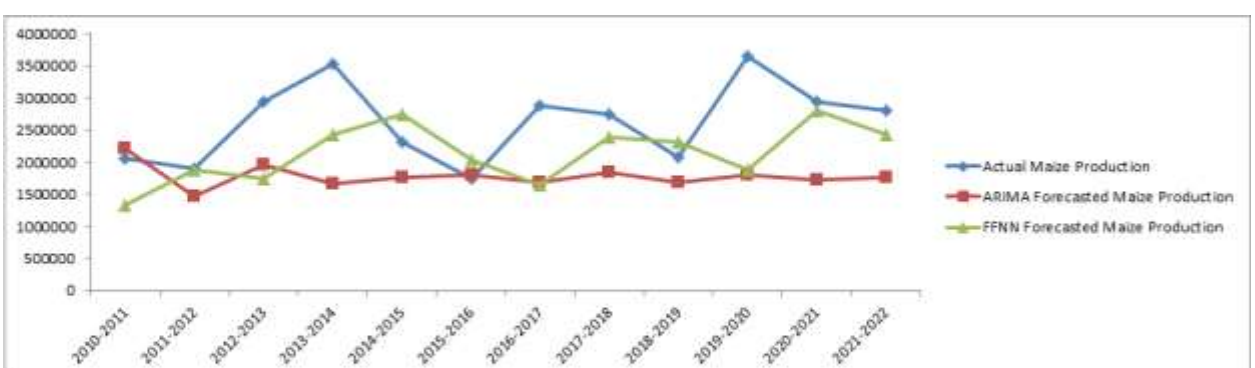

Figure 5.1: Forecasts of ARIMA and FFNN models

A summary of the findings of the study, the FFNN model outperforms the ARIMA model for fitting and forecasting yearly maize output in Telangana. As a result, it is determined that the FFNN model was used to anticipate Telangana annual maize production.

6. 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.