Forecasting the efficiency of Red gram Crop Yield Production in Telangana Using Hybrid Models

V. Raja Venkat Ram, Dr. M. Raghavender Sharma,

Department of Statistics, Osmania University, Hyderabad

ABSTRACT

Red gram, scientifically known as Cajanus Cajan, is an important perennial legume crop widely cultivated in semiarid tropical regions. Statistical Modelling of Linear and Non-Linear statistics has become a large amount of challenge in the combined field of the research analysis. Effective planning and risk management require accurate forecasting and timely policy formulation. A number of popularly used models square measure ARIMA and ANN. This text presents a comparison of ARIMA Time Series Models and Artificial Neural Network (ANN) ensemble with Hybrid Models for predicting the true yield of Red gram crop worth in the state of Telangana. The most important objective of this analysis is to develop a forecasting model to predict the yield for production of the Red gram crop with high accuracy and efficiency. During this paper, a statistic forecasting model, ARIMA and Artificial Neural Networks was developed for forecasting the yearly basis of yields production of the Red gram crop in Telangana. The forecasting performance of the model was evaluated efficiently by showing the accuracy of Root mean squared error (RMSE), Mean Absolute percent Error (MAPE). The yearly forecasts counsel that, the yield production of Red gram crop with a regular deviation of 8% error measure with the accuracy of 92% for the forecasted period of 24 months i.e. 2020-2022 (since Covid period of time).

keywords: *ARIMA, ANN, MAE, RMSE and MAPE*

INTRODUCTION

A statistic is the basic object of the study in varied sectors of analysis. Conventionally statistic modelling includes an underlying assumption that there's a linear underlying relationship between the past and future values of the series.

Agriculture is known to be the spine of the Indian Economy for the last many decades for any kind of growth and price of all the crops. Red gram is cultivated in Two lakhs hectares across the Telangana region making it one of the major crops of the state. It is widely yield growth in Mahbubnagar, Warangal, Nalgonda, and Karimnagar Districts. Yield rotation is crucial in Red gram farming, this helps to utilize nutrients efficiently and to reduce soil-borne diseases. Red gram is more beneficial to human nutrition, and is an important product since it is used in the prices of many foods and ranked second among seed plants after cotton, chilli, soyabean, and sugarcane. Red gram also has a particular economic value since its grams, kernels, shell, and straw can be used commercially as well as extensively throughout the state of Telangana. The yield of the Red gram crop for Telangana is based on the Indian harvested area. Increasing yield would also be an important tool in the development of rural areas of the state by increasing the growth of revenues. Thus, it is important for Telangana to formulate schemes aiming at increasing Red gram price for the future sustainability of all the industry, export revenues, and food safety. An effort is made during this paper to assess the yearly yield of the Red gram crop in the state of Telangana and to forecast the same for a brief tenure by victimization applied mathematics strategies. The subsequent section presents the results that supported the Box-Jenkins methodology of hybrid models and artificial neural networks.

REVIEW LITERATURE

The review of literature on crop production analysis predicts using subset of artificial intelligence which has machine learning and deep learning techniques has a significant growth in current years, with the researchers elaborate various methodologies to develop forecasting accuracy. Moreover, many studies have investigated the use of these technologies to influence past historical data and other significant features for the prediction of crop production.

"Parmar et al. and Usmani et al. introduced the concept of machine learning techniques for predictive analysis of crop production, emphasizing the efficiency of these methods in exploring forecasting accuracy. Billah et al. proposed an enhanced training algorithm for artificial neural networks, elaborates about advancements in algorithmic approaches. Li et al. developed into sentimental analysis about the analysis of crop production using predictive models, henceforth also a deep learning method which considers more predictive analysis of crop production." [8] [9] [11] [12]

"Forecasting analysis of time series approaches have also gained eminent nature of predictive analytics. Mehtab and Sen also focused on applied time series analysis prediction models using Artificial Intelligence techniques. Additionally, Jigar Patel et al explored the fusion of machine learning techniques." [13] [17]

"Ensemble methods has explored for agricultural crop market prediction by Cheng et al., enhancing the merits of combining multiple models for better accuracy. Gunduz et al. applied deep neural networks for prediction of agricultural crop production, which gives the efficacy of advanced models in acquiring crop dynamics."[15] [16]

"Traditional modelling methods, such as Regression and ARIMA models, were also developed. Sujatha and Sundaram used the developed models of regression and neural network for forecasting crop production, while Ariyo et al. elaborated the ARIMA model for forecasting the efficiency of crop production." [20] [21]

In the summary, literature review explores a eminent collection of methodologies for forecasting agricultural crop market, as well as encompassing artificial intelligence components such as ensemble methods of modelling. Researchers persist to explore more innovative technological approaches towards enhancement of forecasts accuracy and capture the challenges in agricultural markets.

ARIMA (Auto Regressive Integrated Moving Average) MODEL USING Hybrid Models

During this section, the modelling of Red gram yield production of Telangana State Box-Jenkins methodology is mentioned. The Box-Jenkins procedure relates to the fitting of the associate ARIMA model of the subsequent type for the given set of information and therefore the general kind of ARIMA (p, d, q) model is given by

$$\Phi(B)\nabla^{d} Zt = \theta(B)at$$

$$\text{Where } \Phi(B) = 1 - \Phi 1B - \Phi 2B^{2} - - - - - \Phi p B^{p}$$

$$\theta(B) = 1 - \theta 1B - \theta 2B^{2} - - - - \theta q B^{q}$$

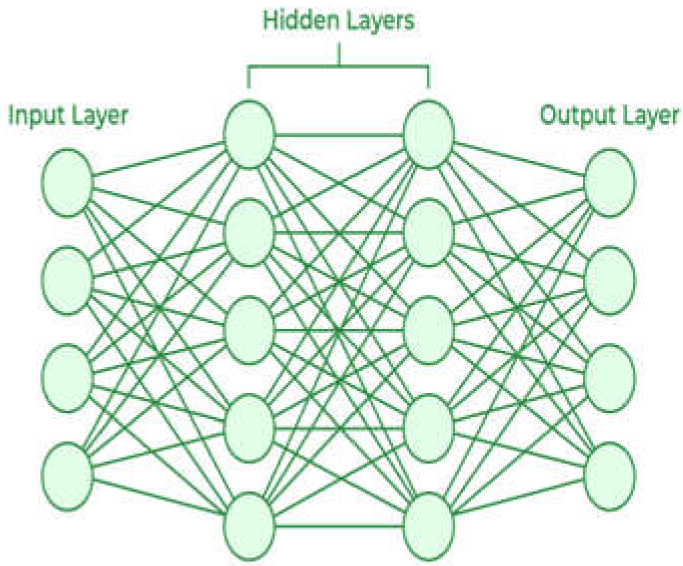
$$\text{And } \nabla^{d} = (1 - B)^{d}$$

We have $B^{K}Zt = Zt–kandat$ is a white noise process with zero mean and variance σ^{2}.The

Box-Jenkins procedure consists of the subsequent four stages. (1) Model Identification, wherever the orders d, p, and q are determined by perceptive the behaviour of the corresponding Autocorrelation function (ACF) and Partial Autocorrelation Function (PACF). (2) Estimation: wherever the parameters of the model are a unit calculable by the most probability methodology. 3) Diagnostic checking by the “Portmanteau Test”, where the adequacy of the fitted model is checked by the Ljung-Box datum, applied to the residual of the model. (4) Forecasts area unit obtained from associate degree adequate model victimization minimum mean square error methodology. If the model is judged to be inadequate, stages 1-3 area unit perennial with completely different values of d, p, and q, till associate the adequate model is obtained (Box et al; 1994).

ANN (Artificial Neural Network) MODEL

An Artificial Neural Network could be a mathematical model that is impressed by the structure and useful aspects of the biological neural network, a powerful predictive model. Associate degree ANN will estimate any nonlinear continuous function up to any desired degree of accuracy. It is widely used in a range of industries, business engineering, and sciences. It has the power to perfectly predict the longer term and is prime to several call processes in designing, scheduling, purchasing, strategy formulation, policy-making, and providing chain operations.



The characteristics of ANN that build it applicable for predictions are its non-linear structure, flexibility, knowledge-driven learning method, and its ability to estimate method universal functions. Neural networks area units precisely shown to possess the universal sensible approximating potential during which they will accurately approximate several varieties of advanced sensible relationships. This can be a very important and powerful characteristic, as any prediction model aims to accurately capture the useful relationship between the variable to be foreseen and different relevant factors or variables. The mixture of the abovementioned characteristics makes ANN a really general and versatile modelling tool for prediction. Finally, ANNs are unit non-linear models. The very fact that globe systems are unit typically non-linear has led to the event of many non-linear statistical models in the last decade. (Hornik,1993; Ramakrishna et al., 2011).

RESULTS AND DISCUSSION

FORECASTING THE EFFICIENCY OF RED GRAM CROP USING HYBRID MODELS

In this paper, the building of prediction models victimization Box-Jenkins methodology for the yearly yield production of Red gram crop is mentioned.

The Mathematical equations of Box Jenkins Methodology states with the development of Predictive models of Red gram crop yield production with the data on yearly basis.

The data on the yearly prices of Red gram were collected from the year 1979 from the Directorate of Economics and Statistics (DAES). The yearly yield of Red gram crop from 1970 to 2018 was used for model building and therefore the yearly Red gram crop from 2019 to 2021 was used for model validation. The prediction models for the prediction of the yearly yield of Red gram crops were developed victimization Box-Jenkins methodology and Artificial Neural Networks. The yearly yield of Red gram crop varied with an average price of 3900 Rs. The subsequent chart shows the time trend of the yearly yield of Red gram crop from 1979. The yearly yield of Red gram crop show a non-stationary time series (**Fig 1**). The average Red gram yield was comparatively low in the year 2019 Jan high in 2019 Oct due to the low and high rain fall during the above years (**Fig 1**). Also shown one of the outliers in the figure to be out of confusion about the prices of response variable.

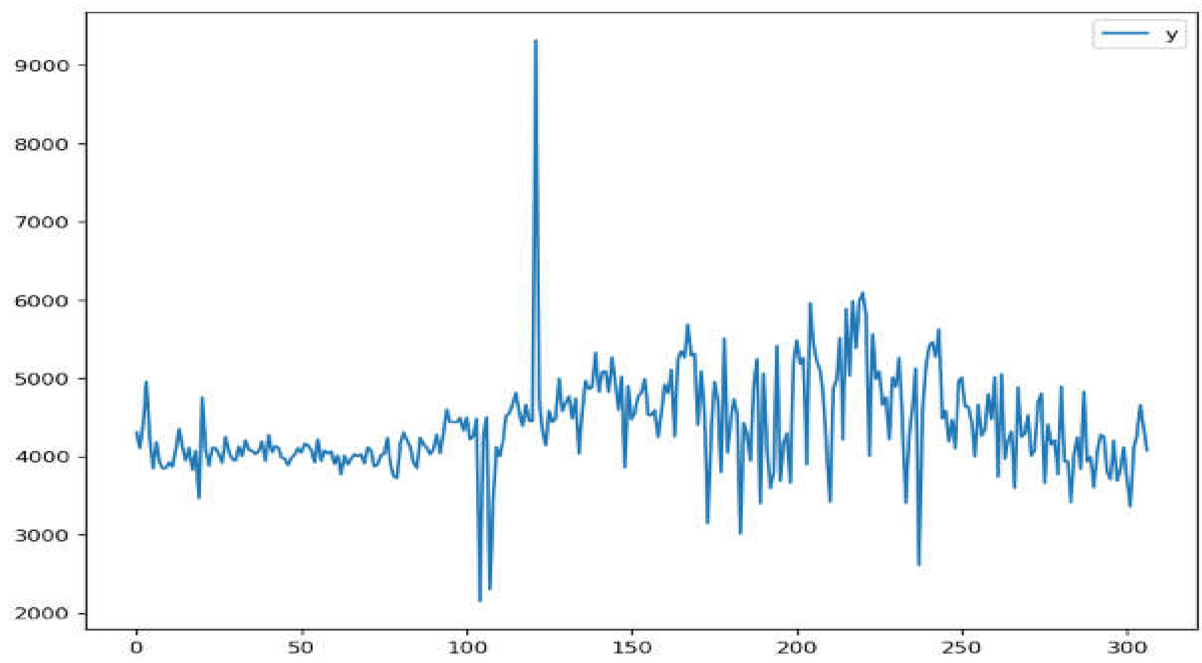


Fig1*:* The yearly average for the yield of Red gram crop (in kgs*)*

The python skilled creator was accustomed to determining the most effective ARIMA model for the yield of Red gram, as this plan automatically determines and estimates the best-fitting ARIMA for one or additional variable series, therefore eliminating the necessity to spot an applicable model through the trial-and-error methodology. It is discovered that the ARIMA (0, 1, 1) model fits the data well as compared with another outcome ARIMA(1,1,0) and therefore the same is tested on the validation set with the original series and orderwise differencing for ACF and PACF. The model parameters square measure is given in the following Table 1 and Table 2

A. Original Series, Differencing, and ACF (Auto Correlation Function)

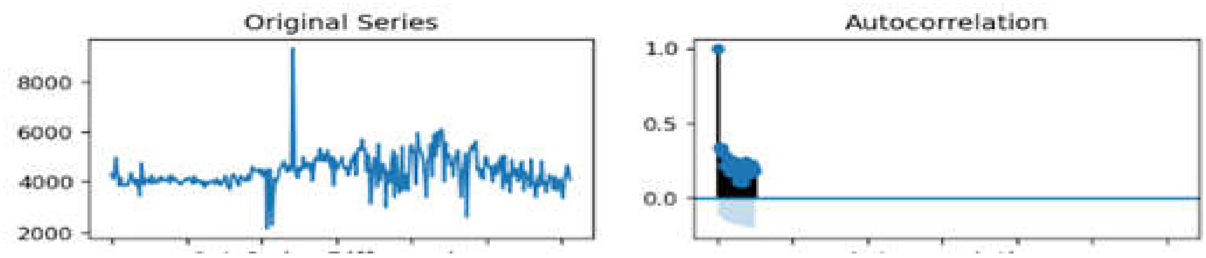


Fig2: Time series of Red gram yield in Telangana Original Series with ACF

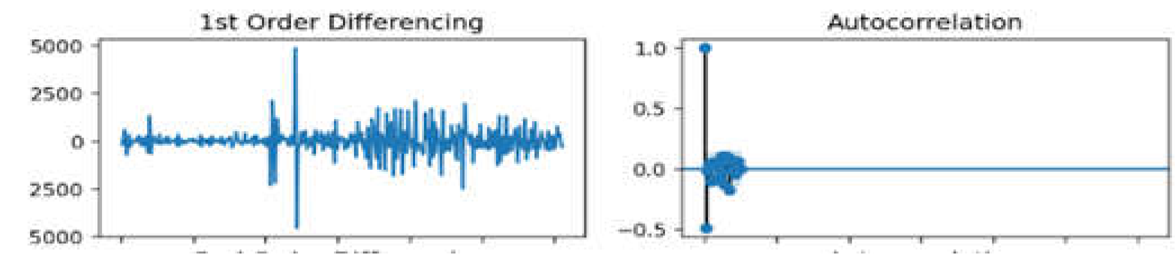


Fig3: Time series of Red gram yield in Telangana 1st Order differencing with ACF

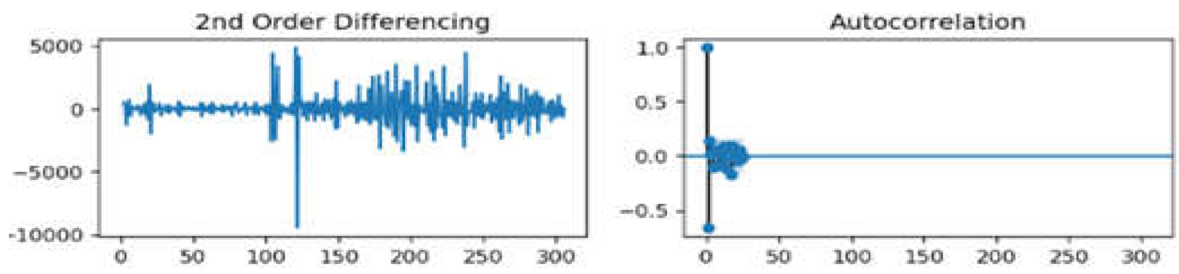


Fig4: Time series of Red gram yield in Telangana 2nd Order differencing with ACF

B. Original Series, Differencing, and PACF (Partial Auto Correlation Function)

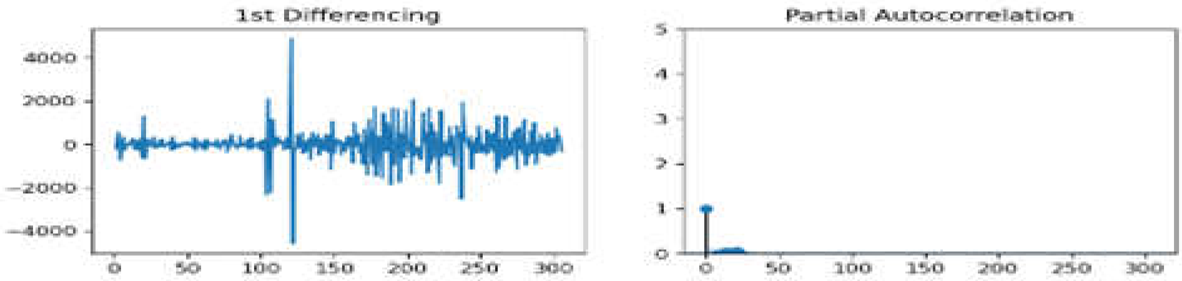
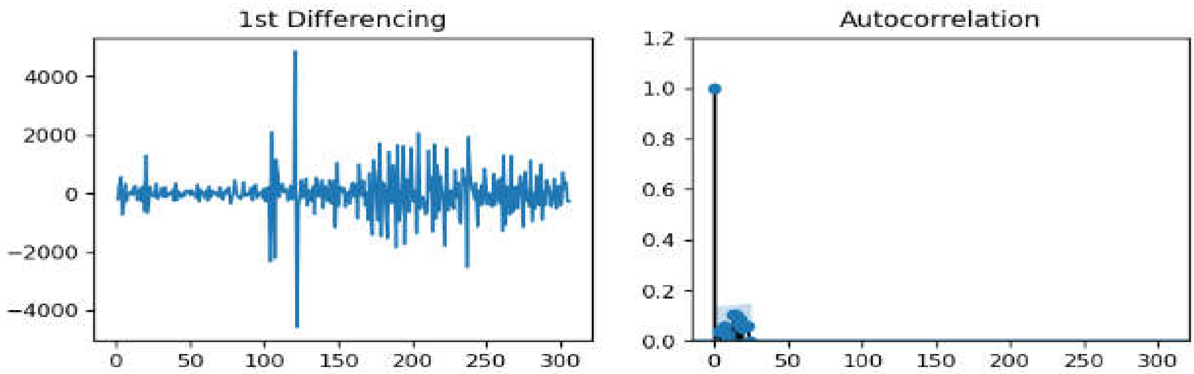


Fig5: Time series of Red gram yield in Telangana, 1st Order differencing with PACF



v

Fig6: Time series of Red gram yield in Telangana 1st Order differencing with ACF

Table 1. ARIMA Model Parameters

Model	a	Parameters Estimate					Goodness of Fit				
		Autoregressive Coefficient			Moving Average Coefficient		R^2	RMSE	MAPE	AIC	BIC
		AR1	AR2	AR3	MA1	MA2					
Area											
(1,1,0)	31.23*	-0.45**					0.86	133.67	12.09	588.14	593.63
(1,1,1)	31.23	-0.45*			-0.01		0.86	133.67	12.10	590.14	597.46
(1,1,2)	30.26	0.10			-0.61**	0.62**	0.88	122.03	11.48	584.54	593.69
(2,1,2)	**30.43***	**0.14****	**-0.13****		**-0.64****	**0.70****	**0.91**	**112.03**	**11.37**	**586.15**	**597.12**
(2,1,1)	35.67	0.34	0.46**		-0.75**		0.86	131.41	11.78	590.71	599.85
(3,1,1)	31.75	-0.40	0.19	0.36**	-0.05		0.88	123.63	11.41	587.37	598.35
(3,1,2)	30.48	-0.37	-0.10	0.23	-0.09	0.39	0.88	121.04	12.03	587.62	600.42
Production											
(1,1,0)	21.85	-0.63**					0.64	178.96	46.87	625.01	630.49
(1,1,1)	21.00	-0.42*			-0.44**		0.70	164.93	46.94	620.53	627.85
(1,1,2)	27.13	-0.18			-0.69**	0.59*	0.73	157.84	44.02	619.57	628.72
(2,1,2)	**25.93***	**-0.60****	**-0.52****		**-0.27***	**0.65****	**0.77**	**147.58**	**43.93**	**616.71**	**627.69**
(2,1,1)	22.10	-0.91**	-0.43*		0.01		0.73	158.75	45.44	619.55	628.70
(3,1,1)	27.94	-0.05	0.41**	0.45**	-0.83**		0.73	156.49	42.59	620.59	631.57
(3,1,2)	25.59	-0.67*	-0.62	-0.09	-0.22	0.68**	0.75	147.66	45.04	618.55	631.35
Productivity											
(1,1,0)	0.07	-0.38**					0.62	119.87	21.95	614.35	619.83
(1,1,1)	2.23	-0.05			-0.72**		0.57	128.9	18.73	602.49	609.80
(1,1,2)	2.14	-0.66*			0.05	-0.61**	0.60	126.92	18.28	603.27	612.41
(2,1,2)	1.14	-0.27	-0.43**		-0.48	0.01	0.57	128.85	17.53	600.49	611.47
(2,1,1)	**1.16***	**-0.28***	**-0.43****		**-0.46***		**0.64**	**118.85**	**17.5**	**598.50**	**607.64**
(3,1,1)	1.16	-0.29	-0.44	-0.01	-0.46		0.59	125.85	17.54	600.49	611.47
(3,1,2)	1.20	-0.66	-0.53	-0.15	-0.08	-0.19	0.58	128.83	17.53	602.48	615.29

Hence, the fitted model for the forecasting the efficiency of Red gram yield in the state of Telangana is ARIMA (2,1,2) and ARIMA (2,1,1) $\nabla Z\text{t} = (1 + 0.052B - 0.119B\ )at$.

The adequacy of the model was checked by exploitation Ljung-Box q test statistic and therefore the same is discovered that Q=13.673 at 20 degrees of freedom. The corresponding p-value of the letter check datum is 0.62 and which is far larger than 0.05, hence, the null hypothesis of the adequate model was accepted and therefore the given ARIMA (0,1,1) model may be an appropriate model for the prediction of the yield of Red gram. Similarly, a synthetic neural networks model was developed for the prediction of the yield of Red gram using python.

FEED FORWARD NEURAL NETWORK USING ANN

ANN (Artificial Neural Network) model may be a feed-forward neural networks model having one input layer, one hidden layer Associate in Nursing, and an output layer. The standardized values of previous observation (Lag-1 or Zt -1) were used as Associate in nursing input for the one-step a head prediction for the yearly yield of Red gram crop during this ANN model. The hidden layer consists of two hidden neurons to capture the nonlinearity element within the statistic. The hyperbolic tangent function is employed as an activation function in the hidden layer and the identity is used as an activation function in the output layer. The subsequent Figure 6 shows a typical feed-forward neural network used for the prediction of Red gram yield crop in Telangana state. The ANN model was trained exploitation back propagation rule until the error measures of the testing sample are smaller than the coaching sample on a trial-and-error basis.

Gradient decent back propagation algorithm: The objective of training is to minimize the error function that measures the misfit between the predicted value and the actual value. The error function which is widely used is mean squared error

$$E = \frac{1}{N}\sum_{n-1}^{N}(e_i)^2 = \frac{1}{N}\sum_{n-1}^{N}\left\{y_t - f\left(\sum_{j-0}^{q}\omega_j g\left(\sum_{i-0}^{p}\omega_{ij} y_{t-i}\right)\right)\right\}^2$$

Where, N is the total number of error terms. The parameters of the neural network are ωj and ωij estimated by iteration. Initial connection weights are taken randomly from uniform distribution. In each iteration the connection weights changed by an amount ωj

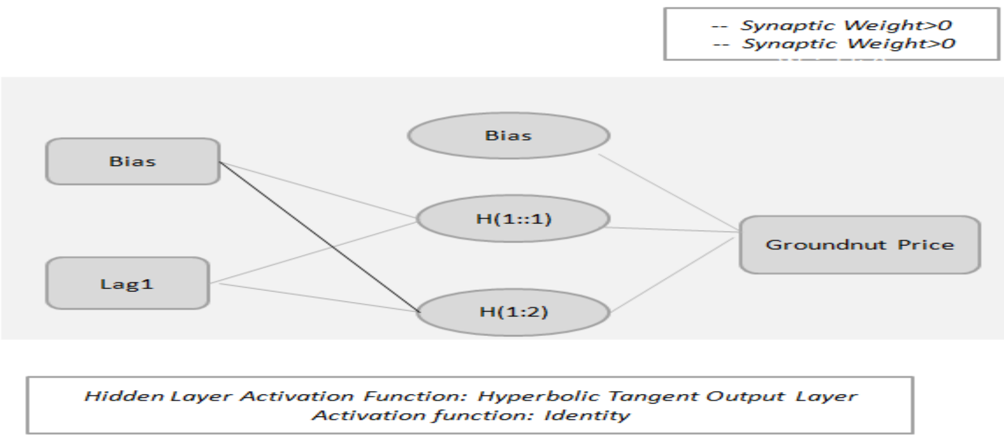


Fig7: Feed forward neural network prediction model for the yearly yield of Red gram.

Table 2. Feed forward Neural Networks Model Parameters

Predictor		Predicted		
		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S_t
Input Layer	(Bias)	0.021	-0.052	
	lag1	0.152	0.267	
Hidden Layer 1	(Bias)			0.119
	H(1:1)			2.213
	H(1:2)			3.104

Hence the ANN Model is here.

$L=(Z_{t-1}\ 32301.15)/3313.543$; $h_1=\tanh(0.021+0.152*L)$; $h_2=\tanh(-0.052+0.267*L)$;
$S_t=0.119+2.213*h1+3.104*$ h2and $Z_t= 32301.52+3313.543*S_t$

COMPARISON OF ARIMA AND ANN MODELS ON ERROR MEASURES

The predictions from the two models were compared in the training sample and testing samples supported the mean absolute error, root means square error and mean absolute percentage errors. The subsequent Table 3 presents the error measures from the ARIMA and ANN prediction models.

Table 3. Comparison of the forecasting performance of ARIMA and ANN models

Measure	Training Sample		Testing Sample	
	ARIMA	ANN	ARIMA	ANN
MAE	323.23	352.31	417.77	415.45
RMSE	519.03	521.11	516.67	514.26
MAPE	0.08	0.06	0.08	0.07

The ANN model has comparatively lower error measures in the testing sample as compared to the ARIMA Model. The ANN fits well stronger than ARIMA within the training and testing samples not show stronger performance in the testing sample. The following figure 7 shows the out-of-sample forecasts supported by ARIMA and ANN models. The ANN model forecasts show a similar trend to the original costs, whereas the ARIMA forecasts show a linear trend over time.

Table 4. Out-of-sample forecasts of Red gram yield crop production in Telangana state using ARIMA and ANN Models

Yield of Red gram in Telangana State Using ARIMA and ANN Models for the 2030 monthly prices

Year - Month	(Rs.)	ARIMA	ANN
2030 - Jan	5950	5954	6044
2030 - Feb	6811	6815	6905
2030 - Mar	9514	9518	9608
2030 - Apr	9650	9654	9744
2030 - May	9851	9855	9945
2030 - Jun	8923	8927	9017
2030 - Jul	8962	8966	9056
2030 - Aug	9453	9457	9547
2030 - Sep	9550	9554	9644
2030 - Oct	9950	9954	10044
2030 - Nov	10488	10492	10582
2030 - Dec	13220	13224	13314
Mean	9360.167	9364.167	9454.167
SD	1793.64	1793.64	1793.64

CONCLUSION

The forecasts recommend that the ANN model predicts well the yield of Red gram crop in Telangana State as compared to the ARIMA model. ARIMA model provides only linear trends whereas the ANN model presents the nonlinear fluctuations within the forecasts and efficiencies with error measures. The conclusion clearly states with the methodology of Box Jenkin's on efficiency of hybrid models from the Literature reviews.

REFERENCES

[1] Box.G.E.P. Jenkins, G.M. and Reinsel, G.C. (1994). Time Series analysis forecasting and control, 3rd ed, Engelalwood cliffs, N.J. PrenticeHall,

[2] BROCKWELL P.J., DAVIS R.A (1996), An introduction to time series and Forecasting, Springer, Verlag.

[3] C.Gershenson, “Artificial Neural Networks for Beginners” Networks, 2003., Vol. cs.NE/0308, pp.8.

[4] Makridakis S., Wheel Wright. S.C., Hyndman R.J., 2003, Forecasting Methods and Applications, John Wiley &Sons.

[5] Hornik, K, (1993), Some new results on neural network approximation, Neural Networks, 6, 1069-1072.

[6] K.Murali Krishna, Dr.M.Raghavender Sharma and Dr.N.Konda Reddy, Forecasting of silver prices using Artificial Neural Networks. JARDCS, Volume 10, 06 issue 2018.R. J. Vidmar. (1992, Aug).

[7] Patra C, Gupta A K and Kumar R R 2020. Forecasting of area, production and yield of jute crop in India using ARIMA model. Journal of Pharmacognosy and Phytochemistry. 9(2):201-207.

[8] Parmar et al., "Agricultural Crop Market Prediction Using Machine Learning", 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), pp. 574-576, 2018.

[9] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, " Agricultural Crop market prediction using machine learning techniques", 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), pp. 322-327, 2016.

[10] GR Mettam and LB Adams, "How to prepare an electronic version of your article" in Introduction to the electronic age, New York:E-Publishing.

[11] M. Billah, S. Waheed and A. Hanifa, " Agricultural Crop market prediction using an improved training algorithm of neural network", 2016 2nd International Conference on Electrical Computer & Telecommunication Engineering (ICECTE), pp. 1-4, 2016.

[12] Li Jiahong, Bu Hui and Wu Junjie, "Sentiment-aware Agricultural Crop market prediction: A deep learning method", 2017 International Conference on Service Systems and Service Management, pp. 1-6, 2017.

[13] Sidra Mehtab and Jaydip Sen, A Time Series Analysis-Based Agricultural Crop market Prediction Using Machine Learning and Deep Learning Models, May 2020.

[14] K. A. Althelaya, E. M. El-Alfy and S. Mohammed, " Agricultural Crop market Forecast Using Multivariate Analysis with Bidirectional and Stacked (LSTM GRU)", 2018 21st Saudi Computer Society National Computer Conference (NCC), pp. 1-7, 2018.

[15] C. Cheng, W. Xu and J. Wang, "A comparison of ensemble methods in Agricultural Crop market prediction", 2012 Fifth international joint conference on computational 516 sciences and optimization (CSO), pp. 755-759, 2012.

[16] H. Gunduz, Z. Cataltepe and Y. Yaslan, " Agricultural Crop market direction prediction using deep neural networks", 2017 25th Signal Processing and Communications Applications Conference (SIU), pp. 1-4, 2017.

[17] Jigar Patel et al., "Predicting Agricultural Crop market index using fusion of machine learning techniques", Expert Syst. Appl., vol. 42, pp. 2162-2172, 2015.
[18] Karol Gryko et al., "Anthropometric variables and somatotype of young and professional malebasketballplayers", Sports, vol. 6, no. 1, pp. 9, 2018.
[19] S. Mehtab and J. Sen, " Agricultural Crop market Prediction Using CNN and LSTM-Based Deep Learning Models", 2020 International Conference on Decision Aid Sciences and Application (DASA), pp. 447-453, 2020.
[20] K. V. Sujatha and S. M. Sundaram, " Agricultural Crop market using regression and neural network models under non normal conditions", INTERACT-2010, pp. 59-63, 2010.
[21] A. Ariyo, A. O. Adewumi and C. K. Ayo, " Agricultural Crop market Using the ARIMA Model", 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp. 106-112, 2014.