The Andhra Agric. J 72 (1): 093-097, 2025
doi :10.61657/aaj.2025.181

An artificial neural network(ANN) model for predicting groundnut price

P D Selvi, A Vani, Shaik Nafeez Umar and M Siva Parvathi
Department of Applied Mathematics, Sri PadmavatiMahilaVisvavidyalayam
S. V. University, Tirupati-517501, Andhra Pradesh, India.

ABSTRACT

The fluctuating nature of agricultural commodity prices presents a significant challenge for farmers, traders, and policymakers. Groundnut, being a vital cash crop, is no exception. This research focuses on developing a predictive model using Artificial Neural Networks (ANN), specifically the Generalized Regression Neural Networks (GRNN), to forecast groundnut prices. By leveraging data spanning five decades (1966-2016) collected from the Chittoor District of Andhra Pradesh, the study evaluates the model's performance using Mean Absolute Percentage Error (MAPE) and Symmetric Mean Absolute Percentage Error (SMAPE). Diagnostic tests, including the Shapiro-Wilk normality test and Grubbs test, confirmed the suitability of the data for modeling, with no significant outliers detected. The GRNN model effectively predicted groundnut prices for 2017 to 2028, showing an upward trend with prices peaking at Rs. 8314 in 2027. The residual analysis yielded MAPE and SMAPE values of 34.9443% and 44.3991%, respectively, indicating reasonable accuracy despite inherent agricultural data volatility. These results provide valuable insights for stakeholders, aiding in better planning and risk management in the agricultural sector.

Keywords: *Generalized Regression Neural Networks (GRNN), MAPE, Prediction and SMAPE*

Agricultural commodity price prediction plays a crucial role in ensuring economic stability and enhancing decision-making for stakeholders involved in the agricultural sector. Groundnut, one of the essential oilseed crops cultivated in India, contributes significantly to the country's agricultural exports and rural economy. However, the price of groundnuts is influenced by several factors, including seasonal variations, market demand, supply conditions, and policy changes. These dynamic factors make accurate price forecasting a complex yet necessary task.
Recent advancements in machine learning and artificial intelligence have provided new opportunities for tackling such challenges. Among various approaches, Artificial Neural Networks (ANN) have emerged as a powerful tool for modeling non-linear relationships in time series data. Generalized Regression Neural Networks (GRNN), a variant of ANN, are particularly suitable for regression problems due to their fast learning and adaptability.

This study aims to develop a GRNN-based model to predict groundnut prices using historical data from the Chittoor District of Andhra Pradesh. The dataset, spanning the period from 1966 to 2016, was obtained from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). By utilizing the GRNN model in R-studio, this research seeks to provide a reliable and robust prediction mechanism, which can benefit farmers, policymakers, and market participants in devising effective strategies for managing production and market risks.
The following sections outline the methodology adopted, the data preparation process, and the implementation of the GRNN model, along with a comprehensive evaluation of its performance using MAPE and SMAPE metrics. The study highlights the potential of machine learning approaches in addressing real-world agricultural challenges and contributing to sustainable development.

MATERIAL AND METHODS

The data used in this study was obtained from the International Crops Research Institute for the Semi-Arid Tropics website, covering groundnut prices in the Chittoor District of Andhra Pradesh from 1966 to 2016. The analysis was conducted using R-studio and the generalized regression neural networks library.

Normality and outlier tests were performed on the dataset. The Shapiro-Wilk normality test was used to determine if the data follows a normal distribution. The Grubbs test was then employed to detect any significant outliers in the time series data, which could potentially impact the accuracy of the predictive model. The GRNN model was selected for forecasting groundnut prices due to its ability to effectively model non-linear relationships in time series data. The key parameter in the GRNN model is the smoothing parameter (σ), which controls the influence of training examples on the final prediction. The model was then used to generate price forecasts for the period from 2017 to 2028.

The Groundnut Price (Rs per Quintal) data collected from icrisat.org in Chittoor District of Andhra Pradesh from the year (1966-2016). The R-studio, generalized regression neural networks (GRNN) library were used for predicting Groundnut price. Normality test- Shapiro-Wilk normality test. The Shapiro-Wilk normality test is a statistical test used to determine if a dataset follows a normal distribution. It compares the data to a normal distribution with the same mean and standard deviation. The test statistic, denoted as W, ranges from 0 to 1, with values closer to 1 indicating that the data is more likely to have come from a normal distribution. The p-value associated with the test statistic provides the probability of observing the given test statistic if the null hypothesis (that the data is normally distributed) is true. If the p-value is less than the chosen significance level (e.g., 0.05), the null hypothesis is rejected, indicating that the data is not normally distributed.

Outliers test- Grubbs test. The Grubbs test is a statistical test used to detect outliers in a dataset. It is named after Frank E. Grubbs, who developed the test in 1969. The Grubbs test compares the largest value in the dataset to the mean and standard deviation of the dataset. The test statistic is calculated as the absolute difference between the suspected outlier and the mean, divided by the standard deviation. The formula for the Grubbs test statistic is:

$$G = \frac{|x_{max} - \mu|}{s}$$

where:

- x_{max} is the largest value in the dataset
- μ is the mean of the dataset
- s is the standard deviation of the dataset

If the test statistic G exceeds a critical value, the suspected outlier is considered a significant outlier. The Grubbs test is particularly useful in identifying outliers in time series data, as it can help ensure the reliability of the data used for predictive modeling. By detecting and removing significant outliers, the Grubbs test can improve the accuracy of the forecasting model. Generalized regression neural networks (GRNN).

Generalized Regression Neural Networks is a type of artificial neural network that is particularly well-suited for regression problems. It is a variation of the Radial Basis Function network, characterized by a fast single-pass learning algorithm.

In a GRNN, the hidden layer consists of RBF neurons, with each neuron representing a training example. The output of each neuron is a measure of the closeness of the input vector to the training example, typically using the multivariate Gaussian function. The output layer then computes a weighted average of the target values, where the weights are determined by the closeness of the input to the training examples.

The key parameter in a GRNN is the smoothing parameter, σ, which controls how much of the training data is used in the weighted average. When σ is large, the result is close to the mean of the training targets, as all targets have similar weights. When σ is small, only the closest training targets to the input vector have significant weights.

The advantages of GRNN include its fast learning, adaptability, and ability to model non-linear relationships in time series data, making it a suitable choice for price forecasting applications like the one presented in the paper.

Time series forecasting using generalized regression neural networks (GRNN). GRNN model associated with a time series and use the model to predict the future values of the time series. It is possible to consult how the prediction has been done

A general regression neural network is a variant of an RBF network characterized by a fast single-pass learning. A GRNN consists of a hidden layer with RBF neurons. Normally, the hidden layer has as many neurons as training examples. The center of a neuron is its associated training example and so, its output gives a measure of the closeness of the input vector to the training example. Commonly, a neuron will use the multivariate Gaussian function:

$$G(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)$$

where x_i and σ are the center and the smoothing parameter respectively (x is the input vector). Given a training set consisting of nnn training patterns (vectors $\{x_1, x_2, x_3, \dots x_n\}$)and their associated n targets, normally scalars ($\{y_1, y_2, y_3, \dots y_n\}$), the GRNN output for an input pattern x is computed in two steps. First, the hidden layer produces a set of weights representing the closeness of x to the training patterns:

$$w_i = \frac{\exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)}{\sum_{j=1}^{n} \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)}$$

Note that the weights decay with distance to the training pattern. The weights sum to one and represent the contribution of every training pattern to the final result. The GRNN output layer computes the output as follows:

$$\hat{y} = \sum_{i=1}^{n} w_i y_i$$

So a weighted average of the training targets is obtained, where the weights decay with distance to the training patterns. The smoothing parameter controls how many targets are important in the weighted average. When σ is very large, the result is close to the mean of the training targets because all of them have a similar weight. When σ is small, only the closest training targets to the input vector have significant weights.Performance of the model MAPE.

Mean Absolute Percentage Error: MAPE is a measure of the prediction accuracy, calculated as the average of the absolute percentage errors of the forecasted values. The formula for MAPE is:

$$MAPE = \frac{1}{n}\sum \frac{|Actual - Forecast|}{Actual}$$

where n is the number of observations, Actual is the true observed value, and Forecast is the predicted value.

MAPE provides a straightforward interpretation of the model's performance, as it represents the average percentage deviation of the forecasts from the actual values. A lower MAPE value indicates higher predictive accuracy, with 0% representing a perfect forecast. Generally, a MAPE value below 20% is considered good, while a value below 10% is considered excellent for time series forecasting. SMAPE. Symmetric Mean Absolute Percentage Error is a performance metric used to evaluate the accuracy of forecasting models. It is a modified version of the Mean Absolute Percentage Error that addresses the issue of MAPE's sensitivity to small actual values. The formula for SMAPE is:

$$SMAPE = \frac{1}{n}\sum \frac{|Forecast - Actual|}{|Forecast| + |Actual|}$$

Where n is the number of observations, Forecast is the predicted value, and Actual is the true observed value. SMAPE provides a symmetric measure of the forecast error, meaning it penalizes both over-predictions and under-predictions equally. The SMAPE value ranges from 0% to 100%, with 0% indicating a perfect forecast and 100% indicating the forecasts are completely off. Generally, a SMAPE value below 10% is considered excellent, below 20% is good, and below 50% is acceptable for time series forecasting.

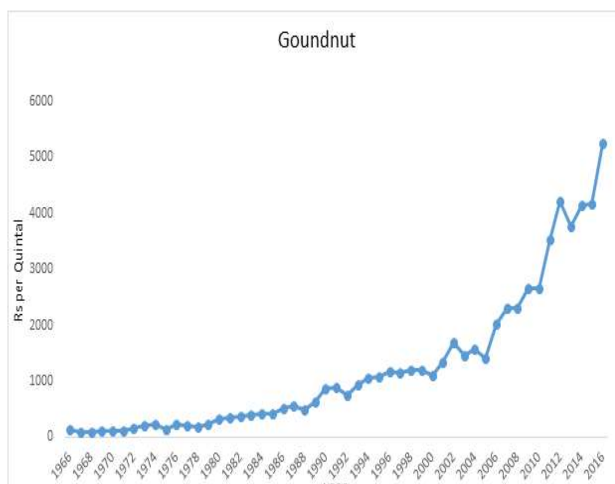


Figure1: Actual trend line of groundnut price (1976-216)

The trend line depicts the actual groundnut prices (Rs per Quintal) over five decades, showcasing a gradual increase from 1966 to 2000. A sharp rise in prices is observed post-2000, reflecting market dynamics such as increasing demand, inflation, and policy changes. This steady upward trend highlights the need for robust prediction models to address the challenges of price volatility and support stakeholders in planning effectively.

Table1: Diagnostic tests in groundnut price

Data Diagnostic test		Test Statistic	p-value
Normality test	Shapiro-Wilk's normality test	0.8017	0.07
Outliers test	Grubbs test	3.1037	0.0286

The Shapiro-Wilk's normality test were applied for data has to normal or not, in this regarding the. The Grubbs test value is given by 3.1037 and p = 0.0286, which is bigger than 0.05. Conclude that there is no evidence of outliers in this time series data.

Shapiro-Wilk's Normality Test: The test statistic of 0.8017 and a p-value of 0.0700 indicate that the null hypothesis (data is normally distributed) cannot be rejected at the 5% significance level. This implies that the groundnut price data follows a normal distribution, which is a desirable property for predictive modeling.

Grubbs Test for Outliers: The Grubbs test yielded a test statistic of 3.1037 with a p-value of 0.0286, which is greater than the threshold of 0.05. This indicates no significant outliers in the dataset. The absence of outliers ensures that the data used for modeling is consistent and reliable, avoiding the undue influence of extreme values on the predictions.

Table2: Prediction of groundnut price (2017 to 2028)

Prediction		
S No	**Year**	**Price (Rs per Quintal)**
1	2017	5569
2	2018	5064
3	2019	5490
4	2020	5553
5	2021	6704
6	2022	6884
7	2023	6541
8	2024	6963
9	2025	6991
10	2026	8140
11	2027	8314
12	2028	7967

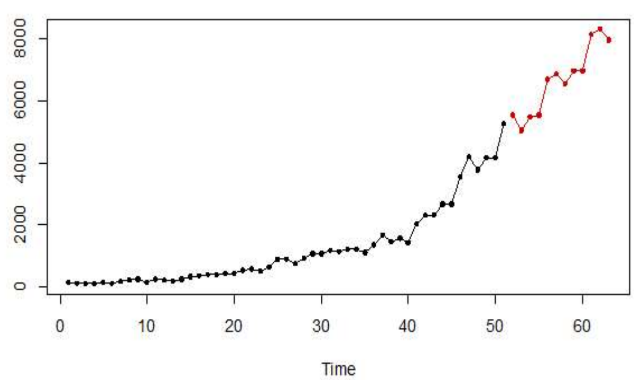
The predicted prices of groundnuts (in Rs per Quintal) indicate a general upward trend over the forecasted years. A moderate increase is observed from 2017 (Rs. 5569) to 2023 (Rs. 6541), showcasing a steady growth phase. Post-2023, the prices show sharper increases, peaking in 2027 at Rs. 8314 before slightly dropping to Rs. 7967 in 2028. This trend likely reflects increasing demand, inflationary pressures, or supply constraints. Such forecasts provide stakeholders a basis for planning production and marketing strategies, though they remain subject to unforeseen market and climatic factors.

Table3: Residual analysis of groundnut price

Performance model	Value
MAPE	34.9443
SMAPE	44.3991

Mean Absolute Percentage Error (MAPE): A MAPE value of 34.9443% indicates a reasonable level of predictive accuracy for the model. Although there is room for improvement, the predictions are within an acceptable range for agricultural data, which is inherently volatile.

Symmetric Mean Absolute Percentage Error (SMAPE): The SMAPE value of 44.3991% highlights the relative accuracy of the model while considering over- and under-predictions symmetrically. This measure reinforces the potential usability of the model despite some variance in predictions.



The graph illustrates the continuation of the upward trajectory from the historical trend (1966–2016). The predicted prices exhibit a smooth transition, aligning with the historical data patterns and indicating the robustness of the GRNN model. This visualization underscores the reliability of the forecasts in capturing the overall market trends.

CONCLUSION

In this paper a computational methodology built on GRNN regression for predicting time series an automatic way has been exhibited. The particular methodology used in straight forward in order to obtain predicting tool. Currently, the time series prediction in the R environment has been applied.This study demonstrates the application of Generalized Regression Neural Networks (GRNN) for predicting groundnut prices, using historical data from 1966 to 2016. The diagnostic tests confirm the normality of the data and the absence of significant outliers, ensuring the reliability of the dataset. The prediction results highlight a consistent upward trend in groundnut prices, providing valuable insights for farmers, traders, and policymakers.

Although the model achieved reasonable accuracy with MAPE and SMAPE values, further refinement could enhance predictive performance. The results underline the potential of machine learning approaches in addressing agricultural challenges and supporting stakeholders in making informed decisions to manage production and market risks effectively. Future research could explore incorporating additional variables or alternative machine learning models to improve forecast accuracy.

LITERATURE CITED

Abady S, Shimelis H, and Janila P 2019. Farmers' perceived constraints to groundnut production, their variety choice and preferred traits in eastern Ethiopia: implications for drought-tolerance breeding. *Journal of Crop Improvement, 33*, 505-521.https://doi.org/10.1080/15427528.2019.1625836

Damor P A, Ram B and Kunapara A N 2023. Stochastic Time Series Analysis, Modeling, and Forecasting of Weekly Rainfall Using Sarima Model. *International Journal of Environment and Climate Change, 13*, 773–782.

Thangavel C, and Sakthipriya D 2024. Machine Learning Ensemble Classifiers for Feature Selection in Rice Cultivars". *Applied Artificial Intelligence, 38*, e2394734.

Crone S F, Hibon M, Nikolopoulos K 2011. Advances in forecasting with neural networks Empirical evidence from the NN3 competition on time series prediction". *Int. J. Forecast. 27(3)*, 635–660

G.P Zhang and M Qi 2005. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research. 160,* 501-514.

Zou K, Kim K S, Kim K *et al,*. 2020.Genetic Diversity and Genome-Wide Association Study of Seed Aspect Ratio Using a High-Density SNP Array in Peanut. *Genes, 12*, 2.

J Hwang and E Little 1996. Real time recurrent neural networks for time series prediction and confidence estimation. IEEE *International Conference on Neural Networks. 4*, 1889-1894.

Sajindra H, Abekoon T, Wimalasiri E M, Mehta D and Rathnayake U2023. Artificial Neural Network for Predicting Groundnut Yield Using Climatic Data. *AgriEngineering*, *5*, 1713-1736. https://doi.org/10.3390/agriengineering5040106

Kumari P, Goswami V, Harshith N *et al,*.2023."Recurrent neural network architecture for forecasting banana prices in Gujarat, India". *International Journal of Information Technology, 16*, 1993–2002.

Deepalakshmi R, Padma Devi S, ShanthalakshmiRevathy J *et al,*. 2019. Scheming a new algorithm for dynamic price prediction of vegetable commodities using statistical price prediction technique. *International Journal of System Assurance Engineering and Management, 14*, 1869–1879.

Received on 17.01.2025 and Accepted on 05.03.2025