



# An analysis the Enhancing agricultural Decision-making with a Hybrid Crop Selection Algorithm (CSA)

Rohit Kumar Raghuwanshi <sup>1 \*</sup>, Ravindra Kumar Tiwari <sup>2</sup>

1. Research Scholar, Department of Computer Application, LNCT University, Bhopal, Madhya Pradesh, India

rohitmca70@gmail.com ,

2. Associate Professor, Department of Computer Application, LNCT University, Bhopal, Madhya Pradesh, India

**Abstract:** Agricultural productivity is highly dependent on the selection of suitable crops based on various environmental, soil, and economic factors. Traditional crop selection methods often lack adaptability and fail to integrate multi-dimensional data for optimized decision-making. This study suggests an algorithm for crop selection. This method uses a number of models to estimate both the level of groundwater & prices of agricultural crops. the strengths of both the linear ARIMA method & nonlinear ANN model, a hybrid model is clearly required to assess the accuracy of price forecasts. We attempted to develop a hybrid technique using multi-classifiers, as single-classifiers were previously the exclusive tool for prediction & forecasting in the agricultural domain.

**Keywords:** Agriculture, Hybrid Technique, ANN, Crop Selection, Groundwater

----- X -----

## INTRODUCTION

Agriculture plays a critical role in global food security, economic stability, and sustainable development. However, farmers often face challenges in selecting the most suitable crops due to varying climatic conditions, soil properties, water availability, and market demand. Traditional crop selection methods rely on empirical knowledge and generalized guidelines, which may not be optimal for maximizing yield and profitability. The need for data-driven, intelligent decision-making in agriculture has led to the integration of advanced computational models into farming practices.

This study suggests an algorithm for crop selection. This method uses a number of models to estimate both the level of groundwater & prices of agricultural crops; with this information, one can choose which crops to plant at which water levels in order to achieve a certain percentage of profit. Predicting agricultural prices has received less academic attention. The accuracy of the predictions is the metric by which the model is measured.

In order to accurately estimate future commodity prices, it is more important to identify when a turning point will occur. There are typically both linear & nonlinear trends in the data on agricultural prices. That is why it is impossible for a single model to account for every possible feature of agricultural price time series data. Taking advantage of the strengths of both the linear ARIMA method & nonlinear ANN model, a hybrid model is clearly required to assess the accuracy of price forecasts. In order to make informed decisions at every level, agricultural price information is essential. Because of the interconnected nature of the world's markets, the necessity for accurate forecasts is growing. In order to provide farmers, traders, or

policymakers with reliable and current price forecasts that take into consideration local information, it is necessary to design a market intelligent system that integrates traditional techniques like neural networks and fuzzy logic. This will allow them to make well-informed decisions about production, marketing, or policy in advance. Given specific regional circumstances, the decision-support system ought to tailor recommendations to specific farmers. We developed a multi-model for agricultural price forecasting due to the aforementioned factors. We attempted to develop a hybrid technique using multi-classifiers, as single-classifiers were previously the exclusive tool for prediction & forecasting in the agricultural domain. In order to refine the crop selection procedure, two models are utilised:

I- Arima Model and II- Artificial Neural Network.

## **PROPOSED ARCHITECTURE**

The authors provide a crop selection technique that combines Arima with an ANN-based model to predict agricultural time-dependent characteristics. When it comes to accurate & versatile time series forecasting, few models can compare to Arima. Arima model does not capture nonlinear patterns, which is a severe shortcoming [G. E. P. Box 1970]. Alternatively, nonlinear patterns in time series can be captured by an ANN model due to its flexibility. There are both linear and nonlinear trends in the time series data from agriculture. Consequently, a multi-parameter combination method is adopted for crop selection. This study makes use of secondary data collected from [www.tnau.com](http://www.tnau.com). The market price prediction takes into account the pricing data of wheat, rice, maize, jowar, & bajra from two seasons. Since each crop is evaluated across two seasons, the value of  $N=10$ .

Dataset characteristics & forecasting goals of time series data dictate the model to be used for time series data forecasting. To make predictions and categorise data, ANN does not employ any particular model techniques. ANN uses an adaptive method to acquire new knowledge. It can be trained or modelled based on the interest in forecasting. The suggested hybrid approach makes use of ANN, which has this benefit. The proposed model is trained using agricultural time series data in order to forecast both groundwater levels and market prices. One of the main algorithms suggested in this study, the Crop Selection Algorithm (CSA), is detailed below.

### **Crop Selection Algorithm (CSA)**

There are four steps to the algorithm.

- 1) Ground water level Prediction
- 2) O/I Transformation
- 3) Market Price Prediction
- 4) Decision making

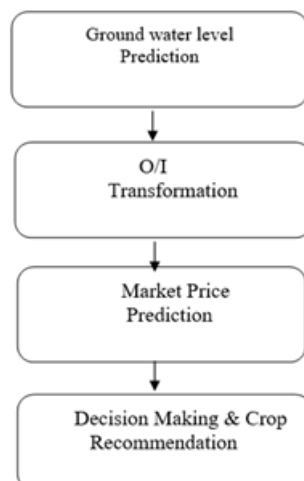


Figure 1 Crop selection Process

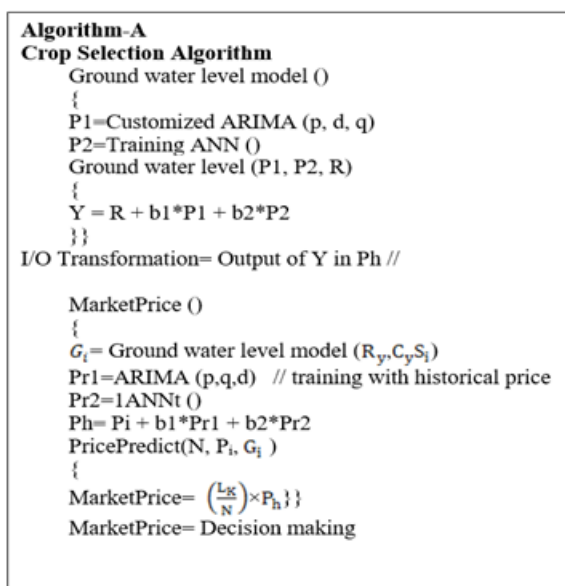
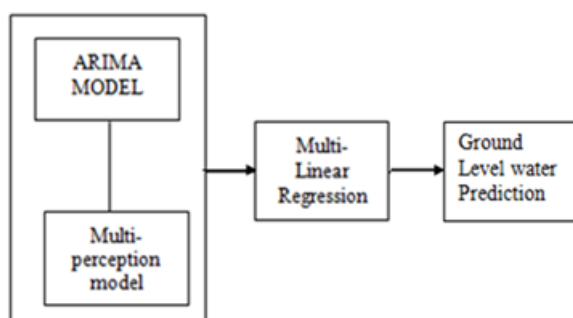


Figure 2 Procedure for Crop Selection Algorithm (A)

## DESCRIPTION OF GROUND LEVEL WATER MODULE

Predicting the amount of groundwater is an essential part of our solution. In contrast to the conventional wisdom, which relies on a single model to forecast groundwater levels, our approach uses a multi-model with a number of parameters, including historical rainfall data, soil type, and temperature.



### Figure 3 Groundwater Level Prediction Model: A Proposed Approach

#### GATHERING DATA AND KEY INDICATORS

The Market Price Module takes rainfall, temperature, and soil type as input variables and produces ground level water as an output.

**Table 1: Output parameter values**

Rainfall data mm	Temp. Degree Celsius	Soil type	Ground water level
10	>35	1	<30
20	>35	2	<30
23	27 to 35	1	<30
34	27 to 35	2	<30
44	>35	1	30 to 60
12	>35	2	30 to 60
18	27 to 35	1	30 to 60
24	27 to 35	2	30 to 60
33	>35	1	>60
31	>35	2	>60
56	27 to 35	1	>60

#### PROPOSED MODEL-1 FOR FORECASTING

Following is a description of the Autoregressive Moving Average model that was used to forecast the value (P<sub>1</sub>).

Rainfall in year= $R_y$

Climate changes in year=  $C_y$

Soil type= $S_i$

ARIMA (p, d, q)

$$P_1 = (\phi(B) (1 - B)^d R_y) - (S_i + C_y)$$

Where  $\epsilon_t$  = white error depend soil and climate.

$C_y$  is seasonal time series data.

Then  $C_y$  in SARIMA (p, d, q) × (P, D, Q)<sub>s</sub>

$$\phi_p(L) \delta_p(L^s) \nabla^d \nabla_s^D C_y = \theta_q(L) \delta_Q(L^s) \epsilon_t$$

$\phi_p(L)$  and  $\theta_q(L)$  are polynomial in L of degree p and q respectively

$\delta_p(L^s)$  and  $\delta_Q(L^s)$  are polynomial in  $L^s$  of degrees P and Q

p is the order of the autoregressive part

d is the order of the differencing

q is the order of the moving-average process

In order to create the Ground water level module, the multi-regression equation will make use of the Predicted P1.

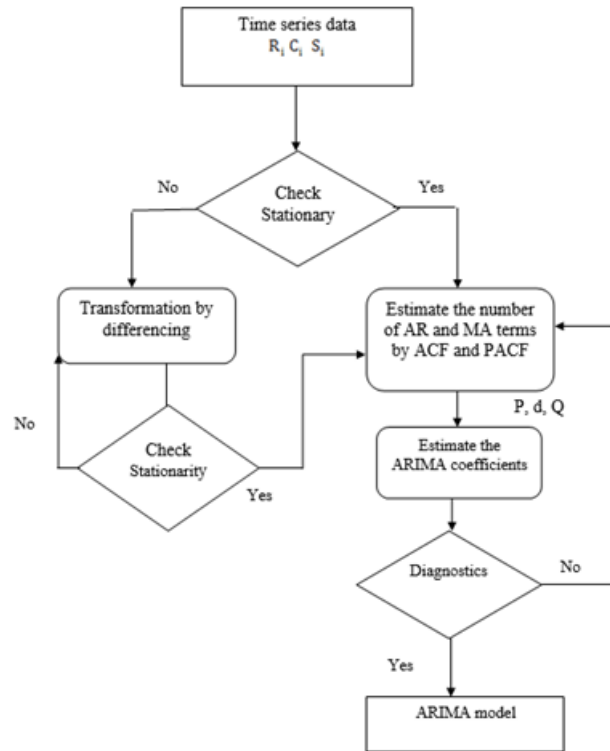


Figure 4 Methodology for finding the parameters of an ARIMA model

## Ground water level determination using a multiparameter algorithm -1

### Description of Algorithm-1

In order to predict future result values from historical data observations, the ARIMA model is adjusted. Here, the method uses a number of input data factors, including soil (Si), rainfall (Ry), & climatic conditions (Cy), to calculate the groundwater level. The raw database is queried using the aforementioned parameters to retrieve the data. The normalisation technique is used to preprocess the data.

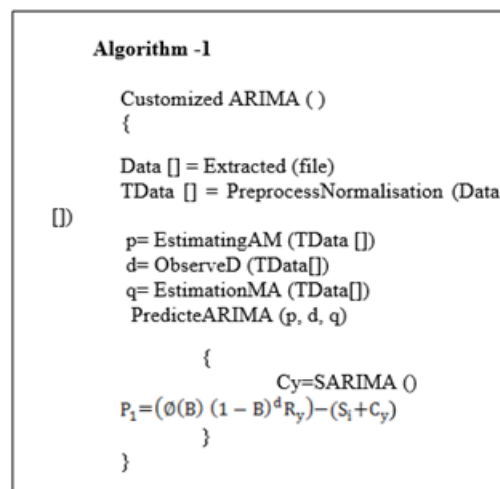


Figure 5 Algorithm-1 for Predicting Groundwater Levels

Although there are numerous stages to preprocessing, the most important ones are data extraction,

transformation, and loading. Improving the efficiency and performance of data mining algorithms, making the data more precisely and accurately represented in a way that humans and machines can understand, facilitating faster data retrieval from databases, and preparing the data for a particular analysis are all benefits of data transformation.

Each of the three dataset parameters—Soil (Si), Rainfall (Ry), and Climatic conditions (Cy)—is subjected to the normalisation approach.

In order to ensure that the data is stationary, the Mean-Max normalisation is done to the database. One normalisation approach is minmax normalisation, which uses the minimum and maximum values in X and Y, respectively, as its parameters. Thus, X is transformed such that its lowest value is mapped to 0 and its highest value is mapped to 1. Therefore, the whole range of X values, from minimum to maximum, are transformed into the integers 0 to 1. Y also has two possible values, 0 and 1.

$$X' = (X_{\max} - X_{\min}) * ((X_i - X_{\min}) / (X_{\max} - X_{\min})) + X_{\min}$$

By using min-max normalisation, the original data values' relationships are maintained.

Determine the mean difference (d) in the dataset. When two or more observations are taken at the same time, a difference is determined. The models are commonly known as ARMA(p,q) models when no differencing is done (d = 0). In order to stabilise a time series' mean, differencing can be used to remove trend and seasonality by reducing fluctuations in the level of the time series.

$$\text{if } d=0: y_t = Y_t$$

$$\text{If } d=1: y_t = Y_t - Y_{t-1} \text{ (differencing in the interval of 1)}$$

$$\text{If } d=2: y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} \text{ (differencing in the interval of 2)}$$

The amount of stationarity in the dataset can be shown by D, which is nature dependent.  $y_t$  is an observation or variable that is measured at a certain moment in time.

P and q, the parameters of the auto-regressive model, should be estimated. The weather is very variable and changes throughout the year. Before calculating the profit (P1), the ARIMA formula is applied to the data set in order to forecast the weather. The multi-regression model makes use of this parameter as well. The model that was built is used to forecast the value of ground level water for the coming year.

## **PROPOSED MODEL-2 FOR FORECASTING**

The created Artificial Neural Model predicts the value (P<sub>2</sub>). When it comes to nonlinear time series data, ANN is the ideal model for predicting. The nonlinear continuous function can be estimated to a desired degree of precision using this approach. When it comes to time series data forecasting, ANN models with a single hidden layer often only have one output. The following is the formulation of a  $p \times q \times 1$  ANN:

$$y_t = \phi_0 + \sum_{j=1}^q \phi_j g \left( \theta_{0j} + \sum_{i=1}^p \theta_{ij} y_{t-i} \right) + \varepsilon_t$$

Where  $\phi_j, (j = 1, 2, \dots, p)$

$\theta_{ij}, (i = 0, 1, 2, \dots, p) \quad j = (0, 1, 2, \dots, q)$  are weight,

$\phi_0, \theta_{0j}$  are bias,  $\varepsilon_t$  is white noise,  $g$  is activation function.

For linear problems, ANN yields mixed results.

The model yields the value of P2, which is subsequently input into the linear regression equation.

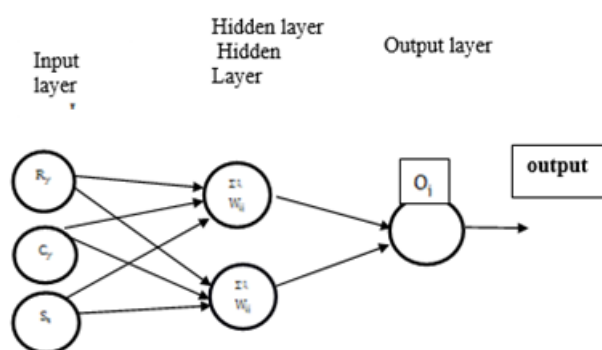


Figure 6 ANN training model with many layers of perceptrons

Proposed Algorithm-2 to determine the Ground water level with the help of multiparameter



**Algorithm- 2**

1. Calculate the summation impute and biased for each connected state from impute.

$$I_j = \sum_i W_{ij} O_i + \theta_j$$

2. The sigmoid activation function for generate the output.

$$O_j = \frac{1}{1 + e^{-I_j}}$$

3. Calculation of error in output. If output equal to target values then terminates the training, otherwise follow step 4

$$\text{Err}_j = O_j (1 - O_j) (T_j - O_j)$$

$$\text{//Error for hidden layer} \quad \text{Err}_j = O_j (1 - O_j) \sum_k \text{Err}_k w_{jk}$$

4. Update the weight and bias corresponding to impute state and hidden layer.

$$\Delta W_{ij} = W_{ij} + (l) \text{Err}_j O_j$$

$$W_{ij} = W_{ij} + \Delta W_{ij}$$

$$\Delta \theta_j = \theta_j + (l) \text{Err}_j$$

$$\theta_j = \theta_j + \Delta \theta_j$$

5. Repeat 1-4 till final target values has been predicated

6. Finally find one target values P2 for further operation multi regression.

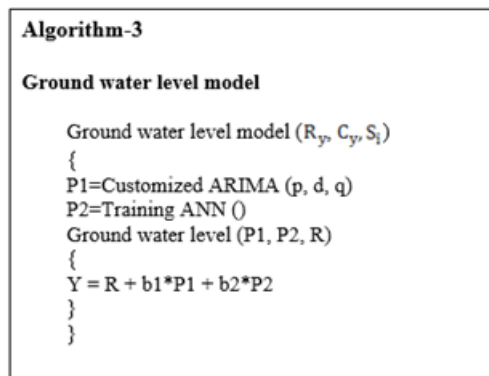
**Figure 7 Groundwater Level Prediction Algorithm-2**

**Explanation of Algorithm**

In this case, we train the data to anticipate non-linear patterns in the dataset using the ANN method. One hidden layer is present in the suggested ANN model. To begin training the ANN model, the following parameters are defined: learning rate (e.g., 0.9), bias (e.g., 0.2), impute (e.g., R\_y, C\_y, S\_i), and weight (e.g., 0.1, 0.2, 0.1, 0.3, 0.4, 0.5).

By adding biased weight on matching states to the weight by mutilation output, the hidden impute values are computed. Moreover, we calculate the activation function. The target values are used to calculate the error. Step 2 continues with updating the weight and bias, and the process ends when the output is equal to the target values. Otherwise, the process continues until the target values (p2) are anticipated. When it comes to nonlinear time series data, ANN is the ideal model for predicting. The nonlinear continuous function can be estimated to a desired degree of precision using this approach. Time series forecasting is often done using ANN models with a single hidden layer, which typically have one output.

**Algorithm-3 Proposed for Using Multi-Parameters to Find the Ground Water Level**



**Figure 8 Algorithm-3 for Predicting Groundwater Levels**

For example, if we have two models that predict values for a future time  $T$ —the ARIMA model with  $P1$  and the multi-Perceptron model with  $P2$ —we can model the ultimate ground water level prediction,  $Y$ , as

$$Y = R + b1 * P1 + b2 * P2$$

Each model's coefficients, denoted as  $b1$  and  $b2$ , are determined using the multi-linear regression technique, with  $R$  representing the starting water level.

### Explanation of the Algorithm

In order to forecast the values of the  $P1$  and  $P2$  variables, respectively, the aforementioned method invokes the ARIMA and ANN models. The  $P1$  model uses the customised Arima Model to make predictions, while the  $P2$  model makes use of the customised Artificial Neural Network Model to make predictions. After that, to forecast the ground level water for  $N+1$  years, the results from the customised models are integrated using a multilinear regression equation. By fitting a linear equation to the observed data, multiple linear regression seeks to model the relationship between a response variable and two or more explanatory variables. The value of  $y$  is proportional to the value of  $x$ , the independent variable. Nevertheless, by determining the model's parameters for a set of observed data, one can estimate the regression model.

### CONCLUSION

The integration of data-driven technologies in agriculture is crucial for improving decision-making and optimizing crop selection. Predicting the turning point was of utmost importance when attempting to anticipate the price of any commodity. Linear and nonlinear patterns were frequently found in agricultural price data. Accordingly, there was no single model that could extract all the features from agricultural price time series data. Taking advantage of the strengths of both the linear ARIMA method and the nonlinear ANN model, a hybrid model was required to assess the accuracy of price forecasts. To make informed decisions at any level, accurate agricultural price data was necessary. Agricultural prediction and forecasting had always relied on single classifiers, but this study attempted to develop a hybrid technique using multi-classifiers. In order to refine the crop selection procedure, two models were utilized: the ARIMA model and ANN. Future research can explore the incorporation of remote sensing data, deep learning models, and real-time IoT-based monitoring to further enhance the accuracy and scalability of the CSA framework.

## References

1. A. Choudhury and J. Jones, "Crop yield prediction using time series models", Journal of Economics and Economic Education Research, Vol. 15, pp. 53-68, 2014
2. Bhuyar, Vrushali, "Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District", Volume 3, Issue 2, March – April 2014, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) pp 200-204
3. Bhuyar, Vrushali, "Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District", Volume 3, Issue 2, March – April 2014, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) pp 200-204
4. DakshayiniPatil, Dr. M. SShirdhonkar, "Rice Crop Yield Prediction using Data Mining Techniques: An Overview", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 5, ISSN: 2277 128X, 2017.
5. Gupta, A. K., and B. D. Mazumdar, "Computational Model for Agricultural Decision Support System", International Journal of Computer Applications, Volume 66– No.15 Page No: 1-6, 2013
6. H. Ansari, "Forecasting Seasonal And Annual Rainfall Based on Nonlinear Modeling with Gamma Test In North of Iran", International Journal of Engineering Practical Research, Vol. 2, pp. 16-29, 2013
7. Jharna Majumdar et. Al. (2017) "Analysis of agriculture data using data mining techniques: application of big data", Journal of Big Data 4 1-15.
8. Khan, M., Islam, M. Z., Hafeez, M., "Irrigation Water Demand Forecasting – A Data Pre-Processing and Data Mining Approach based on Spatio-Temporal Data", In Proceedings of 9th Australasian Data Mining Conference (AusDml1). Ballarat, Australia., CRPIT, 121, pp. 183-194, 2011.
9. M. Ikonya, P. Mwita and A. Wanjoya, "Modeling Export Price of Tea In Kenya: Comparison of ANN and Seasonal Autoregressive Integrated Moving Average", American Journal of Theoretical and Applied Statistics, Vol. 3, pp. 211-216, 2014
10. N. G. Yethiraj, (2012) "Applying Data Mining Techniques in the field of agriculture and allied sciences," International Journal of Business Intelligents, vol. 1, no. 2,
11. S. Bangaru Kamatchi, et. Al. "Improvement of Crop Production Using Recommender System by Weather Forecasts" international conference on recent trends in advanced computing 2019, ictac 2019
12. Zaman, M.S., "Optimization of Crop Water and irrigation requirements under various management condition using Decision support tool", Lecture notes delivered on 16-21 January, 2006 at WALMI, Anand