IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A DATA-DRIVEN FRAMEWORK FOR FORECASTING AGRICULTURAL PRODUCTION

[1] Mrs. D. Poongodi, [2] Dr. K. Haridas

[1]Research Scholar, Department of Computer Science, NGM College, Pollachi.

[2]Associate professor and Head, Department of Computer Applications,

NGM College, Pollachi

Abstract: Agriculture is recognized as a vital branch of the Indian economy, with a significant percentage of the rural population depending on agricultural work for their livelihood. Crop yield prediction is a crucial task in this sector, defined as the forecasting of harvest in kilograms per hectare. Crop yield is estimated based on various factors, including environmental parameters like precipitation, temperature, and evapotranspiration, as well as historic crop cultivation data, fertilizer use, and pesticide use. Due to increasing environmental uncertainty, accurate analysis of historic data is essential for successful farming. Data mining techniques are extensively employed in agriculture to discover knowledge and patterns from large datasets. The primary goal of using data mining in agriculture is to increase the yield, maximize benefits, and decrease the cost of cultivation. Data mining applications include classification, prediction, and clustering, which are used in crop selection systems, fertilizer recommendations, soil data analysis, and crop yield prediction. Prediction of crop yield in advance aids farmers in maximizing production and assists government bodies and stakeholders in planning for procurement, distribution, buffer-stocking, import/export, price fixation, and marketing of commodities.

Keywords: Environmental Factors, Classification, Prediction, Clustering, Soil Data Analysis, Fertilizer Recommendation, Agricultural Decision Support, Precision Farming, Historic Crop Data, Farming Optimization, Agricultural Planning

I. INTRODUCTION

Agriculture is recognized as a vital branch and the backbone of the Indian economy. Farming constitutes an important part of agriculture, and over 58% of the household rural population depends on agricultural work to survive their lives. In India, the agricultural sector contributes approximately 17% to the GDP and is the primary source of food for the world's population.

In agricultural planning, predicting crop yields is one of the more difficult tasks. Crop yield prediction is defined as the forecasting of the harvest in Kilogram per Hectare. Yield depends on several factors, including environmental parameters such as precipitation, temperature, and evapotranspiration, as well as historic crop cultivation data, fertilizer use, and pesticide use. Traditionally, farmers cultivate crops based on previous experience. However, due to increased uncertainty in the environment, and the necessity to feed an expanding global population projected to reach 9.8 billion in 2050, accurate analysis of historic data is crucial for successful farming.

Recent advancements in information technology for agriculture have made estimating agricultural productivity a fascinating research area. Data Mining is the process of discovering knowledge from large datasets. It is a method of extraction, transformation, loading, and predicting meaningful information from huge data to extract patterns and convert it into an understandable structure for further use. The primary purpose of using data mining in agriculture is to increase the yield to get maximum benefit and to decrease the cost of crop cultivation. Data mining techniques, including classification, prediction, and clustering, are widely applied to agricultural problems such as crop selection systems, fertilizer recommendations, soil data analysis, pest and disease prediction, and crop yield prediction.

Prediction of crop yield in advance is a major agricultural concern and is highly useful for various stakeholders. Advanced prediction aids farmers in proactive planning to cut losses and improve yields. It also assists government bodies and policy makers in critical planning related to procurement, distribution, buffer-stocking, import/export, price fixation, and marketing of agricultural commodities. The decision support systems (DSS) implemented for farmers commonly utilize algorithms such as Artificial Neural Networks (ANN), Bayesian networks, and Support Vector Systems.

Given that the performance of predictive algorithms is data-dependent, comparative studies are necessary to identify the most effective techniques. Therefore, research frequently compares the accuracy of different data mining algorithms, such as Naive Bayes, K-nearest neighbor (KNN), Decision Tree, Random Forest, and K-Means clustering, for predicting agricultural crop yields.

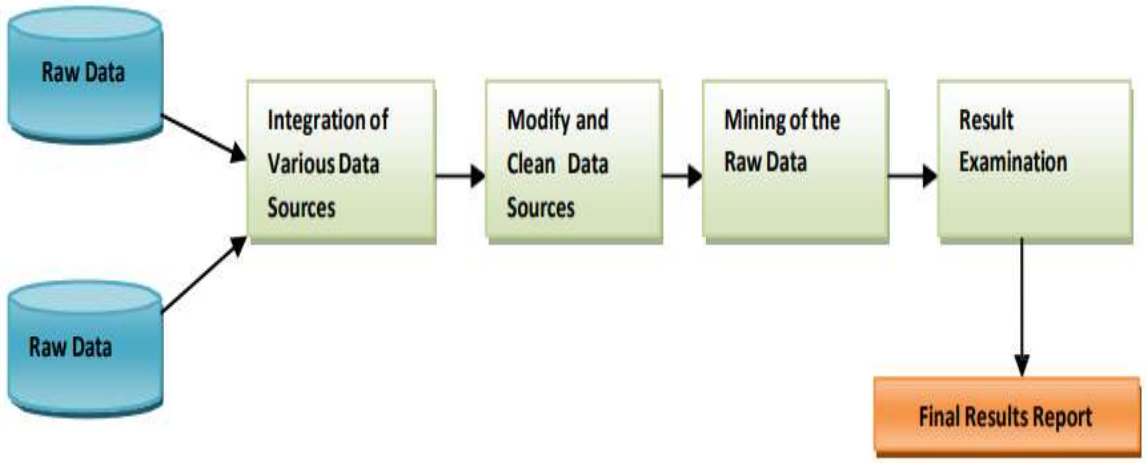

Fig.1 Data Mining Process Diagram

2. Related Work

Various researchers in the past explain the different kinds of data mining techniques in agriculture sector. With the less human interventions, they solve the complex agriculture problems.

AakunuriManjula, Dr. G. Narsimha [3] the study presents the comparison of classification methods like KNN, Bayesian Network, and Decision Tree. Monali Paul, Santhosh K. Vishwakarma, Ashok Verma [4] The process of predicting the crop yield using Data Mining approach uses a data mining approach through which the analyzed soil datasets are predicted. The disadvantage is that the system does not consider the demand existing in the agricultural economy. In our system the crops are suggested to farmers based on the demand by the market prices.

Tng Zhang [5] Crop yield estimation using classification techniques estimates the crop yield and selects the most suitable crop for cultivation using data mining techniques thereby improving the value and gain of farming area. The disadvantages in this system are that the methods to meet the demand and conveying the suggestions to farmers are not feasible. Rossana MC [6] Crop prediction model framework was developed and concluded that climate related variables were not the main determinants of corn yield, rather yield was greatly affected by planting practices, particularly by the application right amount of fertilization.

R. Kalpana [7] presents an overview on information mining techniques in agribusiness. Data processing in Agriculture is connected to raise the analysis field. Application of information mining techniques plays an essential role in agricultural and environmental with the connected areas. In agriculture, some issues like yield estimation and crop productivity remains to be solved with the help of available knowledge. This survey aims to search out suitable data processing models to understand high accuracy and prediction capabilities. It's the opinion that another techniques and algorithms to be studied connected agricultural issues can provide smart lead to agricultural growth. Finally, mistreatment data processing techniques in agriculture could be an up to date technique to search out the answer over the normal and traditional technique.

Powdery Mildew of Mango a devastating disease of mango was predicted by using Decision Tree induction, Rough Sets (RS) and Hybridized Rough Set based Decision Tree Induction (RDT) [12]. The induction algorithms shown better performance over logistic regression. U. Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin [8] In their study analyzed crop yield prediction by using Support Vector Machine (SVM), Multiple Linear Regression (MLR), Ada boost and Modified Non Linear Regression.

Snehal S.Dahikar, et al [14] proposed a crop prediction methodology based on parameters soil, PH, nitrogen, phosphate, potassium, organic carbon etc using the artificial Neural Networks (ANN). In artificial neural network (ANN), basically defined in terms of three parameters i.e. interconnection pattern between different layers of neurons, learning process for updating the weights of the interconnections, the activation function that converts in neurons weighted input to its output activation. In this paper most common neural network architecture feed forward back propagation network is used. The idea of back propagation algorithm is to reduce the error until the ANN learns the training data. Back propagation used the gradient descent method in order to calculate the derivative of squared error function with respect to weights in the network.

The squared error function is as below:

$$E=\frac{1}{2}(t-y)^2$$

E=Squareerror
t=Targetoutput
y=Actualoutputofoutputneuron ——— (1)

$$y=\sum_{i=1}^{n} Wixi$$

n=Numberofinputunittotheneuron
w_i=thei^{th}weight
x_i=thei^{th}inputvaluetotheneuron ———(2)

These equations results true output for a neuron with linear activation function. To get correct outputs a non-linear differential activation function is used, as described below:

$$y=\Phi(net) \qquad net=\sum_{i=1}^{n} Wixi \text{———(3)}$$

These equations will act as foundation for calculating the partial derivative of the error w.r.t a weight wi using chain rule. Further, the author used this ANN for crop yield prediction at rural district based upon above parameters.

Shweta Srivastava et al [15] have designed a model for the prediction of crop where the concept of fitness function is introduced with the variables such as weather and soil using genetic algorithm. Database of all crops was obtained and after collection of data fitness function is defined with the variables defining crop, weather and soil. The fitness function is defined as below:

$$z=y.xa(\Phi + I/0.5)$$

wherez=crop
y=weather
I=typeofsoil ————(4)

Various techniques that are cross over, mutation; reproduction is applied on fitness function for finding optimal solution with the help of MATLAB. In this method geographic information system for the analysis of equation from which information about the area was taken. This information works as input for GA.

Shruti Mishra et al [16] describe the crop yield prediction system by using data mining techniques by doing analysis on agriculture dataset. In the data set three major factors are used that is total production for each crop, area of cultivation, seasons. For the analysis of dataset WEKA tool is used using different classifier algorithms namely J48, LAD Tree, LWL & IBK. Further the classifiers are compared with the values of Root Mean Squared Error, Mean Absolute Error and Relative Absolute Error. Lesser the value of error more accurate the algorithm will work. To evaluate the performance accuracy, the following factors are used: Specificity, Sensitivity, Accuracy, RMSE, MAE, RAE. The equation 5 is used to calculate the Specificity (TNR-True Negative Rate) and Sensitivity (TPR-True Positive Rate) is described as below:

TPR=TP/(TP+FN) TNR=TN/ (FP+TN)

Accuracycanbecalculatedas:(TP+TN)/(TP+FN+FP+TN) ———— (5)

TP=True positive (It will detect the condition which is really present)
TN=True negative (It will not detect the condition which is absent)
FP=False positive (It will detect the condition which is really absent)
FN=Falsenegative(Itwillnotdetectthe conditionwhichisreallypresent)

Different type of classifiers gives different results on the same datasets by using the equation no. (5) The percentage of accuracy is listed in Table 1.

Thegeneralcomparisonsofabovediscussedtechniquesareasbelowin Table1:

S.No.	Reference No.	Models/Algorithm		Accuracy (%)	Results
1.	[14]	Back Propagation Algorithm with Feed Forward Neural Network and Artificial Neural Network		Optimal	Regional soil parameter will be major factor to improve the CYP.
2.	[15]	Genetic Algorithm (Fitness Function)		Optimal	GA has improved/optimal crop yield prediction due to parallel computing in genetic algorithm as compare to
3.	[16]	Classification algorithms based rule based learning K-Neighbour etc.	J48	78.145	Techniques are compared here and IBK has highest accuracy among all.
			LAD Tree	66.225	
			LWL	62.251	
			IBK	80.794	

Table1:ComparisonofDifferentAlgorithms

I. DATAMININGTECHNIQUES

There are numerous data mining techniques available for developing the algorithms for formulations of various models to solve the various issues and generate the useful information and reports. To analyse the big amount of data, data mining came into picture and is also called as KDD process. For the completion of this process various techniques developed so far are explained in this section. N. Gandhi and L. Armstrong studied and analysed agriculture dataset for prediction of rice yield in humid subtropical climate zone and tropical wet and dry climate zone in India [10-11]. The pictorial representation of all the data mining techniques is explained in Fig.2

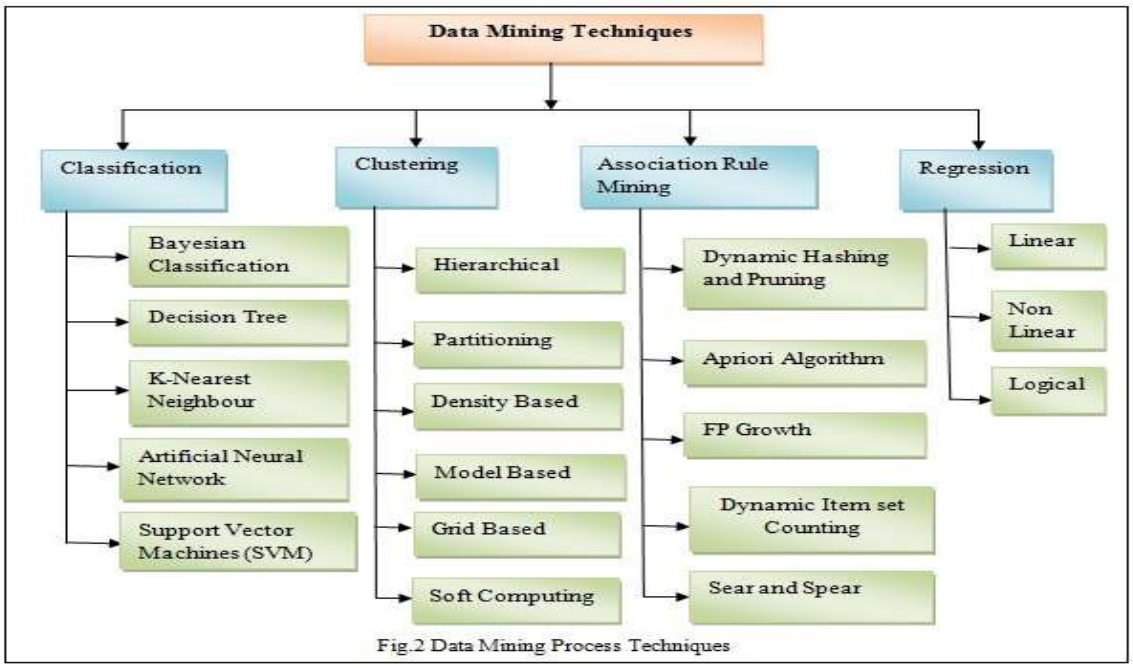

Fig.2 Data Mining Process Techniques

II. DATA MINING METHODOLOGIES IN CROP YIELD PREDICTION

For the agriculture issues data mining techniques are widely used. The purpose of the Data Mining process is to extract the information from a given large data set and convert it into understandable for further use. In this paper, we will analyse the crop yield production by making the use of various data mining techniques for the larger datasets to maximize the crop productivity. For the crop yield prediction, one can consider here the farmer's experience for specific crop or field over the. Nowadays, farmers do not only produce crops but also growing and growing large amounts of data and years. By making the use of GPS technologies useful information can be easily obtained with the help of sensors we can collect data. There are different data mining techniques that can be used for this purpose. Basic Data mining model in crop yield prediction can be described in Fig.3.

A. BASICMODELPROCESSFLOW

1. Identify the Crop to which Data Mining approach with machine leaning algorithms is to be implemented to get a robust CYP Model.
2. CYP for a specific crop based on data set obtained from district agriculture production database.
3. Evaluate different data mining techniques for developing appropriate DSS based on point no 2.
4. Generate various reports for crop yield prediction as visualization tool for farmers

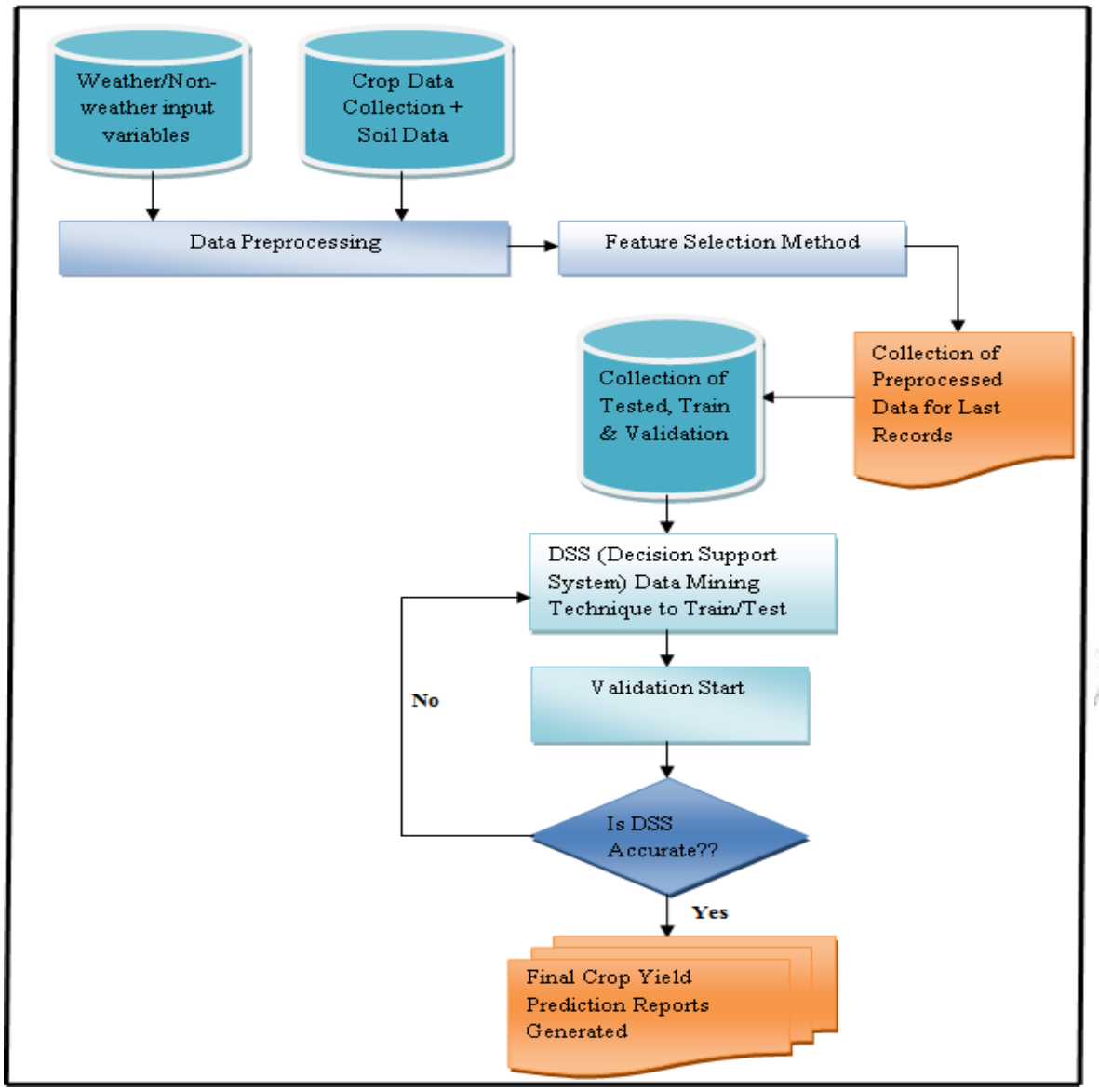

Fig.3CropYieldPrediction

III. RESULT AND CONCLUSION

Agriculture is the most important application area mainly in the developing countries like India. Data mining is used for large data in agriculture and extraction of knowledge is big challenge. The crop yield prediction is still remaining as a challenging issue for farmers, one can make the use of data mining techniques in agricultural field that will creates conditions for mankind satisfactory decisions and with that achieving challenging improvement. The aim of this research is to propose and implement a Decision Support System (DSS) to predict the crop yield prediction from the collection of past data. In this paper, comparison of various data mining & machine learning techniques are made for the smaller datasets and found that higher the accuracy higher will be rate for crop yield prediction. By making use of large datasets, one can improve the results. In future, one can do investigation to understand that how these techniques can

be used with large and complex agriculture datasets and can also be used for crop yield prediction with both seasonally, non-seasonal and spatially by making the use of GIS technologies

REFERENCES

[1] Yogesh Gandge, Sandhya, "A Study on Various Data Mining Techniques for Crop Yield Prediction", 2017 International Conference on Electrical Electronics, Communication, Computer and Optimization Techniques (ICEECCOT).

[2] Bhuvana, Dr.C.Yamini (2015), 'Survey on Classification Algorithms in Data mining.' International Conference on Recent Advances in Engineering Science and Management

[3] AakunuriManjula, Dr.G .Narsimha (2015), 'XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction', Conference on Intelligent Systems and Control (ISCO)

[4] Monali Paul, Santhosh K. Vishwakarma, Ashok Verma, "Prediction of crop yield using Data Mining Approach" Computational Intelligence and Communication Networks (CICN), International Conference 12-14 Dec. 2015.

[5] Tng Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms", Proceedings of the twenty-first international conference on Machine Learning.

[6] AakunuriManjula, G. Narsimha, "XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction", IEEE Sponsored 9th ISCO, 2015.

[7] Rossana MC, L. D. (2013). A Prediction Model Framework for Crop Yield Prediction. Asia Pacific Industrial Engineering and Management Society Conference Proceedings Cebu, Philippines, 185.

[8] R.Kalpana, N.Shanti and S.Arumugam, "A survey on data mining techniques in Agriculture", International Journal of advances in Computer Science and Technology, vol. 3, No. 8,426- 431, 2014.

[9] Umid Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin, "Rice yield prediction model using data mining", International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017, Cox's Bazar, Bangladesh.

[10] P Nancy Newton, Dr. Sapna Singh, "Data Mining in Decision Support System", Proceedings of National Conference on Emerging Trends: Innovations and Challenges in IT, 19 -20, April 2013.

[11] N.Gandhi and L.J. Armstrong, "Applying data mining techniques to predict yield of rice in Humid Subtropical Climatic Zone of India", Proceedings of the 10th INDIACom-2016, 3rd 2016 IEEE International Conference on Computing for Sustainable Global Development, New Delhi, India, 16th to 18th March 2016.

[12] N. Gandhi and L. Armstrong, "Rice Crop Yield forecasting of Tropical Wet and Dry climatic zone of India using data mining techniques", IEEE International Conference on Advances in Computer Applications (ICACA), pp. 357-363, 2016.

[13] Jain Rajni, Minz, S., V. Rama Subramaniam. 2009."Machine learning for forewarning crop diseases". J. Ind.Soc. Agri. Stat. 63(1): pp. 97-107.

[14] Sandeep V.Rode,Snehal S.Dahikar,"Agricultural Crop Yield Prediction using Artificial Neural Network Approach", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering Vol.2,Issue 1, January 2014.

[15] Shweta Srivastava, Diwakar Yagysen,"Implementaion of Genetic Algorithm for Agriculture System", International Journal of New Innovations in Engineering and Technology Volume 5 Issue 1- May 2016.

[16] Shruti Mishra, Priyanka Paygude,Snehal Chaudhary, Sonali Idate, "Use of data mining in crop yield prediction",2018 2nd International Conference on Inventive Systems and Control (ICISC).