

## A Review of Different Techniques Utilized for Crop Yield Prediction

Rabina Dayal<sup>1\*</sup>, Arun Kumar Yadav<sup>2</sup>

<sup>1</sup>Dept. of Computer Science and Engineering, ITM University, Gwalior, India

<sup>2</sup>Dept. of Computer Science and Engineering, ITM University, Gwalior, India

\*Corresponding Author: [rabinadayal013@gmail.com](mailto:rabinadayal013@gmail.com), Mob: +91-75667-96445

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 23/Dec/2018, Published: 31/Dec/2018

**Abstract**— In India, farmers are not getting expected crop yield from their productions. Crop production mostly depends on weather conditions and some statistical methodologies. To get higher crop production yield, farmers sometimes need advices for predicting and analyzing future crop production. This helps farmers to produce a crop with maximum yield. Such methods will be helpful for farmers and government to make a better decision to increase crop production. In this paper present a review on crop yield prediction (CYP) with different data mining (DM) techniques used to evaluate and predict the problem lead to increase CYP. The result analysis is performed on root mean square error (RMSE) and peak signal noise ratio (PSNR).

**Keywords**— RMSE, ANFIS, Data mining, Crop Yield, PSNR.

### I. INTRODUCTION

The contribution of Agriculture in India's economy is 15-17% of Gross Domestic Product (GDP). Farming as a business is interesting harvest creation is reliant on numerous atmosphere and economy factors [1]. The Agriculture sector in India mainly depend on the weather condition, the result of crop yield is reflection of weather. In Agriculture, huge amount of data is generated and this data is much necessary for different purpose. The traditional and predictable analysis of agriculture data purely depends on statics.

Therefore farmers necessarily require continuous advice to predict the production of future crop and analysis must be made which helps the farmers to maximize their crop production [2]. Investigation of how to improve the productivity and economic growth of agricultural production is one strategy which is based on with challenges of climate change and available agricultural land. Most of research and trials have been undertaken to analyze land, soils, and climate.

In agriculture, every farmer main problem is Yield Prediction. The Farmer's always want to know the production of Yield, how much Yield he will obtain from expectation. In past few years, the yield prediction is done using the analysis study of farmer experience over particular crop. Several factors influence the agriculture sector such as Pest, weather and harvest operation. Precise data about history of harvest yield is necessary for making decisions related to agricultural risk management (ARM).

Historical crop production information like crop size and related risks about crop helps these companies to take supply chain decision like scheduling the crop production [3].

DM is a field in Information Technology that deals with finding unknown and hidden patterns from the available data. Applying DM techniques in agricultural field to predict useful crop productivity related information is a noble work [4]. Based on the idea of information being mined there are two classes of function in data mining, Descriptive function that arrangements with general properties of information and Prediction work that distinguishes the patterns in view of accessible information. As far as agriculture is concerned predictive types that include classification, association, clustering and regression are used.

### II. TECHNIQUES OF DATA MINING

#### A. Multiple Linear Regression (MLR)

It is the technique used to show the linear relationship between a dependent variable and at least one free variable(s) [2]. Predicting is called as dependable variable and another one independent variables are known as predictors. It depends on slightest cubes and likely the most generally utilized strategy in climatology for producing systems to remake atmosphere factors from tree ring services.

#### B. K-Mean Method

In DM technique, it is mostly clustering approach. The process of this approach is simple in which certain data partition in K clusters and center of cluster calculate as the

mean of all neighbouring cluster [2]. The mean cluster can be referred as representative of the cluster. In the cluster, center is close to all samples.

#### C. Artificial Neural Network (ANN)

It is a novel method which is used in flood forecast. The benefit of this framework over the other framework is it can demonstrate the rainfall likewise it also predicts the attacks of pest for one week ahead of time [5]. DM devices are starting to indicate an incentive in examining huge informational indexes from confounded frameworks and giving amazing data (White and Frank, 2000). It can utilize yield history with estimated input factors for programmed learning and programmed age of a framework display. In the previous couple of years, a few yield reproduction models have been manufactured.

#### D. Bayesian Network

It is an intense apparatus for managing vulnerabilities and generally utilized as a part of horticulture datasets. It is the graphical system which is used to encrypt the probabilistic relation between variable when it is utilize with genuine method, this model has a couple of inclinations for data inspection. This method unequivocally manages vulnerability of information and connections, and can incorporate both subjective and quantitative variable. It encourages viable correspondence with partners, while advancing an emphasis on key factors and connections of the framework, as opposed to being impeded in points of interest. [6]

#### E. Support Vector Machine (SVM)

It utilized to characterize the given information in both of the 2 groups. It assembles a single hyperplane or an arrangement of hyperplanes and it chooses the hyperplane which gives the widest street between the closest purposes of the 2 groups. Its capacity is to boost the separation among the 2 groups with the goal that the blunder while grouping the given information lessens. In SVM, the streamlining just relies upon the result of the sets of test. It can likewise perform linear classification by kernel function.

### III. CYP MODEL

#### A. Selection of Attribute

It can be hard to find which dependable attribute is. Diverse product yield anticipating and demonstrating technique have been utilized as a part of the past with different victories. Famer needs to go up against the unmistakable issues in view of various factors which impact the orchestrating made by him early. These variables don't have the settled kind of effect, it differs time to time, and year to year relies upon the circumstance, climatic nature, increment in expenses of different requirements under indeterminate condition, equivocalness and ambiguity. Fuzzy Logic (FL) [8] demonstrating gives the definition of scientific displaying to discover the interface brings about indeterminate

circumstances. Factual models regularly don't consider qualities of the plants, the climate, or the dirt properties constraining their convenience. A few models depend on data from only a solitary year or area. At the point when a model is produced utilizing single area or year information, it will have restricted useful applications; in this manner fluctuation from different conditions must be incorporated [9].

#### B. CYP

The CYP includes all important features that are needed for efficient yield of item. This overhauls the game plan outcomes of the gather yield. All the essential parameters are considered as commitments to the model- Adaptive Neuro fuzzy inference system (ANFIS). With everything considered, one of the troubles looked in the want technique is that an immense section of the major parameters that are fundamental to consider for the right figure are not consider. It lessens the effectiveness of the predicted comes about which along these lines prompts nonappearance of fitting assessing of the gather yield. It is besides more mind boggling to predict the enhanced number of information parameters that are to be considered in the figure strategy.

#### C. ANFIS Models

It indicate is one of the capable ways which is used for conjecture, by compelling most of the key parameters as information sources, it upgrades the precision of desire happens which has the property of learning by ANN [10]. It gives some linearity respect to some of its parameters, from now on it assembles the overhead of figuring process without growing the profitability. It fails to streamline the fuzzy rules in it which spoils the execution of figure. Wellsprings of data that are to be considered and picked depend upon the heuristics. It also decreases the execution by ruining the efficiency of the desire procedure.

#### D. Statistical modeling:

This model [14] used MLR model of climate based yield gauging for oil palm. The stepwise MLR system gets contribution as month to month oil palm yields it go about as reliant variable using agro meteorological factors in cumulated time-slack period continuing to collect as the autonomous factors. The essential point of the creator is stressed over specifying of legitimate month to month yield suspecting model over a half year ahead by delineating quantitative association between time-slack meteorological elements and FFB yield of the young create oil palm for the underlying six social event years. Additionally, showing with limited meteorological data is studied as a lively approach for producer while lacking climate estimating instruments. Table 1 demonstrates the Model with critical indicator factors produced for month to month yield estimate (Y) utilizing a stepwise regression approach.

TABLE I. SHOWS SIGNIFICANT PREDICTOR VARIABLES USING MODELS

Duration	Multiple Regression Equation	M2
2005 to 2011	$Y = -1256 + 0.023X_1 + 0.007X_2 + 0.006X_3$	0.68

#### E. ANFIS- Intelligent System

The model exhibited ANFIS [12] for the forecast of wheat yield. This framework used vitality input esteems to deliver yield in light of fuzzy sets. At that point ANN demonstrates was made for the forecast of wheat yield and both the models were analysed as examined in Table.6. The ANFIS framework is the mix of both fuzzy framework and ANN.

TABLE II. SHOWS SUPPORT OF INPUT VARIABLES TO THE OUTPUT FOR WHEAT PRODUCTION

Sensitivity	Yield
N	0.09
P2O5	0.10
K2O	0.03
FYM	0.00
Labour	0.06
Diesel	0.06
Electricity	0.96
Seeds	0.01
Biocides	0.01
Water	0.11
Machinery	0.87

The making of fuzzy guidelines was utilized with appropriate participation capacity to anticipate the wheat yield all the more viably. Be that as it may, it gives some linearity regard to some of its parameters. The results demonstrated that when the amount of commitments for each ANFIS orchestrate lessened and in the meantime, the total number of ANFIS frameworks extended the better results was gotten. The best plan included five frameworks at the essential stage, two frameworks at the second stage and one framework at definitive stage.

#### IV. LITERATURE REVIEW

M. Gunasundari et al. [13] presented a CYP which takes a shot at a versatile group approach over powerfully updated authentic product informational collection to anticipate the harvest yield and enhance the basic leadership in accuracy agriculture. CYP utilizes apiary demonstrating way to deal with examine and characterize the harvest in light of product development design, yield. CYP grouped dataset had been tried utilizing Clementine over existing harvest area learning. The execution of CYP is better other group approaches, which bolsters expansive dataset.

Monali et al. [14] which uses DM strategies with a particular ultimate objective to suspect the characterization of the separated soil datasets. The class, accordingly foreseen will exhibit the yielding of harvests. The issues of foreseeing the

item edit yield is formalized as a portrayal represent, where Naive Bayes and K-Nearest Neighbor (KNN) procedures are used.

Niketa et al. [15] explored to predict rice CYP utilizing Neural Network (NN) in different region of Maharashtra, India. The parameters utilized as a part of examination were least, normal, most extreme temperature, precipitation, territory, reference crop creation and yield for the Kharif season (June to November) for the years 1998 to 2002. The dataset were handling WEKA instrument. The outcomes demonstrated the precision of 97.5% with an affectability of 96.3 and specificity of 98.1. Further, RMSE, relative absolute error and root relative squared error were figured.

Aakunuri et al. [16] proposed a novel framework named eXtensible Crop Yield Prediction Framework (XCYPF) is proposed that is flexible and extensible. It has arrangement for determination of product, reliant and autonomous factors, datasets for edit yield forecast towards exactness agribusiness. The accessible records are utilized alongside precipitation information and surface temperature for trim yield expectation for rice and sugarcane crops.

Papageorgiou et al. [17] presented a model CYP. It utilized cotton crop for tests with the help of fuzzy cognitive map based approach. The cotton CYP proposed by them turned into an essential fixing in choice emotionally supportive network that contributed towards accuracy agriculture. The created FCM show turned out to be proficient in cotton CYP.

Ashwani et al. [18] anticipate the CY and recommend the best product in this way enhance the quality and productivity of the agriculture segment by handling immense volume of information frequently called as Big Data utilizing Hadoop stage. In this, another Agro algorithm which is utilized to anticipate the reasonableness of a yield for a specific soil composes and upgrades the general nature of agrarian creation. This additionally encourages the ranchers to choose a specific harvest to sow contingent upon the climatic condition and gives fundamental data to pick the best climate to do quality cultivating.

S. Poongodi et al. [19] used agricultural related datasets to predict the crop yield in given area through feature extraction and feature selection process. C4.5 with ANFIS classifier is used to generate the rules and classifies the data. This proposed method is used to analysis the suitable area for cultivating a crop based on the data. The experimental results of proposed improved C4.5 with ANFIS classifier algorithm show better accuracy of 92.5%.

Sellam et al. [20] explain the different natural features are the Annual Rainfall (AR), Area under Cultivation (AUC), and Food Price Index (FPI) that impacts the CY and the

relationship among these parameters is built up. Utilizing Regression Analysis (RA), LR the different ecological elements and their punishment on trim yield is examined

Fathima et al. [21] utilizes DM methods on ongoing information that assistance in learning disclosure. They utilize k implies grouping calculation to bunch the ranchers in light of the harvest compose and water system parameters. Apriori calculation is utilized to figure out which two yields are chosen as a successive thing set. They by and large spotlight on the arrangements that administration could outline by the editing practices of agriculturists.

R. Kalpana et al [22] review aim to seek out proper information preparing models to acknowledge high exactness and expectation capacities. The supposition extra strategies and calculations to be examined associated farming issues can give shrewd prompt rural development. At last, abuse information handling systems in farming could be an up and coming strategy to seek out the appropriate response over the typical and customary procedure.

G.N. Fatima et al. [23] explored the utilization of different information digging strategies for learning disclosure in horticulture division. Learning disclosure in budgetary association have been worked to assess their operation and fundamentally to bolster basic leadership utilizing information as key component. Existing programming are wasteful in demonstrating such information attributes. Creators present distinctive shows for finding learning as affiliation principles, bunching, arrangement and relationship appropriate for information attributes. Proposed DM systems, the leader can characterize the extension of agribusiness exercises to enable the diverse strengths in existing farming division. In the agribusiness division, DM can help government to expand yield advantage principally to bolster basic leadership, solid and auspicious data on harvest region, edit creation and land utilize is of incredible significance to organizers and arrangement producers for effective agrarian improvement and for taking choices on obtainment, stockpiling, open dissemination, fare, import and numerous other related issues to contend in the distribute of product example.

TABLE III. COMPARISON OF DIFFERENT METHODS

Author& Year	DM methodologies	Applications
2001,K. Verheyen [25]	Fuzzy set	Yield Prediction in agriculture
2006,Veenadhari, S. [26]	KNN	Scale back procedure burden of KNN algorithm
2012,Sanjay D. Sawaitul [27]	NN	Focuses on weather forecasts

2013, V.R Thakare [28]	fuzzy system	Soil Classification
2015,Aditya Shastry [24]	FL, ANFIS	crop yield in agriculture
2016,Hemageethaa [29]	Naïve Bayes, Apriori algorithm	Market Basket Analysis
2016,Niketa Gandhi [30]	BayesNet	rice crop yield for Maharashtra
2017,Umid Kumar Dey [31]	Nonlinear Regression, SVM	forecast the yield of rice with astute analysis

V. PERFORMANCE MEASURE

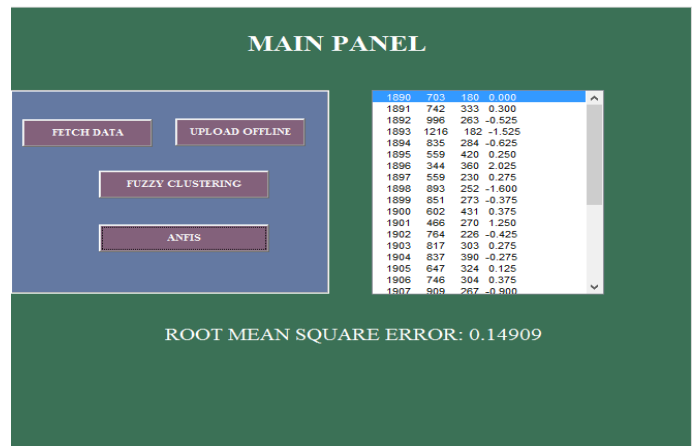


Figure 1. Shows GUI Panel

The fig1 demonstrates the GUI panel Aditya [25] in which GUI set up device is utilized. It comprises of (UI) controls and utilizing those UI controls the previously mentioned GUI is gotten utilizing pushbuttons and boards.

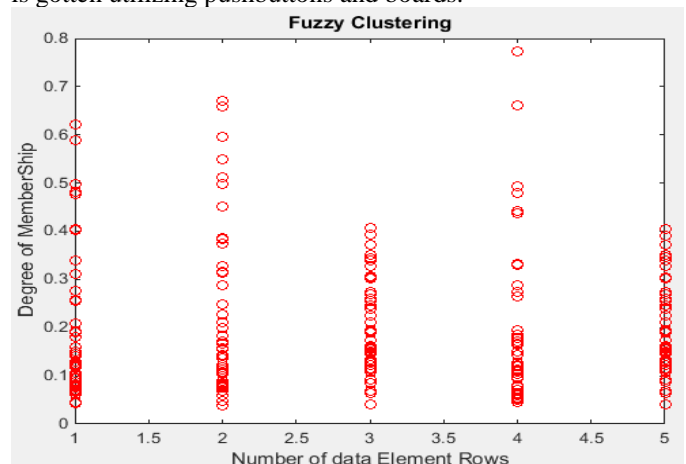


Figure 2. Shows Membership Values of Base System

In Fig2, it shows that the fuzzy clustering in which level of participation which is measured as for the information in the dataset for the yielding of wheat. This strategy is efficient which enables a bit of information to have a place with at least one bunches. The cluster point that is near the inside will have a place with high level of enrollment. It is a likelihood conveyance approach which works iteratively to give every cluster point a correct area.

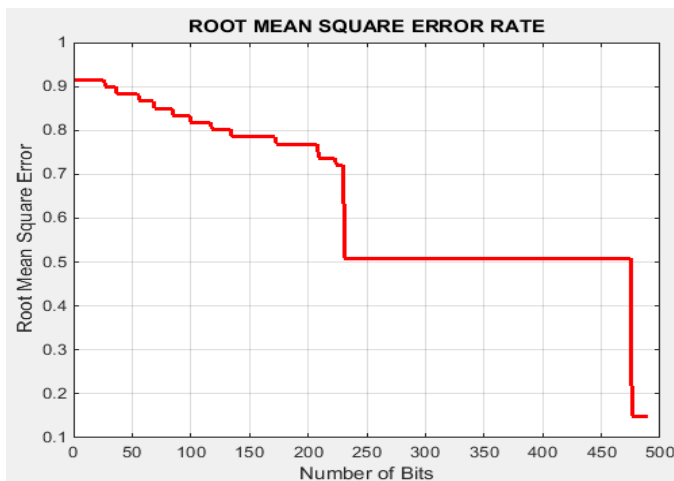


Figure 3. RMSE of Base System

In fig 3 shows that exhibit the learning RMSE rate in which the yielding of wheat batch is evaluated which must be low to have high rate of successfully conjecture of yielding. So the structure is having less blunder rates using ANFIS which is our base approach. The ANFIS deals with the arrangement set as the commitment to terms of level of interests. It amasses a FIS which manages enlistment work constraints which are tuned using back propagation system to perform low blunder rates.

**RMSE of prediction:** It is used to calculate the error rate between predicted and actual values.

$$RMSE_i = \sqrt{k^{-1} \sum_{z=1}^k (a_z - b_z)^2}$$

where  $k$  indicates no. of yrs  $z$ ,  $a_z$  and  $b_z$  represent actual and predicted values.

**PSNR of prediction:** It is used to calculate the improved feature of values. It shows higher then it is better approach. It measured in decibels (dB)

$$PSNR(x) = \frac{10 \times \log((255.^2)) / \log 10}{RMSE(x)}$$

## VI. CONCLUSION

In the recent years, Agriculture sector finds difficulty in improving the productivity of crop yield using natural resources. DM methods solves the problem related to crop selection, crop planning, irrigation planning, water resources management, vegetable production, water resource management. This paper surveys the technical achievements in the field of Agriculture CYP. Discusses methodology, comprehensive survey of various proposed methods to predict crop yield and applications. It also discusses various DM techniques used for prediction of crop yield for rice. Developing better procedures to predict crop profitability in different climatic conditions can help farmer and diverse accomplices in fundamental essential initiative to the extent agronomy and item choice page.

## REFERENCES

- [1] B Vishnu Vardhan and D. Ramesh, O Subhash Chander Goud "Density Based Clustering Technique on Crop Yield Prediction" International Journal of Electronics and Electrical Engineering Vol. 2, No. 1, March, 2014
- [2] Naushina Farheen Imam Shaikh, Prof. R. V. Argiddi "Annual Crop Yield Prediction and Recommend Planting of Different Crops by Using Data Mining Technique" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 10, October 2016
- [3] Chapman P. Gleason Large Area Yield Estimation/ Forecasting Using Plantprocess Models By Chapman P. Gleason For Presentation At The 1982 Winter Meeting Americansociety Of Agricultural Engineers Palmer House, Chicago, Illinois December R 14-17, 1982.
- [4] Jiawei Han, Micheline Kamber, Jian Pie, "Data Mining Concepts and Techniques", Morgan Kaufmann, ASIN B0058NBJ2M
- [5] Raorane A.A., Kulkarni R.V. "Data Mining: An effective tool for yield estimation in the agricultural sector" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 1, Issue 2, July – August 2012
- [6] Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils P.Bhargavi, Dr.S.Jyothi, IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.8, August 2009
- [7] SVM- Muller, K., et al. "An introduction to kernel-based learning algorithms." Neural Network IEEE transactions on 12.2 (2001): 181-201
- [8] Aliyu Muazu, Azmi Yahya, W.I.W. Ishak, S. Khairunniza-Bejo. Yield prediction modeling using data envelopment analysis methodology for direct seeding, wetland paddy cultivation. Agriculture and Agricultural Science Procedia. 2014; 2, 181-190.
- [9] K. Menaka, N.Yuvaraj "A survey on crop yield prediction models" Indian Journal of Innovations and Developments Vol 5(12), October, 2016.
- [10] Kadir, M. K. A., Ayob, M. Z., & Miniappan, N. (2014, August). Wheat yield prediction: Artificial neural network based approach. In Engineering Technology and Technopreneuship (ICE2T), 2014 4th International Conference on (pp. 161-165). IEEE.

- [11] Keong, Y. K., & Keng, W. M. (2012). Statistical modeling of weather-based yield forecasting for young mature oil palm. *APCBEE Procedia*, 4, 58-65.
- [12] Khoshnevisan, B., Rafiee, S., Omid, M., & Mousazadeh, H. (2014). Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs. *Information processing in agriculture*, 1(1), 14-22.
- [13] M. Gunasundari Ananthara, Dr. T. Arunkumar and Ms. R. Hemavathy "CRY – An improved Crop Yield Prediction model using Bee Hive Clustering Approach for Agricultural data sets" International Conference on Pattern Recognition, Informatics and Mobile Engineering, IEEE, 2013.
- [14] Monali Paul, Santosh K. Vishwakarma, Ashok Verma "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" International Conference on Computational Intelligence and Communication Networks, IEEE-2015.
- [15] Niketa Gandhi, Owaiz Petkar and Leisa J. Armstrong "Rice Crop Yield Prediction Using Artificial Neural Networks" International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2016), IEEE.
- [16] Aakunuri Manjula, Dr.G .Narsimha "XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction" 9th International Conference on Intelligent Systems and Control (ISCO) IEEE-2015
- [17] E.T. Papageorgioua,\* , A.T. Markinosb, T.A. Gemtosb " Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application. Elsevier-2011, PP 3643-3657.
- [18] Ashwani Kumar Kushwaha, SwetaBhattacharya "Crop yield prediction using Agro Algorithm in Hadoop" International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 5, No2, April 2015
- [19] S. Poongodi and M. Rajesh Babu "Prediction of Crop Production using Improved C4.5 with ANFIS Classifier" International Journal of Control Theory and Applications volume 10 , Number 21 , 2017.
- [20] Sellam,V., Poovammal, E., "Prediction of Crop Yield using Regression Analysis", *Indian Journal of Science and Technology*, Vol. 9(38), pp.1- 5, 2016.
- [21] Fathima, G.N., Geetha, R., "Agriculture Crop Pattern Using Data Mining Techniques", *International Journal of Advanced Research in Computer Science and Engineering*, Vol. 4, Issue 5, pp.781-786, 2014.
- [22] 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
- [23] G.N.Fatima,R.Geeta ,"Agriculture crop pattern using data mining techniques", *International Journal of Advanced Research in in Computer Science and Software Engineering*, vol. 4, No. 5,781-786 , 2014 .
- [24] Aditya Shastry , Sanjay H A and Madhura Hegde, "A Parameter based ANFIS Model for crop yield prediction", *IACC 2015 IEEE*,pp 253-257.
- [25] K. Verheyen, D. Adriaens, M. Hermy, and S. Deckers, "High resolution continuous soil classification using morphological soil profile descriptions", *Geoderma*, vol. 101, pp. 31-48, 2001.
- [26] Veenadhari, S. 2007, "Crop productivity mapping based on decision tree and Bayesian classification". Unpublished M.Tech Thesis submitted to Makhhanlal Chaturvedi National University of Journalism and Communication, Bhopal.
- [27] Sanjay D. Sawaitul, Prof. K.P. Wagh, Dr. P.N. Chatur, "Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach", *International Journal of Emerging Technology and Advanced Engineering*, Vol. 2, Issue 1, January 2012, pp. 110-113.
- [28] V.R.Thakare and H.M. Baradkar. "Fuzzy System for Maximum Yield from Crops", *International Journal of Applied Information Systems*, pp. 4-9 , March 2013.
- [29] Hemageetha, N., "A survey on application of data mining techniques to analyze the soil for agricultural purpose", 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp.3112-3117, 2016.
- [30] Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong," Predicting Rice Crop Yield Using Bayesian Networks", *ICACCI,IEEE 2016*,pp 795-799.
- [31] Umid Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin, "Rice Yield Prediction Model Using Data Mining", *ECCE, IEEE 2017*, pp 321-326.

### Authors Profile

*Ms. Rabina Dayal* has completed B. Tech. in Computer Science and Engineering. She is currently pursuing his M. Tech. in Computer Science and Engineering from ITM University Gwalior, Madhya Pradesh, India. Her research of interest includes cloud computing and Data Mining.



*Mr. Arun Kumar Yadav* has done B.E. (Computer Science & Engineering) from G.B. Pant Engineering College, Pauri Garhwal, M.Tech (Information Technology) from Sam Higginbottom Institute of Agriculture, Technology & Sciences, Allahabad and pursuing Ph.D. in Computer Science and Engineering from Uttarakhand Technical University, Dehradun. Presently he is working as Associate Professor in the Department of Computer Science & Engineering at ITM University Gwalior, Madhya Pradesh, India. His research interest includes Distributed Database Security, Cloud Computing and Fog Computing. He is a senior member of IACSIT and IAENG Technical Societies.

