International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 12 | Dec 2018 www.irjet.net p-ISSN: 2395-0072

ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUE TO PREDICT ANNUAL YIELD OF MAJOR CROPS

B. Devika[1], B. Ananthi[2]

[1]Research scholar, M.Phil. Computer Science, Vellalar College for Women, Erode12.
[2]Associate Professor and Head, Department of Computer Science, Vellalar College for Women, Tamilnadu, India.

Abstract - *India is generally an agricultural country. Agriculture is the single most important provider to the Indian economy. Agriculture crop production depends on the season, organic, and monetary cause. The prognostication of agricultural yield is challenging and pleasing task for every nation. Nowadays, Farmers are hostile to produce the yield because of erratic climatic changes and scarcity of water resource. The main objective is collecting agricultural data which can be stored and analyzed for useful crop yield forecasting. To predict the crop yield with the help of data mining technique, advanced methods can be introduced to predict crop yield and it also helps the farmer to choose the most suitable crop, thereby improving the value and gain of the farming area.*

***Key Words*: Data Mining, Classification, Crop Yield, Accuracy, K-Nearest Neighbor (KNN), Linear Regression**

1. INTRODUCTION

Data Mining is the process of extract helpful and significant information from huge sets of data. Data Mining in agriculture field is a comparatively novel research field. Yield prediction is a very important agricultural problem. Any farmer is interested in knowing how much yield he is concerning to be expecting. In the earlier period, yield prediction was performing by considering farmer's experience on particular field and crop. In any of Data Mining actions the training data is to be collected from past data and the gathered data is used in terms of training which has to be exploited to study how to categorize future yield predictions. Crop models and decision tools are more and more used in agricultural field to improve production efficiency. The combination of higher technology and agriculture to improve the production of crop yield is becoming more interesting newly. Due to the rapid development of new high technology, crop models and predictive tools might be predictable to become a crucial element of agriculture.

Crop yield is a combined bio-socio-system comprised of complex interaction among the soil, the air, the water, and the crops grown in it, where a comprehensive model is necessary which are possible only through classical engineering expertise. As define by the Food and Agriculture of the United Nations, crop forecasting is the art of predict crop yields and production before the harvest in fact takes place, typically a couple of months in advance. Crop forecasting philosophy is based on various kinds of data collected from different sources: meteorological data, agro-meteorological, soil, remotely sensed, agricultural statistics. Based on meteorological and agronomic data, several indices are derived which are deemed to be relevant variables in determining crop yield, for instance crop water satisfaction, surplus and excess moisture, average soil moisture, etc. Linear Regression model characterizes the mathematical relationships intrinsic to the data set from previous experiments. This method can produce results under various situations assuming extensive information used to expand and test the model. Though, in agricultural data, information is rather sparse and incomplete. Because of this limitation, the linear regression approach is the common approach for predicting yield across large area.

The most investigate statistical crop-yield-weather model are multivariate regression models. Data mining technique aim at finding those patterns or information in the data that are together valuable and interesting to the farmer. A common specific problem to occur is yield prediction.

2. LITERATURE REVIEW

Rajshekhar Borate etc.al [2016] describes and gave the details us for list of used methods, In India there are dissimilar agriculture crops production and those crops depends on the several kind of factors such as environmental science, economy and also the geographical factors covering such methodologies and methods on historic yield of dissimilar crops, it is possible to get info or data which can be supportive to farmers and government organizations for creation well decisions and for make better rules which help to increased production. In this article, our effort is on application of data mining techniques which is use to extract information from the agricultural records to estimate better crop yield for main crops in main districts of India. In our project we found that the precise prediction of dissimilar specified crop yields across different districts will help to farmers of India. From this Indian farmers will plant different crops in different distr [8].

Ramesh, D., and VishnuVardhan, B.,Agrarian et.al [2015] discussed a several subdivision in India is facing rigorous problem to make the most of the crop productivity. More than 60 out of a hundred the crop still depends on monsoon rainfall. Current growths in Information Technology for agriculture field have developed an interesting research area to forecast the crop yield. The problematic of yield prediction is a major problem that remains to be solved based on accessible data. Data mining methods are the better selections for this purpose. Different Data Mining methods are used and evaluated in agriculture

for approximating the upcoming year's crop production. This paper presents a brief analysis of crop yield prediction using Multiple Linear Regression (MLR) method and Density based clustering technique for the particular region i.e. East Godavari district of Andhra Pradesh in India. In this paper an effort is made in command to know the region precise crop yield analysis and it is processed by applying both Multiple Linear Regression method and Density-based clustering method. These models were experimented in respect of all the districts of Andhra Pradesh, then the procedure of evaluation is passed out with only East Godavari district of Andhra Pradesh in India [1].

Veenadhari, S., Bharat Misra, D Singh et.al [2011] discussed that the data mining extraction of unseen predictive information from huge records, is a powerful new technology with great potential to help companies focus on the most significant data in their Data warehouses. Data mining tools predict upcoming trends and performance and growth, allowing businesses to make proactive, knowledge-driven decisions. Though these methods are plausible, theoretically well created, and perform well on extra or less artificial test data sets, they depend on their skill to make sense of real-world data. This article gave us a detail project that is smearing a range of machine learning plans to problems in agriculture and horticulture. They briefly surveyed some of the techniques emerging from machine learning study, define a software workbench for testing with a variability of methods on real-world data sets, and a learning of dairy herd management in that culling rules were inferred from a medium-sized record of herd information. They also defined a range of machine learning plans to problems in agriculture and horticulture. There is a rising number of applications of data mining methods in agriculture and a rising amount of data that are presently available from several resources. This is relatively a novel research field and it is expected to grow in the upcoming. There is a lot of effort to be done on this emerging and interesting study field. The multidisciplinary method of integrating computer science with agriculture will help in predicting managing agricultural crops effectively [10].

Dakshayini Patil etc.al [2017] describes and discover the list of methods and techniques which are used Rice crop creation assumes an imperative part in sustenance safety of India, contributing over 40% to general yield generation. High harvest generation is reliant on appropriate climatic situations. Inconvenient regular atmosphere conditions, for example, low precipitation or temperature extremes can drastically diminish edit yield. Rising well plans to foresee edit efficiency in several climatic conditions can help rancher and different partners in vital basic leadership as far as agronomy and yield result. This article reports utilization of many information mining approaches will anticipate rice trim yield for Maharashtra state, India. To this review, 27 regions of Maharashtra were picked on the establishment of accessible information from openly available Indian Administration records with different atmosphere and yield limitations. This surveys the technical achievements in the field of Rice crop yield prediction. Discuses methodology, comprehensive survey of many proposed approaches to predict rice crop yield and applications. It also discusses various data mining methods used for prediction of crop yield for rice. Rising better plans to foresee crop productivity in various climatic conditions can help farmer and different partners in essential basic leadership as far as agronomy and product decision [2].

Ramesh A. Medar and Vijay. S. Rajpurohit et.al [2014] presented a Precision agriculture (PA) and information technology (IT) are closely interwoven. The former frequently refers to the application of nowadays' technology to agriculture. Due to the use of sensors and GPS technology, in today's agriculture several data are collected. Creation use of those data via IT often leads to dramatic improvements in efficiency. For this purpose, the challenge is to change these raw data into useful data. This paper deals with suitable modeling methods for those agricultural data where the objective is to uncover the surviving patterns. In specific, the use of feed-forward back propagation neural networks will be evaluated and suitable parameters will be projected. In consequence, yield prediction is allowed based on cheaply obtainable site data. In this prediction, economic or environmental optimization of, e.g., fertilization can be passed out. Due to the rapidly advancing technology in the last few decades, ever more of our everyday life has been changed by information technology. Data access, once cumbersome and slow, has been turned into "data at your fingertips" at high speed. Technological breakthroughs have been made in industry and services as well as in agriculture [9].

V. Leemans, M.-F. Destain et.al [2004] describes that the suggested Fresh market fruits like apples are graded into quality groups according to their size, color and shape and to the attendance of defects. The two first quality criteria are actually automatic on industrial graders, then fruits classifying according to the presence of faults is not yet efficient and so remains a manual operation, repetitive, luxurious and not reliable. The classifying of apples using machine vision can be arbitrarily separated into four steps: the images acquisition, their segmentation, their interpretation and in conclusion the fruit classification. This paper presents the three former points on the basis of a literature review, the research outcomes being absorbed on the last point: having extracted data from images acquired on fruits, the paper defines a classifying technique which was implemented on an existing machine and tested on Jon gold apples (bi-color fruits).The first step consists of acquiring images of the surface of the fruit, though it goes through the classifying machine. In order to grade apples, two necessities have to be met: the images should cover the entire surface of the fruit; a high contrast has to be created among the defects and the healthy tissue, while maintaining a low variability for the healthy tissue [4].

Roger J. Brooks, Mikhail. Semenov, Peter D. Jamieson et.al [2001] presented simpler meta-model, which produced very similar yield predictions to Sirius of potential and water-limited yields at two locations in the UK,

Roth Amsted and Edinburgh. This greatly increases the understanding of the nature and consequences of the relationships implicit within Sirius. The study showed that the reply of wheat crops to climate could be explain using a few simple relationships. The meta-model aggregate the three main Sirius components, the computation of leaf area index, the soil water balance model and the evapotranspiration calculation, into simpler equations. These results in a obligation for calibration of fewer model parameters and means that weather variables can be provide on a monthly rather than a daily time-step, because the meta-model can use cumulative values of weather variables. As a result the meta-model is a valuable tool for regional impact assessments when detailed input data are usually not available. As the meta-model was developed from the analysis of Sirius, rather than from statistical fitting of yield to weather data, it must do well for other locations in Great Britain and with different management scenarios [7].

Prof. M.S .Prasad Babu, N.V.Ramana Murty, S.V.N.L.Narayana et.al [2010] describes that the tomato is now the most widely grown vegetable crop in World. It is grownup through the world in farm gardens, small home-gardens, and by market gardeners for fresh consumption as well as for processing purposes. This Tomato crop skilful advisory system is intended at a collaborative venture with eminent Agriculture Scientist and Experts in the area of Tomato Plantation with an excellent team of computer Engineers, computer programmer and creators. This Expert System contains two main parts one is Tomato Info System and the other is Tomato Crop Expert System where in Data system, the user can get all the static information about different species, Illnesses, Symptoms, chemical controls, Preventions, Pests, Virus of Tomato fruits and plants. In Advisory System, the user is having a communication with the expert system online; the user has to answer the questions asked by the Expert System. Depends on the reply by the operator the expert system decides the disease and shows its control measure of disease. This Tomato Crop Data Expert System deals with different varieties of Tomato Crop, Identification of various diseases usually chances to tomato crop based on the symptoms [6].

3. SYSTEM METHODOLOGY

3.1 ARCHITECTURE OF CROP YIELD PREDICTION

The crop yield prediction includes repeatedly all essential parameters that are needed for the well yield of crop. This improves the classification outcomes of the crop yield. All the essential parameters are thought-about as inputs. In common, one in all the issues faced with in the prediction method is that almost all of the required parameters that are essential to consider for the exact prediction are not consider. It decreases the efficiency of the anticipated outcomes which in turn leads to lack of proper forecasting of the crop harvest its additionally tougher to predict the improved predict the improved range of input parameters that are to be considered in the prediction procedure.

Crop prediction is that the art of predicting crop yields and manufacture before the yield really takes place. Before harvest prediction was done by considering the farmer's knowledge on a selected field and crop. This work presents a system that uses data processing strategies so as to predict the analyzed datasets. The anticipated sort can specify the yielding of crops.

Architecture is a system that unites its parts or components into a coherent and purposeful complete. The crop information base consists of farm data like crop varieties, crop year, area and seasonal parameter like Khrif, rabbi and summer crops. The knowledge-based additionally contains of zones furthermore district information, ecological parameter like extreme and lowest temperature value and average precipitation.

The crop yield prediction model that includes associate input module that is in charge of taking input from the farmer. The input module includes crop name, land area, crop year and prediction tons The feature selection model is in charge offset. Selection of associate attribute from crop particulars. The crop yield prediction model used to predict the yield. Once feature selection, the data go to classification rule for grouping similar contents

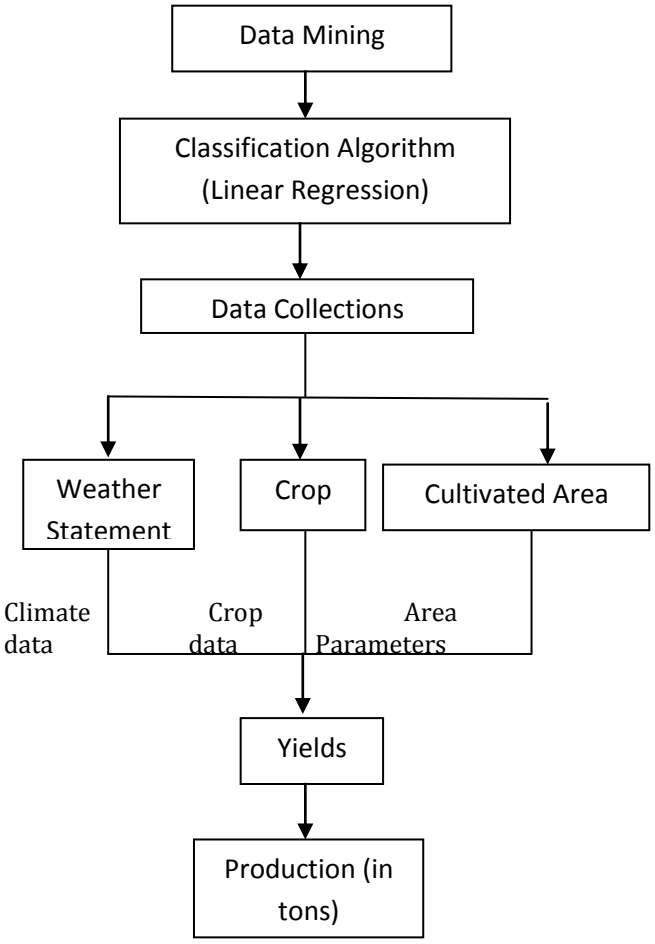

Fig 3.1 Architecture of crop yield Prediction

Climate data and crop parameters used to predict crop growth can be predicted. Then prediction rules are going to be applied to the output of classifying crop particulars in terms of crop name, season and total yield details.

3.2 IMPLEMENTATION

Regression analysis is used to analyze and determine the affiliation between response variable and explanatory variable. The variables considered for analysis during this analysis work are annual prediction, area under cultivation. Crop yield may be dependent on variable which depends on all these ecological factors.

Linear Regression

A Linear regression methodology that's used to analyze a response variable that alterations with the value of the interference variable. A way of predicting the value of a response variable from a given value of the explanatory variable is also referred to as prediction. The least-square fit, that is capable of fitting each linear additionally as polynomial relations, is that the most typically used linear regression. The method of applying model estimate to values outside the domain of the first knowledge is thought as extrapolation. A linear regression model is computed to analyze the yield.

Linear Regression model for crop yield prediction:

To develop the Linear Regression models for crop yield prediction, Linear Regression analysis is majorly used for prediction functions because it provides predicted entity as a function of depended entities.

Steps for crop yield prediction using Linear Regression algorithm as follows: Inputs are given in experimental information set of whether or not data crop information and soil information and their outcomes predicted crop yield for the experimental dataset. Some technique are given

Gather, format and organize the information: only raw information is scarce to work with the model. The data should be gathered, soft out as per the requirement and organization it in such a path, that appropriate results are obtained. Although redoing, additional vital data can be included.

- Collect the dataset then preprocessing the dataset for noise removal process.
- Separate data into testing and training sets: the data information must to go partitioned into two sets. Training set can have greatest rate of the data so as to train most of the examples to create the yield. The samples are collected under training set. Testing set uses the remaining measure of the data to check however they system is performing.

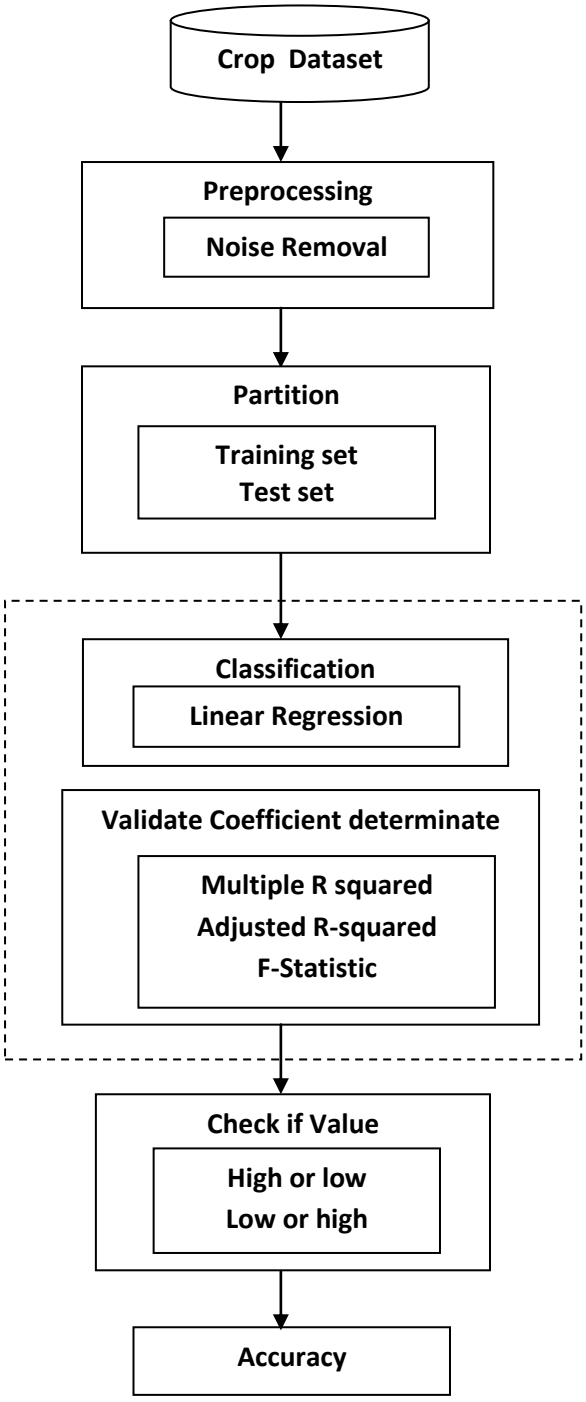

Fig 3.2 Linear Regression Model for Crop Yield Prediction

- Apply Linear Regression on trained sets: the model system depends on upon however complicated the problem is and also the structure likewise should be selected with the requirement. Though altering, the development modeling and structure is adjusted.
- Validation the Multiple R square, Adjusted R-squared and F-Statistic values for this models.
- Apply the trained Linear Regression model on test set and once again calculate the Multiple R square,

Adjusted R-squared and F-Statistic values. Compare the values with completely different models of Linear Regression models. The model that provides to be the best model for crop yield prediction.

4. RESULT AND DISCUSSION

4.1 PERFORMANCE METRICS

Within recent analysis, effectiveness related to carries with it technique reviewed creating use of Espresso at the side of R-Tool 2.35 are widely used to implement with this carries with it methodology. The actual effectiveness carries with it technique through – about once it involves preciseness, recognition moreover as accuracy. The final results, means category effectiveness could be superior through the utilization of LR-PROPOSED being an optimization technique within the category method.

Precision Value: Precision is calculated as the number of true positive predictions divided by the total number of positive predictions.

$$Precisionvalue = \frac{Truepositive}{(Truepositive + Falsepositive)}$$

Recall value: Recall value is specified to as the relevant datasets that are related to the other request Search.

$$Recallvalue = \frac{Truepositive}{(Falsepositive + FalseNegative)}$$

F measure: F measure test's accurateness and is define as the weighted harmonic mean of the precision and recall of the analysis

$$Fmeasure = 2 * \frac{precision * recall}{precision + recall}$$

Analyze to comparison between K-Nearest Neighbor (Existing System)and Linear Regression (Proposed System) with parameter evaluation.

Table 4.1.1 Comparisons of Parameter Values

Algorithm/ Parameter	Precision	Recall	F-Measure
K-NN	0.84	0.85	0.84
LR-Proposed	0.86	0.87	0.89

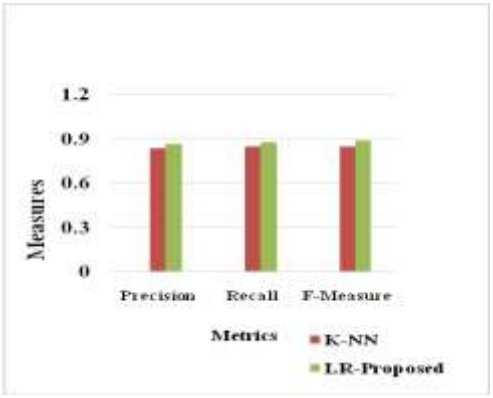

Fig.4.1.1 Graphical Parameters Comparison for Existing K-NN with Proposed LR-Proposed algorithm.

Accuracy: Accuracy gives the required related datasets used for classification. Compute the proportion of true positive and true negative in all calculated cases.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Analyzed crop accuracy compared with K-NN and LR-Proposed

Table 4.1.2 Crop Accuracy Comparison

Algorithm	Cotton	Sugarcane	Turmeric
K-NN	87	84	85
LR-Proposed	95	96	95

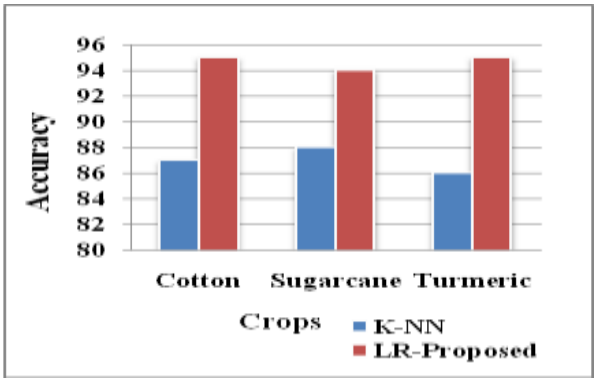

Fig 4.1.2 Crop Accuracy Compared with K-NN and LR-Proposed.

The efficient classification Linear Regression algorithm is used to develop the model. This algorithm is compared, and accuracy is evaluated. From the above table 5.3.2, it is observed that Linear Regression had the best predictive power with high accuracy as compared to K-Nearest Neighbor.

CONCLUSION

In accurate prediction of different specified crop yields across different districts will help to farmers of India. Yield estimation models are utilized in preciseness Agriculture to extend yield production to satisfy demand and to recommend to the government in regard to prediction crop yield on imports of Trichy, Tamilnadu dataset to avoid overlapping. During this work the regression approach were tested in their yield prediction capabilities. The readings were used for model inputs. Linear regression algorithms offered acceptable estimation accuracy, though higher prognostic power could also be obtained by parameters like year, crop, area, production (in tons) and alternative variables, like climate, agricultural practices and soil characteristics are including within the model development. The model using linear regression can be suggested for Ecuadorian conditions. In yield prognostic models are not existent for any crop. From this proposed system the yield of crop (sugarcane, cotton, and turmeric) are predicted in highest level. This model may be reformulated using alternative crop assessments within the future, to develop methods for increasing yield and land territorial management in alternative crops of importance, like wheat, rice.

REFERENCES

1) D Ramesh, B Vishnu Vardhan, "Analysis of Crop Yield Prediction using Data Mining Techniques", International Journal of Research in Engineering and Technology (IJRET),Vol.4, 2015.

2) DakshayiniPatil, Dr. M .S Shirdhonkar, "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.

3) Dr. Rakesh Poonia1 , Sonia Bhargava "Prediction of Crops Methodology using Data Mining Techniques", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 10, October 2017.

4) Leemans V, M F Destain, "A Real Time Grading Method of Apples Based on Features Extracted from Defects", J. Jood Eng., 2004, pages: 83-89.

5) Mehta D R, Kalola A D, Saradava D A, Yusufzai A S, "Rainfall Variability Analysis and Its Impact on Crop Productivity - A Case Study", Indian Journal of Agricultural Research, Volume 36, Issue 1, 2002, pages : 29-33.

6) Prof .M.S.PrasadBabu, N.V.Ramana Murty, S.V.N.L.Narayana, "A Web Based Tomato Crop Expert Information System Based on Artificial Intelligence and Machine Learning Algorithms", IJCSIT, Vol. 1 (1), 2010, 6-15.

7) R J Brooks, "Simplifying Sirius: Sensitivity Analysis and Development of a Meta-Model for Wheat Yield Prediction", European Journal of Agronomy, vol. 14, 2001, pages: 43-60.

8) Rajshekhar Borate., "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India", International Journal of Novel Research in Computer Science and Software Engineering,Vol. 3, Issue 1, pp: (34-37), April 2016.

9) Ramesh A. Medar and Vijay. S. Rajpurohit "A Survey of data mining techniques for crop yield prediction", IJARCSMS, Volume 2, Issue 9, September 2014 pg. 59-64.

10) Veenadhari, S., Bharat Misra, D Singh, "Data mining Techniques for Predicting Crop Productivity – A review article", IJCST, International Journal of Computer Science and technology march 2011.