2016 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2016)

Rice Crop Yield Prediction Using Artificial Neural Networks

Niketa Gandhi, Owaiz Petkar
University Dept. of Computer Science
University of Mumbai
Mumbai, Maharashtra, India
niketa@gmail.com
owaizpetkar@gmail.com

Leisa J. Armstrong
University Dept. of Computer Science, University of Mumbai and School of Science, Edith Cowan University
Perth, Western Australia
l.armstrong@ecu.edu.au

***Abstract*- Rice crop production contributes to the food security of India, more than 40% to overall crop production. Its production is reliant on favorable climatic conditions. Variability from season to season is detrimental to the farmer's income and livelihoods. Improving the ability of farmers to predict crop productivity in under different climatic scenarios, can assist farmers and other stakeholders in making important decisions in terms of agronomy and crop choice. This study aimed to use neural networks to predict rice production yield and investigate the factors affecting the rice crop yield for various districts of Maharashtra state in India. Data were sourced from publicly available Indian Government's records for 27 districts of Maharashtra state, India. The parameters considered for the present study were precipitation, minimum temperature, average temperature, maximum temperature and reference crop evapotranspiration, area, production and yield for the Kharif season (June to November) for the years 1998 to 2002. The dataset was processed using WEKA tool. A Multilayer Perceptron Neural Network was developed. Cross validation method was used to validate the data. The results showed the accuracy of 97.5% with a sensitivity of 96.3 and specificity of 98.1. Further, mean absolute error, root mean squared error, relative absolute error and root relative squared error were calculated for the present study. The study dataset was also executed using Knowledge Flow of the WEKA tool. The performance of the classifier is visually summarized using ROC curve.**

Keywords-artificial neural network; ANN; crop analysis; WEKA; crop yield prediction.

I. INTRODUCTION

Achieving high crop yields is the principle aim of agricultural production. The recognition and management of factors that influence crop yield assist farmers in decision making. There are a number of crop yield prediction models which use either statistical or crop simulation models. Over the last decade it has been observed that Artificial Intelligence (AI) techniques provide a more effective approach to predicting crop yield under different cropping scenarios. The use of artificial neural networks can make models with complex inputs easier to interpret.

This research describes the development of a rice crop yield prediction model through the use of ANNs. The prediction is achieved with the flexible methodology of artificial neural networks (ANNs) using back propagation technique. The most common neural network model is the multilayer perceptron that has been used in the present research.

This approach has been demonstrated by forecasting of rice crop yield prediction for Kharif season from year 1998 to 2002 for Maharashtra state of India, on the basis of different predictor variables including precipitation, minimum temperature, average temperature, maximum temperature, reference crop evapotranspiration and yield. Artificial Neural Networks with Multilayer Perceptron were considered for the present research.

II. RELATED WORK

Biological neural processes of human brain form the basis for the ANN structure. To develop these models the interrelationships of the interconnected processing neurons of the human brain are used. An ANN model develops a formula to ascertain the relationship using large number of input and output examples, to establish model predictions [1]. With little a priori knowledge of the functional relationship, nonlinear relationships which are overlooked by other prediction methods can be determined [2]. An ANN model requires minimum of three layers: the input, hidden and output layers. The input and output layers contain nodes that correspond to input and output variables, respectively. Data move between layers across weighted connections.

A number of studies have described the creation of agronomic based models by using artificial neural networks. Agronomic ANN applications include crop development modeling [3], pesticide and nutrient loss assessments [4], soil water retention estimations [5], and disease prediction [6]. A recent study has reported how aerial pictures have been generally utilized for crop yield prediction before harvesting [7]. These photos were able to provide spatial cloud free data of the yields spectral qualities. Despite the fact that regression model expends more time to be developed, an ANN model can deliver more reliable and accurate crop yield forecasting compared to other methods such as regression models [8].

978-1-5090-0615-1/16/$31.00 ©2016 IEEE

Various studies have described the potential for using ANN for the prediction of crop production. A study by [9] reported the use of ANN for the development of straightforward and precise estimation tool to predict rice yields. In a study using Ward-style ANNs, a year round air temperature prediction model predicting the horizons of 1 to 12 h was developed. These models were found to be useful in the general decision support for farmers [10]. According to a study on predicting soybean rust by [6], the ANN models were found to produce better results than traditional statistical methods. Similarly, in a study more accurate results were generated using ANN model compared to regression techniques to estimate soil water content based on soil physical properties [11]. In another study it was reported that an ANN model performed better ($R2=0.984$) than a regression model ($R2 = 0.780$) while predicting applied nitrogen leaking below the root zone of turf grass [12].

A number of studies have reported on the use of ANN to predict crop yield. Studies have reported the used of back propagation network to predict rice yield based on weather data [13]. Prediction of maize yield based on rainfall, soil and other parameters obtained a testing error of 14.8% [14], and prediction of rice yield using climatic observation data predicted with a maximum of 45-60kg/ha [15]. Other studies have used neural networks to predict rice yield based on soil parameters and achieved a testing error of 17.3% [16] or developed a straightforward and precise estimation to predict rice yields [17].

Crop yield prediction through the use of ANN was also examined for Nepal. In the training phase, the ANN model provided moderate error that showed the accuracy of the model. Further the comparison between the expected and neural net outcome showed the similar result which further evidence the accuracy of the prediction values using this neural network model. The predicted outcome could be utilised for enhancing the paddy yield in Siraha district of Nepal and also in other regions where the topography and vegetation are similar to Siraha district [18]. Other recent studies have shown the effectiveness of a decision support system which used artificial neural network to predict rice production in Phimai district, Thailand [19]. Similar findings were reported in another study which demonstrated the application of using artificial neural network using feed forward back propagation for agricultural crop yield prediction. This study concluded that ANN is a beneficial tool for crop prediction [20].

This paper reports on the use of Artificial Neural Networks to predict the rice crop yield for Maharashtra state, India. The purpose of the research was to (1) check the effectiveness of predicting rice crop yield using Artificial Neural Network (Multilayer Perceptron) for Maharashtra state, India and (2) Evaluate the performance of Artificial Neural Network (Multilayer Perceptron).

III. RESEARCH METHODS

This section will report on the methods used for this study including study area, data set and methodology.

A. Study Area

The study zone for this research was Maharashtra state of India. Maharashtra is situated in the western part of India along the Arabian Sea [21]. It lies between 15° 44' to 22° 6' N and 72° 36' to 80° 54' E. Maharashtra state is the third biggest state in terms of both area and population in India [21]. Agriculture is one of the important industries in the state with both food and trade products are grown in the state. The principal food products of Maharashtra are wheat, rice, jowar, bajra, and pulses [21]. The state is comprised of six regulatory divisions, which are further partitioned into 35 districts, 109 sub-divisions and 357 talukas [21]. For the present study, 27 regions were chosen in particular, Ahmednagar, Amravati, Aurangabad, Beed/Bid, Bhandara, Buldhana, Chandrapur, Dhule, Gadchiroli, Gondia, Hingoli, Jalana, Jalgaon, Kolhapur, Latur, Nagpur, Nanded, Nasik, Osmanabad, Parbhani, Pune, Sangli, Satara, Solapur, Wardha, Washim and Yavatmal.

B. Dataset Used

All the datasets used in the present research were sourced from the publicly available records of the Indian Government for the year 1998 to 2002. The parameters considered for the present research are described below:

- Precipitation (mm): Precipitation is a noteworthy segment of the water cycle, and is responsible for saving the fresh water on the planet. The aggregate precipitation for Kharif season (June to November) for each year of each district was computed from the monthly mean precipitation from year 1998 to 2002 for a specific district.
- Minimum, Average, Maximum Temperature (degree Celsius): Temperature is an essential variable influencing the rate of plant development. Temperatures along with environmental change and the potential for more extreme temperature events will affect plant productivity i.e. crop production. Thus average, minimum and maximum temperature for year 1998 to 2002 of each district was considered for the present research. The average, minimum and maximum temperature for the Kharif season (June to November) were calculated for each district on the basis of monthly mean temperatures.
- Reference Crop Evapotranspiration (mm): Evapotranspiration is the sum of evaporation and plant transpiration from the Earth's territory and sea surface to the air. The reference crop evapotranspiration was calculated on the basis of monthly mean of that year from 1998 to 2002 for each district.
- Area (Hectares): The rice cultivated area of the study area in Maharashtra in Kharif season (June to November) from the year 1998 to 2002 was considered for the present research.

978-1-5090-0615-1/16/$31.00 ©2016 IEEE

- Production (Tonnes): The rice production for the study area for Kharif season (June to November) was considered for the present research.
- Yield (Tonnes/Hectare): An amount of produced rice and the area cultivated for rice in Kharif season from year 1998 to 2002, of each of the selected district from the Maharashtra state, the yield was calculated and considered for the present research.

C. Methodology

Microsoft Office Excel was used to incorporate all the data sets that were acquired for this study. The following steps were followed for processing and preparing the data for applying multilayer perceptron technique.

Step 1: Acquiring each parameter (precipitation, minimum, average, maximum temperature and reference crop evapotranspiration) monthly mean records of each district from 1998 to 2002 from the Indian Government records.
Step 2: Calculating the total precipitation for each year for each district during the Kharif season (June to November).
Step 3: Calculating the average temperature during the Kharif season for minimum, average and maximum temperature for each year for each district.
Step 4: Calculating the average reference crop Evapotrans-piration for each year for each district of Maharashtra state during the Kharif season.
Step 5: Acquiring each districts area, production and rice crop yield details of the year 1998 to 2002 from the publicly available Indian Government records.
Step 6: The raw data set was then collated in single sheet which consisted of the following columns in Microsoft Excel: sr. no, name of the state, name of the district, year, precipitation, minimum temperature, average temperature, maximum temperature, soil type, area, production and yield.
Step 7: For some of the districts particular year's climatic parameters or production data was not available hence those records were omitted. That particular year's data was not used for the current research. Record number was added for each record.
Step 8: For preparing the data set for applying multilayer perceptron technique, unrequired columns were removed. They were sr. no, name of the district and year.
Step 9: The data set was then sorted on the basis of area. Area less than 100 hectares were not considered for the present research. So those records were omitted.
Step 10: The dataset was then sorted on the basis of yield to classify the records in to low, moderate and high yield of rice production. Class low with the range 0.15 to 0.60 tonnes/hectare, class moderate with the range 0.61 to 1.10 tonnes/hectare and class high with the range 1.11 to 3.16 tonnes/hectare.
Step 11: The area and production columns were omitted as yield has been calculated on the basis of them.
Step 12: This data set was then saved in .csv format for further applying multilayer perceptron technique in Weka.

WEKA (Waikato Environment for Knowledge Analysis) [22] is freely available and open source data mining tool available under the GNU General Public License. It was developed at the University of Waikato, New Zealand and was first implemented in its modern form in 1997. Explorer is the principal user interface for Weka but the same functionality is available using the Knowledge Flow interface and from the command line. There is Experimenter which allows the orderly correlation of Weka's machine learning calculations on a gathering of datasets. In addition to the .arff data file format it also supports .csv data file format which makes it possible to export the data into various spreadsheet applications. The data set prepared for the present research was then opened in WEKA and save in .arff format for further processing.

Artificial Neural Network – An Overview

A neural network consists of a set of highly interconnected entities, called nodes or units [23]. Each unit is designed to mimic its biological counterpart, the neuron. Each neuron accepts a weighted set of inputs and responds with an output. Artificial Neural Network (ANN) technique is based upon the model of biological nervous system. The key element of this technique is the novel structure of the information processing system. An ANN is configured for a specific application such as pattern recognition or data classification, through a learning process.

The most important types of Neural Networks for real world problem solving are Multilayer Perceptron, Radial Basis Function Networks and Kohonen Self Organizing Feature Maps [23].

For present research the multilayer perceptron was used as artificial neural network for the prediction of rice crop yield for Maharashtra state, India.

IV. PERFORMANCE EVALUATION

A confusion matrix was generated after running the current data set on Multi Layer Perceptron using WEKA tool. All measures can be calculated based on four values, namely True Positive (TP, a number of correctly classified that an instances positive), False Positive (FP, a number of incorrectly classified that an instance is positive), False Negative (FN, a number of incorrectly classified that an instance is negative) and True Negative (TN, a number of correctly classified that an instance is negative). These values are defined in Table I below.

TABLE I. A CONFUSION MATRIX

		Observed	
		True	**False**
Predicted	**True**	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

978-1-5090-0615-1/16/$31.00 ©2016 IEEE

From the values of Table I sensitivity, specificity, accuracy, F1 Score and MCC were calculated.

$$\text{Sensitivity} = TP / (TP + FN) \quad (1)$$

$$\text{Specificity} = TP / (TP + FP) \quad (2)$$

Correctly classified instances are known as sensitivity and incorrectly classified instances are known as specificity and can be calculated using Eq. 1 and Eq. 2 described above.

Accuracy is defined as the overall success rate of the classifier and computed by Eq. 3

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (3)$$

The F1 score measures exactness utilizing the insights precision p and recall r. Accuracy is the proportion of true positives (TP) to all predicted positives (TP + FP). Recall is the proportion of true positives to every single actual positive (TP + FN).

The F1 Score can be calculated by using the Eq. 4.

$$F1 = (2TP)/(2TP + FP + FN) \quad (4)$$

The Matthews Correlation Coefficient (MCC) is a measure of the nature of paired (two-class) arrangements.

The MCC can be calculated by using the Eq. 5.

$$MCC = ((TP*TN)-(FP*FN)) /\surd(TP+FP)(TP+FN)\ (TN+FP)\ (TN+FN) \quad (5)$$

The study dataset was also executed using Knowledge Flow in WEKA tool. The Knowledge Flow provides an alternative to the Explorer as a graphical front end to WEKA's core algorithms. The Knowledge Flow presents a data flow inspired interface to WEKA. The WEKA components were selected from the tool bar and placed them on a layout canvas and connected them together in order to form a knowledge flow for processing and analyzing current dataset. The components selected in Knowledge flow are discussed below:

- DataSources tab:
  - ArffLoader: Used to load the arff dataset file for the current study
- Evaluation tab:
  - ClassAssigner: Assigns a column to be the class for current study data set, training set or test set.
  - ClassValuePicker: Chooses a class value to be considered as positive class. This is useful for generating data for Receiver Operating Characteristic (ROC) curves that is generated for the current study.
  - CrossValidationFoldMaker: Splits the current data set, training set or test set into folds.
  - Classifier Performance Evaluator: Evaluates the performance of batch trained/tested classifiers of the current study.
- Classifiers tab:
  - MultiLayer Perceptron: Classifier to be used for the present study.
- Visualization tab:
  - Text Viewer: Component shows textual data. Shows data sets and classification performance statistics.
  - Mode lPerformance Chart: Component used for visualising threshold ROC curves.

The Text Viewer was used to see the results and Model Performance Chart was used to visualise the Receiver Operating Characteristic (ROC) Curve. This was used to compare the results of classifier for accuracy, sensitivity, and specificity. ROC curve is a graphical plot that shows the execution of a parallel classifier framework as its separation edge is shifted. The curve is achieved by plotting the true positive rate (TPR) also called as sensitivity or recall against the false positive rate (FPR) also called as fall out at different edge settings.

V. EXPERIMENTAL RESULTS

This section presents the results obtained after running the Multi Layer Perceptron technique on rice crop yield data set of Maharashtra state, India. WEKA was used to construct the algorithms. Different parameter set for Multi Layer Perceptron were as follows: GUI=true; autoBuild=true; debug=false; decay=false; hiddenLayers=a; learningRate=0.3; momentum=0.2; nominalToBinaryFilter=True; normalizeAttributes=True; normalizeNumericClass=True; reset=True; seed=0; trainingTime=500; validationSetSize=0; validationThreshold=20.

The algorithm achieved the accuracy of 97.54%, sensitivity of 96.33% and specificity of 98.12%. A three layer feed-forward back propagation neural network was obtained after running the technique as shown below in Figure 1 with 6 neurons in Input layer, 4 neurons in hidden layer and 3 neurons in output layer showing the predicted yield as low, moderate and high.

For training and testing the ANN, 10-fold cross validation method was used to subset the data. The data were randomly divided into 10 parts of which one part was used for testing while the remaining parts were used as the training data. The processes were repeated 10 times and the results were average. The back propagation algorithm was applied for training the feed-forward back propagation neural network.

The model was also evaluated and following parameters were computed which resulted in mean absolute error of 0.0526, root mean squared error of 0.1527, relative absolute error of 12% and root relative squared error of 32%. Further it gives F1 score of 0.96 and MCC of 0.94.

978-1-5090-0615-1/16/$31.00 ©2016 IEEE

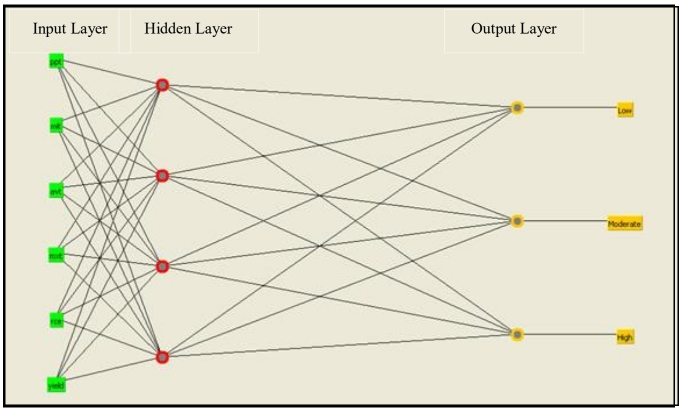

Figure 1 Neural Network Diagram

Figure 2 below shows the Knowledge Flow layout used for the current dataset. The dataset was loaded using an ArffLoader and results were generated and seen in TextViewer.

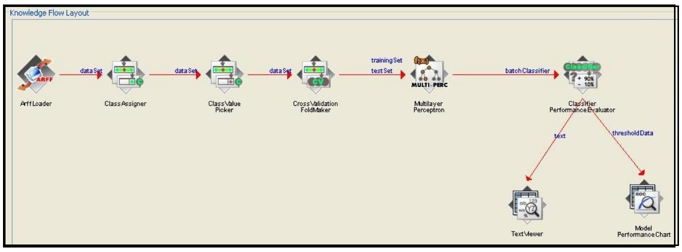

Figure 2 Knowledge Flow Layout

Further in this exploration the ROC curve was visualised using ModelPerformanceChart for three classes Low, Moderate, High which represents the sensitivity and specificity of those classes as shown in Figure 3 below.

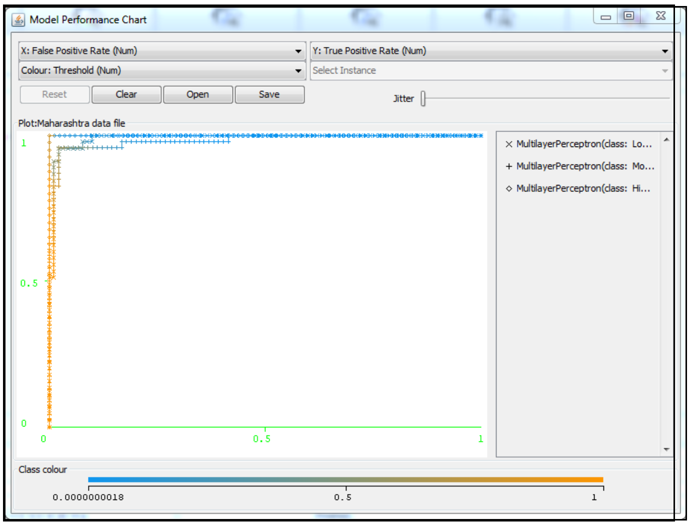

Figure 3 Model Performance Chart

Figure 3 shows the model performance chart. The curve which is generated in continuous form represents 100% sensitivity (no false negatives) and 100% specificity (no false positives). And the curve which generated in discrete form carries false negatives and false positive values which do not have 100% sensitivity and specificity.

VI. DISCUSSION AND CONCLUSIONS

To show the interactions between the factors affecting the crop yield, a non-linear method is required to interpret the relationship. A. linear method such as linear regression was considered to be insufficient to show the interactions of the factors and crop yield due to the complexity of the factors affecting crop yield. . ANN was considered one alternative to traditional linear regression techniques for predicting crop production.

The accuracy of these models show the potential to predict the rice crop yield for Maharashtra as well as India. Other data mining techniques applied for the current study area have been proved to be less accurate compared to the ANN model discussed in this paper [24,25]. It is possible to widen the effectiveness and improve the prediction capabilities of the ANN based rice crop yield prediction by considering additional parameters that effect the crop production for Maharashtra state, India.

REFERENCES

[1] Y.A. Pachepsky, D. Timlin, and G. Várallyay, “Artificial neural networks to estimate soil water retention from easily measurable data”, Soil Science Society of America Journal, vol. 60(3), pp.727-733, 1996.

[2] D.A. Elizondo, R.W. McClendon and G. Hoogenboom, “Neural network models for predicting flowering and physiological maturity of soybean”, Transactions of the ASAE, vol. 37(3), pp.981-988, 1994.

[3] D.A. Elizondo, R.W. McClendon, and G. Hoogenboom, “Neural network models for predicting flowering and physiological maturity of soybean”, Transactions of the ASAE, vol. 37, pp. 981–988, 1994.

[4] C.C. Yang, S.O. Prasher, S. Sreekanth, N.K. Patni and L. Masse, “An artificial neural network model for simulating pesticide concentrations in soil”, Transactions of the ASAE, vol. 40, pp. 1285–1294, 1997.

[5] M.G. Schaap and W. Bouten, “Modeling water retention curves of sandy soils using neural networks”, Water Resources Research, vol. 32, pp. 3033–3040, 1996.

[6] W.D. Batchelor, X.B. Yang and A.T. Tshanz, “Development of a neural network for soybean rust epidemics”, Transactions of the ASAE, vol. 40, pp. 247–252, 1997.

[7] G.A. Thomas, G. Taylor and J.C.Wood, “Mapping yield potential with remote sensing”, Precision Agriculture, vol. 1, pp.713–720, 1997.

[8] M. Kaul, R.L. Hill and C. Walthall, “Artificial neural network for corn and soybean prediction”, Agricultural System, vol. 85, pp. 1-18, 2005.

[9] B. Ji, “Artificial neural networks for rice yield prediction in mountainous regions”, Journal of Agricultural Science, vol. 145, pp. 249–261, 2007.

[10] B.A. Smith, G. Hoogenboom and R.W. McClendon, “Artificial Neural Networks for Automated Year round Temperature Prediction”, Computers and Electronics in Agriculture, vol. 68, pp. 52–6, 2009.

[11] M. Schaap and W. Bouten, “Modeling water retention curves of sandy soils using neural networks”, Water Resour. Res., vol.32, pp. 3033–3040, 1996.

978-1-5090-0615-1/16/$31.00 ©2016 IEEE

[12] S.K. Starrett and G.L. Adams, “Using artifi - cial neural networks and regression to predict percentage of applied nitrogen leached under turfgrass”, Commun. Soil Sci. Plant Anal., vol. 28, pp.497–507, 1997.

[13] J. Liu, C. Goering and L. Tian, “A neural network for setting target corn yields”, Transaction of the ASAE, 44(3), pp. 705-713, 2001.

[14] M. O‘Neal, B. Engel, D. Ess, J. Frankenberger, “Neural network prediction of maize yield using alternative data coding algorithms”, Biosystems Engineering, 83(1), pp. 31-45, 2002.

[15] S. Puteh, M. Rizon, M. Juhari, J. Nor Khairah, S. Siti Kamarudin, B. Aryati, R. Nursalasawati, “Back propagation algorithm for rice yield prediction”, Proceedings of 9th of the Ninth International Symposium on Artificial Life and Robotics, Beppu, Japan, Oita, pp. 586-589, 2004.

[16] W. Ji and J. Cui, “Application of Geographical Information System (GIS) in Agricultural Land classification and Grading”, 2011 International Conference on Agricultural and Natural Resources and Engineering Advances in Biomedical Engineering, 3(5), pp. 201-205, 2011.

[17] B. Jietal, “Artificial neural networks for rice yield prediction in mountainous regions”, Journal of Agricultural Science, 145, pp. 249–261, 2007.

[18] T. Ranjeet and L. Armstrong, “An architecture of a decision support system for Western Australian Agriculture Industry”, Proceedings of the 9th Conference of the Asian Federation for Information Technology in Agriculture “ICT’s for future Economic and Sustainable Agricultural Systems”, Perth, Australia, 2014.

[19] S. Jabjone and S. Wannasang, “Decision Support System Using Artificial Neural Network to Predict Rice Production in Phimai District, Thailand”, International Journal of Computer and Electrical Engineering, 6(2), pp. 162-166, 2014.

[20] S. Dahikar and S. Rode, “Agricultural crop yield prediction using artificial neural network approach”, International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 2(1), pp. 683-686, 2014.

[21] Report on Economic Survey of Maharashtra 2012-2013, Directorate of Economics and Statistics, Planning Department, Government of Maharashtra, Mumbai (2013).

[22] Weka 3:Data Mining Software in Java, Machine Learning Group at the University of Waikato, Official Web: http://www.cs.waikato.ac.nz/ml/weka/index.html, accessed on 26th March 2016.

[23] A. Abraham, Artificial Neural Networks: Handbook of Measuring System Design, John Wiley & Sons, Ltd., 2005,ISBN: 0-470-02143-8.

[24] N.Gandhi, L.J. Armstrong and O. Petkar, “Predicting Rice Crop Yield Using Bayesian Networks”, communicated, 2016.

[25] N.Gandhi, L.J. Armstrong and O. Petkar, “Rice Crop Yield Prediction in India using Machine Learning Techniques”, communicated, 2016.

978-1-5090-0615-1/16/$31.00 ©2016 IEEE