ISSN No. 0976-5697

Volume 8, No. 5, May – June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Wheat Crop Yield Assessment Using Decision Tree Algorithms

Rupinder Singh
(Research Scholar)
Department of Computer Engineering
Punjabi University
Patiala, India

Gurpreet Singh
(Assistant Professor)
Department of Computer Engineering
Punjabi University
Patiala, India

Abstract: Prediction of agriculture yield with the help of data mining tools and techniques is one of the emerging research domains from past couple of years. This paper focuses on studying the effects of rainfall and relative humidity on three different stages of wheat crop and thus generating rules about crop yield using decision tree induction technique. A comparison of different classification algorithms is also done to discover the best rules about wheat crop yield with maximum accuracy and least errors. These study findings will help the farmers in future to predict their wheat crop yield in advance before crop growing season.

Keywords: Data mining; Wheat crop yield prediction; J48; Random Tree; Naïve Bayes.

I. INTRODUCTION

Nearly 43% of geographical area in India has been covered with the aid of agriculture area. Agriculture could be very vital occupation for us as it offers not only food to people but also provides raw materials to different industries. Agriculture sector performs very powerful and essential part in the Indian financial gadget. More than 75% of people living in India rely directly or indirectly upon on agriculture and agriculture related jobs for their livelihood [30]. During 2015-16, agriculture's percentage in Gross domestic Product (GDP) of India dropped to 17.4% as compared to 18.3% during 2013–14 [18]. Despite the decline in share of agriculture sector, yet agriculture is the biggest contributor to India's GDP. Estimation of different crop yields is one of major concerns among agriculture scientists and government policy makers in order to analyze development of agriculture sector in India.

According to United Nations survey report 2015, the recent population of India is about 1.3 billion [38]. To feed such a massive population, we require millions of tons food grains. Wheat is one of the major food grains in India. We should know what quantity of wheat we are able to have and what quantity we want to import or what amount we are able to export. For this purpose, we want to estimate wheat crop yield earlier in order to answer such questions.

Earlier than crop growing season, the primary purpose of a farmer is to realize about how much yield will be obtained beneath available circumstances. For such situation, there are different data mining techniques like classification, clustering, regression and association rule mining are available for agriculture yield estimation. Data mining plays very effective role in crop yield analysis and prediction. The main objective of data mining process is to uncover some meaningful patterns from given data. All these patterns discovered by different data mining techniques should be new, valid, useful and easily understandable to the users. An introduction of different data mining techniques that we used during our study is presented below.

A. EM (Expectation Maximization) Clustering

Clustering is one of data mining techniques which create clusters by grouping the similar items together. Items present in a cluster are way dissimilar to items present in other clusters. There are different methods used for clustering, a few are known as Partitioning based method, Density based method, Model based method, Hierarchical method and Grid based method [10]. Among them EM clustering method is used during our study. EM clustering is a Model based clustering technique which discovers the maximum likelihood parameters of a model. The equations present in such a model cannot be resolved in straight manner. Normally those models demand latent variables along with the regarded records observations and some unknown parameters that mean occasionally, there might be some missing values in given information [10]. EM clustering method evaluates probabilities of cluster memberships primarily based on one or more probability distributions rather than allocating instances to clusters for maximizing the variations in means for continuous variables. The aim of the EM clustering method is to maximize the general opportunity or probability of the records.

B. Classification

Classification is a method of organizing data into predefined categories with distinct class labels. Classification process consists of predicting a certain output based on some specific input [19]. Classification is one of the different techniques of data mining that assigns objects in a group to target categories. The intention of category is to correctly predict the goal class for every case within the records. There are different methods available for classification task, some of them are known as Bayesian classification, classification by Decision Tree Induction, Rule based classification, classification by Back Propagation, Associative classification etc. Under these methods various algorithms like J48, ID3, Random Tree, JRip, NNge, OneR, PART, Random Forest, Naïve Bayes, KStar etc. available for classification work. Among them J48, Random Tree and Naïve Bayes algorithms are used by us for wheat crop yield analysis during the study.

C. J48 Classifier

J48 classification algorithm is known as an optimized version of the C4.5 algorithm used in WEKA environment [40]. J48 also referred to as a statistical classifier that creates a decision tree from a fixed of training records with the use of measures like entropy and information gain [15]. The J48 classification technique recursively splits datasets consistent with checks on attributes in order to split the possible outcomes. The J48 algorithm uses the top down greedy method to create decision tree for classification [29]. The algorithm considers all of the viable checks which could break up a dataset and pick up a test which has the optimum information gain value. For every discrete variable, one take a look at with outcomes as many as the variety of different values of the attribute are considered. For every continuous variable, binary assessments concerning every different value of the variable are considered [40].

D. Random Tree Classifier

Random Tree models were considerably advanced in the discipline of machine learning within the current years [29]. With k random characteristics at every node, a random tree is a form of decision tree obtained at random from a set of feasible trees. In this situation, the word "random" means that every tree in the set of trees has a same threat of being sampled. Every other manner of announcing this context is that the distribution of all trees is "uniform". Random trees can be generated efficaciously and the aggregate of big units of random trees commonly results in precise models [40].One can also say that Random Tree is used for building a tree that considers k randomly selected attributes at every node. It does not perform pruning process and also has an option to permit approximation of probabilities of classes primarily based on extended set. This phenomenon is known as back-fitting [15].

E. Naïve Bayes Classifier

Naive Bayes is an unsophisticated approach for building classifiers models which are used for assigning class labels to problem instances represented in the form of vectors of characteristic values [12]. Naïve Bayes algorithm makes use of the normal distribution to model numeric attribute values. The probabilistic Naïve Bayes classifier is implemented by Naive Bayes algorithm. NB classifier uses kernel density approximates which expand performance if the normality hypothesis is generally accurate; it could also manage numeric values with the use of supervised discretization [23].

F. Introduction to Wheat

Wheat is grown all over the world in 217 million hectare area along with total production of 632 million tones. Wheat is the most essential meals-grain of India and is the staple food of thousands of Indian people. Cultivation of wheat has been carried out for thousand years in India. Cultivation of wheat crop has exhibited a sturdy boom fashion for the reason that onset of the green Revolution in 1968 [7]. Now, India is the second biggest producer country of wheat in the international market after China having approx 12% share in total global wheat production. In India, cultivation of wheat is carried out in an area of approximately 30 million hectare along with a production of 93 million tones [37]. Wheat is the second one most essential food grain in India subsequent to rice. Wheat is known as Rabi crop or winter crop. Wheat is sown within the starting of winter season and harvested in the starting of summers. In Punjab, it is sown in the month of November and harvested in the month of April.

G. Different Growth Stages of Wheat Plant

Wheat is the main food grain crop of Punjab. It requires almost 145 – 160 days for maturing [26]. Wheat growth stages can be broadly divided into several unique levels viz. germination, tillering, stem extension, head development, flowering, yield formation and ripening [35]. All these levels of wheat development are further combined into three main phases or stages which are classified as vegetative, reproduction and maturation [2] [34] [36]. These three stages have three main components that influence wheat crop yield. Those components are mentioned includes height of plant, number of productive tillers on plant and number of grains per spike [27].

1) Vegetative Period (Stage): The vegetative phase consists of approximately 65 to 75 days. This phase is characterized by a slow increase in the height of plant from germination till tillering [2] [34] [36]. Sub stages occur during the vegetative period are germination, seedling development, brairding and tillering [2] [13] [14].

2) Reproduction period (Stage): The reproduction stage consists of approximately 30 to 35 days. This phase is characterized by jointing, emergence of flag leaf, booting, culm elongation and flowering [20] [36]. The sub stages occur during the reproduction period includes stem elongation, booting, head emergence and flowering [13] [14].

3) Maturation and Ripening Period (Stage): The maturation and ripening stage consists of approximately 35 to 40 days. This phase is characterized by growth of grains, increase in gain size and grain weight [36]. The main sub stages occur during the maturation and repining period are milky stage, hard dough stage and ripening [2] [13].

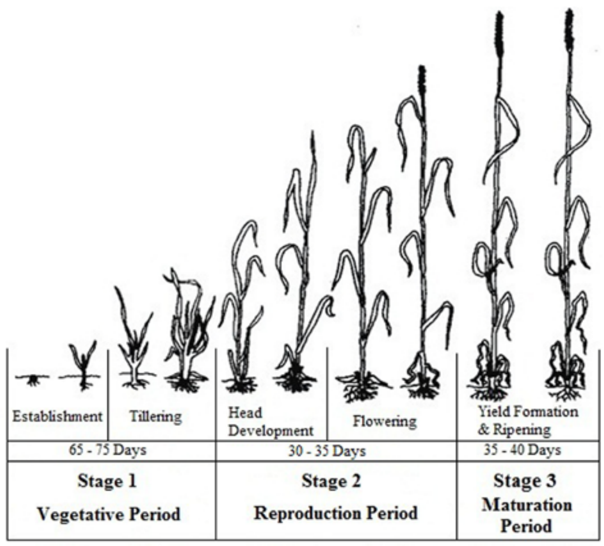

Figure 1. Different growth stages of wheat plant

H. Parameters Responsible for Variable Crop Yield

Crop yield may rely on many factors which can be unbiased of one another consists of geographical, meteorological, biological factors and financial policies. The crop yield also relies upon other factors like weeds, diseases, pests, time of harvesting etc [1]. Among all of them, meteorological parameters have our greater concern in regard with wheat crop yield as discussed below.

1) Temperature: Wheat requires cool weather for proper growth. At the time of sowing of wheat the best temperature is among 10°C to 15°C and at the time of harvesting, most suitable temperature range is 21°C to 26°C [8]. But if temperature remains below -10^0C for more than 12 hours then it would be difficult for wheat to survive under such extreme temperature conditions [31].

2) Rainfall: Wheat flourishes well in regions receiving annual rainfall of 75cm approximately. Annual rainfall should not be more than 100cm for wheat cultivation. At the time of harvesting, heavy rains cause damage to wheat crop yield [8]. Also during growing season too much rainfall can cause heavy rusts to the crop [26].

3) Relative Humidity: Relative Humidity is the ratio of water vapour content to the saturated water vapour content at a given point of temperature and pressure. It is expressed using percentage (%) sign. 50 to 60% humidity is finest for proper wheat growth but during maturation stage, it needs less humidity [33]. If relative humidity reaches below 50 % then it results in moisture loss and it outcomes in extra water requirement for irrigation. Excessive relative humidity along with high temperature results in disorder infestation [16].

II. Related Work

This section provides a short description of different data mining techniques used for analysis and prediction of various crops yield. A number of the most pertinent study findings are presented in this section along with authors' name.

Anshal Savla et al. [4] presented the survey of different classification techniques for formulating yield prediction accuracy in precision agriculture. Datasets from years 1961 to 2013 of soybean crop with parameters yield, seeds, area harvested and total production have been considered for study. Four classification techniques Neural Network, Support Vector Machines (SVM), Bagging and REP Tree have been applied for experimental work. Different performance measures have been used to calculate the performance of above four classification algorithms. Authors concluded that Bagging algorithm among four techniques of classification is most suitable for soybean crop yield prediction.

Ashwinirani and B. M. Vidyavathi [9] presented in their paper about the ameliorated methodology for the design of sugarcane yield estimation using Decision tree method. Different datasets consist of crop yield, soil type, seed type, fertilizer consumption, pest control techniques, land preparation methods, cropping pattern, labor utilization and different climate parameters have been collected for proposed work. PCA filtering and smoothing functions has been used for data preprocessing. By using decision tree algorithm they built up a model that can give almost accurate estimation of crop yield and can also identify those important variables that are responsible for variable crop yield.

A.T.S. Ahamed et al. [5] applied different data mining techniques to predict yield of major crops in various districts in Bangladesh. Datasets of rice, wheat and potato crop yields in 15 districts for two seasons have been used during study. A wide range of meteorological and biotic parameters were considered for experimental work. Different clusters have been formed with K Means clustering technique according to the type of parameters selected. Two datasets have been formed; one was training dataset and other was test dataset. Three different data mining techniques linear regression, K nearest neighbor (KNN) and neural network have been applied for the prediction of different crop yields. Accuracy of above applied techniques lie between 90 to 95 percent. However prediction of crop yields was not as accurate as expected by authors. It happens due to the use of small training dataset during study.

K. K. Pandey, V. N. Rai, B. V. S. Sisodia and S. K. Singh [22] discussed about the effects of climate variables on rice crop in eastern Uttar Pradesh. They stated that climate is one of the maximum essential elements influencing crop increase. Weather could have an effect on yield directly via affecting the crop growth structural traits of crop along with plant population, number of tillers leaf region and many others and circuitously through its impact on occurrence of pests and other diseases. They said that volume of weather influence on crop productions depends not only on magnitude of climate parameters, but additionally on their frequency distribution. They also stated that at the time of sowing of rice, low temperature results in poor germination, decreased tillering, early flowering and exposing the floral elements to the recent damage and thus reducing the crop yield. Step wise regression technique has been applied to measure the effect of climate parameters on rice yield. After the results they concluded that all the climate parameters have made tremendous effect on rice yield. The effect of climate variable varies with the different phase of crop yield.

N. Gandhi and L. J. Armstrong [24] presented the effect of seasonal rainfall on rice yield using association rule mining. Datasets of forty three years from 1960 to 2002 of 29 districts in Rajasthan state were considered for results generation. Whole kharif season is divided into three distinct divisions to analyze the impact of seasonal rainfall on rice yield in better way. Apriori algorithm has been applied for results generation by using WEKA tool. Results indicate that an average rainfall is good for rice yield. A very poor rainfall at the end of season may be responsible for poor rice yield.

N. Gandhi, O. Petkar, L. J. Armstrong and A. K. Tripathy [25] predicted rice crop yield in India with the help of Support Vector Machines (SVM). Datasets of 27 districts of Maharashtra state for years 1998 to 2002 were collected for analysis work. Sequential Minimal Optimization (SMO) classifier has been applied using WEKA tool to portray the relationship between different climate parameters and rice crop yield. Different metrics have been used to measure the performance of SMO classification algorithm on given dataset. After the analyses, it was found that other classification algorithms perform much better than SMO classifier on same dataset.

S. Shanmuganathan, P. Sallis and A. Narayanan [28] studied the impact of daily extreme weather conditions on grapevine and perennial crop yield in New Zealand with the help of different data mining techniques. Datasets from years 1997 to 2009 of maximum, minimum and grass minimum temperature on daily basis and annual grapevine yield data has been considered for experimental work. Self-Organizing Maps (SOM) and χ^2 test has been applied to generate some meaningful relationships between meteorological parameters and grapevine crop yield. Results obtained from SOM method shows relationship between daily maximum temperature intervals 14.1-23^{0}C from week 1 to 29 and >26^0 C from week 35 to 45 whereas χ^2 test depicts relationship between daily maximum temperature intervals <23^0 C for week 32 and 33 and >26^0 C for week 44 and 45 with low crop yield.

Y. Vagh and J. Xiao [39] discussed about the dual effect of rainfall and temperature on wheat crop yield in Australia. Data samples of rainfall and temperature were collected from years 2001 to 2010. The collected data samples of 10 years were averaged using R script individually. Data mining techniques regression and classification GP algorithm were applied using WEKA tool to find out the relationships between meteorological variables and wheat crop yield for selected time period. Results indicate a complex relationship between wheat crop yield and temperature as yield expected to be high at higher temperature and low at lower temperature. GP algorithm showed a positive correlation between temperature and wheat crop yield as yield expected to increase with an increase in temperature and decrease in rainfall. Authors also suggested further investigation to discover the influence of other factors like soil type on wheat crop yield in order to get better results.

Table I. Various techniques and parameters used for agriculture yield prediction

Sr. No.	***Authors***	***Crops Under Study***	***Parameters Under Study***	***Techniques Applied***
1	A. Savla et al. [4]	Soybean	Area harvested, seed	Neural Network, Support Vector Machines (SVM), Bagging, REP Tree,
2	Ashwinirani and B. M. Vidyavathi [9]	Sugarcane	Temperature, solar radiation, humidity, rainfall, soil type, seed variety, seed rate, fertilizer type, weed management, pest management etc.	PCA function filtering and Decision Tree
3	A.T.S. Ahamed et al. [5]	Rice, Wheat and Potato	Temperature, sun shine, humidity, rainfall, soil pH, soil salinity, irrigated area etc.	K Means Clustering, Linear Regression, K Nearest Neighbor (KNN) and Neural Network
4	K. K. Pandey, V. N. Rai, B. V. S. Sisodia and S. K. Singh [22]	Rice	Rainfall, wind velocity and sun shine	Step Wise Regression
5	N. Gandhi and L. J. Armstrong [24]	Rice	Rainfall	Apriori algorithm
6	N. Gandhi, O. Petkar, L. J. Armstrong and A. K. Tripathy [25]	Rice	Rainfall, minimum, average and maximum temperature	Support Vector Machines (SVM)
7	S. Shanmuganathan, P. Sallis and A. Narayanan [28]	Grapevine and Perennial	Maximum, minimum and grass minimum temperature	Self-Organizing Maps (SOM) and χ^2 test
8	Y. Vagh and J. Xiao [39]	Wheat	Temperature and rainfall	Regression and Classification function GP

III. MATERIALS AND METHODS

A. *Study Area*

The area selected for research work is Patiala district of Punjab state, India, which lies approximately between $29^0$49’ and $30^0$47’ North latitude and between $75^0$58’ and $76^0$54’ East latitude [11]. The weather here is standard of Punjab plain i.e. particularly warm in summer season and cold during winter season. The district is usually remains dry and hot during summer season except having a monsoon season of three months approximately starting from mid June to mid September. The district has very hot summer season starting from mid April till mid July and a very cold winter season starting from December to mid February. The district experiences annual average rainfall of 688 mm [6].

B. *Input Variables*

1) Wheat crop yield: Data of yield per hectare year wise of Wheat crop from year 1993 to 2016 in Patiala district was collected from Agriculture Department of Patiala.

2) Climate Data: Meteorological data with parameters Rainfall and Relative Humidity on daily basis for wheat growing season from year 1993 to 1998 of Patiala district was collected from India Meteorological Department, Meteorological Centre Chandigarh and remaining meteorological data on daily basis from year 2000 to 2016 was collected from India Meteorological Department, Meteorological Centre Punjabi University Patiala.

C. *Data Preprocessing*

Collected raw data of wheat crop yield and meteorological parameters has been put into Microsoft Office Excel for data cleaning and data preprocessing functions. Here missing values were omitted from raw data and data is stored in different files in Comma Separated values (CSV) format so that further analysis on data can be performed by using WEKA tool.

D. *Introduction to WEKA*

The full form of WEKA is Waikato Environment for Knowledge Analysis. It was developed at the University of Waikato in New Zealand. WEKA is written in Java programming language and dispensed below the phrases of the GNU General Public License. It can be worked on almost every system platform and has been examined under Microsoft Windows, Linux, and MAC operating systems. WEKA provides a consistent platform to various machine learning algorithms, together with techniques for data preprocessing and methods for comparing the results of various data mining techniques on a given dataset [17].

E. *Methodology Adopted*

Methodology of our study is mainly divided into three parts. First part includes the formation of clusters using EM clustering technique whereas second part includes the prediction of wheat crop yield with the help of decision tree classification method. Third part covers the performance evaluation of different classification algorithms using some specific metrics.

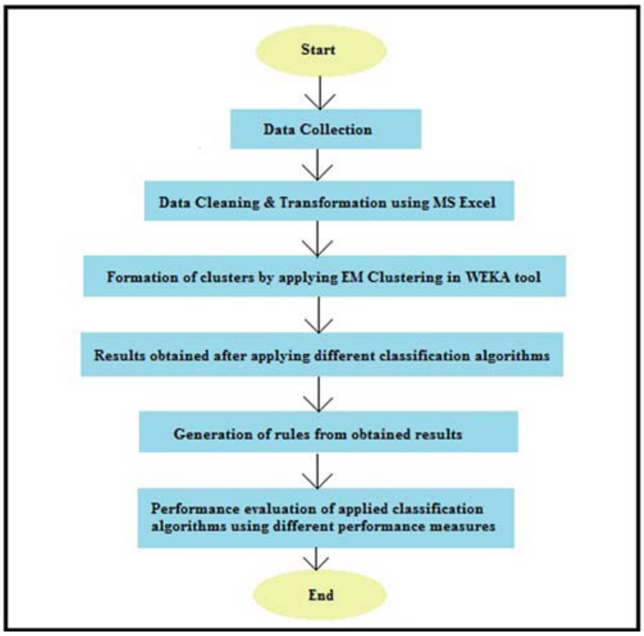

Figure 2. Methodology.

1) Clustering of Wheat crop yield and climate variables: EM clustering is applied to wheat crop yield data and relative humidity and rainfall occurred between three growth stages of wheat crop. Three growth stages of wheat plant are better known as vegetative period, reproduction period and maturation & repining period. In this situation, EM clustering performs discretization of data as number of clusters formed during the process represents low, medium and high values respectively.

2) Prediction of wheat crop yield using different classification methods: After the completion of discretization process, Decision Tree classifiers and Bayesian classifier has been applied to data for prediction of wheat crop yield. A couple of rules about wheat crop yield are generated by using Decision Tree Induction method.

3) Performance evaluation of classification algorithms: Various classifier accuracy measures for evaluating the performance of different classification algorithms used during study are explained below.

a) Accuracy: The accuracy of a classifier denotes the capability of a given classifier to properly predicting the class label of attributes without any class label. Accuracy of a classification algorithm described as the percentage of dataset attributes that are correctly classified by the classification algorithm [19].

b) Kappa Statistic: The Kappa statistics of inter observer agreement may be formulated to look into the selected styles or patterns of disagreement concurrently by way of selecting similar sets of weights which reflect the function of every response class in the provided agreement index [21]. The calculation of Kappa statistics works on principle of the difference between observed agreement and expected agreement [3].

c) Mean Absolute Error (MAE): The MAE measures the average upon the verification set of absolute values of the differences between predicted values and the corresponding observed values [32].

d) Root Mean Squared Error (RMSE): First of all every value of the difference between predicted values and actual values is squared and then average of all values is calculated over the given dataset and at the end, square root of that average is taken for the calculation of RMSE.

e) Relative Absolute Error (RAE): The RAE first measures the total absolute error of training data and then divides it by the total absolute error of the predictor for normalization.

f) Root Relative Squared Error (RRSE): The Relative Squared Error (RSE) first calculates the overall squared errors and then divides it by the total squared error of the predictor for normalization. We take the square root of the RSE to Find the RRSE.

IV. RESULTS

Results obtained after applying three different classification algorithms with the help of WEKA tool are presented in this section. Details of attributes representing meteorological parameters and wheat crop yield of Patiala district are presented below in the form of table.

Table II. Details of different attributes used during work

Sr. No.	*Attribute*	*Detail*
1	VR	Rainfall during vegetative period
2	RR	Rainfall during reproduction period
3	MR	Rainfall during maturation period
4	VH	Relative humidity during vegetative period
5	RH	Relative humidity during reproduction period
6	MH	Relative humidity during maturation period
7	WY	Wheat crop yield

EM clustering is applied to wheat crop yield data and relative humidity and rainfall occurred between three growth periods of wheat crop. With the help of clusters formed during EM clustering process, rainfall, relative humidity and wheat crop yield data are categorized into low, medium and high values.

1) Results obtained for rainfall and wheat crop yield by using J48 classification algorithm: A pruned decision tree was constructed using J48 classification algorithm taking into consideration of rainfall occurred at three growth stages of wheat crop and their influence on wheat crop yield.

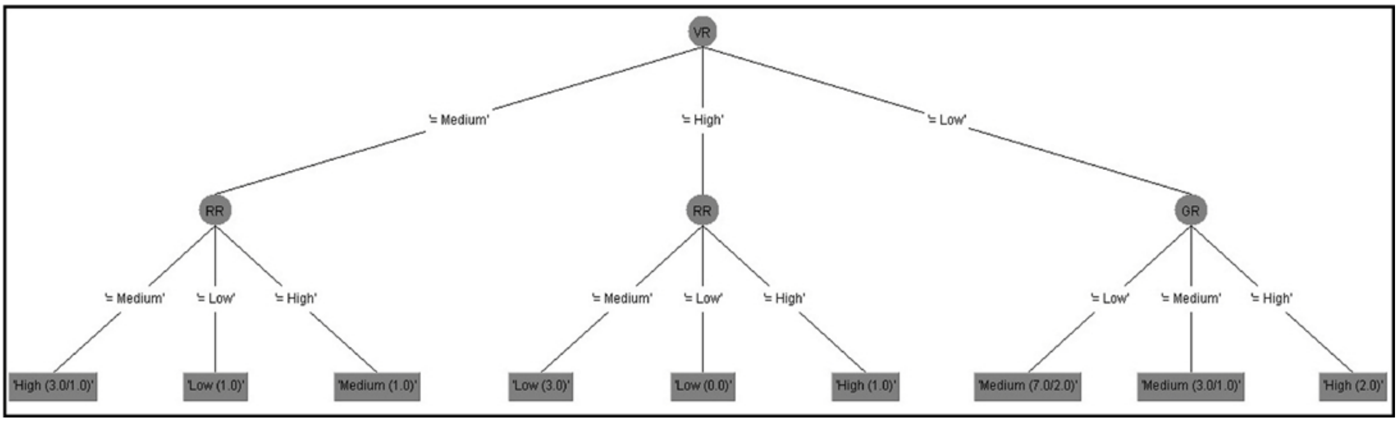

Figure 3. J48 pruned tree for rainfall and wheat crop yield.

J48 classifier achieves accuracy of 80.95% with kappa statistic of 0.71. It gives a mean absolute error (MAE) of 0.175, root mean squared error (RMSE) of 0.296, relative absolute error (RAE) of 39.69% and root relative squared error of 63.03%.

2) Results obtained for rainfall and wheat crop yield by using Random Tree classification algorithm: For rainfall and wheat crop yield dataset, Random tree classifier achieves accuracy of 80.95% with kappa statistic of 0.71. It gives a mean absolute error (MAE) of 0.169, root mean squared error (RMSE) of 0.291, relative absolute error (RAE) of 38.32% and root relative squared error of 61.93%.

3) Results obtained for rainfall and wheat crop yield by using Naïve Bayes classification algorithm: Naïve Bayes classifier achieves accuracy of 66.67% with kappa statistic of 0.49 on same dataset. It gives a mean absolute error (MAE) of 0.325, root mean squared error (RMSE) of 0.388, relative absolute error (RAE) of 73.59% and root relative squared error of 82.65%.

4) Results obtained for relative humidity and wheat crop yield by using J48 classification algorithm: Another pruned decision tree was constructed using J48 classification algorithm taking into consideration of percentage of relative humidity occurred at three growth stages of wheat crop and their influence on wheat crop yield.

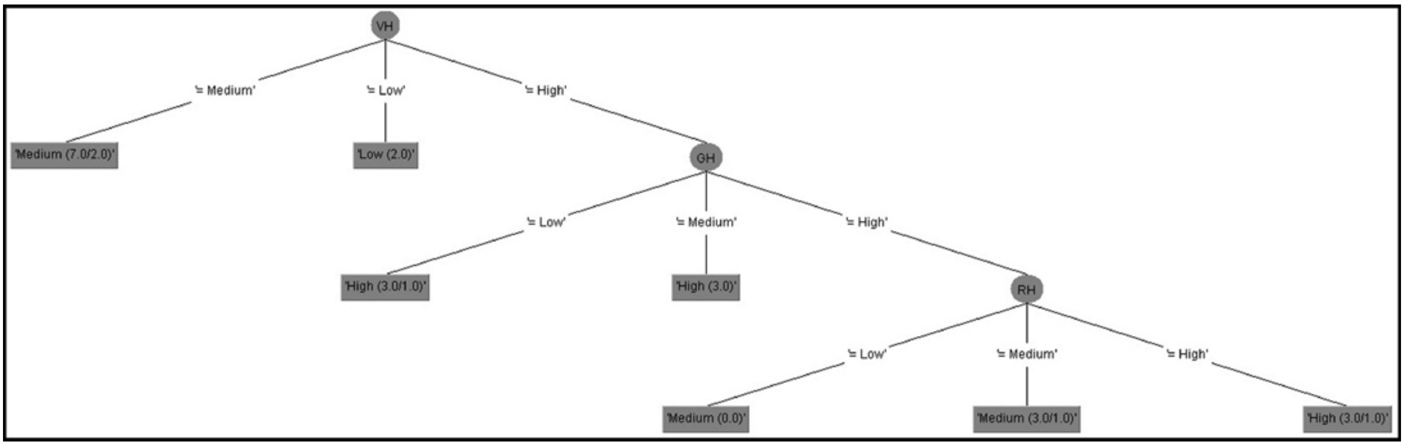

Figure 4. J48 pruned tree for relative humidity and wheat crop yield.

J48 classifier achieves accuracy of 76.19% with kappa statistic of 0.65. It gives a mean absolute error (MAE) of 0.217, root mean squared error (RMSE) of 0.329, relative absolute error (RAE) of 49.27% and root relative squared error of 70.22%.

5) Results obtained for relative humidity and wheat crop yield by using Random Tree classification algorithm: For relative humidity and wheat crop yield dataset, Random tree classifier achieves accuracy of 76.19% with kappa statistic of 0.65. It gives a mean absolute error (MAE) of 0.179, root mean squared error (RMSE) of 0.299, relative absolute error (RAE) of 40.72% and root relative squared error of 63.84%.

6) Results obtained for relative humidity and wheat crop yield by using Naïve Bayes classification algorithm: Naïve Bayes classifier achieves accuracy of 76.19% with kappa statistic of 0.64 on relative humidity and wheat crop yield dataset. It gives a mean absolute error (MAE) of 0.323, root mean squared error (RMSE) of 0.375, relative absolute error (RAE) of 73.03% and root relative squared error of 79.78%.

V. DISCUSSION

During present study, three classification algorithms have been used for wheat crop yield analysis and prediction. Number of rules generated by decision tree induction method to find out the influence of rainfall and relative humidity are listed below separately.

A. Rules Generated from J48 Pruned Tree for Rainfall and Wheat Crop Yield

1) IF VR = Medium AND RR = Medium THEN WY = High

2) IF VR = Medium AND RR = Low THEN WY = Low

3) IF VR = Medium AND RR = High THEN WY = Medium

4) IF VR = High AND RR = Medium THEN WY = Low

5) IF VR = High AND RR = High THEN WY = High

6) IF VR = Low AND MR = {Low, Medium} THEN WY = Medium

7) IF VR = Low AND MR = High THEN WY = High

Total seven rules are generated by J48 classification algorithm to detect the impact of rainfall occurred during three growth stages of wheat crop and its consequences on wheat crop yield. It is clear from above rules that rainfall occurred during vegetative period has more influence on wheat crop yield as compared to rainfall occurred during reproduction and maturation period. It is also found that high rainfall during vegetative and reproduction period leads to high wheat crop yield whereas low rainfall during vegetative and maturation periods results in medium wheat crop yield.

B. Rules Generated from J48 Pruned Tree for Relative Humidity and Wheat Crop Yield

1) IF VH = Medium THEN WY = Medium

2) IF VH = Low THEN WY = Low

3) IF VH = High AND MH = {Low, Medium} THEN WY = High

4) IF VH = High, MH = High AND RH = High THEN WY = High

5) IF VH = High, MH = High AND RH = Medium THEN WY = Medium

Total five rules are generated by J48 classification algorithm to detect the impact of relative humidity occurred during three growth stages of wheat crop and its consequences on wheat crop yield. It is clear from above rules that percentage of relative humidity occurred during vegetative period has more influence on wheat crop yield as compared to reproduction and maturation period. It is also found that medium percentage relative humidity during vegetative period leads to medium wheat crop yield whereas low percentage relative humidity during vegetative period leads to low wheat crop yield. However, high value of relative humidity throughout whole wheat growing season leads to high wheat crop yield.

A comparison of all thee algorithms used during study is also done by using different metrics.

C. Comparison of Classification Algorithms used for Meteorological Parameters and Wheat Crop Yield Data Analysis

Table III. Comparison Based On Classifier Performance Measures (For Rainfall and Wheat Crop Yield)

Sr. No.	*Classifier*	*Accuracy*	*Kappa Statistic*	*MAE*	*RMSE*	*RAE*	*RRSE*
1	J48	80.95%	0.71	0.175	0.296	39.69%	63.03%
2	Random Tree	80.95%	0.71	0.169	0.291	38.32%	61.93%
3	Naïve Bayes	66.67%	0.49	0.325	0.388	73.59%	82.65%

Table IV. Comparison Based On Classifier Performance Measures (For Relative Humidity and Wheat Crop Yield)

Sr. No.	*Classifier*	*Accuracy*	*Kappa Statistic*	*MAE*	*RMSE*	*RAE*	*RRSE*
1	J48	76.19%	0.65	0.217	0.329	49.27%	70.22%
2	Random Tree	76.19%	0.65	0.179	0.299	40.72%	63.84%
3	Naïve Bayes	76.19%	0.64	0.323	0.375	73.03%	79.78%

After the performance evaluation, Random Tree has proved to be most accurate among all three classification algorithms used for study.

VI. CONCLUSION

The current work has been based on detecting the influence of rainfall and relative humidity on wheat crop yield using decision tree algorithms. By using decision tree classification technique, some useful information has been extracted in the form rules from provided data of meteorological parameters and wheat crop yield. Classification techniques that have been used during study portray some relationships between climate parameters and wheat crop yield. The salient conclusions drawn from the work done are:

- Decision tree results indicated that rainfall and relative humidity has more influence over wheat crop yield during vegetative period as compared to reproduction and maturation period.
- Rules generated from decision tree analysis will help the users to predict the conditions responsible for variable wheat crop yield under given meteorological parameters.
- After the performance evaluation of three classifiers used during study, it can be concluded that Random Tree classification algorithm is most suitable for wheat crop yield prediction under current dataset.

VII. FUTURE ENHANCEMENTS

In future, the present study can be extended upon following basis:

- More number of attributes such as soil salinity, soil pH, area harvested, fertilizer consumption, pesticides consumption, temperature, solar radiation etc. can be used to form better wheat crop yield prediction model.

- Performance of these classification algorithms can be evaluated on some bigger test datasets in order to get better results.

REFERENCES

[1] A. A. Raorane and R. V. Kulkarni, "Review- Role of Data Mining in Agriculture," International Journal of Computer Science and Information Technologies, ISSN: 0975-9646, Vol. 4, No. 2, pp. 270–272, 2013.

[2] A. Bauer, D. Samika and A. Black, "Correlation of Five Wheat Growth Stage Scales Used in the Great Plains," Agriculture Research Service, ISSN: 0193-3701, pp. 1-16, 1983.

[3] A. J. Viera and J. M. Garrett, "Understanding Inter Observer Agreement: The Kappa Statistic," Family Medicine, Vol. 37, No. 5, pp. 360-363, May 2005.

[4] A. Savla et al., "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," 2nd International Conference on Innovations in Information,Embedded and Communication systems, 2015.

[5] A. T. S. Ahamed et al., "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh," IEEE SNPD, Takamatsu, Japan, June 2015.

[6] About Patiala. [Online]. Available: http://patiala.gov.in/html/about_the_district.htm

[7] Agriculture. [Online]. Available: www.iisc.ernet.in/insa/ch21.pdf

[8] Agriculture in India. [Online]. Available: www.nos.org/media/documents/SecSocSciCour/English/Lesson-12.pdf

[9] Ashwinirani and B. M. Vidyavathi, "Ameliorated Methodology for the Design of Sugarcane Yield Prediction Using Decision Tree," COMPUSOFT: An International Journal of Advanced Computer Technology, ISSN: 2320-0790, Vol. 4, Issue 7, pp. 1882-1889.

[10] B. S. Binni and T. Mathew, "Clustering and Regression Techniques for Stock Prediction," Procedia Technology, Vol. 24, pp. 1248-1255, 2016.

[11] DBAS – Patiala (Punjab). [Online]. Available: https://cdm.unfccc.int/Projects/DB/DNVCUK1179324777.88/ReviewInitialComments/OLEJNKZCUBK3FLFPWKGZNFB3NM64L0

[12] D. Sindhuja and R. J. Priyadarsini, "A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder," International Journal of Computer Science and Mobile Computing, ISSN: 2320–088X, Vol. 5, Issue 5, pp. 483-488, May 2016.

[13] Growth and Development Guide for Spring Wheat. [Online]. Available: http://www.extension.umn.edu/agriculture/small-grains/growth-and-development/spring-wheat/

[14] Growth Stages of Wheat. [Online]. Available: https://www.usask.ca/agriculture/plantsci/winter_cereals/winter-wheat-production-manual/chapter-10.php

[15] H. Chauhan, V. Kumar, S. Pundir and E. S. Pilli, "A Comparative Study of Classification Techniques for Intrusion Detection," International Symposium on Computational and Business Intelligence, pp. 40-43, 2013.

[16] Humidity – Climatic Requirement for Wheat. [Online]. Available: http://agropedia.iitk.ac.in/content/humidity-climatic-requirement-wheat

[17] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques. 2nd ed., San Francisco, CA: Morgan Kaufmann, 2005.

[18] India Economic Survey 2015-16 – Key Highlights. [Online]. Available: https://home.kpmg.com/content/dam/kpmg/pdf/2016/04/KPMG-Flash-News-India-Economic-Survey-2015-16%E2%80%93Key-Highlights-3.pdf

[19] J. Han and M. Kamber, Data Mining: Concepts and Technologies. 2nd ed., San Francisco, CA: Morgan Kaufmann, 2006.

[20] J. R. Haun, "Visual Quantification of Wheat Development," Agronomy Journal, Vol. 65, pp. 116-119, January 1973.

[21] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," Biometrics, Vol. 33, No. 1, pp. 159-174, March 1977.

[22] K. K. Pandey, V. N. Rai, B. V. S. Sisodia and S. K. Singh, "Effect of Weather Variables on Rice Crop in Eastern Uttar Pradesh, India," Plant Archives, ISSN: 0972-5210, Vol. 15, No. 1, pp. 575-579.

[23] M. N. Amin and M. A. Habib, "Comparison of Different Classification Techniques Using WEKA for Hematological Data," American Journal of Engineering Research, ISSN: 2320-0936, Vol. 4, Issue 3, pp. 55-61, 2015.

[24] N. Gandhi and L. J. Armstrong, "Assessing impact of seasonal rainfall on rice crop yield of Rajasthan, India using Association Rule Mining," International Conference on Advances in Computing, Communications and Informatics, Jaipur, India, pp. 1021-1024, September 2016.

[25] N. Gandhi, O. Petkar, L. J. Armstrong and A. K. Tripathy, "Rice Crop Yield Prediction in India using Support Vector Machines," 13th International Joint Conference on Computer Science and Software Engineering, 2016.

[26] Package of Practices for Crops of Punjab – Rabi. [Online]. Available: www.kvkfaridkot.com/pp_rabi_2013.pdf

[27] R. Dehgahi, A. Joniyas and M. D. Latip, "Rainfall Distribution and Temperature Effects on Wheat Yield in Torbate Heydarie," International Journal of Science Research in Knowledge, ISSN: 2322-4541, 2(Special Issue), pp. 121-126, 2014.

[28] S. Shanmuganathan, P. Sallis and A. Narayanan, "Data Mining Techniques for Modelling the Influence of Daily Extreme Weather Conditions on Grapevine, Wine Quality and Perennial Crop Yield," Second International Conference on Computational Intelligence, Communication Systems and Networks, pp. 90-95, 2010.

[29] S. Subramanian, V. B. Srinivasan and C. Ramasamy, "Study on Classification Algorithms for Network Intrusion Systems," Journal of Communication and Computer, Vol. 9, pp. 1242-1246, November 2012.

[30] S. Thenmozhi and P. Thilagavathi, "Impact of Agriculture on Indian Economy," International Research Journal of Agriculture and Rural Development, ISSN: 2319-331X, Vol. 3, No. 1, pp. 96-105, December 2014.

[31] S. Veenadhari, "Crop Advisor: A Software Tool for Forecasting Paddy Yield," Bonfring International Journal of Data Mining, ISSN: 2277-5048, Vol. 6, No. 3, pp. 34-38, July 2016.

[32] Weka Data Analysis. [Online]. Available: www.cs.usfca.edu/~pfrancislyon/courses/640fall2015/WekaDataAnalysis.pdf

[33] WHEAT. [Online]. Available: http://nsdl.niscair.res.in/jspui/bitstream/123456789/502/1/WHEAT%20-%20formatted.pdf

[34] Wheat. [Online]. Available: http://www.fao.org/land-water/databases-and-software/crop-information/wheat/en/

[35] Wheat Growth Stages. [Online]. Available: http://prairiecalifornian.com/wheat-growth-stages/

[36] Wheat Production Handbook. Kansas State University, Manhattan, May 1997.

[37] Wheat – The Crop for Ensuring Indian Food Security. [Online]. Available: farmer.gov.in/imagedefault/pestanddiseasescrops/wheat.pdf

[38] World Population Prospects – Key Findings & Advance Tables. [Online]. Available:https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf

[39] Y. Vagh and J. Xiao, "A Data Mining Perspective of the Dual Effect of Rainfall and Temperature on Wheat Yield," International Journal of Computer and Communication Engineering, Vol. 1, No. 4, pp. 358-364, November 2012.

[40] Y. Zhao and Y. Zhang, "Comparison of Decision Tree Methods for Finding Active Objects," Advances in Space Research, Vol. 41, pp. 1955–1959, 2008.