Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 4777 – 4787

doi:10.3906/elk-1612-361

A predictive modeling approach for improving paddy crop productivity using data mining techniques

Anitha ARUMUGAM*
Department of Information Technology, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

Received: 31.12.2016 • **Accepted/Published Online:** 03.10.2017 • **Final Version:** 03.12.2017

Abstract: Agriculture has a great impact on the economy of developing countries. To provide food security for people, there is a need for improving the productivity of major crops. Rapidly changing climatic conditions and the cost of investment in agriculture are major barriers for small-holder farmers. The proposed research aims to develop a predictive model that provides a cultivation plan for farmers to get high yield of paddy crops using data mining techniques. Unlike statistical approaches, data mining techniques extract hidden knowledge through data analysis. The data set used in this research for mining process is real data collected from farmers cultivating paddy along the Thamirabarani river basin. K-means clustering and various decision tree classifiers are applied to meteorological and agronomic data for the paddy crop. The performance of various classifiers is validated and compared. Based on experimentation and evaluation, it has been concluded that the random forest classifier outperforms the other classification methods. Moreover, classification of clustered data provides good classification accuracy. The outcome of this research is the identification of different combination of traits for achieving high yield in paddy crop. The final rules extracted by this research are useful for farmers to make proactive and knowledge-driven decisions before harvest.

Key words: K-means clustering, decision tree induction, crop productivity, classification accuracy, cultivation plan

1. Introduction

Forecasting crop productivity is a scientific technique of predicting crop yield before harvest. Research on crop productivity is essential to ensure adequate availability of food throughout the year irrespective of climate and market dynamics. To retain existing farmers and to acquire new farmers, there is an urgent need for developing agricultural risk management models. This research aims to develop a prediction model that protects farmers from agricultural risks. In this paper, a framework that helps scientific decision making in agriculture is proposed.

Using this model, farmers can plan the cultivation process well in advance. To prevent loss, farmers can identify suitable combinations of traits like soil nutrition, seed quality, water availability, and amount of fertilizer. It is a scientific model that provides suitable cultivation plans to farmers in accordance with the changing agronomic factors. Existing models can only perform crop yield prediction by statistical analysis [1]. Therefore, there is an urgent need for scientific analysis and reporting. The proposed model is based on data mining techniques. Data mining is a technique of extracting hidden, useful, and interesting patterns from raw data. In this work, discovery of knowledge regarding suitable cultivation plan is done in various stages such as data cleaning, integration, attribute selection and transformation, data mining, pattern evaluation, and knowledge presentation.

*Correspondence: dr.aanitha@yahoo.com

Paddy is the most important crop in south India. Yield of paddy crop depends on various meteorological and agronomic factors such as seed quality, amount of rainfall, soil fertility, crop rotation, land preparation, type of fertilizer, and amount of fertilizer. In order to assess the relationship between these factors and crop yield and to identify the input variables influencing the productivity of paddy, a real data set collected from farmers cultivating paddy along the Thamirabarani river basin is used in this research.

Raw agricultural data are preprocessed and only the essential factors are identified by filtering. The major data mining techniques used in this research are K-means clustering and decision tree classifiers such as decision stump, j48, REPTree, and random forest. Performances of these algorithms are compared based on classifier accuracy measures. The final knowledge regarding the cultivation plan is discovered, evaluated, and presented.

Integration of clustering along with classification is the key feature of this research. Clustering eliminates the outliers. Outliers are the representatives for degradation in quality of final knowledge. Since outliers are eliminated by clustering, only the relevant instances are considered for classification process.

2. Literature survey

Uno et al. [2] proposed an artificial neural network (ANN)-based framework for predicting corn yield. In this study the growth of crop under study was monitored for different amounts of nitrogen application and with different weed control methods. The predictive power of various machine learning algorithms like ANN and step-wise multiple linear regression (SMLR) is compared. Airborne hyperspectral imagery data of corn plots in Eastern Canada have been used for analysis. Performance of ANN and SMLR is found almost equal. A prediction error rate of 20% root mean squared error (RMSE) has been obtained. Therefore, the authors suggested a need for future work to reduce the prediction error rate. The data set used for experimentation is obtained from a single corn cultivator.

El-Telbany and Warda [3] proposed a methodology using the C4.5 decision tree algorithm to discover classification rules for Egyptian rice diseases. Only 7 agronomic attributes are considered for experimentation. Entropy and gain are used as attribute selection measures. The performance of the neural network and C4.5 decision tree algorithm is compared. Accuracy attained with the neural network and C4.5 algorithm is 96.4% and 97.25%, respectively.

Mucherino et al. [4] applied data mining techniques such as K-means, K nearest neighbor, ANN, and support vector machines in the field of agriculture. K-means has been proposed for predicting wine fermentation into five clusters namely blue, red, pink, brown, and green. Data from a winery in Chile's Maipo Valley have been used. Results for good and bad fermentation are 67% and 33%, respectively. Neural networks are used for classifying sounds made by pigs into classes such as coughs, metal clanging, grunting, and noise. The data set used for experimentation had 354 sounds. The data set was created by an expert farmer through experimental observation. Accuracy of 90% is achieved for classification of sounds by pigs.

Nasrin Fathima and Geethar [5] proposed a data mining methodology for decision making in agriculture. A data set collected from the agriculture department, Perambalur, has been used for mining process. Techniques such as K-means clustering, a priori algorithm, kNN classifiers, and correlation measures are used for identifying suitable land, crop variety, crop name, etc. to get high yield. Accuracy of 76% is achieved.

Diriba and Borena [6] proposed a technique for predicting crop productivity. The data set used for analysis has been obtained from statistical analysis and research department of land tenure system in Ethiopia. Totally, 11 agronomic attributes have been identified for experimentation. Classification algorithms such as j48, REPTree, and random forest have been applied over the Ethiopian agricultural data set. Prediction accuracy of 83% is attained. Moreover, amount of fertilizer is identified as an important factor influencing crop yield.

Tomidal and Yamaguchi [7] developed a different application of decision trees for identifying student voltage drop. Data set used for research was obtained for 6 years from various institutions such as University of Narino, Colombian Institute for Development of Higher Education, Colombian National Registry of Civil Status, National Bureau of Statistics, etc. Hidden causes for drop out of students from higher education have been identified. Results proved the confidence level as 85%.

Papageorgiou et al. [8] developed a fuzzy cognitive map-based approach to improve the productivity of cotton crop. A causal relationship between soil properties and cotton yield is identified and represented as a cause–effect relationship. A data set has been created using 6 years' data from the cotton fields of Central Greece. The FCM model suggested has nodes representing factors such as texture and organic matter, and directed edges connecting them. It predicted the yield of cotton into two classes: low and high. Classification accuracy for 6 years ranges from 67% to 78%.

Amara Singha et al. [9] developed a simulation model and predicted crop and water productivities by considering climatic factors. Data were collected from the Field Crops Research and Development Institute, Sri Lanka. The authors concluded that supplementary irrigation enhances rice productivity over usual rain-fed conditions. Finally the simulated productivity of Oryza rice crop is compared with actual productivity. The results proved that simulated yield has 95% confidence over observed yield.

Ashwinirani and Vidyavathi [10] applied PCA and decision tree classifier for predicting yield of sugarcane. By this framework, farmers can identify faults at an early stage to reduce loss. In this research basic agricultural parameters and weather data are considered for analysis.

A predictive model for agricultural yield was developed by Marinkovic et al. [11]. In this work data mining techniques are applied over an agricultural data set and a prediction model has been developed for forecasting annual yield of maize, soya bean, and sugar beet. The data set used for this research is collected from main fields in the Serbian Province of Vojvodina for 10 years from 1999 to 2008. An improved model tree for soya bean was built in three ways: (i) without attribute selection, (ii) with attribute selection using best first search, (iii) with attribute selection using a genetic algorithm. The highest accuracy is attained with a genetic algorithm, which has RMSE of 30%.

In a prediction model developed by Kashish et al. [12] a comparative analysis on the performance of various classifiers over a dengue data set was made. Experimentation was done using 3 Weka interfaces: Explorer, Experimenter, and Knowledge flow. The dengue data set used in this research had 108 instances and 18 attributes. Knowledge flow analysis was performed to comparatively study the performance of various classifiers like naïve Bayes, j48, and REPTree. Naïve Bayes and j48 classifiers produced 100% accuracy.

A predictive model has been developed by Ramya and Lokesh [13] to improve the productivity of mulberry crop. The authors developed an automated system for analyzing soil characteristics for increasing productivity of mulberry and cocoons. The data set used for this research was created by conducting survey in major districts of Mysore and Bangalore. Soil testing was done at the Sericulture R&D Institute, Bangalore. Fertility of the soil was checked and ideal values were identified by applying various classification algorithms such as decision stump, j48, random forest, and REPTree. Accuracies attained by these classifiers were 75%, 100%, 100%, and 99%, respectively.

Veenahadri et al. [14] proposed a decision tree-based model for improving the yield of soya bean crop. It has been determined that humidity, temperature, and rainfall are the major factors influencing productivity of soya bean. Agricultural parameters are not considered for analysis. The data set used for this research was climatic data influencing soya bean productivity in Bhopal district. The outcome of this work is a decision tree depicting values for different parameters to get high yield and low yield of soya bean. These rules were better understood by farmers.

In summary, while applying neural networks and K-means, only 90% prediction accuracy was achieved. kNN classifiers produced 76% accuracy. REPTree has produced 83% accuracy on prediction of Ethiopian crop productivity. Fuzzy cognitive map (FCM)-based method and decision stump classifier produced only 75% prediction accuracy. J48, naïve Bayes, and random forest when applied for dengue disease prediction and student voltage drop prediction produced prediction accuracy of 99% and 100%. Therefore, it is proposed to use classification techniques such as j48, REPTree, and random forest in this research, as these are efficient in terms of prediction accuracy and interpretability over other techniques such as linear regression, neural networks, and kNN classifiers. From the survey of all works discussed above, it is inferred that no significant study was conducted to provide a complete cultivation plan involving soil characteristics, seed characteristics, rainfall, and agronomic variables to predict the yield of paddy crop. This research is aimed to provide complete guidance to farmers at an early stage. Data mining techniques such as clustering and classification are used to predict the yield of crop. Performances of various decision tree-based classification techniques are compared, the results are validated, and a final solution is provided as knowledge easily interpretable by farmers.

3. Proposed system

Statistical methods involve determining explicit information only. They do not involve any validation mechanism. However, data mining methods can discover implicit knowledge through data analysis. In this research, forecasting the crop yield and identification of ideal condition for getting high yield of paddy crop is carried out scientifically by a data mining approach.

The predictive modeling framework proposed in this research involves 6 major steps as shown in Figure 1.

3.1. Data

The proposed predictive model is implemented in WEKA (Waikato Environment for Knowledge Analysis) software. The agricultural data set involving attributes such as seed quality, soil fertility, land type, and crop variety is collected from farmers cultivating paddy along the Thamirabarani river basin.

It is a real data set constructed from 200 questionnaires distributed to farmers having paddy fields in Tirunelveli and Tuticorin districts irrigated by the Thamirabarani river.

3.2. Data preprocessing

Preprocessing is a technique to improve the quality of data to be presented to the mining process. High quality input data will provide highly useful and interesting knowledge. In the proposed system, data preprocessing is done in 4 major ways: (i) data cleaning, (ii) attribute selection, (iii) transformation, and (iv) integration.

3.2.1. Data cleaning

Data cleaning is a method of replacing incomplete, inconsistent, and noisy data. Some attributes have null data or missing data. To obtain high quality knowledge, cleaning is done by eliminating null values.

Attributes with few null values are updated by finding average value (e.g., amount of fertilizer). This is done by using a filter in WEKA - filters.unsupervised.attributes. replacemissingvalue. Attributes with too many null values (e.g., type of pesticide) are removed using a filter in WEKA - filters. unsupervised.attributesremove.

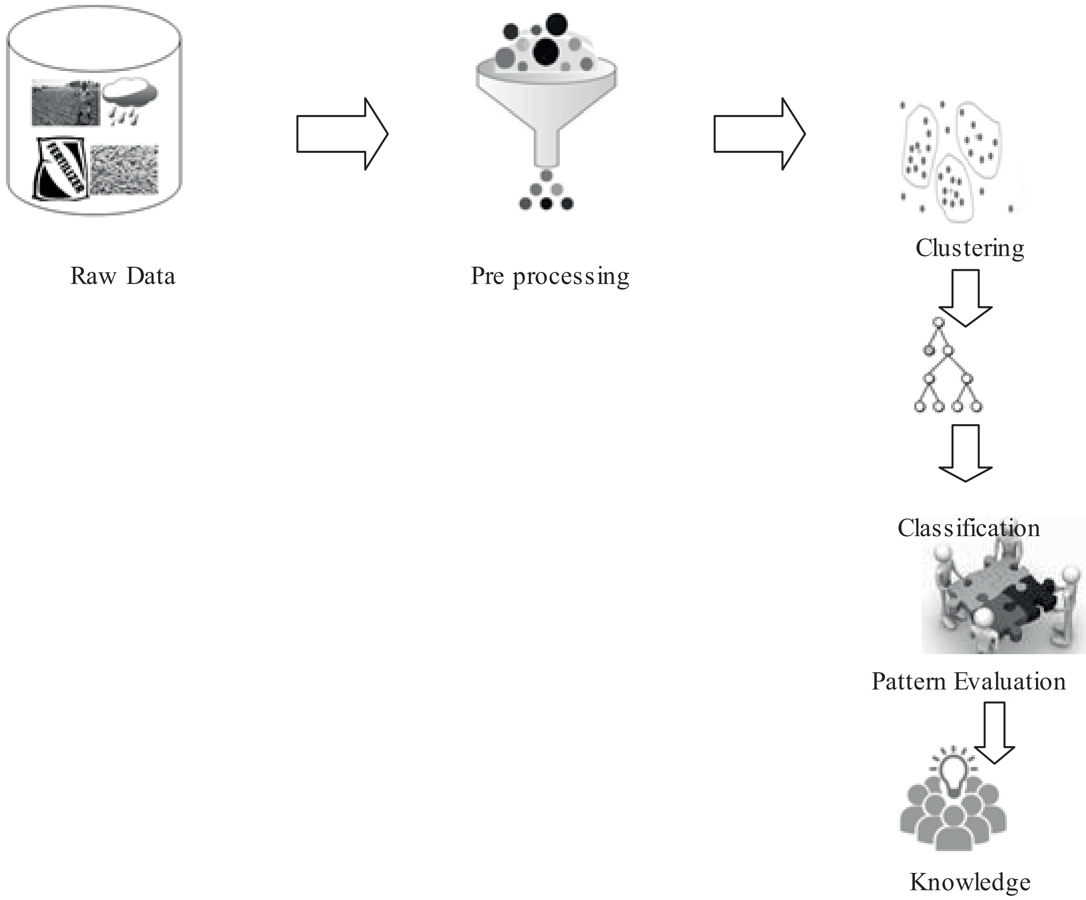

Figure 1. Proposed predictive modeling framework.

3.2.2. Attribute selection

Attributes that contribute more to mining process are identified in this process. Attribute selection is done by eliminating irrelevant and redundant attributes. Using an attribute selection filter in WEKA, attributes are arranged based on information gain.

Attributes with very low scores for information gain such as crop spacing, irrigation method, and seed quality are removed by attribute selection. The input parameters used in this research are listed in Table 1.

3.2.3. Transformation

The process of converting the data to the form suitable for mining task is called transformation. Attributes with numeric values are converted to categorical attributes (e.g., temperature) using Filter.unsupervised.attributes. numerictonominal option in WEKA.

3.2.4. Data integration

The agricultural data obtained through questionnaires distributed over farmers of different regions along Thamirabharani river basin are integrated.

Table 1. List of selected attributes.

S. no.	Attribute	Values	Description
1	Type Of Soil	{Black Soil,Red Soil,Alluvial Soil,Clay}	Nature Of Soil
2	Crop Variety	{Ambai-16,Ir-20,Aduthurai-45,KarnatakaPonni,AST-16,Tk,Samba,Arikkinaari,BPT-5402,COVAI-45}	Type of paddy
3	Seed Quality	{Medium,Good}	Same crop variety can be cultivated using seeds of different quality
4	Seed Rate	{<=25,26–30,>30}	Price of seed per kg
5	Season	{Kaar,Advanced Kaar,Pishanam}	Period of cultivation
6	Amount Of Fertilizer Used Per Acre	{Numeric - In Kg Per Acre}	Total amount of fertilizer used from sowing to harvest
7	Amount Of Rainfall	{Low,Medium,Good}	Amount of rainfall during cultivation period
8	Land Preparation Method	{Tillage,Planting}	Procedure followed for land preparation before sowing
9	Sowing Procedure	{Manual,Mechanical,Kulimuthu}	Methodology followed for sowing
10	Fertilizer Type	{Dung,Complex,Urea,Pottasium, Sulphide}	Type of fertilizer
11	Amount Of Pesticides Used Per Acre	{Numeric - In Liters Per Kg}	Total amount of pesticide used from sowing to harvest
12	Crop Rotation	{Urad,Cucumber,Banana,Cotton, Groundnut,Chilli}	Crop cultivated for rotation purpose
13	Natural Manure Added	{Yes,No}	Addition of organic matters
14	Soil Fertility	{Nitorgen,Phosphorus,Potassium-Low,Medium,High}	Amount of minerals present in the soil
15	Temperature	{Low,Moderate,High}	Amount of temperature
16	Yield Per Acre	{Good,Not Good}	If productivity >2000 kg yield is good, otherwise not good

3.3. Clustering

Clustering is the process of grouping data objects such that objects in the same cluster are highly similar and objects belonging to different clusters are highly dissimilar. There are high intracluster similarities and low intercluster similarities. Clustering presents high quality and useful patterns for subsequent phases of knowledge discovery by eliminating outlier objects. Outliers are objects whose behavior does not comply with the general behavior of other data objects. These are representatives for degrading the quality of mined knowledge. Therefore, it is very much essential to do clustering before classification.

In the proposed work, K-means clustering is performed to identify useful patterns for the classification task. The output of the clustering phase will be a set of closely related cultivation patterns.

3.3.1. K-means clustering

1. Fix the number of clusters k.
2. Randomly fix k data objects as initial cluster centroids.
3. (Re)assign each data object to the cluster, to which its Euclidean distance with the cluster centroid is minimum.

4. (Re)compute the cluster centroid.
5. Repeat steps (3&4) until cluster centroid converge.

If k value is too high, number of clusters will be more. It will lead to the elimination of many useful objects as outliers, thereby reducing the usefulness of final knowledge. Instead, if k value is set too small, number of clusters will be very less. It causes outliers to be included as part of the cluster. These outliers, if considered for mining process will affect the accuracy of results. Therefore, the initial value of k is set as 2, using domain knowledge. Cluster 0 is a set of values that provide good yield. Cluster 1 is set of values that provide low yield. Euclidean distance is used as distance measure. By implementing k-means clustering, the data objects are clustered.

3.4. Classification

Classification is a supervised learning process by which data objects are grouped into classes of known labels. It involves two phases: learning phase and classification phase. In the learning phase, training data are analyzed and a classifier model is built. In the classification phase, the test data are used to estimate the accuracy of classification.

In this research, a popular classification technique called decision tree induction is used to predict the classes. In decision tree induction, classification is done in top down fashion. The decision tree is a flowchart-like tree structure in which interior nodes represent the test condition or splitting criterion, edges represent the outcome of the test, and leaves represent class labels. All the tuples represented by leaf node are as pure as possible, i.e. they must belong to the same class. Based on the values of agronomic and meteorological parameters, class labels can be of low yield, medium yield, and high yield. Cultivation plans are extracted as rules from decision trees. They are easy to interpret and understand by farmers. The parameters considered for experimentation involve both numeric and nominal data. Decision trees can efficiently handle these data.

Decision tree induction is a two-step process. In the first step the decision tree is constructed. In the second step it is pruned. Decision tree construction is shown in Figure 2.

The training set and test set are separated by 10-fold cross validation. In this method, the given data set is divided into 10 blocks. Initially, the first 9 blocks are used as a training set for constructing the classifier model and the last block is used for testing. This process is repeated for other blocks, such that each block is used as a test set exactly once. The final performance is measured by finding the average [4]. Since the average is taken over 10 partitions of data, it lowers variance and predictive error rate.

Decision trees can represent knowledge in a visually appealing and descriptive manner. They perform classification by decision analysis [6]. There are three types of decision trees. If categorical values are predicted, it is called a classification tree. Instead, if numeric values are predicted, it is called a regression tree. If both categorical and numeric values are predicted, it is called a classification and regression tree [3].

In this research, information gain is used as an attribute selection measure. It is the difference between actual information required to the information needed to classify the tuple using attribute A.

$$Gain = Info\,(D) - Info_A(D) \tag{1}$$

$$Info\,(D) = -\sum_{i=1}^{m} p_i log_2(p_i) \tag{2}$$

$$p_i = \frac{|C_{i,D}|}{|D|}, \tag{3}$$

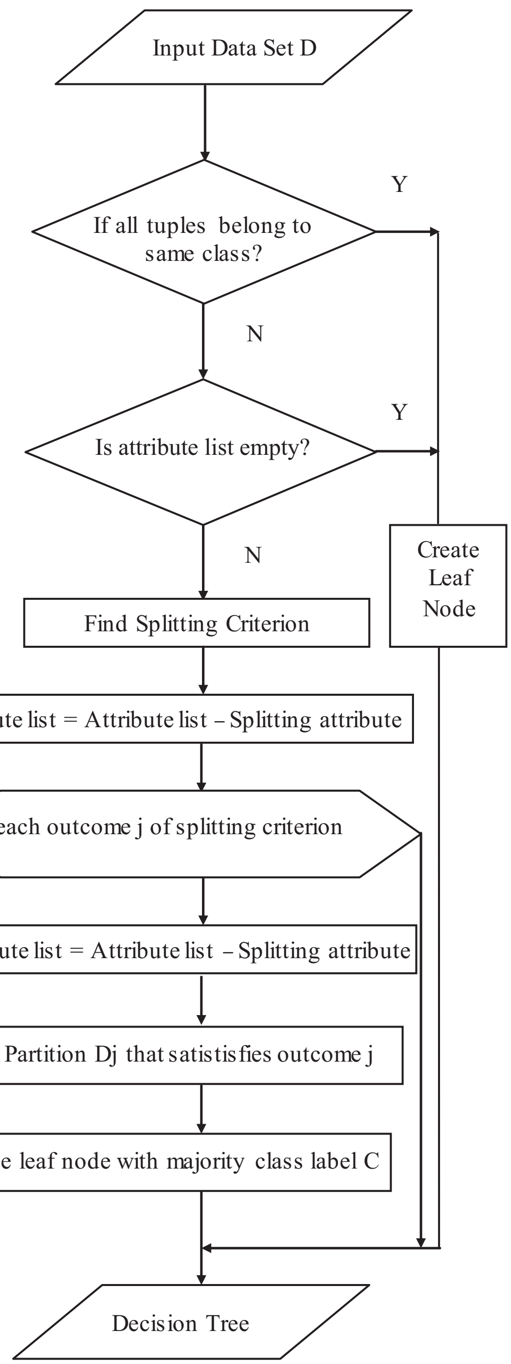

Figure 2. Flowchart of decision tree construction algorithm.

where p_i is the probability of arbitrary tuple in D belonging to class C_i

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{D} Info(D_j), \tag{4}$$

where $\frac{|D_j|}{D}$ is the weight of the jth partition.

4. Experimental results and discussion

The data set used in this research is constructed from 200 questionnaires distributed to various farmers cultivating paddy along the Thamirabarani river basin. Three districts, namely Tirunelveli, Tuticorin, and Kanyakumari, are identified for analysis.

From the raw data collected from farmers of the above three districts, a pre-processing operation is carried out. Based on the value of information gain, 13 agronomic factors and 3 meteorological factors are identified as key attributes. On the preprocessed data, k-means clustering is performed. The clustered data set is given as input of the classifier. The classification method identifies suitable class as (yield = good) or (yield = not good). Accuracy values achieved from various decision tree classifiers such as J48, decision stump, random forest, and REPTree are shown in Table 2. The random forest algorithm classified 195 out of 200 instances correctly, resulting in an accuracy rate of 97.5%.

Table 2. Classification accuracy results.

	J48	Decision Stump	Random Forest	REP Tree
Correctly Classified Instances	188	163	195	184
Incorrectly Classified Instances	16	47	5	16
Kappa statistic	0.78	0.51	0.93	0.78
Mean absolute error	0.09	0.26	0.10	0.12
Root mean squared error	0.26	0.36	0.17	0.26
Relative absolute error	24.6	68.89	26.27	30.5
Root relative squared error	59.97	83.31	39.7	58.97
Total Number of Instances	200	200	200	200

The results prove that performing clustering before classification can identify only the most relevant data needed by the classifier.

Therefore, Figure 3 shows good accuracy is achieved when clustering is integrated with classification. The highest classification accuracy is achieved with random forest. Here the number of clusters is set as 2, in order to group the data into two categories, namely high yield and low yield. The high yield cluster had 148 instances and the low yield cluster had 52 instances. For the problem concerned, the high yield cluster is contributing more. Therefore, 148 instances from the high yield cluster are considered for the mining process.

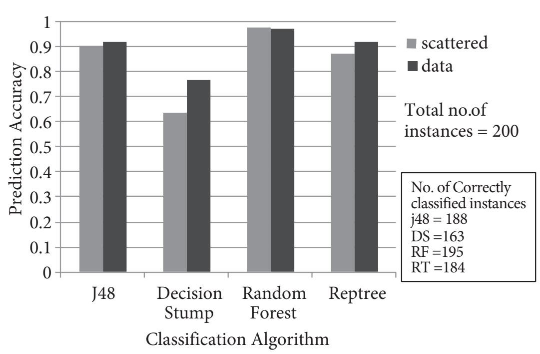

Figure 3. Comparison of classification accuracy.

Classification accuracy values achieved with the classification algorithms j48, decision stump, random forest, and REPTree are 92%, 76.5%, 97.5%, and 92%, respectively. Corresponding results are shown in Table 3. Values for TP rate, FP rate, precision, recall, F-measure, and ROC are shown in Table 4.

Table 3. Classification summary.

Algorithm	Classification Accuracy	
	On scattered data	On clustered data
J48	90.5%	92%
Decision Stump	63.5%	76.5%
Random Forest	98%	97.5%
Reptree	87.5%	92%

Table 4. Detailed accuracy.

	TP Rate	FP Rate	Precision	Recall	FMeasre	ROCArea
J48	0.92	0.18	0.92	0.92	0.92	0.91
Decision Stump	0.77	0.13	0.85	0.76	0.78	0.78
Random Forest	0.97	0.06	0.97	0.98	0.98	0.99
REPTree	0.92	0.17	0.92	0.92	0.92	0.95

The decision tree generated by j48 is the cultivation plan. A sample interpretation of knowledge represented by the decision tree is shown in Table 5.

Table 5. Extracted cultivation plan

Cultivation Plan
IF(amount of rainfall=good) ˆ (fertilizertype=Complex,urea,potassium,sulphate) ˆ (cropvariety=AMBAI-16) ˆ(seedrate=26-30)ˆ (temperature=moderate) THEN yield=HIGH
IF (amount of rainfall=good)ˆ (fertilizer type=dung,Complex,urea,potassium,sulphate) ˆ(sowing procedure=kulimuthu)ˆ(typeofsoil=blacksoil) THEN yield=HIGH

5. Conclusion

The major finding of this research is the accurate extraction of hidden knowledge about a cultivation plan involving major agronomic and meteorological factors. This knowledge is useful for getting high yield of paddy crop. Performances of classifiers such as j48, decision stump, random forest, and REPTree are evaluated in terms of classification accuracy and root means squared error values. It has been concluded that the random forest classifier has high predictive power with an accuracy rate of 97.5%, ROC = 0.99, precision = 0.97, and recall = 0.98.

Also in this research, K-means clustering is integrated with decision tree classifiers in order to improve classification accuracy. Through experimentation it has been proved that classification accuracy is improved by 4.6% on average. Even with worst-case performance using decision stump, integration of k-means clustering has improved the accuracy from 63.5% to 76.5%. As the clustering process brings closely relevant instances for the classification process, there is good improvement in classification accuracy.

References

[1] Aishwarya ABR. A literature study on application of data mining tools for rice yield prediction. International Journal of Innovative Technology and Research 2016; 4: 2757-2759.

[2] Uno Y, Prasher SO, Lacroix R, Goel PK, Karimi Y, Viau A, Patel RM. Artificial neural networks to predict corn yield from compact airborne spectrographic imager data. Comput Electron Agr 2005; 47: 149-161.

[3] El-Telbany M, Warda M, El-Borahy M. Mining the classification rules for egyptian rice diseases. Int Arab J Inf Techn 2006; 3: 303-307.

[4] Mucherino A, Papajorgji P, Pardolas PM. A survey of data mining techniques applied to agriculture. Open Resource International Journal 2009; 9: 121-140. DOI: 10.1007/s12351-009-0054-6.

[5] Nasrin Fathima G, Geetha R. Agriculture crop pattern using data mining techniques. International Journal of Advanced Research in Computer Science and Software Engineering 2014; 4: 781-785.

[6] Diriba Z, Borena B. Application of data mining techniques for crop productivity prediction. HiLCoE Journal of Computer Science and Technology 2013; 1: 151-155.

[7] Tomida K, Yamaguchi Q. Using data mining to identify patterns of student voltage drop. Kuroshio Journal 2015; 8: 138-146.

[8] Papageorgiou EI, Gemtos TA. Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture. Appl Soft Comput 2011; 11: 3643-3657.

[9] Amarasingha RPRK, Suriyagoda LDB, Marambe B, Gaydon DS, Galagedara LW, Punyawardena R, Silva GLLP. Simulation of crop and water productivity for rice (*Oryza sativa* L.) using APSIM under diverse agro-climatic conditions and water management techniques in Sri Lanka. Agr Water Manage 2015; 160: 132-143.

[10] Ashwinirani, Vidyavathi BM. Ameliorated methodology for the design of sugarcane yield prediction using decision tree. Compusoft - An International Journal of Advanced Computer Technology 2015; 4: 1882-1889.

[11] Marinkovic B, Crnobarac J, Brdar S, Antic B, Jacimovic G, Crnojevic V. Data mining approach for predictive modelling of agricultural yield data. In: BioSense 2009 Sensing Technology in Agriculture, Forestry and Environment Workshop; Oct 2009; Novi Sad, Serbia.

[12] Shakil KA, Alam M. Dengue disease prediction using WEKA data mining tool. OALib Journal 2015; 2: 1-24.

[13] Ramya MC, Lokesh V, Majunath TN, Hegadi RS. A predictive model construction for mulberry crop productivity. Procedia Comput Sci 2015; 45: 156-165.

[14] Veenadhari S, Mishra B, Singh CD. Soyabean productivity modelling using decision tree algorithms. International Journal of Computer Applications 2011; 27: 11-15. doi: 10.5120/3314-4549.