

Design and Development of IOT Based Intelligent Agriculture Management System in Greenhouse Environment

C. Kishore Kumar, V. Venkatesh



Abstract: The introduction of Internet of Things (IoT) has a significant impact on shaping the communication and internetworking landscapes. The upcoming IoT researches are linked with design of standards and open architectures still requiring a global attention before deployment. The main objective is to design and develop a framework on Internet of Things (IoT) for precision agriculture using Machine learning techniques, where it surges the efficiency in farming by minimizing the loss of water and studying the fertility of the field. Libelium Smart Agriculture is used to connect to the IoT which uses Waspote module. Waspote is the plug and sense platform which is programmed using Waspote IDE configured to connect with the available Local Area Network (LAN). With the help of Machine learning techniques like Classification And Regression Technique (CART) and Linear Support Vector Machine (SVM), the amount of water required by the crops can be estimated. In this paper, various regression such as stochastic gradient descent and boosted tree regression techniques are compared and results were obtained. Although each model applied in this paper performed well in predicting whether the crop needs to be irrigated, the optimal prediction accuracies were acquired by Boosted Tree Regression (BTC). It is compared by the fold numbers, Root Mean Squared Error (RMSE) and coefficient of Determination (CoD). The accuracy of the boosted tree regression came out to be 91.93% and the stochastic gradient descent prediction model delivered 62.95% accuracy. The amount of water required for the irrigation is then sent to appropriate actuator like solenoid valve and motor can be turned on for that particular period of time. Calibration test results and Measurements are represented to enhance the accuracy and success rates of Precision Agriculture (PA).

Index Terms: Decision Support System (DSS), Libelium, Machine learning, Smart Agriculture.

I. INTRODUCTION

Agriculture is an enormous space which primarily includes developing, maintaining, monitoring and yielding of crops like sugarcane, paddy, tomatoes, etc. A thousand years back when human beings got a colossal improvement for the formation of sustenance where nature helped them to nurture, build technologies around them[1]. Nowadays due to the progression of industrialization, agriculture has taken up its circumstance in business promote. The art of soil monitoring and conservation of water has greatly contributed

to production of high yielding variety of crops. However, the use of fertilizers and pesticides has affected the agricultural field. So to reduce the use of harmful fertilizers and to expand the yield, numerous strategies have been plotted and executed by the agronomists. Considering the availability of soil, water and appropriate conditions of the field, agronomists came up with the idea of precision agriculture. Precision agriculture is a farm management technique which comprises of observing, monitoring, measuring and responding to the variable parameters of the crops[2]. India has been contributing 30% of total organic production in the world but accounts to only 2.59% of total cultivation area of 57.8 million hectares. This is because of the lack of awareness of precision agriculture and poor farming techniques. Also because of the prevailing high input cost and inadequate technical infrastructure, farmers are not being able to meet the expected cultivation. So, agriculture can be efficiently improved by consolidated utilization of precision agriculture techniques based on Internet of Things (IoT) and Machine Learning (ML) approach. Based on these lines, agriculture can be possible more accurate, cost effective and high yield. In this study, Machine learning approaches such as Linear Support Vector Machine (SVM), Classification and Regression Technique (CART) has been compared. Regression such as Stochastic Gradient Descent and Boosted Tree Regression are used and they are compared based on parameters such as Mean absolute error (MAE), Relative absolute error (RAE), Relative squared error (RSE) and Coefficient of determination (CoD), Root mean squared error[3]. The main aim of the research is to create a prototype that creates the unstructured Decision Support System (DSS) by applying ML algorithms on the libelium smart agriculture sensor information comprising field of environmental and meteorological data for irrigation recommendations using agronomist inferred knowledge. The scientists have proposed many techniques to handle different parameters such as temperature, humidity of water and soil proposed by [4] "Remote Sensing control of an irrigation system using a distributed wireless sensor Network". Decision Support system briefly explained by [5].

II. RELATED WORK

Scholarly articles for Decision Support System (DSS) prediction for irrigation exist in many of the literature. They have been used for multiple intends corresponding to ameliorating the irrigation through various restraints and topologies [6], organizing [7],

Manuscript published on 30 July 2019.

* Correspondence Author (s)

C.Kishore Kumar, Embedded Systems, SASTRA Deemed to be university, Thanjavur, India.

V.Venkatesh, CSE Department, SASTRA University, Thanjavur, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

computing requirements for the irrigation [8], Monitoring consumption of water [9]. Irrigation DSS models for ML approach has been given in [10], where they used an approach to reduce the wastage, allocating and optimizing the water consumption for one single crop in multiple time variants. They are not plot specific characteristics. For specific sectors and plot data, [11] explained another generic approach based on electromagnetic mapping where they predict water table depth and water status of the soil. Most of the existing models (irrespective of the ML usage), DSS is provided by calculating water balance models and net evapotranspiration based on FAO-56, Penman-Monteith equation.

III. SYSTEM ARCHITECTURE

Fig. 1 shows architecture for design and development of environmental monitoring system for agricultural application based on Internet of Things and machine learning. Libelium Smart agriculture comprises of temperature, humidity and pressure sensor probe, solar radiation probe are used. For finding the values of wind direction, wind Speed and the amount of precipitation are calculated using wind vane, anemometer, wind vane and pluviometer respectively. Libelium is a plug and sense platform which can be installed directly on the field. Battery can be charged with the help of photovoltaic cells. It consumes low power, can be used in rugged environments, portable water monitoring. These data can be collected as a Java Script Object Notation (JSON) and it can be converted to comma separated value (CSV) file.

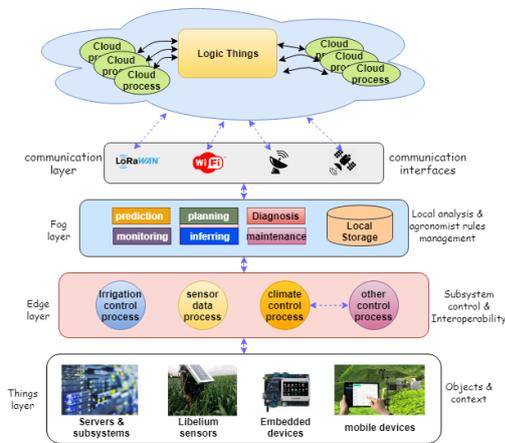


Fig 1. Architecture for design and development of environmental monitoring system for agricultural application based on internet of things

IV. METHODOLOGY AND MATERIALS

Many machine learning procedure are present to direct the modeling issue, where the problem is missing data. Those procedures involve the analysis of problems arising when the decision has to be made for weather forecast where it is very hard to quantify the factors which are process-specific which affects the decision. The shadow output of temperature value obtained from the Libelium smart agriculture is shown in fig 2.

A. Dataset

Libelium smart agriculture consists of sensor probes and waspmote plugs which is used to obtain different Last update: Sep 26, 2018 5:47:21 PM +0530

```
Shadow state:
1 {
2   "desired": {
3     "id": "54569",
4     "id_wasp": "AGRI_MODE_1",
5     "id_secret": "33806F057C105448",
6     "sensor": "TC",
7     "value": "26.59",
8     "datetime": "2018-09-26 17:47:08"
9   },
10  "delta": {
11    "id": "54569",
12    "id_wasp": "AGRI_MODE_1",
13    "id_secret": "33806F057C105448",
14    "sensor": "TC",
15    "value": "26.59",
16    "datetime": "2018-09-26 17:47:08"
17  }
18 }
```

Fig 2. Shadow output of the Libelium smart agriculture in JSON notation

parameters implemented as shown in fig 3. The datasets involves parameters such as soil temperature, soil moisture, wind direction, water vapor pressure, solar radiation, precipitation, mean wind speed are used for prediction of soil irrigation. These parameters are obtained from the Libelium smart agriculture kit and meteorological data are collected. Fig 4. represents the framework of the proposed model where the sensor based environment data and meteorological data comprised to form a new dataset. CART approach is done on the dataset and the amount of rainfall and amount of water required to the crops are predicted. Then the dataset is used in ML techniques and output is predicted by finding the rate of true estimation error.



Fig 3. Implementation of libelium connected to various sensors like pluviometer, wind vane, anemometer

In this study CART technique is used. Classification And Regression Technique (CART) involves the nodes and trees which help us to understand how the paramaters are mutually relatable to one another. For the analysis of linear Support Vector Machine (SVM), Anaconda and jupyter notebook is used. The regression techniques can be programmed using python in the jupyter notebook. The estimator implies regularized linear models with Stochastic Gradient Descent regression (SGD). The advantages of SGD are efficiency and ease of implementation. But SGD needs a hyperparameter such as the regularization parameter and the iterations. It is very susceptible to feature scaling.

The slope of the loss is determined at each sample at a time and the model can be updated along with a decreasing learning rate. It also allows out of core learning and partial fit method. The best model can be obtained when the data

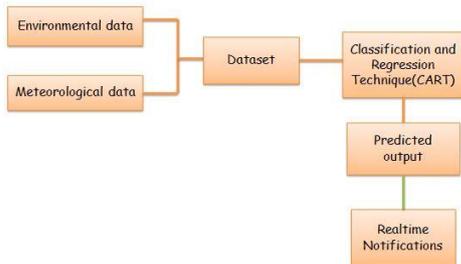


Fig 4. Proposed model of the framework

have unit variance and zero mean. To obtain the measure of misfit of the data, root-mean-squared error (RMSE) value is calculated. The effect of feature values and feature extraction of the particular variable can be explained by the linear regression model.

With respect to output variable, homogeneous samples comprised of heterogeneous group of features. Structure basically comprises of a tree having a parent node which split into two child node. It is said to be the best split when they divide the given data into two predominant categories. Number of nodes is called as “size” of the tree and depth of the tree is called as “levels”. The best split also defined by purity of sub-groups and nodes of similar size. Those dense structures can be optimized by pruning. The data can be a raw one and unit conversions are not required. There are no prior assumptions of data required.

Microsoft azure machine learning studio helps to analyse the characteristic difference between Stochastic Gradient Descent and Boosted Tree Regression. Fig 5. represents the framework of Boosted Tree regression where columns of the data have been selected, splitted, trained, scored and evaluated and visualized based on the parameters. Fig 6. represents the framework of cross validation done among the linear, boosted tree and poisson regression on Microsoft azure platform.

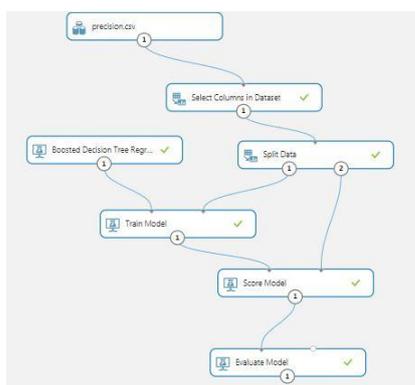


Fig 5. Boosted Decision tree regression framework on Microsoft azure

B. Classification And Regression Trees (CART)

CART algorithm has been already used by [12] and also known as binary recursive partitioning. The each of the parent node is split into child nodes and they additionally splitted furthermore. The comprised dataset is segmented to form a homogeneous sample with same features. Binary response of the categorical variable is obtained: either it is

‘Yes’ or ‘No’. Generally, there are four major procedures in CART:

Building a tree: The data separated into columns are placed at root node. Amidst training, the starting variable is divided according to all the possible values of the data. The child

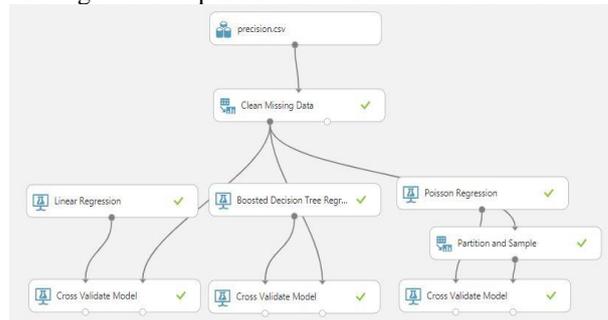


Fig 6. Cross validation of dataset among different regression techniques

nodes with corresponding binary responses are obtained, a ‘yes’ and a ‘no’ for each split. For each node, a class is assigned regardless of splitting of child nodes. Goodness-of-split criterion is then used in terms of reduction in impurity or heterogeneity. It can be done by calculating Gini Index given as:

$$g(t) = \sum_{j \neq i} p(j|t)p(i|t)$$

where $p(i|t)$ and $p(j|t)$ are the probabilities of i and j , respectively, at the particular node. Highest reduction of impurity is identified and the parameter is then said to be best split. The Gini index then be calculated for all the variables. The above steps are repeated for every nodes of the non-terminal child.

Stopping criterion: At a particular time interval, new large tree is produced with which it overfits the data comprised in the learning dataset. The splits are then abruptly hold when there is no improvement in the obtained predicted result. Then the resubstitution of various error rate of the regression is estimated where it predicts the case misclassified on the features used through testing. It is given in [13]:

$$R(d) = \frac{1}{N} \sum_{i=1}^N X(d(x_n) \neq j_n)$$

where, X is the sign function and $d(x)$ is the classifier such that,

$$X=1, \text{ if } X(d(x_n) \neq j_n) \text{ is true and } \\ X=0, \text{ if } X(d(x_n) \neq j_n) \text{ is false.}$$

This particular pointer function is biased and underrates the value of the error.

Pruning of the tree: CART removes the unimportant nodes and reduces the number of splits using the 10-fold cross validation. In this study 10 fold cross validation is done and in each fold 68 samples are used. The percentage of accuracy during prediction has to be high and more honest when comparing prediction error rate to the resubstitution error. The graphical representation of prediction error against frequency is plotted and the minimum error is selected.

Selection of optimal tree: The optimal tree is the one which do not overfit the parameters of training dataset to be selected among the order of pruned trees evaluated from the crossvalidation (cv). Breiman et al. (1984) suggested that optimal tree is the least possible tree which has the minimum standard error among the estimated error rate.

Then the test sample is evaluated and results obtained based on such as Mean absolute error (MAE), Relative absolute error (RAE),

Relative squared error (RSE) and Coefficient of determination (CoD), Root mean squared error between the predicted and the observed data. Table 1 and Table 2 shows the crossvalidation evaluation results by fold for Stochastic Gradient Descent Regression and crossvalidation evaluation results by fold for Boosted Tree Regression respectively.

This score model is based on the development of each parameter compared to the primary spitting variables. The variable with the maximum sum of improvement has the highest score and rest of the variables scored low with the descending order. CART does not use training based data performance like C4.5 for the selection of the tree. But the performance is observed based on independent test data and crossvalidation [14].

C. Algorithm

Step 1: Take marked input data with the list of target variables and list of independent variables.

Step 2: Best split for each independent variables

- ❖ Split data: Training set- 70%, Testing Set -30%
- ❖ Maximum no. of leaves used in single tree- 20
- ❖ Maximum no. of training set to cohere a leaf -10
- ❖ Learning rate- 0.2
- ❖ Total no. of trees built -100

Step 3: Selecting the best variable for the split

- ❖ Column names: Precipitation(In), Solar Radiation(kJ m day), Wind Direction (Degrees), Mean Wind Speed(MPH),Water vapor Pressure(mbar), soil moisture, soil temperature.

Step 4: perform 2nd and 3rd steps till the stopping criterion comes for each of the nodes of the tree.

Step 5: prune the decision tree and evaluate it for the crossvalidation and find true estimate error.

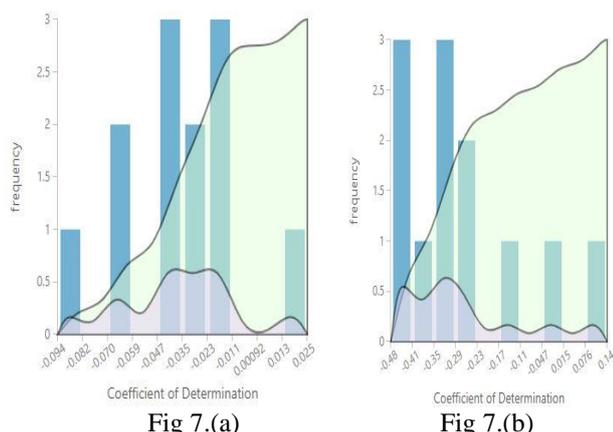


Fig 7. (a) Coefficient of determination obtained for Stochastic Gradient Descent with cumulative distribution and probability density

Fig 7.(b) Coefficient of determination obtained for 10 fold cross validation model for Boosted Tree regression with cumulative distribution

V. RESULTS AND DISCUSSION

A 10 fold cross validation done on the datasets which are

classified into 68 samples and the accuracy of cross validation is almost equal to the re-substitution accuracy in both the SGD classifier and Boosted Tree classifier. Table 1 and Table 2 show the cross validation results evaluated by the respective algorithms and prediction for tree analysis is presented. It is performed to determine whether the classifier is being over fitted. In this study, performance is satisfactory and promising. Compared to the Stochastic Gradient Descent classifier, Boosted tree regression model achieved considerably better results. The first fold in SGD gives

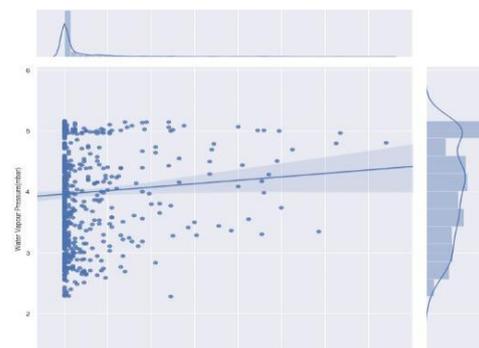


Fig 8. Overfitting and coefficient analysis of amount of rainfall occurred against water

minimum RMSE of 0.068575, relative squared error of 1.0942 and standard deviation of coefficient of determination of 0.629532. The first fold of Boosted tree regression gives minimum RMSE of 0.253672, relative squared error of 1.04291 and standard deviation of coefficient of determination of 0.919301. The cutoff point of coefficient of determination (COD) for a good model always range between 0 to 1 and closer to 1, Boosted tree regression model outperforms the SGD classifier because the standard deviation of coefficient of determination is 91.93%. Negative sign is ignored because COD is R^2 . Fig 7.(a) and fig 7.(b) are the Coefficient of determination obtained for 10 fold cross validation model for Stochastic Gradient Descent and boosted tree regression model compared with cumulative distribution and probability density function. In fig 8, the majority of the folds provide the value over 0.3 and the performance is much better than the SGD classifier.

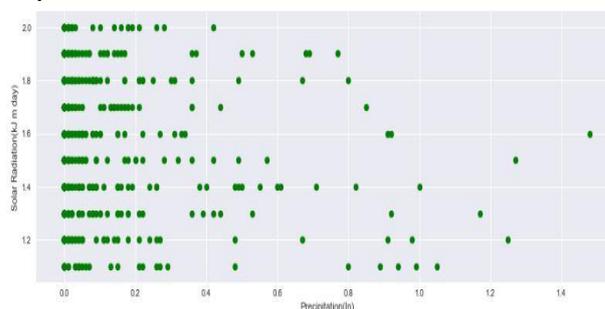


Fig 9.Scatter plot of solar radiation

The dataset is used to find the correlation analysis of the features and correlation coefficient is obtained. Correlation analysis is used to study the interrelationship between the continuous and measured parameters.

The positive correlation coefficient is obtained for the water vapor pressure and precipitation and the value is 0.071. Regression is used against these parameters and the graphical representation is shown in fig 8.

Stochastic Gradient Descent (Regression)						
Fold Number	Number of examples in fold	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
0	67	0.056386	0.068575	1.246386	1.094204	-0.094204
1	68	0.086935	0.170228	0.890973	1.025417	-0.025417
2	68	0.105687	0.208345	0.846992	1.042622	-0.042622
3	68	0.108527	0.230637	0.811392	1.041999	-0.041999
4	68	0.124266	0.245158	0.812255	1.063234	-0.063234
5	68	0.098259	0.206148	0.860702	1.020011	-0.020011
6	68	0.127493	0.231151	0.81981	1.063481	-0.063481
7	68	0.093314	0.179748	0.863382	1.020457	-0.020457
8	68	0.09716	0.192012	0.84449	1.021874	-0.021874
9	68	0.103445	0.205229	0.856666	1.026991	-0.026991
Mean		0.100699	0.193151	0.890953	1.037791	-0.037791
Standard Deviation		0.019822	0.049272	0.134635	0.629532	0.629532

Table 1. Cross validation evaluation results by fold for Stochastic Gradient Descent Regression

The amount of water required for the crop also predicted by analyzing the meteorological data such as solar radiation and precipitation. So the amount of rainfall occurred has been studied against solar radiation and the scatter plot between these variables have been obtained and shown in fig 10. Maximum range of rainfall has been recorded from 0.1 to

0.3. The amount of water required for the irrigation calculated apart from these ranges.

Boosted Tree Regression						
Fold Number	Number of examples in fold	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
0	67	0.208397	0.253672	0.974881	1.042911	-0.042911
1	68	0.241715	0.298532	1.172814	1.476256	-0.476256
2	68	0.25133	0.311862	1.095543	1.438459	-0.438459
3	68	0.248456	0.296087	0.983107	1.119229	-0.119229
4	68	0.263589	0.312476	1.103219	1.322007	-0.322007
5	68	0.283198	0.324574	1.212506	1.420703	-0.420703
6	68	0.257034	0.302401	1.087596	1.290431	-0.290431
7	68	0.251942	0.311553	1.038676	1.301709	-0.301709
8	68	0.24655	0.304255	1.053258	1.271529	-0.271529
9	68	0.253054	0.309316	1.132083	1.375477	-0.375477
Mean		0.120111	0.194968	1.114044	1.287985	-0.287985
Standard Deviation		0.012621	0.032478	0.385961	0.919301	0.919301

Table 2. Cross validation evaluation results by folds for Boosted Tree Regression

VI. CONCLUSION

Comparing the regression results of the framework with supervised linear SVM for the various parameters regression based on dataset generated. The Classification And Regression Technique (CART) framework outperforms the supervised linear SVM with basic tuning techniques, correlation analysis and feature selection. CART algorithm clusters the dataset by which it can provide augmented knowledge about the model and the given input data. It can handle incomplete or corrupt datasets for regression. The proposed model can be extended with adding Decision trees and sophisticated regression techniques. Loading and preprocessing data can be achieved through overfitting and grid search algorithms. MQTT protocols can be enhanced with the help of Firebase and Amazon Web Services (AWS).

ACKNOWLEDGMENT

The authors are grateful to the Department of Science & Technology, New Delhi, India for the financial support: Grant numbers: (SR/FST/ETI-71/2014) and (SR/FST/MSI-107/2015). They also wish to acknowledge SASTRA University, Thanjavur for extending infrastructural support to carry out the work.

REFERENCES

1. Tomo Popovic, Nedeljko Latinovic, Ana Pešić, Zarko Zec'evic', Boz'o Krstajic', Slobodan Djukanovic., "Architecting an IoT-enabled platform for precision agriculture and ecological monitoring: A case study", Computers and Electronics in Agriculture, vol.no: 140, 2017, pp. 255-265.

2. Evans, R., & Bergman, J., "Relationships between cropping sequences and irrigation frequency under self-propelled irrigation systems in the Northern Great Plains (NGP)", USDA Annual Report, 2007, pp. 466-479.
3. Anat Goldstein et.al, "Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge", Precision Agriculture, vol.no: 19, 2018, pp.421-444.
4. Kim, William M.Iverson, Robert G.Evans, "Remote Sensing control of an irrigation system using a distributed wireless sensor Network", vol.no:57, , 2008, pp.1379-1387.
5. Junyan Ma, Xingshe Zhou, Shining Li , Zhigang Li, "Connecting Agriculture to the Internet of Things through Sensor Networks", 2011, pp.19-29.
6. Allen, R. G., "Crop evapotranspiration: Guidelines for computing crop water requirements", (FAO) irrigation and drainage paper. Rome: FAO, 1998, pp.344-356.
7. Vellidis, G., Tucker, M., Perry, C., Reckford, D., Butts, C., Henry, H., et al., "A soil moisture sensor-based variable rate irrigation scheduling system", Precision agriculture '13. The Netherlands: Wageningen Academic Publishers, 2013, pp. 487-609.
8. Thysen, I., & Detlefsen, N. K., "Online decision support for irrigation for farmers", Agricultural Water Management, vol.no: 86, 2006, pp.269-276.
9. Mateos, L., Lopez-Cortijo, I., & Sagardoy, J. A., "SIMIS: The FAO decision support system for irrigation scheme management", Agricultural Water Management, vol.no: 56, 2002 , pp.193-206.
10. Y. Liu, Z-H. Ren, D. M. Li, X. K. Tian, Z. N. Lu, "The research of precision irrigation decision support system based on genetic algorithms," in Proc. 5 Int. Conf. on Machine Learning and Cybernetics, Dalian, 2006, pp.56-67.
11. Hedley, C. B., Roudier, P., Yule, I. J., Ekanayake, J., & Bradbury, S., "Soil water status and water table depth modeling using electromagnetic surveys for precision irrigation scheduling", Geoderma, vol. no:199, 2013, pp.22-29.
12. Breiman, L., "Random Forests and Machine Learning", vol. 45, 2001, pp. 5-32.
13. Hill, T., Lewicki, P., "STATISTICS: Methods and Applications", Statsoft, 2007, pp.678-690.
14. Andriyas, S., & McKee, M., "Recursive partitioning techniques for modeling irrigation behavior", Environmental Modeling & Software, vol.no:47, 2013, pp.207-217.

AUTHORS PROFILE



C.Kishore Kumar is currently pursuing Masters of Technology at SASTRA University, Thanjavur, India. His research interests are in the areas of Smart agriculture, Embedded systems, Internet of Things, Wireless sensor networks, Machine learning, etc. He has already published papers on smart irrigation and farming using Node MCU ESP8266.



Dr.V.Venkatesh received the PhD degree in Computer Science from the SASTRA university. He is an assistant professor of computer science at the SASTRA university. He has conducted research in edge analytics and Wireless sensor Networks. His latest publications are as follows:
A system for real time monitoring and imparting emergency circumstances, Data fusion: An energy

efficient perspective for device-to-device (D2D) communication, Effective fire alarm system with real time multi sensor data fusion. For further publications detail <http://orcid.org/0000-0002-0625-0161>