A Proficient Microclimate and Assisted Sustainable Agriculture using Data Analytics

Dr.P.TAMIJE SELVY[1], G.POORANI[2], S.AKILA[3], P.S.DHEEBIKA[4], M.DHIVYALAKSHMI[5]
Professor[1], Assistant Professor[2],Student [3,4,5]
Department of Computer Science and Engineering[1,2,3,4,5]
Sri Krishna College of Technology[1,2,3,4,5], INDIA
[1]p.tamijeselvy@gmail.com, [2]pooranilalithb07@gmail.com ,[3]dhivyalakshmi.m03111996@gmail.com

Abstract: - A proficient microclimate and assisted sustainable agriculture is done using data analytics. The climatic conditions of different places vary with respect to the environment. This is called as microclimatic condition. The changes in the microclimate may affect the agricultural growth. Understanding the behavior of the complex ecosystem, monitor only the climatic conditions and displays the nearby location of shadow is done in the existing system (Used in smart city environment).Only the Urban Heat Island(UHI) effect[15] is monitored. The drawback of the existing system is, the usage of these details in not up to the greater extend. The goal of the proposed system is to provide solution to improve the growth of agriculture when there is a microclimatic change. The project in new methodology is to implement proficient microclimate analysis, to display type of crop and level of irrigation. The climatic conditions of locations such as, Coimbatore, Madurai are collected from Tamil Nadu Agricultural University (TNAU). The collected dataset is processed in RStudio using Boruta algorithm and fuzzy rule based system to address the above challenges. The proposed system helps to improve the agricultural growth of the nation. The final system displays the type of crop to be planted and level of irrigation to be provided to improve the crop plantation and will be disseminated to farmers via website.

Key-Words: - Data analytics, Sustainable agriculture, Urban Heat Island effect (UHI), Boruta algorithm, Fuzzy rule based system, irrigation.

1 Introduction

Agriculture, an important sector of our economy accounts for 14 per cent of the nation's GDP and about 11 per cent of its exports. Rice, wheat, cotton, sugarcane, milk and potatoes are the major agricultural products produced. The major field agriculture is comprised with soil fertility and irrigation. This is so because plant agriculture is mainly concerned with crops and their economic production. Sustainable agriculture helps to permit sophisticated resource proficiency. They produce greater agricultural production using lesser land, water and energy, confirming profitability for the farmer. These essentially include methods that, among other things, protect and enhance the crops and the soil and improve water absorption. The data analytics refers to breaking a whole dataset into its separate component from individual analysis. From the analysis of agricultural data the required information will be provided to the farmers to increase agricultural growth.

1.1 Data Analytics in Agriculture

Data analytics can prevent spoilage by moving products faster and more efficiently. It has made such a widespread impact in the agriculture industry that it is hard to identify all of its effects, and harder still to predict what changes it might bring. The insight provided by data analytics allows farmers to start and harvest their crops at the optimum time, which maximizes crop yields and minimizes stress. Farmers are using drones with advanced sensors to survey their crops, update their data and notify them of areas that need improvement. As the technology continues to progress and the farmers can expect drones to move from surveying to planting and harvesting themselves.

- Collection of data about the status of the soil, insects, crops, livestock, water and weather
- Agricultural drones monitoring fields
- Soil and crop sensors sending feedback on soil composition, soil moisture and plant health

1.2 Feature Selection

Feature selection also known as subset selection is a process commonly for high dimensional datasets, wherein subsets of the features available from the data are selected for application of a learning algorithm. The best subset contains the least number of dimensions and we discard unimportant dimensions. This is an important stage of preprocessing and is one of two ways of avoiding the dimensionality problems.

Feature selection methods can be classified into three types such as Filter method, Wrapper method and Embedded method.

1.2.1 Filter Method

Filter method is used to select the features on the basis of their scores in various statistical tests for their relationship with the outcome variable. Filter methods use statistical methods for valuation of a subset of features. The filter methods do not remove multicollinearity.

- Pearson's Correlation: It is used for measuring the linear dependence between two continuous variables X and Y.
- LDA: Linear discriminant analysis is used to find a linear combination of features that characterizes or separates two or more classes (or levels) of a categorical variable.
- ANOVA: ANOVA stands for Analysis of variance. It is similar to Linear Discriminant Analysis that it is operated using one or more categorical independent features and one continuous dependent feature. It provides a statistical test of whether the means of several groups are equal or not. It is used to check the similarity between the variables.
- Chi-Square: It is a statistical test applied to the groups of qualified features to evaluate the likelihood of association between them using their frequency distribution.

1.2.2 Wrapper Method

Wrapper methods are based on the implications from the previous model, features can be included or removed from the subset [13]. The problem is essentially reduced to a search problem. These methods are computationally very expensive.

- Forward Selection: Forward selection is an iterative method starts with no feature in the model. Features are included which best improves the model till an addition of a new variable does not improve the performance of the model. Attribute with higher importance is used to select the attributes. Training dataset with higher importance attributes are used.
- Backward Elimination: In backward elimination, the least significant features are removed at each iteration. It will help to improve the performance of model. Repeat this until no improvement is observed on removal of features. They add or remove variables one-at-a-time until some stopping rule is satisfied. The operation starts with full set of attributes based on the performance.
- Recursive Feature elimination: It is a greedy optimization algorithm is used to find the best performing feature subset. At each iteration, it creates models and separates the best or the worst performing feature. It builds the upcoming model with the remaining features and rank the features based on the order of their elimination.

1.2.3 Embedded Method

Embedded methods[18] combine the qualities of filter and wrapper methods. It is executed by algorithms that have their own incorporated feature selection methods.

- LASSO (Least Absolute Shrinkage and Selection Operator) regression performs L1 regularization that adds consequence equivalent to entire value of the magnitude of coefficients. It is generally used as regulation method.
- Ridge regression performs L2 regularization that adds consequence equivalent to square of the magnitude of coefficients.

2 Related Works

Punit Rathore et al [1] proposed Real-time environment monitoring and analysis is an important research area of Internet of Things (IoT). Understanding the behavior of the complex ecosystem requires analysis of detailed observations of an environment over a range of different conditions. One such example in urban areas includes the study of tree canopy cover over the microclimate environment using heterogeneous sensor data. There are several challenges that need to be addressed, such as obtaining reliable and detailed observations over monitoring area, detecting unusual events from data, and visualizing events in real-time in a way that is easily understandable by the end users (e.g., City Councils). In this regard, we propose an integrated geovisualization framework, built for real-time wireless sensor network data on the synergy of computational intelligence and visual methods, to analyze complex patterns of urban microclimate. A Bayesian maximum entropy based method and a hyper ellipsoidal model based algorithm have been built in our integrated framework to address above challenges. The proposed integrated framework was verified using the dataset from an indoor and two outdoor network of IoT devices deployed at two strategically selected locations in Melbourne, Australia. The data from these deployments are used for evaluation and demonstration of these components' functionality along with the designed interactive visualization components. IOT devices were used to monitor the temperature, using those details shadow regions were predicted.

Dijana et al [2] proposed the intrinsic relationship between dataset characteristics and feature selection techniques performance. Thus, the research experimentally examined how dataset characteristics affects the accuracy and the time complexity of feature selection. Using accessible datasets, the functioning of feature selection techniques are assessed with different features, widespread experiments with five feature selection methods, three types of classification algorithms and all possible permutations of dataset. The decision tree method is applied to evaluate the interdependencies between dataset characteristics and performance. The results of the study shown the intrinsic relationship between dataset characteristics and feature selection techniques performance.

Harine RajaShree et al [3] proposed the common problem existing among the Indian farmers are they don't choose the right crop based on their soil requirements. Due to this they face a serious setback in productivity. This problem of the farmers has been addressed through precision agriculture. Precision agriculture is a modern farming technique that uses research data of soil characteristics, soil types, crop yield data collection and suggests the farmers the right crop based on their site-specific parameters. This reduces the wrong choice on a crop and increase in productivity. In proposed system, this problem is solved by proposing a recommendation system through an ensemble model with majority voting technique using Random tree, CHAID, K-Nearest Neighbor and Naive Bayes as learners to recommend a crop for the site specific parameters with high accuracy and efficiency.

Aakash G Ratkal et al [4] proposed, about half of the population of India depends on agriculture for its livelihood, but its contribution towards the GDP of India is only 14 per cent. One possible reason for this is the lack of adequate crop planning by farmers. There is no system in place to advice farmers what crops to grow. In proposed system crop yield is predicted and price that a farmer can obtain from his land, by analyzing patterns in past data. We make use of a sliding window non-linear regression technique to predict based on different factors affecting agricultural production such as rainfall, temperature, market prices, area of land and past yield of a crop. The analysis is done for several districts of the state of Karnataka, India. Our system intends to suggest the best crop choices for a farmer in order to address the prevailing socio-economic crisis facing many farmers today.

Kuljit Kaur et al[5] proposed about that the prediction of agriculture yield is a job that requires unification of knowledge from several areas such as data mining, statistics and agriculture. Subject of crop yield prediction has been very popular among various organizations working in agriculture, producers etc. Prediction of crop yield helps in managing the storage of crops as well as it directs the transportation decisions, and risk management issues related to crops. Pattern of rainfall and temperature are dynamic due to global warming, and resulting in undergoing impingement on crop productivity. Data Mining focuses upon methodologies for extracting useful knowledge from data and there are several tools to extract the knowledge that is it is a proficiency of examining the dataset such that the end results can be deduced easily and rapidly from the dataset. The knowledge gathered can be used to forecast the paddy yield. But farmers do not use any knowledge discovery process approach on paddy yield data. Data mining can be used in agriculture for decision making. In this study, we collected data from different government organizations, after preprocessing and discretization of data applied Predictive Apriori algorithm using Data Mining tool (WEKA) for analysis of daily temperature, daily rainfall and paddy yield to predict the paddy yield and to analyze the effect of temperature and rainfall on the paddy yield.

Federico Viani et al [6] proposed a decision support system based on the combination of the wireless sensor and actuation network technology to support the irrigation management in agriculture using fuzzy logic set theory. Using irrigation details fuzzy rule set is generated and the outputs of numerical soil and crop models are used to provide optimized irrigation schedule. The suggested actions are devoted to reduce the waste of water and to maximize the crop yield according to the weather conditions and the real water needs. The proposed methodology is embedded in the Network gateway. It helps to suggest the irrigation details using fuzzy rule based system implemented using network gateway making the system, a truly smart and autonomous wireless decision support system. Lala Septem Riza et al [7] proposed the ideas about fuzzy rule-based systems. They are based on fuzzy concepts to address real-world problems. The models are constructed from data and they are implemented. In this way, accurate and interpretable systems can be built for data analysis and modeling tasks. Some examples are provided on the usage of the package and a comparison with other common classification and regression methods available in R.

Atta-ur-Rahman et al [8] proposed a user behavior classification using Fuzzy Rule Based System (FRBS). Using this technique a network user can be monitored and his/her behavior can be classified depending on his/her activities like unauthorized websites usage, attempting to breach in network security, firewalls, unauthorized services access and frequency of attempts etc. The information about a user is obtained by his/her web, database, hardware and other applications logs.

FRBS classifies a user to one of the predefined categories based on the information extracted from user logs. This would great help in network security and privacy as well as users may be guided for sincere mistakes and other measures may be taken by the organization. Significance of the proposed scheme is shown by examples and results. The attempts used to reveal the functionality of frbs function and its weight factor to be determined.

Dr.P.Tamije Selvy et al [9] proposed performance of classification methods to predict the accuracy levels. Data can be classified according to any criteria, not only relative importance or frequency of use. Classification plays a major role in disease diagnosis. The paper contains brief discussion of various classification methods that includes Case Based Reasoning, decision trees, K-nearest neighbour classifier, naïve bayes classifier and neural network. The paper also discusses some applications of classification model. The performance of the classification methods are observed where the CBR classification model results in 90.7% of specificity, 92.3% of sensitivity and 95.5% of prediction accuracy.

The pruning methods that first utilizing all features to train a model and then attempt to eliminate some features by setting the corresponding coefficients to 0, while maintaining model performance such as recursive feature elimination using support vector machine (SVM) [10]. Features with coefficients that are close to 0 are then eliminated. Due to good performance, regularization models attract increasing attention.

2.1 Motivation

The existing microclimatic system involves the identification of shadow regions and understanding the behavior of the complex ecosystem. To identify only shadow regions in forest areas will not provide adequate support to the farmers to increase the agricultural production. So, a proficient system based on geo-visualization is proposed which will suggest the type of crop to be planted and level of irrigation to be provided to that particular crop using feature selection algorithm[11] and fuzzy rule based system[12].

3 Proposed Scheme

The proposed system is designed to display the type of plantation and irrigation level. The system aims mainly for the improvement of agriculture in the microclimatic condition. This is achieved by extracting the temperature and humidity from the TNAU dataset[19]. Fuzzy rule based system is used to generate the set of rules to display crop details depending upon temperature and humidity. The proposed system proposes monitoring solution for farmers with high level, lost cost assistance to accelerate the efficiency of their cultivations and improving the environmental sustainability of the whole process and quality of the production. The proposed system is implemented using R Studio. Figure 1 represents the Architectural Design of Proposed System.

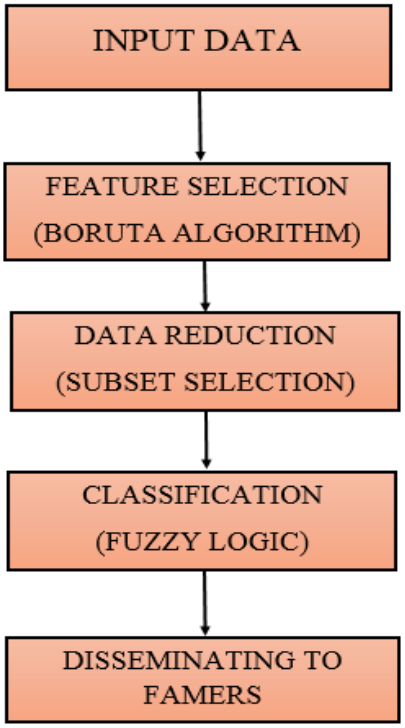

Figure 1 Architectural Design of Proposed System

3.1 Feature Selection

Feature selection is the process of selecting a subset of relevant features for model development. TNAU dataset provides weather on a specific date, year, month and day which comprises weather condition, temperature, humidity and pressure. The Boruta algorithm is a feature selection[17] wrapper algorithm, capable of comparing original attributes importance with importance achievable at random performs a top-down search for relevant features. In proposed system, for feature selection Boruta algorithm is used. Precisely, it works as a wrapper algorithm around Random Forest. It adds randomness to the given data set by creating shuffled copies of all features also called as shadow features. It applies a feature importance measure to evaluate the importance of each feature where higher means more important. At every iteration, it checks whether a real feature has a higher importance than the best of its shadow features and constantly removes features which are deemed highly unimportant. The algorithm terminates when all features tagged as confirmed or rejected or it reaches a specified limit.

3.1.1 Boruta Algorithm

1. Duplicate copies of all independent variables are created. When the number of independent variables in the original data is less than 5, at least 5 copies are created using existing variables.
2. The values of added duplicate copies to remove their relationships with the target variable. It is called shadow features.
3. The original ones are combined with the shuffled copies
4. Run a random forest classifier on the combined dataset and performs a variable importance measure (the default is Mean Decrease Accuracy) to evaluate the importance of each variable where higher means more important.
5. Then Z score is computed. Z score is defined as mean of precision loss divided by standard deviation of precision loss.
6. The maximum Z score among shadow attributes (MZSA) is calculated.
7. Tag the variables as 'unimportant' when they have importance expressively lower than MZSA. Then we permanently remove them from the process.
8. Tag the variables as 'important' when they have importance expressively higher than MZSA.
9. Repeat the above steps for predefined number of iterations (random forest runs), or until all attributes are either tagged 'unimportant' or 'important', whichever comes first.

3.2 Data Pre-Processing

Data pre-processing is a process that involves transforming raw data into an understandable format. Real data is often incomplete, inconsistent, and lacking in certain behaviors and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues. Data preparation and filtering steps can take significant amount of processing time. Data pre-processing is used database-driven applications such as customer relationship management and rule-based applications. Data pre-processing includes cleaning, order selection, normalization, transformation, feature extraction and selection, etc. The product of data pre-processing is the final training set. Data Pre-Processing includes data cleaning, data integration, data transformation, data reduction[16] and data discretization. Data reduction is the process of minimizing the amount of data that needs to be stored in a data storage environment. Data reduction can increase storage competence and shrink costs.

3.3 Fuzzy Rule Based System

Fuzzy rule-based systems is based on fuzzy concepts to report complex real problems. Fuzzy sets allow for degrees of set membership, defined by a value between zero and one. A degree of one means that an object is a member of the set, a value of zero means it is not a member, and a value somewhere in-between shows a partial degree of membership. The grade of membership of a given element is defined by the so-called membership function.

Fuzzy set theory provides the tools to effectively represent linguistic concepts, variables, and rules, becoming a natural model to represent human expert knowledge. A key concept is that of a linguistic variable, defined as a variable whose values are linguistic terms, each with a semantic described by a fuzzy set. A linguistic value refers to a label for representing knowledge that has meaning determined by its degree of the membership function. They are expressed in the form “IF A THEN B” where A and B are fuzzy sets.

FRBS employs a supervised classification approach to do the unsupervised classification analysis[14]. It tries to automatically explore the potential classifications in the data patterns and identify them with some interpretable fuzzy rules. Simultaneous classification of data patterns with these fuzzy rules can reveal the actual boundaries of the classifications. The analysis will start by considering the two main conceptual components of these systems, knowledge, and reasoning, and how they are represented. Then, a review of the main structural approaches to fuzzy rule-based systems will be considered. Hierarchical fuzzy systems will also be analyzed.

3.3.1 Model Space Partition

FRBS based on weight factor in Model Space partition. The steps to be followed:

1. Split equally the input and output spaces of the given numerical data into fuzzy regions as the database. In this case, fuzzy regions refer to intervals for the linguistic terms.
2. Fuzzy IF-THEN rules are created for the training data using the database. Calculate membership function for all values in the training data. Repeat the process for all cases in the training data to construct fuzzy rules covering the training data.
3. Grouping the degrees of membership functions in the antecedent and consequent parts are determined.
4. Attain a final rule base after deleting redundant rules. Considering the degrees of rules, redundant rule with a lower degree is deleted.

4 Experimental Results

This section presents the experiment results after testing the proposed system. The proposed system suggests the type of crop to be planted and level of irrigation to be provided for that particular plant depending upon the temperature and humidity in the environment which may increase the agricultural growth. The dataset for a particular location is collected from the TamilNadu Agricultural University comprises weather condition, temperature, wind, humidity and pressure as attributes. The dataset taken from the TNAU is analyzed and preprocessed using Boruta algorithm and the important features are extracted from the dataset. The dataset is reduced and the report sheet is generated using the boxplot images. Figure 2 shows the TNAU dataset.

	date	time	temperature	wind	humidity	pressure
1	2016-01-01	30:00	21	11	83	1019
2	2016-01-01	01:30	21	9	83	1018
3	2016-01-01	02:30	20	7	88	1018
4	2016-01-01	03:00	19	6	88	1018
5	2016-01-01	03:30	19	6	88	1018
6	2016-01-01	04:00	19	6	88	1017
7	2016-01-01	04:30	19	6	88	1018
8	2016-01-01	05:00	19	6	88	1018
9	2016-01-01	05:30	18	7	94	1018
10	2016-01-01	06:00	18	6	94	1018
11	2016-01-01	06:30	18	6	94	1018
12	2016-01-01	07:00	18	Calm	94	1019
13	2016-01-01	07:30	18	Calm	88	1019
14	2016-01-01	08:00	19	6	88	1020
15	2016-01-01	08:30	21	7	78	1020

Figure 2 TNAU Dataset

Boruta algorithm is used to tag the important attributes from the TNAU dataset. Attributes that have significantly worst importance than shadow ones are dropped. The dataset taken from TNAU is analyzed and pre-processed using Boruta algorithm and the important features are extracted from the dataset. The dataset is reduced and the report sheet is generated using boxplot images. Figure 3 shows the output of Boruta algorithm.

```
Boruta performed 84 iterations in 2.697986 secs.
 3 attributes confirmed important: humidity, temperature, time;
 2 attributes confirmed unimportant: date, pressure;
               meanImp medianImp     minImp   maxImp normHits  decision
date          0.000000  0.000000  0.0000000  0.00000 0.0000000  Rejected
time         14.739207 14.712438 12.2769336 17.82592 1.0000000 Confirmed
temperature 15.526399 15.572492 12.4427415 19.60841 1.0000000 Confirmed
humidity    13.665398 13.642024 11.7199598 16.00126 1.0000000 Confirmed
pressure     3.117188  2.931303 -0.1841798  7.04401 0.3333333  Rejected
```

Figure 3 Output of Boruta Algorithm

Boxplots are used to measure how well the data is distributed in a data set. It divides the data set into three quartiles. This graph represents the minimum, maximum, median, first quartile and third quartile in the data set. It is also useful in comparing the distribution of data across data sets by drawing boxplots for each of them. Figure 4 shows the attribute importance graph.

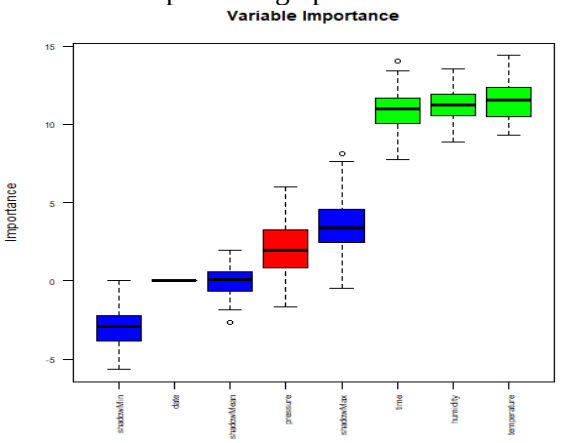

Figure 4 Attribute importance graph

Data reduction is the process of minimizing the amount of data that needs to be stored in a data storage environment. Data reduction can increase storage competence and shrink costs. The reduced set of attributes are displayed in the console for reference. Figure 5 depicts the reduced dataset with relevant attributes.

	time	temperature	humidity
1	30:00:00	21	83
2	01:30	21	83
3	02:30	20	88
4	03:00	19	88
5	03:30	19	88
6	04:00	19	88
7	04:30	19	88
8	05:00	19	88
9	05:30	18	94
10	06:00	18	94
11	06:30	18	94
12	07:00	18	94
13	07:30	18	88
14	08:00	19	88
15	08:30	21	78

p p

Figure 5 Reduced dataset

FRBCS using Ishibuchi's method with weight factor is used to generate IF-THEN rules. It implements the second type of FRBCS which has certainty grades

(weights) in the consequent parts of the rules. The antecedent parts are then determined by a grid-type fuzzy partition from the training data. The consequent class is defined as the dominant class in the fuzzy subspace corresponding to the antecedent part of each fuzzy IF-THEN rule. Considering temperature and humidity IF-THEN rules are generated to suggest the type of crop to be planted[20]. Figure 6 shows the Fuzzy IF-THEN rules.

```
The fuzzy IF-THEN rules:
   V1          V2 V3      V4  V5       V6 V7    V8   V9    V10 V11 V12
1  IF temperature is medium and humidity is small THEN plant  is  10
2  IF temperature is medium and humidity is small THEN plant  is  11
3  IF temperature is medium and humidity is small THEN plant  is   2
4  IF temperature is  large and humidity is small THEN plant  is   5
5  IF temperature is  large and humidity is small THEN plant  is  12
6  IF temperature is  large and humidity is small THEN plant  is  11
7  IF temperature is  small and humidity is small THEN plant  is  10
8  IF temperature is  large and humidity is small THEN plant  is  10
9  IF temperature is medium and humidity is small THEN plant  is  12
10 IF temperature is  small and humidity is small THEN plant  is   7
11 IF temperature is  small and humidity is small THEN plant  is  13
12 IF temperature is  large and humidity is small THEN plant  is   2
13 IF temperature is  small and humidity is small THEN plant  is   6
14 IF temperature is  small and humidity is small THEN plant  is   2
15 IF temperature is  small and humidity is small THEN plant  is   9
16 IF temperature is medium and humidity is small THEN plant  is   3
17 IF temperature is  small and humidity is small THEN plant  is  12
18 IF temperature is  small and humidity is small THEN plant  is   4
19 IF temperature is medium and humidity is small THEN plant  is   5
20 IF temperature is  small and humidity is small THEN plant  is  11
21 IF temperature is  small and humidity is small THEN plant  is   8
22 IF temperature is  small and humidity is small THEN plant  is   1
23 IF temperature is medium and humidity is small THEN plant  is   9
24 IF temperature is medium and humidity is small THEN plant  is  13
25 IF temperature is  small and humidity is small THEN plant  is   5
26 IF temperature is  large and humidity is large THEN plant  is  12
27 IF temperature is  small and humidity is small THEN plant  is   3
28 IF temperature is medium and humidity is small THEN plant  is   8
29 IF temperature is medium and humidity is small THEN plant  is   7
30 IF temperature is medium and humidity is small THEN plant  is   6
31 IF temperature is  large and humidity is small THEN plant  is   3
32 IF temperature is medium and humidity is small THEN plant  is   4
```

Figure 6 Fuzzy IF-THEN rules

The execution of Boruta algorithm is triggered using event-action button in shiny webpage. The graph picturize that three attributes are confirmed. Figure 7 shows the output graph of Boruta algorithm.

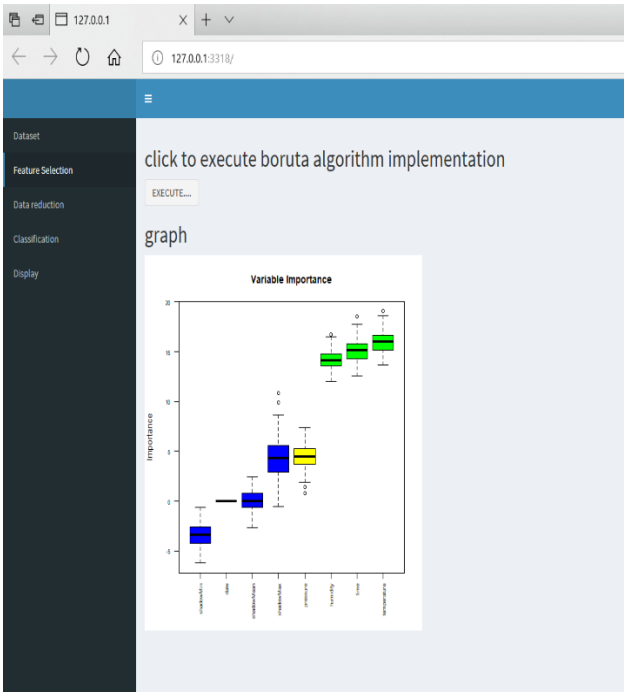

Figure 7 Output graph of Boruta algorithm

The execution of model space partition algorithm is triggered using event-action button in shiny webpage. Figure 8 shows the output of FRBCS model.

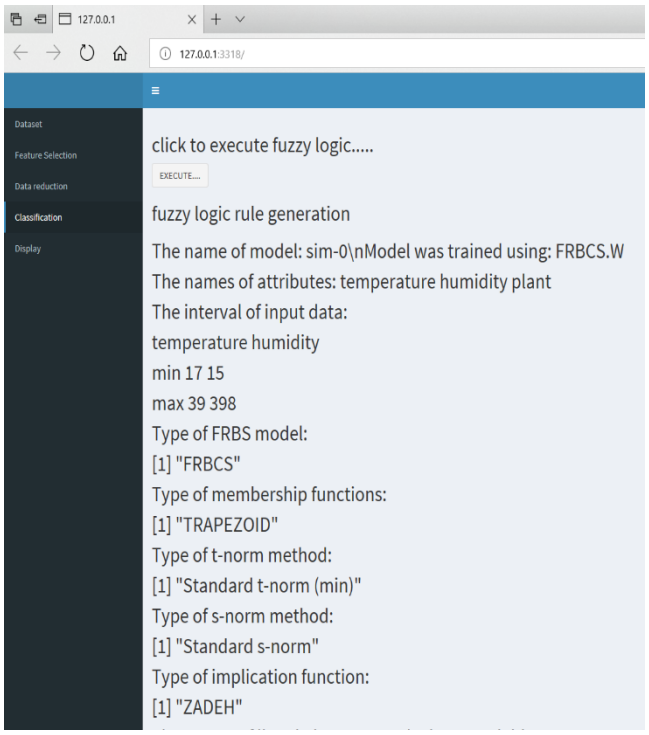

Figure 8 Output of FRBCS model

Using generated Fuzzy IF-THEN rules, based on temperature, humidity and selected soil, the type of crop and the level of irrigation is suggested using shiny webpage. Figure 9 shows the type of crop and respective level of irrigation is displayed.

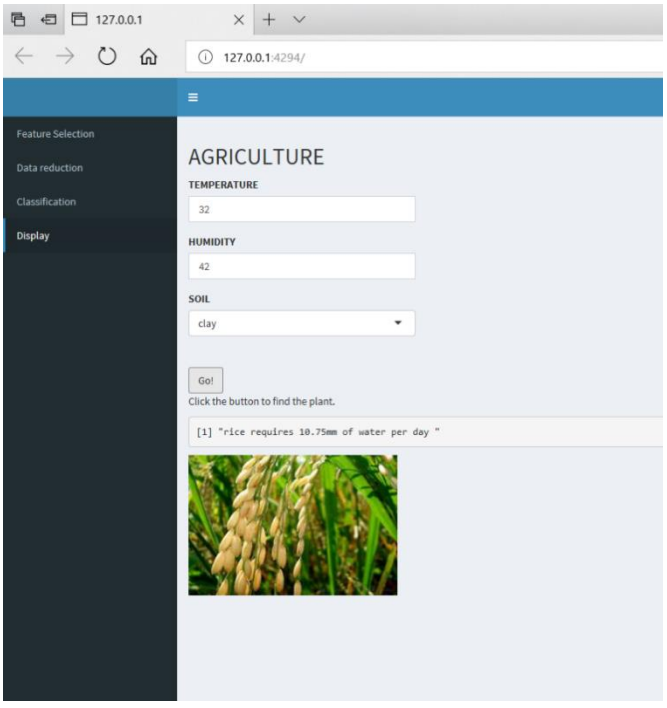

Figure 9 Type of Crop and level of irrigation

5 Conclusion

In proposed system, monitoring solution for farmers with high level, lost cost assistance to accelerate the efficiency of their cultivations and improving the environmental sustainability of the whole process and quality of the production. The presented system is simple and yet efficient. The type of crop to be planted and level of irrigation to be provided is suggested which will improve the agricultural production. The result will be displayed in the webpage and it can be accessed by the farmers.

The result presented in proposed system is promising but several improvements on both materials and methods will be carried out to reach the requirements of fully automated suggestion of crops and irrigation level.

In future development, the type of crop to be planted and the level of irrigation details will be disseminated to the farmers via SMS and mobile applications in the smartphones.

References:

[1] Punit Rathore, Aravinda S.Rao, Sutharshan Rajasegarar, Elena Vanz, Jayavardhana Gubbi, Marimuthu Palaniswami, ”Real-time Urban Microclimate Analysis Using Internet of Things”, *IEEE Internet of Things Journal*, vol. 1, 2017.

[2] Dijana Oreski, Stjepan Oreski, Bozidor Klicek, “Effects of dataset characteristics on the performance of feature selection techniques”, *IEEE Internet of Things Journal*, 2016.

[3] Harine RajaShree, S.Pudumalar, E.Ramanujam, “Crop recommendation system for precision agriculture”, *IEEE Internet of Things Journal*, 2017.

[4] Aakash G Ratkal, Gangadhar Akalwadi, Vinay N Patil, Kavi Mahesh, “Farmer’s Analytical Assistant”, 2016.

[5] Kuljit kaur, Kanwalpreet, Singh Attwal, “Effect of temperature and rainfall on paddy yield using data mining”, 2017.

[6] Federico Viani, Michael Bertolli, Marco Salucci, “Low-Cost Wireless Monitoring and Decision Support for Water Saving in Agriculture”, 2017.

[7] Lala Septem Riza, Christoph Bergmeir, Francisco Herrera, Jos´e Manuel Ben´ıtez, “Fuzzy Rule-Based Systems for classification and Regression in R”, 2015.

[8] Atta-ur-Rahman,Dur-e-NajafZaidi, Muhammad Hamad Salam, “User behaviour classification using Fuzzy Rule Based System”, 2015.

[9] Dr.P.Tamije Selvy,Dr.V.Palanisamy, S.Elakkiya, “Evaluation of classification algorithms for disease diagnosis”, *Journal of Global Research in Computer Science* Volume 4, No. 4, 2013.

[10] I.Guyon, J.Weston, S.Barnhill,V.Vapnik, “Gene selection for cancer classification using support vector machines”, 46(1-3):389–422, 2002.

[11] Miron B. Kursa, “Feature Selection with the Boruta Package”, University of Warsaw Witold Rudnicki, *Journal of Statistical Software* September 2010, Volume 36, Issue 11.

[12] Jeff Heaton, “Fuzzy Logic in R”, *Forecasting & Futurism*, July 2014 , Issue 9.

[13] Anam Mustaqeem,Syed Muhammad Anwar, Muhammad Majid, Abdul Rashid Khan, “Wrapper method for feature selection to classify cardiac arrhythmia”, Engineering in Medicine and Biology Society (EMBC), 39th *Annual International Conference of the IEEE*, 2017.

[14] AntanasVerikas, Jonas Guzaitis, AdasGelzinis, and MarijaBacauskiene, “A general framework for designing a fuzzy rule-based classifier”, Knowledge and Information Systems, 10.1007/s10115-010-0340-x, pp. 1-19.2010.

[15] Yupeng Wang, Umberto Berardi, Hashem Akbari, "The Urban Heat Island Effect in the City of Toronto", *Procedia Engineering*, Volume 118, 2015.

[16] Pritesh Vora , Bhavesh Oza, "Improved Data Reduction Technique in Data Mining", *International Journal of Computer Science and Mobile Computing*, Vol.2 Issue. 5, May- 2013,

[17] Isabelle Guyon,Andre Elisseeff, "An Introduction to Variable and Feature Selection",Journal of Machine Learning Research,2003

[18] A.Shen, M.Kuzlu, M.Pipattanasomporn, S.Rahman, L.Che, "A performance testing method for embedded software platforms", IEEE International Conference on Cyber Technology in Automation, 2016

[19] http://en.tutiempo.net/records/vocb,“TNAU-WEB” [online ,Accessed: 1 Jan.2016]

[20] http://agropedia.iitk.ac.in/.[Agropedia]