Asian Journal of Engineering and Applied Technology
ISSN: 2249-068X (P) Vol.12 No.1, 2023, pp.45-50

DOI: https://doi.org/10.51983/ajeat-2023.12.1.3657

# Agronomic Analysis and Nutrient Application Guidance Using Machine Learning

**K. Sai Saketh Reddy[1], Sengar Vikrant Pratap Singh[2], Bingi Sai Keerthan[3] and B. Ravinder Reddy[4]**
[1,2&3]UG Scholar, [4]Assistant Professor,
Department of Information Technology, Sreenidhi Institute of Science and Technology, Telangana, India
E-mail: 19311a12d9@sreenidhi.edu.in, 19311a12f0@sreenidhi.edu.in, 19311a12f4@sreenidhi.edu.in, ravinderreddyb@sreenidhi.edu.in
(Received 18 April 2023; Revised 30 April 2023; Accepted 23 May 2023; Available online 1 June 2023)

***Abstract*** **- The Indian economy has always been based primarily on agriculture. The industry has met the entire nation's food consumption needs while ranking among the top exporters of agricultural products worldwide. Although this industry has faced its fair share of difficulties lately, few of them are as pressing as they were before Covid-19's limits on local and international travel. Agriculture and farming enterprises have an opportunity to expand as more people become aware of the value of wholesome food and the reasons why terms like "organic" actually matter. Online marketing is a simple way to improve sales, brand awareness, and market share in the farming and agricultural sectors. One of the most significant factors that makes websites useful is that unlike in the past, this growth is enabling business owners and farmers to specify how they must operate. But because of the shortage of labour caused by the reverse labour migration, which hindered harvesting, farmers need a strategy to boost production. Farmers must grow the appropriate crops and apply the appropriate fertilisers in order to maintain a constant output. Many farmers plant crops that are inappropriate for a particular piece of land because they are unaware of this, wasting a tonne of time and resources. Farmers can therefore use web tools to help them find the best crop to cultivate on their land in order to overcome this dilemma.**

***Keywords:*** **Machine Learning, Decision Tree, Random Forest, Database, Python**

## I. INTRODUCTION

It is crucial to safeguard the world's food supply because INDIA has a vast population, and the weather might change at any time. When there is a drought, framers experience serious problems. The kind of soil has a considerable impact on crop productivity. Fertiliser usage advice could help farmers choose the best option for their farming situation. There are many studies that predict crop yield using information and communication technology (ICT). We can also predict agricultural production thanks to data mining. By carefully examining the historical data, we may advise the farmer to plant a better crop for a larger yield. Considerations for a better yield include soil type, soil fertility, as well as one of the key elements, rainfall and groundwater availability. Cash crops are preferable on dry soil, while wheat and sugarcane are preferable on wetlands. In India, there are 15 agroclimatic areas that are separated according to the kind of terrain. Some particular crops can be grown in each agroclimatic area. Based on it, we must advise the farmer on the ideal crop among those native to each region's climate. The endeavor's ultimate aim is to produce the most crop with the lowest output. Precision farming's primary objectives are economic success and environmental sustainability. Agriculture has supported our nation since the dawn of civilization.

Today's climate varies drastically on a regular basis. As a result, understanding weather patterns is necessary for cultivating crops. We must use technology to uncover or appreciate agricultural facts and teach farmers how to create crops properly. Another essential component in the growth of crops is fertiliser. If fertiliser is used excessively or sparsely in the field, the soil may lose its fertility, which could hinder crops from producing as intended. As a result, fertiliser also becomes important. Since we can use the weather to improve the economy of India, understanding the weather is very crucial. Understanding temperature conditions is critical for India most importantly because crop forecasting can help the nation's economy. Crops have a big impact on the Indian economy.

## II. METHODOLOGY

*A. Pre-Processing:* A small number of variables are missing from the given data set; these values could cause erroneous predictions to be made. Python is used to delete the missing values. To obtain accurate findings, it is also crucial to delete replicated data from the data set. To scale the data to a given range, normalisation is utilised.

*B. Support Vector Machine:* One supervised machine learning technique that is frequently utilised in classification and regression issues is the support vector machine. Each data point is portrayed as a point in n-dimensional space, with the value of each attribute substituting in for a particular coordinate.

*Step 1:* Once more, two separate sets are created from the entire training set. (Holdout and train)
*Step 2:* Using the first part (train), train the chosen base models.
*Step 3:* Try them out using these conditions.(holdout)
*Step 4:* Now, the accuracy is assessed using the results of the prediction test.

The two groups of data points can be divided using any number of hyperplanes. The widest margin, or the most significant difference between data points from the two classes, is what we look for in a plane. In order to increase categorization accuracy for upcoming data points, the margin distance should be maximized. Hyperplanes, which act as lines of separation for decision-making, aid in categorising the data points. It is possible to group the data points that are situated on each side of the hyperplane. The overall number of features also affects how big the hyperplane is. The hyperplane resembles a line if there are just two input features. When three input features are available, the hyperplane degenerates into a two-dimensional plane. It is more difficult to picture when there are over three features. The positioning and direction of the vectors of support, or data points, closest to the hyperplane are impacted.. These support vectors are used to raise the tolerance of the classifier. The aforementioned ideas form the framework of our SVM Support Vector Machine.

*C. Output:* The Support Vector Machine Used in this project gives the accuracy over 90.01%.The following information can be entered by the user or the farmer via the online application to obtain the prediction as shown.

## III. MODELING AND ANALYSIS

The user must register for an account and log in. Crop and fertiliser recommendations are made to users using the random forest algorithm and fertiliser retrieval from a database. Users can access their accounts to see suggested crops and fertilisers and to buy them. We employ the apriori algorithm, which generates sets of frequently bought items, to suggest which fertilisers to purchase in pairs. Given the variety of uses it may be put to, this application is highly helpful because it offers both product recommendations and the option to buy them. The selected crops in this work are rice, wheat, sugarcane, onion, and sunflower, jowar, cotton, tobacco, soybeans, and chilli, among others. Crop yield data from the previous five years was gathered from a variety of sources. The suggested system's objective is to help farmers grow crops with a higher yield.

The proposed work consists of three steps.
*1. Soil Classification:* To categorise soil, one might use information on its nutrients. For classifying dirt, machine learning approaches Random Forest and support vector machine learning are utilised. The result of the two algorithms will be a classification, a display of the Recall, Precision, confusion matrix, f1-score, average values, and correctness in %.

*2. Crop Yield Prediction:* Data on crop yields, nutrients, and geographic locations can all be used to predict agricultural yields. The Random Forest and Support Vector Machine algorithms get these inputs. Based on current inputs, these algorithms will forecast the yield.

*3. Fertilizer Recommendation:* Leveraging information on agricultural products, locations, and fertilisers allows for fertiliser prescriptions. This section suggests suitable crops as well as the proper fertiliser for each commodity. Third-party applications display information about the weather, relative humidity, temperature, air pressure, and a basic description.

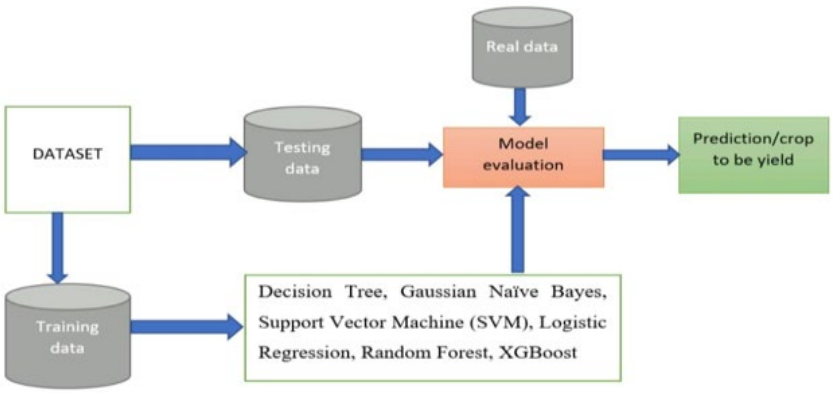


Fig. 1 Data Flow Diagram

*A. Support Vector Machine (SVM)*

SVM creates a hyperplane or collection of hyperplanes in a high-dimensional or infinite space that can be utilised for characterising, detecting relapse, or performing a number of other tasks. Since, generally speaking, the greater the edge, the smaller the classifier's speculative error, the hyperplane with the longest distance to the nearest preparation information purpose of any class achieves a good partition. Since The SVM strategy makes use of translations to make sure that every bit of data is computed up to the variable's position in the first degree, a bit capacity of k(x, y) was used to attain the optimal computational period. A reasonable computational burden must be imposed.

*Advantages*

1. A regularisation over-fitting is avoided by an SVM calculation's parameter.
2. To create expert knowledge for the issue, the component capture is implemented in SVM calculations.

*B. Random Forest*

Based on ensemble learning, random forest is a supervised machine learning technique. You can combine different sorts of algorithms or employ the same strategy more than once in ensemble learning to create a framework for forecasting that is more accurate. The random forest algorithm combines a number of related methods. Classification and regression issues can both be solved using the Random Forest approach.

*Advantages*

1. Provided that there are numerous trees, and they are all instructed using a portion of data, the random forest approach is not biassed.
2. Need to an additional data point be included in the dataset, the Random Forest method is stable and remains unaffected.

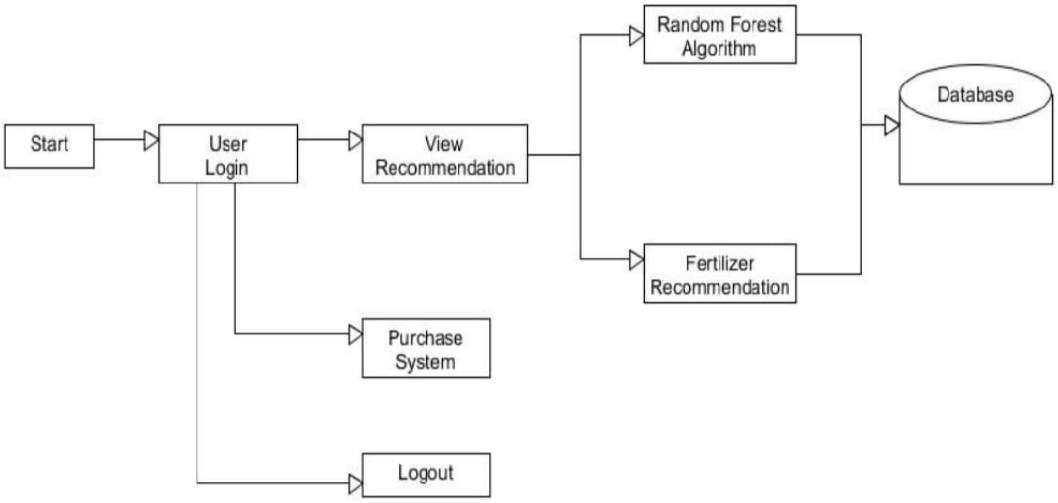


Fig. 2 The architecture of proposed system

The method for predicting crops based on data. The dataset is made up of various data that are taken into account when creating a dataset for crop prediction. The document contains details on previous year's crops, the quantity of rainfall in that area, the kind of soil, crop seasons, temperature, humidity, and agricultural production over the preceding few years.

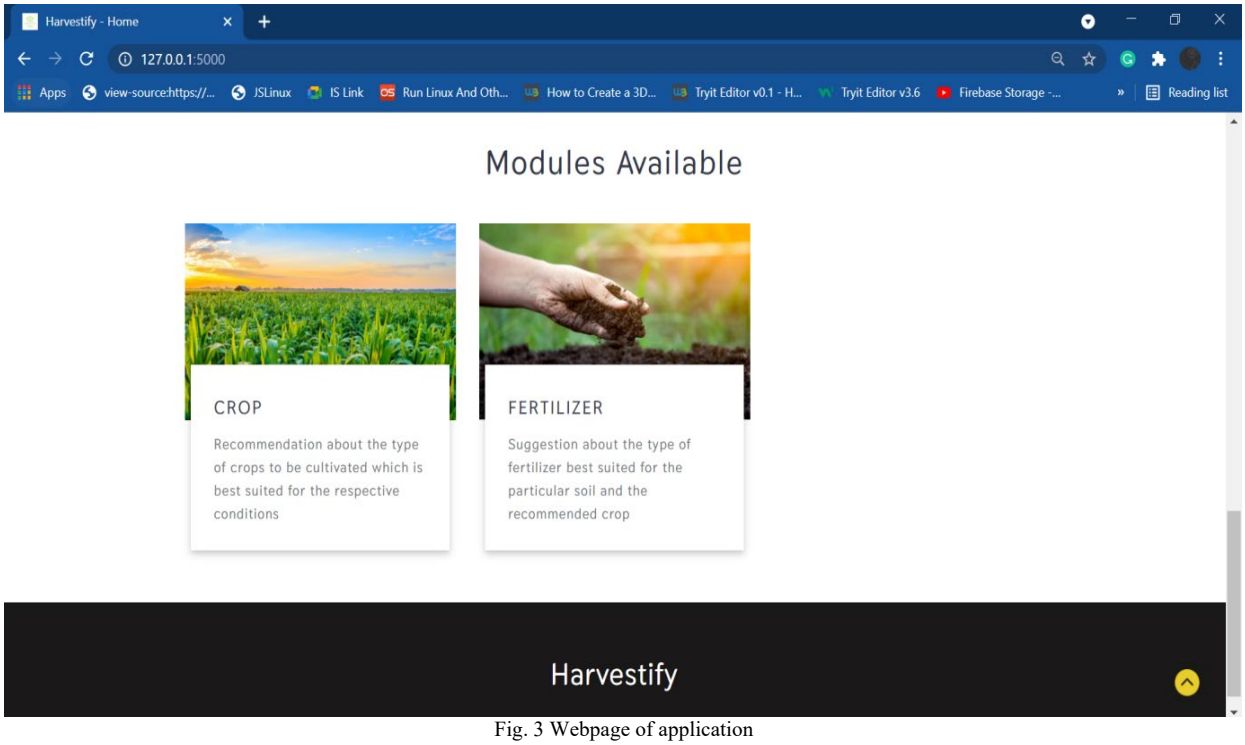


Fig. 3 Webpage of application

The external crop datasets are loaded into the crop prediction technique. After browsing the dataset, pre-processing is carried out in a number of steps as described in the information Pre-processing section. After pre-processing the material, train the models using the coaching set for the call tree classifier. We frequently think about a wide variety of elements, including temperature, humidity, soil hydrogen ion concentration, and anticipated precipitation, while predicting the crop. These are the input variables that can be manually inputted or gathered by sensors for a system. Values for the input parameters and the predicted precipitation are added to a list.

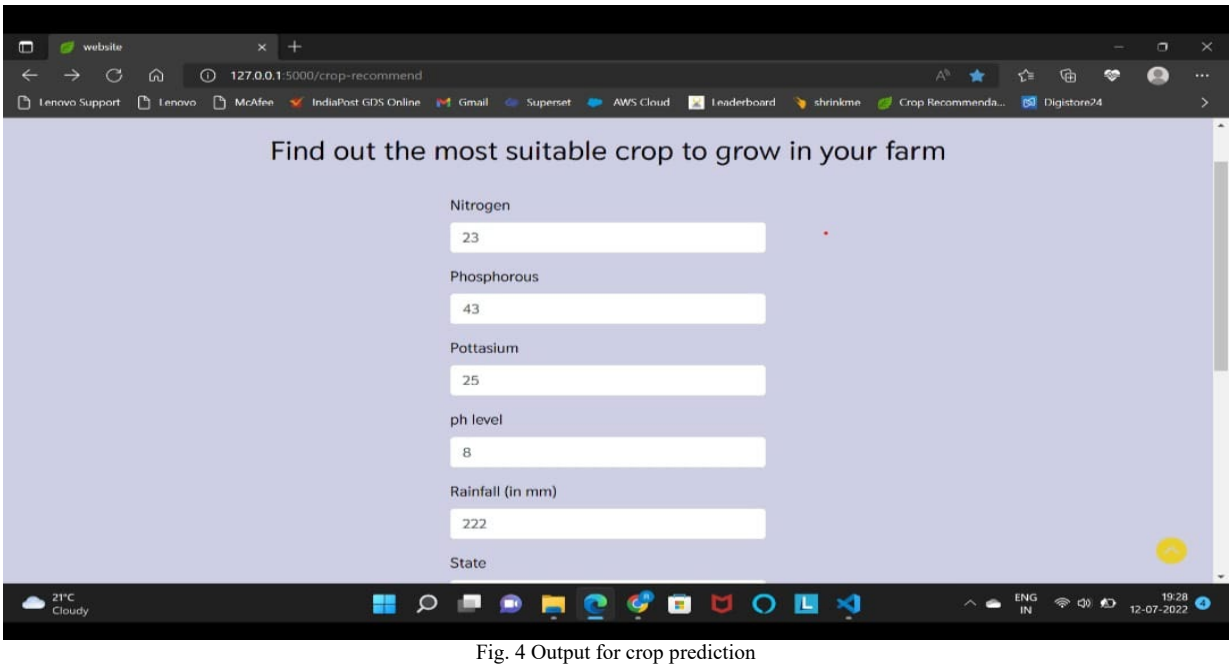


Fig. 4 Output for crop prediction

System applies random forest algorithm using knowledge base in ontology as training set for crop recommendation. Random forest algorithm is used as the accuracy of it is found to be higher than ID3 algorithm for a given dataset. This is because ID3 algorithm constructs only a single tree. If one node/crop is not included into the tree accurately, the entire prediction may be wrong.

A random number of trees is constructed by random forest algorithm and output of random trees are calculated. Final output of random forest algorithm is aggregation of output random trees. Decision criteria for crop recommendation is based on production quantity of the crop and market price of the crop in the specific area. District, state, season are the input parameters for random forest. Random forest generates many numbers of decision trees by extracting training data stored in ontology, each tree predict a crop for given test data. Final output is calculated as the probability of a particular crop predicted by a random tree. Figure 4 shows example of predicted probabilities for crops suitability.

Recommendation of fertilizers is based on Nitrogen, Phosphorous and Potassium measurements from soil. Nitrogen in the soil is responsible for color of leaves. If low quantity of nitrogen is found in the soil, then plants will have slight yellowish leaves and if quantity is moderate or high, it will have greener leaves. The phosphorous content in the soil is responsible for the reproductive system of the plant. Its value will predict the growth of fruits and flowers of the plants. The potassium content of soil is responsible for its overall growth. Its value will predict how stronger the plant roots will be and will also determine the overall growth process of the plant.

Online marketing is a simple way to improve sales, brand awareness, and market share in farming and agricultural sectors. One of the most important elements that makes websites useful is that unlike in the previous, this growth has allowed business owners and farmers to specify how they must operate. But because of the shortage of labour caused by the reverse labour migration, which hampered harvesting, farmers need a strategy to boost production. Farmers must grow the appropriate crops and apply the necessary fertilisers in order to maintain a constant output. Many farmers plant inadequately grown crops for a particular piece of land because they are unaware of this, wasting a tonne of time and resources. On this page, farmers get tips for fertility and crops. The user can add their own soil data in the crop concept module, and its system will then recommend which crop to grow.

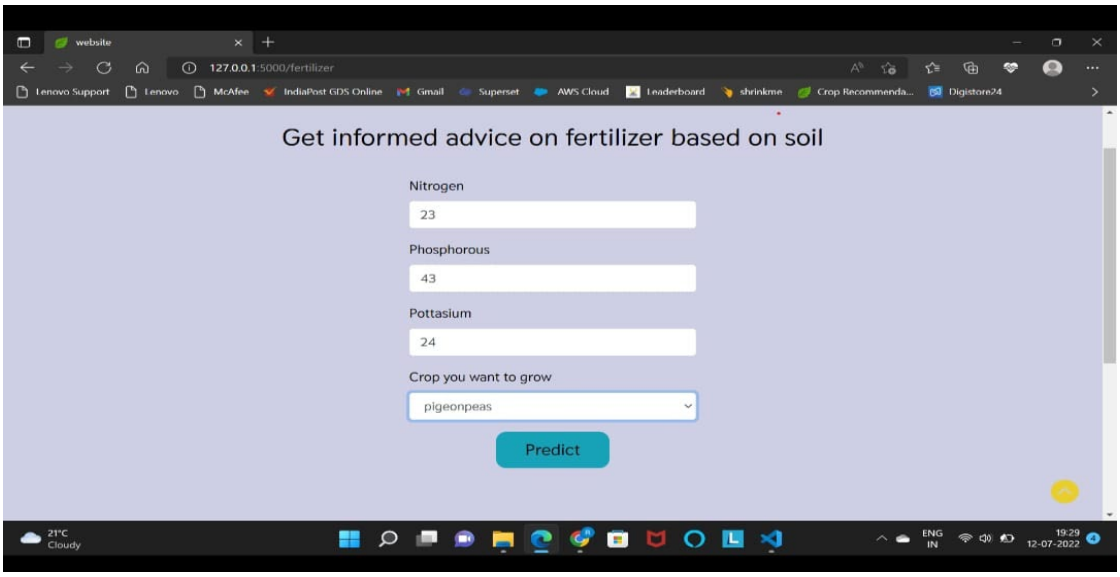


Fig. 5 Output for Fertilizer prediction

## IV. PERFORMANCE ANALYSIS

There are several approaches to evaluate the performance of the model; nevertheless, we decide to use measures like R2 score and MAE. The R2 score is used to evaluate a model's performance. The degree of change in the output dependent characteristic(s) is predicted from the input independent variable.

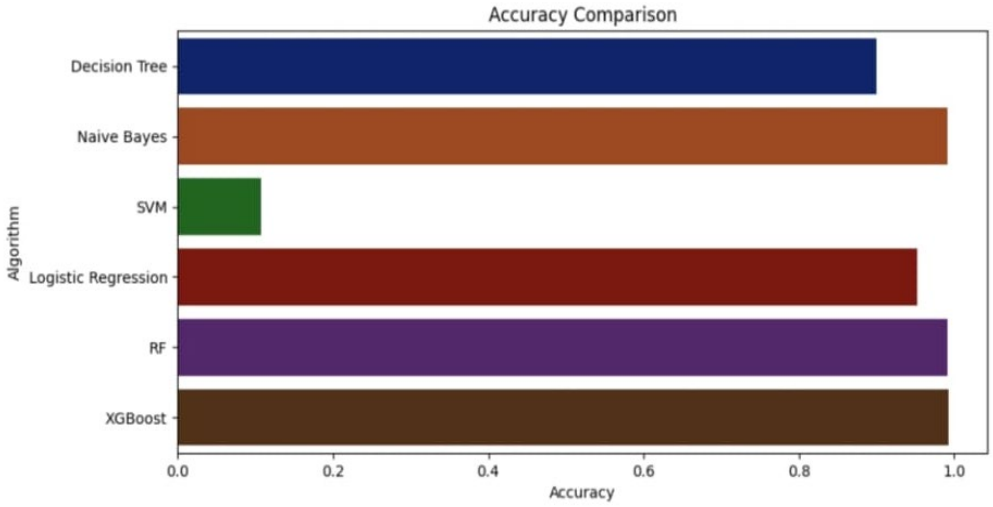


Fig. 6 Bar Plot of Model Accuracy

Analysis of crop yield and classification of soil using graphs. Random Forest's accuracy is 86.40%, whereas Support Vector Machine's accuracy is 74.32%, making it a good algorithm for classifying soil. SVM algorithm accuracy for yield prediction is 99.47%, while RF accuracy is 97.48%. SVM method is therefore effective for predicting agricultural yield.

## V. CONCLUSION

The primary goal of this project is to forecast a crop's nutritional requirements, which depend on the nutrient content of the soil and the growth environment. Farmers will be helped by this method in choosing the best crop for their particular plot of land and in applying the proper quantity of fertiliser to get the highest yield. Based on the previously processed crop data, the Support Vector Machine method aids in the precise prediction of the crop. The newcomers will also benefit from this system's assistance in selecting a crop that will flourish in their region and bring in a healthy profit. More people will become involved in agriculture if there is a respectable quantity of profit. Additionally, crop development depends on the local climate, and since the seasonal monsoons that occur now are unexpected, it is simple for farmers when the outcome of a prediction is also based on the local climate. In addition to assisting users in predicting crop water requirements, live weather forecasting will also assist farmers in reducing crop damage brought on by rain or drought.

## REFERENCES


[1] G. Rushika, K. Juilee, M. Pooja, N. Sachee and R. L. Priya, "Prediction of Crop Yield using Machine Learning," *IRJET,* No. 02, pp. 2337-2339, 2018.

[2] T. Ruchita, B. Shreya, D. Prasanna and C. Anagha, "Crop Yield Prediction using Big Data Analytics," *IJCMS,* Vol. 6, No. 11, 2017.

[3] P. Monali, K. Santosh, Vishwakarma and V. Ashok, "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach," *ICCICN,* 2015.

[4] N. Hemageetha and G. M. Nasira, "Analysis of Soil Condition based on pH value using Classification Techniques," *IOSR-JCE,* Vol. 18, No. 6, pp. 50-54, 2016.

[5] E. Manjula, and S. Djodiltachoumy, "Data Mining Technique to Analyze Soil Nutrients based on Hybrid Classification," *IJARCS,* Vol. 8, No. 8, 2017.

[6] Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare, "Crop Recommendation System to maximize Crop yield using Machine Learning," *IRJET,* No. 12, 2017.

[7] S. Veenadhari, Bharat Misra and C. D. Singh, "Machine Learning Approach for Forecasting Crop Yield based on Climatic Parameters," *ICCCI,* pp. 1-5*,* 2014.

[8] A. Sadia, K. Abu Talha, Mahia Mahrin, A. Wasit and M. R. Rashedur, "Analysis of Soil Properties and Climatic Data to Predict Crop Yields and Cluster Different Agricultural Regions of Bangladesh," *IEEE ICIS,* pp. 80-85, 2018.