



# Implementation of Effective Crop Selection by Using the Random Forest Algorithm

R.Karthikeyan<sup>1</sup>, M.Gowthami<sup>2</sup>, A.Abhishek<sup>3</sup>, P.Karthikeyan<sup>4</sup>

Assistant Professor<sup>1</sup>, Assistant Professor<sup>2</sup>, UG Scholar<sup>3</sup>, UG Scholar<sup>4</sup>  
Department of Computer Science and Engineering<sup>1,2,3,4</sup>

Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi, Tamilnadu

\*Corresponding author E-mail: [karthikeyanr@velhightech.com](mailto:karthikeyanr@velhightech.com)<sup>1</sup>, [gowtamim@velhightech.com](mailto:gowtamim@velhightech.com)<sup>2</sup>, [abhishekanantharaman99@gmail.com](mailto:abhishekanantharaman99@gmail.com)<sup>3</sup>, [keyan1319@gmail.com](mailto:keyan1319@gmail.com)<sup>4</sup>

## Abstract

Accurate predictions of crop yield are critical for developing agriculture. We have provided a machine-learning method, Random Forests which has a ability to predict crop yield corresponds to the current climate and biophysical change. We have collected a large crop yield data from various sources. These data are used for both for the model training and testing. RF was found huge capable of predicting crop yields and over performed MLR standards in every performance statistics that were compared. From various results that shows that RF is an efficient machine-learning algorithm for crop predictions at current condition and has a huge accuracy in data analysis.

**Keywords:** Crop production, Machine learning, Random Forests, Weather Conditions

## 1. Introduction

The agricultural sector is going to face enormous challenges in order to feed the 9.6 billion people that the FAO predicts are going to inhabit the planet by 2050: food production must increase by 70% by 2050, and this has to be achieved in spite of the limited availability of arable lands, the increasing need for fresh water and other less predictable factors, such as the impact of climate change, which, according a recent report by the UN could lead, among other things, to changes to seasonal events in the life cycle of plant and animals. Farmers in many parts of India are largely dependent on timely rainfall for harvest and subsequent profits. Uncertainty surrounding this phenomenon has also haunted them since the beginning of civilization. Over time however, this uncertainty had reduced significantly as farmers back in the day could almost accurately plant crops based on previous experience with weather conditions. This wisdom has been passed on from one generation of farmers to the other. Gradual onset of global warming and climate changes, over the last century, have slowly-yet steadily put this wisdom out of use. As for rain-fed farmers preparing for agriculture, soil-water equation is fragile and any delay in rainfall could easily mar the harvest.

One way to address these issues and increase the quality and quantity of agricultural production is using sensing technology to make farms more "intelligent" and more connected through the so-called "precision agriculture" also known as 'smart farming' with machine learning.

Much like software, improvements in machine learning have seemingly endless possibilities. Researchers in modern agriculture

are testing their theories at greater scale and helping make more accurate, real-time predictions. Modern agriculture has the potential to discover even more ways to conserve water, use nutrients and energy more efficiently, and adapt to climate change. Machine learning in agriculture allows for more accurate disease diagnosis all the while, helping eliminate wasted energy and resources from misdiagnoses. Farmers can upload field images taken by satellites, UAVs, land based rovers, pictures from smartphones, and use this software to diagnose and develop a management plan.

Recent ensemble classifiers such as RF are machine learning algorithms that construct a set of classifiers instead of one classifier, and then classify new data points by taking a vote of their predictions. RF classifier can be defined as the collection of all tree-structured classifiers. RF is an advanced version of bagging. Instead of partitioning each node using the best split among all the variables, this splits every node using the best among a minimal set of predictors which is randomly chosen at that node. A fresh training data set is originated from the actual data set with replacement. Then, the multiple decision tree is grown using random forest algorithm. This strategy of producing 'N' decision tree with 'N' test result makes RF a better in accuracy when it is compared with other existing all algorithms.

## 2. Literature Survey

"Y. Xiaoxia"[1] stated that An prediction algorithm was proposed for soil dampness in view of enhanced BP. The expectation calculation in view of BP neural system and molecule swarm enhancement (PSO) was presented for foreseeing the time arrangement of soil dampness data gained from remote sensor



systems. The time arrangement parameters of BP were resolved and the weight and edge of BP were enhanced by utilizing particle swarm optimization (PSO) calculation.<sup>[1]</sup> The dirt dampness time arrangement was anticipated by BP strategy. Be that as it may, the mean square mistake for BP strategy is high.

“R.V.Chandrashekar “ [2] The multi-gastly and FTIR methods were examined for estimation of soil dampness. The data of soil tests were gathered by utilizing randomized sampling method. Each example was set in fixed polyethylene packs and exchanged for soil dampness and soil surface investigation. The dirt surface was broke down by utilizing hydrometer strategy

“C.Notarnicola”.[3] The dirt dampness was estimated by utilizing gravimetric technique. The impact of the reflectance on various dampness conditions was examined by FTIR spectroscopy technique. The triangulation technique was additionally connected for fitting estimation of soil dampness. Be that as it may, registering power was high and FTIR have single bar while other dispersive techniques have two fold pillar.

“ Rakesh Kumar1”,[4]A fluffy rational based strategy for learning procurements is produced and utilized for recovery of fleeting cases for a situation based thinking (CBR) framework. The strategy is utilized to procure information about what remarkable highlights of persistent vector, one of a kind worldly cases demonstrate huge similitude between cases. Such information is encoded in a closeness estimating capacity what's more, consequently used to recover k closest neighbors (k-nn) from an extensive database. Forecasts for the present case are produced using a weighted middle of the results of closely resembling past cases (i.e., the k-nn, or the simple outfit).

“Ozlem Akar”, [5] stated that Past cases are weighted by their level of closeness to the present case. As needs be, a fluffy k-nn based expectation framework, called WIND-1, is tried,proposed and executed.

‘M. Z.Younis’,[6] stated that is one of a kind segment is an expertly-tuned fluffy k-nn calculation with a fleeting measurement.

“J. Dong “,[7] said that tried with the issue of creating 6-hour expectations of cloud roof and perceivability at an air terminal, given a database of more than 300,000 continuous hourly air terminal climate perceptions (36 years of record).

“A. Balenzano”,[8] stated that forecast precision is estimated with standard meteorological insights and contrasted with a benchmark forecast strategy, tirelessness climatology (PC). In reasonable reenactments, WIND-1 is altogether more exact. It takes around one moment for WIND-1 to create an estimate.

### 3. Existing System

The advancement in information storage is providing vast amounts of data. A huge data set of crop database is extracted. The database contains measurements of soil data from various locations.In

addition to the research establishes whether Soils are Classified Using various data mining techniques, comparison was made between Naive Bayes classification and analyse the most effective technique.

The Naive Bayes classifier assumptions that the present of a particular feature of a class is unrelated to the presence of any other feature. Applying Naive Bayes Data Mining Technique for Crop selection will depending on the nature of the Naive probability model. It can be trained very easy in a supervised learning section. In several practical applications, parameter estimation for naive Bayes uses the method of naive Bayes model with believing in Bayesian probability or using any Bayesian methods.

### 3.1 Disadvantages

1. Must have Knowledge on Bayesian probability or Bayesian methods.
2. Time taken for the process is larger
3. Based on the assumption that features have same statistical relevance.

## 4. Proposed System

This system proposes an approach to analyse large data set. Firstly, this proposal gives a introduction to application of analysis in the massive data analysis in the field of agronomy.Data about weather, irrigation, and yield from several other sources (e.g. meteorological station and irrigation-plan records) for past few decades are collected and analysed to produce an output which has the highest productivity of each grains in their respective geographical conditions.Simultaneously, the data about weather, soil conditon, moisture content, due factor etc. are recorded.From these records the random forest model are trained to evaluate the perfect crop for the current geographical conditions.

### 4.1 Advantages:

1. Powerful and accurate, good performance on prediction.
  2. User able to know the predicted crops values so, they can by plant more effectively.
- Can able to deploy different types of crops by selecting them in the same window

## 5. System Architecture

The diagrammatic representation of our proposed system is as follows:

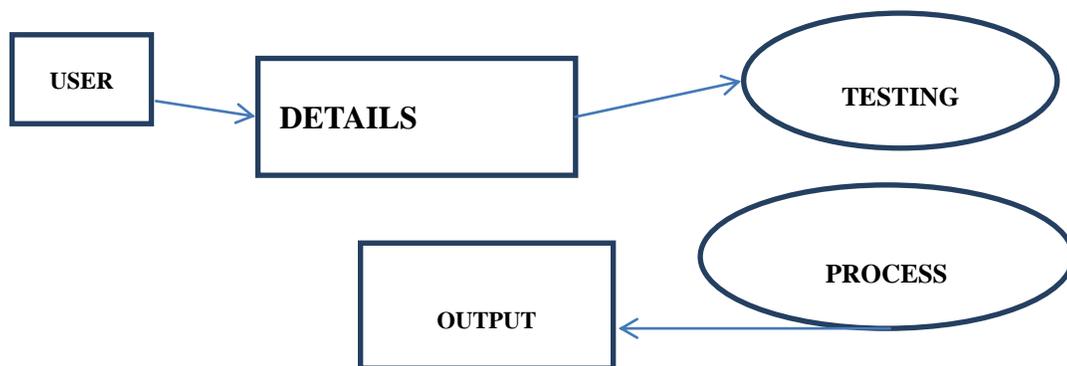


Figure (1): Architectural Diagram

In the above Figure (1) represents that through this system, the user can able to take decision about consulting which crop to deploy by providing exact information of feature of crop . We created a database from various sources. Using this information, the system can provide accurate prediction of particular crop based on learning algorithm.

### 6. Project Modules

Our proposed system is categorized into three modules. They are pictorially represented as follows:

#### 6.1 Module – I: Gathering and Analysing Data

- The large data set has been collected from various sources on crops and their climatic and soil requirement and analysed.
- In this module the User have to register first, then only he/she has to access the account

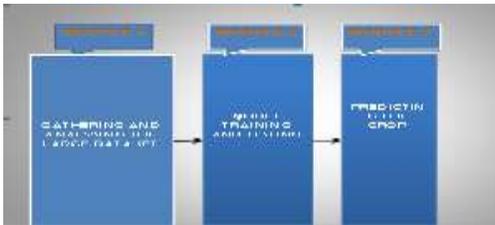


Figure (2): Module Description

- The Registration involves in getting the details of the users who wants to use this application

#### 6.2. Module – Ii: Data Clustering Using C-Means

- After Successful login, User location is fetched and the temperature is calculated
- Random Forest model are trained and tested to produced an high defined accuracy in crop selection.

#### 6.3. Module – Iii Predicting the Crop

- After testing, the next step is predicting the crop.
- The user knows the exact crop to be deployed for the current geographical status.

### 7. Screenshots

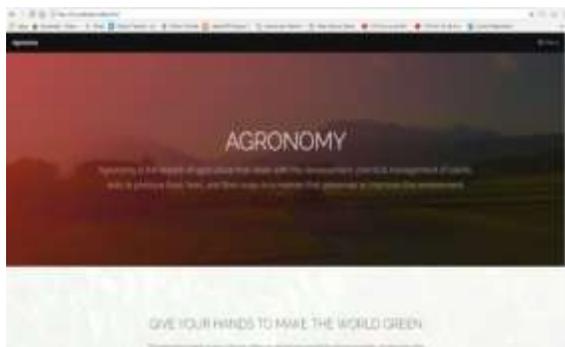


Figure (4): Index page with description about agronomy

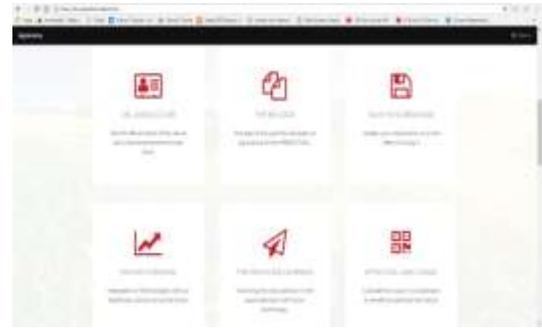


Figure (5): Process of predictive analysis using agricultural data



Figure (6): Description about agricultural productivity

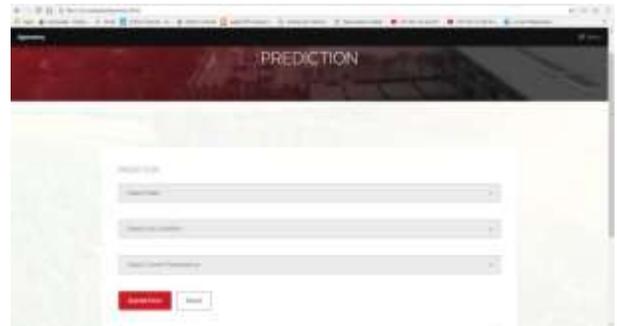


Figure (7): Prediction page



Figure (8): output of predictive analysis

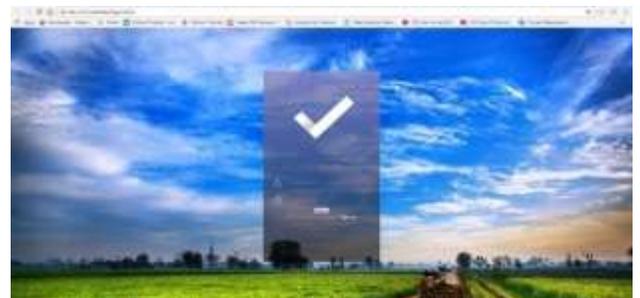


Figure (9): user login

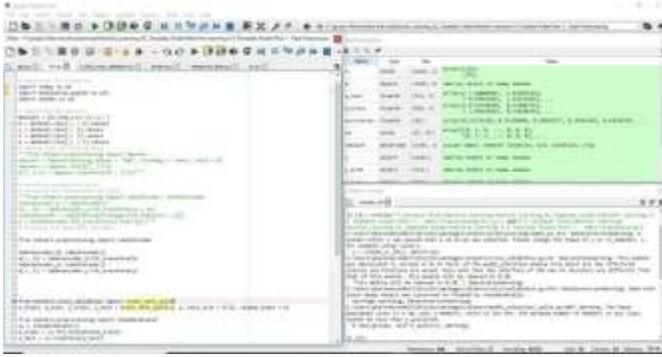


Figure (10): Algorithm for random forest using predictive analysis

retrieval over agricultural crops', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2011,4(2)

## 8. Conclusion

This project is employed to search out the gain the knowledge about the crop that can be deployed to make a efficient and useful harvesting. While the existing system, lags a bit in accuracy in prediction. The ultimate result that was foretold wasn't that accurate with the original. To overcome this drawback, we have a proposed a project to replacement technique with randomforest. It takes only less time for its processes and the accuracy of the prediction is high.

## 9. Suggested Future Work

In future, by using upcoming software's and latest technology, the processing time can be reduced. Efficiency of our module will be increased. Similarly, the same method can be implemented using advanced machine learning algorithm, if it is introduced in near future.

## References

- [1] Y. Xiaoxia, Z.Chengming,'A soil moisture prediction algorithm base on improved BP', 5th International Conference on Agro-geoinformatics (Agro-geoinformatics),2016.
- [2] R.V.Chandrashekar,'Development or end of Agriculture? Implications on Agriculture' , Indian Journal of Innovations and Developments, 2012,1(8), 653-658.
- [3] C.Notarnicola,' A bayesian change detection approach for retrieval of soil moisture variations under different roughness conditions', IEEE Geoscience and Remote Sensing Letters, 2014, 11(2), 414-418
- [4] Rakesh Kumar<sup>1</sup>, M.P. Singh<sup>2</sup>, Prabhat Kumar<sup>3</sup> and J.P. Singh<sup>4</sup>, 'Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique', 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015,pp.138-145.
- [5] Evaluation of random forest method for agricultural croppclassification, Asli Ozdarici Ok, Ozlem Akar & Oguz Gungor European Journal of Remote Sensing , 439-450.
- [6] S. M. Z.Younis, J.Iqbal,'Estimation of soil moisture using multispectral and FTIR techniques', The Egyptian Journal of Remote Sensing and Space Science, 2015,18(2), 151-161.
- [7] J. Dong, S.C. Steele-Dunne, T. E. Ochsner, N.van de Giesen,'Determining soil moisture by assimilating soil temperature measurements using the Ensemble Kalman Filter', Advances in Water Resources, 2015, 86, 340-353.
- [8] Balenzano, F. Mattia,G. Satalino, M. W. Davidson,' Dense temporal series of C-and L-band SAR data for soil moisture