


# Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods

Mamata Garanayak, School of Computer Engineering, KIIT University (Deemed), Bhubaneswar, India

Goutam Sahu, Department of Computer Science and Engineering, Centurion University of Technology and Management, Bhubaneswar, India

Sachi Nandan Mohanty, Department of Computer Engineering, College of Engineering Pune, Pune, India

 <https://orcid.org/0000-0002-4939-0797>

Alok Kumar Jagadev, School of Computer Engineering, KIIT University (Deemed), Bhubaneswar, India

## ABSTRACT

Agriculture is a foremost field within the world, and it's the backbone in the Republic of India. Agriculture has been in poor condition. The impact of temperature variations and its uncertainty has engendered the bulk of the agricultural crops to be overripe in terms of their manufacturing. A correct forecast of crop expansion is a vital character in crop forecast management. Such forecasts will hold up the federated industries for accomplishing the provision of their occupation. ML is the method of finding new models from giant information sets. Numerous regressive ways like random forest, linear regression, decision tree regression, polynomial regression, and support vector regression will be used for the aim. Area and production are among the meteorological information that's made by necessary data. This paper figures out the yield recommendation of the crop by the accurate comparison of numerous machine learning ML regressions where the overall percentage improvement over several existing methods is 3.6%.

## KEYWORDS

Crop Yield, Decision Tree (DT) Regression, Linear Regression (LR) Prediction, Machine Learning, Polynomial Regression (PR), Random Forest (RF) Regression, Support Vector Regression (SVR)

## INTRODUCTION

Agriculture is the leading hold up and the paramount territory of the Indian wealth. The manufacturing of farming is very little. As the ultimatum for daily bread is heighten epidemically, the farmers, investigators, analysts, scientists, and government attempt to site further attempts and schemes to heighten the agricultural manufacturing to lodge the needs (Shastry et al., 2017). India is generally stubby despite being a huge sector and yields of crops per hectare. Correct productivity of crops hangs on numerous parameters such as properties of soil, irrigation, terrain, and climate.

Owing to several components such as change of climate, tumble levels of water, accidental rainfall, imprudent utilize of bio-pesticides, etc., the intensity of agricultural manufacturing is diminishing in India. The majority of farmers do not attain awaited crop yield for a variety of grounds (Kumar et al., 2018) To acknowledge manufacturing intensity, yield manufacturing is carried out which

DOI: 10.4018/IJAEIS.20210101.oa1

This article, published as an Open Access article on February 26th, 2021 in the gold Open Access journal, the International Journal of Agricultural and Environmental Information Systems (converted to gold Open Access January 1st, 2021), is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

requires forecasting the yield of the crop relied on the existing information. Formerly, crop production estimations were based on farmer's certain crops and experience of cultivation.

There are numerous methods to amplify and upgrade crop production and standards. Varieties of researches have been conducted to grow a well-organized technique for yield forecasting but focal points have been consistently on analytical techniques and not much has been carried out in machine learning (ML) approaches. The crop manufacturing hangs on different components (Renuka & Terdal 2019) which varies with every square meter and hangs on:

1. Geographical region;
2. Climate (Temperature, moistness, precipitation);
3. Types of loam (saline, alkaline, and non-alkaline, etc.);
4. Composition of loam (ph, N, P, OC, Zn, F, K, and EC, etc.).

Varieties of subsets of the above specifications are applied in several forecasting models for different crops. Forecasting models are generally of 2 types.

1. Analytical models that employ a solitary forecast function that considers every single space of samples.
2. Technologies of machine learning, a recent technology for knowledge explore that associates input and output.

Learning the machine without elucidated computer programming is one of the capabilities of machine learning techniques, so it enhances machine production by discovering and distinguishing the stability and design of operating information. Machine learning can be categorized into 3 broad categories according to the methods of learning –Supervised, unsupervised, and Reinforcement learning (Singh et al., 2017). In our paper, we are building the work with supervised algorithms to forecast crop production. These categories of algorithms assist to construct the most precise and effectual model as here, the learning information occurs with labels or required outputs and the objective is to discover a common rule of depicting input to output. It presumes to construct a machine learning model that is relied on labeled samples.

This paper predicts the accuracy of the future production of five different crops such as rice, ragi, gram, potato, and onion crops using various supervised machine learning approaches in of Andhra Pradesh region and recommends the crop to yield. The dataset is collected from the statistical and agricultural department of Andhra Pradesh, it consists of precipitation, yield, cloud cover, vapor pressure, season, production, and area dataset. The Linear, Decision trees, Random Forest, Polynomial, and Support Vector Regressions have been utilized for crop production forecasting.

An Agricultural farmer is always interested to know whether how much yield he/she is about to produce. In the past times, Predictions on crop yield on different crops was performed by considering farmer's experience on specific field and crop. The crop production is effected by variegated seasonal, biological and economical constituents but unforeseeable changes in these constituents lead to a huge loss to farmers. In Most of the Cases, Farmers even commit suicide because of not able to pay the bank loans taken for farming due to production loss. These risks can be minimized when significant mathematical or statistical methodologies are applied on data related to soil, weather as well as past yield and using these Methods, We can recommend the Best Crop to farmer for his Agricultural land so that it helps to get maximum profit.

The paper is collocated as: Segment 2 presents the associated work or literature review, whereas the proposed approach is discussed in Segment 3. Then, the experimental results and performance analysis on agricultural information are discussed in Segment 4. In the end, the conclusion is given in Segment 5.

## LITERATURE REVIEW

Aditya Shastry, H. A. Sanjay, and E. Bhanusree (2017) describe that the regression techniques can be utilized for yield prediction for the area with satisfactory results. The attempt exhibits that the regression method can be used for crop forecasting for the geographical region with adequate outcomes. India is remarkable amongst most of the countries in generating production in Asia and the usage of different crops in several pieces of our country is perceived widely. The regression model, the forecast of the generation of wheat, maize, and cotton is done for chosen years. The outcome demonstrates that the proposed regression model is a method for foreseeing yield prediction. Here the result different models are compared based upon the root mean square, R<sup>2</sup> statistics, and percentage prediction error. The prototype which permits the bottom root mean square, percentage prediction error, and elevated R<sup>2</sup> statistics outcome is considered to be the finest prototype model for crop production forecast.

Arun Kumar, Naveen Kumar, and Vishal Vates (2018) uses the ideas of descriptive analytics in the agricultural area. The analysis work comes up with the data concerning whatever may be the applied data analytics on sugarcane crop datasets. In this paper 3 crucial algorithms of supervised learning such as k-NN, SVM, and LS-SVM to train and construct the model. This paper is principally according to the relative learning of several procedures when we appeal this method on datasets and it manifests the correctness of each method to train the datasets and also mean squared error (MSE) at the cross-validation phase of the test data. Again this experimentation effort can be intensifying to the succeeding stage. From this, they can construct a recommender system of crop yield and dispensation for farmers. By which agronomists can build a conclusion in which season which crop should disperse so that they can get increased profit. This practice is a service for a structured dataset. In later we can implement data independent practice also. It means whatever the format of data, our system should serve with some coherence.

Renuka, Sujata Terdal (2019) manifests the consequences acquired after implementation of the algorithms of machine learning on sugar cane crop information set of Karnataka, India. The algorithms k-NN, SVM, and decision tree (DT) are appealed on production information sets, soil information sets, and rainfall information set. To evaluate the best performing method verities of machine learning algorithms have been implemented on agricultural information. Here, they applied 3 verities of supervised learning algorithms, such as SVM, k-NN, and DT, and also furnished data on how to appeal information analysis to the information set of sugar cane. This procedure will assist to diminish the complications faced by agronomists and will set out as an arbiter to impart agronomists with the particulars they need to collect towering payback and maximize net surpluses.

Vaneesbeer Singh, Abid Sarwar, and Vinod Sharma (2017) gives an approach which utilizes several techniques of machine learning to forecast the classification of the production relied on macro and micro-nutrients repute in the information set. Machine learning algorithms are applied after analysis to forecast the classification of production. The classification, thus forecasts will identify the production of crops. The difficulty of yielding the crop is constructed as a classification in which several classifier algorithms are applied.

S. MamathaJajur, Soumya N. G. and G. T. Raju (2019) labor will well-being agronomists to amplify yielding in agriculture, decreases soil degradation in cultivated grasslands and decreases fertilizer applied in crop yield by prescribing the correct crop considering several attributes. The suggested labor aids agronomists incorrectly choosing the crop for cultivation and accomplish unceasingly. In the hereafter, the suggested set-up can be expanded to contemplate retail insistence and availability of retail infrastructure, awaited surplus and possibilities, and upright ingathering storage and processing mechanization. This would furnish radical forecasting based on geographical area, climatic situation, and the profit-making aspects.

ZeelDoshi (2018) considered 5 significant crops (maize wheat, jawar, bajra, rice) and 15 insignificant crops (pulses, ragi, potato, tur, rapeseed and mustard, jute, gram, barley, cotton, groundnut, sesame, soybean, sugarcane, tobacco, sunflower) and uses the DT, NN, RF, and k-NN

approaches to discover the percentage of accuracy and also discovers that the percentage of accuracy is elevated for NN (91%).

Z.H. Khalil, S.M. Abdullaev (2020) discovered the crash of weather trends on winter productions of Al-Diwanyah Iraq, whenever wheat and barley are representing the important crops during this area. It's been marked that the environmental circumstances fluctuate in Diwaniyah, which are identified by soaring temperatures and minimum causative, have any apprehension on crops. This could end to the temperatures trend soar, minimum precipitation, and wilted winds origin a rise within the evaporation rate and so the looks of issues of soil. Despite the elevated than, the barley and wheat crops manifest an increase in tendency, however, this growth wistfully didn't ceil the rate of import, basically in wheat, thus this analysis results is much more important for organizations and agriculture centers in Al Diwanyah for completely exploit climate changes and fluctuations for enhancing and maximize crop productions.

Teresa Priyanka, PratishtaSoni, C. Malathy (2018) describes that the production of crop prediction still concerns to be a tough issue to face by the agronomist. The focus of this work is to the purpose and implement a rule-based system to forecast the crop production from the pool of collected ancient information. This has been attained by applying both ANN and CNN using satellite imagery on agriculture to anticipate the several crop yields.

S. Veenadhari, Dr. Bharat Misra, and Dr. CD Singh (2014) proposed the potential utility of data mining methods in forecasting the crop production relied on the climatic parameters. The evolved webpage is customer friendly and therefore the correctness of forecast is more than 75% in all the crops and districts chosen in the paper. The customer-friendly Online page evolved for forecasting crop production are often employed by any of the customer and their alternative of the crop by providing environmental surroundings of that area.

Alberto Gonzalez-Sanchez, Juan Frausto-Solis, Waldo Ojeda-Bustamante (2014) suggested a work that collates the forecast accuracy of machine learning regression methods for crop production forecast in datasets of 10 crops. Multiple regression, M5-regression trees, perceptron multilayer NNs, support vector regression, and k-nearest neighbor techniques were queued. Out of all 4 accuracy metrics were applied to uphold the models: the root means square (RMS) error, root-relative square error (RRSE), mean absolute error (MAE), and correlation factor (R). The outcomes show that M5-Prime and k-nearest neighbor techniques acquire the less average RMSE errors (5.14, 4.91), the less RRSE errors (79.46%, 79.78%), the bottom average MAE errors (18.12%, 19.42%), and therefore the high rise average correlation factors (0.41 and 0.42).

Jichong Han, Zhao Zhang, Juan Cao, YuchuanLuo, Liangliang Zhang, Ziyue Li, Jing Zhang (2020) forecast winter wheat production at the county scale based on multi-source information and the multiple numbers of machine learning models. It had been found that SVM, RF, and GPR anticipated wheat productions with rising accuracy, and RF incontestable the simplest generalization ability among the 3 methods. In China, the RF model will estimate wheat production correctly before the harvesting days. Moreover, here they looked over the clash of the pulse window chosen on the forecast correctness and initiate the window has towering forecast correctness all over the heighten interval. They also established that the forecast correctness diverted by farming sectors and step by step procedures, and the topographical dissimilarity will influence the crop production forecast correctness. In summation, EVI is the most supreme predictor utilized in this learning for winter wheat production. They are confident that the substructure to predict winter wheat production by diverse-source data and the GEE manifesto is comprehensive and relevant to other crops in the globe.

Phusanisa Charoen-Ung and Pradit Mittra piyanuruk (2018) proposed 2 separate techniques for differentiating the sugarcane production category at the plot level from the information about characteristics of the plot. The first technique is based on RF where implementation is based on the RF of scikit-learn and the second technique is relied on GBT (Gradient Boosting Tree) where implementation has relied on the XG-Boost library. The correctness of these two machine learning relied upon forecast techniques are 71.83% and 71.64%. Several feasible matters for subsequent effort

can be recorded as below. The earliest one can learn the upgrade in the forecasting correctness in the instance that every day the temperature statistics in the regional region and more statistics about the clay features of each plot are accessible as in some connected service. Second, an inclusive study about the results of hyper-parameter adjusting, attributes engineering, and attributes selection on the dataset. Third, we can appeal a prototype stacking method to upgrade the forecast correctness. Finally, we can work out the issue as the regression complication.

Yvette Everingham, Justin Sexton, Danielle Skocaj, and Geoff Inman-Bamber (2016) presented a sugarcane production forecast technique by using Random Forest. The characteristics applied in this paper include the index of biomass, information of climate (e.g., rain falls, radiations and temperature, etc.) and productions from 2 preceding years. 2 predictive tasks are representing here: (1) the classification problem for forecasting whether the production will be above or below the observed median production (t/ha), and (2) the regression downside for forecasting the production estimates in 2 completely separate times. The motive of this article was to ascertain if a data mining perspective could offer up to date perception that can elucidate sugarcane production in the Wet Tropics, Australia. Forecasting the size of the crop can impact farmland resolution such as how much nitrogen fertilizer to utilize and assist millers carefully scheme conservation and exertion plan of action to be prepared for the beginning of the milling condition. Marketers can appeal the identical grip to enhance industry gain across additional potent and chosen ahead selling plan and logistical strategy. These advancements in profit-making and climatic outcomes are the principal components for conveying imperishable emulsions to the Production Company.

Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler, Dennis J. Timlin, Kyo-Moon Shim, James S. Gerber, Vangimalla R. Reddy and Soo-Hyung Kim (2016) proposed the efficacy of RF regression using machine learning as a standard to model complex production reactions of wheat, grain maize, potato, and silage maize at universal and areal scales. The RF algorithm program has different blessings to regress complicated systems of the crop, however isn't nonetheless being widely utilized in this area. They are incontestable that RF provides superior performance in forecasting the production of all crops and areas tested. The result of this manifests sturdy potential for the implementation of an RF algorithm as an alternate applied mathematics modeling method for crop production forecasting. Moreover, the outcome of this learning manifests sturdy dormant for the execution of an RF procedure as another statistical prototyping method for crop production predictions. It is pointed out that RF has the possibility of overfitting information for the state where training data are intense while its correctness can reduce where training data were scarce. Similarly by appealing RF regression to figure outside training data proportion should be kept away. Here their outcomes support that RF regression can be an effective method for forecasting crop production at the worldwide and regional scale with the attentive selection of training data.

Ms. Fathima, Ms. Sowmya K, Ms. Sunita Barker, Dr. Sanjeev Kulkarni (2020) proposed the possible utilization of data mining methods in forecasting the crop production relied on the insert attributes mean rainfall and area of land. Here the expanded webpage is client-friendly and the correctness of forecasting is beyond 90%. The districts selected in the learning indicating elevated correctness of forecast. The customer-friendly web page evolved for forecasting crop production can be utilized by any customer by providing mean rainfall and land of that area. The method was acquired for all the land to upgrade and validate the validity of production forecasting which are practical for the agronomists for the forecast of a particular crop. The later service pointed at the inspection of the whole set of data and will be loyal to the worthy master plan for better the efficiency of the suggested method. Utilize such kind of prototype to predicting is not limited to agriculture only. The clustering and regression are one of the capable technologies in the area of data mining which can be utilized in several processes.

Amit Sharma, Pradeep Kumar Singh, Ashutosh Sharma, Rajiv Kumar (2019) suggested an architecture that is capable of observing the required environmental variables and therefore revealing

an event at its primary stage. The nodes that are used for sensing continuously keep track of variables until the occurrence of fire succeed. The work attempts to find the chances of wireless sensor network and un-manned aerial vehicles in the applications of management of disaster. The principal aim of this work is to design a system that perform correct detection along with it confirmation through images in real time from un-manned aerial vehicles.

Mamata Garanayak, Sachi Nandan Mohanty, Alok Kumar Jagadev, Sipra Sahoo (2019) proposed a movie recommender system by utilizing item-based collaborative filtering and K-Means clustering techniques which gives near about 80% accuracy. Here they consider m number of users, n number of items and presenting a recommendation model to fabricate the movie recommendation by calculating the similarity between the movies (items).

Mamata Garanayak, Sipra Sahoo Sachi Nandan Mohanty, Alok Kumar Jagadev (2020) proposed an automated recommender system for educational institute in India which predicts the ranks for 2019 for each branch of every IITs. This paper also illustrates prediction using Time Series Forecasting and recommendation algorithm using classification techniques. Also a comparative study of Random Forest Classification and K-Nearest Neighbor classification has been done and finally, the recommendation algorithm shown reliable results with 94.11% accuracy in prediction model.

## **PROPOSED APPROACH**

The Proposed Model is shown in Figure 1. In the suggested approach, first, we collect the data set from the Andhra Pradesh Govt. then pre-processing and extractions of the feature phase have been done. In the pre- the processing of the data phase, we load the missing values, the appropriate data range and use the RF classifier technique to choose the area and production attribute as our final attribute for implementation. Finally, we apply various supervised machine learning regression methods to build a prototype, which comes up with the accuracy and forecast of each crop production and recommendation. The system is organized in the following phases such as a collection of information sets, preprocessing phases, selection of features, and appealing different ML regressions such as linear regression, Decision Tree (DT) regression, Random Forest, Polynomial and Support Vector Regression methods.

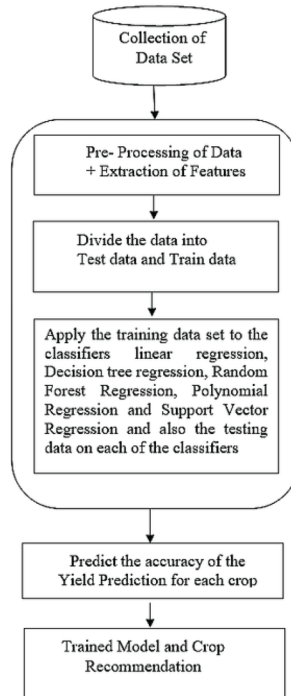
### **Collection of Data Sets**

A dataset is a gathering of information. In the instance of horizontal data, a data set communicates to one or an additional database board, where every column of a board constitutes a specific variable, and every row constitutes an identified record of the data set in the query. Data for five different crops such as rice, ragi, gram, potato, and onion are collected from the Andhra Pradesh Government and the data sets are prepared for processing. The five data sets for five different crops used in this paper are crop yield data having attributes like precipitation, cloud Cover, area, vapor pressure, season, yield and production, etc.

### **Preprocessing Step**

In machine learning, this phase is a very crucial phase that assists in increasing the standard of the data to develop the uprooting of significant insight from the data. Preprocessing of the five data sets of five different crops (rice, ragi, gram, onion, and potato) consists of loading the missing values, the proper range of the data, and the functionality extraction. In this paper, we have applied the IsNull () technique for checking null values and label encoder () for transforming the absolute data into numerical data.

Figure 1. Proposed Model



## Features Selection

Feature extraction is the activity where we manually chose those attributes which donate most to the forecast variable or outcome in which we are attentive. Having immaterial attributes in the data can lessen the correctness of the prototypes and build the prototypes to learn to rely on immaterial characteristics. Feature extraction of five data sets of five different crops (rice, ragi, gram, potato, and onion) should simplify the amount of data involved to represent a huge data set. We have applied the RF Classifier () technique for the selection of attributes. This technique picks the attribute which is having increased entropy value as the main feature for the accurate forecast of crop production forecast.

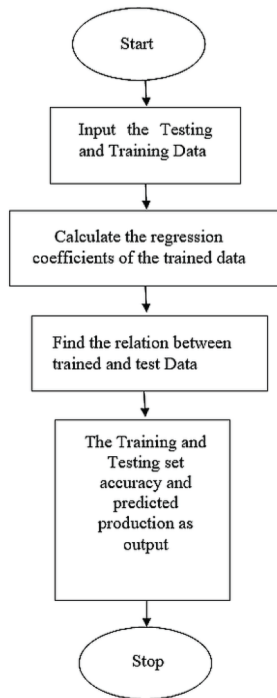
## Break the Datasets Into Train Sets and Test Sets

This phase covers training and testing data of the five different crops such as rice, ragi, gram, potato, and onion as input data. The loaded data of all the five crops are split into 2 sets, such as train sets data and test sets data, with a division ratio of 70% or 30% (0.7 or 0.3). In the training sets, a classifier known as an RF classifier is applied to build the obtainable input data. In this phase, we create the classifier's reinforce information. During the testing stage, the data are tested and the concluding data are formed during pre-processing and are refined by machine learning.

## Appealing ML (Machine Learning) Regression Components

In this paper, we have used five several crops (such as rice, ragi, gram, potato, and onion) and five several supervised ML techniques such as linear regression, decision tree regression, random forest regression, polynomial regression and support vector regression for accurate calculation of each crop's production forecast as described below.

Figure 2. Linear Regression



### Linear Regression

In LR (linear regression), the depiction is a linear equation that unites a particular set of input value (m), to the solution to which is the forecasted output for that set of output value (n). Both the input value and the output value are numerical. In the linear regression method first, the testing and training data are inputted to calculate the regression coefficients of the trained data and find the relation between the test and train data. The training and testing accuracy were calculated as output. The flowchart of Linear Regression is given in Figure 2.

The linear equation allocates one scale factor to every input value or column, called a coefficient and let that is constituted by C. One supplementary coefficient is also attached, giving the line a supplementary degree of freedom and is called the intercept coefficient or bias coefficient. For example, in a very statistical procedure drawback (an isolated x & one y), the shape of the representation of equation would be:

$$y = C_0 + C_1 * x \tag{1}$$

Here we got that the production grades classification accuracy of five different crops on the training and the testing sets which is shown in Table 1.

### Decision Tree (DT) Regression

Decision Tree (DT) is a foreboding representation that works by examining states at every level of tree and moves towards the bottom of the tree where several decisions are recorded. The state depends on the application and the result might be in terms of the decision. In this method, we calculate the information gain of every attributes such as precipitation, cloud cover, area, vapor pressure, and



Table 1. Prediction accuracy of different crops using linear regression

Crops	Accuracy (%)	
	Training Set	Testing Set
Rice	86.61	91.90
<u>Ragi</u>	93.10	92.76
Gram	91.62	82.56
Potato	94.10	87.69
Onion	94.73	93.56

production. The two most important attributes tie with the same value of information gain are area and production and by using these selected attribute a decision tree was made and then expanded.

Here we got that the production grades classification accuracy on the training and the testing sets of all the five crops which is as shown in Table 2.

#### *Random Forest (RF) Regression*

This is a group of the methodology of learning which is usually applied for each regression and classification. To coach the representation to achieve forecast exploitation of this rule, the test characteristics must proceed through the rules of each tree created randomly. As an outcome of this method, a distinct target will be forecasted by each random forest for the same test characteristics. Then, votes are counted based on each forecasted target. The concluding forecast of the algorithm is the highest vote forecasted target. The random forest algorithm can able to handle gone values efficiently.

Table 2. Prediction accuracy of different crops using decision tree regression

Crops	Accuracy (%)	
	Training Set	Testing Set
Rice	99.99	81.80
<u>Ragi</u>	99.99	87.24
Gram	99.97	85.43
Potato	99.67	73.55
Onion	99.71	93.43

The random subset of training information is sampled during the training of the RF model; there will be 30% of training instances that are not applied in each Decision Tree building. These illustrative are called OOB (Out Of Bag) instances. We can also apply this OOB illustrative to calculate the forecasting accuracy of the RF model by averaging the evaluations of OOB instances of the decision tree of the random forest model. In this paper, we apply the RF model in scikit-learn. The input features are area and production for five different crops.

Here we got that the production grades classification accuracy on both the training and the testing set of five different crops which is as shown in Table 3.

Table 3. Prediction accuracy of different crops using random forest regression

Crops	Accuracy (%)	
	Training Set	Testing Set
Rice	97.84	86.86
Ragi	99.17	90.30
Gram	98.29	84.95
Potato	99.12	73.91
Onion	98.86	94.78

### Polynomial Regression (PR)

In this proposed model PR could be a type of regression where the unconventional variable is taken as area (x) and the dependent variable is taken as production (y) and we attempt to notice the connection between them. PR fits relationship which is nonlinear between the x value and therefore the corresponding mean (conditional) of y, represented by  $E(y|x)$  (Figure 5).

In this case, we got that the production grades classification accuracy on both the training and the testing set of five different crops as shown in Table 4.

### Support Vector Regression

Support Vector Regression splits the specified data into the decision surface. The decision surface is again split the data into two classes of the hyperplane. Training points define the supporting vector that defines the hyperplane. Probably, a hyperplane with the maximum span to the closest learning data point typically has huge errors and superior margins because of the superior margins; the classifier's generalization is weak. The flow chart for support vector machine is given in figure 6 which first break the data into train and test sets and after that by using the training data sets of five different crops SVM find the separate hyper-plane and finally using the testing data sets it to apply the separating the hyper-plane to the data sets and check if the classification was correct or not. After all, phases lastly calculate the accuracy of every crop (Figure 3).

Here we found that the yield grades classification accuracy on both the training set and the testing set of each crop as shown in Table 5.

Table 4. Prediction accuracy of different crops using polynomial regression

Crops	Accuracy (%)	
	Training Set	Testing Set
Rice	87.06	91.58
<u>Ragi</u>	94.92	93.13
Gram	92.68	85.02
Potato	97.22	83.85
Onion	95.66	84.02

### Majority Voting (MV)

In the legislative course of action, the phrase “majority” clearly signifies “more than half.” As it allied to a referendum, a majority vote is nothing more than half of the referendums throw. Non-voting or expressionless are prohibited in cunning a majority vote.

The majority voting or MV method is one of the methods of incorporating the class labels acquired as an outcome of the unconventional classifiers. In this strategy, the classification of an unmarked occurrence is achieved. This methodology takes much more time as at consolidating scheme for stare recently the class that acquired the most amazing number of votes. This method is also called as the

Figure 3. Flowchart of SVM algorithm

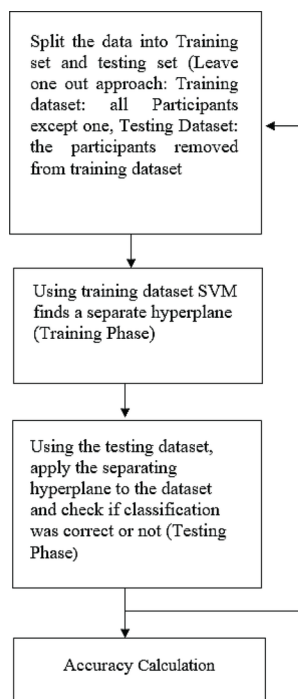


Table 5. Prediction accuracy of different crops using support vector regression

Crops	Accuracy (%)	
	Training Set	Testing Set
Rice	87.68	92.04
<u>Ragi</u>	94.01	93.60
Gram	90.45	78.48
Potato	96.83	84.02
Onion	91.92	68.64

vote of plurality discovered schemes. This is the most commonly applied combiner. Mathematically, it is written as:

$$\text{Class}(m) = \arg \max_{c_i \in \text{dom}(n)} \left( \sum_k g(n_k(m), c_i) \right) \quad (2)$$

where  $n_k(m)$  is the classification of the  $k^{\text{th}}$  classifier and  $g(y, c)$  is an indicator function explained as  $g(n, c) = 1$  if  $n=c$  and 0 if  $n \neq c$ .

### Model Training and Crops Recommendation

Training a prototype conveys determining adequate values for all the weights and the racism from tag samples. In supervised instruction, a machine learning procedure constructs a prototype by surveying numerous instances and attempting to discover a prototype that reduces loss; this procedure is known as empirical risk minimization.

After appealing the five different data sets of five different crops (rice, ragi, gram, potato, onion) to several supervised machines learning regression techniques such as linear regression, random forest regression, polynomial regression, decision tree regression, and support vector regression, we procure models and find the training and testing set accuracies of the five different crops that are finally trained for the crop recommendation for cultivation on the priority basis that is which crop to yield in that area first based on the production accuracy. The weights of this model can then be conserved, and the agronomists can comfortably make use of crop recommendations by providing their area's precipitation, cloud cover, vapor pressure, and production as the input to the system.

## EXPERIMENTAL RESULT AND PERFORMANCE ANALYSIS

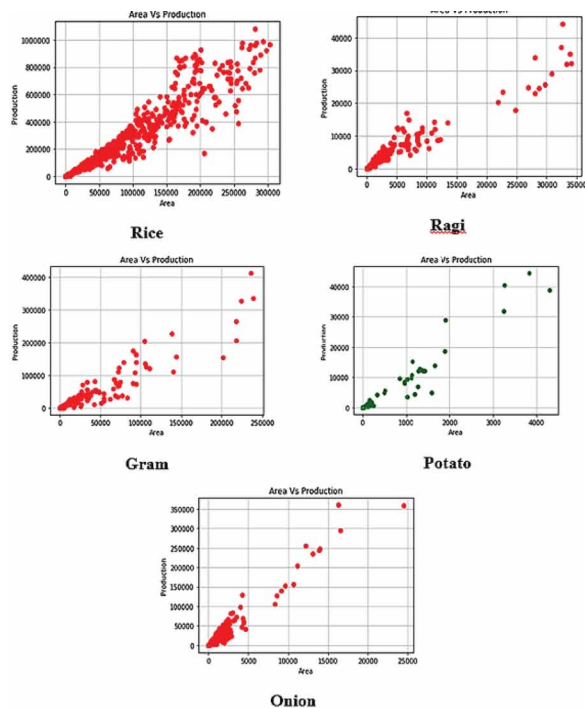
### System Setup and Configurations

Here we used the most useful library for machine learning in Python i.e. Scikit-learn for Calculating accuracy, Splitting the Training and Testing Set and for importing and building Machine Learning Models like Linear Regression, Decision Tree Regression, Random Forest Regression, Polynomial Regression and Support Vector Regression.

## Data Exploration of Different Crops

In the data exploration stage, we examine each variable of each crop and do the philosophical experiments about their meaning and importance for our vision. We focus on the dependent variable and independent variable and understand that. Then clean the data sets of each crop and handle the missing data, outliers, and categorical variables and finally check if our data sets meet the assumptions needed by multivariate techniques or not. Outliers are the values in the dataset which are exceptional from the idle part of the data. The outliers can be a consequence of delusion or inaccuracy in perusal, liability in the system, manual delusion, or misleading. In this work, we use the Z-score method for the removal of outliers in python. Z score is a salient quantifying or outcome that informs how many standard deviations exceeding or beneath a number are from the mean of the dataset. Positive Z score conveys the no. of standard deviation exceeding the mean and a negative outcome means no. of standard deviation beneath the mean. The Data Exploration of different crops is shown in Figure 4.

Figure 4. Data exploration



## Linear Regression of Different Crops

The predicted output of different crops using Linear Regression is shown in Figure 5.

## Decision Tree Regression of Different Crops

The predicted output of five different crops (rice, ragi, gram, potato, and onion) are using Decision Tree Regression is shown in Figure 6.

Figure 5. Linear Regression

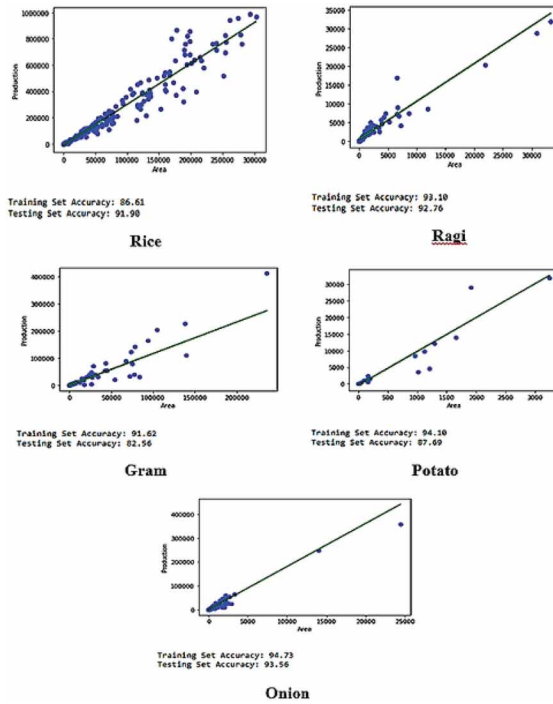
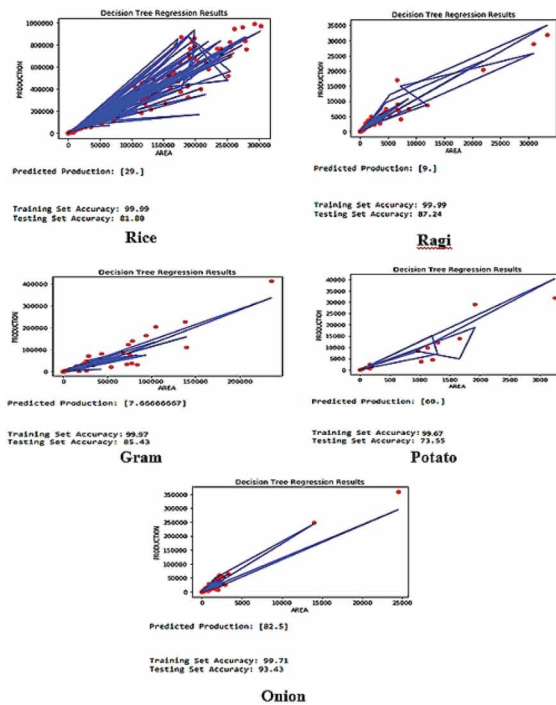


Figure 6. Decision Tree Regression



## Random Forest Regression of Different Crops

The predicted output of five different crops (rice, ragi, gram, potato, and onion) are using Random Forest Regression is shown in Figure 7.

## Polynomial Regression of Different Crops

The predicted output of five different crops (rice, ragi, gram, potato, and onion) are using Polynomial Regression is shown in Figure 8.

## Support Vector Regression of Different Crops

The predicted output of five different crops (rice, ragi, gram, potato, and onion) are using Support Vector Regression is shown in Figure 9.

## PERFORMANCE ANALYSIS

We have implemented different Machine Learning models like Linear Regression, Decision Tree Regression, Random Forest Regression, Polynomial Regression and Support Vector Regression for accuracy prediction of 5 Different Crop Yields like Rice, Ragi, Gram, Potato and Onion. We initially started with Data Exploration to know the Relation between the variables and to develop an appropriate model. The proposed model contains two phases: The training phase and the test phase. We have taken 70% of Data for training the Dataset and the remaining 30% for testing the Dataset. The Splitting of Dataset is done by Train\_Test\_Split function which splits the data arrays into two subsets that is training data and testing data. We do not need to divide the dataset manually. Train\_Test\_Split function will make the random partition for two subsets. Then it will create A Regression Object. In the training set, the data was collected. After that, regressor fit () is used to train the model

Figure 7. Random Forest Regression

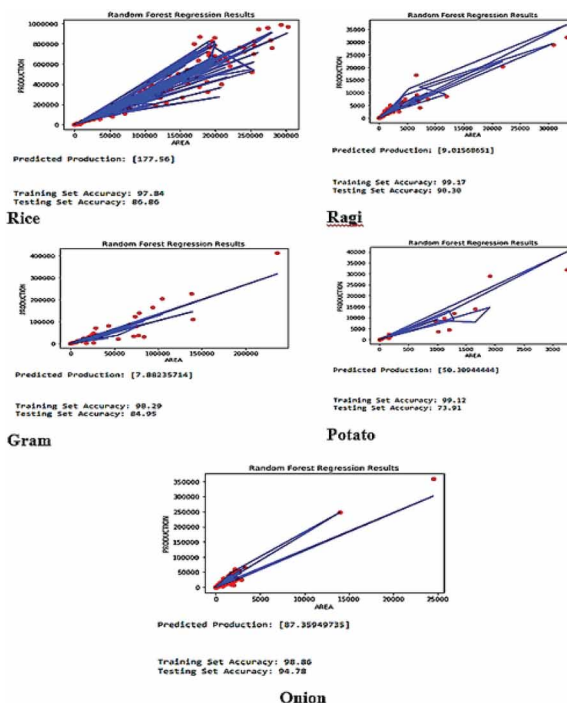


Figure 8. Polynomial Regression

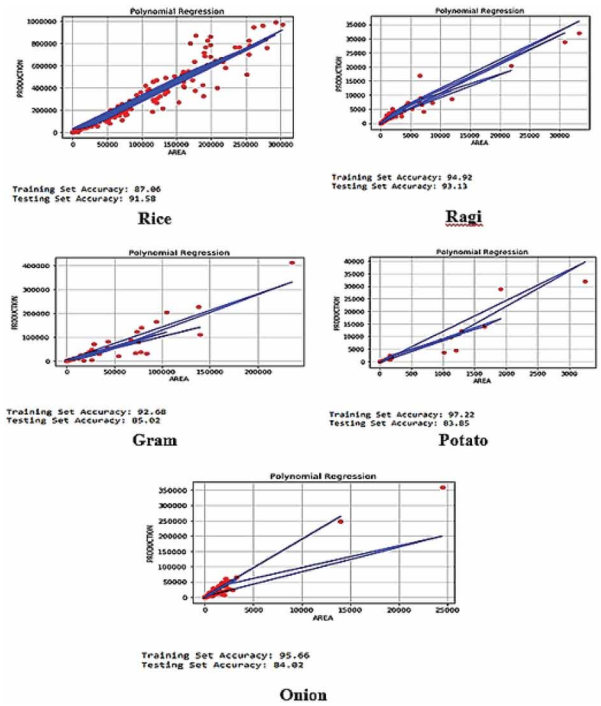
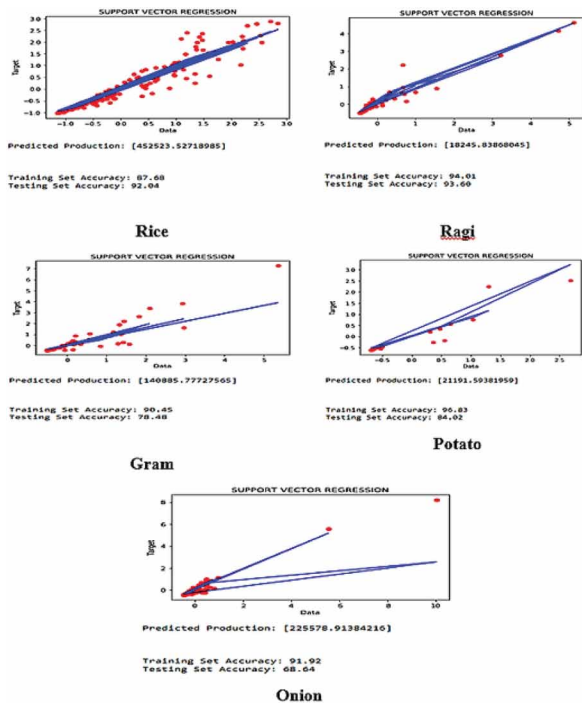


Figure 9. Support Vector Regression





using the training set. Then regressor. predict () is used to make predictions using the testing set. Finally, regressor. score () is used to understand how the model is making predictions and to calculate the prediction accuracy. After Knowing the Prediction Accuracy of Different Crops using Different ML Models, we have selected The Best Predicted Accuracy for Each Crop using different Models and Recommended that Crop get better production as compared to other crops.

From the average crop yield prediction accuracies we found the recommended crop rank wise as shown in Table 6 and the graphical representation is shown in Figure 10.

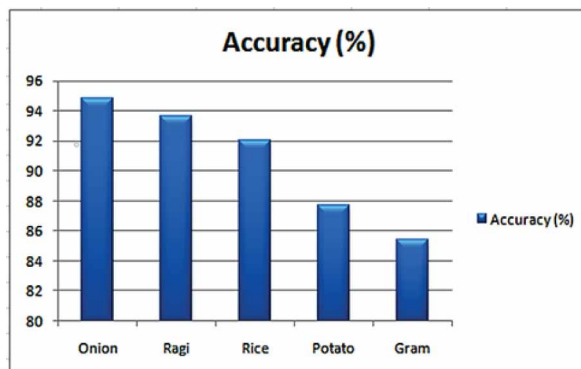
## CONCLUSION

The system assists the agronomists in picking an appropriate crop for their farming land based on the essential variables. The system is to plan and grow a recommendation model to create the recommendations for crops relied on geological and climatic attributes using machine learning procedures. The recommendation crop system has been planned that takes into consideration the dataset concerning the five several crops such as rice, ragi, gram, potato, and onion. The dataset of these five crops is initially preprocessed and then several regression methods such as linear regression prediction, decision tree regression, random forest regression, polynomial regression, and support vector regression are applied to forecast the accuracy. Finally, the majority voting (MV) method has been applied as the combination technique to provide the final accuracy. The final accuracy derived

Table 6. Recommended crops rank wise

Crop	Accuracy (%)
Onion	94.78
Ragi	93.60
Rice	92.04
Potato	87.69
Gram	85.43

Figure 10. Crops recommendation rank wise



using these above techniques are 94.78%. Therefore, the derived work imparts a serving hand to the agronomists in the proper selection of the crops for agriculture. This constructs an exponential earn in the crops yield which in succession boosts the wealth of India.

## REFERENCES

- Doshi, Z. (2018). Agro Consultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms. *Institute of Electrical and Electronics Engineers*, 3(2), 123–135.
- Everingham, Y., Sexton, J., Skocaj, D., & Bamber, G. I. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, 36(2), 27–35.
- Fathima, M., Sowmya, K., Barker, S., & Kulkarni, S. (2020). Analysis of crop yield prediction using data mining technique. *International Research Journal of Engineering and Technology*, 7(5), 7708–7713.
- Gandhi, N., Armstrong, L. J., Petkar, O., & Tripathy, A. K. (2016). Rice crop yield prediction in India using support vector machines. *International Joint Conference on Computer Science and Software Engineering*, 3(2), 1–5. doi:10.1109/JCSSE.2016.7748856
- Garanayak, M., Mohanty, S. N., Jagadev, A. K., & Sahoo, S. (2019). Recommender system using item based collaborative filtering (CF) and K-Means. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 23(2), 93–101. doi:10.3233/KES-190402
- Garanayak, M., Mohanty, S. N., Jagadev, A.K., & Sahoo, S. (2020). An Automated Recommender System for Educational Institute in India. *EAI Endorsed Transactions on Scalable Information Systems*, 20(26): e9. doi.org/, pp-1-13.10.4108/eai.13-7-2018.163155
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random Forests for Global and Regional Crop Yield Predictions. *Plus One*, 11(6), 46–59. doi:10.1371/journal.pone.0156571 PMID:27257967
- Khalil, Z. H., & Abdullaev, S. M. (2020). The sensitivity of winter crops to climate variability in the irrigated subtropics of Iraq (AI- Diwaniyah). *International Conference on Computational Intelligence and Data Science*, 16(7), 1066–1079.
- Kumar, A., Kumar, N., & Vats, V. (2018). Efficient crop yield prediction using machine learning algorithms. *International Journal of Research in Engineering and Technology*, 05(6), 3151–3159.
- MamathaJajur, S., Soumya, N. G., & Raju, G. T. (2019). Crop Recommendation using Machine Learning Techniques. *International Journal of Innovative Technology and Exploring Engineering*, 9(25), 658–661.
- Priyanka, T., Soni, P., & Malathy, C. (2018). Agricultural Crop Yield Prediction Using Artificial Intelligence and Satellite Imagery. *Eurasian Journal of Analytical Chemistry*, 13(7), 6–12.
- Renuka & Terdal, S. (2019). Evaluation of Machine learning algorithms for Crop Yield Prediction. *International Journal of Engineering and Advanced Technology*, 8(6).
- Sanchez, A. G., Solis, J. F., & Bustamante, W. O. (2014). Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research*, 12(2), 313–328. doi:10.5424/sjar/2014122-4439
- Sharma, A., Singh, P. K., Sharma, A., & Kumar, R. (2019). An efficient architecture for the accurate detection and monitoring of an event through the sky. *International Journal of Computer Communication*, 148(1), 115–128. doi:10.1016/j.comcom.2019.09.009
- Shastry, A., Sanjay, H. A., & Bhanushree, E. (2017). Prediction of crop yield using Regression Technique. *International Journal of Computing*, 12(2), 96–102.
- Singh, V., Sarwar, A., & Sharma, V. (2017). Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach. *International Journal of Advanced Research in Computer Science*, 8(5), 1254–1259.
- Ung, P. C., & Mittrapiyanuruk, P. (2018). Sugarcane Yield Grade Prediction using Random Forest and Gradient Boosting Tree Techniques. *International Joint Conference on Computer Science and Software Engineering*, 15(4), 85–91.
- Veenadhar, S., Misra, B., & Singh, C. D. (2014). Machine learning approach for forecasting crop yield based on climatic parameters. *International Conference on Computer Communication and Informatics*, 11(3), 13–24. doi:10.1109/ICCCI.2014.6921718

*Mamata Garanayak is a research scholar at KIIT Deemed to be University, Bhubaneswar, Odisha. She is presently working as an Assistant Professor in Computer Science and Engineering department of Centurion University of Technology and Management, Bhubaneswar, Odisha.*

*Goutam Sahu is a Student in the School of Computer Science & Engineering at the Centurion University of Technology and Management where he will be graduating in the year 2021. At his 3rd Year of Under-graduation, He selected Machine Learning and Artificial Intelligence as a Domain of his Personal Interest. Goutam Sahu completed many Projects and written some Research Journal Papers related to Machine Learning and Artificial Intelligence. His research interests lie in the area of Agriculture, Optometry, etc. ranging from theory to design to implementation. He has actively collaborated with researchers in several other disciplines of computer science to make his Work successful. The Age of Goutam Sahu is 20 and Currently, He is Working on Hyperspectral imaging in the field of Agriculture.*

*Sachi Nandan Mohanty (PhD), received Ph.D. from IIT Kharagpur in the year 2015, with MHRD scholarship from Govt of India. Prof. Mohanty research areas include Data mining, Big Data Analysis, Cognitive Science, Fuzzy Decision Making, Brain-Computer Interface, and Computational Intelligence. Prof. S N Mohanty has received 3 Best Paper Awards during his Ph.D at IIT Kharagpur from International Conference at Benjing, China, and the other at International Conference on Soft Computing Applications organized by IIT Rookee in the year 2013. He has published 20 SCI Journals. As a Fellow on Indian Society Technical Education (ISTE), The Institute of Engineering and Technology (IET), Computer Society of India (CSI), Member of Institute of Engineers and IEEE Computer Society, he is actively involved in the activities of the Professional Bodies/Societies. He has been bestowed with several awards which include "Best Researcher Award from Biju Pattnaik University of Technology in 2019", "Best Thesis Award(first Prize) from Computer Society of India in 2015", "Outstanding Faculty in Engineering Award" from Dept. of Higher Education, Govt. of Odisha in 2020. He has received International Travel fund from, SERB, Dept of Science and Technology, Govt. of India for chair the session international conferences USA in 2020. Dr. Mohanty currently reviewer of many journal like journal of Robotics and Autonomous Systems (Elsevier), Computational and Structural Biotechnology (Elsevier), Artificial Intelligence Review (Springer), Spatial Information Research (Springer). Ten Edited book, published by Wiley, CRC, and Springer Nature, and four authors' book on his Credit.*

*Alok Kumar Jagadev (PhD) is currently working as Professor in the School of Computer Engineering, KIIT Deemed to be University. He has obtained his Master degree from Utkal University in the year 2001 and also obtained Ph.D. degree for his work in the field of Wireless Adhoc Networks from Siksha 'O' Anusandhan University in the year 2011. He has contributed more than 40 papers in various journals and conferences of international repute. He has also contributed three Book Chapters in international published edited Volumes. He has authored/co-authored four textbooks in the field of computer science. He has also edited books for different international publications like IGI global, Springer etc. He has involved in organizing many international conferences and workshops. He has already supervised a dozen of Master students and guided three Ph.D. scholars. His research interest includes Soft Computing, Data Mining, Bio-informatics, etc.*