International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Crop Prediction System using Machine Learning Algorithms

Pavan Patil[1], Virendra Panpatil[2], Prof. Shrikant Kokate[3]

[1]Pavan Patil SPPU, (Pimpri Chinchwad College of Engineering)
[2]Virendra Panpatil SPPU, (Pimpri Chinchwad College of Engineering)
[3]Professor Shrikant Kokate, Dept. of Computer Engineering, PCCOE, Maharastra, India

---

Abstract - *As we are aware of the fact that, most of Indians have agriculture as their occupation. Farmers usually have the mindset of planting the same crop, using more fertilizers and following the public choice. By looking at the past few years, there have been significant developments in how machine learning can be used in various industries and research. So we have planned to create a system where machine learning can be used in agriculture for the betterment of farmers. The surveyed research papers have given a rough idea about using ML with only one attribute. We have the aim of adding more attributes to our system and ameliorate the results, which can improve the yields and we can recognize several patterns for predictions. This system will be useful to justify which crop can be grown in a particular region*

***Key Words*: Machine Learning in Agriculture, Classification Algorithms, Decision Tree, KNN**

1. INTRODUCTION

Crop production may be a complicated development that's influenced by soil and environmental condition input parameters. Agriculture input parameters vary from field to field and farmer to farmer. Collection such info on a bigger space may be a discouraging task. However, the environmental condition info collected in Republic of India at each 1sq.m space in numerous components of the district is tabulated by Indian meteoric Department. The massive such knowledge sets may be used for predicting their influence on major crops of that individual district or place. There are completely different foretelling methodologies developed and evaluated by the researchers everywhere the globe within the field of agriculture or associated sciences. A number of such studies are: Agricultural researchers in alternative countries have shown that tries of crop yield maximization through pro-pesticide state policies have LED to hazardously high chemical usage. These studies have reported a correlation between chemical usage and crop yield [1]. Agriculture is associate trade sector that's benefiting powerfully from the event of detector technology, knowledge science, and machine learning (ML) techniques within the latest years. These developments return to satisfy environmental and population pressures round-faced by our society, wherever reports indicate a requirement for robust international agriculture yield increase to produce food for a growing population on a hotter planet. Most of the work tired the sector of yield foretelling via cubic centimeter makes use of some kind of remote sensing knowledge over the farm. Agriculture seeks to extend and improve the crop yield and therefore the quality of the crops to sustain human life. However, within the current time, folks tend to require a lot of like a shot appreciated jobs. There are fewer, and fewer folks concerned in crop cultivation. additionally, the continual increase of human population makes the cultivation of the crops at the proper time and right place even a lot of vital, because the climate is dynamic and therefore the shifts from traditional weather pattern are a lot of frequent than before manufacture. Food insecurity may be a drawback that can't be avoided, and humans should build use of latest innovative technologies to create use of existing soil, water and air conditions to get larger crops. The information gap between ancient ways that of cultivating and new agricultural technologies may be overcome if the computer code may be designed to model the interactive impact of climate factors, particularly the impact of maximum events (e.g. heat, rainfalls and excess water) occurring at completely different growing phases of crops. The temperature change undoubtedly affects the native and world food production, therefore planning computer code to model crop predictions needs new methodology for temperature change studies, situations for temperature change adaptation, and policymakers which will limit the devastating effects of weather on food provide. Experimental proof is employed to form environmental condition zones that have seen changes in weather and water, the 2 most significant factors in guaranteeing a in crop. The soil sort will modification over time because of weather and pests, therefore crop management must manage a fancy quantity of information, directly or indirectly associated with one another. It will therefore by considering a simplified reality, to permit a quick assessment of the impact of temperature change in agriculture. Agriculture should adapt to those climate changes, and it will do therefore by developing models which will in theory optimize management practices, maximize the rotations of the new crop to manage the changes of soil, novel breeding programs. By maximizing the worth of foretelling, the seasonal climate changes may be ascertained and recorded in an exceedingly timely manner. Later on, by victimization computer code supported machine learning, one will timely assess the temperature change impact and check attainable situations that incorporate ascertained changes in climatic conditions and water distribution. data {processing} is that the process of analyzing the experimental knowledge collected over a amount and varied locations from completely different views, extract trends or patterns {of data of knowledge of info} and switch them into helpful information for users. Users will then additionally reason and/or summarize the relationships ascertained from the collected knowledge, and

typically predict what knowledge to expect. Machine-learning techniques are a part of data processing and knowledge exploration and focus exclusively on characteristic correlations or patterns among massive datasets or massive relative databases. The patterns, associations, or relationships among all this knowledge will additional be reborn into information that's offered to the user as historical patterns and future trends. This information provided by machine learning will facilitate farmers with crop cultivation by predicting probabilities of crop losses or stop losses altogether.

2. Literature Review

Nowadays many experts are applying automated farming. Since Decision Tree is an well-known algorithm it was used for prediction which is a supervised learning algorithm and multiple linear regression which is generalized prediction model. An attempt has been made to research the influence made by decision tree induction technique of climatic parameters on soybean productivity. For easy understanding of end-user different kind of rules were created from the Decision tree. The paper from Md. Tahmid Shakoor & co paper helped us for selecting various attributes like land capability classification, soil depth, slope, drainage, texture, erosion, and permeability [4]. Two supervised classification machine learning algorithms has been implemented in this study. Our system takes the necessary weather and soil properties data for a given coordinate automatically from an appropriate source. Another advantage is that their system worked on large regions, and provides forecasts at a resolution compatible with best input data resolution, which in the case is originally from the soil data. The ability of forecasting crop before the beginning of the crop season. This provides users with the capability to perform strategy changes, like choosing a more robust genetic variation before planting or even changing the crop type, in order to accommodate for extreme climatic variations further ahead in the crop cycle [2].

The algorithm developed introduces a data-driven model to predict and forecast crop yield using joint dependencies of soil and climate features. Although there are several techniques existing to obtain rainfall predictions, the algorithm discussed in this paper succeeded in emphasizing on Rainfall along with the crop yield prediction. This designed model took into account the most relevant environment as well as soil parameters that affect the crop growth, in a way that each of those parameters received equal weight in the final prediction. The outcomes of this research can benefit the agriculturists/farmers by knowing the investment capital on the crop to be sown, even before the sowing season begins. The predictive pattern of the algorithm can benefit local self-government and financial institutions to allocate suitable funds or fiscal loans to farmers. Naive Bayes is used for the large dataset can also be beneficial. Use of naïve Bayes and decision tree makes the model very efficient in terms of computation. The system is scalable as it can be used to test on different crops. From the yield graphs, the best time of sowing, plant growth and harvesting of the plant can be found out. Also, the optimal and worst environmental condition can also be incurred. The model focuses on all type of farms, and smaller farmers can also be benefitted. This model can be further enhanced to find the yield of every crop, and for pesticide recommendation. Also, it can be modified to suggest about the fertilizers and irrigation need for crops.

3. Related Work

1. A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast:

The system projected during this work is created by a neural network wherever inputs area unit treated on an individual basis. Static soil information in handled by fully-connected layers whereas dynamic meteorological information is handled by continual LSTM layers. This explicit design was trained with historical information for many soil properties, precipitation, minimum and most temperature against historical yield labels at county level. When training, the model was tested in an exceedingly separate information set and showed comparable results with existing yield prognostication ways that create use of in-depth remote sensing data. the most important lesson learnt from our experiments is that it's attainable get ascendable yield forecast as a result of the projected neural network model will notice and exploit redundant info each within the soil and within the weather information. To boot, the model might be able to learn AN implicit illustration of the cycles of the crops evaluated during this paper, considering the seasonal atmospherically information used as input.

2. Machine learning approach for forecasting crop yield based on climatic parameters

The present study provides the potential use of information mining techniques in predicting the crop yield supported the environmental condition input parameters. The developed webpage is user friendly and therefore the accuracy of predictions square measure higher than seventy-five per cent all told the crops and districts designated within the study indicating higher accuracy of prediction. The user-friendly web content developed for predicting crop yield may be utilized by any user their alternative of crop by providing environmental condition knowledge of that place.

3. Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model

The planned work introduces efficient degree economical crop recommendation system. Use of naïve mathematician makes the model terribly economical in terms of computation. The system is scalable because it may be wont to take a look at on totally different crops. From the yield graphs the simplest time of sowing, plant growth and gather of plant may be known. Conjointly the best and worst condition may also be incurred. The model focuses on all style of farms, and smaller farmers may also be benefitted.

This model may be more increased to seek out the yield of each crop, and for chemical recommendation. Conjointly it may be changed to recommend concerning the fertilizers and irrigation want of crops.

4. Evaluation of Predictive Data Mining Algorithms in Soil Data Classification for Optimized Crop Recommendation.

In this study, we've given the analysis potentialities for the classification of soil by mistreatment well-known classification algorithms as J48, BF Tree, and OneR and Naïve Bayes; in data processing. The experiment was conducted on information instances from Kasur district, Pakistan. We have ascertained the comparative analysis of those algorithms have the various level of accuracy to determine the effectiveness and potency of predictions. However, the advantages of the higher understanding of soils classes will improve the productivity in farming, reduce dependence on fertilizers and build higher prognostic rules for the advice of the rise in yield. In the future, we have a tendency to contrive to form a Soil Management and

5. Agricultural Production Output Prediction Using Supervised Machine Learning Techniques

Two supervised classification machine learning formula has been enforced during this study. the choice Tree Learning-ID3 (Iterative Dichotomiser 3) and KNNR discover the patterns within the knowledge set containing average temperature and precipitation worth obtained throughout the cropping amount of six major crops in 10 major cities of Bangladesh for the past twelve years and provides the prediction. ID3 uses the choice tree table that consists of the ranges of the precipitation, temperature and yield knowledge. The research provides an answer to the current downside that was much required for farmers in People's Republic of Bangladesh. Though the research is restricted to some mounted dataset, the long run ahead promises addition of a lot of knowledge which will be analyzed with more machine learning techniques to come up with crop predictions with higher exactness. Moreover, the analysis will result in profits and invention of advanced farming techniques which will improve our economy and can facilitate United States stand out as a technologically advanced country.

4. EXISTING SYSTEM

An agro-based country depends on agriculture for its economic growth. When a population of the country increases dependency on agriculture also increases and subsequent economic growth of the country is affected. In this situation, the crop yield rate plays a significant role in the economic growth of the country. So, there is a need to increase crop yield rate. Some biological approaches (e.g. seed quality of the crop, crop hybridization, strong pesticides) and some chemical approaches (e.g. use of fertilizer, urea, potash) are carried out to solve this issue. In addition to these approaches, a crop sequencing technique is required to improve the net yield rate of the crop over the season. One of existing system we identified is Crop Selection Method (CSM) to achieve a net yield rate of crops over the season. We have taken example of CSM to demonstrate how it helps farmers in achieving more yield

Crop can be classified as:

a) Seasonal crops— crops can be planted during a season. e.g. wheat, cotton.

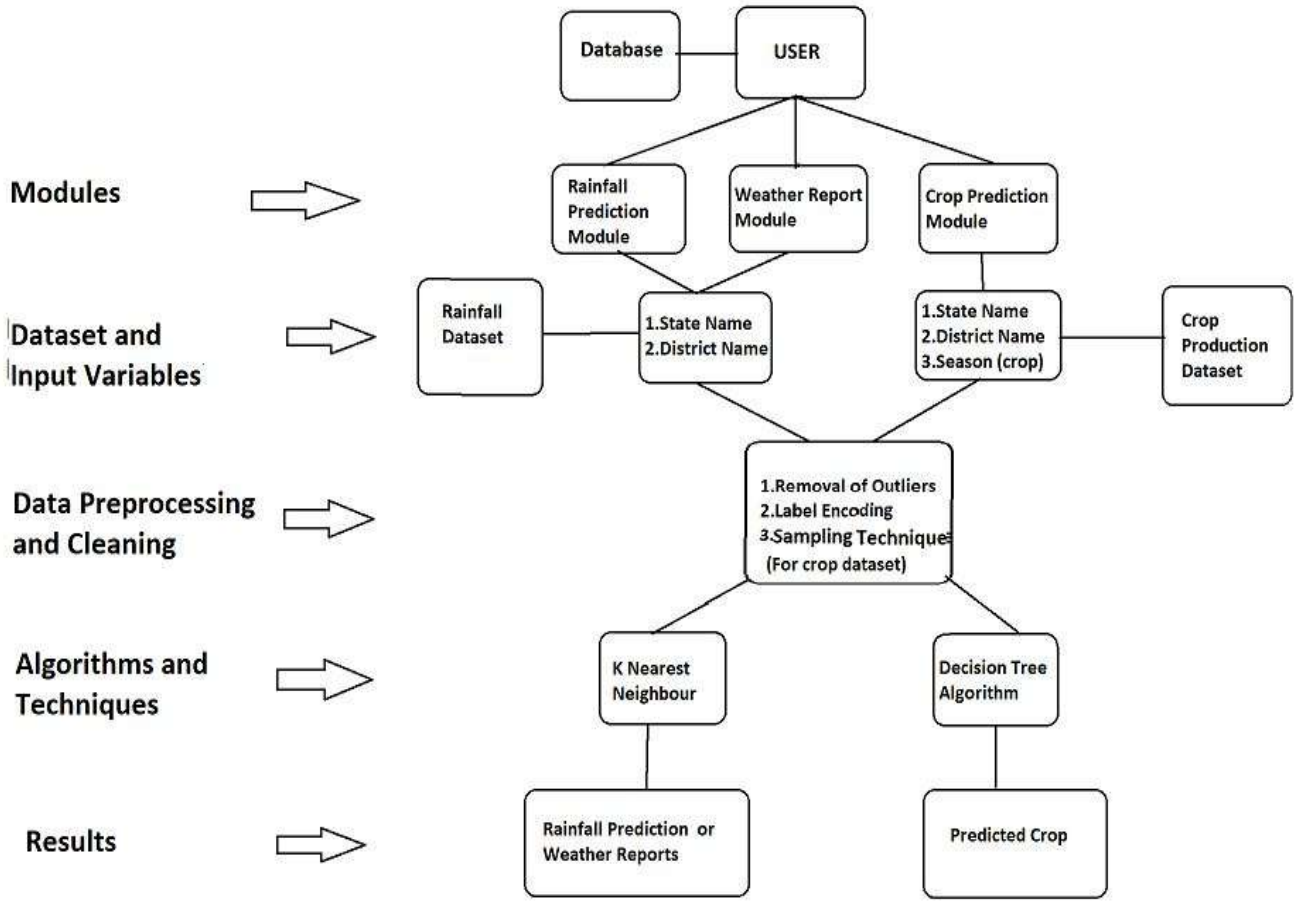
b) Whole year crops— crops can be planted during the entire year. e.g. vegetable, paddy, Toor.

c) Short time plantation crops— crops that take a short time for growing. e.g. potato, vegetables, ratio. d) Long-time plantation crops— These crops take a long time for growing. e.g. sugarcane, Onion. A combination of these crops can be selected in a sequence based on yield rate per day. Illustrates sequences of crops with cumulative yield rate over the season. CSM method, shown in may improve the net yield rate of crops using the limited land resource and also increases re-usability of the land.

Basically, in crop selection method makes use of technique where it recommends different set of crops for same area over the years. There are various options are available to select for farmers. They can choose one of the options and observe the results. The combination which will give high yield for same area is generated as output for that area. In this way CSM method tries to predict the suitable crops for given area. Farming Systems in India are strategically utilized, according to the locations where they are most suitable. The agricultural systems that significantly follows to the agriculture of India are subsistence farming, organic farming, industrial farming. Regions all over India differ in types of farming they use; some are based on horticulture, ley farming, agroforestry, and many more. The surveyed research papers have given a rough idea about using ML with only one attribute. We have the aim of adding more attributes to our system and ameliorate the results, which can improve the yields and we can recognize several patterns for predictions. This system will be useful to justify which crop can be grown in a particular region

5. PROPOSED SYSTEM

In our system we are making use of a classification algorithms to improvised the crop yields.



5.1. Data Acquisition:

Dataset must have following attributes

- Soil Parameters:
  Soil Type
  Soil Ph value
- Climatic Parameters:
  Humidity
  Temperature
  Wind
  Rainfall
- Production
  Cost of cultivation
  Previous year yield details for that region

In this project we are performing crops prediction for district level. So main aim is to find the dataset which contains production details of past 10-12 years also details about climatic parameters and soil parameters like rainfall, temperature, moisture, soil contents etc. details. These factors will help in the prediction of the crops by using various classifiers on the given dataset. Thus, various factors are assessed and the factors strongly leading to accurate prediction of the crops

5.2 Preprocessing:

The dataset that is used needs to be pre-processed because of the presence of redundant attributes, noisy data in it. Initially, data cleaning operation is performed where the redundant factors are determined and are not considered for the prediction of crops. Over18 which are either having the same values for all the employees or are completely unrelated to the prediction task. As part of the exploratory data analysis, the categorical factors are split and are assigned values as 0 and 1 based on whether the factor is present or not. These assigned values assist in further classification based on that particular factor.

5.3 Classifier Models:

5.3.1 Decision Tree Classifier:

The decision tree is method of selecting best root nodes until we get elements of same class we keep on splitting the tree on the basis of attributes. With versatile features helping

actualize both categorical and continuous dependent variables, it is a type of supervised learning algorithm mostly used for classification problems. What this algorithm does is, it splits the population into two or more homogeneous sets based on the most significant attributes making the groups as distinct as possible. The decision tree algorithm will give us best split on different features for selection of most suitable crop among the population. The feature selection methodology of Decision tree classifier makes it suitable for prediction of suitable crops. The Selection attributes of Decision tree classifier are as follow.

5.3.1.1 Gini Index

Gini index says, if we select two items from a population at random then they must be of same class and probability for this is 1 if population is pure. Used to calculate impurity for the features of given classes.

5.3.1.2 Entropy

A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogeneous). If the sample is completely homogeneous the entropy is zero and if the sample is equally divided then it has entropy of one.

5.3.1.3 Information Gain

The information gain is based on the decrease in entropy after a dataset is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain (i.e., the most homogeneous branches. This attribute selection methods will play vital role in prediction of crop.

5.3.1.4 C4.5 Algorithm

The C4.5 algorithmic program uses info gain as ripping criteria. It will handle numerical and categorical information similarly as missing values. To handle continuous values, it generates threshold and so divides attributes with prices quite the edge price and values up to the edge value. It offers the subsequent edges. They're explicable, in contrast to different classifiers, that need to be seen as a recorder that has a class to a given input instance. Call trees will be envisioned as tree graphs wherever nodes and branches represent the classification rules learnt, and leaves denote the ultimate categorizations.

5.3.3 KNN

KNN may be a variety of instance-based learning, wherever the performance is barely approximated regionally and every one computation is delayed till it's the classification. Both for classification and regression, a helpful technique will be to assign weights to the contributions of the neighbors, in order that the nearer neighbors contribute additional to the typical than the additional distant ones.

6. RESULTS AND ANALYSIS

We tested decision tree, naïve bayes classifier, and KNN classifier with sample dataset containing attributes like crop name, cost of cultivation, cost of irrigation, cost of production which are independent variables and yield per hectare is dependent variable.

The result obtained are represented using confusion matrix which shows relation between prediction of algorithms and actual values obtained when sample are tested after training dataset.

1. Decision Tree Classifier:

Confusion matrix for decision tree classifier:

[[221, 49],

[67, 163]]

Accuracy for Decision-Tree: 76.8%

Precision for Decision-Tree:0.767

Specificity for Decision-Tree:0.708

2. KNN Classifier:

Confusion matrix for KNN:

[[269, 30],

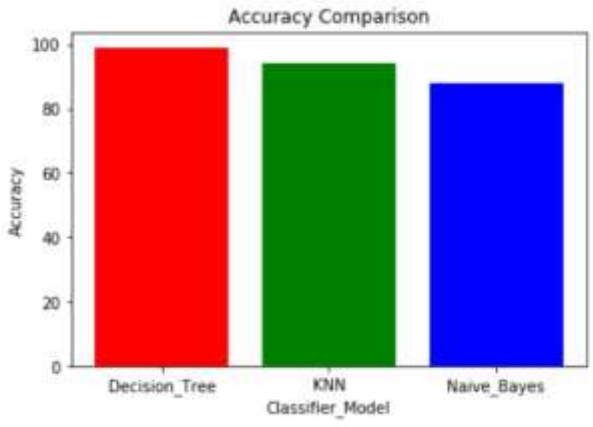
[23, 178]]

Accuracy for KNN: 89.4%

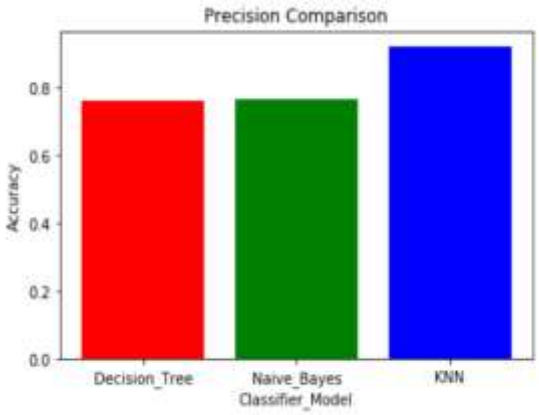
Precision for KNN:0.921

Specificity for KNN:0.8825

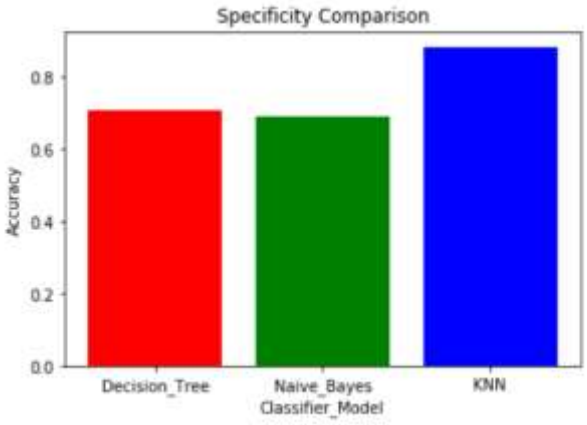
Accuracy Comparison:



Precision Comparison



Specificity Comparison:



7. CONCLUSION

The project work introduces an efficient crop recommendation system using classifier models. The system is scalable as it can be used to test on different crops. From the yield graphs the best time of sowing, plant growth and harvesting of plant can also be found out along with prediction for crops. Decision tree shows poor performance when dataset is having more variations but naïve bayes provides better result than decision tree for such datasets. The combination classification algorithm like naïve bayes and decision tree classifier are better performing than use of single classifier model.

REFERENCES

[1]. S.Veenadhari, Dr Bharat Misra, Dr CD Singh.2019."Machine learning approach for forecasting crop yield based on climatic parameters.".978-1-4799-2352-6/14/$31.00 ©2014 IEEE

[2]. Igor Oliveira, Renato L. F. Cunha, Bruno Silva, Marco A. S. Netto.2018."A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast.".978-1-5386-9156-4/18/$31.00 ©2018 IEEE DOI 10.1109/eScience.2018.00131

[3]. Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein."Prediction of Crop Cultivation".978-1-7281-0554-3/19/$31.00©2019 IEEE

[4]. Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty.2017."Agricultural Production Output Prediction Using Supervised Machine Learning Techniques".978-1-5386-3831-6/17/$31.00 ©2017 IEEE

[5]. G Srivatsa Sharma, Shah Nawaz Mandal, Shruti Kulkarni, Monica R Mundada, Meeradevi.2018."Predictive Analysis to Improve Crop Yield Using a Neural Network Model".978-1-5386-5314-2/18/$31.00 ©2018 IEEE

[6]. Rashmi Priya, Dharavath Ramesh.2018."Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model". 978-1-5386-5314-2/18/$31.00 ©2018 IEEE

[7]Talha Siddique,Dipro Barus,Zanntual Fredous,Amitabh Chakravarti. 2017. "Automated Farming Prediction".0978-1-5090-6182-2/17/$31 @2017 IEEE

[8]Takeshi Yoshida Noriyuki Murakami and Hiroyuki Tauiji.2017. Hybrid Machine Learning Approach to Automatic Plant PhenotypingFor Smart Agriculture". 978-1-5090-5888-4/16/$31.00 @IEEE 2016

BIOGRAPHIES

Prof. Shrikant Kokate
Qualification: ME Comp
Area of Interest: Data Science, Web Technologies

Name: Pavan Patil
Qualification:BE Computer(Pursuing)

Name: Virendra Panpatil
Qualification:BE Computer(Pursuing)