

## Crop Yield Prediction Using Supervised Machine Learning Algorithm

Hardik Joshi<sup>1</sup>, Monika Gawade<sup>2</sup>, Manasvi Ganu<sup>3</sup>, Prof. Priya Porwal<sup>4</sup>

<sup>1</sup>(Computer Engineering, Atharva College of Engineering/ University of Mumbai, India)

<sup>2</sup>(Computer Engineering, Atharva College of Engineering/ University of Mumbai, India)

<sup>3</sup>(Computer Engineering, Atharva College of Engineering/ University of Mumbai, India)

<sup>4</sup>(Computer Engineering, Atharva College of Engineering/ University of Mumbai, India)

**Abstract :** Agriculture is considered as the main and the foremost culture practiced in India. Nowadays, some farmers don't have awareness about the crop which will give them good production. They usually plan the cultivation process based on their experiences. But due to the lack of precise knowledge about cultivation, they may end up cultivating undesirable crops which will not give them an expected production rate. Hence, some of the farmers choose the suicide over it for paying off the loan which is not the right solution for it. There is an invention of new technologies like ML, IOT, AR but it's not getting used properly in the agricultural field. So our paper proposes a software application to predict crop yield from past data. This can be done by applying a machine learning algorithm on that data. The information about the farmer's state, district, season, land area, crop type is used for the estimating production rate of crops. Thus the project focuses on the creation of a prediction model which may be used to predict the maximum production rate of the crop before sowing.

**Keywords:** Agriculture production, Machine learning Prediction, Supervised learning, Unsupervised learning

### I. Introduction

According to recent survey, around 272.82 million farmers dwell in Maharashtra. It has been also found that suicide rates of farmers is also increasing from last few years. Therefore, to help the farmers take decisions that can make their farming more efficient and profitable. So in this we are proposing a system that will estimate crop production before harvesting depending on certain parameters. If they are having an idea about the amount of production they can expect, they can contact to their crop production contractor in advance of harvesting, often assuring a more competitive price. And moreover there is no such application exist as of now. Our paper proposes a system to predict crop yield from previous data using the concept of machine learning.

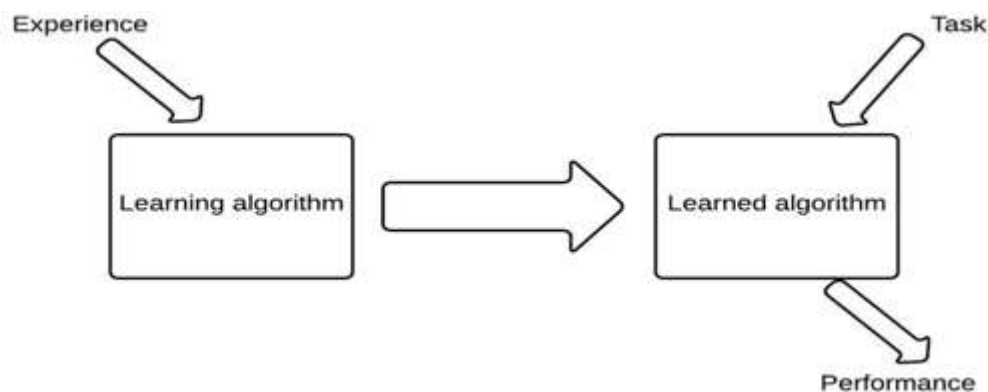


Fig. 1. Machine Learning Concept

Learning is the concept of improving the regular performance or action that one performs based on the experience and knowledge gain through regular activities. Using this same concept for building a system that can automatically learn and improve through experience. Machine learning is based on the similar concept of learning, Here a computer program is said to learn through the knowledge E (experience ie. past data) regarding some tasks T and performance P. Computer program is said to learn if its performance P improves at task T with experience E [14]. In simple words machine learning is the concept where machine learns by it own without needed to be programmed explicitly. It is the application i.e. subset of artificial intelligence that have enabled the system to automatically learn and improve. Programs are also generated in way by integrating input and output of the program. There is methodological difference between machine learning and statistics. One

difference is that machine learning focus on the optimization and performance and statistics concentrate on inference.

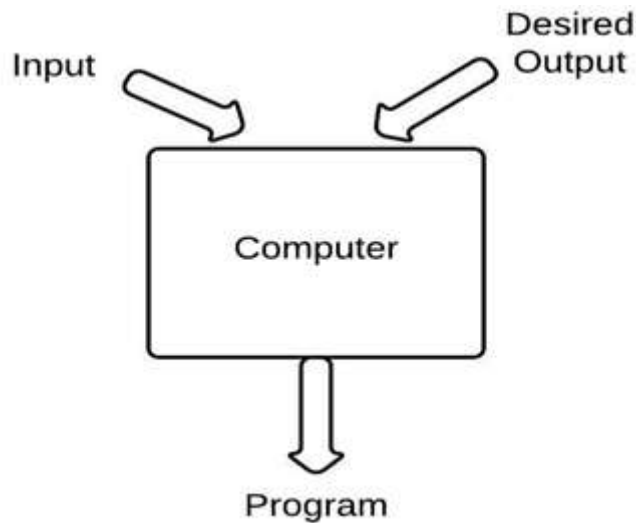


Fig. 2. Machine Learning Solution

The above figure show that for the machine learning solution the input and desired output or the target solution we want is to be given to the computer then the program or model is created to work on the test values or new data[15].

Since in our system, learning is done based on past data and then outcome is generated based on the test values, which is also termed as supervised learning algorithm in machine learning.

### 1.1. Supervised Learning

In supervised learning each input pattern will be having the corresponding output pattern that can be the targeted or desired output pattern. Comparative study is made between the actual output of the system and the desired output to find out the 'error'. Training the system or program in this continues till the system is able to give the expected or desired response[16].

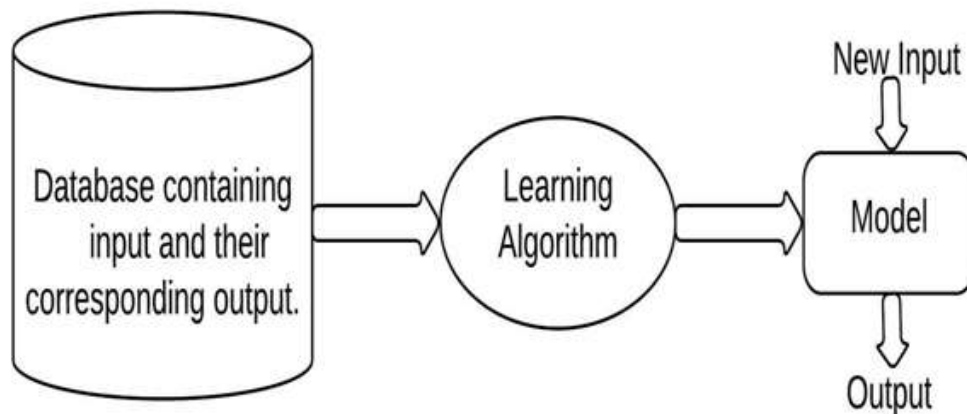


fig. 3. Supervised Learning

## II. Literature Review

Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene [1]

The system proposed in this paper helps farmers to cultivate proper crops for better yield of production. The system uses supervised and unsupervised machine algorithms for precise and accurate prediction by analyzing nutrients present in soil and crop productivity based on location. The result of two algorithms is compared and one giving the best and accurate output is selected. Hence, in this paper supervised and unsupervised algorithms are compared and supervised algorithm is found to have more score of accuracy than unsupervised algorithm.

Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty [2]

The study in this provides list of profitable crops in a particular area using decision making algorithms. This prediction is made using the K-nearest Neighbour algorithm and Decision Tree Learning, ID3 which analyzes production output of different crops in previous years. By analyzing data, algorithm gives result which predicts preferred profitable output.

Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh [3]

In this paper selection of crop is done to improve net yield rate of crops to be planted over the season. Prediction yield rate depends on parameters like weather, soil type, water density, crop type etc. Performance and accuracy depends on predicted value of influenced parameters, so there was need to adapt a prediction method for more accuracy and high performance.

S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh [4]

This paper demonstrated the potential use of data mining techniques in predicting the crop yield based on climatic input parameters. Based on only climatic condition we can't predict the crop production and quality.

K. Menaka, N.Yuvaraj [5]

This paper discusses various models like artificial neural network, Adaptive Neuro-Fuzzy inference system, Fuzzy logic and Multi-linear regression are analyzed to improve the prediction of crop yield production. The study has been proved that the Adaptive Neuro Fuzzy Inference System (ANFIS) prediction model provides better result than other approaches.

Vinciya, P., and Dr A. Valarmathi [6]

This paper presents a brief analysis of crop yield prediction using Multiple Linear Regression (MLR) technique for the selected region. It is mainly focused on analyzing the agriculture analysis of organic farming and inorganic farming, time cultivation of the plant, profit and loss of the data and analyze the real estate business land in a specific area. It concentrates organic, inorganic and real estate data sets from which the prediction in agriculture will be achieved.

Deodhar, Mayuresh [7]

For successful production of crops, we must ensure whether a particular crop will yield in particular soil and weather condition. And also, if crop is not yielding properly, that means it must have some disease. So, this study primarily focuses on two parts: One part takes input factors such as weather condition, soil properties, etc. and uses Bee Hive clustering approach to give output as crop name which will be suitable in particular condition. In second part application will take diseased crop image as an input and with the help of image processing technique and it will detect which might be the possible disease that have caused to the crop.

Buchade, Omkar [8]

This paper provides a solution for Smart Agriculture by monitoring the agricultural field which can assist the farmers in increasing productivity to a great extent. This work presents a system, in form of an android based application, which uses data analytics techniques in order to predict the most profitable and suitable crop in the current weather and soil conditions. The proposed system integrates the data obtained from repository, weather department and by applying machine learning algorithm: Multiple Linear Regression. Thus, the paper proposes a system that by integrating data from various sources, data analytics, prediction analysis which can improve crop yield productivity and increase the profit margins of farmer helping them over a longer run.

Dr.N.Suma, Sandra Rhea Samson, S.Saranya, G.Shanmugapriya, R.Subhashri [9]

This paper theorizes an explorative data analysis and discusses about designing of various predictive model. A sample data set is considered and different regression techniques are applied in order to identify and analyze the properties of each. Different regression techniques which are discussed in this paper are Linear, Multiple Linear, non-Linear, Logistic, Polynomial and Ridge regression. Using this paper, comparative study of various data analytics algorithm is done. This helps us to judge which algorithm best suits our proposed system

N. Himageetha [10]

This paper proposes a system for prediction of crops production in the current year. In order to know the crop production, it make use of a data mining algorithm K-Means. This system also uses prediction mechanism in form of fuzzy logic. Fuzzy logic is a rule-based prediction logic wherein a set of protocols are applied on the land for farming, rainfall and production of crops. Using this paper, a clear view of how K-Means

can be used to analyze data sets is obtained. Similar to set of rules as they have applied in form of fuzzy logic, we will be applying the set of rules to predict which crop will yield maximum profit based on previous years cost of crops and current soil and weather data.

Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar, Chandan Mehta [11]

This paper provides a way for smart agriculture by monitoring the agricultural field which can assist the farmers in increasing productivity to a great extent. It also applies machine learning and prediction algorithm like Multiple Linear Regression to identify the pattern among data and then process it as per input conditions.

E.Manjula, S.Djodiltachoumy [12]

The purpose of this research paper is to propose and implement a rule based system to predict the crop yield production from the collection of past data. This has been achieved by applying association rule mining on agriculture data from 2000 to 2012.

M.C.S.Geetha [13]

This paper discusses on different data mining applications in solving the different agricultural problems. It combines the work of different authors in one place so it is useful for researchers to get information of current scenario of data mining techniques and applications in context to agriculture field. It also provides a survey of various data mining techniques used in agriculture which includes Artificial Neural Networks, K-nearest neighbor, Decision tree, Bayesian network, Fuzzy set, Support Vector Machine and K-means.

### **2.1. Report on present investigation**

- About 60 % of the land in the India is used for agriculture in order to fulfil the needs of 1.2 billion people. And population is also increasing day by day. Thus, the modernization of agriculture is needed that will lead the farmers of our country towards profit and solve many problems of them[8].
- In Existing system farmers not connected with any technology and analysis.
- In conventional system farmer uses “trial and error” method. Farmer test on land with different crops, water availability, etc. and after many such “tries”, farmer probably gets the crop production as expected.
- Many papers have done the survey by taking into consideration different parameters.
- There are some systems which helps in selection of crop but no system is accurate[3].
- Some papers use data mining techniques in predicting the crop yield based on climatic input parameters. But only climatic conditions are not sufficient to predict crop yield[4].
- Some survey papers have analysed different machine learning algorithms that can be used in prediction of crop[5].
- There are many review papers on crop yield prediction that stated different algorithms that can be used for prediction. But no such system exists as of now. So, there is need to implement such system so that farmers will get benefit from it.

## **III. Proposed System**

In order to describe any system first problem is to be carried out in the current system and need to be described in detailed manner which is achieved by describing the problem statement of the system and then providing solution to that problem statement and future discussion or conclusions can be made based on the given solution.

### **3.1. Problem Statement**

Earlier Crop yield prediction was performed by the farmer by considering his experience on a particular field and crop. But that won't always be correct. Thus, our proposed system takes some parameters as input. For these parameters, the information about the farmer's state, district, season, land area, crop type is required. Based on the information given by the farmers, crop yield is predicted for the year of harvesting. The proposed system applies machine learning and prediction algorithm like Multiple Linear Regression to estimate crop yield as per input conditions. As past year production is also considered, the prediction will be more accurate.

The application should be able to take and convey the proper information. The system should help farmers to cultivate proper crop for better yield production. User friendly interface is to be provided to farmers

### 3.2. Solution

The system is intended to introduce the use of software technology in the field of agriculture. The aim of this system is developing an android application for farmers so that they can get the approximate idea about the crop production for the year in which crop is to be harvested. Objective is that they can get estimation of crop yield in advance. If they are having an idea about the amount of production they can expect, they can contract their crop production in advance of harvesting, often assuring a more competitive price than if they have to wait for harvesting the crop. The system will require the information about the farmer's state, district, season, land area, crop type is required. Based on only those information given by farmer, prediction can be made for production which will be a optimised and feasible prediction as several parameters are considered. The android application will take and convey these information with user friendly interface for farmers.

Advantage of this proposed system are System is extremely useful for farmers, the main advantage of using this application is that farmers can get the idea about the crop production based on which they can plan their farming activity and market income expectations, industry also can get advantage from yield predictions by better managing and scheduling the regimentation of their business, simple and easy to install and configure, most of the technologies utilized in this project are available freely, minimal changes on the system may result in new better system functionality.

The above proposed system as a solution to the problem statement can be depicted very well in the form of the context level diagram as given in Fig 4. which provide a better understanding of the system components and the working of the proposed system in simple and easy manner. From the figure we can see that first the input to the system is taken through the android application from the user of the application with the help of well defined and familer user interface. Then the application will process the input for validation and if validated then pass it future to API of ML logic for processing and applying the ML algorithm for give the results to the user. Then the result will be processed and report will be generated by android application to present the data in well defined manner to the user again through the help of user interface.

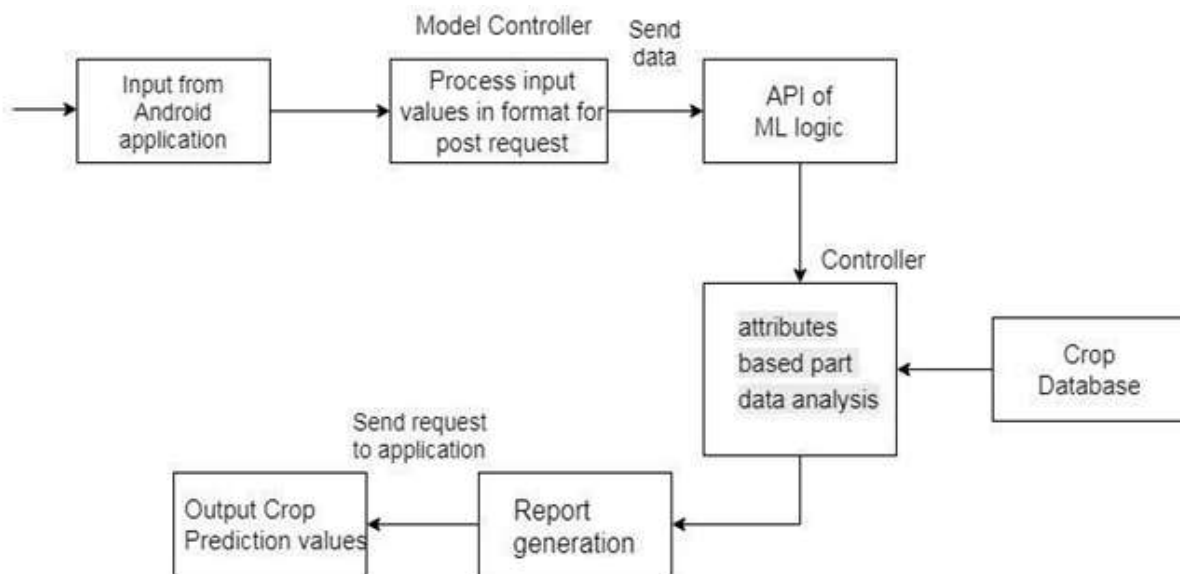


Fig 4. Propose system workflow

### 3.3. Need

Farmers usually plans their cultivation process based on their previous experiences. Due to the lack of precise knowledge about cultivation, farmers end up cultivating undesirable crops which results in loss of production rate. To help farmers take decisions that can make their farming more efficient and profitable, there is need to implement a system which predicts the production of crops before harvesting. We are proposing an idea for an android application for farmers so that they can get the approximate idea about the crop production for the year in which crop is to be harvested. They can get estimation of crop yield in advance.

Application of this system can be:

- Contract farming
- Agriculture
- Industries logistic planning

### 3.4. Market Potential and Competitive advantages

As there are no such free application to help farmers in agriculture. Farmers will get better advantage of their agriculture products by making more profits from their production in market. In addition, the prediction of the crop production certain new features can be added later in updates like weather prediction, crop diseases, etc.

### IV. Experimental Analysis

The prediction for the production of the crop depending on past data is made to help farmers and industries by providing them information about crop production. Using the supervised learning algorithm of machine learning data analysis is done in efficient manner. Data will be collected from the user of the application. Based on the given input data algorithm of machine learning will predict the crop yield. To enable the system to predict it is to be trained based on the past data about the crop yield in reference to the input parameters. Only after training the system application can make prediction based on the test values and parameters.

Initially the machine is to be trained on the dataset using sklearn model of machine learning in python. After training the model the predictions can be made by using the valid parameter values with the help of supervised algorithm of Machine Learning. The language we are using for implementing the logic of machine learning is Python.

The data are taken in 6 input variable. The variable are state name, district name, crop year, season, crop, area. Table 1 shows the details of input variables.

**Table 1: Dataset description**

Variable	Description
State Name	The data was collected from 29 states in India
District Name	Data is collected from all the 640 districts and 5924 sub districts in the states of india.
Crop_Year	The data was taken from 1997 onwards.
Season	Seasons like kharif, whole year, rabi, etc crop cultivation seasons.
Crop	Plants like arecanut, rice, banana, cashewnut, coconut, etc
Area	Area of agriculture plants regions in hectares.
Production	The production of crop in specific year in kilogram per hectares.

The proposed system is tested on collected data obtained for the years from 1997 to 2014. Data is collected from all the 640 districts and 5924 sub districts in the states of india. It contains nearly 200000 records. The overview of data set is already explained Table 1 shows the description of agriculture data set. Table 2 shows the sample dataset.

**Table 2 - Sample dataset**

State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254	2000
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2	1
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102	321
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176	641
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720	165
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168	65100000

The collected data set was preprocessed and some records are removed from the data. Now the dataset contains nearly 200000 records. After preprocessing step, the data set was fed to the polynomial regression algorithm.

#### 4.1. Algorithm used

- Polynomial Regression: In statistics, polynomial regression is a type of regression analysis in which the relationship between the independent variable (x) and the dependent variable(y) is modeled as an nth degree polynomial in x.
- Polynomial Regression in the form of mathematical equation:  $y$  (Dependent Variable) =  $a + b * x$ (Independent Variable) ^ n(degree)
- Here, the model becomes the non-linear amalgamation of attribute variables i.e., there can be exponential variables, sine, cosine, etc. Polynomial Regression is the kind of Linear Regression model. There are different cases of Polynomial Regression are given below.
- Linear Regression with multiple variable :  $h(x) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$
- In this kind of Linear regression, the forecast value( $h(x)$ ) depends on multiple variable i.e.,  $x_1, x_2, \dots, x_n$ .
- Polynomial (Linear) Regression :  $h(x) = b_0 + b_1x_1 + b_2x_1^2 + b_3x_1^3 + \dots + b_nx_1^n$
- In this kind of Linear regression, the forecast value( $h(x)$ ) depends on the power of the single variable i.e.,  $x_1, x_1^2, x_1^3$  and so on. Polynomial Regression is evaluated as a special instance of Linear Regression.

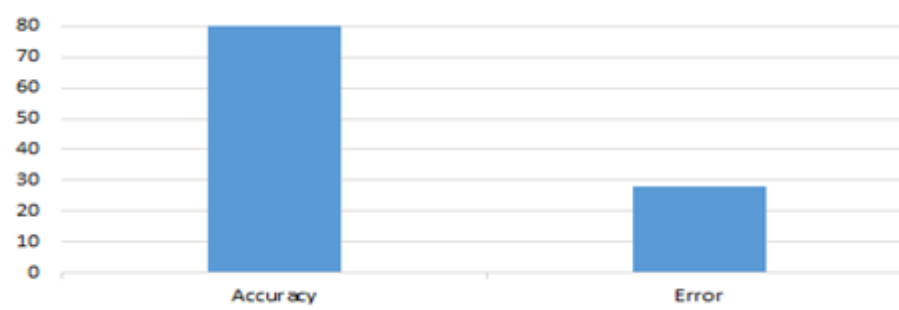


Fig 5. Overall accuracy and error rate

The crop production was predicted based on the input parameters as described in the table 1. The state, district, year, season and crop is input to the prediction model. Figure shows the overall accuracy and error rate of prediction result.

#### V. Conclusion

In current system farmers are not connected with any technology and analysis. So there are many chances of loss of money. Sometime wrong selection of crop will effect on their income. To reduce these we have proposed to develop an android application, which will predict crop yield production using past years data.

The prediction made which the application will make will be more precise as we are taking several parameters into consideration and the algorithm to be used for the prediction is supposed to be supervised learning algorithm so that their will be minor or no chance of error as the training is guided by the training model.

#### 5.1. Future Scope

- Crop diseases detection can be added with system
- Real time data can be collected using IoT automatically for more precise results.
- Weather and climate conditions can be shown for the particular crop for farmers help.

#### References

##### Papers from Conference Proceedings (Published):

- [1] Prediction of Crop Yield using Machine Learning by Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L. IRJET Volume: 05 Issue: 02 ,Feb-2018.
- [2] Agricultural Production Output Prediction Using Supervised Machine Learning Techniques by Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty, 978-1-5386-3831-6/17/\$31.00©2017 IEEE.
- [3] Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique by Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).

- [4] Machine learning approach for forecasting crop yield based on climatic parameters S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh. 2014 International Conference on Computer Communication and Informatics (ICCCI -2014), Jan. 03 –05, 2014, Coimbatore, INDIA.
- [5] A survey on crop yield prediction models by K. Menaka, N.Yuvaraj in Indian Journal of Innovations and Developments Vol 5(12), October, 2016.
- [6] Vinciya, P., and Dr A. Valarmathi. "Agriculture Analysis for Next Generation High Tech Farming in Data Mining." (IJARCSSE) (2016).
- [7] Deodhar, Mayuresh, et al. "Android Application For Crop Yield Prediction And Crop Disease Detection."
- [8] Buchade, Omkar. "Machine Learning based Crop Prediction System Using Multi-Linear Regression." International Journal Of Emerging Technology and Computer Science 3.2 (2018)
- [9] Dr.N.Suma, Sandra Rhea Samson, S.Saranya, G.Shanmugapriya, R.Subhashri, "IOT Based Smart Agriculture Monitoring System", Feb 2017 IJRITCC.
- [10] N. HemaGeetha. "A survey on application of data mining techniques to analyze the soil for agricultural purpose." IEEE, 2016.
- [11] "Crop Prediction System using Machine Learning" by Prof. D.S. Zingade, Omkar Buchade, Nileshe Mehta, Shubham Ghodekar, Chandan Mehta, Dec 2017 International Journal of Advance Engineering and Research Development
- [12] "A Model for Prediction of Crop Yield" by E.Manjula, S.Djodiltachoumy, International Journal of Computational Intelligence and informatics, Vol. 6: No. 4, March 2017 ISSN:2349-6363298
- [13] "A Survey on Data Mining Techniques in Agriculture" by M.C.S.Geetha Feb 2015 Journal of Innovative Research in Computer and Communication Engineering

**Books:**

- [14] Alex Smola, S.V.N. Vishwanathan. (2004). Introduction to Machine Learning. (1st ed.) [Online]. Available: <http://alex.smola.org/drafts/thebook.pdf>
- [15] Tom M. Mitchell. (March 1, 1997) . Machine Learning. (1st ed.) [Online]. Available: <https://www.cs.ubcluj.ro/~gabis/ml/ml-books/McGrawHill%20-%20Machine%20Learning%20-Tom%20Mitchell.pdf>
- [16] Shai Shalev-Shwartz, Shai Ben-David. (2014). Understanding Machine Learning: From Theory to Algorithms. (1st ed.) [Online]. Available: <https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>