

Prediction of Crops based on Environmental Factors using IoT & Machine Learning Algorithms

Ashok Tatapudi, P Suresh Varma

Abstract: India being an agricultural country, the most part of economy is depends on yield growth. Agriculture is largely dependent on rainwater and also relies on different soil variables such as aspects of nitrogen, phosphorus, potassium and climate such as temperature, precipitation, etc. The technological growth in agriculture will increase the crop productivity. Remote sensing systems like IOT systems are being more widely used in smart farming systems, these systems produce generous amount of data. Machine learning is an ongoing work that has been filed to forecast the plant based on data trends. The proposed system would integrate sensors such as Ph., Moisture, Rainfall, Temperature and Humidity to analyze the information from these sensors and to implement machine learning algorithms: Linear Regression, Decision Trees, Random Forest, and GDBoost. The most desirable crops are predicted according to the current environment. This work gives farmers a better prediction to plant what kind of crops in their farm field based on the criteria listed above in order to improve Smart Farming's productivity.

Keywords: Agriculture, IoT, Machine Learning, Data Analytics, Prediction

I. INTRODUCTION

In this paper the proposed system is a form of plant forecasting program to increase production based on key technologies: the internet of things and techniques of machine learning. Sensor technology has been developed and sensor forms such as humidity, temperature, soil moisture sensor, and pH sensors are used to detect the appropriate plant prediction elements. Machine Learning technology predicts the crop based on the sensor data. Uses of these technologies helpful to the farmer for better production rate in agriculture.

1.1 AGRICULTURE

Improving the development and value of crops while increasing operating costs and degradation of the atmosphere is a key objective of agriculture. Potential growth and yield depends on many different attributes of development such as environment, surface properties, and control of irrigation and fertilizer. Agriculture is every economy's backbone. In a nation such as India, which is undergoing ever-increasing food demand due to rising growth, developments in the agricultural sector are needed to meet the needs. Since ancient times, farming has been known as India's primary and

paramount society. Ancient people in their own territory grow the plants and have been tailored for their needs. Therefore, many people such as human beings, animals and birds have grown the natural crops and used them. The area of farming is increasingly deteriorating owing to the development of new innovative technology and techniques. Because of this, plentiful citizens with creativity were focused on developing synthetic goods that are composite items which contribute to an unhealthy life. Nowadays, there is no knowledge among modern people regarding increasing the crops in the right time and place. Such cultivation strategies often adjust the seasonal climatic conditions against the fundamental resources such as land, water and air which contribute to food insecurity. After considering all of these problems and concerns such as climate, temperature, and various variables, there is no proper solution or engineering to solve the problem they face. In India, there are several ways to boost agricultural economic growth. There are several directions in which plant production and crop value can be enhanced and strengthened.

1.2 IOT IN AGRICULTURE

The Internet of Things is an interconnection between electronic systems, physical or virtual computers, artifacts with unique identifiers and the ability to transfer information over a network without human-to-human or user-to-computer contact. In the field of agriculture, few researchers have suggested IoT-based systems with machine learning to forecast plant type. This device which manages the plant growth form. The proposed system helps to make accurate judgments and analyze data obtained from the sensors and processed in the server and evaluated using machine learning algorithms. The data sensed by the sensor from crop yield for different parameters of humidity, temperature, precipitation, pH quality. Etc. is processed in storage by IoT systems that are then used to forecast plant varieties that have a direct impact on crop growth after a predictive decision is made to forward to the end user for further intervention that will help the end user.

1.3 MACHINE LEARNING

Machine learning is a tool that is commonly used to tackle agricultural problems. It is used in the analysis of large data sets and in the data sets to establish useful classifications and patters. The overall goal of the Machine Learning system is to derive the knowledge from a set of data and turn it for further use into an understandable framework. This paper's main objective is to design a system that can predict the type of crop based on the properties of soil and weather.

Revised Manuscript Received on November 22, 2019.

* Correspondence Author

Ashok Tatapudi*, Assistant Professor, Department of Computer Science & Engineering, University College of Engineering, AdiKavi Nannaya University, Rajamahendravaram, India. Email: ashok.tatapudi@gmail.com

P Suresh Varma, Professor, Department of Computer Science & Engineering, University College of Engineering, AdiKavi Nannaya University, Rajamahendravaram, India. Email: sureshvarmap@gmail.com

Prediction of Crops based on Environmental Factors using IoT & Machine Learning Algorithms

As the population is growing in today's world and it is supposed to be in billions as the years go by and we need to improve crop production to feed those billions of people. The population is increasing and on the other hand the agricultural land is declining due to various reasons such as industrialization, retail markets and residential buildings are being built on these agricultural lands and in order to feed these billions we need to increase production and this can be achieved by introducing appropriate technology in agriculture. Smart agriculture is the most important thing required in everyday life.

Section 2 outlines related work, while Section 3 analyzes the proposed system and structure, Section 4 presents the findings, and Section 5 finishes the document.

II. RELATED WORK

The numerical approach, namely the methodology of multiple linear regression and the process of data mining, namely the clustering strategy dependent on size, were used to predict plant yields [1]. Kalman filter (KF) is used with predictive analysis in the proposed technique to acquire quality data without any noise and to transmit this data for cluster-based WSNs. Decision tree using predictive analytics for crop yield prediction, seed identification, soil classification, weather prediction, and crop disease prediction for decision making. This platform combines IoT modules such as and cube (IOT Gateway) and Mobius (IOT Service Platform) to provide consumers with a smart crop growth tracking solution [2]. In machine learning algorithm is developed using logistic regression to process raw data and forecast outcomes. It provides the result but is less reliable than other algorithms [3]. The use of spatial data mining in the agricultural field has been clarified by authors [4]. They used the K-means algorithm along with incremental improvement of the optimization approach for the study of spatial connections. As initial spatial information, temperature and precipitation are provided and evaluated to increase crop yield and growing crop losses. In [5] authors consider the problem of predicting the average yield of a type of crop (e.g., soybean) for a region of interest based on a sequence of remotely sensed images taken before the harvest and convolutional neural networks applied data to predict the type of crop. This [6] addresses scientific advancements over the past 15 years on machine-based learning techniques for precise crop yield forecasting and approximation of nitrogen status, and concludes that rapid advances in sensing technology and ML techniques will provide cost-effective and detailed solutions for improved crop and environmental condition estimation and decision-making. Remote Monitoring System (RMS) is introduced in this system [7], a hybrid solution to internet and wireless communications. The main goal is to capture real-time agricultural production environment information that provides easy access to agricultural facilities such as warnings through Short Messaging Service (SMS) and weather patterns, plants, etc.

III. PROPOSED WORK

The proposed research work focuses on the use of effective IOT devices and decision learning for prediction. In system design, we have included flow of communication between

different system components and input and output for different modules present in system. Sensed information is contrasted to data set that is stored on past experience and is generated as a consequence. The system architecture is shown in Fig. 1. Main components of system are

- IOT devices
- Machine learning algorithm for prediction.

As per the result obtained from the analysis, the farmer must settle on the process selection of best crop for that particular soil for increase the production rate of the crop.

3.1 IOT DEVICES

In IoT-based smart agriculture, sensors (light, humidity, temperature, soil moisture, etc.) are used to track the plant field and optimize the irrigation system. The farmers were able to monitor the conditions of the field from anywhere.

This work aims to provide a device for tracking specific plant temperature & humidity values, soil and water pH sensor and fertilizer control, soil moisture sensor for detecting soil moisture levels. All these devices are used as well as producing the information to track the plant. IoT system interfaced with the sensor:

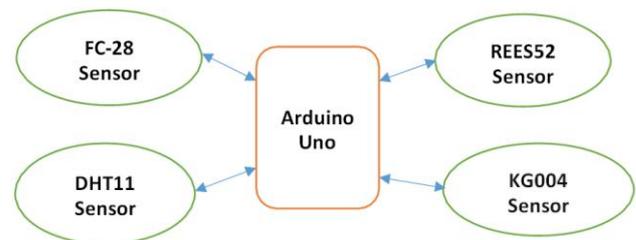


Fig. 1. IoT Architecture

3.1.1 FC-28 SOIL MOISTURE SENSOR

The sensor for soil moisture is quite clear to use. The two wide uncovered pads serve as detector sensors, functioning as a variable resistor together. The more moisture in the ground, the greater will be the conductivity between the surfaces, resulting in lower resistance and higher output of SIG. [8]

Table- I: FC-28 Soil Moisture Sensor

Input Voltage	Output Voltage	Input Current	Output Signal
3.3-5V	0-4.2V	35mA	Analog & Digital

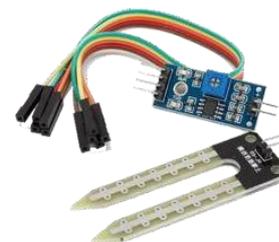


Fig. 2.FC-28 Soil Moisture Sensor

3.1.2 DHT 11 HUMIDITY & TEMPERATURE SENSOR

DHT11 Temperature & Humidity Sensor uses an electronic signal performance optimized temperature & humidity sensor system. This guarantees high reliability and outstanding long-term durability through the use of the proprietary digital signal processing methodology and sensing technologies for temperature & humidity. It device incorporates a resistive moisture measurement module and an NTC temperature measurement element and connects to an 8-bit high-performance microcontroller providing excellent quality, fast response, anti-interference capabilities and cost-effectiveness. [9]

Table- II: DTH -11 Sensor

Item Measurement	Range	Humidity Accuracy	Temperature Accuracy	Resolution	Package
DHT11	20-90 % RH 0-50 °C	±5% RH	±2°C	1	4 Pin Single Row

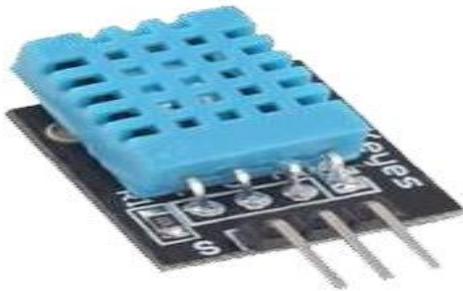


Fig. 3.DTH-11 Sensor

3.1.3 KG004 RAIN DROP SENSOR

The unit of the rain detector is an easy tool to measure water. When raindrop falls through the rainy board and also to measure rainfall rate, it can be used as a switch. [10]

Table- III: KG004 Rain Drop Sensor

Item Measurement	Range	Output Voltage	Output Signal
KG004	-0.3 to +36 V	+36 V	Digital and Analog



Fig. 4. KG004 Rain Drop Sensor

3.1.4 REES52 PH SENSOR

PH is an indicator of a solution's acidity and alkalinity, the pH level ranges between 0 and 14. The pH shows the concentration in certain solutions of hydrogen [H]⁺ ions. A detector that detects the potential difference between two electrodes can reliably calculate it: a reference electrode (silver / silver chloride) and a hydrogen-sensitive glass electrode. That's what the test was. We also need to use an electronic circuit to properly condition the signal, and with a micro-controller like Arduino we can use this sensor. [11]

Table- IV: REES52 pH Sensor

Supply Voltage	Current	Consumption	Working Temperature
5V	5-10mA	≤ 0.5 W	10-50 °C



Fig. 5. REES52 pH Sensor

IV. MACHINE LEARNING ALGORITHM FOR PREDICTION

Machine learning is widely applied to aspects in agriculture. It is used in the study of large data sets and in the data sets to define valuable classifications and patterns. The overall goal of the Machine Learning process is to extract the information from a set of data and transform it for further use into an understandable structure.

Based on available data, this paper analyzes the crop yield form. The methodology of machine learning was used to forecast crop yields in order to maximize crop profitability. Figure 6 demonstrates the stream of estimation of the expected crop yield.

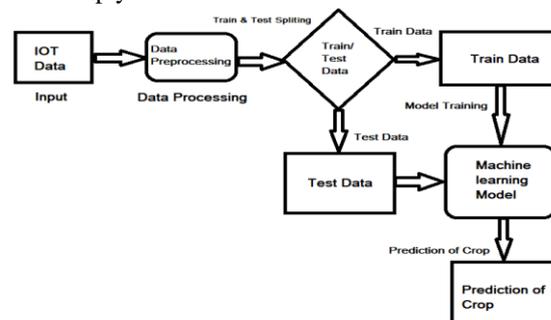


Fig. 6. System Architecture

Prediction of Crops based on Environmental Factors using IoT & Machine Learning Algorithms

As shown in the previous example, sensors are installed on the farm to detect data related to humidity, temperature, precipitation, and pH. Logistic regression, Decision Tree, Random Forest, and GDBost algorithms are used to characterize sensed data. The forecast result shows which soil may be ideal for different crops and the state of soils.

4.1. OVER VIEW OF DATA

Data was gathered from different outlets and create datasets in this process. And for analysis, such databases are used. Internet sources such as Data.gov.in then indiastat.org are several.to produce the data. The below table having the sample data values, we almost cover all the crops which are helpful to the farmer. Duration shows that crop duration in months, Min & Max attributes are showing the range temperature which is required for the crop. N, P, K Values are fertilizers for specific crop. pH min & max values for soil management. Rain fall is the range of that area.

Table V. Overview of Sample Data Points

Crop	Duration	Min Temp	Max Temp	PH min	PH max	RF min	RF max	N	P	K
Bajra	3	18	30	3	8	350	750	L	L	M
Banana	4	15	35	6.5	8.5	450	750	M	VL	VL
Barley	4	12	32	3	8	800	1100	VL	VL	M
Bean	2	14	32	5.5	6.5	300	500	L	VL	M
Black pepper	6	23	33	5.5	6.5	1200	2500	H	VL	M
Black gram	2	23	35	5	7	500	700	L	H	VL
Bottle Gourd	2	24	27	6.5	7.5	400	650	VL	VL	VL

Table VI. State Wise Sample Data Points

State	N	P	K	ppm(parts per million)		
A& N	VL	VL	L			
A P	L	VH	M	Nitrogen	Phosphorus	Potassium
Karnataka	H	M	M	VL: 10	VL: > 5	VL: >100
Assam	M	L	VL	L:10-20	L:5 -10	L: 100-150
Bihar	VL	VH	H	M:20-30	M:10-20	M:150-250
Goa	M	VL	M	H:30-40	H: 20-30	H: 250-300
Gujarat	VL	VH	H	VH:40+	VH: 30+	VH: 300+

Table VI. Sample Temperature Data Points

Station	Name	Month	Max Temp	Min Temp	Mean RF
Bangalore	Karnataka	January	27.4	15	4.9
Bangalore	Karnataka	Feb	30.1	16.6	7.9
Bangalore	Karnataka	March	32.8	19	10
Bangalore	Karnataka	April	33.9	21.3	43.9
Bangalore	Karnataka	May	33.1	21.1	111.9

Bangalore	Karnataka	June	29.4	19.8	79.7
Bangalore	Karnataka	July	27.7	19.3	109.7

4.2. MACHINE LEARNING MODELS

Machine learning is the method of finding correlations in large datasets that were previously unknown and potentially increasing. The extracted data is used as a statistical and classification template to reflect. Datasets obtained from IoT & Multiple Sources tend to be considerably more complex than the database used in machine learning historically. Machine learning is specifically defined as machine learning that is analytical or predictive. Yet predictive data mining is mostly used in the agricultural area. Classification and clustering are two primary methods. Some of the methods below are used to extract the response from the data collected.

The technique suggested includes two phases: the stage of preparation and the phase of evaluation. The data was collected and pre-processed during the training phase. The learning cycle utilizes pre-processed information to train the template. The yield value is calculated in the testing phase based on the rules developed. Work begins with phase of pre-processing. The data collected were pre-processed in this phase. Several information has been omitted from the data set in the pre-processing. Some of the field was inadequate for crop production. So the data will be deleted. Models used in the phase of training and testing described below:

A. Logistic Regression

The method used to relate a dependent variable to one or more independent variable is logistic regression. The dependent variable is sometimes called predictors, and predictors are called independent variable. Regarding plant type prediction(c) as variable based and temperature and humidity disparity, soil moisture, pH rate as variable independent. The formula that has been established is in the

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 \quad (1)$$

B. Decision Trees

One of the classification algorithms that can be used in machine learning is the decision tree. The model for inductive learning is the educational decision tree. According to some parameters, a model is created from information or observations. The design is aimed at discovering from the experienced instances a general rule. Therefore, decision trees may execute two different tasks based on whether the goal feature is distinct or constant. In the case of the wood, a classification tree will lead in the creation of a regression tree as in the second case. 12] [13]

C. Random Forest

Random forests are a mixture of tree predictors so that each tree relies on the values of a self-sampled random variable with the same distribution for all forest trees. Forest generalization error converges a.s. The number of trees in the woods is increasing to a peak.

A tree classifier forest's generalization error depends on the strength of the specific forest trees and their comparison. Using a random selection of features to separate each node produces error rates that are more stable in terms of noise relative to Adaboost. Internal measurements calculate variance, frequency and consistency, and these are used to show the response to increase the number of characteristics used in the splitting. External measurements are also used for parameter importance estimation. Specific ideas apply to regression as well. [14]

D. Gradient Boosting

The main causes of variation between real and expected values are noise, uncertainty, and discrimination as we try to predict the goal factor using any machine learning methodology. Ensemble allows certain variables to be growing. Gradient Boosting slowly, additively or sequentially teaches multiple prototypes. By using gradients in the loss function ($y = ax + b + e$, e needs a special note as it is the error term), gradient boosting does the same. The loss concept is a measure of how the coefficients of the good model fit the underlying data. A rational interpretation of the function of loss will rely on what they seek to maximize. [15]

4.3. IOT ANALYTICS

IoT Analytics is a fully managed product that allows advanced analytics easy to run and operationalize on massive volumes of IoT information without having to worry about the cost or difficulty usually required to build an IoT analytics system. It is the best way to run IoT data analytics and gain insights to create better and more accurate IoT apps and machine learning use cases decisions.

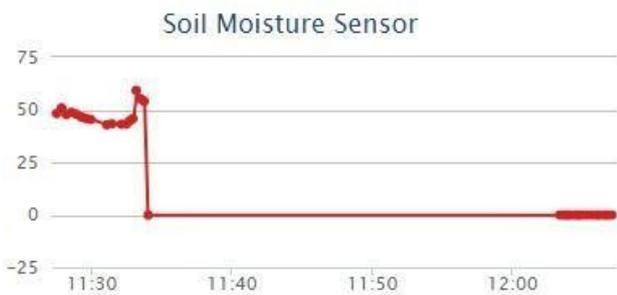


Fig.7. Soil Moisture Data



Fig.8. Temperature Sensor Data

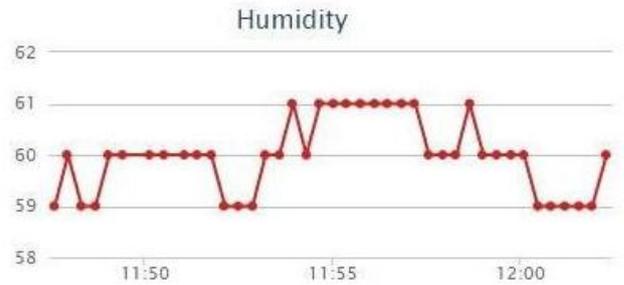


Fig.9. Humidity Sensor Data

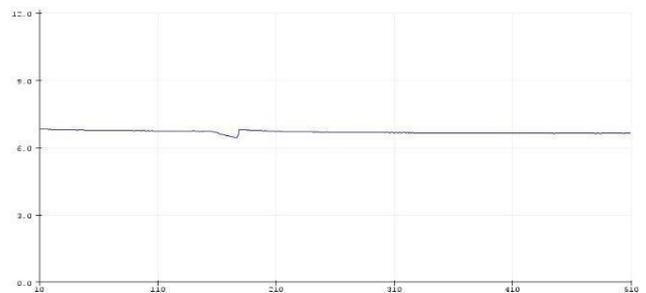


Fig 10. pH Sensor Data

V. RESULT AND DISCUSSION

We can easily predict the type of crop based on the features we consider after extracting the features from the represented information. In order to approximately evaluate the efficiency and effectiveness of our experiment, we compare ensemble, classification algorithms on the data for predictions. Classification Algorithms are Logistic Regression and Decision Tree algorithms, these generates 85.2%, 94.14% accuracies respectively. Ensemble algorithms are Random Forest and GDBoost. Random Forest is ensemble method to control the bias and variance in data, it is improves its accuracy when the depth of the tree is high and low when its depth is low. To avoid the over fitting of the model we put a threshold values to our model and that's value 14. With that depth its generated 96.32% accuracy with 3.68% error rate Fig showed below. Gradient Boost is also an ensemble method to boost up the model performance, this method generated highest accuracy with the boosting technique and that is 96.69% and the error rate is 3.61. Accuracies table showed below.

Table VII. Accuracies of Machine Learning Models

Classifier	Error Rate	Accuracy
Logistic Regression	14.8	85.2
Decision Tree	5.86	94.14
Random Forest	3.68	96.32
Gradient Boost	3.61	96.69

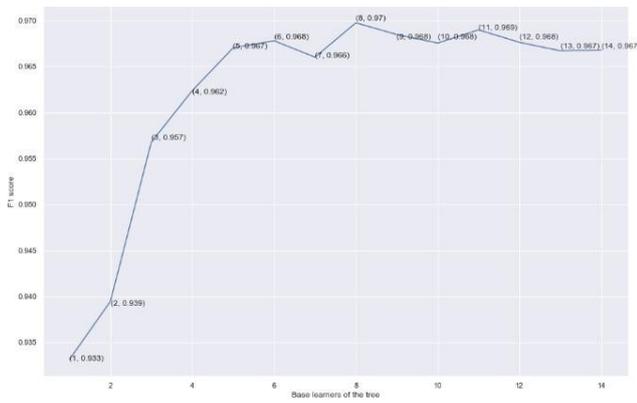


Fig 11. Random Forest Algorithm Depth Generation

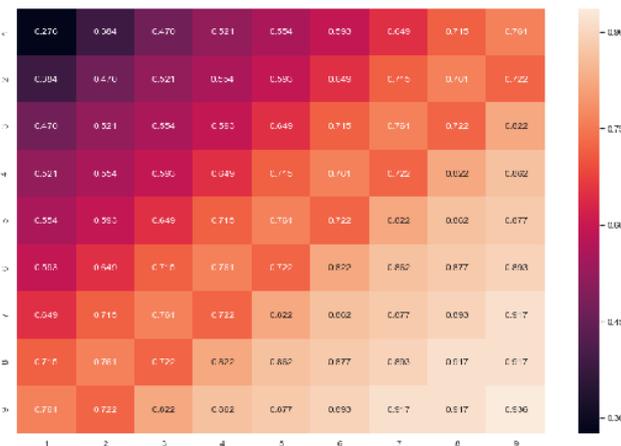


Fig.12. Scoring Values of Models on Predicted Value

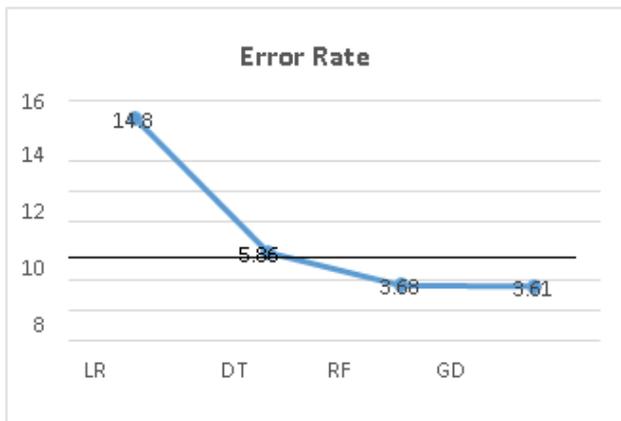


Fig.13. Error Rate of Models

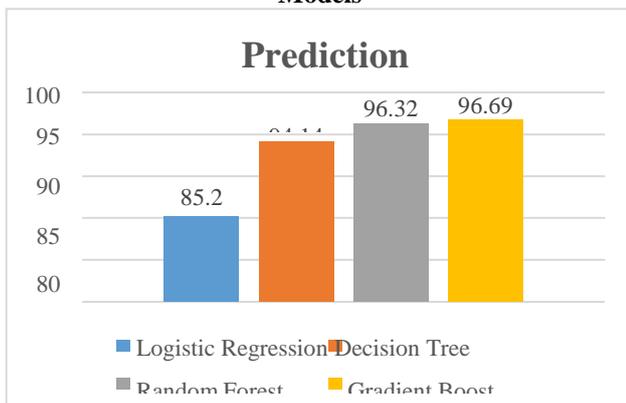


Fig.14. Accuracies of Machine Learning Models

VI. CONCLUSION

The proposed system lists all potential crops that are feasible in a given area, allowing the farmer to determine which plant to produce. The program has carried out a careful examination of the climate, weather and pH information and recommends which are the most viable crops that can be grown in the correct environmental situation. This program often analyses the past data output that will enable the farmer gain insight into the market demand and value of different crops. Because total plant varieties under this program will be protected, farmers can know about the crop that may never have been produced. IOT leads to connection of all farming devices together with help of internet in. Different types of sensors employed in farm is give real time data of farm condition and the devices can be used to increase the moisture, acidity, etc. accordingly.. Further the best profitable crop can also be found in light of the monetary and inflation ratio.

ACKNOWLEDGMENT

I wish to express my sincere appreciation to those who have contributed to this thesis and supported me in one way or the other during this amazing journey. I am extremely grateful to my thesis guide, Professor Suresh Varma, for his guidance and all the useful discussions and brainstorming sessions, especially during the difficult conceptual development stage. His deep insights helped me at various stages of my research.

REFERENCES

1. M.K. Gayathri, Dr.G.S. Anandha Mala, J Jayasakthi, "Providing Smart Agricultural Solutions to Farmers for better yielding using IoT", TIAR 2015.
2. Patil, S., and S. R. "INTERNET OF THINGS BASED SMART AGRICULTURE SYSTEM USING PREDICTIVE ANALYTICS". Asian Journal of Pharmaceutical and Clinical Research, Vol. 10, no. 13, Apr. 2017, pp. 148-52, doi:10.22159/ajpcr.2017.v10s1.19601.
3. Thomas Truong; Anh Dinh; Khan Wahid. An IoT environmental data collection system for fungal detection in crop fields [M]//2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)
4. D, Rajesh. (2011). Application of Spatial Data mining for Agriculture. International Journal of Computer Applications. 15. 10.5120/1922-2566.
5. YOU, J.; LI, X.; LOW, M.; LOBELL, D.; ERMON, S.. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. AAAI Conference on Artificial Intelligence, North America, feb. 2017.
6. Anna Chlingaryan, Salah Sukkarieh, Brett Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review, Computers and Electronics in Agriculture, Volume 151, 2018, Pages 61-69, ISSN 0168-1699.
7. K. A. Patil and N. R. Kale, "A model for smart agriculture using IoT," 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), Jalgaon, 2016, pp. 543-545.
8. <https://www.electroniccomp.com/soil-moisture-sensor-module-india>.
9. <https://www.mouser.com/ds/2/758/DHT11-Technical-Data-Sheet-Translated-Version-1143054.pdf>.
10. https://www.rhydolabz.com/sensors-weather-sensors-c-137_147/raind-rop-sensor-module-p-2336.html
11. <https://scidle.com/how-to-use-a-ph-sensor-with-arduino/>
12. Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Estimation of neural network parameters for wheat yield prediction. In Max Bramer, editor, Artificial Intelligence in Theory and Practice II, volume 276 of IFIP International Federation for Information Processing, pages 109–118. Springer, July 2008.

13. Iv'an Mej'ia-Guevara and 'Angel Kuri-Morales. Evolutionary feature and parameter selection in support vector regression. In *Lecture Notes in Computer Science*, volume 4827, pages 399–408. Springer, Berlin, Heidelberg, 2007.
14. Breiman, L. (2001) Random Forests. *Machine Learning*, 45, 5-32.
15. <https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab>