Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4193 - 4202
Received 25 April 2021; Accepted 08 May 2021.

A Farmer Support System using IoT and Machine Learning

Lydia D. Isaac[1*], Janani.I[2]
[1]Assistant Professor, Department of Information Technology, Sona College of Technology, Salem
[2]Assistant Professor, Department of Information Technology, Sona College of Technology, Salem
*tulip.lydia@gmail.com

ABSTRACT

The Cultivation process is usually planned by the farmers based on their previous experiences. Sometimes they end up in cultivating undesirable crops due to lack of optimal weather and as their predictions are not precise. Application of Artificial Intelligence in this field can bring a boom in the income of farmers by changing the scenario through cultivating the most optimal crops. This research focuses on introducing a farmer support system for predicting coffee yield in the Eastern Ghats region of Tamilnadu by applying different Machine Learning (ML) techniques and Internet of Things (IoT). The IoT module is used for sensing the temperature, humidity and amount of rainfall whereas the ML module is used for predicting the coffee yield with the input from IoT module and Soil test samples.

Keywords: IoT, Machine Learning, Farmer Support System, Coffee yield Prediction

Introduction

Internet of Things is an integrated system consisting of computing, mechanical and digital devices that have high capabilities of transferring data over a network without Human –computer interactions. IoT has a wide variety of things that can collect information and send, that can receive information and act upon them and even that can do both. Taking a simple example from farming, a soil moisture sensor when connected to the irrigation system can detect the moisture level of the soil and if necessary can turn on the irrigation system to water the crops. Even if this system is connected to a weather information system, can take a decision whether to water the crops now or whether there would be a rainfall now. Thus integrating different types of sensors related to crop cultivation like those for temperature, humidity, rainfall and so on can collect huge amount of data that is necessary for analyzing various aspects of cultivation.

Machine learning paves way for finding out patterns in data and help in applying them automatically to take good decisions and also make predictions about the future. Classification is one similar method in machine learning that involves predicting a target variable in data that has not been observed previously. The end point of classification is predicting a class (target variable) by building a classifier using training data and predicting the class using test data.
This process involves, finding out the relationship between the input attributes and the target attributes that is between independent variables and dependent variables. it belongs to supervised learning. There are many applications that involve classification like internet traffic interception, medical diagnosis, email filtering and so on. In this paper we tried applying it to predict the yield of coffee in the region of Yercaud, located in the Eastern Ghats region of Tamilnadu.

Climate and the soil composition play a vital role in coffee cultivation. Hot, humid climate with temperature around 15°C to 28°C and annual rainfall of about 150 to 250 cm is required for obtaining good yields of coffee. It does not endure snowfall, frost, temperature above 30°C and is generally grown in shady areas. Stagnant water and prolonged drought is injurious to coffee and hence it is grown at elevations around 600 to 1600 meters above sea level. Loamy soil with good

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4193 - 4202
Received 25 April 2021; Accepted 08 May 2021.

deal of minerals and humus are ideal for cultivating coffee. To increase productivity and to replenish fertility, the soil must be properly fertilized with proper manure. Thus the proposed farmer support system integrating the advantages of the two recent domains, IoT and Machine Learning would help these small growers in predicting their yield by sensing soil moisture, temperature and humidity through sensors and using the collected data for making useful decisions.

Literature Review

Agriculture is the recent area of focus in which various domains are integrating to provide useful solutions to enhance production and in turn the profit. The authors [1]
have developed a smart farming system that is affordable using IoT and Machine Learning which on deployment would provide clear cut details on the real time crop conditions to the farmers. The project involves sensors that are connected to the cloud, where real time data is being collected for analysis and suggests that Machine Learning algorithms could be used for predicting the future conditions of the crops. They have employed a pre-trained SVM Classifier which takes the crop name and sensor values as input to predict, is the condition optimal for growth of the crop or not. Hot, thirsty and optimal are the classes used with regard to soil moisture. Almost all categories of crops such as cereals and pulses and sugarcane have been considered for the research by training a logistic regression model by using Tensor flow.

Researches on the applications of various data mining methods or techniques for making critical decisions in the agricultural domain has evolved in the recent years [2]. Support Vector Machines, Bayes theorem and Artificial Neural Networks contribute to it. A naïve ratio algorithm has been proposed by authors [3] for providing faster predictions about the yield to the farmers and has used cross validation to analyze the performance of the model on an unseen dataset. Similar studies have been made [4] for predicting the yield of rice crop in the state of Maharashtra, India with SVM classifier and suggests SVM for crop predictions.

Deep learning which uses multiple layers to extract high level features has led to significant breakthroughs in most of the predictive applications [10]. The authors have described a semi-parametric approach that combines two different domains for prediction. They have admitted that their approach would be suitable for almost any prediction problem that involves imperfect prior knowledge mapping input functions to outcomes.

The authors of [11] have proposed a system that takes into account soil data, weather data and past years production data to suggest the best crop to be cultivated so that it ends up profitable. It is designed in such a way that it would list down all the possible crops a farmer could cultivate based on the current environment condition. They also have suggested connecting all farming devices to the internet through IoT.

A plant-driven crop management system monitoring soil, climate and crop providing decision support like treatments to be delivered like irrigation to be done, fertilizers to be applied and so on has been proposed by the authors [13]. They have applied ML techniques to find out the decision rules. The ML model that they have suggested for discovering knowledge is based on iterative process and based on this, patterns and associations that arise internally are analysed

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4193 - 4202
Received 25 April 2021; Accepted 08 May 2021.

again to expand knowledge. They have focused only towards the prevention of impacts due to water stress on the plants.

From all these related studies, we tried to combine different methods that have been used so far in the area of agriculture. Thus we have tried to use the power of IoT and ML in predicting the yield of coffee in Tamilnadu.

Methodology

Coffee yield is dependent on both climatic conditions as well as soil composition. Geographic parameters such as elevation, aspect and factors of environment like rainfall, temperature and humidity influence the economic production of coffee more than the soil factors. Hence the dataset used for analysis consisted of parameters related to environment, geographic parameters and soil specific parameters.

The proposed approach involves two main modules:

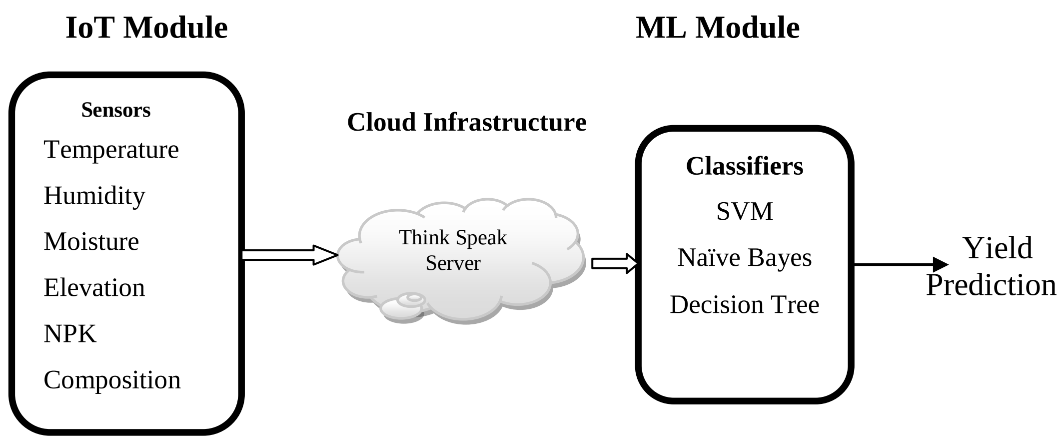

Fig.1 Proposed Model

A. IOT Module

In order to obtain the environmental parameters like temperature, amount of rainfall, humidity and geographical parameters like elevation and aspect specific sensors like optical sensors, temperature, soil moisture and humidity sensors are employed. The data that is sensed with the help of sensors is being stored in an array. At attaining a particular time span (some x seconds), every time the sensed data from the sensors is sent for analysis to the server in cloud. The value x is chosen based on the bandwidth and speed of the network at the farm. As we know that the network speed will be varying from time to time, the value of x will not be constant. It would vary according to the speed of the network and this will be checked hourly or when required.

B. *Machine Learning Module*

Machine learning (ML) has come into sight with big data technologies and other

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4193 - 4202
Received 25 April 2021; Accepted 08 May 2021.

technologies like high-performance computing thereby promoting larger opportunities to unravel and understand various processes that involve big data in the operational environments related to agriculture. Quite a number of relevant researches have been done in the area of crop yield prediction. The efficiency of Support Vector Machine (SVM) models and Bayesian Network models and Decision tree models seems to be notable in predicting the yield of crops.

1. Support Vector Machines (SVM)

It is a supervised learning method which mainly contributes in solving classification and regression problems. It normally classifies data samples in disjoint clusters. SVM has been employed in solving many real-world problems such as text categorization, Classification of images, classification of satellite and biological data and many more. It finds a hyper-plane that clearly classifies data points in an N-dimensional space. N is the number of features. Results shown in paper [6] reflect that SVM is more accurate with respect to crop yield predictions.

2. Naïve Bayes

Naive Bayes classification is probabilistic. The basis of this classifier is the Bayes theorem that involves a hypothesis of autonomy. Its dependent on the precise nature of the model. This could also be described as an independent feature model; as it assumes that the presence or absence of one feature is independent of the other. This makes it more efficient for supervised learning. The most interesting part is that a minimal training data is sufficient to estimate mean and variance that are necessary for classification [8].

3. Decision Tree

An example of a predictive model is a decision tree which lists out the various decisions at the bottom most level as the tree is constructed. At every level a condition is checked which entirely depends on the application. The algorithm begins with an original set as the root node and during each iteration it hits upon each unused attribute among the set of attributes and computes entropy and gain [9]. Then the set is divided into selected attributes. At each subset the algorithm is reproduced considering only the attributes which were not selected previously. Applying decision trees could depict logical patterns related to the yield, even with complex data involving high degree of details.

Experimental Setup

This section describes the experimental setup necessary for successful predictions. It throws light on the algorithmic steps, various hardware and software components involved.

A. Hardware Components:

a. ADS1115:

This is a prodigious analog to digital converters that can be easily used with Raspberry Pi and Arduino. This can be achieved by interfacing with its I2C communication bus. It is a 16-bit

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4193 - 4202
Received 25 April 2021; Accepted 08 May 2021.

ADC with 4 channels of higher precision. Each channel is used to collect data related to temperature, humidity and soil moisture. Signals can also be amplified to obtain data with higher precision.

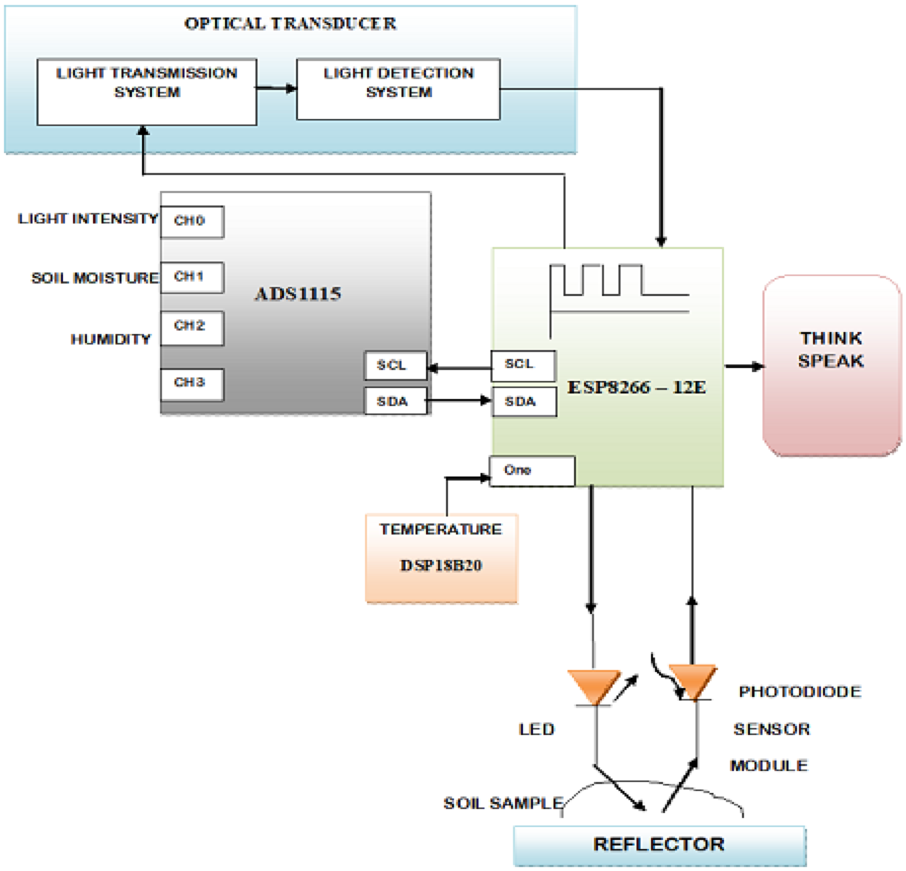

Fig. 2 Field-side architecture

b. ESP8266-12E:

It is a cost effective WiFi module that supports serial communication. It is also compatible with a lot of development platforms especially Arduino. AT-commands or Lua Script can be used to program it. It is used here to transfer data from Arduino to the Think Speak IoT analytics platform.

c. Sensors:

i. DHT-11:

This sensor is for reading the temperature and humidity values at the cultivation area. The data will be received as digital waves.

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4193 - 4202
Received 25 April 2021; Accepted 08 May 2021.

ii. FC-28:

This sensor is for detecting the level of moisture present in the soil. It consists of two electrodes that can be placed into the soil. The conductivity between these two electrodes would vary according to the level of moisture contained in the soil. The output obtained will be in the form of an analog signal. Therefore it is necessary to use a converter to make it a digital signal to interface it with the Arduino.

iii. LM-393:

This sensor monitors the amount of sunlight received by the plant and this will help us in determining whether the crop receives optimal amount of sunlight required for its growth.

B. Software:

Python: To apply machine learning to data for prediction analysis Python 3 can be installed on the server.

C. Cloud Storage:

An Amazon AWS T2.Micro EC2 instance should be framed-up and the deployment of the web application is to be done on this server. This would act as a repository or in other words a database to store the received data. This stored data can then be used for training and testing the machine learning models to make the ultimate predictions on the yield of coffee.

As an alternative, the entire web application can also be put in on a single raspberry pi. And Raspberry Pi should be connected to the LAN network in the farm. But this scheme has a major drawback. When the farmer is outside the farm, he would be unable to have full control over the equipment and hence real-time access on the data remains unachieved. A cost effective solution is yet to be found for ensuring global access to data in absence of the internet.

D. Algorithm

1. The data pushed to the stack from various sensors is collected.

2. Preprocessing is done on the collected data by
Normalizing it and treating the missing values.

3. Split the data into training and testing sets such that 70% is used for training and 30% is used for testing.

4. Train using Decision Tree Classifier, Naïve Bayes Classifier and SVM.

5. Test the accuracy

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4193 - 4202
Received 25 April 2021; Accepted 08 May 2021.

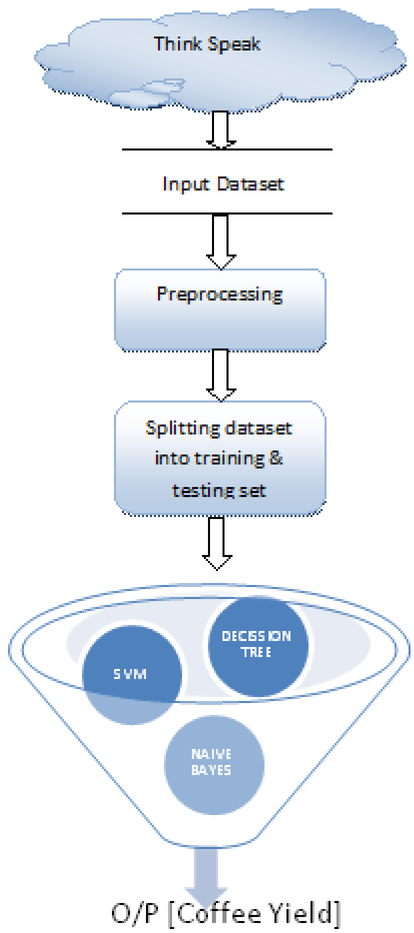

Fig.3 Workflow of ML Module

Results

Dataset: The dataset collected involves year, average rainfall, altitude, humidity, temperature, blossom showers, backing showers and the Nitrogen, Phosphorous and Potassium [NPK] composition of the soil as features.

The dataset was split into training set and testing test and were used to predict the yield of coffee using Decision Tree Classifier, Naïve Bayes Classifier and SVM.

A. Decision tree

An accuracy of 0.75 was achieved. In order to select the root node the value of gini index was used which is being calculated as shown below.

$$Gini = 1 - \sum_{i=1}^{c} (p_i)^2$$

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4193 - 4202
Received 25 April 2021; Accepted 08 May 2021.

If the value of Gini Index is 0.5, it denotes that there is equal distribution of elements into the classes specified.

p_i– probability of an object getting classified to a class.

Decision trees would prefer selecting the feature with least Gini index as the root.

Fig.4 Decision Tree obtained

B. Naïve Bayes

When we trained the Naïve Bayes Classifier and tested, an accuracy of 0.1 was achieved. This was because the predictors were independent. This algorithm works on Bayes theorem as shown below.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

C. Support Vector Machine

When we trained using Support Vector Machine, precision value of 0.78 was obtained. This denotes the ratio of no. of positive observations that has been correctly predicted to the total predicted positive observations.

$$Precision = \frac{TP}{TP + FP}$$

Recall obtained was 0.50, which is nothing but the ratio between no. of positive observations that has been correctly predicted to the all actual observations.

$$Recall = \frac{TP}{TP + FN}$$

F1 score obtained was 0.68, which is nothing but the weighted average of Precision and Recall. This score takes into account both false positives and false negatives.

$$F1\ Score = \frac{2 * (Recall * Precision)}{Recall + Precision}$$

Conclusion and Future Enhancement

Crop yield predictions can contribute a lot to food production sectors installed world-wide. The rapid advancement in sensing technologies in hand with the typical capability of ML based systems in processing large number of input data to derive predictions has been applied to design a cost-effective farmer support system, which would help in predicting coffee yield in the region of Yercaud, Tamilnadu. From the experimental results we were able to conclude that the Naïve Bayes Classifier can be preferred for studies related to yield predictions of crops. The proposed research was restricted with a small set of sample data due to unavailability of related precise datasets. Our future work would be to eradicate the limitations faced in this research and to build hybrid systems by integrating Machine Learning with more targeted sensor platforms and signal processing for best predictions in the field of agriculture.

References

[1] Reuben Varghese, Smarita Sharma , “Affordable Smart Farming Using IoT and Machine Learning”, Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3.

[2] Niketa Gandhi, Leisa J. Armstrong, “A review of the application of data mining techniques for decision making in agriculture”, 2nd International Conference on Contemporary Computing and Informatics, 2016.

[3] Ms. K.R. Sri Preetha, KanagaPriya, K.DivyaPrabha, S. Dharanipriya, “Crop Rotation And Yield Analysis Using Naive Ratio Classification”, International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017, 29 ISSN 2229-5518.

[4] N. Gandhi, L. Armstrong , O. Petkar and A. Tripathy, “Rice Crop Yield Prediction in India using Support Vector Machines”, Proceedings of IEEE, 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016.

[5] Sunil Kumar, Vivek Kumar, R. K. Sharma, “Rice Yield Forecasting Using Support Vector Machine”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019.

[6] SanjaBrdar , DubravkoCulibrk ´ , BrankoMarinkovic´ , Jovan Crnobaracand Vladimir Crnojevic´, "Support Vector Machines with Features Contribution Analysis for Agricultural Yield Prediction", https://www.researchgate.net/publication/301216536, April 2011.

[7] Gonzalez-Sanchez Alberto, Frausto-Solis Juan, OjedaBustamante W. Predictive ability of machine learning methods for massive crop yield prediction. Span J Agric Res.2014;12(2):313–2.

[8] P.Bhargavi, Dr.S.Jyothi, "Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils", IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.8, August 2009.

[9] Renuka, Sujata Terdal, "Evaluation of Machine Learning Algorithms for Crop Yield Prediction", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019.

[10] Andrew Crane, "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture", Andrew Crane – Droesch, 2018 Environ. Res. Lett. 13 114003.

[11] Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar, Chandan, "Crop Prediction System using Machine Learning", International Journal of Advance Engineering and Research Development (IJAERD) Special Issue on Recent Trends in Data Engineering, Volume 4, Special Issue 5, Dec 2017.

[12] Dr.N.Suma, Sandra Rhea Samson, S.Saranya, G.Shanmugapriya, R.Subhashri, "IOT Based Smart Agriculture Monitoring System", IJRITCC, Feb 2017.

[13] Lydia D. Isaac, Janani. I, "Integrating Feature Selection and Multiclass Classification for Sport Result Prediction", 'International Journal of Innovative Technology and Exploring Engineering (IJITEE)', ISSN: 2278–3075 (Online), Volume-9 Issue-1, November 2019.

[14] Savvas Dimitriadis, Christos Goumopoulos, "Applying Machine Learning to Extract New Knowledge in Precision Agriculture Applications", Panhellenic Conference on Informatics, 978-0-7695-3323-0/08 © 2008 IEEE, DOI 10.1109/PCI.2008.30.