Automated Control System for Crop Yield Prediction using Machine Learning Approach

Meeradevi1 and Monica R Mundada2
1Dept. of CSE, M S Ramaiah Institute Technology, Bangalore, India.
2Dept. of CSE, M S Ramaiah Institute Technology, Bangalore, India.

Abstract:
There have been various studies and research done on analyzing the patterns that can be observed in the yield production and predicting data of similar nature. To overcome the flaws of traditional agriculture such as large labor requirement, poor real time data acquisition, small monitoring area, Wireless Sensor Network based precision agriculture and prediction for increasing crop yield is done using data mining techniques. The proposed model uses only few sensors for experimental use. These nodes sense the environmental parameters like temperature, humidity, pH, NPK values. Analysis is done on sensed data which is captured from field and stored in server for further analysis. Thus, there is a need for a unique model that does prediction of the crop yield but uses these agriculture parameters to give results that help improving the crop yield. The farmer or researcher can do efficient farming using this technology.

Keywords: Data Mining, Yield, Irrigation, Prediction, Linear-Regression, Sensors.

1. INTRODUCTION
Farmer’s in India play major role to feed the growing population of the country. This makes crop analysis and prediction as important as crop production. Food crops in India are mostly seasonal and require monitoring of soil parameters like the amount of soil pH, Temperature, Humidity, NPK values of soil content. Rainfall and Temperature are the two environmental factors that impact the crop yield.

Using the past data of the said soil parameters associated with a piece of land or continuous sensing of data using sensors in a particular season for a particular crop, the yield can be predicted for future. New models of soil regeneration can be formed by analyzing the data [3]. Various algorithms have been designed in the domain of data analysis when it comes to predictions. In previous decades, significant transformation of agricultural sector is happened. In past, agriculture was driven by bid, but today it is driven by demand. Various classification algorithms are designed for prediction. The proposed method includes the result that helps in betterment of the yield. A farmer learns by experience as to how can the yield be improved. But, in extreme unexpected cases like a possible flood, this project when tied up with the weather forecasts can help by letting the farmers realize much faster as to how to protect their living, by giving away predictions that take into account these extreme scenarios. New information has to reach end-users very fast in order to use potential opportunities and achieve benefits. Information on seed, water, nutrients and plant protection is one of the main factors for successful farming. The proposed system discuss about predicting yield production. Yield prediction is a very important in agriculture sector for the farmers, prediction will help to overcome situation like drought, rainfall etc. Some data mining techniques such as K-Means and Multiple Regression methods can be used to provide the solution for predicting yield production. This aims to data models that achieve a high accuracy and a high generality in terms of yield prediction capabilities [1].

Wireless Sensor Network used for agriculture application for sensing environmental parameters and data analytics algorithm is used for predicting the crop yield and improving the crop yield and quality based sensed data. Due to wrong prediction farmers are suffering from loss of their yield. This real time monitoring of data will help to increase their yield by providing right amount of fertilizer and irrigate the crop only when required. Sensors can be deployed in the field of any size and clustering algorithm is used to form cluster and cluster head selection is done which aggregates the data from deployed nodes and sends this aggregated data to base station and from base station to end user mobile and user can monitor his crop being away. Data aggregation and transmission need more energy and energy is one of the important factors and another important factor is distance between the nodes and Base Station. Clustering algorithm will help to use energy efficiently.

2. PROPOSED SYSTEM
2.1. Crop Parameters
As shown in Figure1, The connections between the sensors and the microcontroller Atmega 328P are shown along with the connections between relay, battery and solar panel. Values sensed by the respective data pins of the sensors will be sent to the microcontroller to transfer the data to the user’s system and mobile.
The entire system is powered by solar energy. Various sensors are used in the system like Temperature sensor, Humidity sensor, pH sensor, soil Moisture sensor.

A relay is used to switch ON/OFF the motor which works on the power generated by the solar energy.

Setting up the pH sensor, which is read in the array variable “pHArray” such that multiple analog values of the same crop can be taken (in order to remove the errors) using the pH sensor and then accurately determining the correct analog pH value of the crop. This is done by average array() function. And the pH value is converted in terms of voltage and finally in terms of standard pH value using:

\[
voltage = \text{avergearray}(\text{pHArray}, \text{ArrayLenth}) \times \frac{5.0}{1024}
\]

\[
pHValue = 5 \times voltage + \text{Offset}
\]

The same is displayed on the LCD.

- Solar panel

Solar panel is used to charge the battery cells using sunlight for future use in case power is not available. Solar panel is made of solar cells. This provides increased efficiency by sunlight that falls on the panel. The panel has a rated capacity of 12 volts and 5 Watt power.

**Figure 1.** Circuit diagram

**Figure 2.** Block diagram
The values obtained by the sensors are sent to the microcontroller which is the driving element. Depending on the moisture content value obtained from the sensor, the motor is switched ON/OFF and the same status is displayed on the LCD and sent to the user via SMS using the GSM module driven by the microcontroller. The sensors status is sent to the server using GPRS module to predict the yield using linear regression. The status of the system is periodically displayed on the LCD and keeps the user updated about the field’s status. All the data collected from the sensors are stored in the database for analysis of the yield. The power to run the system is from the renewable source of energy i.e., solar energy.

3. EXPERIMENTAL SITE

3.1. Implementation

The system is completely automated irrigation systems designed for better implementation to suit almost all conditions and scenarios of testing [2]. The system is portable and real-time effective system. Arduino Software(IDE) is used to implement the system’s functionality. The data collected from all the sensors will be sent to the dedicated server on a user’s system.

The separate server is established using XAMPP which is a free and open source cross-platform web server. It consists of Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages. The data is sent to the dedicated server using the public URL that is established using ngrok which is a multiplatform tunneling, reverse proxy software that establishes secure tunnels from a public endpoint such as internet to a locally running network service while capturing all traffic for detailed inspection and replay. These data sent to the server will be used for further analysis and prediction of crop yield using Linear Regression algorithm. System is managing the water use because it provides irrigation as per the requirement of the crop. System monitors the motor and report by SMS using GSM. It has fast response time and easy in use. Solar based system is able to save energy and it is also environment friendly.

3.2. Linear Regression Algorithm

The implementation of key algorithm here is Linear Regression. Linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X.

Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

- The main goal is prediction. So, linear regression is used as predictive model for observed data set of y and X values. After developing such a model, an additional value of X is given without its accompanying value of y; the fitted model can be used to make a prediction of the value of y.

As shown in figure 6, the generation of a straight line between thousands of scattered point’s linear regression algorithm requires calculation. The steps for calculations are as shown below.

Step 1: root mean square error: \( \varepsilon = (\text{predicted} - \text{actual values}) \) and then finding root mean square of the same.

Step 2: mean: \( \text{sum(values)/float(len(values))} \) i.e Mean of the values dataset

Step 3: covariance: \( (x[i] - \text{mean}\_x) \times (y[i] - \text{mean}\_y) \)

Step 4: variance: \( \text{sum}((x - \text{mean})^2) \) for all x values dataset) for the prediction of the crop yield.

Let us consider crop is dependent on temperature. Let the crop yield be Y. Now, consider a dataset for learning or training model. So, job is to learn from the sensed dataset and predict future crop yield at any given temperature range.

Step 5: The equation \( Y = aT + \varepsilon \), where Y is crop yield, T is temperature and \( \varepsilon \) is error. We use a dataset to train the algorithm to generate a set of truth values which can be used for further analysis of new dataset created for crop yield, given the temperature. Then we find mean of all the crop yield values, along with variance and covariance.

Software Serial library will make pin 2 and 3 as Rx &Tx Include Liquid Crystal for LCD. Define variables for motor, soil moisture sensor, LED and other components. Then in void setup() function, initialize serial communication to 9600 bps and directions are given to the various PIns. Call gsmInit function to initialize the GSM module. The reading read by the sensors .In void loop() function, the pH of the soil is determined using the following calculations:

\[
\text{voltage} = \text{average(pHArray, ArrayLength)} \times \frac{5.0}{1024}; \text{pHValue} = 5 + \text{voltage} + \text{Offset};
\]

\[
\text{pHValue} = 5 + \text{voltage} + \text{Offset};
\]

Readings are taken in pHArray such that multiple values of the same crop is taken in order to remove errors using the pH Sensors and using average() to accurately determine the correct analog pH value of the crop. Then convert the value from voltage to standard pH value. The motor is turned on or off according to the moisture sensor status and a SMS is sent to the user using sendSMS function and the status of the entire module and functionality is continuously displayed on the LCD using lcd.print(). If the value of digitalRead(Mois_sensor)=0 then the moisture level in the soil is normal, motor will be off else if the value of digitalRead(Mois_sensor)=1 then the moisture level in the soil is low, so motor will be on as shown in figure 4. And this status is displayed on LCD accordingly. The data collected from all the sensors will be sent to the user and also to the dedicated server on a user’s system in order to perform regression analysis to predict the Yield. Check the various functions in full code given at the end. Here the
gsmInit() and initGPRS() functions are important and are used to initialize the GSM module and GPRS Internet connection.

3.3. Implementation of Modules

Atmega 328 P uses another algorithm to extract and send the data from the sensors to the server on the dedicated system through GPRS module and an SMS to the user’s mobile through GSM functionality.

4. RESULTS

4.1 System Setup

The soil moisture sensor keeps track of the moisture content of the soil continuously. The below figure 3 shows when soil moisture is low. When the moisture content of the soil becomes low, a signal is sent by the soil moisture sensor to AtMega 328P board and the AtMega 328P board sends a signal to the relay to close the circuit and the water motor switches ON. The status of the of the motor is displayed on the LCD display and a SMS is sent to the user using GSM/GPRS module. Below is the image which depicts the LCD display when moisture content is low.

Figure 3. Soil moisture content low

Figure 3 shows soil moisture content is normal. The soil moisture sensor keeps track of the moisture content of the soil continuously. When the moisture content of the soil becomes normal, a signal is sent by the soil moisture sensor to AtMega 328P board and the AtMega 328P board sends a signal to the relay to break the circuit and the water motor switches OFF. The status of the of the motor is displayed on the LCD display and a SMS is sent to the user using GSM/GPRS module.

Below is the image which depicts the LCD display when moisture content is normal.

Figure 4. System Setup

Along with the soil moisture content, various other values like pH of the soil, Temperature and humidity is sent to the user and displayed on the LCD. And these all values are also sent to the main server where the data is collected and used for predicting the crop yield.

4.2. Explanation og Module

After collecting data from the microcontroller through the server on the system, the python script will analyze and generate crop yield based on the trained crop yield values and straight line along the region of the scattered points on the graph between temperature and crop yield. The output shown on the webpage in Figure 5, has timestamp, temperature, humidity, moisture, pH of the soil. Data is sensed starting from 28th April 2018 as shown in below table that consists of 10 entries out of 281 entries. The yield expected is 87% after prediction using the proposed system.

Figure 5. Real Time Sensed Soil parameters

The system uses Linear regression to predict the crop yield based on the parameters collected from the microcontroller, the data collected through server will be stored in a dedicated
system and will be accessed by the python script, in which the algorithm is implemented. The algorithm will use the dataset to plot graphs between temperature and the crop yield to be predicted and a straight line which is close to all the plotted points will be generated as the linear line which can be used for predicting crop yield given a temperature range[4]. Temperature being the dependent variable and crop yield being the explanatory or independent variable.

After collecting data from the microcontroller through the server on the system as shown in below figure 6, the python script will analyze and generate crop yield based on the trained crop yield values and straight line along the region of the scattered points on the graph is between temperature, crop yield and Moisture, crop yield drawn using linear regression.

5. CONCLUSION
A prediction model is built to predict agriculture production according to the collected data to identify a hidden pattern, and to observe variation in the parameters of the soil to make better decisions. The system is simple and reliable. It is portable and real-time effective system. The main goal of this system is to predict and improve crop yield, and save time, manpower & energy. Main advantage of this design to provide automated detection and alerting system. System is managing the usage of water because it provides irrigation as per the requirement of the crop. System operates the water motor based on soil moisture and reports by SMS using GSM module to the user’s mobile. It has fast response time and easy in use. Solar based system is able to save energy and it is also environment friendly.

ACKNOWLEDGEMENT
We are grateful to Ramaiah Institute of Technology, Bangalore for providing immense support in guidance and testing of the research project. We are thankful for our students Subramaniyan S, Tallam Amar Kumar, Vikas M S, Vivek R for helping in completion of the research work.

REFERENCES

