AUTOMATED IRRIGATION AND CROP PREDICTION

1st Komal Menger
Information Technology
Vidyalankar institute of Technology
Mumbai,India
komal.menger@gmail.com

2nd Sowmya Iyer
Information Technology
Vidyalankar institute of Technology
Mumbai,India
iyer.sowmya18@gmail.com

3rd Srushti Jamsandekar
Information Technology
Vidyalankar institute of Technology
Mumbai,India
srushtijamsandekar1998@gmail.com

4th Prof. Rohit Barve
Information Technology
Vidyalankar institute of Technology
Mumbai,India
rohit.barve@vit.edu.in

Abstract—This project is based on a national issue, of shortage of water in the villages and the problem that exists in irrigation. The idea came from an interview of Hon. Prime Minister of India, Mr. Narendra Modi who once addressed the issue of water supply to the crops, the basic idea being, the water supplied to the crops should be in adequate quantity such that there should not be any wastage of water. A solution to it is by designing a water supply system which will take into consideration the moisture level of the soil and dispense the water according to the need of the soil, which would not damage the crop and also water utilization won't be excessive. Also, Selection of crops is an important issue for agriculture planning.This selection should depend upon various parameters of the soil such as the soil type, its texture, and the amount of nutrients present in the soil. It is very essential to ensure that the nutrients required by the crops are either already present in the soil or are added by the farmers in appropriate quantity.

Keywords—Agriculture, Automatic Irrigation, IoT, PH Sensor, Arduino, Fertilizer prediction, Machine learning, Linear Regression, Decision tree, NPK, Production prediction, Crop Prediction.

I. INTRODUCTION

This project is designed basically by keeping in mind our farmers who strive a lot and also works day and night to provide us healthy food. Now-a-days farmers are adapting a revolutionary change in using modern agricultural implements and machines to enhance his ability and capability of working. This project will help him to notice the moisture of water in time and provide water to his field without taxing himself of going to the field personally, but it will be done automatically. This may give him relief to a certain level. If this machinery is implemented and practised it will be definitely liked by many farmers and will help to grow economically. Aim is to develop a controlled smart irrigation system to provide an irrigation system which is automatic for the plants which help in saving water and money. The main objective is to improve the health of the plants and at the same time reduce the undue usage of water. It has to made sure that the crops are neither over irrigated nor under irrigated.By using cost efficient sensors, we can continuously monitor the water content in the soil and as and when the water level will decrease and cross a certain threshold, the sensors will activate the sprinkler or a pump connected to it leading to water supply to the crops. This setup will help the small scale farmers to improve the crop yield which result in growing th Indian economy. The system design includes soil moisture sensors, the real time sensed data will be compared with predetermined threshold values of various soil and specific crops. The deployed sensors data are fed to the Arduino Uno processor which is linked to the cloud storage. The data received by the cloud is used to perform data analysis using data mining techniques and machine learning algorithm will be applied which predicts the suitable crops that can be grown in that particular soil. It will also predict the production of crops. Farmers can get to know the type of soil and can plough the crops depending on the type of soil. Also if a farmer wants to yield a particular crop then the improvement of soil can be done by adding the required nutrients in the soil.

II. LITERATURE SURVEY

Sadia Afrin, Abu Talha Khan, Mahrin Mahia, Rahbar Ahsan Analysed the soil Properties and Climatic Data to predict crop production and also performed Cluster analysis on different agricultural regions of Bangladesh by using methods like K-means, PAM and DBSCAN. Along with these four linear regression methods to predict crop yields were used.[1]

K. L. Ponce Guevara1, J. Palacios Echeverra1 developed Green Farm-DM: A tool that analyses data of vegetable crops from a greenhouse and applied data mining techniques.Algorithm C4.5 uses a decision tree based on the data entropy was used to determine the results and results were visualized graphically.[2]

Dharesh Vadalia, Minal Vaity, Krutika Tawate, Dynaneshwar Kapse, “Real Time soil fertility analyzer and crop prediction”- The proposed system determines the basic parameters of soil like pH and EC (electrical conductivity) which majorly affect the quality of soil. This system includes pH and Electrochemical sensors connected with Arduino board. PH value is used to determine the contamination of Nitrogen , Phosphorus and potassium which determines the soil quality.[3]

Mrs. N. Hemageetha, Dr. G.M. Nasira “Analysis of soil condition Based on pH value using Classification Technique” They applied data mining algorithms on agriculture data. The dataset was collected from agriculture university which contained various attributes like soil type, pH value, NPK and electrical conductivity. Four different classification algorithms on preprocessed dataset were applied and compared all the results and then chose one classifier which gave the best result. According to their system, J48 classifier gave the best result with highest accuracy.[4]

III. METHODOLOGY

Datasets:

The data sets used in this project have been obtained from a site that already exists for the farmers by the government of India. For prototyping our model, we have used 34 major crops that are grown in major parts of Maharashtra.The parameters used for classifying the soil are as follows:

1.Rainfall

2.Temperature

3.pH

The original dataset consisted of more than 1,00,000 data values, after applying various data pre processing techniques and removal of outliers and noise the data was consolidated to 1500 data points for further analysis.

	A	B	C	D	E	F
1	Rainfall	Temperat	Ph	Crop	Production	
2	400.1508	20	3	Bajra	0.0069	
3	400.1633	20	3.2	Bajra	0.00747	
4	400.1639	20	3.2	Bajra	0.00749	
5	400.1797	20	3.2	Bajra	0.00822	
6	400.1958	20	3.2	Bajra	0.00895	
7	400.2905	20	3.2	Bajra	0.01328	
8	400.3887	20	3.2	Bajra	0.01777	
9	400.3989	20	3.2	Bajra	0.01824	
10	400.4159	20	3.2	Bajra	0.01902	
11	400.4561	20	3.2	Bajra	0.02085	
12	400.4604	20	3.2	Bajra	0.02105	
13	400.4672	20	3.2	Bajra	0.02136	
14	400.5277	20	3.2	Bajra	0.02413	
15	400.542	20	3.2	Bajra	0.02478	
16	400.5644	20	3.2	Bajra	0.02581	
17	400.5686	20	3.2	Bajra	0.026	
18	400.6572	20	3.2	Bajra	0.03005	
19	400.6943	20	3.2	Bajra	0.03175	
20	400.7698	20	3.2	Bajra	0.0352	
21	400.7811	20	3.2	Bajra	0.03571	
22	401.0274	20	3.2	Bajra	0.04698	
23	401.3183	20	3.2	Bajra	0.06028	

regressiondb

Fig. 1. - Dataset collected for regression

This data is then used to predict the crop that can grow in a particular soil and also to predict the production of a particular crop.

Installation of software and libraries :

We have used the software PyCharm for programming in python. To perform analysis and prediction using machine learning, we have installed a few modules in Python. To plot graph and produce output, various packages like - Numpy, Pandas, Sklearn, Seaborns, Statsmodels, Matplotlib and CSV.

III. PROPOSED SYSTEM

We propose a system that consists of the following five steps:

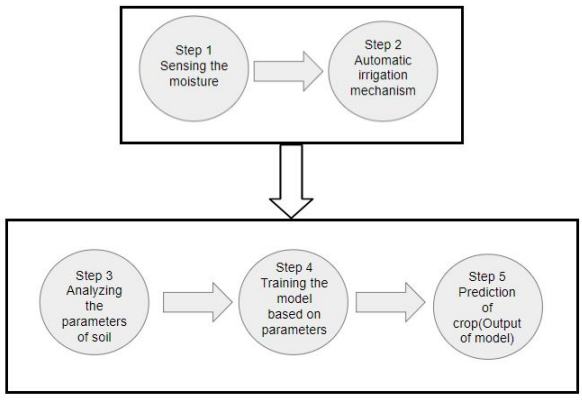

Fig. 2. - System flow

Step 1: Sensing the data
Step 2: Automatic irrigation mechanism
Step 3:Analyzing the parameters of soil

Step 4: Training the model based on parameters
Step 5: Prediction of crops and its production (Output model)

IV. MODELS AND TECHNIQUES

We have used two different machine algorithms for different purposes.

1. Regression model:

This model is used to obtain relationships between the production of crops and soil parameters. This algorithm will help us in predicting the amount of production for a particular crop.
We are using linear regression to identify the relationship of production with each of the parameters.

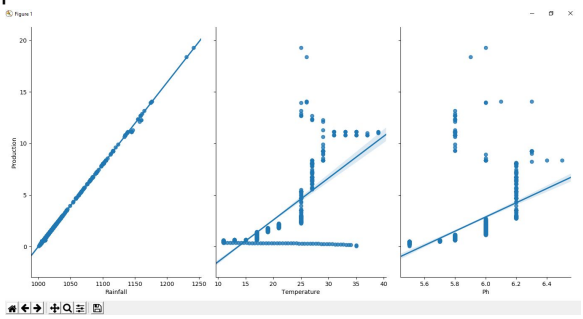

Fig. 3. - Representation of relationship between production of crops and different parameters.

The algorithm then gives us a linear equation. Coefficients of the linear equation $y=\beta 0 + \beta 1 \times Rainfall + \beta 2 \times Temperature + \beta 3 \times Ph$ are then used for generating the linear line in the above graph.

2. Decision tree:

Decision tree is a model which trains a model based on the observation of its features, forming the structure of a tree in the process thereby giving meaningful result .Greedy approach is used by the decision tree which chooses the attribute at first step. Also it overfit the training data which can give poor results for unseen data.The result obtained from this model is usually better than the result from any one of individual models.

V. EXPERIMENTAL SETTING

We are using three main sensors in this project, that is soil moisture sensor, pH sensor and DHT sensor. They are used to sense various parameters of the soil.

Fig. 4. - pH sensor

This sensor senses the pH of the soil, thereby indicating if it is acidic or basic and we can predict if the soil needs fertilizers or not.

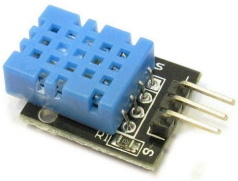

Fig. 5. - DHT sensor

This sensor detects the humidity and temperature of the soil .

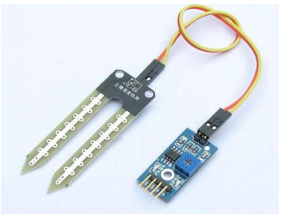

Fig. 6. - Soil moisture sensor

This will sense the moisture content of the soil which will contribute to irrigating the soil whenever the sensor detects a lower value.

All these sensors together contribute to giving real time parameters of the soil. Depending on the analysis that has been done using our Machine learning algorithms like Decision tree and linear regression, we can automatically irrigate the soil as well as predict the crops and their production.

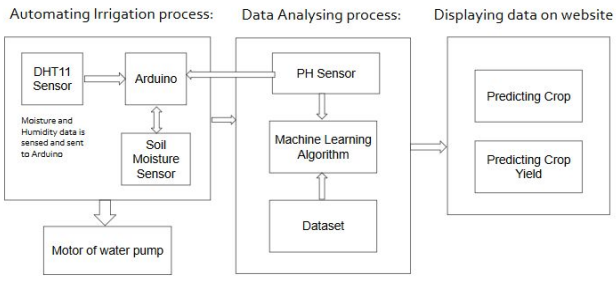

Fig.7.-.Implementation

VI. RESULT AND DISCUSSIONS

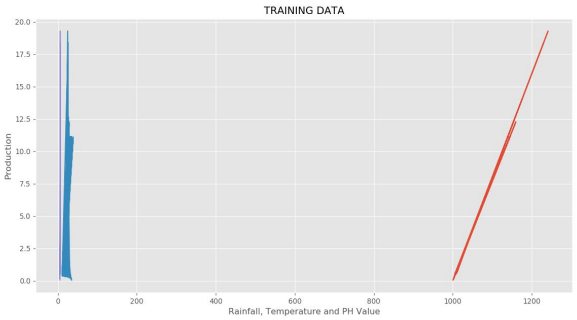

Fig.8. -Representation of training data for linear regression of Crop Production with Rainfall,Temperature and Ph Value

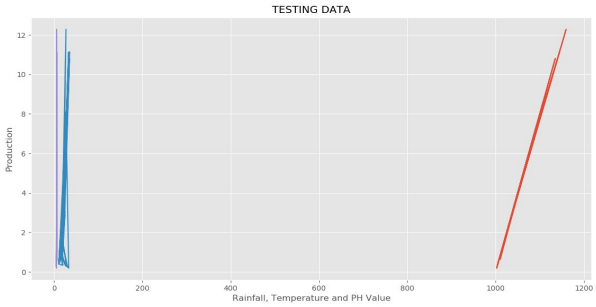

Fig.9.- Representation of testing data for linear regression of Crop Production with Rainfall,Temperature and Ph Value
Grey line: Relation between Ph Value and Crop Production
Blue line: Relation between Temperature and Crop Production
Red line : Relation between Rainfall and Crop Production

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Fig.10.-Formula for accuracy

Root Mean Squared Error (RMSE) is used for evaluating the model.

To check the accuracy in multiple linear regression , use the variance. If variance score is
near about 1 is perfect prediction.The figure below is the screenshot of the accuracy of our model as well as the predicted crop.

```
C:\Users\komal\PycharmProjects\untitledl\venv\Scripts\python.exe C
   temperature   humidity       ph    rainfall label
0    20.879744  82.002744  6.502985  202.935536  rice
The data present in one row of the dataset is
   temperature   humidity       ph  ...  rice  watermelon  wheat
0    20.879744  82.002744  6.502985  ...     1           0      0

[1 rows x 34 columns]
The accuracy of this model is:  89.56989247311829
The predicted crop is jute
```

Fig.11.-Accuracy and prediction

VI. CONCLUSIONS AND FUTURE SCOPE

This model currently works with temperature, humidity, rainfall, area and total amount of production as the working parameter.In the future,we are planning to consider the other parameters of the soil that affects the growth and production of crops so as to obtain a more accurate prediction.Suggesting fertilizers for crops based on contamination in soil is remaining.We hope to contribute in the betterment of the country by doing and carrying our research forward in the future.Our focus will also be on creating a proper database for the research and make it open source and available

to the public.Doing so would enable any researcher who wants to contribute in this field, have easier access to data.Future researchers would not need to go through the lengthy and hectic process of primary data collection and data processing. Our current dataset can be considered as the base of the future work and any future contribution can be made based on this structure.

VI. REFERENCES:

[1] Afrin, S., Ahsan, R., Mishal, Khan, A. T., Mahia, M., Rahman, R. M., M. R., Ahmed, W. (2018). Analysis of Soil Properties and Climatic Data to Predict Crop Yields. 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS).

[2] Ponce-Guevara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E.,Dominguez Limaico, H. M., Suarez-Zambrano, L. E., Rosero-Montalvo, P.D., Alvarado-Perez, J. C. (2017). GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial). 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM).

[3] Dharesh Vadalia, Minal Vaity, Krutika Tawate, Dynaneshwar Kapse, “Real Time soil fertility analyzer and crop prediction”, IRJET, Volume 4, Issue 3, March 17.

[4] Mrs. N. Hemageetha, Dr. G.M. Nasira, “Analysis of soil condition Based on pH value using Classification Technique”, IOSR-JCE, Volume 18, Issue 6, Nov 16.