An Efficient Method for Soil Analysis using FARM-IT Approach

Melvin Martis
St. John College Of
Engineering and Management,
Palghar

Paul Lobo
St. John College Of
Engineering and
Management,
Palghar

Shreesh Chavan
St. John College Of
Engineering and Management,
Palghar

ABSTRACT
Agriculture has benefited from improvement in data mining and automation, in recent times. Based on training dataset, this paper suggests a data mining method to analyze soil content such as pH, Field Capacity, Wilting Point, Nitrogen, Potassium, Phosphorus, etc. and helps determine the amount of fertilizer required. Additionally, an irrigation facility based on automated detection of moisture can help optimize the use of water. This system uses a server. The sensors output to server through a distributed network connected wirelessly via a tool such as, ZigBee. This project aims at optimizing the use of fertilizer and water, and accordingly, suggest proper actions.

General Terms
Soil moisture, Sensors, Soil analysis, Crop analysis and yield prediction, Fertilizers, Soil temperature, Classification, et al.

Keywords
Naïve Bayes, Iterative Dichotomiser 3 (ID3), WEKA tool, Data mining, NPK values, Decision tree, Algorithms.

1. INTRODUCTION
In India, an agrarian country, it has become indispensable to improve issues to fulfill the demand of a large population. Excessive use of water and fertilizers is a leading factor for decline in yield of crops. The financial losses are immense since the crops are half developed.

A farming assistant can guide the farmer to be active in his farm at the right time. The goal of this project is to develop a software tool that accepts input of soil (N, P, K and moisture values) and determines amount of fertilizer needed. It also predicts crop yield based on training data acquired from various sources [1]. This ensures optimum use of fertilizers and prevents crop destruction. The algorithm and interface used are collectively named as FARM-IT.

2. RELATED WORK
A literature survey related to our project has helped us to better understand algorithms and techniques required for our project. Also, based on the research, we have come across sensors that can help us automate moisture control for irrigation facility. Naïve Bayes is an algorithm that has been widely used, based on our survey of soil and agriculture related projects.

2.1 Automatic irrigation system on sensing soil moisture content [2]
This report concentrates on automated irrigation. The project's goal is to automate irrigation. The water pump is switched on or off after reading the soil moisture level.

Employing this technique ensures minimum human intervention, still providing appropriate irrigation. It uses arduino board, programmed to determine moisture level in soil and automatically activate the pump if it goes below the threshold value.

This project assists us in implementing a similar system as a part of our project.

2.2 Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction [3]
According to this paper, data mining can be used to determine which class a soil sample belongs to. It also uses regression to predict attributes. WEKA tool is used to classify soil, comparing it against labels. Result of this project is decrease in the price of soil tests, which can save efforts and time of Indian soil testing laboratories.

This paper helps us to study algorithms used, as well as, tools used for implementing data mining.

2.3 A data mining approach to soil temperature and moisture prediction [4]
Soil moisture and temperature datasets are used to train the data mining algorithms to predict based on its knowledge base. Inappropriate watering techniques by farmers, lead to decrease in yield. This research uses regression method to predict level of moisture in the soil sample, as well as, temperature based on its knowledge base. This in turn helps the farmers to know and decide the steps to be taken for efficient irrigation.

Since, the training data set is not varied, this report lacks accuracy, to some extent. It doesn't consider the observation as a single database.

2.4 Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils [5]
This project classifies a large amount of soil data into classes using Naïve Bayes approach. The classification is done primarily based on the texture and type of soil. Soil data is acquired from many sources and analyzed before classifying iteratively.

Since, Naïve Bayes approach is used for classification of large amount of soil data, this research assists us in better understanding the algorithm.

3. KNOWLEDGE PRE-REQUISITES

The classification of soil data sets is complex due to large number of variables involved. A table has been generated to generally classify the attributes of soil. [6]

Table 1. Classification of soil attributes

Input Data Required	Description		
pH Value	Acidic (0-6.9)	Neutral (7.0)	Basic (7.1-13.9)
Moisture	Moisture values depend on: Field Capacity and Wilting point		
Soil Content	Nitrogen	Phosphorus	Potassium
Soil Type	Medium Black Soil		
Crops	Cash Crops	Around the year crops	Vegetables

Nitrogen, Phosphorus, Potassium (NPK) values are input based on laboratory tests of soil samples. These inputs along with moisture and temperature, play an important role in deciding classification of crops and what fertilizers are required. The above general classification is used to create a more detailed classification in the form of decision tree.

Here technique such as clustering can be implemented using WEKA tool to create clusters of the soil based on their input values. Further classification is done using ID3 algorithm.

4. PROPOSED METHODOLOGY

The proposed methodology consists of ID3 for constructing decision trees followed by Naïve Bayes for classification purpose.

4.1 ID3

ID3 (Iterative Dichotomiser 3) is used to create decision trees based on the classes (e.g. low, high, medium). This information is used by Naïve Bayes to train the algorithm and apply it to a test input data.

Attributes are classified in ID3, into various classes such as, low, medium or high. The nodes with child classes are called as "decision nodes". The child nodes that cannot be further divided (leaf nodes), are "classes" to which the attributes belong, given a condition. [7]

For example, if N-value (Nitrogen-value) is less than 40 percent it is classified as "low". Here, N-value is the decision node and "low" is the class to which the soil sample belongs.

Entropy is a function of how unordered input data set is. Hence, the amount of data required for classification is directly proportional to increase in Entropy. Entropy, should therefore be as low as possible. This can be achieved by filtering the data manually before providing it as input to the algorithm.

Entropy of data is given by the following formula:

$$Entropy = -p(a) * \log[p(a)] - p(b) * \log[p(b)]$$

Information is the opposite of Entropy. **Information gain (I.G)** is the value left after classifying the instances. It is the difference between Entropy after the classification and before.

$$I.G = (Entropy)_{after} - (Entropy)_{before}$$

Algorithm:

Step 1: if (attribute-value < threshold)

    classify as "low"

  else if (attribute-value between range)

    classify as "medium"

  else

    classify as "high"

Step 2: Repeat for all attributes

ID3 algorithm iteratively classifies all the data.

4.2 Naive Bayes

Naïve Bayes is a classification technique. It assumes that the input data is independent. [8]

Naive Bayes classifier can be trained efficiently and quickly. The algorithm belongs to supervised learning, that is, it learns from the training data.

$$p(y|x) = \frac{p(x|y). p(y)}{p(x)}$$

Conditional probability is given by the formula: [9]

where,

$p(y|x)$ = Posterior Probability that attribute y belongs to class x

$p(x|y)$ = Likelihood that the class is x in the event that attribute is y

$p(x)$ = Probability that class type is x (out of all possible values of x)

$p(y)$ = Probability that attribute is y

This can be given to the whole dataset as training data. Consider an example, how data is received as input in Naïve Bayes.

Example: Training Set:
{("Rice", low, high, high), ("Cotton", low, medium, medium), …}

This training set determines what crop is best suited for respective values of N, P, and K.

Table 2. Sample Input to Naïve Bayes Algorithm

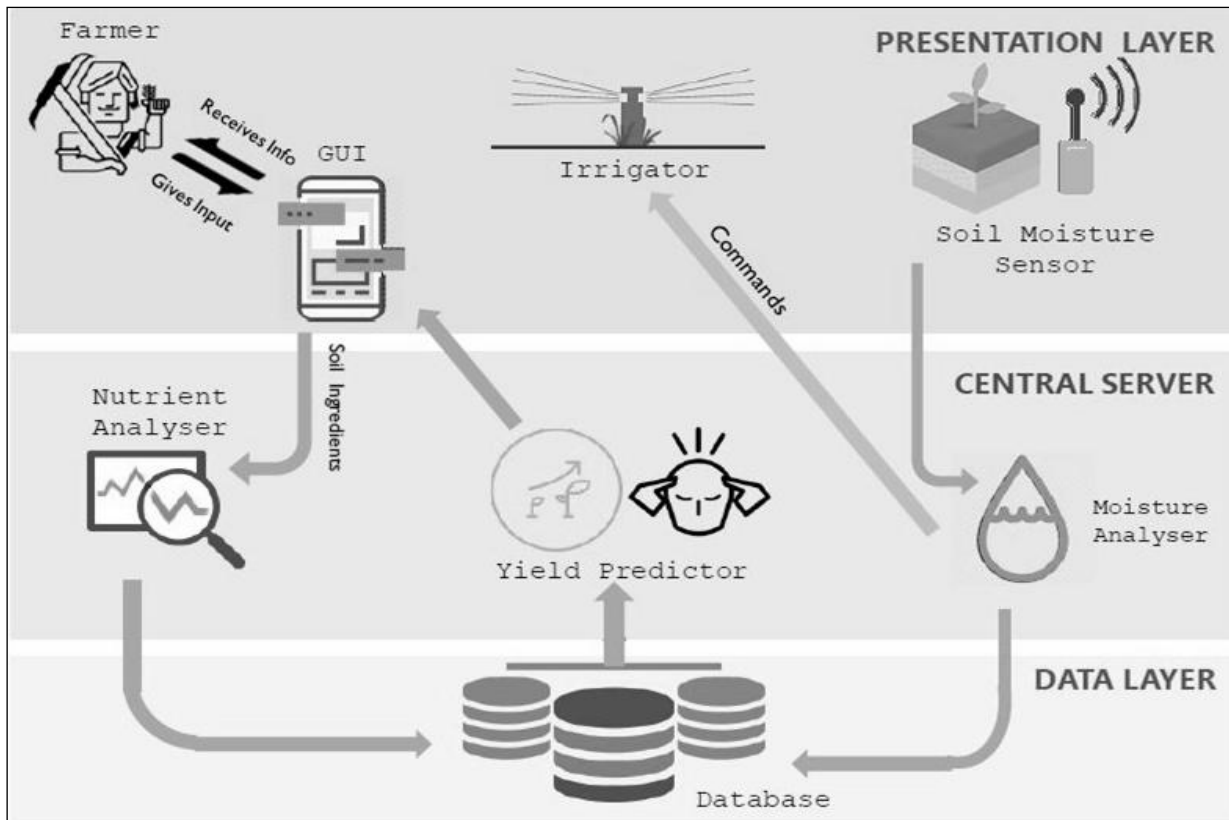| Attribute → Class ↓	Nitrogen	Phosphorus	Potassium
Rice	low	high	high
Cotton	low	medium	medium
Rice	low	medium	high

5. ARCHITECTURE

Fig 1: Proposed Architecture

The proposed architecture consists of 3 layers:

1. Presentation Layer
The presentation layer of the proposed method consists of Farmer, Irrigator, GUI and Sensors. The GUI is a proposed web interface, since a web platform provides portability and support across any Operating System.

2. Central Server
The central server deals with the storage and retrieval of the soil data which consists of 3 parts:

  i. **Nutrient Analyzer**: The inputs received from user are analyzed by nutrient analyzer, which further stores the data into a database.

  ii. **Yield Predictor**: It predicts yield using the proposed methodology and displays it on the GUI.

  iii. **Moisture Analyzer**: It uses the input from moisture sensor to analyze and interpret the moisture content obtained from the soil.

3. Data Layer
The data layer serves as a storage medium for all the inputs received from the central server and provides output to the yield predictor.

The data layer receives data from 2 sources:

  i. Nutrient Analyzer

  ii. Moisture Analyzer

6. CONCLUSION

This project would help a regular user to maintain his farm better by providing fertilizers in the right amount. It can also help the user by suggesting a schedule. The predictions and suggestions are made based on past records and present input. It also makes use of feedback from the user to improve the prediction. The network of moisture sensors provides real time irrigation and minimizes the load of watering. This helps save water and also, prevents over saturation of soil.

FARM-IT is therefore an efficient and important tool for a regular farmer, that can help him in his day-to-day agricultural activities and also help improve his farming habits.

7. FUTURE WORK

Due to lack of real-time sensors for analyzing soil content, such as, N, P and K, suggestions cannot be provided immediately. Real time sensors for detection of soil components, can be used to receive NPK values in real time, like the moisture sensor. This would greatly help the user as, regular visits to the lab for soil testing can be avoided. Since, these sensors are currently unavailable, future scope of this project could include, making use of these sensors as and when they are available. Besides, more factors such as detection of carbon content (organic matter), accurate weather forecast can be used, depending upon the location.

Therefore, the project can be improved by using in-situ sensors and considering more elements for measuring soil content in real time.

9. REFERENCES

[1] Department of Agril. Chem. & Soil Science, Dr.B.S. Konkan Krishi Vidyapeeth, Dapoli, Dist. Ratnagiri., MH, IN.

[2] Nagarajapandian M , Ram Prasanth U , Selva Kumar G, Tamil Selvan S, "Automatic irrigation system on sensing soil moisture content", International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering, Vol. 3, Issue 1, January 2015.

[3] Jay Gholap, "Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction" Asian Journal of Computer Science and Information Technology Vol. 2(8) (2012) pp. 251– 252.

[4] William Myers, Seth Linden, and Gerry Wiener, "A data mining approach to soil temperature and moisture prediction" Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[5] 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.

[6] "Soil attributes as predictors of crop production under standardized conditions", Stenberg, B. Biol. Fertil. Soils (1998) 27: 104.

[7] Hitarthi Bhatt, Shraddha Mehta, Lynette R. D'mello, Dept. of Computer Engineering, D.J. Sanghvi College of Engineering, "Use of ID3 Decision Tree Algorithm for Placement Prediction" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5), 2015, 4785-478

[8] Haiyi Zhang, Di Li Jodrey School of Computer Science Acadia University, Canada, "Naïve Bayes Text Classifier" 2007 IEEE International Conference on Granular Computing DOI 10.1109/GrC.2007.40

[9] Ray, S (2015, September 13). 6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python). Retrieved from https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/