International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 213–220

Soil Fertility Prediction Using Data Mining Techniques

Rajat Chaudhari [1], Saurabh Chaudhari [2], Atik Shaikh [3], Ragini Chiloba [4], Prof.T.D.Khadtare[5]

Department of Computer Engineering, Savitribai Phule Pune University

rajatchaudhari281997@gmail.com
saurabhchaudhari3014@gmail.com
shaikhatikrajjak@gmail.com
raginichiloba1233@gmail.com
tanajikhadtare@gmail.com

Abstract

Agriculture industry in India is the greatest sector for employment but lack of research in this sector is the reason behind less productivity. It's important to implement computational research, Machine learning techniques in Agriculture industry to make India better quality and quantity producer in food sector. Machine Learning techniques are useful in abstracting patterns and establishing relationships between varied data sets and predicting reasonable outputs. It can be efficiently applied in Agriculture industry to improve efficiency in this sector. We have discussed application of Machine learning techniques in Agriculture sector to analyze fertility of soil. Agriculture industry has been always one of the interested areas of research. This study venture to analyze soil data depending upon various factors, classify it and improve efficiency of each model using different combinations. Agricultural research has been profited by technical advances such as automation, data mining. Today, data mining is used in a vast areas and many off-the-shelf data mining system products and domain specific data mining application soft wares are available, but data mining in agricultural soil datasets is a relatively a young research field. The large amounts of data that are nowadays virtually harvested along with the crops have to be analyzed and should be used to their full extent.

Keywords: ANN; SVM; Decision Tree; kNN; Soil;

I. INTRODUCTION

Agriculture industry in India is the greatest sector for employment but lack of research in this sector is the reason behind less productivity. It's important to implement computational research, Machine learning techniques in Agriculture industry to make India better quality and quantity producer in food sector. Machine Learning techniques are useful in abstracting patterns and establishing relationships between varied data sets and predicting reasonable outputs. It can be efficiently applied in Agriculture industry to improve efficiency in this sector. The paper considered Indian soil data to predict the fertility of soil considering various elements like nitrogen, oxygen, Phosphorous, porosity and various environmental elements like air, temperature. India has diverse weather conditions as well as soil type across different regions.

Soil fertility defines growth of plants when other environmental factors like light, water, temperature are favorable. Soil fertility is influence by several factors like Climate, irrigation (Soil water), Soil, acidity, Soil

ISSN: 2233-7857 IJFGCN

alkalinity, Nutrition in Soil. Globalization, changing weather condition, urbanization, higher use of pesticides is the reason of decreasing quality of soil in India. Deficient soil type lead to less agricultural production and ultimately higher cost of food products. Different soil types are used to analyze fertility of soil. The ultimate goal of applying technology in Agricultural with minimal impact in fertility of soil and quality of food product [2]. Machine Learning techniques are useful in abstracting patterns and establishing relationships between varied data sets and predicting reasonable outputs. Agriculture industry in India is the greatest sector considered for employment and has been part of research. Machine learning techniques can be efficiently applied in Agriculture industry to improve research. In current scope of project, we have developed a model for fertility of soil based on different soil type. After receiving fertility depending upon various soil type, a comparative study of machine learning techniques such as ANN, linear regression, SVM and Decision tree is carried out.

Field	Description
Ph	pH value of soil
EC	Electrical conductivity, decisiemen per meter
OC	Organic Carbon, %
P	Phosphorous, ppm
K	Potassium, ppm
Fe	Iron, ppm
Zn	Zinc, ppm
Mn	Manganese, ppm
Cu	Copper, ppm

Fig-1: Soil Nutrients Data set Description

2. LITERATURE SURVEY

The literature survey is divided into 2 main themes. In the first section, traditional methods were analyzed but these methods did not give perfect results. So, in the second section data mining methods were introduced to give better output.

1. **Literature review on Traditional Approaches of soil fertility prediction**

ISSN: 2233-7857 IJFGCN

Dharesh Vadalia, Minal Vaity [3], their proposed system determined the basic constituents of soil like pH and electrical conductivity which majorly affect the quality of soil. This system includes portable device which is made up using pH and EC sensors and Arduino board along with the analog to digital converter. Sensors sensed the pH and EC of particular soil sample gives the value to the Arduino board in real time. Analog to Digital Converter is used to convert analog ph. value to digital value. Arduino board requires 9V power supply which is given by adapter and sensors require 3.3V-5V power. With the help of Arduino, pH value is converted into Nitrogen, Phosphorus and potassium which determines the soil quality. Arduino displayed NPK values on display screen and farmer have to manually enter NPK values in his own remote device application. Application will give digitally generated fertility report which contain suitable crops and required fertilizer.

S. Sivachandran and K. Balakrishnan [4] Embedded Based Soil Analyzer is used to analyze various soil nutrients with the help of pH value. As per the availability of nutrients, recommendations of cultivating the particular crop will be given. This project uses microcontroller which determines PH of diluted soil sample using Glass ph. electrode. PH is defined as the negative logarithm (base 10) of the activity of hydronium ions (H+or, more precisely, H3O+aq) in a solution. It ranges from 0 to 14, with 7 being neutral. A pH below 7 is acidic and above 7 is basic. The optimum pH range for most plants is between 5.5 and 7.0. The system includes Microcontroller Unit, Signal conditioning, Sensors, Display, Thermal Printer and Power supply. Output from pH sensors and EC sensor are analog in nature to process these analog signals to the system A/D converters are used. PIC18F458 Microcontroller has the inbuilt analog to digital converter and no need to connect the external A/D converter for the system. In this system, keypad is used to connect the user and the system.

Siuli Roy, SomprakashBandyopadhyay,[5] With the help of wireless sensor network agricultural parameter guarantees increase in production and lower input costs in precision farming by real time monitoring of location based specific environmental and soil conditions. It also not only improves crop management but reduces waste and labor costs also. The test bed implementation of a wireless sensor network has been presented in this paper for automatic and real-time monitoring of soil and environmental parameters which influences crop production. The paper demonstrates technical challenges including energy management scheme, placement of sensors in outdoor environment, the integration of sensors, actual power consumption rates and remaining practical issues.

Apurva C. Pusatkar, Vijay S. Gulhane [6], the advanced development in wireless sensor networks was used in monitoring various parameters in agriculture. In this context, due to the advancements in small-scale sensor devices with wireless technologies, one is able to remotely monitor humidity, temperature and moisture. In this paper it was proposed to implement a wireless sensor network connected with centralized basic node using ZigBee, which was connected to a Central Monitoring Station (CMS) through Global System for Mobile (GSM) technologies or General Packet Radio Service (GPRS). The system acquires Global Positioning System (GPS) parameters related to the field then transfers them to a central monitoring station. This system was presumed to evaluate soil conditions and act accordingly in order to help farmers. This system implies monitoring various factors such as humidity, soil moisture and provide remote monitoring using ZigBee which sends data wirelessly to a central server which collects data store it and allow it to be displayed as needed and also be sent to the client mobile.

ISSN: 2233-7857 IJFGCN

2. Literature review on Data Mining approaches of soil fertility prediction

Monali Paul and Santosh K. Vishwakarma [7] used Data Mining Techniques for the prediction of crop yield. K-Means algorithm was used to forecast the pollution in the atmosphere and then applied for simulating for daily precipitations. The researcher proposed soil profile descriptions for classifying soils in combination with GPS based technologies. K-Means approach was applied to it for soil classification. Crop classifications using hyperspectral data were carried out by adopting one of the data mining approaches which is Support Vector Machines. The researchers worked on rainfall variability analysis and its impact on crop productivity. The effect of observed seasonal climatic conditions such as rainfall and temperature variability on crop yield prediction was considered through an empirical crop model.

Amol D. Vibhute [8] used Hyperspectral remote sensing has been widely used for mapping of soil. The researcher first pre-processed the satellite image which was necessary to reconstruct the degraded image for making an appropriate original view and to improve quality for better feature extraction and classification. For this Support Vector Machine (SVM) is used. SVM offers the training dataset which depends on closest pixels of the separation hyperplane or boundary between classes. The satellite Hyperspectral data was trained using acquired real time ground truth points which were the support vectors in the classification. Support vector machine classifier was evaluated for the mapping and classification of the Hyperspectral data.

Li Dongming and Zhang Lijuan [9] focuses on the analysis of the basic meaning in data mining and the structure of the decision tree uses the decision tree algorithm—C4.5 to establish a soil quality grade prediction model and combines the soil composition to be a training sample. First, the decision tree carries out data learning, according to each data of soil quality levels in the training sample, then gets the decision tree that accords with learning rules. The Decision Tree shows the relationship between the composition of soil and soil quality grade. The decision tree C4.5 algorithm provides a method to solve this difficulty, and it provides a reliable theory basis for precision fertilization.

P. Vinciya, Dr. A. Valarmathi [10] used Multiple Linear Regression [MLR] technique for crop analysis. Decision tree algorithm and Classification is used to perform analysis of over 362 datasets and provide result. The training dataset here is classified into as organic, inorganic and real estate for predicting the type of soil. Three methods are used which includes Decision tree, Naive Bayes Classifier, and KNN Classifier which analyses soil and predicts crop yield, however rule-based induction and SVM can be used for more accuracy as results are not accurate.

Jay Gholap and Shailesh Gargade [11] used Data Mining Techniques for the soil quality analysis. The ID3(Iterative Dicotomizer) Algorithm is a non-incremental algorithm means it derives its classes from a fixed set of training instances. The dataset which contains attributes like phosphorous, nitrogen, oxygen etc. with class labels. By using this training data set we are going to classify the sample of soil into different fertility levels by taking its attribute. So, for this we need to create a decision tree based on training dataset by using Iterative Dicotomizer 3 algorithm. A decision tree is a multiple nodes or binary nodes tree which contains nodes which are all of given dataset branches values leaf node which are the class labels. While creating decision tree first we need to decide root node. For this the information gain of all attributes are calculated and the attributes with highest IG (Information Gain) is selected as root node then the value becomes branches of tree and repeatedly using this formula Decision Tree is constructed. After constructing

ISSN: 2233-7857 IJFGCN

the decision tree, the attributes of the soil like phosphorous, nitrogen, hydrogen which we want to test for quality or fertility of soil and finally the class label for that unclassified record will obtained. In this way this the soil quality analysis was achieved by using the algorithm of Data Mining Iterative Dicotomizer 3.

Ankita Kharat and Samiksha Dhulap [12] used the Techniques of Data Mining for categorization of soil. Soil Classification is the important one. It influences many properties and significance of land use and management. Soil classification deals with the systematic categorization of soils based on distinguishing characteristics as well as criteria that dictate choices in use. The classification of Soil data [2] using GA Tree technique, which is a decision tree builder which is based on Genetic Algorithms (GAs). GA Tree offers some unique features that are not found in any other tree inducers while at the same time it can produce better results for many difficult problems. The Genetic Algorithms, the binary Decision that is built first for soil categorization. To generate the decision tree for the dataset of soil which include data of soil attributes it takes more time and then the rules are classified with that decision tree. Genetic based Decision rules are effective for soil data categorization.

Dr. S. Jyothi [13] includes soil classification using Fuzzy Rules. This paper is to discuss how fuzzy classification rules are generated for soil data. The fuzzy rules represent the chemical properties like P, N, K of soil. Soil classification can effectively reduce the complexity of information and help us to understand the main feature of soil attributes . In this paper, we generate fuzzy rules from training data to deal the soil data classification problem, by first defining the membership functions for the input attributes of the soil data, and then generating the member functions defined for the attributes of the soil data and then we merge these fuzzy rules in order to generate definitive fuzzy rules for suitable classification of soil.

3. Proposed System

The system aims to help farmers to predict the soil fertility for better yield production. To be precise and accurate in predicting fertility, the project analyze the nutrient contents present in the soil. For our system we will be training soil database and we will classify that particular soil sample into particular range using classification algorithms . It can be achieved using unsupervised and supervised learning algorithms, like Decision tree, kNN, SVM, Naive bayes. Dataset will then trained by learning algorithms. It compares the accuracy obtained by different learning techniques and the result will be delivered to the end user.

ISSN: 2233-7857 IJFGCN

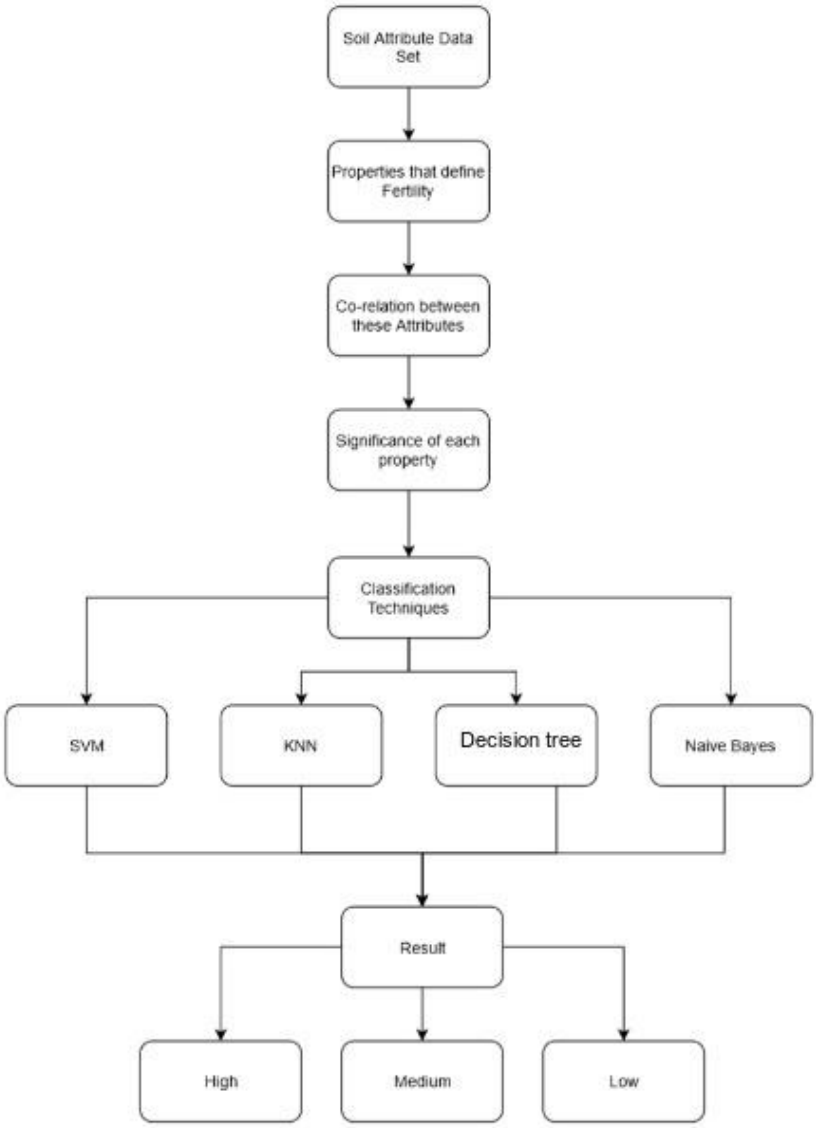

Fig-2: Soil Fertility Precision conceptual Flow

4. RESULT

We have applied four important Data mining techniques of classification which are decision tree, Navie bayes, kNN and SVM. For each algorithm data is divided in test and train. The performance of different classifier is compared on fact that how each of them has classified data correctly.

ISSN: 2233-7857 IJFGCN

Classification Techniques	Accuracy
Decision Tree	60.0
SVM	44.61538461538462
KNN	55.8974358974359
Naive Bayes	43.07692307692308

Fig-3: Comparison of Accuracy

From the above we can conclude that Decision tree performs best among all the technique.

5. CONCLUSION

After analyzing a number of papers, we conclude that soil fertility prediction will help to reduce the difficulties faced by the farmers and act as a medium to provide the farmers efficient information required to get high yield and thus maximize profits which in turn will reduce the suicide rates and lessen his difficulties. A model is implemented to predict the fertility of soil. The system uses supervised and unsupervised Machine learning algorithms and gives best result based on accuracy. The results of the four algorithms will be compared and the one giving the best and accurate output will be selected. The model has been tested by applying different kinds of machine learning algorithm. Generally, SVM shows good accuracy but among all the classifiers, Decision tree gives the highest accuracy in soil classification. The proposed model is justified by a properly made dataset and machine learning algorithms.

ISSN: 2233-7857 IJFGCN

REFERENCES

[1] Nikhil Deshmukh, Shailesh Rathi, Shital Thakare, Siddharth Chaudhary, "Predicting Fertility of Soil Using Data Mining Techniques in School of Computing", National College of Ireland, 2018.

[2] Rahman, S. A. Z., Chandra Mitra, K., & Mohidul Islam, S. M, "Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series", (ICCIT), 2018.

[3] Dharesh Vadalia, Minal Vaity, Krutika Tawate, Dynaneshwar Kapse, "Real Time soil fertility analyzer and crop prediction", IRJET, 2017.

[4] S. Sivachandran and K. Balakrishnan, "Real Time Embedded Based Soil Analyzer", International Research Journal of Engineering and Technology (IRJET), 2014.

[5] Siuli Roy, Somprakash Bandyopadhyay, "A Test-Bed on Real-Time Monitoring of Agricultural Parameters Using Wireless Sensor Networks For Precision Agriculture", 2015.

[6] Implementation of Wireless Sensor Network for Real Time Monitoring of Agriculture. International research journal of engineering and technology (IRJET), 2016.

[7] Paul, M., Vishwakarma, S. K., & Verma, A, "Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach", International Conference on Computational Intelligence and Communication Networks (CICN), 2015.

[8] Vibhute, A. D., Kale, K. V., Dhumal, R. K., & Mehrotra, S. C., "Soil type classification and mapping using hyperspectral remote sensing data", International Conference on Man and Machine Interfacing (MAMI), 2015.

[9] Dongming, L., Yan, L., Chao, Y., Chaoran, L., Huan, L., & Lijuan, Z,"The application of decision tree C4.5 algorithm to soil quality grade forecasting model.", First IEEE International Conference on Computer Communication and the Internet (ICCCI), 2016.

[10] P. Vinciya, Dr. A. Valarmathi, "Agriculture Analysis for Next Generation High Tech Farming in Data Mining", IJARCSSE, 2016.

[11] Gholap, J., Ingole, A., Gohil, J., Gargade, S., & Attar, "Soil data analysis using classification techniques and soil attribute prediction", 2012.

[12] Komal Varadkar, Ankita Kharat, Anushree Deshmukh, "Soil Quality Analysis using Data Mining Comparison of Two Algorithm ID3 and Naïve Bayes), IJARCSSE, 2019.

[13] P. Bhargavi, S. Jyothi, "Soil Classification by generating Fuzzy rules", IJCSE, 2010.

ISSN: 2233-7857 IJFGCN