Classification of Soil and Crop Suggestion using Machine Learning Techniques

Mrs. N. Saranya
Assistant Professor
Department of Computer Science and Engineering
Sri Shakthi Institute of Engineering and Technology,
Coimbatore, Tamil Nadu ,India

Ms. A. Mythili
PG Scholar
Department of Computer Science and Engineering
Sri Shakthi Institute of Engineering and Technology,
Coimbatore, Tamil Nadu ,India

Abstract-Agriculture is the major source for living for the people of India. Agriculture research is the major source of economy for the country. Soil is an important key factor for agriculture .There are several soil varieties in India. In order to predict the type of crop that can be cultivated in that particular soil type we need to understand the features and characteristics of the soil type. Machine learning techniques provides a flexible way in this case. Classifying the soil according to the soil nutrients is much beneficial or the famers to predict which crop can be cultivated in a particular soil type. Data mining and machine learning is still an emerging technique in the field of agriculture and horticulture. In this paper we have proposed a method for classifying the soil according to the macro nutrients and micro nutrients and predicting the type of crop that can be cultivate in that particular soil type. Several type of machine learning algorithms are used such as K-Nearest Neighbour (k-NN), Bagged tree, Support vector machine(SVM) and logistic regression.

Keywords- Machine learning, agriculture, soil, classification, nutrients, chemical feature, accuracy.

I. INTRODUCTION

Data mining has been used for analyzing large data sets and establish classification and patterns in the datasets. The techniques are used to elicit significant knowledge that can be easily predictable by individuals. Data mining is a challenging technology in the field of agriculture. Nowadays data mining has been used in the field of agriculture for soil classification, wasteland management, crop and pest management[1]. In assessed the association rules of affiliation methods in DM and applied into the soil science to anticipate the significant connections and gave association rules to different soil types in agriculture. The agriculture factors such as rain, weather, soil type, pesticides and fertilizers are the main responsible to increase the production. The main reason for agriculture is to grow crops. Crop cultivation depends on the nature and the nutrients of the soil increasing the cultivation of land which brings a loss of supplements present in the soil. In the crop cultivation soil plays an important role. It is important for plants, animals rocks and living organisms. All of these helps in managing the fertility of the soil.

A soil test is carried out to identify the nutrients content, composition and other components contained in the soil.Soil tests are mainly conducted to measure the fertility and other defencies present in the soil so that suitable measure can be taken to resolve it.

Machine learning is a field of computer science where new developments evolve at recent times , and also helps in automating the evaluation and processing done by the mankind ,thus by reducing the burden on human power. Machine learning is the field of Artificial Intelligence by the dint of which computers can be taught without explicit programming. In simple terms, the meaning of machine learning is basic algorithms can provide information about a dataset without writing code to solve this program manually. Instead of writing code you provide data or the basic algorithm and it forms its own conclusions based on this data. In machine learning agriculture, the methods are derived from learning process. Those methodologies need to learn through experiences to perform a particular task. Once the learning is completed then the model can then be used to make an assumption to classify and to test data. The data is achieved after gaining the experience of the training process.

Classification is the main problem in data mining. Classification is a data mining technique based on machine learning which is used to categorize the data item in a dataset into a set of predefined classes. It helps in finding the diversity between the objects and concepts. It also provides necessary information for which research can be done in a systematic manner.

II. LITERATURE REVIEW

In a research carried out by Zaminur Rahman a comparative study of several machine learning techniques has been carried out. They have carried out the classification using the data of Bangladesh. Considered the six district soil data and used the geographical features for classification. They have used k Nearest Neighbour, Bagged tree and SVM finally compared the results of three algorithms and brought out a model for classifying the soil types and the suitable crop that can be cultivated in that particular soil type[3].Among the used three algorithms SVM has obtained the average accuracy.

In a research carried out by Leisa J.Armstrong a comparative study of data mining algorithms. They have used a large dataset extracted from the Australian Department of Agriculture and Food(AGRIC) to conduct the research[4].

In an approach carried out by Jay Gholap carried out a modal to classify the soil based on fertility. The dataset was collected from the soil testing laboratories of Pune District. They have used WEKA tool for developing an automated system[5].

Chiranjeevi M. N carried out a research for classifying the soil types so that it can be useful for the farmers for analyzing the type o soil and the crop that can be cultivated so that there will a good yield and profit. They have considered the data mining algorithms for classifying the soil.They have used algorithms such as J48 decision tree classifier and Naïve bayes classifier among these two algorithms Naïve bayes has obtained the maximum accuracy of 98%[6].

III. DATA MINING PROCESS

The methodology involves the following steps:
- Dataset collection
- Pre-processing
- Classification
- Prediction
- Result

A .Data collection

Most of the research papers carried out the model using the chemical parameters , water content, electrical conductivity, organic content and the fertility. The values of these are taken as inputs for the algorithm[7].

B. Pre-processing

For a successful completion of a model a huge set of data is required. The data that is collected from real world might be in raw format .It may contain some missing values, inconsistent and noisy values. In this step such redundant values should be filtered. The data is made normalized.

C. Classification

It is one of the data mining .This is used to analyze the data and allocates it into a separate class. In pre processing step a prototype is developed. In classification the removed prototypical is tested against the pre defined dataset. That is to quantify the prototypical trained performance and accuracy.

D. Prediction

The presentation of classification algorithm associated based on accuracy and performance analysis and will provide a suggestion for the farmers to cultivate in a particular soil type.

E. Results

The final result gives the suggestion of crops.

IV. CLASSIFICATION ALGORITHMS

Some of the classification algorithms used are k-Nearest Neighbour, SVM and logistic regression.

A.   K-Nearest Neighbour classifier

It is one of the method used for classification and regression. An object is classified by a plurality vote of its neighbours, with the object being assigned to the class most common among bits k nearest neighbours.It is widely used in real life scenarios since it is a non parametric meaning it does not make any underlying assumptions about the distribution data[8] .
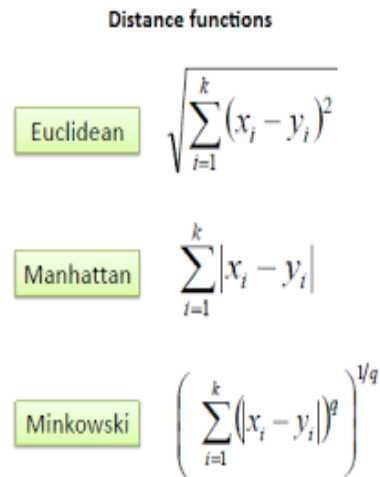
Distance functions

Distance functions

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}\left(|x_i - y_i|\right)^q\right)^{1/q}$

Fig.1.Distance functions

B.   Support Vector Machine

Support vector machine is a supervised machine learning model that uses classification problems. After giving an SVM model sets of labelled training data for either two categories, they are able to categorize new examples. It works based on the decision planes that defines the boundaries. The decision plane separates one object from another object of different class. The data points that are nearer to the hyper plane are called as support vectors[9]. Kernel function is used to separate non linear data by transforming input to a higher dimensional space .Gaussian radial basis function kernel is used in the model

$K(X_i ,X_j) = e^{- \|x_i , x_j\|^2/2\varphi^2}$

where $K (X_i, X_j)$ =input vectors in input space, $\| X_i , X_j \|^2$= higher dimensional space of X and Y coordinate and σ is a free parameter.

C.   Logistic Regression

It is one of the predictive analysis. It is one of the predictive analysis. It is a classification algorithm used to assign observations to a discrete set of classes. The hypothesis of logistic regression tends it to limit the cost function between 0 and 1.

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}}$$

Fig 2.Formula

V. PROPOSED METHODOLOGY

The system architecture of the proposed model is shown in the below figure:
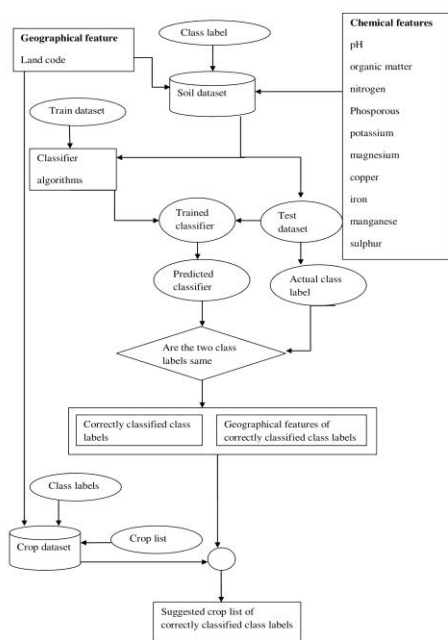
Fig. 3.Proposed architecture

The proposed system involves two phases the training and testing phase. It uses two database the soil database and crop database[10]. The soil database includes the chemical features and geographical features of the soil. Table 1 shows the chemical features of the soil.

Table 1.Chemical attributes

Attribute	Details
pH	pH value of soil
Salinity	Ds/meter
Organic Matter %	Percentage
Potassium	Mili equivalent/100 gram soil
Sulphur	Microgram/per gram soil
Zinc	Microgram/per gram soil
Boron	Microgram/per gram soil
Calcium	Mili equivalent/100 gram soil
Magnesium	Mili equivalent/100 gram soil
Copper	Microgram/per gram soil
Iron	Microgram/per gram soil
Manganese	Microgram/per gram soil

VI. RESULT ANALYSIS

The proposed model is based on soil and crop database. Several machine learning algorithms are used to classify the soil type. For a particular soil type suitable crop is suggested. From the experimental result ,We see that SVM has obtained the maximum accuracy. The classification accuracy is tabled below.

Table 2.Result

	Method	Accuracy
Ramesh V .,et.al [9]	J48	91.9%
Gholap Jay	J48	92.7%
Proposed work	SVM	96%

VII. CONCLUSION AND FUTURE ENHANCEMENT

A model is proposed for predicting the soil type and suggest a suitable crop that can be cultivated in that soil. The model has been tested using various machine learning algorithms such as kNN, SVM and logistic regression. The accuracy of the present model is maximum than the existing models. In future suitable fertilizers are suggested for the well growth of the crop cultivated. The present models deals with available old data whereas the future model contain the real time a data that is directly received from agricultural land that is placed with sensors .The sensors senses the soil fertility and other minerals contained in the soil.

REFERENCES

[1] V. Rajeshwari and K. Arunesh, " Analyzing Soil Data using Data Mining Classification techniques," Vol 9(19),May 2016.
[2] Jay Gholap , Anurag Ingole , Jayesh Gohil, Shailesh Gargade, Vahida Attar (2013), "Soil data analysis using classification techniques and soil attribute prediction,".
[3] Sk Al Zaminur Rahman, Kaushik Chandra Mitra ,S.M. Mohidul Islam(2018),"Soil classification using Machine Learning Methods and Crop Suggestion based on Soil Series".
[4] L.Armstrong , D.Diepevven & R. Maddern(2004),"The Application of Data Mining Techniques to categorize agricultural soil profiles".
[5] Chiranjeevi .M .N , Ranajana B Nadagoundar(2018)," Analysis of Soil Nutrients using Data Mining Techniques".
[6] Ramesh Vamanan, K.Kumar (2008),"Classification of Agricultural Land Soils A Data Mining Approach".
[7] Chandrakar PK , Kumar S, Mukherjee D(2011), "Applying classification techniques in Data Mining in agricultural land soil".
[8] Campus –Valls G , Gomez –Chova L , Calpe –Maravilla J, Soria –Olivas E, Martin –Guerreo JD, Moreno J(2003) Support vector machines for crop classification using hyperspectral data.
[9] Bhuyar V(2014) , "Comparative analysis of classification techniques on soil data to predict fertility rate for Auranagbad District ".
[10] T .Mathavi Parvathi ,"Automated soil testing process using combined mining process"Manonmaniam Sundaranar University.