



Soil Classification and Harvest Proposal Implemented using Machine Learning Techniques

N.Saranya, A.Mythili

Abstract: The major source of living for the people of India is agriculture. It is considered as important economy for the country. India is one of the country that suffer from natural calamities like drought and flood that may destroy the crops which may lead to heavy loss for the people doing agriculture. Predicting the crop type can help them to cultivate the suitable crop that can be cultivated in that particular soil type. Soil is one major factor or agriculture. There are several types of soil available in our county. In order to classify the soil type we need to understand the characteristics of the soil. Data mining and machine learning is one of the emerging technology in the field of agriculture and horticulture. In order to classify the soil type and Provide suggestion of fertilizers that can improve the growth of the crop cultivated in that particular soil type plays major role in agriculture. For that here exploring Several machine learning algorithms such as Support vector machine(SVM),k-Nearest Neighbour(k-NN) and logistic regression are used to classify the soil type.

Keywords: Machine Learning, Mining, Nutrients, Agriculture, Chemical Feature Classification, Accuracy, Prediction.

I. INTRODUCTION

Data mining has been used for analyzing large datasets and establish the classification and relationship among the data in the dataset. The techniques are used to elicit the knowledge that can be easily predicted by individuals. Nowadays mining is a challenging technology in the field of agriculture[1]. Agricultural and biological researches has used various mining techniques that could be used to classify the soil type based on the feature. Soil is one of the key factor that is most essential for the growth of a crop. The soil contains various chemical features such as the micro and macro nutrients. The micro nutrients are already available in the soil such as zinc, potassium, copper, magnesium and sulphur. These nutrients are available in the soil as default. When they are available in lesser percentage it can be boosted up by adding the fertilizers whether natural or artificial fertilizers. The agricultural factors considered are rainfall, weather, soil type, fertilizers and pH. The main reason for agriculture is to cultivate crops. Several people of India depends on agriculture for food. Crop cultivation majorly depends on the climatic conditions and the nutrients present in the soil which increase the cultivation of the crop. Soil is a major ingredient for the crop for its growth[2].

Revised Manuscript Received on October 30, 2020.

* Correspondence Author

Mrs. N.Saranya, Assistant Professor Department of Computer Science and Engineering Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India E-mail.nsaranya@siet.ac.in

Ms. A.Mythili, PG Scholar Department of Computer Science and Engineering Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu ,India E-mail.mythilijoe20@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Soil should be classified according to its texture and its chemical feature. A soil dirt test is completed to distinguish the supplements substance, creation and different parts contained in the soil. Soil tests are for the most part led to quantify the ripeness and different deficiencies present in the soil with the goal that reasonable measure can be taken to determine it. Machine learning is a field of software engineering where new advancements advance at ongoing occasions, and furthermore helps in robotizing the assessment and preparing done by the humankind, subsequently by decreasing the weight on human force. AI is the field of Artificial Intelligence by the dint of which PCs can be educated without express programming. In straightforward terms, the importance of AI is fundamental calculations can give data about a dataset without composing code to fathom this program physically[3]. Rather than composing code you give information or the essential calculation and it shapes its own decisions dependent on this information. In AI agriculture, the strategies are gotten from learning process. Those philosophies need to learn through encounters to play out a specific undertaking. When the learning is finished then the model would then be able to be utilized to make a supposition to characterize and to test information[4]. The information is accomplished in the wake of picking up the experience of the training procedure. Classification is the fundamental issue in data mining. Order is an information mining strategy dependent on AI which is utilized to arrange the information thing in a dataset into a lot of predefined classes. It helps in finding the decent variety between the items and ideas. It additionally gives essential data to which research should be possible in a deliberate way.

II. MACHINE LEARNING TYPES

A. Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed[5]. Machine learning focuses on the development of computer programs that can access data and use it for learn themselves.

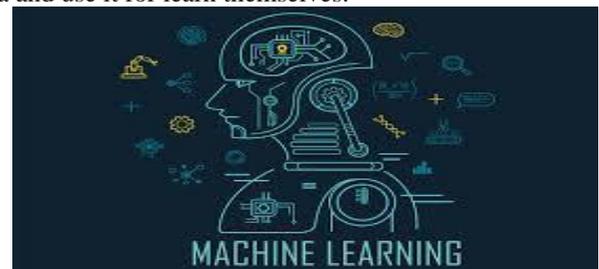


Fig 1.Machine Learning



B. Supervised learning

The supervised machine learning algorithms are those algorithms which needs outside help. To be classified and each branch represents a value that the node can take. The input dataset is separated into train and test dataset. The train dataset has output variable which should be predicted or grouped. All calculations take in an examples from the preparation data set and apply them to the test data set for forecast or classification. The work process of supervised machine learning algorithm is given in Fig.2

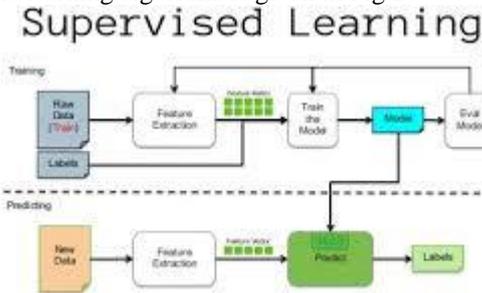


Fig.2 Supervised learning

C. Unsupervised learning

Exactly when new data is displayed, it uses the as of late learned features to see the class of the information. It is for the most part utilized for prediction and clustering. A work process of unsupervised learning is given in Fig.3.

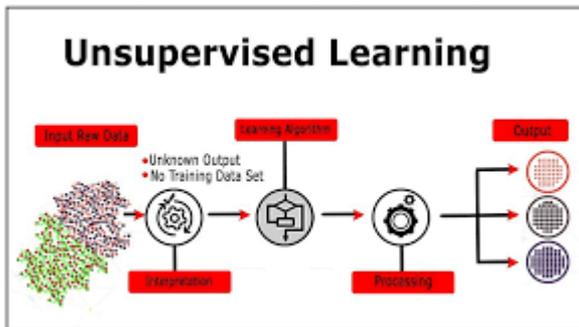


Fig.3 Unsupervised machine learning

Two of the main methods used in unsupervised learning are principal component and cluster analysis. Cluster analysis is used in unsupervised learning to group, or segment, datasets with shared attributes in order to extrapolate algorithmic relationships. Cluster analysis is a branch of machine learning that groups the data that has not been labelled, classified or categorized. Instead of responding to feedback, cluster analysis identifies commonalities in the data and reacts based on the presence or absence of such commonalities in each new piece of data. This approach helps detect anomalous data points that do not fit into either group.

D. Semi Supervised learning

Semi-supervised learning algorithm is a strategy which consolidates the intensity of both regulated and solo learning it will in general be natural item full in those zones of AI and data mining where the unlabelled data is starting at now present and getting the named data is a monotonous system. There are numerous classes of semi-supervised learning.

E. Reinforcement Learning

Reinforcement learning is a sort of realizing which settles on choices dependent on which moves to make to such an extent that the result is increasingly positive. The student has no information which moves to make until it's been given a circumstance. The move which is made by the student may influence circumstances and their activities later on. reinforcement adapting exclusively relies upon two criteria: experimentation search and deferred result. The common model for reinforcement learning is portrayed in Fig 4.



Fig.4.Reinforcement learning

In the figure 4, the specialist gets an information I, current state s, state change r and info work I from nature. In light of these data sources, the operator produces a conduct B and makes a move a which creates a result. E. Perform Multitask learning which has a straightforward objective of helping different students to do better. When perform multiple tasks learning calculations are apply to undertaking, it recalls the methodology how it tackled the issue or how it scopes to the specific end. The calculation at that point utilizes these means to discover the arrangement of other comparable issue or errand. This encouraging of one calculation to another can likewise be named as inductive exchange system. In the event that the students share their involvement in one another, the students can adapt simultaneously instead of separately and can be a lot quicker. F. Outfit Learning at the point when diverse individual understudies are merged to outline only a solitary understudy then that particular sort of learning is called assembling learning. The individual understudy may be Naïve Bayes, choice tree, neural system, and so on. Outfit learning is a hotly debated issue since 1990s. It has been seen that, an assortment of students is frequently better at making a specific showing as opposed to solitary students.

III. METHODOLOGY

The methodology used for developing a model that classify the soil type according to its chemical features and gives a suggestion o crops that can be cultivated in that suitable soil type[6]. Finally it gives a suggestion of fertilizers that can be helpful for the growth of the crop that is cultivated.

Some of the classification algorithms that are been used or classification of soil are listed:

A. K-nearest Neighbour

The purpose of the k Nearest Neighbours (kNN) algorithm is to use a database in which the data points are separated into several separate classes to predict the classification of a new sample point.

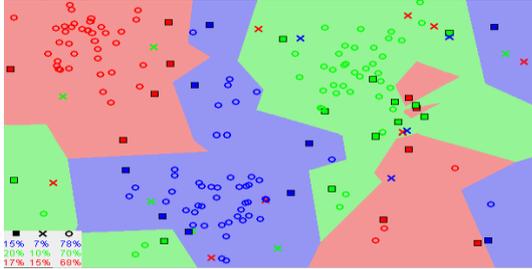


Fig 5. K nearest neighbour

Image showing how similar data points typically exist close to each other here similar data points are close to each other[7]. The KNN algorithm hinges on this assumption being true enough for the algorithm to be useful. KNN captures the idea of similarity.

The KNN Algorithm

1. Load the data
2. Initialize K to your chosen number of neighbours
3. For each example in the data

Steps that are followed are

- i. 3.1 Calculate the distance between the query example and the current example from the data.
- ii. Add the distance and the index of the example to an ordered collection
- iii. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances
- iv. Pick the first K entries from the sorted collection
- v. Get the labels of the selected K entries
- vi. If regression, return the mean of the K labels
- vii. If classification, return the mode of the K labels

B. Support vector machine

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane[8]. In other words, given labeled training data (*supervised learning*), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side. In real world application, finding perfect class for millions of training data set takes lot of time. This is called regularization parameter. Here we define two terms regularization parameter and gamma. These are tuning parameters in SVM classifier.

B. Regularization

The Regularization parameter (often termed as C parameter in python’s sklearn library) tells the SVM optimization how

much you want to avoid misclassifying each training example.

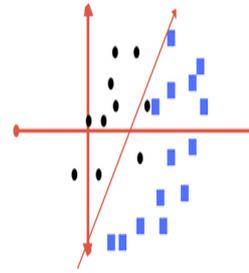


Fig .6 Lower regularization value

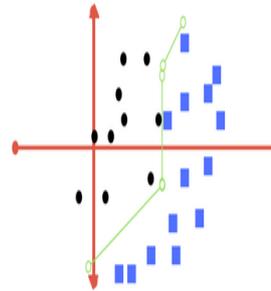


Fig.7 Higher regularization value

D. Gamma

The gamma parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’. In other words, with low gamma, points far away from plausible separation line are considered in calculation for the separation line[9]. Where as high gamma means the points close to plausible line.

E. Logistic Regression

Logistic regression is one of the most popular machine learning algorithms for binary classification. This is because it is a simple algorithm that performs very well on a wide range of problems.

F. Logistic Function

The logistic function is defined as:
transformed = $1 / (1 + e^{-x})$

Where e is the numerical constant Euler’s number and x is a input we plug into the function.

G. Logistic Regression Model

The logistic regression model takes real-valued inputs and makes a prediction as to the probability of the input belonging to the default class (class 0).If the probability is > 0.5 we can take the output as a prediction for the default class (class 0), otherwise the prediction is for the other class (class 1)[10].

IV. IMPLEMENTATION

The work flow of the proposed methodology is shown in the below figure



Soil Classification and Harvest Proposal Implemented using Machine Learning Techniques

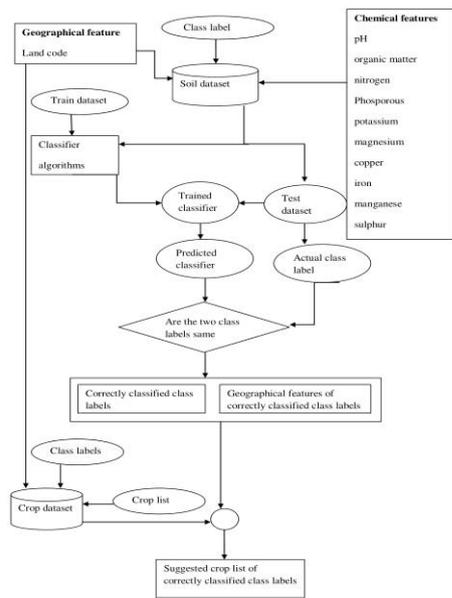


Fig. 8 Proposed architecture

The proposed system consists of two phases the first phase is the soil database and the next phase is the crop database. The soil is classified according to the chemical features such as the pH, potassium, sulphur, zinc, calcium, magnesium, copper and manganese.

V. RESULT ANALYSIS

The proposed model consists of two datasets the soil and crop datasets. Many algorithms have been used to classify the soil types. Among them the logistic regression has obtained the maximum accuracy. The suitable crop type is suggested for a particular soil type. The below table gives the accuracy of the used algorithms.

Table 1. Result

Algorithm	Accuracy
Support vector machine	885
K Nearest neighbour	92.3%
Logistic regression	96.4%

VI. CONCLUSION AND FUTURE WORK

A model is developed for classifying the soil series using suitable machine learning algorithms. A suitable crop type is suggested for a particular soil type. The developed model gives more accuracy than the existing model. In the future work the suitable fertilizer will be suggested which will increase the growth of the crop. Then automatically the cultivation increases. A suitable sensor will be deployed in the agricultural land and it automatically senses the soil and

send notification about the soil series so the suitable measures can be taken. The sensor mainly senses the minerals contained in the particular 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. Diepeveen & 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.

AUTHORS PROFILE



Saranya N done her Masters in Computer science and Engineering at Sri Shakthi Institute of engineering and Technology. Currently doing her doctorate and working as an Assistant Professor in same college. She has a working experience around 8 years. Her area of interests are operating systems, Database management systems, Data mining algorithms and big data.



Mythili A is a citizen of India. She is a student of Master's in Computer Science and Engineering at the department of Computer science and Engineering at Sri Shakthi Institute of Engineering and Technology Tamil Nadu India. Her area of interests are machine learning, Big data and data mining algorithms.