EAI Endorsed Transactions
on Energy Web

Research Article EAI.EU

Classification and Analysis of Soil Types using Bayesian Models for the Crop Agronomy

A. Zakiuddin Ahmed[1] , Dr. T. Abdul Razak[2]

[1]*Research Scholar, Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University Tiruchirappalli.*
[2]*Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli.*

Abstract

The core objective of this research paper is to classify soil of a particular region and also perform the scientific study for the prediction of suitable crops, which will yield more profit to the agronomist. The production of agriculture is mainly depends on soil type, cultivation seasons (climate), irrigation method (such as surface, sprinkler, drip/trickle, subsurface etc) and fertilizers (to increase the crop productivity). Soil is classified based on its physical properties (color, texture, structure, porosity, density etc) and chemical properties (phosphorous, nitrogen, carbon, calcium, magnesium, sodium, pH, potassium, sulfur etc). In this research paper soil types are classified by applying Bayesian models to decision tree algorithm and performed various analysis to verify that soil types are correctly classified and as wells to predict suitable crop cultivation during Kharif (monsoon), Rabi (winter) and Zaid (summer) seasons along with suitable irrigation methods and proper use of fertilizers to increase the crop productivity. The proposed algorithm offers unique features by adopting other tree inducers which produce optimum results even though the conditions are leads to chaotic behavior. The final results obtained are presented which illustrates optimum results and generates best decision tree for classification of soil type from the soil dataset. The Bayesian approach guarantees that the classifications of soil types are more accurate than the existing Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Decision Tree (DT) algorithms.

Keywords: Soil types, Decision tree, Machine learning, Bayesian models, agronomy

Received on 19 February 2023, accepted on 25 October 2023, published on 30 October 2023

doi: 10.4108/ew.4271

*Corresponding author Email: azaki2020@gmail.com

I . Introduction

In this research paper, soil dataset of a particular region is downloaded , machine learning techniques are used with our proposed model for the classification of soil types. Further scientific studies are performed to predict the suitable crop cultivation when the appropriate irrigation methods and fertilizers are utilized. Here we do the agricultural process by adopting scientific techniques for the production of most suitable crops during Kharif, Rabi, Zaid seasons along with exact fertilizer using the classified soil types.

In general constraints based regulations, boundaries, mathematical and other functions may define the classifications. The relationship between a known class constraints based regulations, boundaries and the attributes are to be classified [5]. The classification is

unsupervised then clustering, association and reduction of dimension are the major tasks of classification technique [2].

SVMs (Support Vector Machine) were incorporated into the sampling technique to broaden the coverage (large samples) of the realizable design space and create realizability limits. The black-box simulator's successful or unsuccessful convergence indicates whether a certain processedclass to be classified was feasible [7]. This sample values are called as initial values. Further it is demonstrated with the Bayesian classifier and this classifier is the best for minimizing the classification error probability [6]. The Bayesian classification, closest neighbour classifiers,Naive Bayes classifiers are holding the major contriution in the reccent years. This paperutilized Bayesian probabilistic principles to address the soil classification problem with the intention of reducing the likelihood of the classification error.

Using probability the uncertainty about the classiciation, Bayesian classification is the optimal probabilistic method to learn and infer. These things areanalyzed basedon the different understanding of what to learn as well as what to classify from the existing data. Our assumptions about the relationship between the existing classified datas are analyzed through probability distribution. After reviewing the facts, a posterior distribution captures our updated classified datas. With the use of observed data and Bayesian reasoning, it is possible to create probability distributions to arrive the optimality [11]. Machine learning brings together the ideas of computer science and statistics to couple that predictive power. One such widely used model in machine learning is Bayesian model. This model is used to decide the progress of likelihood of a state space.

KNN is a straightforward classifier used to solve many classification issues. There is no learning carried out during the training period. The testing step makes use of the training data. When a unknown class is evaluated, the KNN method determines its K nearest neighbours and distributes the class among them. Only K and the distance metric are used in the KNN algorithm to finetune and achieve high classification accuracy[10].

Decision Tree (DT) is the most powerful predictive model that divides the rooted directed tree recursively into subspaces. The internal nodes in decision trees have edges [12]. A series of hierarchical classification and precedence features (enters into the next level) are used by decision trees.

According to these we constructed a mathematical model to analyze the differences of soil data sets and cultivation of crops. We analyzed a large samples related to affirmative and reverse decisions and the differences of opinions among the cases through Bayesian approach (to identify the irrelevant portion, and produces inspectable models). This method is used to better prediction performance and classification of a sample data and training set data. Parametric and non - parametric tests demonstrated the distinct prediction of data.

II Literature Review

Yethiraj NG et al., [13] as described the significant techniques by using various algorithms and techniques such as, KNN, ID3, the K-means, Artificial Neural Network to classifies the data.

Vamanan R et al., [9] explained Data mining classification techniques which are applied in classification of soil in a particular region database. Barghavi.P et al., [4], assessed the soil data.

Anju Rathee et al [3] discussed the performance of Classification and Regression Tree (CART). Among all the three algorithms they concluded that C4.5 clustering algorithm is the best algorithm.

Amir Ahmad et al [1] exhibited a method called random projection and applied these random projection and random discretization in the soil data set. Further they proposed Random Discretization method to analyze random discredited features. It has been realized that Random Projection and Random Discretization Ensembles are used in robust analysis of the noisy data related to lossy and lossless compression. V. Rajeswari et al [8] examined various classification algorithms.

III Proposed Bayesian model

The proposed algorithm for classification of Soil type using Bayesian approach to Decision Tree is designed and simulated through python language. In general various sorting techniques are used to split the input dataset through n number of partitions. Decision Tree is referred to as a predictive modeling technique from the subfield of machine learning.

In this section we considered n different soils while choosing each multinomial variables admits only two possibilities like, there is differences of opinion in cultivation of crops , its specification , performance and there is no differences of opinion in its specification, performance and cultivation of crops.

Bayesian Model

Applying Bayes rule to the unknown variables of a data modeling problem is called Bayesian modeling.

Bayesian statistics is used for organizing an optional method for developing predictive connections between climate and soil types. Mapping soil properties within this probability model is a method by which values of a given soil resources at each location are extracted from a set of specified soil resources groups.

Deduction or induction inference methods used by Machine Learning algorithm can make the computer to be trained from existing data or theories. To provide as training data, only small amount of samples are needed to form knowledge base by Machine Learning algorithm. It is easier than the direct extraction from the higher expert. This algorithm derives the rules and measures to connect the two parts and predictable output is related with set of examples. In this analysis part, the decision tree and Bayesian modeling are used and it needs training to retrieve the information. Usually soil sampling is measured as proficient technique to get such kind of results. Without any personal constraints and non-financial background, making an effort to collecting samples, are easier said than done.

Naive Bayesian algorithm provides the way of calculating posteriori probability say $P(c/x)$ and this is calculated using $P(c), P(x)$ and $P(x/c)$ and is denoted as

$$P(c/x) = \frac{P(x/c)P(c)}{P(x)} \quad (1)$$

Where

$P(c/x)$ is the posteriori probability

$P(c)$ is the priori probability of class

$P(x)$ is the priori probability o predictor class and

$P(x/c)$ is the likelihood of predictor probability

Further Naive Bayesian classifies the prediction as follows

$$P(c/x) = \frac{n_c+mp}{n+m} \quad (2)$$

Where

$n =$ number of training data set

$m =$ sample size

$p =$ priori estimate of the probability and

$n_c =$ number of instances with the attributes (related to class 0 to class 5)

The optimum prediction is arrived through Naive Bayesian classifier

$$P(optimum\ class) = log_2 P(c/x) - log_2 P(c) \quad (3)$$

The proposed method (Bayesian approach to Decision Tree)

- Get the soil dataset (from Kaggle) input and import it into the proposed algorithm.
- Process the soil dataset
- Extract its features
- Define training datasets and test data sets.
- Form leaf and branch required for an Decision Tree
- Apply the Bayesian Model with the Decision Tree and this form the final Decision Tree.

In this research paper initially, 80% training data set sample is given and trained for the prediction and the remaining 20% testing data are utilized Bayesian and decision tree model using machine learning technique. Bayesian approach to decision tree is given below:

Step1: The 5200 dataset is given as input to our proposed model.

Step2: Extraction processes reduce the dimensions by the Machine learning (ML) techniques. The given dataset is divided into the Training Dataset 4160 (80%) and Test Dataset 1040(20%).

Step3: Using the above Soil Dataset, our proposed decision tree model is constructed by implementing the Bayesian approaches.

Step 4: Finally, soil type is predicated, which we can be realized with the help of parameters like precision, recall, F1-Score and accuracy.

Further we proposed the following algorithm to adopt and recognize the pattern of differences between the various types of soils

Step 1: Starts with the initial probability and estimates the model.

Step 2: Taking expectations into account, classify the transformation with the model parameters.

Step 3: Note down the changes occurred.

Step 4: Re-estimate the possibilities iteratively until convergence.

Step 5: If there is no affirmative change in the model then leave, else recognize the changes pattern and analyze its hidden origin.

IV Results and Discussion

The proposed algorithm Bayesian approach is given in the Table- 1 given below. We have taken 5200 datasets for this paper. This 5200 dataset which is divided into 80% Training set (4160) and 20% Test datasets (1040). The proposed algorithm is used the

measures such as Recall, F-measures, Precision and accuracy. These four metrics are necessary to mark the judgment of the classification task.

Class classification	Precision	Recall	F1-score	Support
Class 0	**0.85**	**0.86**	**0.86**	**562**
Class 1	0.00	0.00	0.00	**0**
Class 2	0.00	0.00	0.00	**0**
Class 3	**0.84**	**0.80**	**0.82**	**470**
Class 4	0.00	0.00	0.00	**0**
Class 5	0.00	0.00	0.00	**0**
Micro average	0.83	0.83	0.83	**1040**
Macro average	0.58	0.62	0.48	**1040**
Weighted average	0.85	0.83	0.84	**1040**
Accuracy	**0.840508**			

Table1: Performance analysis of Proposed algorithm

In the above Table 1, Class 0 denotes Red Soil (4 sub groups), Class 1 denotes Black Soil (3 sub groups), Class 2 denotes Sandy Soil (2 sub groups), Class 3 denotes Laterite Soil, Class 4 denotes Red loamy/clay soil and Class 5 denotes Alluvial Soil (2 sub groups as per ICAR 1980). Table 2 represents Suitable Crops against the Soil types. Further table3 represents the comparisons between the existing algorithms and proposed algorithm.

Agricultural production (crop cultivation) of the Karnataka state is depends on three different season: 1.Kharif (July to October) , 2. Rabi (October to March) and 3. Zaid (summer).

Soil Type	Season	Suitable Crop
Class 0 (Red Soil)	**Kharif (monsoon)**	Rice, maize, cotton, black gram, ground net
	Rabi (winter)	Wheat, bajra, rice, oil seeds (Rapeseed oil and mustered oil)
	Zaid (summer)	Tomato, watermelon, musk melon, cucumber, etc.,
Class 3 (Laterite Soil)	**Kharif (monsoon)**	Rice, pulse (dry beans, dry peas, Faba Beans, Lentils etc), cotton, rice, tea, coffee.
	Rabi (winter)	Wheat, tea, coffee, pepper etc
	Zaid (summer)	Coconuts, pulses, vegetables.

Table- 2t Crops agai st the Soil types

The above shc wn s crops can be cultivated during the Kharif (monsoon), Rabi (v inter) , Zaid (summer) seasons by ad pting suitable irrigation method (either surface, sprinkler, drip/tricle or subsurface) and use the ertilizers to ir crease the crop productivity. Here if th soil is Red Soil and the season is Kharif then suggested crops are rice, maize, cotton, black gram, groudnet ; and if the sea son is Rabi then suggested crops are wheat, bajra, c ilseeds and in Zaid the suggested crop are tomato, watermelon, cucumber and so on.

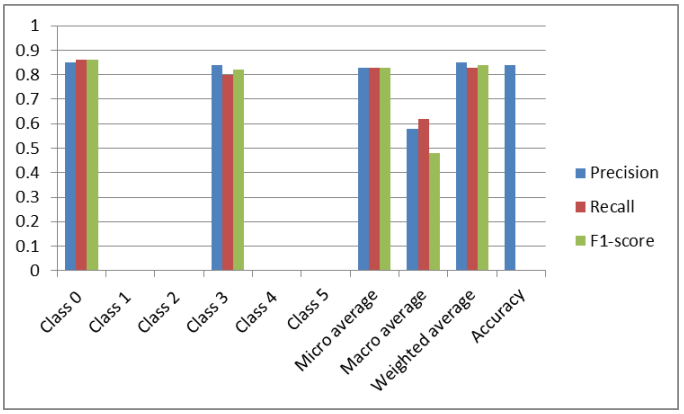

Figure 1: Precision, Recall and F1 score outlook of all classes including averages and accuracy of the proposed algorithm

Algorithms	Precision	Recall	F1-Score	Accuracy
KNN	0.79	0.78	0.79	0.80
SVM	0.80	0.81	0.83	0.81
Decision Tree	0.82	0.83	0.84	0.82
Proposed Algorithm	0.86	0.85	0.85	0.84

Table 3: Comparisons between the existing algorithms and proposed algorithm

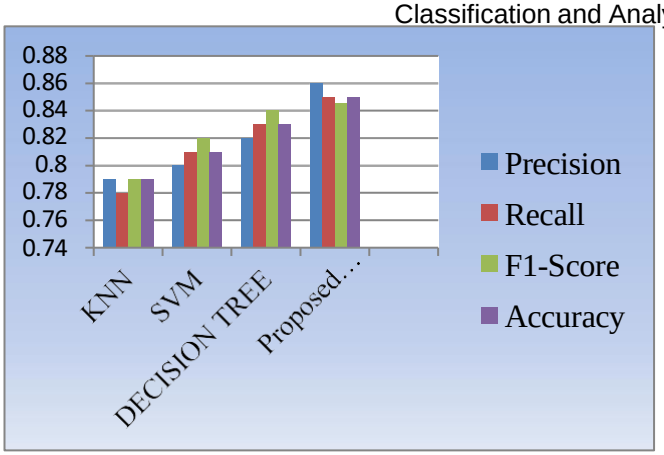

Figure 2 : Comparisons between the existing algorithms and proposed algorithm

Figure 2 compare the outputs of the existing algorithms and Bayesian approach to Decision Tree - proposed algorithm. In addition to that figure 3 represents the Precision, Recall and F1 score outlook of all classes including averages and accuracy of the Bayesian approach to Decision Tree - proposed algorithm. Figure 4 classifies the Class 0 and Class 3 - Soil map of Karnataka State.

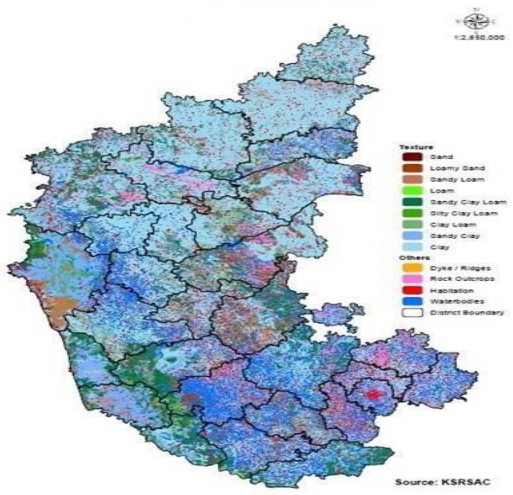

Figure 3: Class 0 and Class 3 - Soil map of Karnataka state

Element	Normal Rate ppm (Optimum Temperature)
Phosphorus	20 – 30
Sulfur	5 – 10
Zinc	1 – 2
Iron	7– 16
Manganese	5 – 10
Copper	0.8 – 0.9
Potassium	150 – 200
Calcium	1300 and above
Magnesium	100 or higher
Sodium	70 – 110

Table 4: Class 0 and Class 3 soil classification

The numerical values shown in the blow tables, table 5 and table 6 are estimated using the equations (1) and (2).

Class	Cultivation of crops		P(Cultivation of crops)
	Yes	No	
Class 0	4	0	0.31
Class 1	0	3	0.115
Class 2	0	2	0.15
Class 3	1	0	0.28
Class 4	0	1	0.08
Class 5	0	2	0.15

Table 5 : Soil data set for cultivation of crops

The Table 5 clearly shows that the proposed algorithm predicts only two soil types class 0 (Red Soil) and Class3 (Laterite Soil) and these two soil types are most suitable for cultivate the crops shown in the above table 2.

Class outlook	Cultivation of crops	
	P(Cultivation of crops) for Yes	P(Cultivation of crops) for No
Class 0	**0.845**	0.155
Class 1	0.115	0.885
Class 2	0.075	0.925
Class 3	**0.96**	0.04
Class 4	0.04	0.96
Class 5	0.075	0.925

Table 6: Soil data set likelihood test table

Above Table 6 also proves that class 0 and class 3 are the soil types found in the Karnataka state which are the most suitable for cultivation of crops shown in table 2

V Conclusion

Classification of Soil types for the given datasets of a region with more parameters leads to achieve the classification tasks successfully and accurately. Bayesian methods are adapted to perform classification of Soil type by making use of machine

learning techniques to progress the accuracy and efficiency of the input data. In this research paper, analysis of soil types by using the Proposed Bayesian approach to Decision Tree Algorithm has been compared and concluded that the proposed Algorithm gives better result than the other three existing algorithms. Also predicted the suitable crop can cultivated using each soil type during three different seasons Kharif (monsoon crops), Rabi (winter crops) and Zaid (summer crops) when the proper irrigation method and fertilizer are used. The result obtained using this algorithm can be further improved by implementing ensemble learning techniques and multilayer neural network methods with decision tree algorithms.

References

[1] Amir Ahmad and Gavin Brown "Random Projection Random Discretization Ensembles—Ensembles of Linear Multivariate Decision Trees", IEEE Transactions on Knowledge and data Engineering vol.26, No.5, pp.1225-1239May2014, ISSN:1041 – 4347.

[2] D. Anantha Reddy, B. Dadore, and A. Watekar, Crop recommendation system to maximize crop yield in Ramtek region using machine learning, International Journal of Scientific Research in Science and Technology 6 (2019), pp. 485–489. doi:10.32628/IJSRST196172.

[3] Anju Rathee, Robin Prakash Mathur, "Survey on Decision Tree Classification Algorithms for the Evaluation of Student Performance", International Journal of Computers and Technology (IJCST), Vol. 2 No. 4, pp.244-247, April 2014, ISSN:2277-3061.

[4] P.Barghavi and S.Jyothi, "Applying Naïve Bayes Data Mining Technique for Classification of Agricultural Land Soils," International Journal of Computer Science and Network Security, Vol.9, No.8, pp.117-122, August2013.

[5] R. Jahan, Applying Naive Bayes classification technique for classification of improved agricultural land soils, International Journal for Research in Applied Science & EngineeringTechnology (IJRASET) 6 (2018), pp. 189–193. doi:10.22214/ijraset.2018.5030.

[6] Nadimani G.V, Hemalatha. M, "An Evaluation of Clustering Technique over Intrusion Detection System", International Conference on Advances in Computing, Communication and Informatics (ICACCI), pp. 1054-1060, August 2012.

[7] Rahman S.A.Z., Mitra, K.C., Islam, S.M.M, Soil classification using machine learning methods and crop suggestion based on soil series. In: International Conference of Science and Technology Computer (2018)

[8] Rajeswari.V and Arunesh P.K, "Analysis of Soil data using data mining Classification Techniques", Indian Journal of Science and Technology, Vol.9(19), pp.1-4, May2016, ISSN:0974-6846

[9] R. Vamanan and K. Ramar, "Classification of agricultural land soils a data mining approach," International Journal of Computer Science and Engineering, Vol.3, No.1,pp.379-384,January2016,ISSN:0975-3397.

[10] Train, K. E. Discrete Choice Models with Simulation. Cambridge University Press, Cambridge, 2009.

[11] Vogel, S.; Ney, H.; Tillmann C, HMM-Based Word Alignment in Statistical Translation. Available online: http://aclweb.org/anthology/C96-2141

[12] Yoon, B.J. Hidden Markov Models and their Applications in Biological Sequence Analysis. Curr. Genom, 2009, 10, 402–415.

[13] N.G.Yethiraj, "Applying Data Mining Techniques in the field of agriculture and allied sciences", International Journal of Business Intelligent, Vol.01, Issue2, pp72-76, December 2012, ISSN 2278 – 2400.