https://doi.org/10.56572/gjoee.2022.34.2.0011 Gujarat Journal of Extension Education Vol. 34 : Issue 2 : December 22

ANALYZING SOIL FERTILITY USING DATA MINING TECHNIQUES

R. S. Parmar[1], V. I. Mehra[2] and G. J. Kamani[3]

1 Professor, College of Agricultural Information Technology, AAU, Anand -388110
2 & 3 Asst. Prof., College of Agricultural Information Technology, AAU, Anand -388110
Email: rsparmar@aau.in

ABSTRACT

Data mining classification techniques play a significant role in agriculture. To increase yield production, many parameters are responsible like weather, fertilizers, pesticides, and soil. Soil is an important key aspect of agriculture, as an increase in yield can only be attained by ensuring that the soil provides a balanced and adequate supply of nutrients. The main aim of the work is to predict soil fertility using data mining classification techniques. Soil fertility is predicted using data mining classification techniques such as J48, RandomForest, DecisionTable, PART, and Naive Bayes. These classifier algorithms are used to extract soil fertility knowledge from soil data. A comparison of different classifier algorithms concerning prediction accuracy revealed that the RandomForest is the best classifier to classify the soil fertility dataset. The RandomForest classifier algorithm can produce more consistent results of soil data for the Anand District of Gujarat State and the Kappa Statistics in the prediction were improved.

Keywords: *data mining, soil fertility, yield production, classification, agriculture*

INTRODUCTION

Indian economic system directly depends on agricultural production. Agriculture is entirely dependent on soil quality but as time passes more and more agricultural production results in the loss of nutrients available in the soil. The need of recognizing techniques that will slow down the elimination of nutrients and also return the required nutrients in the soil, so that farmers keep receiving high quality and good quantity crop productions. Data mining techniques are likely to continue as a major driver of agriculture development in India. Data mining in agriculture is an emerging research topic. Data mining is a process of discovering previously unknown patterns used for strategic decision-making. There are various steps involved in the mining process such as Data Integration, Data Selection, Data Cleaning, Data Transformation, Data Mining, Pattern Evaluation, and Knowledge Presentation and Use of Discovered Knowledge. Rajeswari and Arunesh (2016) made a comparative analysis of three algorithms like Naïve Bayes, JRip, and J48. JRip classification algorithm gives better results for the experimental dataset and has correctly classified the maximum number of instances as compared to two others. JRip is recommended to predict soil types. Chiranjeevi and Ranjana (2018) conducted a comparative analysis of two algorithms like Naive Bayes and J48. Naive Bayes gives better results for the experimental dataset and correctly classified the maximum number of instances compared to J48. Naive Bayes Algorithm can be recommended to predict crop growth in a particular soil sample. Raunak (2018) studied three algorithms namely Naive Bayes, zeroR, and stacking. Naive Bayes classification algorithm gives a superior effect on the experimental dataset and correctly classified the maximum number of instances compared to the remaining classifiers. Naive Bayes can be suggested to predict soil types. Kalekar *et al.* (2018) studied the J48 decision tree algorithm and showed an accuracy of 87.5% in classifying the soil according to its fertility which can be used to recommend appropriate fertilizers. Bhuyar (2014) studied different classifiers algorithms to predict fertility rates. The study indicates that J48 classifier performs better to predict the fertility index. Observation also shows that the fertility rate for Aurangabad district is medium. This will help the decision-maker recommend fertilizers accordingly. Elhamayed (2016) studied algorithms like J48, Decision Table, PART, Naive Bayes, and. By analyzing the overall experiment results of the production dataset, it is concluded that J48 algorithm has produced the best classification performance than IKB and NaïveBayes and has produced a slight difference in performance with Decision Table, PART classifiers. Dolas and Joshi (2018) studied classifying the soil using the CART algorithm helps to know the overall health of the soil and the content of nutrients in the soil. Jethva *et al.*(2018) made a comparative analysis of algorithms like J48, Naïve Bayes, JRip, and K-Means classifier algorithms and concluded that the decision tree algorithm performs the best to analyze soil fertility conditions. Baskar *et al.* (2013) conducted a comparative study for soil classification of naïve bayes, JRip and J48. They found J48 to be the best method

for soil classification. Prajapati *et al*. (2019) proposed three approaches to a nonparametric classifier: fast k-NN, training set reduction techniques, and a hybrid approach. It was observed that the suggested techniques are better than the existing K-nearest neighbors *(*KNN) classification technique. Data mining techniques are useful in abstracting patterns and establishing relationships between varied data sets and predicting reasonable outputs. In the current scope of the study, we have developed a model for the fertility of soil based on different soil types. After receiving fertility depending upon various soil types, a comparative study of Data mining techniques is carried out. The present research will be useful for researchers to know data mining techniques can be efficiently applied in the Agriculture industry to improve research (Kamani *et al*. 2019, Kamani *et al*. 2021, Parmar *et al*. 2022).

OBJECTIVE

To investigate J48, RandomForest, DecisionTable, PART, and Naive Bayes classifier algorithms and compare different fitted classifier algorithms.

METHODOLOGY

In this research, soil fertility datasets were collected from Laboratory for Micronutrients Soil and Water Testing, Anand Agricultural University, Anand, Gujarat. The Dataset consists of soil fertility data from 34 villages in the Anand district of Gujarat. The Dataset has7 attributes namely the value of soil (pH), Electric conductivity(EC), Organic Carbon (OC), Nitrogen (N), Phosphorous (P_2O_5), Potassium (K_2O), Fertility Index (FI). The soil Dataset consists of a total of 204 instances. The fertility index is the class label which is categorized as L- Low, M- Medium, H- High, A-Acidic, N- Neutral, and A-Alkaline. Data mining tool WEKA version 3.8.1 was used for this research. This dataset was prepared in an Excel sheet with .CSV extension. Normalization is a scaling technique or a pre-processing stage. Where we can find a new range from an existing one range. Min-Mix Normalization technique was used to normalize (shown in eq.1)the experimental dataset. As per the Min-Max normalization technique,

$$v' = \frac{v - \min_A}{\max_A - \min_A}(new\_\max_A - new\_\min_A) + new\_\min_A \quad (1)$$

Where v is the respective value of the attribute

v' is Min-Max Normalized data one

$\min_A$ is the respective Minimum of the value of the attribute

$\max_A$ is the respective Maximum value of the attribute

Classification is a data mining technique that forms the common stepping stone for various recursive algorithms and methods designed for data mining. The classification method makes use of statistical techniques like decision trees, linear programming, neural network, statistic-oriented approach, etc. The following section describes classification algorithms such as J48 tree classifier, RandomForest tree classifier, Decision Table rule classifier, PART rule classifier, and Naive Bayes Bayesian classifier.

J48 Classifier

The C4.5 algorithm for building decision trees is implemented in Weka as a classifier called J48. Classifiers, like filters, are organized in a hierarchy: J48 has the full name weka.classifiers.trees.J48. The classifier with default parameters is read as J48 –C 0.25 –M 2.

Random forest classifier

Random Forest is an improvement over bagged decision trees that disrupts the greedy splitting algorithm during tree creation so that split points can only be selected from a random subset of the input attributes. This simple change can have a big effect on decreasing the similarity between the bagged trees and in turn the resulting predictions. Click the "Choose" button and select "RandomForest" under the "trees" group.

Decision table classifier

A Decision Table is an accurate method for numeric prediction from decision trees and it is an ordered set of If-Then rules that have the potential to be more compact and therefore more understandable than the decision trees. The reason to explore the decision table is its simplicity, and less compute-intensive algorithm than the decision-tree-based approach. The algorithm is found in the Weka classifiers under Rules.

PART classifier

Part technique comes under the rules classification. It obtains the rules from a partial tree built using the J4.8 classifier technique. Therefore most of the time part and J4.8 gives the same result for the classification of given data. The algorithm, PART is found in the Weka classifiers under Rules.

Naive Bayes classifier

A Naive Bayes classifier is one of the classifiers in a family of simple probabilistic classification techniques in machine learning. It is based on the Bayes theorem with independence features. Each class label is estimated through

the probability of a given instance. It needs only a small amount of training data to predict the class label necessary for classification. The algorithm, Naive Bayes is found in the Weka classifiers under Bayes.

RESULTS AND DISCUSSION

The experiment was performed on the Weka tool kit. The results are shown with 10-fold cross-validation to avoid overlapping. From the Weka tool kit, five classifiers are studied namely tree-based J48 and RandomForest, rules-based DecisionTable, and PART and Bayesian-based Naive Bayes. The performance of each classifier is examined in terms of correctly classified instances, incorrectly classified instances, Kappa statistic, mean absolute error, root mean squared error, and root relative squared error.

Table 1: Performance error of different classifier algorithms

Performance error	Different Classifier Algorithms				
	Tree based		Rules-based		Bayesian based
	J48	RandomForest	DecisionTable	PART	NaiveBayes
Kappa statistic	0.84	0.87	0.84	0.84	0.52
Mean absolute error	0.01	0.01	0.04	0.01	0.03
Root mean squared error	0.13	0.09	0.13	0.10	0.15
Root relative squared error	56.92 %	49.37 %	71.62 %	56.60 %	80.96 %

The results presented in Table 1 indicated that the tree-based and rule-based classifiers have better performance than the Bayesian-based classifier. In tree-based classifiers, two algorithms are examined namely J48 and RandomForest. RandomForest has a better performance than J48. In rules-based classifiers, two algorithms are studied namely DecisionTable and PART. PART has better performance than Decision Table. In general, it could be observed that random forest is better than PART, which revealed that RandomForest is the best classifier to classify the soil fertility dataset. Fig. 1 shows that J48, RandomForest, DecisionTable, PART, and Naive Bayes are correctly classified with 86.70 %, 89.66 %, 87.19%, 86.21, and 62.07 respectively.

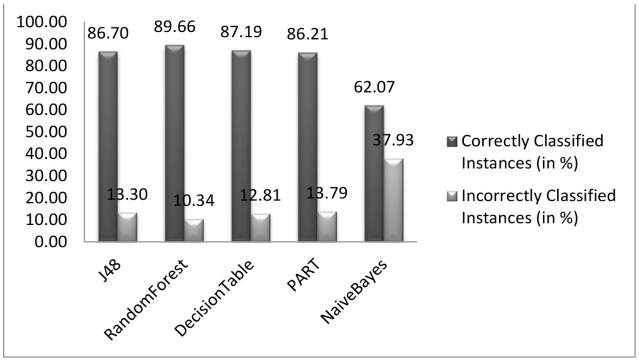

Fig. 1: Accuracy of Different Classifier Algorithms

Fig. 2 depicted that the weighted average of True Positive Rate were 0.867, 0.897, 0.872, 0.862, and 0.621 for J48, RandomForest, DecisionTable, PART, and Naive Bayes respectively. Based on the training soil data set it is concluded that the tree-based RandomForest classifier classified the data set in an important sense.

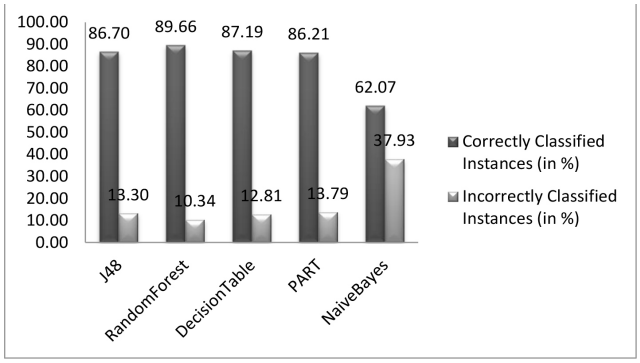

Fig. 2: Weighted Average of True Positive Rate

CONCLUSION

Data mining techniques, which are applied to find out suitable information from the huge dataset, can be utilized to discover useful knowledge from the database such as soil fertility data.

This study reveals an analysis of data mining classification techniques used in soil data and also compares the J48, RandomForest, DecisionTable, PART, and Naive Bayes classifier algorithms. The tree-based RandomForest algorithm performs best to analyze soil fertility conditions. This study helps the researcher for getting an efficient algorithm selection for the fertilizer recommended based on soil fertility data.

CONFLICT OF INTEREST

No conflict of interest among researchers.

REFERENCES

Baskar, S.S.; Arockiam, L. and Charles, S. (2013). Applying Data Mining Techniques on Soil FertilityPrediction.

International Journal of Computer Applications Technology and Research., 2(6):660-662

Bhuyar, V. (2014). Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District. *International Journal of Emerging Trends & Technology in Computer Science.,* 3(2) : 200-203

Chiranjeevi, M.N. and Ranjana, B.N. (2018). Analysis of Soil Nutrients using Data Mining Techniques. *International Journal of Recent Trends in Engineering & Research.,* 4(7) :103-107

Dolas, V.M. and Joshi, U. (2018). A Novel Approach for Classification of Soil and Crop Prediction. *International Journal of Computer Science and Mobile Computing.*, 7 (3): 20-24

Elhamayed, S.A. (2016). Enhancement of Agriculture Classification by Using DifferentClassification Systems. *International Journal of Computer Applications*.,3(1) :08-12

Jethva, J. M.; Gondaliya, N. and Shah, V. (2018).A Review on Data Mining Techniques for Fertilizer Recommendation. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology.,* 3(1): 2456-3307

Kalekar, A. ; Vispute, S.; Kokane, P.; Kamble, M. and Bokefode, K. (2018). Automated Generation and Analysis of Soil Health Card and Calculation of the Village Soil Fertility Index. *International Journal of New Technologies in Science and Engineering.*, 5(3): 118-125

Kamani G. J., Parmar R. S., Ghodasara Y. R. (2019), Data normalization in data mining using graphical user interface: a pre-processing stage, *Guj. J. Ext. Edu,* 30(2): 106-109

Kamani G. J., Parmar R. S., Ghodasara Y. R. (2021), Machine learning models for productivity trend of Rice crop, *Guj. J. Ext. Edu.,* 32(2): 189-196

Parmar R.S. , Vegad N.M, Mehra V.I. (2022). An attitude of PG scholars of agricultural extension towards Application of mobile technology using artificial intelligence technique, *Guj. J. Ext. Edu,* 33(1): 136-143

Prajapati, B.P. and Kathiriya, D.R., (2019). A Hybrid Machine Learning Technique for Fusing Fast k-NN and Training Set Reduction: Combining Both Improves the Effectiveness of Classification. In Progress in Advanced Computing and Intelligent Engineering. Springer, Singapore. 174: 229-240.

Rajeswari, V. and Arunesh, K. (2016).Analyzing Soil Data using Data Mining Classification Techniques. *Indian Journal of Science and Technology.,* 9(19) : 1- 4

Raunak, J. (2018). Applying Naive Bayes Classification Technique for Classification of Improved Agricultural Land soils. *International Journal for Research in Applied Science & Engineering Technology*., 6(5): 189-193

Received : September 2022 : Accepted : December 2022