ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 7, July 2015

An Enhanced Technique to Predict the Accuracy of Soil Fertility in Agricultural Mining

Dr.S.Hari Ganesh[1], Mrs. Jayasudha[2]

Assistant Professor, Dept. of Computer Science, Bishop Heber College (Autonomous), Tiruchirapalli, India[1]

M.Phil. Scholar, Dept. of Computer Science, Bishop Heber College (Autonomous), Tiruchirapalli, India[2]

Abstract: The techniques of data mining are extremely popular in the area of agriculture. Today, data processing is employed in a very large areas and plenty of ready-to-wear data processing system product and domain specific data processing application software's are obtainable, however data processing in agricultural soil knowledge sets may be a comparatively a young analysis field. In this paper we offer internet base answer for the soil testing laboratories yet as free messages for the farmer that contains data like soil testing code, chemical that is important for the crop and additionally the knowledgeable recommendation. Additionally farmers specify there next crop whereas they furnish their sample to scantiest therefore in keeping with next crop the chemical can recommend. The result's supported the classification of contains that should be gif tin soil and in keeping with result report are generated.

Keywords: Data Mining, Soil Testing, classification.

I. INTRODUCTION

Data mining is that the computational method of discovering patterns in massive data sets involving ways at the intersection of computer science, info systems. The goal of the data mining method is to extract information from a data set and rework it into an obvious structure for more use. The particular data processing task is that the automatic or semi automatic analysis of huge quantities of data to extract previouslyun known interesting patterns like teams of data records (cluster analysis), uncommon records (anomaly detection) and dependencies (association rule mining).A soil take a look at is that the analysis of a soil sample to see nutrient content, composition and alternative characteristics. Tests area unit typically performed to measure fertility and indicate deficiencies that require to be remedied. The soil testing laboratories area unit given appropriate technical literature on numerous aspects of soil testing, together with testing methods and formulations of plant food recommendations. The outcome of this research can result into substantial diminution within the value of those tests, which can save lots of efforts and time of Indian soil testing laboratories.

II. RESEARCH METHODOLOGY

A. Dataset collection

Data set required for this analysis. These datasets contain varied attributes and their many values of soil samples taken from literature review. Dataset has ten attributes and a complete 1988 instances of soil samples. Table one shows attribute description. The dataset has 9 attributes.

Table 1: Attribute & Description

Attribute	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

B. Automatic System

Classification of soil system is prime demand for identification of soil properties. Specialist system will be an awfully potent tool in identification of soils quickly and accurately. The normal approach of classification systems was tables, flow-charts. This type of arrangement (manual system) takes plenty time, and therefore it's necessity of automatic system as fast, reliable system for higher utilization of technician's time. We suggest an automatic system that has been developed for classifying soils supported fertility. Being rule-based system, it depends on facts, concepts, and theories that area unit needed for the implementation of this method. Rules for soil classification were collected from soil testing laboratory. The soil sample instances

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 7, July 2015

were categoryfield into the fertility class labels as: terribly High, High, Moderately High, Moderate, Low, and extremely Low. These category labels for soil samples were obtained with the assistance of this auto matic system and that they are used extra for proportional study of classification algorithms.

III. COMPARATIVE STUDY OF SOIL CLASSIFICATION

Soil classification was measured serious to review because of relying upon the fertility class of the soil domain information consultants determines that crops ought to be taken on it specific soil and that fertilizers ought to be used for an equivalent. The subsequent section describes Naive bayes, J48, JRip algorithms in short.

A. Naive Bayes

A Naive Bayes classifier may be a straight forward probabilistic classifier supported applying Bayes theorem with strong independence assumptions. Looking on the precise nature of the likelihood model, naive Bayes classifiers are often trained terribly with efficiency in an exceedingly supervised learning setting. A plus of the naive Bayes classifier is that it solely needs a tiny low quantity of training data to estimate the parameters necessary for classification.

B. J48 (C4.5)

J48 is an open source Java implementation of the C4.5 alg orithm in the Wake data processingtool.C4.5 could be a program that makes a choice tree supported a collection of labelled input file. This decision tree will then be checked against unseen labelled test data to quantify however well it generalizes. This algorithm was developed by Ross Quinlan. It's an extension of Quinlan's earlier ID 3algorithm. C4.5 uses ID3 algorithm that accounts for continuous attribute value ranges, pruning of decision trees, rule derivation, and soon. The decision trees generated by C4.5 are often used for classification, and for this reason, C4.5 is usually mentioned as a statistical classifier.

C. JRip

This algorithm implements a propositional rule learner, continual progressive Pruning to produce Error Reduction (RIPPER), that was proposed by William W. Cohen as an optimized version of IREP. In this paper, 3 classification techniques (naïve bayes, J48 (C4.5) and JRip) in data processing were evaluated and compared on basis of time, accuracy, Error Rate, True Positive Rate and False Positive Rate. Multiple cross-validations were utilized in the experiment. Our studies showed that J48 (C4.5) model turned out to be the most effective classifier for soil samples.

Table 2: Comparisons of Regression Algorithms

Algorithm	Linear Regression	Least Median Square Regression
Time taken to build the model	0.17s	10.84s
Relative Absolute Error	10.76%	10.07%
Correlation Coefficient	0.9511	0.9865

IV. PREDICTION OF UNTESTED ATTRIBUTES

Algorithms almost like regression; Least Median square, simple regression completely different attributes were expected. owing to these outcome the values of phosphorous attribute was found to be most in truth predicted and it depends on slightest number of attrib utes. Once all attributes area unit numeric, regression coul d be a natural and easy technique to think about for numeric prediction; however it suffers from disadvantage of dimensionality. If data exhibits non-linear dependency, it's going to not offer good results. In this case, least median square technique is employed. Median regression techniques incur high process value which regularly makes them impossible for practical issues. Several regression tests were applied mistreatment weka data processing tool to predict untested numeric attributes. Linear-Regression take a look at for predicting phosphor gave the simplest and correct results. These predictions are often used to determine phosphor content while not taking ancient chemical tests in soil testing labs, and this may eventually save plenty of time. Applied math results of those tests area unit given in Table 2.There was terribly restricted variations amongst the predicted values of phosphor attribute. Though the smallest amount Median of Squares algorithms are understood to provide higher results, we have a tendency to observe that the accuracy of regression as shown in Table 3 was comparatively reminiscent of that of least median of squares algorithm.

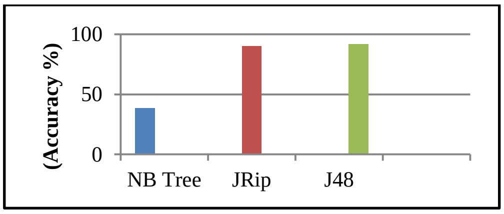

Figure 1: Comparison among NB, JRip and J48algorithms

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 7, July 2015

Table 3: Contrast of classifiers

Classifier	NB Tree	JRip	J 48
Correctly Classified Instances	822	1898	1988
Incorrectly Classified Instances	1265	201	149
Accuracy (%)	3865%	90.24%	91.90%
Mean Absolute Error	0.321	0.0423	0.0283

The observation of above reading shows the Relative Absolute Error is almost identical for each the prediction formula. Conjointly Least Median square regression offers higher numeric predictions however the time taken to make the model is 67 times that of linear regression, thus computational cost used by linear regression is much lower than that of least Median Square.

V. CONCLUSION

In this paper, we've got suggested an analysis of the soil information using completely different algorithms and prediction technique. In spite the very fact that the smallest amount median squares regression is known to produce higher results than the classical regression technique, from the given set of attributes, the foremost accurately foreseen attribute was “P” (Phosphorous content of the soil) and that decided using the regression technique in lesser time as compared to Least Median Squares Regression. In this paper we have demonstrated a comparative study of varied classification algorithms i.e. Naïve bayes, J48 (C4.5), JRip with the assistance of data mining tool rail. J48 is incredibly easy classifier to form a decision tree.

REFERENCES

[1] Cunningham, S. J., and Holmes, G. (1999). Developing innovative applications in agriculture using data mining. In the Proceedings of the Southeast Asia RegionalComputer Confederation Conference,1999.
[2] A Thesis by D.Basavaraju, Characterisation and classification of soils in Chandragiri mandal of Chittoor district,Andhra Pradesh. Jiawei Han and Micheline Kamber, Data Mining concepts and Techniques, Elsevier..
[3] Minaei-Bidgoli, B., Punch, W. Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System. Genetic and Evolutionary Computation, Part II. 2003. pp.2252–2263. Quinlan, J.R. C4.5: Programs for Machine Learning. Morgan Kaufman. 1993.
[4] Isbell, R. F. (1996). The Australian Soil Classification.Australian soil and land survey handbook. (Vol. 4).Collingwood, Victoria, Australia: CSIRO Publishing.
[5] Ibrahim, R. S. (1999). Data Mining of Machine Learning Performance Data. Unpublished Master of Applied Science (Information Technology), Publisher; RMITUniversity Press.
[6] Soil Survey Staff 1998, Keys to soil taxonomy. Eight Edition, Natural Resource Conservation Services, USDA, Blacksburg, Virginia. Soil Survey Staff 1951, Soil Survey Manual. US Department of Agricultural Hand book No. 18.
[7] P. Bhargavi, and Dr. S. Jyothi , Soil Classification Using GATree. International journal of computer science & information Technology (IJCSIT) Vol.2, No.5, October 2010.
[8] P. Bhargavi, and Dr. S. Jyothi , Soil Classification by Generating Fuzzy rules. International Journal on Computer Science and Engineering, Vol. 02, No. 08, 2010, 2571-2576.
[9] P. Bhargavi, and Dr. S. Jyothi , Fuzzy C-Means Classifier for Soil Data. International Journal ofComputer Applications (0975 – 8887), Volume