An Approach to Predict Soil Fertility in Agriculture using J48 Algorithm

Dr.S.Hari Ganesh, D.PrittyCindrella

Abstract—Data mining involves the systematic analysis of huge information sets, and data processing in agricultural soil datasets is exciting and fashionable analysis space. The productive capability of a soil depends on soil fertility. Achieving and maintaining applicable levels of soil fertility, is of utmost importance if agricultural land is to stay capable of nutrient crop production. During this analysis, Steps for building a prophetic model of soil fertility are explained. This paper aims at predicting soil fertility category mistreatment call tree algorithms in data processing. Further, it focuses on performance calibration of J48 call tree algorithmic program with the assistance of meta-techniques like attribute choice and boosting.

Index Terms—agricultural mining, DM, classification ,J48.

I. INTRODUCTION

Data processing may be a comparatively young and knowledge domain field of computing, is that the method that tries to get patterns in massive knowledge sets. It utilizes strategies at the intersection of AI, machine learning, statistics, and info systems. The goal of the data mining method is to extract information from a knowledge set and rework it into an obvious structure for additional use. A soil take a look at is that the analysis of a soil sample to see nutrient content, composition and different characteristics. Tests area unit sometimes performed to live fertility and indicate deficiencies that require being remedied. During this analysis, soil dataset containing soil take a look at results has been accustomed apply numerous classification techniques in data processing. Soil fertility may be a crucial attribute that is taken into account for land analysis , additionally achieving and maintaining necessary levels of fertility is vital for nurturing crop production, thus this paper includes steps for building associate degree economical and correct prognostic model of soil fertility with the assistance of J48 rule.

II. RELATED WORK

2.1. Dataset Collection

Data set needed for this analysis. These datasets contain Varied attributes and their several values of soil samples taken. Dataset has ten attributes and a complete 1988 instances of soil samples. Table one shows attribute description.

Table1: 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

2.2. Comparison of J48, NB Tree and Simple Cart for Soil Fertility Prediction

Soil fertility is taken into account be one among the essential attributes for deciding cropping pattern specially space. During this section, results of assorted call tree algorithms on dataset area unit shown. Supported these, the most effective classifier is chosen and any used for standardization its performance. The subsequent section explains call tree algorithms like J48, NBTree and Simple Cart briefly.

2.2.1. J48 (C4.5)

J48 is associate degree open supply Java implementation of the C4.5 algorithmic data processing tool. C4.5 may be a program that makes a call tree supported a collection of labeled input file. This algorithmic rule was developed by Ross Quinlan. The choice trees generated by C4.5 will be used for classification, and for this reason, C4.5 is usually noted as a applied math classifier.

2.2.2. NB Tree

This algorithmic rule is employed for generating a call tree with naive Bayes classifiers at the leaves (Kohavi R., 1991).

2.2.3. Simple Cart

It's a non-parametric call tree learning technique that produces either classification or regression trees, looking on whether or not the variable is categorical or numeric, severally. It is used for implementing stripped-down cost-complexity pruning (Breiman L. et al. 1984) during this paper, 3 call tree techniques (J48 (C4.5), NBTree and Simple Cart) in data mining were evaluated and compared on basis of accuracy and Error Rate. Denary cross-validation was employed in the experiment. Our studies showed that J48 (C4.5) model clothed to be best classifier for soil samples.

Table2: Comparison of different classifiers

Classifier	NB Tree	Simple Cart	J 48
Correctly Classified Instances	1700	1824	1827
Incorrectly Classified Instances	288	164	161
Accuracy (%)	85.51	91.75	91.90

III.ENHANCED PERFORMANCE OF J48 ALGORITHM

Accuracy of J48 rule for predicting soil fertility was highest; thence it absolutely was used as a base learner. Now, the aim was to extend its accuracy with the assistance of other meta-techniques like attribute choice and boosting.

3.1 With attribute selection
Attribute choice reduces data set size by removing irrelevant/redundant attributes .It finds minimum set of attributes such ensuing chance distribution of information categories is as shut as potential of original distribution. Attribute authority methodology – CfsSubsetEva was used, that evaluates the price set of attributes by considering the individual prophetic ability of every attribute (Hall M.A., 1998).Following square measure the results exploitation Attribute Selected Classifier with base learner as J48.

Table3: Using Attribute Selected Classifier with J48 as Base Learner

Correctly identified instances	1853	93.2093 %
Incorrectly identified instances	135	6.7907 %

It may be clearly seen that accuracy has been exaggerated from 91.90 to 93.20 when application of attribute choice technique.
3.2 Combining attributes choice and boosting technique
Boosting may be a machine learning meta-algorithm for playacting supervised learning. It will boost performance of weak learner and convert it into a powerful learner. It will increase the weights of incorrectly known instances and reduces the weights of properly known instances over its iterations.
Adaboost is weka implementation of boosting technique that is employed for enhancing a nominal category classifier (Freund Y. and Schapire R. 1999). Following are the results when exploitation combination of attribute choice and Adaboost with J48 as base learner.

Table4: Results after using combination of attribute selection and boosting with J48 as base learner

Correctly identified instances	1923	96.7304 %
Incorrectly identified instances	65	3.2696 %
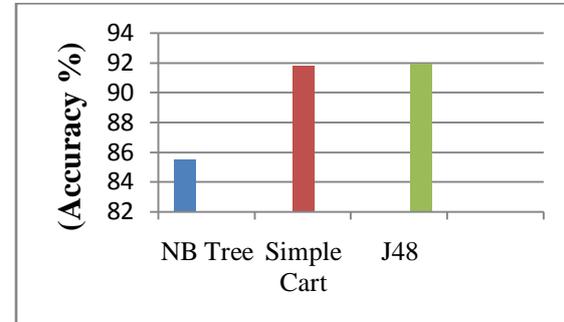
Here, accuracy is enhanced up to 91.90% which makes this predictive model to be more accurate.

IV.CONCLUSION

The large amounts of knowledge that area unit these days just about harvested at the side of the crops have to be compelled to be analyzed and will be wont to their full extent. Varied call tree algorithms will be used for prediction of soil fertility. My studies showed that J48 offers 91.90 the concerns accuracy; therefore it will be used as a base learner. With the assistance of different meta-algorithms like Attribute choice and boosting, J48 offers accuracy of 96.73% that makes a decent prophetical model.

V. REFERENCES

[1]. Ashok Kumar. D,Kannathasan. N, "*A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining*" IJCSI International Journal of Computer Science Issues,Vol. 8, Issue 3, No. 1, May 2011 ISSN (Online): 1694-0814.

[2]. Chandrakanth. Biradar, Chatura S Nigudgi, *"An Statistical Based Agriculture Data Analysis*" International Journal of Emerging Technology and Advanced EngineeringVolume 2, Issue 9, September 2012 (ISSN 2250 – 2459)

[3]. Darcy Miller, Jaki McCarthy, Audra Zakzeski, "*A Fresh Approach to Agricultural Statistics: Data Mining and Remote Sensing*" National Agricultural Statistics Service 3251 Old Lee Highway, Fairfax, VA 22030 - JSM 2009.

[4]. KamalakKannan. P,Hemalatha. H, "*Agro Genius: An Emergent Expert System for Querying Agricultural Clarification Using Data Mining Technique*" International Journal of Engineering and Science ISBN: 2319-6483, ISSN: 2278-4721, Vol. 1, Issue 11 December 2012.

[5]. Leisa J. Armstrong, Dean Diepeveen and Rowan Maddern, *"The application of data mining techniques to characterize agricultural soilprofiles"*, Volume 70,pages 85-100.AustralianComputer society,Inc.Darlinghurst,Australia,ISBN-978-1-920 682-51-4.

[6]. Lingxian Zhang, Xinxing Li, *"The corn disease remote diagnostic system in China"* Journal of Food, Agriculture & Environment Vol.10 (1): 617-620,2012

[7]. Ramesh Vamanan and K.Ramar, *"Classification of Agricultural Land Soils A Data Mining Approach"* International Journal on Computer Science and Engineering (IJCSE)-2012.

[8]. Sally Jo Cunningham and Geoffrey Holmes, *"Developing innovative applications in agriculture using data mining".* Department of Computer Science, University of Waikato, Hamilton, New Zealand.

[9]. Quinlan, J.R. *C4.5: Programs for machine learning.* Morgan Kaufmann, SanMateo, 1993,CA.

[10]. Wang, Y. and Witten, I.H. *"Induction of model trees for predicting continuousclasses."*Proceedings of the Poster Papers of the EuropeanConference on MachineLearning,Prague, 128-137,1997.

[11]. Witten, Ian H., and Frank, Eibe *Data Mining: "Practical machine Learning Toolsand Techniques with Java Implementations",.* MorganKaufmann,SanFransico, 1999.