Efficient Classification of Agriculture Land Soils In Statewise From India Using Data Mining With Weka

Mrs. T. Seeni Selvi

Associate Professor, PG & Research Department of Computer Science, Hindusthan College of Arts and Science, Coimbatore, Tamilnadu, India

ABSTRACT

Agriculture is the ultimate field for human society. And also it is the backbone of India. The production of agriculture products is continuously going decrease in recent years. There is a need for an immediate concentration for this agriculture field. The knowledge acquisition is possible by using existing agricultural database with data mining. Classification is the familiar methods that are used to provide the important information. This paper shows the different classification methods were applied into the dataset and produce the results using WEKA.

Keywords : Data Mining, Classification, Classifier, Decision Tree

I. INTRODUCTION

Data mining is the process to find interesting knowledge from large amounts of data [1]. The aim of the data mining process is to extract knowledge from an existing data set and transform it into a human understandable formation for advance use. It is the process of analyzing data from different view and encapsulates it into useful information. There is no restriction to the type of data that can be analyzed by data mining.
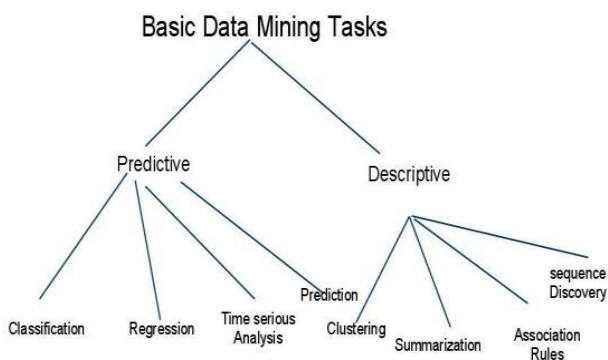

Figure 1. Data Mining Techniques

Data mining is broadly categorized into two main aspects that are Predictive and Descriptive model

respectively. Descriptive data mining is to identify the general properties of the data in the database. Predictive data mining is used to predict explicit values based on patterns decided from known results.

CLASSIFICATION

Classification is one of the predictive model. Classification consists of assigning a class label to a set of unclassified case. There are two types in classification.

1) Supervised Classification
2) Unsupervised Classification

Supervised Classification

The set of possible classes is known in advance. For Classification, there are many classifiers are used.

Unsupervised Classification

Set of possible classes is not known in advance. After classification we can try to assign a name to that class. Unsupervised classification is called Clustering.

DECISION TREE

Decision tree is the widely used classification algorithm. Decision Tree builds the classification model in a tree structure. It breaks down a dataset

into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

AGRICULTURE

Indian Agriculture is known for its diversity which is mainly result of variation in resource and climate, to topography and historical, institutional and socio economic factors. Data mining is a essential feature to discover the agricultural related knowledge such as soil fertility, yield prediction and so on. Classification algorithms involve finding rules that partition the data into disjoint groups.

II. REVIEW LITERATURE

S.Manimekali et al[4] presented a model for focalized on soil type. The number of techniques and its strategies were analyzed from her survey. The techniques like Naïve Bayes, Genetic Algorithm, Neural Network and J48 were basically used for higher classification. The set of rules JRip was applied and validated in her research using WEKA. It was located that 67% of correctly classified insatnces of soil utilized in cropyielding. All the soil parameter values are taken from into consideration earlier than cropping.

V.Rajeswari et al[5] compared and produced the results of data mining techniques. The data set was taken from the soil testing lab., Viruthunagar District. They had taken 110 data which contains the attributes such as Village Name, Soil type or Color, Soil Texture, PH, EC, Lime Status, Phosphorous. They took three classification algorithms JRip,J48 and Naïve Bayes. Based on Training Dataset it is concluded that weighted average of True Positive Rate of JRipclassifer is 0.982. Naïve Bayes classifier produces True Positive rate 0.97 and J48 produces the True Positive rate is 0.86.

Ramesh Vamanan[6] at al presented a work that compared the different classifiers and the outcome of this research could improve the management and systems of soils uses throughout a large number of fields that include agriculture, horticulture, environmental and land use management. The experiments conducted analyzed small number of traits contained within the dataset to determine their effectiveness when compared with standard statistical techniques. The agriculture soil profiles that are used in this research were selected for completeness and for classification of soils. The recommendations arising from this research implies that data mining techniques may be applied in the field of soil research in the future as they will provide research tools for the comparison of large amount of data. Data mining techniques when applied to an agricultural soil profile, may improve the verification of valid soil profile, may improve the verification of valid patterns and profile classification when compared to standard statistical analysis techniques.

P.Bhagavathi et al[7] presented a comparative study of data mining algorithms like GA Tree, Fuzzy Classification Rules and Fuzzy C-Means algorithm. The achieved performances were compared and analyzed on the collected Supervised and Unsupervised Soil data. GA Tree and Fuzzy Classification rules were used for supervised learning. Classification based on Fuzzy rules gave much performance than GA Tree. For Unsupervised learning Fuzzy C-Means algorithm was used for classifying the soil data. This helps to classify soil texture based on soil properties efficiently, which influences fertility, drainage, water holding capacity, aeration, tillage and bearing strength of soils.

## III. RESULT AND DISCUSSION

A) DATASET

In this work, we collected the dataset from India Crop Production- State Wise. The name of the dataset is Crop Production.CSV.

B) ENVIRONMENT

WEKA Tool is used for the classification of crop production.

C) METHODOLOGY

Decision tree is the famous classification algorithm which is widely used for prediction. For this research, Random Forest Tree and Decision Stump algorithms are used and the results are compared.

RANDOM FOREST TREE CLASSIFICATION
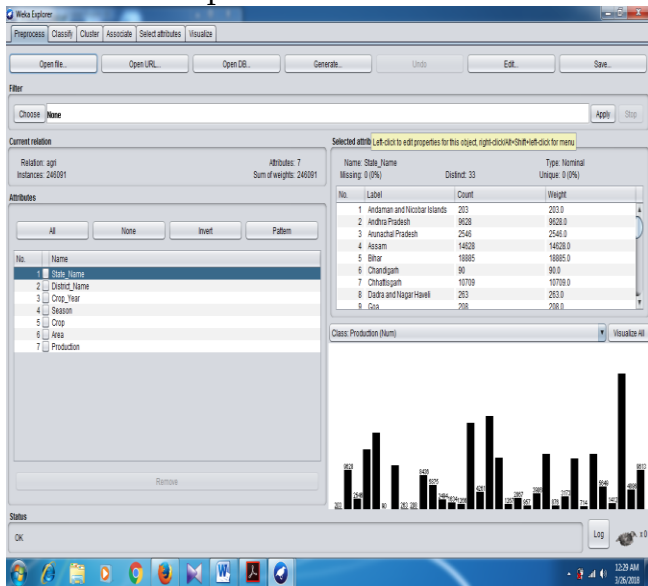
Time taken to build a classification model is 5.56 seconds. There are 7 attributes used in this dataset. That are State_name, District_Name, Crop_year, Season, Crop, Area and Production.
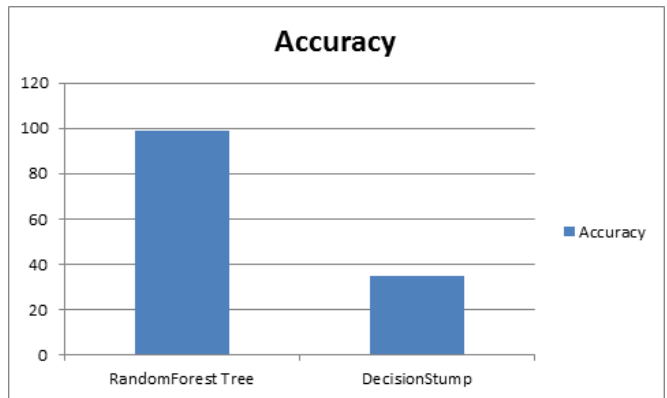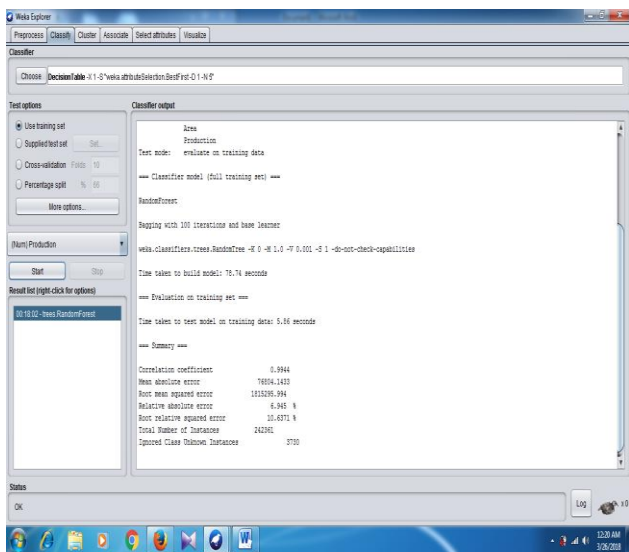
DECISION STUMP
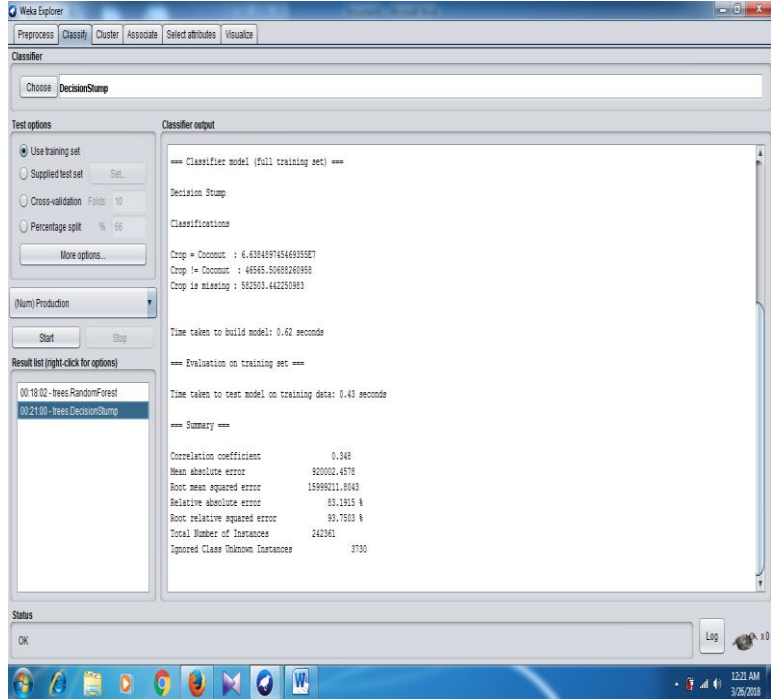
Time taken to build a model using Decision Stump is 0.43 seconds.

COMPARISION WITH THE RESULTS

The Accuracy of predicting the classes using this two model

Decision Tree Model	Accuracy
RandomForest Tree	99
DecisionStump	35

IV. CONCLUSION

In this paper, the comparative analysis of two algorithms like Random Forest Tree and Decision Stump is projected. Dataset is downloaded from India Crop Production- State Wise. Random Forest tree gives better results of this dataset. Random Forest Tree Produced 99% accuracy where the DecisionStump produced only 35% accuracy.

V. REFERENCES

[1]. Jiawei Han, MichelineKamber, ‖Data Mining: Concepts and Techniques‖, 2nd edition, Morgan Kaufmann, 2006.

[2]. K.Thenmozhi, P.Deepika," Heart Disease Prediction Using Classification with Different Decision Tree Technique" International Journal of Engineering Research and General Science, Volume 2, Issue 6, October-November, 2014.

[3]. K.Thenmozhi, P.Deepika "Different Datamining Techniques Involves In Heart Disease Prediction-A Survey" in International Journal of Scientific Research ISSN NO. 2277-8179 September 2014.

[4]. S.Manimekalai,K.Nandhini, " Survey on Classification Techniques for Soil Data Prediction to Better Yielding of Crops", International Journal of Computer Science and Engineering, Volume-6, Issue-1.

[5]. V. Rajeswariand K. Arunesh, "Analysing Soil Data using Data mining Classification Techniques", Indian Journal of Science and Technology, Vol.9(19), May 2016.

[6]. Ramesh Vamanan, K.Ramar, "Classification of Agricultural Land Soils a Data Mining Approach", International Journal on Computer Science and Engineering, Vol-3 No.1, Jan 2011.

[7]. P.Bhagavathi, Dr. S. Jyothi, "Soil Classification Using Data Mining Techniques: A Comparative Study", International Journal of Engineering Trends and Technology, July to August Issue, 2011.

[8]. Geetha MCS. Implementation of association rule mining for different soil types in agriculture. International Journal of Advanced Research in Computer and Communication Engineering. 2015 Apr; 4(4):520–2.

[9]. Solanki J, Mulge Y. Different techniques used in data min ing in agriculture. International Journal of Advanced Re search in Computer Science and Software Engineering. 2015 May; 5(5):1223–7.

[10]. Bhuyar V. Comparative analysis of classification techniques on soil data to predict fertility rate for Aurangabad District.International Journal of Emerging Trends and Technology in Computer Science. 2014 Mar-Apr; 3(2):200–3.

[11]. Fathima NG, Geetha R. Agriculture crop pattern using data mining techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 2014 May; 4(5):781–6.

[12]. Suman, Naib BB. Soil classification and fertilizer recom mendation using WEKA.International Journal of Comput er Science and Management Studies. 2013 Jul; 13(5):142–6.

[13]. Ramesh D, Vardhan VB. Data mining techniques and ap plications to agricultural yield data. International Journal of Advanced Research in Computer and Communication Engineering. 2013 Sep; 2(9):3477–80.

[14]. Tsai F, Lai JS, Chen WW, Lin TH. Analysis of topographic and vegetative factors with data mining for landslide verifi cation.Ecological Engineering.2013 Dec; 61:669–77.