

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

Third Generation Agricultural Support System Development Using Data Mining

Jeysenthil.KMS¹, Manikandan.T², Murali.E³

Research Scholar, Dept. of Computing Science and Engineering, VIT university, Chennai, Tamilnadu, India²

Student, M.E CSE, Apollo Engineering College, Chennai, Tamilnadu, India¹

Assistant Professor, Dept of CSE, Apollo Engineering College, Chennai, Tamilnadu, India³

Abstract: Agriculture is a multi-dimensional sector directly linked to the malnutrition and food security of the country. This paper provides an effusive farmer supporting system called “Third generation agricultural support system development” where it assess the various classification techniques of data mining and apply them to a soil science database to establish if meaningful relationships can be found. In data mining conception, clustering and classification technique make efficient knowledge exploration and knowledge acquisition from integrated organic farming and produces better solution to the farmers about their cultivation. The agricultural support system development consists of collecting soil information’s with computational approach which is followed in organic way of farming huddling using clusters(k-means) and gives best result using classification(orthogonal array) from the databases to the farmers along with beneficial microbes to release nutrients to crops for increased sustainable production in an ecological – intensification and pollution free environment.

Keywords:Support System, Classification, Clustering, Sustainable, Exploration, Exploitation.

I. INTRODUCTION

Data mining is the process of discovering previously unknown and potentially interesting patterns in large datasets. Recent technologies in data mining are nowadays able to provide a lot of information on agricultural-related activities, which can then be analyzed in order to find important information. India is principally an agrarian economy. Agriculture sector in India contributes 16% of GDP and 10% of export earnings. So agricultural development should be considered obligatory for the country. The analysis of agricultural data sets with various data mining techniques may yield outcomes useful to researchers in the Agricultural field.

This research encompass the delopment of agricultural supprt system which uses decision induction technique, k-means clustering and orthogonal array implementations by collecting the informations of major crops that prevails over tamilnadu and specifies a integrated organic farming method that requires less water, involves less expenditure and gives more yield which is useful for small and marginal farmers.Sustainable integrated organic farming collect soil attributes and imposes farmers to use systematic process of cultivation by adopting proper seed selection and use of good seeds, manure application, use of growth promoters in regular intervals, adoption of water and pest management practices will help farmers to spend less on cultivation and to gain more profit. In addition, adoption of organic techniques places a less strain on natural farming ecosystems.

A [1] survey of the available literature on data mining and pattern recognition for soil data mining is presented. Data mining in Agricultural soil datasets is a relatively novel research field. Efficient techniques can be developed and tailored for solving complex soil datasets using data mining.The [2] goal of this paper is to provide a comprehensive review of specific decision tree classifier, discrete wavelet transformation for statistical agricultural data available and graphical user interface method for the purpose of data mining to enhance energy forecasting analysis. The [3] cutting-edge techniques to produce agricultural statistics used to guide policies and standards that bring our nation’s food from the farm to shelf have sprouted from the innovation of statisticians at the USDA’s National Agricultural Statistics Service (NASS). Recent use of remote sensing has been the crux of the development of a “Census by Satellite” (Cropland Data Layer Program), and data mining techniques have excavated statistics utilized in ensuring the quality of data from collection to estimation. In paper[4], an expert system exclusively for the integrated disease management in

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

finger millet is being presented by incorporating fuzzy logic method to frame the rules and apply defuzzification to attach a value to the severity of the disease identified, based on which the control and remedial measures are suggested. One of the paper[5] designed and implemented a corn disease remote diagnostic system, which is focused on the prevention, diagnosis and control of diseases that affect China corn production. The knowledge acquisition process was conducted based on the knowledge obtained from the literature and experts.

The work in this paper is divided into two models. (1) Classification, (2) Clustering. The first model includes soil registration and accounts where data are collected and classified using decision induction techniques. The data values are collected from tamilnadu agricultural university cuddalore where the farmers test their field soils in it. In clustering, the K-means algorithm is used to cluster the values which are classified and gives best combination of clustered values. Paper is organized as follows. Section II describes modules, classification techniques and system design. Section III represents clustering and analysis, where one of the efficient algorithm – k-means has been implemented and in Section IV the Result shows the table of combination which are analysed as best combinations.

II. MATERIALS AND METHODS

2.1 Soil Registration and accounts

In soil record-keeping and accounts, the various types of soil nutrients such as macro nutrients (NPK, Ph.), micronutrients (Fe, Mg, Zn, B, Cu) and microbes (basso bacterium) are collected and stored in a database based on location and pin codes. The hydrogeology charts which displays watershed managements across cuddalore and specifies aquifers directions and water irrigation magnitudes to various places of cuddalore. The tamilnadu agricultural university (TNAU) records all soil samples across cuddalore district and has a history of successful grown paddy crops for the past 10 years (2003-2013). The figure 1 illustrates farmer’s accounts are classified as existing and new accounts which consists of eminent features to come transversely.

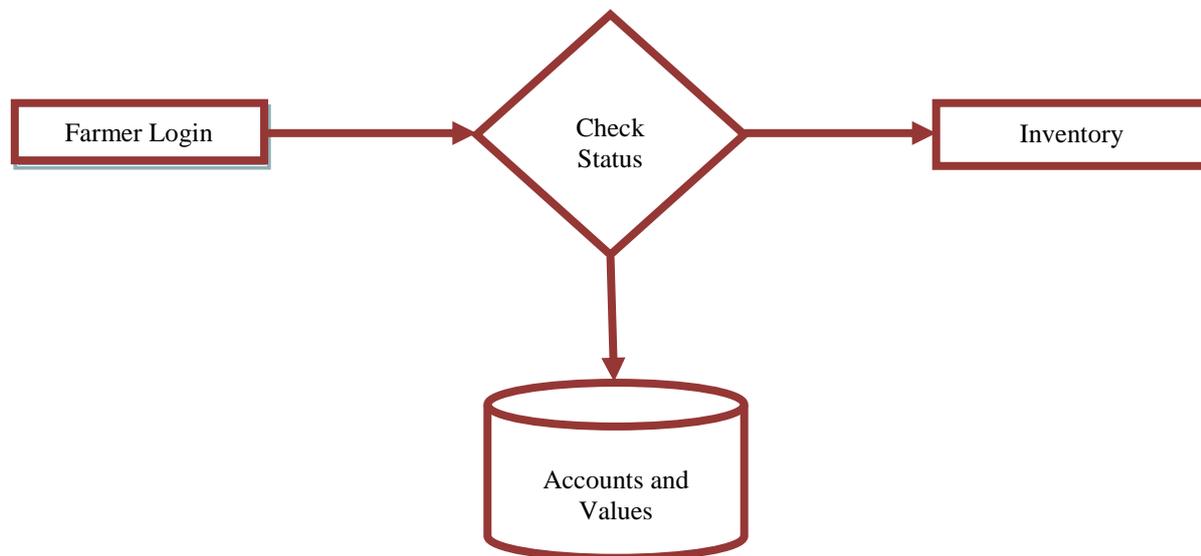


Fig.1Record Keeping and Accounts

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

2.2 Clustering and Formations

In Clustering and formations the predicted informations are clustered using K-Means algorithm where the given soil values are grouped according to their derived types. From the the groups the centroid points are created to retrieve best combinations of data and the results are shown in clusters observations into k groups, where k is provided as an input parameter. . It then assigns each observation to clusters based upon the observation’s proximity to the mean of the cluster. The cluster’s mean is then recomputed and the process begins again. The centroid represents the most typical case in a cluster. For example, in a data set of customer ages and incomes, the centroid of each cluster would be a customer of average age and average income in that cluster. The centroid is a prototype. It does not necessarily describe any given case assigned to the cluster. The k-means algorithm is an evolutionary algorithm that gains its name from its method of operation. The algorithm clusters observations into k groups, where k is provided as an input parameter. It then assigns each observation to clusters based upon the observation’s proximity to the mean of the cluster. K-means clustering is an algorithm used for partitioning (clustering) N data points into K disjoint subsets so as to minimize the sum-of-squares criterion:

$$J = \sum_{j=1}^K \sum_{n \in S_j} |x_n - \mu_j|^2$$

where x_n is a vector representing the n^{th} data point and μ_j is the geometric centroid of the data points in S_j . The number of clusters K must be selected at onset. The data points are assigned at random to initial clusters, and a re-estimation procedure finally leads to non-optimized minima. Despite these limitations, and because of its simplicity, k-means clustering is the most popular clustering strategy. The figure 2 shows K-Means clustering intends to partition n objects into k clusters in which each object belongs to the cluster with the nearest mean. This method produces exactly k different clusters of greatest possible distinction. The best number of clusters k leading to the greatest separation (distance) is not known as a priori and must be computed from the data.

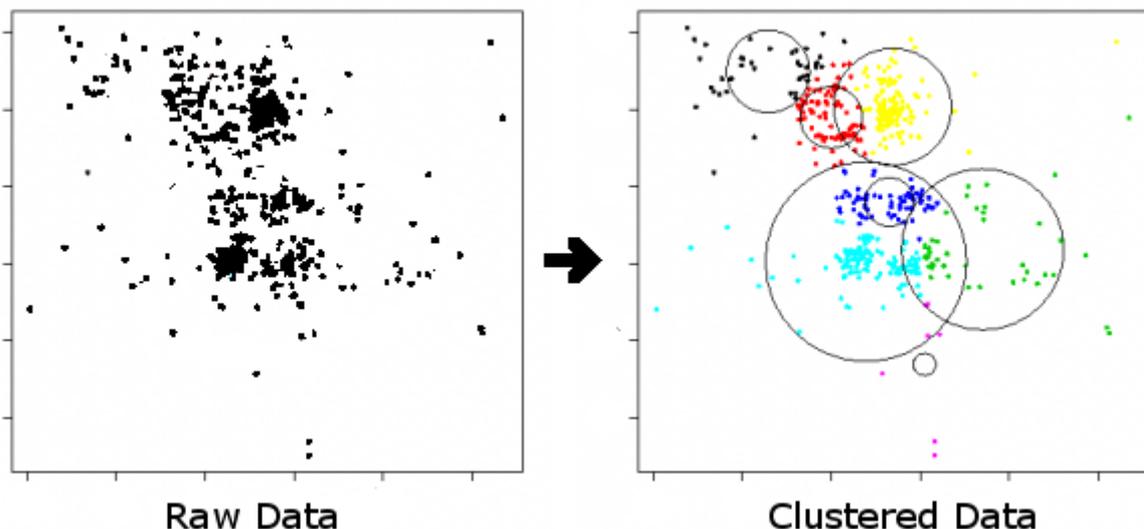


Fig. 2 K-Means Clustering

2.3 Soil Classification and Inventory

In soil classification and inventory, the values are analyzed using decision induction technique and classified to develop innovative approaches to predict the best combination for cultivating crops. A decision tree is a flow-chart-like tree

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

structure, where each internal node denotes a test on an attribute, each branch typifies an outcome of the test, and leaf nodes typify classes or class distributions. The top most node in a tree is the root node. In order to classify an unknown sample, the attribute values of the sample are tested against the decision tree. A path is discovered from the root to a leaf node that holds the class prediction for that sample. Decision trees were then coinciding to classification rules using IF-THEN-ELSE. The basic algorithm for decision tree induction is an insatiable algorithm that constructs decision trees in a top-down repetitive divide-and-conquer manner where each combinations are specified with three levels of values as stored in an orthogonal array implementations and decisions are taken based on that values. The figure 3 illustrates the decision taken when the soil types are classified and clustered.

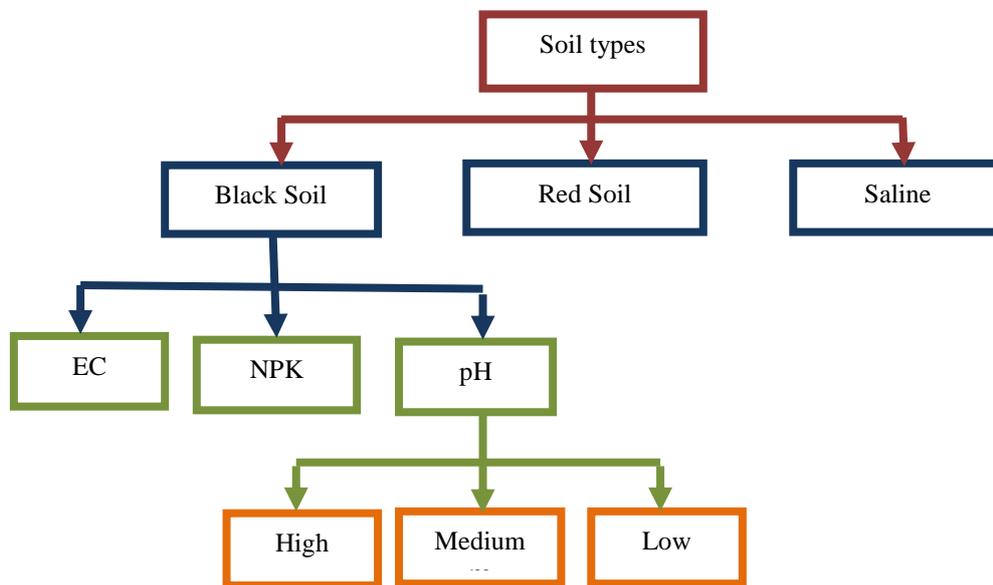


Fig. 3 Decision Tree for Soil Classification

2.4 Best combinations and Methods

The best combinations are obtained by permuted statistical analysis where the support system operates the input given to the inventory and displays soil type, crops supported, rainfall level, pest management issues and weed control type to the farmers. The profile displays the 6 organic methods its implementation and the crops supported with its type. So the farmers could refer the better combinations prevailed over their region and can implement the organic way of farming as enforced. The figure 4 shows the best combinations of clustered data.

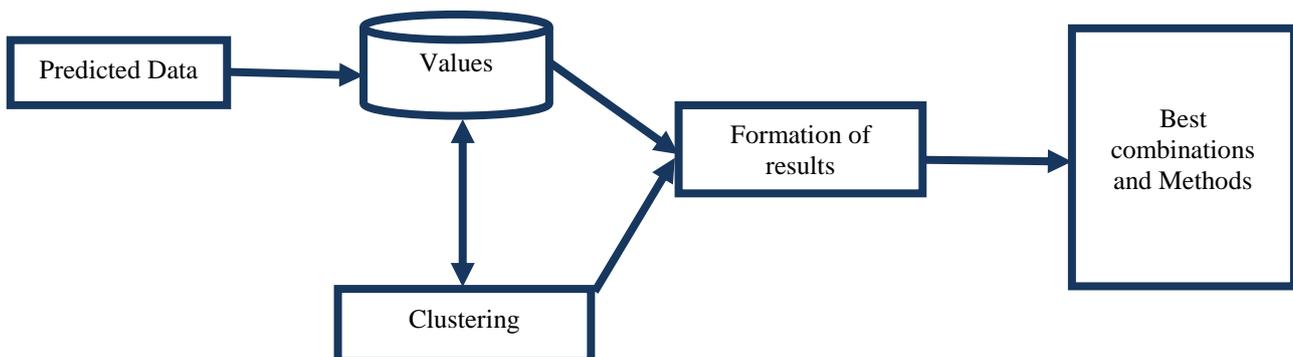


Fig. 4 Best Combinations

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

III. CLUSTERING AND ANALYSIS

3.1 Data Analysis:

The farmer's location and pin code of the cuddalore district are manipulated and process statistical and permuted values by classification, clustering and formation analysis to specify a desired combinations and methods of organic farming. The farmers in the cuddalore district are the prominent cultivators of paddy and sugarcane. They are cultivating sugarcane to about 355 tonnes in 30 hectares of land and rice to about 316 tonne in 115 hectares of land. The figure 5 map explores the total land under cultivation of sugarcane and paddy.

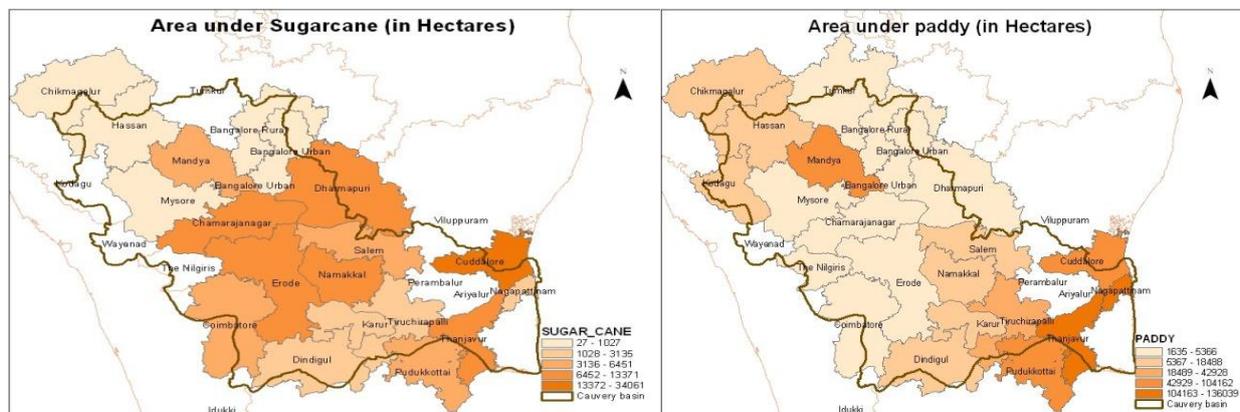


Fig. 5 Area Cultivation in Hectares

3.2 Data Collection:

The various locations and its soil attributes are collected with their corresponding pincode and matched the mean value as shown in the table 1. The table denotes 7 blocks of cuddalore district and their mean attributes where sugarcane, paddy, cotton, groundnut and black gram are the major crops cultivated in and around these blocks. the soil attributes such as pH, EC, NPK, Zinc, Copper, Iron, Manganese are recorded and manipulated to support the blanket recommendations of the major crops cultivated.)

The soil attribute charts which displays soil attributes across cuddalore and specifies the mean value within the blocks at the various places of cuddalore.

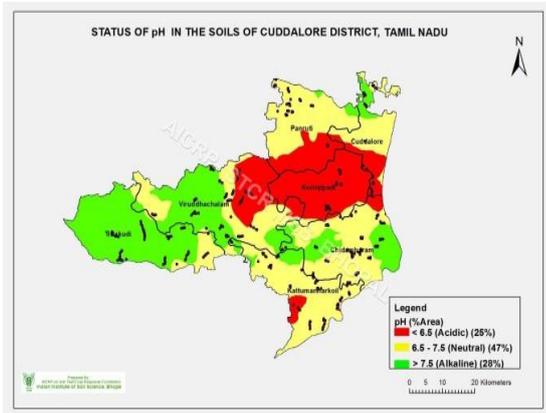
Location	Pincode	Major Soil	pH	EC	N	P	K	Zn	Fe	Mn	Cu
Cuddalore	607001	Black	7.0	0.5	56	11	70	0.67	17.43	7.92	1.26
Panruti	607106	Red	5.9	0.32	63	11	70	0.66	26.05	8.63	1.15
Kurijipadi	607302	Black	5.4	0.31	59	7.5	95	0.72	14.85	9.42	1.20
Chidambaram	608001	Black	7.3	1.10	77	25	110	1.31	10.36	8.79	0.81
Kattumannarkoil	608301	Black	8.2	0.36	67	7.5	75	0.65	6.45	3.42	1.15
Viruthachalam	606001	Alluvial	6.8	0.42	69	5	90	0.48	22.15	11.1	1.30
Titakudi	606106	Red	7.7	0.34	62	7.5	77	0.86	18.36	11.43	0.92

Tab. 1 Mean Results

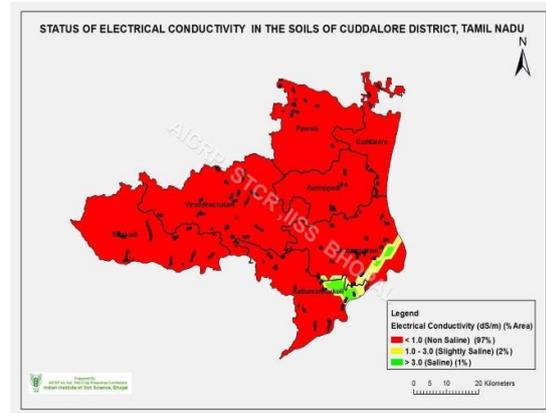
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

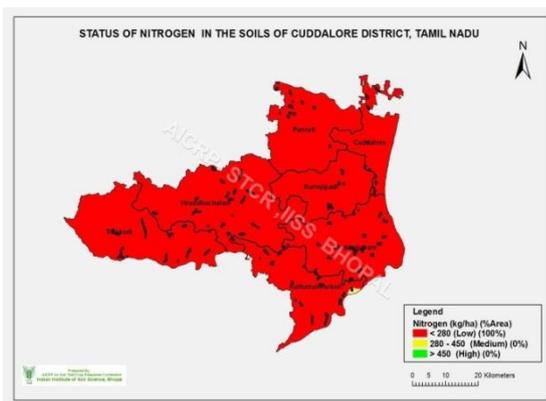
Vol. 3, Issue 3, March 2014



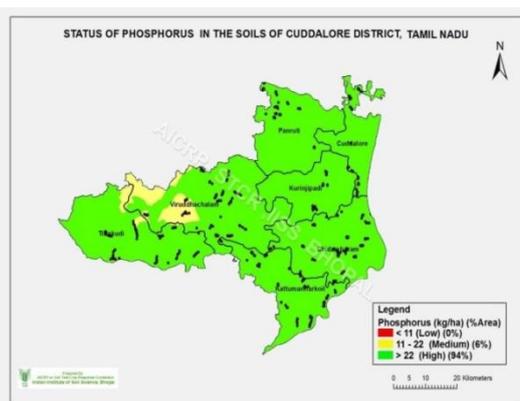
(a)



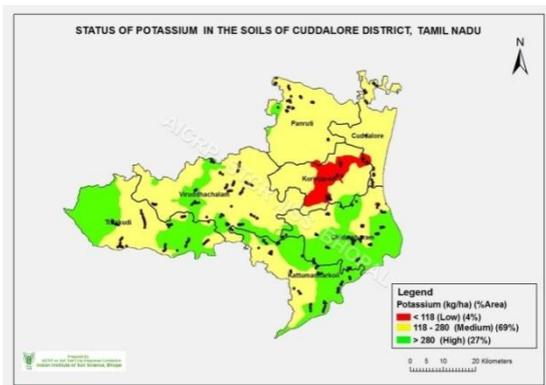
(b)



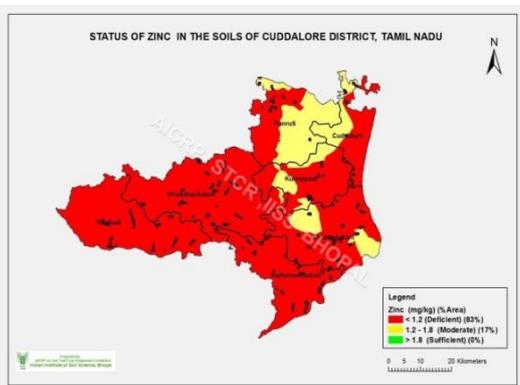
(c)



(d)



(e)



(f)

Fig 6 Maps in Cuddalore showing (a) pH (b) EC (c) Nitrogen (d) Phosphorus (e) Potassium (f) Zinc

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

IV. RESULT AND DISCUSSION

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.

L₉ (3⁴) Standard Orthogonal Array :

Experiment no	Soil Type	Crop Type	Rainfall Level	Pest Management
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

Tab. 2 Experimentation Combinations

The table 2 denotes the experimental combinations suggested to the farmers. The research activities involved a process to establish if classification could be found in the data. These processes involved the statistical manipulation of the data set in SQL server.

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. The aim of the research was to determine if a relationship or correlation can be established with soil data set. The process involved the creation of analysis tools and charting the data so that the classification of soils is displayed and experts can interpret the findings. The initial screen provided a set of information that is required by the researchers and took a large amount of time to complete with the current statistical methods.

V. CONCLUSION

This project propose not only awareness, but also daunting integrated organic methods to the farmers for the eco-friendly and pollution free environment which prevents vigorous causes such as Global warming, salt accumulation in farms, soil erosion, nutrients deficiency in soils etc. This paper provides Support system for agriculture. This support system provides basic information about organic agriculture for the beginners in farming, giving the best combination for cultivate the crops and creating awareness about the organic farms .This support system advices and suggestions in the area of crop field by providing facilities like dynamic interaction between expert peoples and the user without the need of expert (crop) at all times. The inclusion is simple but effective techniques will help in development of the

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

agriculture and industrial fields. This work performs the minimum statistics on agricultural data more efficiently and easily.

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 Engineering-Volume 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] Kamalak Kannan. 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 soil profiles" Volume 70, pages 85-100. Australian Computer Society, Inc. Darlinghurst, Australia, ISBN-978-1-920682-51-4.
- [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 (IJCSSE)-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, San Mateo, 1993, CA.
- [10] Wang, Y. and Witten, I.H. "Induction of model trees for predicting continuous classes." *Proceedings of the Poster Papers of the European Conference on Machine Learning*, Prague, 128-137, 1997.
- [11] Witten, Ian H., and Frank, Eibe (1999) *Data Mining: Practical machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco.