IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Soil Prediction and Disease Prediction of Plant

[1]Raj Singh, [2]Naveed Ellahie , 3 Prof. Sagar G. mohite

[1]UG Student, [2]UG Student , 3 Assistant Professor

Department of Computer Engineering

Bharati Vidyapeeth (Deemed to be University) College of Engineering Pune

Abstract - Crop management of certain agriculture region is depends on the climatic conditions of that region because climate can make huge impact on crop productivity. Real time weather data can helps to attain the good crop management. Data mining is the process of discovering of new pattern from large data sets, this technology which is employed in inferring useful knowledge that can be put to use from a vast amount of data, various data mining techniques such as classification, prediction, clustering and outlier analysis can be used for the purpose. Real time weather data can helps to attain good crop management. Weather is one of the meteorological data that is rich by important knowledge. In this project we include the hybrid model to improve the agriculture productivity by using data mining techniques.

Yield prediction is very popular among farmers these days, which particularly contributes to the proper selection of crops for sowing. This makes the problem of predicting the yielding of crops an interesting challenge. Earlier yield prediction was performed by considering the farmer's experience on a particular field and crop. This work presents a system, which uses IOT techniques in order to predict the analyzed soil datasets. The category, thus predicted will indicate the yielding of crops.

Keywords – Crop yield prediction, Bee Hive clustering algorithm, crop growth , precision Agiculture .

1. INTRODUCTION

As agriculture struggles to support the rapidly growing global population, plant disease reduces the production and quality of food, fiber and bio fuel crops. Losses may be catastrophic or chronic, but on average account for 42% of the production of the six most important food crops. Losses due to postharvest disease can be disastrous, especially when farms are a long way from markets and infrastructure and supply chain practices are poor. Many postharvest pathogens also produce toxins that create serious health problems for consumers.

Farmers spend lot on disease management, often without adequate technical support, resulting in poor disease control, pollution and harmful results. In addition, plant disease can devastate natural ecosystems, compounding environmental problems caused by habitat loss and poor land management.

A crop prediction is a huge problem that occurs. A farmer had an attention in understanding how much produce he is going to expect. Traditionally farmers decide this based on permanent experience for specific yield, plants and weather conditions. Character directly thinks about produce prediction rather than concerning on crop prediction. If the correct crop is expected then yield will be better. Problem of crop and yield prediction using modified k-means clustering algorithm thereby creating better earnings for berry farmers. Clustering is the process of grouping the data into classes or groupings, so that objects within a cluster have high similarity in agreement to each other but are incredibly dissimilar to objects in option clusters.

This project is focused on research aimed at improving food security by reducing crop losses, particularly for low-income farmer.

LITERATURE SURVEY

Many cluster algorithms model combination. C.Preisach et.al [1] proposed a generic relational ensemble model which improves cluster accuracy of scientific publications by combining the probability distribution of several relational attributes and local attributes. Relational attributes probability is determined using graph representation and local graph using traditional text cluster. Heterogeneity becomes a major issue since models are combined in specific format. Xiaoxin Yin et.al [2] proposed cross mine tree and cross mine rule. Both cluster approaches make use of Tuple ID propagation which helps in virtual join among relation rather than physical join. In Tuple ID selection, key attributes are used for spamming among relations. For non target relation, all relations are joined together for computing foil gain. However both the methods are unable to handle database imbalances for complex application.

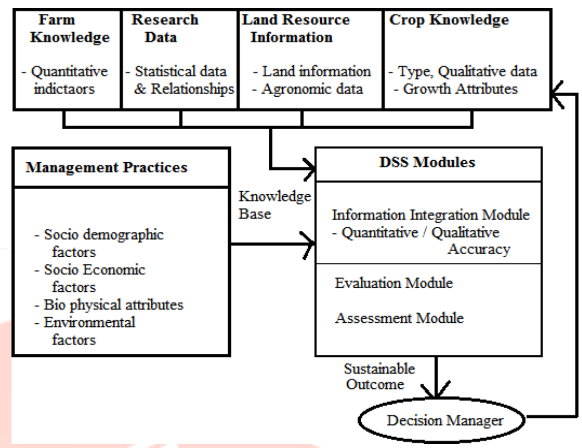

Figure 1 Crop Knowledge Base Frame work

2. CROP KNOWLEDGE BASE

A crop knowledge base is constructed with set of relations (Crop_details, Region and Cropping_info). Each relation contains at least one or more key attributes (primary or foreign). Crop Type contains attributes Crop_ID, CropType, CropVariety, Yield Duration and Favorable Seasons. Crop_ID is guaranteed to be unique and used as primary key. CropType represents type of crop. Crop Variety represents different variety of crops, seed type (e.g. Co43, IR20, ADT36). Season represents suitable seasons for cropping particular crop. Region of crop growth contains attributes Region_ID, Region Name, Soil Type, Soil Ph, Water Ph, Sunlight and Rainfall. Region_ID is assigned to uniquely to each RegionName. SoilType represents different type of soils in particular region (e.g. black soil, clay soil, alluvial soil).

3. CROP YIELD PREDICTION ARCHITECTURE

Subset selection of attributes from crop knowledge base is handled by feature selection for robust learning. The cluster algorithm classifies crop yield into various classes for a particular region. Classified yield information helps for finding better crop for a region and crop yield is prediction done by prediction rules.

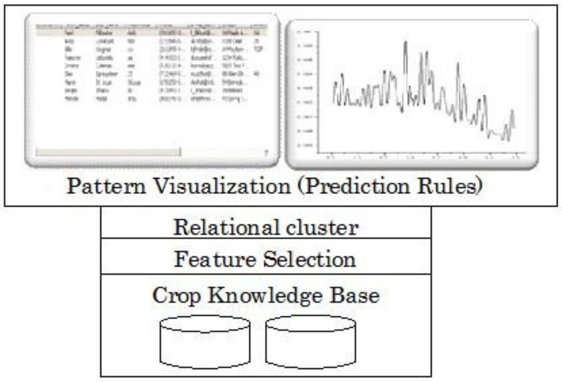

Figure 2 Crop yield prediction architecture

In a crop the growth parameters like optimum LAI and CGR at flowering have been identified as the major determinants of yield [10]. A combination of these growth parameters can demonstrate the varying yields which is better than any individual growth variable [14].

Feature Selection

Feature selection plays a major role in selecting attributes from crop knowledge base. To classify crop yield feature selection filters necessary parameters on basis of

∀Region x Crop yield (Query –A)

∀Region x SoilpH x Water pH x Rainfall (Query –B)

∀ [Region x Crop yield] (Query –C)

Region x Avg (Soil pH) x Avg (Water pH) x Avg (Rainfall)] (Query –D)

4. END WHILE

The algorithm starts with an initial population of n scout bees. Each bee represents a potential clustering solution as set of 'z' cluster centres or zones.

Step1- The initial locations of the zones are randomly assigned to the bees list 'n'. The Euclidean distances between each data object and all centres are calculated to determine the cluster to which the data object belongs (i.e. the cluster with centre closest to the object). Hence initial clusters can be constructed, while clusters centres are replaced by actual centroids of the clusters to define a particular clustering solution (i.e. a bee). This initialization process is applied each time new bees are to be created.

Step 2- The fitness computation process Fi is carried out for each site visited by a bee using clustering metric 'e', which is inversely related to fitness.

Step 3- 'm' sites with the highest fitnesses are designated as "selected sites" and chosen for neighbourhood search.

Step 4- Conducts search around the selected sites 'm', assigning more bees to search in the vicinity of the best 'e' sites.

Step 5- Selection of the best sites can be made directly according to the fitnesses associated with them. Alternatively, the fitness values Fi are used to determine the probability of the sites being selected.

Step 6- Searches in the neighbourhood of the best e sites- those which represent the most promising solutions are carried out recruiting more bees for the best 'e' sites than for the other selected sites. Together with scouting, this differential recruitment is a key operation of the Bees Algorithm.

Step 7- For each patch 'g', only the bee that has found the site with the highest fitness (the "fittest" bee in the patch) will be selected to form part of the next bee population. This restriction is purposefully introduced to reduce the number of search sites to be explored.

Step 8- Remaining bees in the population are randomly assigned around the search space to identify for new potential solutions.

Step 9- Each iteration ends with the colony having two parts to its new population, such as representatives from the selected patches, and scout bees assigned to conduct random searchs These steps are repeated until a stopping criterion 's' is met.

5. ALGORITHM

- Generate colour transformation structure.
- Convert colour values from RGB to the space specified in that structure.
- Apply K means clustering for image segmentation.
- Masking of green pixels (masking green channel).
- Eliminate the masked cells present inside the edges of the infected cluster.
- Convert the infected cluster from RGB to HIS.
- Generation of SGDM matrix for H and S.
- Calling GLCM function in order to calculate the features of it.
- Computation of texture statistics.
- Configure k-n (classifier) for recognition.

6. EXPERIMENTAL RESULTS AND DISCUSSIONS

This experimental test bed uses the data provided by director of economics and statistics of India [15], food and agriculture organization [19] and TNAU agro-tech portal [20] to build the crop knowledge base. The proposed CRY cluster algorithm is implemented using Clementine with knowledge base maintained and updated using Oracle database. Decision analysis and sustainability of crop yield is verified using the clustered dataset are in Fig. 5 & Fig. 6 yields of paddy and sugarcane are shown over set of region. In the graph mean value for a particular crop.

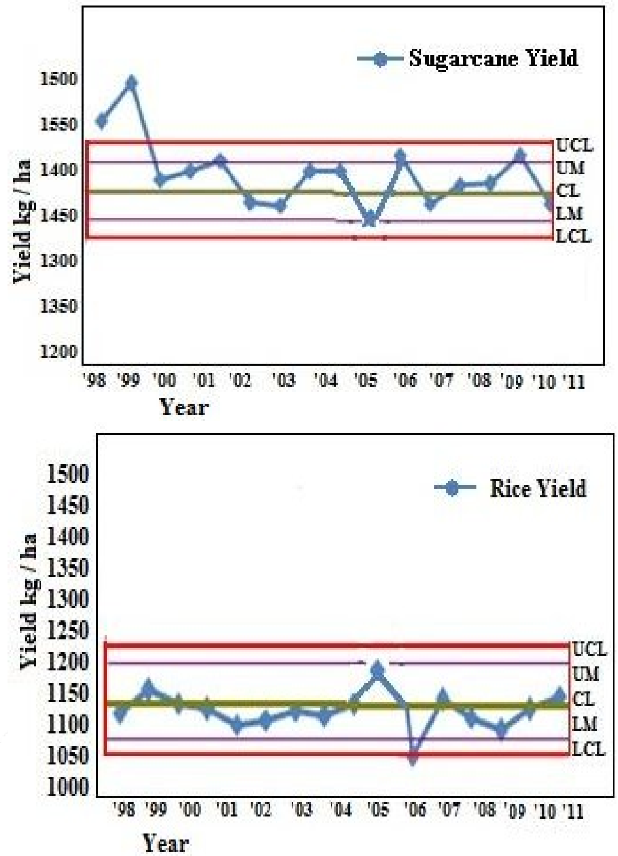

Figure 3 Rice Yield Analyses

7. PERFORMANCE EVALUTION:

In order to test the performance of the algorithm new agricultural datasets of records from crop knowledge base are tested with proposed CRY algorithm and cluster and regression tree algorithm. Cluster and regression tree algorithm is modeled in Clementine® and trained with large number of datasets from crop knowledge base.

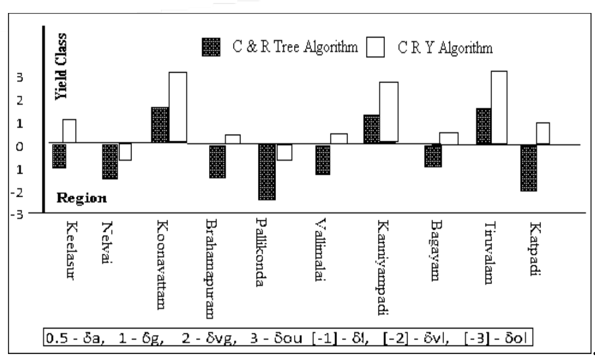

Fig. 4 Improved Yield of crop analysed using CRY algorithm

8. CONCLUSION:

- Agriculture is the most important

application area particularly in the developing countries like India.

- Use of information technology in agriculture can change the scenario of decision making and farmers can yield in better way.
- For decision making on several issues related to agriculture field; data mining plays a vital role.
- In this paper we have discussed about the role of data mining in perspective of agriculture field.
- We have also discussed several data mining techniques and their related work by several authors in context to agriculture domain. This paper also focuses on different data mining applications in solving the different agricultural problems.

9. REFERENCES

[1] Christine Preisach, Lars Schmidt-Thieme “Ensembles of Relational Classifiers” Springer Knowldge Information System vol.14, Issue-3, pg.no-249–272.

[2] Xiaoxin Yin, Jiawei Han, Jiong Yang, Philip S. Yu “Efficient Cluster Across Multiple Database Relations: A Crossmine Approach ” IEEE Transactions On Knowledge And Data Engineering, vol. 18, Is

[3] H. Guo, H. L. Viktor “Multirelational Cluster: A Multiple View Approach” Springer Knowldge Information System vol.17, Issue-3, pg.no-287–312.

[4] Takoi K. Hamrita, Jeffrey S. Durrence, George Ve L L Idis, “Precision Farming Practices” IEEE Industry Applications Magazine Mar- Apr 2009

[5] G. Delgado, V. Aranda, J. Calero, M. Sánchez-Maranón, J.M. Serrano,D. Sánchez, M.A. Vila “Using fuzzy data mining to evaluate survey data from olive grove cultivation” computers and electronics in agriculture 2009 vol.65 pg.99–113

[6] Aída Jiménez, Fernando Berzal ,Juan-Carlos Cubero “Using trees to mine multirelational databases” Data Mining and Knowledge Discovery ISSN: 1384-5810 pp-1-39

[7] Mucherino A, Papajorgji P.J, Pardalos, P, “Data Mining in Agriculture”, Springer 2009.

[8] B. Taskar, E. Segal, and D. Koller. “Probabilistic clustering in relational data” In IJCAI-01 pages 870--876, 2001.

[9] J. V. Stafford, “Implementing precision agriculture in the 21st century,’’ Computer Electronics in Agriculture. vol. 76, no. 3, pp. 267– 275, July 2000.

[10] H.Auernhammer, “Precision farming: The environmental challenge,” Computer Electronics in Agriculture., vol. 30, no. 1–3, pp. 31–43, 2001

[11] de Jong R, Bootsma A, Huffman T, Roloff G (1999) “Crop yield variability under climate change and adaptive crop management scenarios”. Final project report submitted to the Climate Change

[12]Xingang Xu, Bingfang Wu, Jihua Meng, Weifeng Zhou, "Research of crop yield models in China " Geoscience and Remote Sensing Symposium. IGARSS '05. Proceedings. 2005 IEEE International Volume: 2

[13]Luc De Raedt , Kristian Kersting, “Probabilistic Inductive Logic Programming” Probabilistic Inductive Logic Programming Theory and Applications vol. 4911, pp. 1–27

[14]Kazuhisa Tsunoyama, Ata Amin, Michael J. E. Sternberg , Stephen H. Muggleton,“Scaffold Hopping in Drug Discovery Using Inductive Logic Programming” J. Chem. Inf. Model., 2008, 48 (5), pp 949–957

[15]Director of economics and statics of India http://dacnet.nic.in/eands/latest_2006.htm

[16]Yi Lin “Support Vector Machines and the Bayes Rule in Cluster” Springer, Data Mining and Knowledge Discovery, 6, 259–275, 2002

[17] T.M. Mitchell, Machine Learning. McGraw Hill, 1997

[18]C.J.C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” Data Mining and Knowledge Discovery, vol. 2, pp. 121- 168, 1998.

[19]Food and Agriculture Organization http://faostat.fao.org/site/339/default.aspx

[20]TNAU agritech portal http://agritech.tnau.ac.in/