

Novel Data Mining Techniques for Agriculture Data Sets

Gagana H S¹ Gouthami H S²

^{1,2}Department of Computer Science

^{1,2}Mangalore University, Jnana Kaveri PG Centre, Kushalnagar, Karnataka, India

Abstract— Data mining is the ever-green growing field of research in the domain of computer science engineering. The concept of this paper shows about the role of the data mining in perspective of yield/crop analysis of farming. Mining is the process of extracting implicit and potentially needful data from the data information and it's the process of extract data information from huge data. The data mining concept include collections of data, data extractions, data analyzing and statistic of data. At present days the most important field in the real world is agriculture and farming is the major occupation and backbone of our country and In India more than 70% peoples are depending on agriculture. Data mining is domain concerned with the extraction of the useful and meaningful data from large databases or data warehouses. Now-a-days data mining in Agriculture is one of the trending research fields for researchers. Data mining's important task is to choose the proper data mining method. Here we using Clustering method, it's one among the oldest method used in Mining techniques. Cluster analyzing is the process of identifying data that are similar to each other. Crop prediction is the major task in agriculture field for high productivity.

Keywords: Data Mining, Agriculture, Crop Productions, Clustering, Yields, Soil

I. INTRODUCTION

The study of agriculture is known as agricultural science. Current day's data mining is one of the interesting fields for research. One of the important tasks in Data Mining are to selections of the right data mining method. Cluster method is the oldest method type using in Data Mining concept. Data cleaning, data integration, data selection, transformations of data, data evaluations are the major function or part of data mining technology. Clustering is technique of identifying data that are similar to each other. This method can help to understanding the data difference and data similarity b/w the data information. And the concept is called data segmentation & it is very helpful for data user to understand what's going on with in the dataset or databases. In the field agriculture soil type and rainfall is the major key elements to decide the crop production i.e. Cultivation of the crop is based on the soil and rain. Another important work in agriculture is selecting soil is based on the weather condition and also when to harvest the crop for the best cultivation.

II. EXISTING SYSTEM

In the existing system the yield and crop prediction is done with the single algorithm based on classification in which the single algorithm is used with the single dataset by that we cannot get the proper expected output. It has only the single output with the single dataset.

III. PROPOSED METHOD

At present day, a few no. of farmers are uses the current new technology, new methods, advanced tools & techniques in agriculture field for good crop productions. Data mining can be used for forecasting the future trends of agricultural processes. Now a day everyone is using smart phones with internet access. Farming is undergoing a digital revolution. Even small farmers are using smartphones and collecting the information about soil conditions (government provide Soil health card to farmers), weather conditions (ex: rain alert), new and effective advanced fertilizer data, variety of seeds available in shops, agriculture advanced vehicle info, foreign farming technology information etc. data are available in digital form. Using the aggregating and clustering method to manage and analyses these data in data mining technique.

Data mining concept is very popular and best method in the Area of agriculture. The major technique for data mining includes Associations rule, classifications, clustering and regressions. Data mining technique can be broadly classified into two models i.e. Predictive models consist of tasks such as classifications, Regressions, analysis of Time series, predictions, and Descriptive model comprises of clustering, summarizations, association rule, and sequence discovery. Clustering techniques used to figure out the crop data in different data mining techniques that are in the use for crop yield estimation in agriculture. Crop prediction is major task done in agriculture field for high productivity.

IV. DATA COLLECTION AND ANALYSIS METHODOLOGY

Data collection and analysis is a fundamental rule to discovering needful information from the datasets by applying the dissimilar MLT (machine learning techniques). Some types are the natural data collection methods shown in below:

- 1) Observations and analyzing
- 2) Surveys and Questionnaires
- 3) Documentations reviews
- 4) Face-to-face Interviews
- 5) Focus on groups
- 6) Case studies
- 7) Illustrated presentations and seminars
- 8) Collection of related ideas
- 9) Other visual representations, etc.

V. DATA MINING USING CLUSTERING TECHNOLOGY

Clustering means it is the process of grouping a set of physical or abstract object into classes of similar objects. Or it is a method used to group of similar data on the basis of feature. It does not have any labels. Cluster is a form represented as class of object. The different type clustering techniques are Hierarchical technique, Partitioning, Density

and model based Clustering, Grid based and Soft computing Methods, networks data and Cluster graph.

Data mining techniques to predict the type of soil is a primary key factor of agriculture. Soil type is an major role in agriculture. Data mining techniques in agriculture field plays an important role in Indian economy. Decision making is the main element for farmer to improve the crop yield, so data mining is the domain suitable for agriculture mining. Using clustering, the crop yield data are stored in the cluster wise formats. Multiple farming indicates to two or more crops together. Crop rotation refer to crop one after another in fixed rotation to maintain the soil fertility this takes an year to complete selection of crop based on the soil. The above data are stored like clusters and managed by the concept of clustering techniques.

Data mining techniques used in agriculture field for crop yield, it's based on parameters like Soil condition, weather condition, seeds and type of fertilizer, Diseases and pests, Harvesting method and other biological and geological factors etc. Less no. of application software is available for agriculture, soil is the important resource for agriculture to test the soil fertility, and with use soil dataset information can predict the crop yield for the present season. Soil testing and classification plays important role in fertility.

VI. CROP YIELD: KARNATAKA STATE

A. Introduction:

Karnataka is one of the states in India. Karnataka has 30 districts & 176 taluks with a total of 27,481 villages. Agriculture or farming is the heartbeats of the rural population. According to the 2011 population Census, agriculture supported 13.7 million workers, of whom 24 per cent were cultivators and 26 per cent were agricultural workers. A total of 123,100 square km of land is under cultivation in Karnataka, constituting 65 percentile of its total geographical area. 5 million hectare are under rain fed cultivation. The Karnataka state is the 2nd highest, after the Rajasthan state in India.

The main application of Data mining method in agriculture especially on soils, weather conditions, seeds and fertilizer etc. revising the current situations of pledge making and improving cultivation yields in a good way showing in the below figure 1.

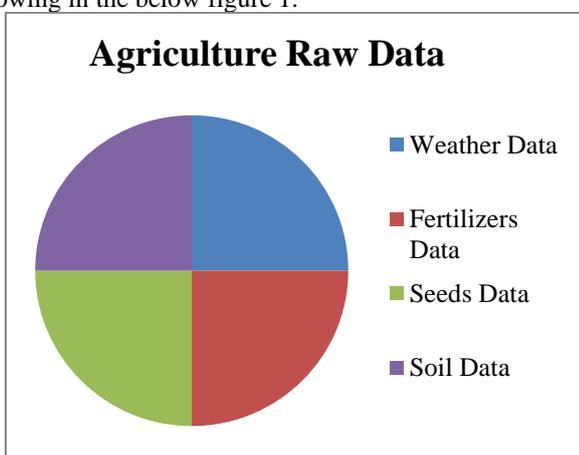


Fig. 1: Agriculture raw data Graph

The analysis of soils is an important role for resolution making on several issues related to agriculture field. The important component of soil health Mission are collections and analysis of soil samples, providing soil health data, prepare digital soil fertility maps and provide soil health card to farmers and finally creating awareness to our farmers about their soil status. Agricultural productions depend on qualities and availabilities of natural resources like water, soil and weather conditions. Farming outcomes can be evaluated by promoting conservation and sustainable use of these scarce natural resources through appropriate location and specific measures. The main component of soil is water and air, minerals, organics matter, and the proportions of which vary and together form a system for plants growth. Soil type are studying and classifies according to their uses. Mainly Soil Testing is a important stage of the discipline of what type of chemical/Fertilizer Using and Management.

B. Different types of soil found in Karnataka state

There are seven types of soil identified in all over Karnataka. These are listed below

- Red soil: Red gravelly loam soil, Red loam soil, Red gravelly clay soil, Red clay soil.
- Black cotton soil: gravelly soil, loose, black soil, basalt deposits.
- Lateritic soils: Lateritic gravelly soil, Lateritic soil.
- Black soils: Deep black soil, Medium deep black soil, Shallow black soil.
- Alluvio-Colluvial Soils: Non-saline, saline.
- Forest soils: Brown forest soil.
- Coastal soils: Coastal laterite soil, Coastal alluvial soil

NATURE OF SOIL	%	CROPS GROWN
Black cotton soil	29.6%	Cotton, Jower, Wheat, Sugarcane, Linseed, Gram, Fruit & vegetable.
Red Soil	28%	Wheat, Rice, Millets, Pulses
Laterite Soil	2.62 %	Unsuitable for agriculture due to high content of acidity and inability to retain moisture. Cashew and tropical grow well on it.
Forest Soil	7.94%	Plantations Crop like tea, coffee, spices and tropical fruits.
Peaty & other Organic Soil	2.17 %	Not conducive to cultivation

Table 1: Different Type of Soil with % and Crops Data

Farming in the Karnataka State is covering over 3 season, these are kharif season (July-October), rabi or winter season (October-March), and summer season (April-June). The weather conditions of the State permit the cultivation of different types of crops. The major crops are rice, sorghum (jowar), pearl millet (bajra), maize, wheat, finger millet (ragi), and other minor millets, etc.

C. Nutrients Status of Soils

Some major key elements of farming is varieties of seeds, application of fertilizers and assured irrigation has resulted in substantial growth in food grain production. However, extensive use of fertilizers and use of improved varieties has resulted in greater mining of soil nutrients which intern resulted in depleting soil fertility, decline in water table, and decrease in organic matter content and deterioration in soil health. For example the crops are follows some pre-determined conditions like its PH value, temperature, average rainfall, type of fertilizers, etc. and the different types of crops (like Cotton, Sugarcane, Wheat, rice, Bajra.) with its basic functions.

D. Soil Testing Program in Karnataka

There are 29 static soil testing laboratories are working in Karnataka State with analyzing capacity of 2 lac soil samples per annum. The basic functions of soil samples are analyzed for pH, EC, Major nutrients for examples like nitrogen, prosperous, potassium, secondary nutrients like Sulfur and Magnesium and micronutrients such as Zn, Fe, Mn, Cu and Boron. Water samples are also analyzed for pH, EC, Sodium Absorption ratio (SAR), Ca, Mg, Co₃, HCO₃, Cl₂ and the quality of water for irrigation. Well qualified and experienced personnel are working as analysts to ensure the quality of analysis. The quality analyst test the soil with various methods in soil testing center and produce the proper results and given to the farmer in form of soil health card as well as same data are collected from raitamitra website, farmer can access the data with internet. The progress achieved in respect of soil sample target and achievement for the present year. After analysis of soil samples suitable recommendations are given to the farmers to enable them for application of fertilizers in order to get better yields besides maintaining of soil fertility.

VII. WEATHER CONDITIONS

Weather change is a highly effected part in agriculture sector or farming. Present days a very less no. of farmers are uses the advanced techniques, methods, tools of agriculture for good productions. The mining technique can be used to predict the future values of agriculture related process. Data mining of this process results in the finding of new patterns in huge datasets. Real time climate change data is difficult to analyze and manage so we are using various algorithm and related statistical methods are used to maintain agriculture data. The crop loss/gain depends on climate of that region. Real time weather data can helps to the farmer for planting the particular and variety of crops on time. The government weather forecasting department provides correct climate data (data like: rainy, wind, sunny, snow) to farmers. The main purpose of mining method is to extractions of knowledges from the previous datasets. Data Scientist are collecting and evaluating the soil data to improve their state of how soil contributed to weather condition change by releasing greenhouse gases, as well as how soil data might aid in adapting to climate or weather condition change. Collections and analyzing this type of data information's is risky, but data scientist believed it could full fill in

knowledge gaps associated with the co-ordination b/w soil and climate change.

VIII. RESULTS

In this article Weather condition (Rainfall, Temperature, etc.), Seeds data, Fertilizers data and Soil health are the most important things for good crop yield production.

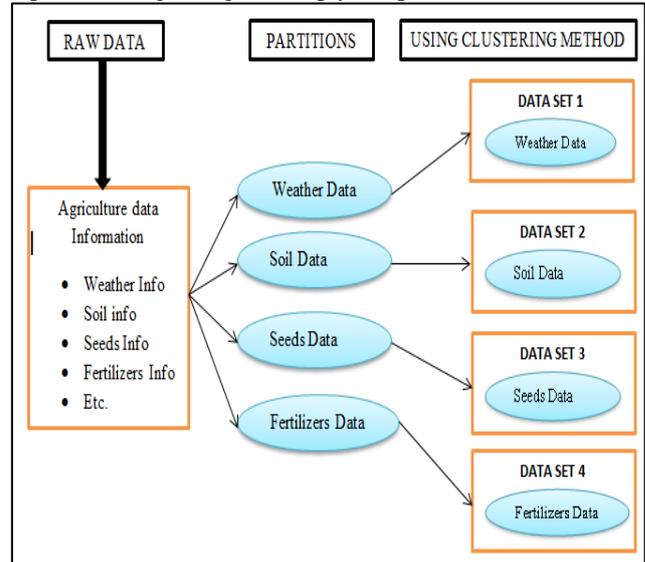


Fig. 2: Applying clustering technique to agriculture data

The above architecture represents the raw agriculture data is partitioned into sub parts using clustering technique, the different data information are stored in individual datasets like dataset1 to dataset4. Agriculture field plays the main part in developing countries like India.

The after using clustering method, the data is partitioned into four sets like datasets 1 to 4. This is shown in the figure 3. The datasets are indicated in different colors like sky blue indicate weather dataset, green indicates seeds datasets, and brown indicates fertilizer dataset, blue indicates soil datasets. All data are stored in data warehouse and easily accessible by the authorized user.

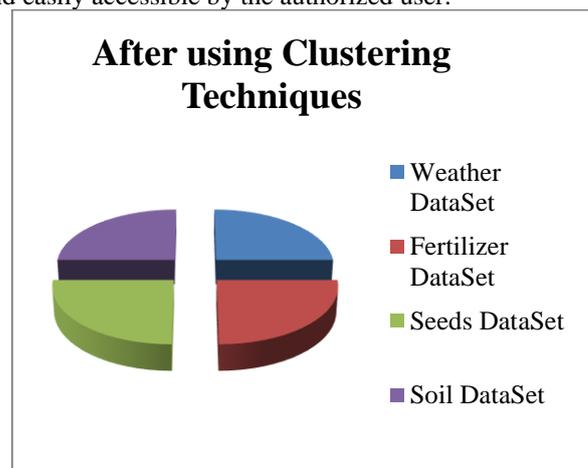


Fig. 3: Data sets after using Clustering method

IX. CONCLUSION AND FUTURE SCOPE

Data mining technology is the process of extracting the useful information from the raw data in such a manner that the system is able to achieve good performance with

minimum error/failure rate probability. Data mining techniques can be developed to perform predictions on possible future crop and yields. Using this for good management with amount of available information it is easier to make the decision on agriculture. This research paper deals with the efficient data mining technology to achieve high crop yield production with minimum error/failure rates. Our future method deals with the effective & efficient clustering technique of data extraction. Finally the classification approach and the performance are evaluated. So our proposed approach is able to provide the useful data to our farmers.

REFERENCES

- [1] <http://raitamitra.kar.nic.in>
- [2] https://en.wikipedia.org/wiki/Data_mining
- [3] www.karntaka.gov.in
- [4] <http://www.yourarticlelibrary.com/soil/soils-of-india-six-different-types-of-soils-found-in-india/12779/>
- [5] N. Hema Geetha, "A Survey on Application of data mining techniques to analyze the soil for agriculture purpose", IEEE, 2016
- [6] Shastry, H. A. Sanjay, and M. Hegde, "A parameter based AN-FIS model for crop yield prediction", In Advance Computing Conference (IACC), 2015 IEEE International, pp. 253-257. IEEE, 2015.
- [7] S. Veenadhari, B. Misra, C. D. Singh. "Machine learning approach for forecasting crop yield based on climatic parameters." In Computer Communication and Informatics (ICCCI), 2014 International Conference on, pp. 1-5. IEEE, 2014.
- [8] B.D. Lobell, M. B. Burke, "On the use of statistical models to predict crop yield responses to climate change", *Agricultural and Forest Meteorology* 150, no. 11, pp: 1443-1452, 2010.
- [9] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavva, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture." In Advanced Computing (ICoAC), 2016 Eighth International Conference on, pp. 32-36. IEEE, 2017.
- [10] T. Palosuo, K. C. Kersebaum, C. Angulo, P. Hlavinka, M. Moriondo, J. E. Olesen, R. H. Patil et al. "Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models." *European Journal of Agronomy* 35, no. 3, pp: 103-114, 2011.
- [11] Chipanshi, Y. Zhang, L. Kouadio, N. Newlands, A. Davidson, H. Hill et al. "Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape." *Agricultural and Forest Meteorology*, pp: 137-150, 2015.
- [12] S. Asseng, F. Ewert, P. Martre, R. P. Rötter, D. B. Lobell, D. Cammarano, B. A. Kimball et al, "Rising temperatures reduce global wheat production." *Nature Climate Change* 5, no. 2, 2015.
- [13] B.D. Lobell, M. B. Burke, "On the use of statistical models to predict crop yield responses to climate change", *Agricultural and Forest Meteorology* 150, no. 11, pp: 1443-1452, 2010.
- [14] <http://www.originlab.com/doc/Origin-Help/Multi-Regression-Algorithm>.
- [15] Tejas S. Mehta, Dr. Dhaval R. Kathiriya, "Survey of Data Mining Techniques in Precision Agriculture", IJSR, 2015
- [16] https://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm