ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

An Analytical approach on Data Mining in Agri-Clinics-A Survey

[1] K.P.Mangani, [2] Dr. R.Kousalya
[1] Ph.D Scholar, PG & Research Department of Computer Science, [2] Research Supervisor, Professor & Head, Department of Computer Applications.
[1][2] Dr.N.G.P Arts and Science College, Coimbatore-641048, India.

Abstract: **-- "Most things except Agriculture can wait" is the statement by Jawaharlal Nehru. To utilize the agriculture in a better way this paper handshakes with the Information technology and establish Agri Clinics which are envisaged to provide expert advice and services to farmers on various aspects like cropping practices, plant protection, crop insurance, post-harvest technology, clinical services for animals, feed and fodder management and crop prices to enhance productivity of crops and increase the incomes of farmers. The goal of this survey provides a review of different types of data mining approaches and problems being addressed in agriculture related fields .The agriculture related data are recorded but they are not digitized so they are integrated in standardized format which deals with big data that bridge the knowledge of the data to the yield estimation and gives an analytical approach to support decision making system and recommendations for farmers in agriculture related fields.**

Key words: **Data mining, Classification, support-vector machines, Regression, Agriculture, Bigdata.**

I. INTRODUCTION

To improve the farmers socio economy in agriculture productivity this paper uses data mining concepts and analytics to explore its application to change the farmer's perception and lifestyle through the technology. To apply the data mining concepts the system should collect data of all kinds by farmers and Agropreneurs and make those datasets linked together or being used to extract information which could help to improve the agriculture industry. Huge volumes of data are being collected throughout the food production chain from agriculture to the end product. This enormous data from multiple sources could present a "Big Data" opportunity for agriculture. Big Data is not about just one piece of equipment; it's a new approach to bring together everything that goes into producing the finished product and applies permaculture[15] which is an intelligent system uses free sustainable energies and resources. It also uses the mobile applications that involves real-time data collection of production performance. The Predictive analytics in the data mining can be used to make smarter decisions in agriculture by collecting real-time data on weather, soil type,Average rainfall, land fertility, demand and availability of labour which is known as precision agriculture. Bigdata plays an important role in precision agriculture for managing real-time data analysis with structured massive data. The data Analytical efficiency and throughput would be a challenge with the substantial increase in size of big data. The unstructured data received from different agricultural sources would contain multiple dimensions and the data mining techniques are applied for performing critical analysis. Analytics helps to allow the agriclinic modeling to adopt and obsorbs the data with ICT Technologies and focus on the globalized need of Agropreneurs.

II. RELATED WORKS

From the Research article [2] V. Ramesh and K. Ramr explains the comparison of different data mining classifiers and the outcome of this research could improve the management and systems of soil uses throughout a large fields that include agriculture, horticulture, environmental and land use management.

In the case study on "Rainfall variability analysis and its impact on crop productivity"[1] the authors collected the weekly rainfall data and number of rainy days recorded at the main Dry farming research station for 39 yrs. The correlation and regression studies were worked out using rainfall(x) as independent variable and yield(y) as dependent variable to derive information on rainfall-yield relationship and to develop yield prediction model for important crops.
From the research article " Big Data analysis on geographical segmentations and resource constrained scheduling of production of agricultural commodities for better yield[3] " the authors describes the procedure to find out the crop

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

associations and patterns under climatic influence for each geographical segmentation, to give better yields and implements the crop rotation and irrigation schedule.

From "Big Data Analytics Architecture for Agro Advisory System"[4] ,the author propose an open source, cost-effective and scalable big data analytics architecture for an Agro advisory system.An Analytic framework for big data application development is built and implemented in this paper. The author suggests a prototype application for crop yield prediction that is implemented for cotton crop in Ahmedabad district, Gujarat, India. All the above works are done with the methodology.

III. METHODOLOGY USED

A. Agriclinics

Agriclinics[11,12] tells how the product is produced and processed along the entire value chain from land to retail outlets like traditional grocery stores, restaurants, food service providers.Such products includes additive or antibiotic free, organic products, locally and/or family-farmer grown, animal treatment/welfare production practices, sustainable production
processing,distribution systems etc. Data and information systems are required to monitor and measure these processes and activities at each stage of the food supply chain. This data and information must be tagged or linked to the physical product (boxes of grains, cuts of meat, etc.) that flows along that supply chain so that final product can be marketed and certified to fulfill the consumer's demand and desire. To augment the production the big data technologies is applied in the following sub-sectors of agriculture

• Green Houses

Analyzing data from connected devices,sensors and disease forecasting with climatic reports from multiple devices that are used in agri-green houses.The big data analytics is applied which helps the Agripreneurs to predicate and maintain the crop yield .

• Food Supply chain

Since India ranks 97th of 118 in global hunger index the Analytics helps to solve the hunger and nutrition problems prevailing in india with the help of sensors and smart technologies. It also forecasts what information insights are essential to supply chain to increase production efficiency and profitability that focuses on food safety and quality from producer to consumer and the alternative control technologies is to be used.

• Livestock Production

The information technology helps to serve the growth condition, determinants and constraints on the growth that can be executed . The analytics with the IOT (Internet of Things) technology helps to use real time data which enhance and control the biological growth in accordance with the temperature. It also deals with the output units of production and prediction with the input units like nutrition, location and spatial details etc .

B. Agro-functional units

Agrimetrics[13] supports organisations across the agri-food value chain to develop their data modelling and analytics capabilities to estimate the productivity of different crops. These activities are undertaken by the government departments in order to monitor the progress of the agriculture sector & provide insurance to the sector. Revenue, agriculture & Economics & Statistics departments are jointly involved in the estimation process. Researchers & many other agencies uses the data generated by the Government departments and strengthen the Agri-Marketing infrastructure with grading and standardization methodology.For using agricultural product as raw material the previous historical crop yield information is important for supply chain operation of companies engaged in industries. Livestock, food, animal feed, chemical, poultry, fertilizer pesticides, seed, paper and many other industries uses these information for an accurate estimate of yield production. some activities are done at agriclincs are

- Cropping practices
- Plant protection
- Crop Insurance
- Post-harvest Technology
- Clinical services to animals
- Feed and fodder Management

This paper discusses data mining techniques such as regression, classification and association which is a process model for analyzing data and describes the DSS (Decision Support system) needed for modeling Agriclinics. It also includes the following processes

C. Data cleaning with SVM

The data could be in the form of aggregates or duplicates, we need to clean the available data set. Once Data is collected and identified that there are lack of information, the data that must be cleaned to reveal the original context of data, preserving the integrity of the Data. The Studies conducted by agricultural researchers have shown that the attempts of crop yield maximization through pesticide state policies have

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

a negative correlation between pesticide usage and crop yield. The Study [14] had shown that how agricultural data including weed scouting, pesticide usage and meteorological recordings is useful for optimization of pesticide usage and classify the weeds from crops to form clusters. These clusters reveal interesting patterns of farmer practices helps to identify the crop patterns and composition of pesticide usage.

D. Hadoop Environment

The data cleaning process and the prediction can be done in the hadoop environment which is a Big data Analytics tool and platform for developing and running large massive data.The core of Apache Hadoop consists of a storage part, known as Hadoop Distributed File System (HDFS)[15] and a processing part which is a MapReduce programming model. Hadoop splits files into large blocks and distributes them across nodes in a cluster. HDFS provides storage services for massive amounts of data, while MapReduce provides computing services for data. HDFS can handle very large files, has been able to handle the PB (Petabyte) level of data, and has a high degree of fault tolerance characteristics.Distributed programs written on the basis of the MapReduce model can distribute a task to a cluster of large numbers of computers and process large numbers of data sets in a highly fault tolerant manner to extract the needed features and derive the yield results and comparable results with the previous history. It also suggests the better production level of farming and the alternate model for the problems prevailing in the agriculture and in its subsector patterns.

IV.APPLICATION OF DATA MINING TECHNIQUES

The more common model functions in current data mining practices which may apply for agriclinics are

1.Classification : It classifies a data item into one of several predefined categorical classes. the k-means approach is used for classifying soils in combination with GPS-based technologies [6], k-means approach to classify soils and plants and SVMs to classify crops [5] from weeds.

2.Regression: It maps a data item to a real valued prediction variable . A crop yield prediction model is built and deployed on Spark. The Multiple Linear Regression (MLR) model[7,4] is used for cotton crop yield prediction based on weather indices like average temperature and rainfall.

3.Clustering: It maps a data item into one of several clusters, where clusters are natural groupings of data items based on similarity metrics or probability density models and helps to cluster the crops based on the criteria defined[8] and gives the availability and demand of crops.

4.Discovering association rules: describes association relationship among different attributes like crop pattern associations and predicting the alternate crops in case of crop failure[9].

5.Scheduling Algorithm for crops based on geospatial data : scheduling algorithm on Big data analytics helps to estimate the irrigation schedule for the major field crops during the period of peak water demand; the schedules are given for different soil types and different climates for the given geographies[3]

6.Summarization : provides a compact description for a subset of data and give a picture of yield production .

7. Sequence analysis: It models sequential patterns, like time-series analysis. The goal is to model the states of the process generating the sequence or to extract and report deviation and results over time in terms of graphs.

V. DATA ANALYTICS IN AGRICLINICS

The multidisciplinary approach to integrate agriculture with spatial data mining by methods to extract interesting and regular knowledge from large spatial databases of agriculture and its related fields. The workflow of Big Data Project is shown in fig-a that gives a formulation of problem with data collection and how to store , transfer and analyze the data. It finally gives the reports with data visualization.Now we discuss how we apply the Big data Analytics in the following sections of agriclinics.

- Big data analytics [16] on geographical segmentation of agricultural area is done to examine the spatial distribution of crops and to investigate the subsectors of agricultural activities that are carried out on each geographical segment. Based on the data of a particular segment, the big data is applied for analyzing the demand and need in the variations of crops that are suitable for that segment and the types of yield received by the producers.

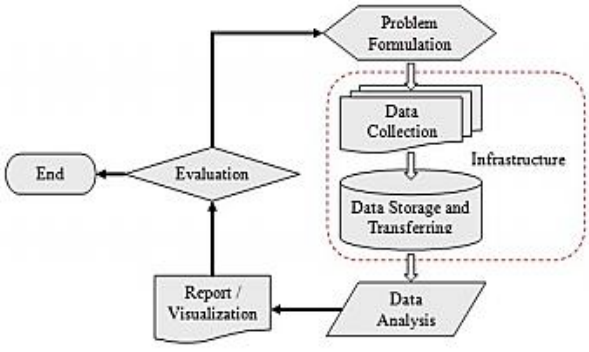

Fig. a: Workflow of Big Data Project

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

- The challenges in the sustainability of agriclinics leads to the possibility for the farmers or agripreneurs to adopt a new combination on agricultural crop patterns when it gives better yield than the current ones, that are predicted by data analytics.
- The yield performance of various crops in a country or region is not uniformly distributed. There are interregional, intra-regional, intra-village and intra-farm variations in the production and productivity of different crops. Based on geographical and farming constraints the yield may differ and the system makes a study to find out the determining factors. Our Big data analytics on agricultural data, facilitates a diagnosis at the micro level (household and field level) causes of existing agricultural backwardness and then recommends the suitable strategies to enhance the productivity.
- In agribusiness the agriculture has been considered as an industry where the data centric analysis makes an attempt to identify the impediments which are coming in the way of making this occupation as an agribusiness with increasing profit by establishing green houses as a structured project.
- Big data applications in agriculture can also be used for weather analysis and prediction , climate change impact, determining change in cropping patterns, creation of benchmarking datasets, data integration and interoperability of data generated by sensors and machines used by greenhouses in precision agriculture.It also utilizes Agri-robots, text, voice, image and video processing for Agro-Advisory Systems[4], irrigation and water distribution . It helps to identify disease,fertilizers, pesticides and its schedule and usage, geo-spatial analysis for agriculture, real time reporting, finance and crop insurance, marketing, supply chain management and agribusiness.
- Data mining techniques are applied with the personalized agricultural extension systems (AES) which is developed into farmer-specific AES (FAES) and cluster-specific AES (CAES) by considering the delivery of agriculture scientist advice at the farmer-level and cluster-level respectively. In FAES, personalized advice is delivered to each farm by setting of IT Kiosks in rural areas for access to various agriculture related portal.In CAES, the farms are grouped into clusters and advice is delivered for each cluster[10].
- Analytics can be made on setting up Storage Structure and Rural Godowns and Retail marketing outlets for processing Agri-products to give value added Products.

IV. COMPARATIVE RESULTS

A. Scheduling Algorithm model

The Table shown in figure 1 is the sample output that uses the scheduling algorithm on Big data analytics to estimate the crop rotation and irrigation schedule for the major field crops during the period of peak water demand; the schedules are given for different soil types and different climates for the given geographies

Geography 1:
Soil Type : Red soil,Region:Tropical,Fertility :67%,
Rainfall : 10 cm, Avg. Temp : 35 deg. Cel.

1st year	Fall wheat & rye	Spring oats peas barley beans lentils	Fallow
2nd year	Fallow	Fall wheat & rye	Spring oats peas barley beans lentils
3rd year	Spring oats peas barley beans lentils	Fallow	Fall wheat & rye

Figure 1: Sample output on crop rotation schedule for a geographical region of Rajasthan

The time series graph is drawn below in figure 2 to represent the number of months and crop's schedule .The bar graph represents such crop limited within 12 months for rotation.

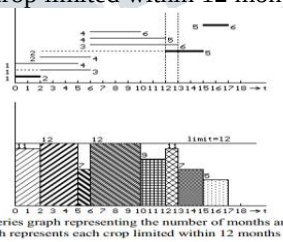

Figure 2: The time series graph representing the number of months and crop's schedule. The bar graph represents each crop limited within 12 months for rotation

B. Multiple Linear Regression (MLR) model

It is used for cotton crop yield prediction based on weather indices like average temperature and rainfall and also used to predict yield using a single weather parameter. The yield prediction results are shown in table 1.It also shows the difference in the results obtained.

TABLE I. EXPERIMENTAL RESULTS

R-squared	Actual yield in 2011 (Kg/Hactor)	Predicted yield in 2011 (Kg/Hactor)	Difference (Kg)
0.7814	283	193	90

C. Association mining and prediction model

Based on the above two model discussed, this paper applies the data mining techniques of association and prediction model to predict the fertility and pest control rate with respect to cropping pattern[17] which is shown in the below table.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

1) The below table shown in figure 3 gives the crop pattern schedule with seasonal rainfall, temperature and soil type.It uses the association mining algorithm to find the asssociate crops for the suitable crop patterns and regression technique that gives the predicted fertility rate and pest control rate .

Soil type:black-alkanic Rainfall= 15 cm Temp:32 deg.cel
Soil Fertility :60% pest control: 30%

Cropping pattern	Crop combination	Predicted fertility rate	Predicted pest control rate
Mixed Intercropping	Sorghum Pearl millet cowpea	60%	30%
Row intercopping	Maize+Green gram (1:1) Maize+black gram (1:1) Groundnut+Redgram (6:1)	54%	40%
Alley cropping	Soyabean or hay planted in between rows of black walnut or pecan trees	62%	45%
Strip cropping	Groundnut +Redgram(6:4)	63%	35%

Figure 3: Sample table on crop pattern schedule for crops

The bar graph shown in figure 4 represents the fertility rate with different types of cropping pattern which is taken as annually .

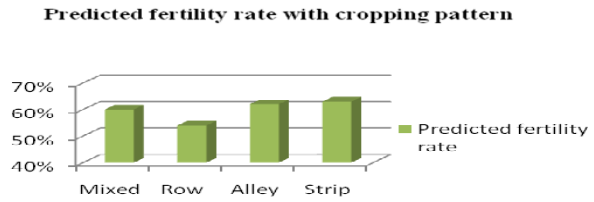

Figure 4: Predicted fertility rate with cropping pattern

The Pie chart shown in figure 5 represents the pest control rate with the corresponding cropping pattern taken annually.

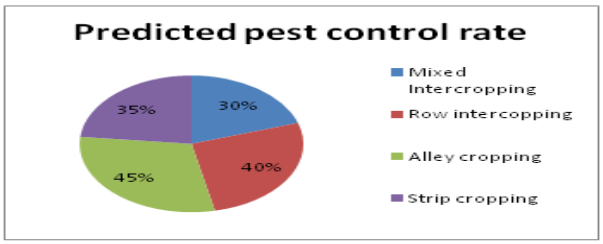

Figure 5: Predicted pest control rate with cropping pattern

V. CONCLUSION

In this paper we discussed about how data mining algorithms and the data analytics are applied over agriclinics that helps to implement the model which reduce the technological gap between rural communities and information through recommendations and decision support system. And we have discussed with the variability of crop patterns with pest control and fertility which is given in the table to understand the predictable model of data mining.

REFERENCES

[1] D.R. Mehta, AD. Kalola, D.A Saradava and AS. Yusufzai ,"Rainfall variability analysis and its impact on crop productivity",Indian agriculture research journal 2002 29,33.,8) Main Dry Farming Research Station, GAU, Targhadia - 360 003, India

[2] V. Ramesh and K. Ramr ,"Classification of agricultural land soils: A data mining approach" International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 3 No. 1 Jan 2011 379

[3] Senthilvadivu,Vinu Kiran,Prasanna Devi , Manivannan,"Big Data Analysis on geographical segmentations and resource constrained scheduling of production of agricultural commodities for better yield".

[4] Purnima Shah, Deepak Hiremath and Sanjay Chaudhary,"Big Data Analytics Architecture for Agro Advisory System", 2016 IEEE 23rd International Conference on High Performance Computing Workshops

[5] Karimi Y, Prasher SO, Patel RM, Kim SH," Application of support vector machine technology for Weed and nitrogen stress detection in corn ", Computer Electronics Agriculture 51:99–109, 2006

[6] Verheyen K, Adriaens D, Hermy M, Deckers S High-resolution continuous soil classification using morphological soil profile descriptions. Geoderma 101:31–48, 2001.

[7] D Ramesh, B.Vishnuvarthan, "Analysis of crop yield prediction using data mining techniques",International journal of Research in Engineering and Technology,

[8] A Mucherino , Petraq Papajorgji , P. M. Pardalos, "A survey of data mining techniques applied to agriculture"

[9] Farah Khan,Dr.DivakarSingh, "Knowledge discovery on Agricultural dataset using Association Rule Mining", International Journal of Emerging Technology and Advanced Engineering,Vol 4,Issue 5,2014

[10] E-agriclinic: an IT based approach for cost-effective and personalized agricultural extension system,P.Krishna Reddy, G.Syamasundar Reddy, B.Rammohan Rao,D.Guruva Reddy, BV Ratnam, P.Vasudeva Reddy and M.Kumaraswamy,

International Institute of Information Technology,Gachibowli, Hyderabad, Andhra Pradesh, India

[11] Website of www.agriclinics.net.

[12] Shoji Lal Bairwa,Present Status of Agriclinics and Agribusiness Centers Scheme in India: An Analysis , IJSRM,volume 2,issue 9 september

[13] Website of www.agrimetrics.co.uk

[14] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh ,Data mining Techniques for Predicting Crop Productivity – A review article ,IJCST Vol. 2, Issue 1, March 2011

[15] Website of http://www.permaculturevisions.com/

[16] D.P. Acharjya, Kauser Ahmed P, A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, International Journal of Advanced Computer Science and Applications,Vol. 7, No. 2, 2016

[17]https://abhijeetgautam.wordpress.com/2015/09/21/cropping-pattern- in-india/