IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 31-35

Survey of Data Mining Techniques and Applications of New Techniques in Agriculture

Mythili. R, Pradeep Raj. D
Assistant Professor, Department of Computer Applications, Indo-American College, Cheyyar Thiruvannamalai dt.Tamilnadu, India.
Assistant Professor, Department of Computer Applications, Indo-American College, Cheyyar Thiruvannamalai dt, Tamilnadu, India.

Abstract
Data Mining is the process of finding new patterns from large data set and converting them into meaningful information which can be used for any business logic. This paper focuses on the survey of Data Mining techniques used and application of new techniques for the better crop yield production. Application of Data Mining techniques in agricultural fields can revolutionize the crop yield production and its analysis. Data mining techniques such as K-Mean and K nearest neighbor are used in this paper. The application of Data Mining techniques enables us to automate the prediction and analysis of various problems and helps farmers to make critical farming decisions based on the conditions, soil fertility, crop duration, disease detection and other crucial factors that can result in low yield production. The hidden pattern in these data helps us to explore new possibilities in maximizing the yield production.
Keywords: *agriculture, data mining techniques.*

I. Introduction

Data Mining is the process of extracting useful and important information from large sets of data. This paper describes an overview of Data Mining techniques their applications agricultural field. Agriculture is dependent on some factors like soil, climate, cultivation, irrigation, fertilizers, temperature, rainfall, harvesting, pesticide weeds etc. Many industries use agricultural products as raw material, livestock, food, animal feed, chemical, poultry, fertilizer, pesticides, seed and paper. An accurate estimate of crop production and risk helps these companies in planning supply chain decision like production scheduling. Business such as seed, fertilizer, agrochemical and agricultural machinery industries plan production and marketing activities based on crop production estimates [1, 2]. There are 2 factors which are helpful for the farmers and the government in decision making namely:

1. It helps farmers in providing the historical crop yield record with a forecast reducing the risk management.
2. It helps the government in making crop insurance policies and policies for supply chain operation.

II. Literature Review

A variety of data mining techniques are used in different fields to enhance or gain knowledge. Now a day the researchers, data analysts and scientists has more concentrated on how mining and machine learning techniques are used to analyze various soil profiles to enrich the field of agriculture [1].

Crop Productivity Mr. Narsi Reddy Gayam stated in his research learning ―A study of crop yield distribution and crop insurance‖ which takes the input data from INDIA relating sugarcane and Soybean. After his research he concluded crop yield are not normally distributed [2].

Application of DM Techniques in Agriculture Dr. Bharat Misra, et al., [3] observed the research studies on application of data mining techniques in the field of agriculture. Some of the techniques, such as ID3 algorithms, the k-means, and the k-nearest neighbor, artificial neural networks and support vector machines applied in the field of agriculture were presented. Data mining in application in agriculture is a relatively new approach for forecasting / predicting of agricultural crop/animal management.

Spatial DM Vashovardhankelkar, et al,[4] Surveyed and says that data selection is the data relevant to the analysis is decided and retrieved from the various data locations. Data preprocessing is the process of data cleaning and data integration is done. Data Mining is the crucial step in which clever techniques are applied to extract potentially useful patterns. The decision is made about the data mining technique to be used. Interpretation and Evaluation is interesting patterns representing knowledge are identified based on given measures. The discovered knowledge is visually presented to the user [5].

Soils Clustering A. Banumathi, A. Petalakshmi (2012) has explained how data mining can be applied on large datasets to discover patterns using the method of clustering. They have analyzed the Fuzzy C-Means

algorithm and have made inferences that initial seed value selected either sequentially or randomly have effect on the value of the ensuing cluster. Amrender kumar (2004) [6] explains the techniques for forecasting of crops. The various data mining and neural network techniques are used for forecasting the result. These techniques are decision tree, rule induction, navie-bayes, neural network, ANN, radial basis function, recurrent network, multilayer perceptron and RBF techniques and predict the result. Anwiti Jain, Anad Rajavat, Rupali Bhartiya (2012) [7] has explained about the clustering mechanism which is an unsupervised technique of learning. They used optimization formulation of problem in designing the algorithm together with novel iterative method.

III. Existing Data Mining Techniques In Agriculture

Neural networks

Sanjay D.Sawaitul et al., focuses the information about weather and are observed and stored. The recorded parameters are used to forecast weather. If there is a change in any one of the recorded parameters like wind speed, wind direction, temperature, rainfall, humidity, then the upcoming climatic condition can be predicted using artificial neural networks, back propagation techniques. The increase in signal range will work in large areas as well[8].

Somvanshi, V.K. et al., deliberate the modeling and prediction of rainfall using artificial neural networks and Box- Jenkins methodology. Other applications of artificial neural networks in hydrology are forecasting daily water hassle and flow forecasting[9].

K-means

K. Verheyen et al., Data Mining is the process of discovering meaningful patterns and trends by shifting through huge amount of data, using pattern detection technologies as well as statistical and mathematical techniques. As an example, the K-Mean approach is used for classifying soils in combination with GPS based techniques[10].

Fuzzy set

Jagielska et al., describe applications to agricultural related areas. Such as Yield prediction is a very important agricultural problem. Any farmer might be interested in knowing how much yield is expected. In the past, yield prediction was achieved by considering farmer's experience on particular field, crop and climate condition. We have discussed additional information about data like probability in probability theory, grade of membership in fuzzy set theory[11].

Tellaeche et al., detecting weeds in precision agriculture. The paper summarize an automatic computer vision system for the detection and differential spraying of Avena sterilis, a toxic weed growing in cereal crops. With such purpose it have been designed a hybrid decision making system based on the Bayesian and Fuzzy k-Means classifiers, where the a priori probability required by the Bayes framework is supplied by the Fuzzy k-Means[12].

Decision tree

Veenadhari, S. Influence of climatic factors on major kharif and rabi crops production in Bhopal District of Madhya Pradesh State was considered. The findings of the study revealed that the decision tree analysis indicated that the productivity of soybean crop was mostly influenced by comparative humidity followed by temperature and rainfall. The decision tree analysis shows that the productivity of paddy crop was mostly inclined by Rainfall followed by comparative Evaporation and humidity. The result of decision tree were confirmed from Bayesian classification. The rules formed from the decision tree are useful for identifying the conditions intended for high or low crop productivity[13].

K-nearest neighbour

Altannar Chinchulunn et al.,The k-nearest neighbor classification algorithmic rule may be divided into 2 phases: coaching section and testing section. Bermejo associated Cabestany urged a reconciling learning algorithmic rule to permit fewer information points to be utilized in coaching information set. Several different techniques are projected to scale back procedure burden of k-nearest neighbor algorithms[14]. Rajagopalan and U. Lal A number of studies have been carried out on the application of data mining techniques for agricultural data sets. For example, the K-Nearest Neighbor is applied for simulating daily precipitations and other weather variables[15].

Support Vector Machine

S.Veenadhari et al.,The main plan of Support Vector Machine (SVM) is to classify information samples into 2 disjoint categories. The essential plan behind is classifying the sample information into linearly severable.

Support Vector Machine (SVM) area unit a group of connected supervised learning ways used for classification and regression[16]. Tripathi, S et al., The SVM-based DM is applied to future climate predictions from the second generation Coupled Global Climate Model (CGCM2) to obtain future projections of precipitation. The results are then analyzed to assess the crash of climate change on rainfall over India. It is shown that SVMs provide a promising alternative to conventional artificial neural networks for statistical downscaling, and are appropriate for conducting climate impact studies[17].

IV. Data Mining Based On Bayesian Networks For Agriculture

Example of application of Bayesian networks: The data set contains 2 000 cases, part of the data and the variables with their status showed in Table1. The domain problem has 6 variables; each of them has several attributes. And one variable responds to one node in the model respectively.

The variables and their implications describe as follows.

1) Soil texture: the relative proportions of sand, silt, and clay particles in a mass of soil.

2) Organic matter: consists of plant and animal material that is in the process of decomposing.

3) Gradient: A measure of slope (soil-surface), i.e. the rate of inclination for land topography changes.

4) Drainage: The capability of draining off the water when the field doesn't need the water. The means of draining collectively, as a system of conduits, trenches, etc.

5) Soil pH: indicates the acidity of the soil, it can be determined by having a soil analysis carried out, and has a range approximately from 0 to 14;

6) Land grades: the quality of the agricultural land measured by the natural and economic characteristics.

Table1:

Land Code	Soil texture	Organic matte	Gradient	Drainage	Soil pH	Land grades
0616	sand	2.06	3	2	6.45	III
0896	silt	1.38	3	2	6.55	IV
1025	sand	2.06	2	2	6.42	III
1380	sand	1.38	2	3	6.42	IV
1620	silt	1.38	2	2	6.42	IV
1698	clay	1.95	2	3	6.42	I
1806	clay	1.38	3	3	6.50	II
1912	clay	1.30	3	3	6.85	II

Identifying the main drivers of agricultural growth

Gross Domestic Product (GDP) in agriculture is affected by large number of factors like public and private investment for infrastructure creation and input intensification, use of productivity enhancing inputs like fertilizer, quality seeds, irrigation, timely availability of credit, superior technology interventions, conducive climatic conditions, etc. Favorable Terms of trade (TOT) of agriculture vis-a-vis non-agriculture is another important determinant that triggers growth in the sector. Among the climatic factors, rainfall plays a crucial role in determining annual output in agriculture, as deficiency in rainfall or untimely rainfall can significantly lower crop output. A regression analysis is attempted here to identify the main factors affecting GDP in agriculture and to measure the effect of each factor in final output, in terms of elasticities.

V. Applications In Agriculture

Relationship between sprays and fruit defects

Fruit defects are often recorded (for a multitude of reasons, sometimes for insurance reasons when exporting fruit overseas). It may be done manually or through computer vision (detecting surface defects when grading fruit). Spray diaries are a legal requirement in many countries and at the very least record the date of spray and the product name. It is known that spraying can have affect different fruit defects for different fruit. Fungicidal sprays are often used to prevent rots from being expressed on fruit. It is also known that some sprays can cause russeting on apples.[18] Currently much of this knowledge comes anecdotally, however some efforts have been in regards to the use of data mining in horticulture.[19]

Growth of sheep from genes polymorphism using artificial intelligence

Polymerase chain reaction-single strand conformation polymorphism (PCR-SSCP) method was used to determine the growth hormone (GH), leptin, calpain, and calpastatinpolymorphism in Iranian Baluchi male sheep. An artificial neural network (ANN) model was developed to describe average daily gain (ADG) in lambs from input parameters of GH, leptin, calpain, and calpastatin polymorphism, birth weight, and birth type. The results revealed that the ANN-model is an appropriate tool to recognize the patterns of data to

predict lamb growth in terms of ADG given specific genes polymorphism, birth weight, and birth type. The platform of PCR-SSCP approach and ANN-based model analyses may be used in molecular marker-assisted selection and breeding programs to design a scheme in enhancing the efficacy of sheep production.[20]

Optimizing pesticide use by data mining

Recent studies by agriculture researchers in Pakistan (one of the top four cotton producers of the world) showed that attempts of cotton crop yield maximization through pro-pesticide state policies have led to a dangerously high pesticide use. These studies have reported a negative correlation between pesticide use and crop yield in Pakistan. Hence excessive use (or abuse) of pesticides is harming the farmers with adverse financial, environmental and social impacts. By data mining the cotton Pest Scouting data along with the meteorological recordings it was shown that how pesticide use can be optimized (reduced). Clustering of data revealed interesting patterns of farmer practices along with pesticide use dynamics and hence help identify the reasons for this pesticide abuse.[21]

Explaining pesticide abuse by data mining

To monitor cotton growth, different government departments and agencies in Pakistan have been recording pest scouting, agriculture and metrological data for decades. Coarse estimates of just the cotton pest scouting data recorded stands at around 1.5 million records, and growing. The primary agro-met data recorded has never been digitized, integrated or standardized to give a complete picture, and hence cannot support decision making, thus requiring an Agriculture Data Warehouse. Creating a novel Pilot Agriculture Extension Data Warehouse followed by analysis through querying and data mining some interesting discoveries were made, such as pesticides sprayed at the wrong time, wrong pesticide sused for the right reasons and temporal relationship between pesticide usage and day of the week.[22]

Analyzing chicken performance data by neural network models

A platform of artificial neural network-based models with sensitivity analysis and optimization algorithms was used successfully to integrate published data on the responses of broiler chickens to threonine. Analyses of the artificial neural network models for weight gain and feed efficiency from a compiled data set suggested that the dietary protein concentration was more important than the threonine concentration. The results revealed that a diet containing 18.69% protein and 0.73% threonine may lead to producing optimal weight gain, whereas the optimal feed efficiency may be achieved with a diet containing 18.71% protein and 0.75% threonine.[23]

VI. Conclusion

Data mining plays a crucial role for decision making on several issues related to agriculture field. Several authors discussed about the role of data mining and its applications in solving the different agricultural problems in their work. In this survey paper, it is discussed about usage of K-means algorithm in soil classification and residue regions of interest by color images, detecting weeds in agriculture. This paper is useful for researches to analyse the different authors' work and get information about data mining techniques and its applications in agriculture field. Analysis of our parameters like soil, weather, water give impact to the agriculture. Adaptive strategies can help minimize negative impacts. These need research,

Reference

[1]. 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 2011379.

[2]. Rainfall variability analysis and its impact on crop productivity Indian agriculture research journal 2002 29,33.,8) SPRS Archives XXXVI-8/W48 Workshop proceedings: Remote sensing support to crop yield forecast and area estimates GENERALIZED SOFTWARE TOOLS FOR CROP AREA ESTIMATES AND YIELD FORECAST by Roberto Benedetti A, Remo Catenaro A, Federica Piersimoni B.

[3]. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, ―Data mining Techniques for Predicting Crop Productivity – A review article‖, International Journal of Computer Science and technology, march 2011.

[4]. D.Rajesh, ―Application of Spatial Data mining in Agriculture International Journal of Computer Applications, Volume 15, February 2011.

[5]. International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 2, February 2012) 275 ―Survey on Data Mining ―VibhaMaduskar and Prof. Yashovardhankelkar.

[6]. Amrender Kumar, ―WEATHER BASED CROP FORECASTING TECHNIQUES‖, Agricultural Statistics Research Institute.

[7]. Anwiti Jain, Anad Rajavat, Rupali Bhartiya ―Design, Analysis and Implementation of Modified K-Mean Algorithm for Large Data-Set to Increase Scalability and Efficiency‖, IEEE Explore December 2012.

[8]. An Approach", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, Issue 1, January 2012, pp. 110-113.

[9]. K. SOMVANSHI, ET AL., "MODELING AND PREDICTION OF RAINFALL USING ARTIFICIAL NEURAL NETWORK ANDARIMA TECHNIQUES", J. IND. GEOPHYS. UNION, VOL. 10, NO. 2, PP. 141-151, 2006.

[10]. K. Verheyen, D. Adriaens, M. Hermy, and S. Deckers, "High resolution continuous soil classification using morphological soil profile descriptions", *Geoderma,* vol. 101, pp. 31-48, 2001.
[11]. Jagielska, C. Mattehews, T. Whitfort, "An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and roughsets to automated knowledge acquisition for classification problems", Neurocomputing, Vol. 24, pp. 37-54,
[12]. Tellaeche, A., BurgosArtizzu, X. P., Pajares, G., &Ribeiro, A. (2007), "A vision-based hybrid classifier for weeds detection in precisionagriculture through the Bayesian and Fuzzy k-Means paradigms", In Innovations in Hybrid Intelligent Systems (pp. 72-79). Springer BerlinHeidelberg.
[13]. Veenadhari, S. 2007, "Crop productivity mapping based on decision tree and Bayesian classification". Unpublished M.Tech Thesis submittedto MakhanlalChaturvedi National University of Journalism and Communication, Bhopal. AltannarChinchulunn, PetrosXanthopoulos, Vera Tomaino, P.M.Pardalos, "Data Mining Techniques in Agricultural and EnvironmentalSciences", International Journal of Agricultural and Environmental Information Systems,1(1),26-40,January-June2010.
[14]. B. Rajagopalan and U. Lal, "A K-nearest neighbor simulator for daily precipitation and other weather variable", *Water Resources*, vol. 35, pp.3089-3101, 1999.
[15]. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, "Data mining Techniques for Predicting Crop Productivity – A review article", InternationalJournal of Computer Science and Technology IJCST Vol. 2, Issue1,March2011.
[16]. Tripathi, S., Srinivas, V. V., &Nanjundiah, R. S. (2006), "Downscaling of precipitation for climate change scenarios: a support vector machineapproach", Journal of Hydrology, 330(3), 621-640.
[17]. *"Apple russeting". www.extension.umn.edu. Retrieved 2016-10-04.*
[18]. Hill, M. G.; Connolly, P. G.; Reutemann, P.; Fletcher, D. (2014-10-01). "The use of data mining to assist crop protection decisions on kiwifruit in New Zealand". *Computers and Electronics in Agriculture*. **108**: 250–257.
[19]. Mojtaba, Tahmoorespur,; Hamed, Ahmadi, (2012-01-01). "neural network model to describe weight gain of sheep from genes polymorphism, birth weight and birth type". *Livestock science*. ISSN 1871-1413.
[20]. Abdullah, Ahsan; Brobst, Stephen; Pervaiz, Ijaz; Umar, Muhammad; Nisar, Azhar (2004). *Learning Dynamics of Pesticide Abuse through Data Mining* (PDF). Australasian Workshop on Data Mining and Web Intelligence, Dunedin, New Zealand.
[21]. Abdullah, Ahsan; Hussain, Amir (2006). "Data Mining a New Pilot Agriculture Extension Data Warehouse" (PDF). *Journal of Research and Practice in Information Technology, Vol. 38, No. 3, August 2006*: 229–249. Archived from the original (PDF) on 2010-09-23.
[22]. Ahmadi, H.; Golian, A. (2010-11-01). "The integration of broiler chicken threonine responses data into neural network models". *Poultry Science*. **89** (11): 2535–2541. doi:10.3382/ps.2010-00884. ISSN 0032-5791.