

A Survey on Data Mining Techniques in Agriculture

Shalin Paulson
St.Joseph's College
Irinjalakuda

Abstract—Data mining is a fast emerging and highly rising research oriented field in agriculture for formulating and analysing various conditions on crop yield. In this paper our focus is on studying and experimenting the applications of data mining techniques in agricultural field. Data mining plays a significant and unique role for making decision on several issues related to agriculture field. Data mining in agriculture can provide help in predicting yield, forecasting weather and rainfall, quality of seed and soil, production of crops. Predictive data mining technique is used to predict future crop, pesticides and fertilizers to be used, revenue to be generated for proper growth and function of crops in agriculture. Several data mining techniques such as k-means(KM),k-nearest neighbour(KNN), artificial neural network(ANN), support vector machine(SVM) are implemented to solve and help in searching various different ways to improve the growth of agriculture. Each data mining technique has its own different way of visualizing various problems and leading to give us a proper solution for every causing problem in agriculture. In this paper we gained to achieve proper knowledge and correct information of each data mining technique thereby using them in all related issues. It summarises the whole data by combining and executing all the techniques mentioned along with by implementing some new technique thereby improving the way of agriculture planning leading to proper agricultural growth.

Keywords - *K-means-nearest neighbour, Artificial neural network, support vector machine, WEKA tool.*

I. INTRODUCTION

Data mining is a task of extracting more and more data from known and existing data (discovery of new data from set of databases). By doing data mining we can able to generate and build meaningful and knowledgeable data. Data mining includes various techniques that implement the use of data mining. Data mining techniques are used in many fields to enhance the quality and importance of useful data. It plays a vital and key role in each respective field. The main aim of data mining is to collect the processed data and transform it into human understandable format[4][14].

Data mining in field of agriculture is a vast and valuable concept. Agriculture is the domain of our Indian nation. Agriculture is considered as a quality occupation and has a special role in each lifestyle. The productivity of agriculture is very low as compared with others, so as the demand of food is increasing daily, the researchers, farmers, agricultural scientists and government are trying to

put extra effort and techniques for more production. And as a result, the agricultural data increases day by day. As the content of data increases, it requires proper way for these data to be extracted or produced when needed.

The intention of this paper is to give details about different data mining techniques in field of agriculture and try to implement other techniques that would be more beneficiary along with existing techniques.

II. LITERATURE SURVEY

A. Data mining: concepts and techniques

This paper covers the basic information about the concept of data mining and its various techniques that help to act as the basic system to implement only basic techniques applied to various fields so thereby implementing many improving solutions to large problems. This paper summarizes overall content and several methods for implementing data mining and minimizing the causes and leading to good solutions. The main advantages covered here is that it is simple and easier to implement, easily understandable. The disadvantages of this paper is that there no future work and does not consist of a new proposed system.

B. Data mining and applications to agricultural yield data

This paper tries to gain several knowledge and useful data for predicting yield production. Several data mining techniques such as k-means, k-nearest neighbour, artificial neural networks and support vector machine are implemented for proper functioning and smart growth of agriculture. Each techniques are used for performing forecast of pollution in atmosphere, weather scenarios, classify sounds of birds, classify soils and predicting rainfall. It is considered as efficient since it include all necessary details in computing agricultural related problems and finds solutions for each. The disadvantage mentioned here is that it cannot evolve itself into a new and more precise system.

C. Survey of data mining techniques applied to agriculture

The paper also focuses on data mining techniques for solving complex agricultural problems using data mining and enhances several applications in agricultural fields. This paper is classified on clustering and classification mechanisms. This survey paper does not prove efficient here because it uses only general methods of

classification and clustering not the basic ones so it cannot be easily accessed for future work.

D. A survey on data mining technique in agriculture

The survey created on agriculture mistreatment data processing techniques reveals the importance to extract helpful data from dataset. Data processing techniques like K-Means, K-NearestNeighbor (KNN), Support Vector Machine and Artificial Neural Network square measure recent applications in agriculture. This survey aims to search out correct data processing models to realize high accuracy and prediction capabilities. It's the opinion that additional techniques and algorithms to be studied connected agricultural issues can provide smart lead to agricultural growth. This survey offer recommendations for future analysis directions in agriculture-related fields. Data processing and therefore the varied methodologies related to it will scale back the complexity of the information sanctionative farmers to create choices additional simply.

E. A survey on data mining techniques in agriculture

This paper discusses about the role of data mining in agriculture field and also focuses about several data mining techniques and their related work by several authors in context to agriculture domain. It also discusses on different data mining applications in solving the different agricultural problems. It integrates the work of various authors in one place so it is useful for researchers to get information of current scenario of data mining techniques and applications in context to agriculture field. This paper provides a survey of various data mining techniques used in agriculture which includes Artificial Neural Networks, K - nearest neighbor, Decision tree, Bayesian network, Fuzzyset, Support Vector Machine and K – means. The main advantage is that it completely suits with the current procedure in agriculture and the drawback is that it by implementing future works and implementing new reviews we can make this paper in precise form.

F. Recent developments in data mining and agriculture

This survey covers some very recent applications of data mining techniques in the field of agriculture, we first present two applications in this field in particular, we consider the problem of discovering problematic wine fermentations at the early stages of the process, and the problem of predicting yield production by using sensor data information. Secondly, we briefly describe other problems in the field for which we found very recent contributions in the scientific literature. By using sensor networks this paper can be reviewed and made more better than the current.

G. A brief survey of data mining techniques applied to agricultural data

The paper provides a brief review of a variety of Data Mining techniques that have been applied to model data from or about the agricultural domain. The Data Mining techniques applied on Agricultural data include k-means, bi clustering, k nearest neighbor, Neural Networks (NN)

Support Vector Machine (SVM), Naive Bayes Classifier and Fuzzy c-means. This survey summarizes the application of data mining techniques and predictive modeling application in the agriculture field. Drawback is that it needs to predict the future trends of agricultural processes. The main aim is to analyse data in effective way that requires understanding of appropriate techniques of data mining. The intention of this paper is to give details about different data mining techniques in perspective of agriculture domain so researchers can get details about appropriate data mining technique in context to their work area.

III. DATA MINING TECHNIQUES IN AGRICULTURE

Different techniques were proposed for mining of data over the years. In this paper we present some of the most used general Data Mining techniques in the field of agriculture such as k-means(KM), k-nearest neighbor(KNN), artificial neural network(ANN), support vector machine(SVM) are implemented to solve and help in searching various different ways to improve the growth of agriculture.

| Methodology | Applications |
|-------------------------|--|
| K-means | Forecasts of pollution in atmosphere Classifying soil in combination with GPS |
| k-nearest Neighbor | Simulating daily precipitations and other weather variable |
| Support Vector Machine | Analysis of different possible change of the weather scenario |
| Decision Tree Analysis | Prediction soil dept |
| Unsupervised Clustering | Generate cluster and determine any existence of pattern |
| WEKA Tool | Classification system for sorting and grading mushrooms. |

Fig1: Data mining methodologies and its uses in agriculture

A. k-means algorithm in the field of agriculture:

The k-means algorithm is used in classifying soils using GPS-based technologies. An amount of sensor data can be easily collected, so that farmers do not only harvest crops but also growing and growing amounts of data. Performs classification of the plant, soil, and residue content, monitoring water quality changes, detecting weeds used in agriculture, to perform forecast of the pollution in the atmosphere, prediction of wine fermentation problems can be performed by using a k-means approach. Knowing in advance that the wine fermentation process could get stuck help the enologist to correct it and ensure a good fermentation process.

B. k-nearest neighbor algorithm in the field of agriculture:

The k-nearest neighbor classification algorithmic may be divided into two phases: coaching section and testing section. 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.

C. Artificial neural networks in field of agriculture:

This algorithm mainly focuses on the information gained about weather and is observed and stored for later use. The recorded data are used to forecast weather. If there is a change in any one of the recorded data like wind speed, wind direction, temperature, rainfall, humidity, then the future climatic condition can be predicted using artificial neural networks. It can also help in prediction of rainfall using artificial neural networks and, other applications of artificial neural networks in hydrology are forecasting daily water hassle and flow forecasting. ANNs are used even to classify eggs as fertility and uses computer vision to recognize cracks in eggs.

D. Support Vector Machines in field of agriculture:

The main goal of Support Vector Machine (SVM) is to classify information samples. The essential plan behind is classifying the sample information into linearly severable. The SVM-based DM is applied to future climate predictions from these code generation Coupled Global Climate Model (CGCM2) to obtain future projections. The results obtained are then analyzed to present the crash of climate change on rainfall over India. SVMs can be used to classify the sound of birds and other sort of different sounds, to classify pizza sauce spread and for detecting weed and nitrogen stress in corn. Similarly, sensors are used to smell milk that is classified using SVMs.

IV. APPLICATIONS OF DATAMINING IN AGRICULTURE

Data mining has many applications which are related to each technique in their own respective areas. Using each data mining techniques, applications are implemented so as to improve the technology. Recent technologies are nowadays able to provide a lot of information on agricultural-related activities, which can then be analyzed in order to find important information. Various applications increases the usage of data mining and leads to creating innovative and interesting ideas.

A. Prediction of problematic wine fermentations

Wine is usually produced all around the world. The process of fermenting the wine is very important, because it can implement the productivity of wine-related industries and also the quality of wine. If we could predict how the fermentation is going to be at the early stages of the process, we could interfere with the process in order to guarantee a regular and smooth fermentation. Fermentations are studied by using different techniques, such as, for example, the k-means_algorithm—and a

technique for classification based on the concept of biclustering.

B. Detection of diseases from sounds issued by animals

The detection of animal's diseases in farms can impact positively the productivity of the farm, since sick animals can cause contaminations and spread contagious diseases. Moreover, the early detection of the diseases can allow the farmer to cure the animal as soon as the disease appears. Sounds made by pigs can be analyzed for the detection of diseases. In particular, their coughs can be studied, because they indicate their sickness content. A computational system is under development which is able to monitor pig sounds by microphones installed in the farm, and which is also able to discriminate among the different sounds that can be detected.

C. Optimizing pesticide use by data mining

Recently conducted studies by agriculture researchers 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 procedure. Hence excessive use of pesticides is harming the farmers with adverse financial situations, environmental and social impacts. By data mining the cotton Pest Scouting data 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.

D. Sorting apples by water cores

Before reaching the market, apples are checked and the ones showing some defects are removed. However, there are also unseen defects, that can spoil the apple flavor and look. An example of unseen defect is the watercore. This is an apple disorder that can affect the longevity of the fruit. Apples with few or mild watercores are sweeter, but those apples with moderate to severe degree of watercore cannot be stored for long length of time. In addition, a few fruits with severe watercore could spoil a whole batch of apples. Due this reason, a computational system is undergoing study which takes X-ray photographs of the fruit while they run on conveyor belts, and which is also able to analyse (by data mining techniques) the taken pictures and estimate the possibility that the fruit contains watercores.

V. PROPOSED SYSTEM

Proposed system is to generate more effective and improving way that reviews new creation as compared to the current existing system. In short producing a new system by overcoming the existing one. In this paper under the existing system at present we mainly focus on data mining techniques and their uses in each respective fields. In the existing system by using several data mining techniques we help to solve problems occurred in agriculture related fields. In the proposed system i am

trying to implement the concept of E-Agriculture. The main aim of this paper is to reach farmers for their awareness, usage and perception in e-Agriculture. e-Agriculture supports marketing of agricultural products.

The agriculture sector in India is currently facing a difficult phase. India is moving towards an agriculture emergency due to inadequate investment in irrigation and agriculture infrastructure, lack of attention, ineffective land management, non-given of fair prices to farmers for their crops and insufficient land reform in India, etc. Food production and productivity in India is declining while its food consumption is increasing. The situation has further been worsening due to use of food grains because of demand of bio fuels.

By the use of this technique, India's food production and productivity has been increased for agricultural purposes. The developed nations are using technology of laser in place of tractors to plough lands. This helps in optimizing the use of a range of inputs parameter such as water, seeds, fertilizers, etc. The problem occurs here is that Indian farmers cannot pay for this technology. In addition, power and electricity also cause a major problem for Indian farmers.

“E-Agriculture” is highly emerging and fastest growing field in the connection of agricultural informatics, development and entrepreneurship which focuses to agricultural services and duties, technology distribution and information delivered or developed through the Internet and associated technologies. Mainly, it focuses on the design, development, assessment and application of innovative ways to use active or emerging information and communication technologies (ICTs).

E-agriculture is a vast field for enhancing existing agriculture and food security through enhanced processes for knowledge access and switch using information and communication technologies. In short e-Agriculture will connect all concerned persons starting from farmers to researchers together. Farmers can get there required information at any time from any part of world and they can also get the help from experts viewing their problem immediately by without moving anywhere.

This paper estimates in benefiting the Indian farmers, if the market and weather information is delivered to their mobile phones. For improving agricultural productivity an expert agricultural advice is given to the farmers both in timely and personalized situations. Here, in this system agricultural experts generate the advice by using the modern agriculture which is highly knowledge intensive which also requires timely, reliable and accurate information on natural resource endowments and their usage patterns at present and future technology available for their utilization and other information about markets, weather, insurance, subsidy, etc. The news releases from the government does not reach the farmers in time, therefore an alert system is being built for daily releases and for seasonal releases.

VI. MAIN APPROACHES ON E-AGRICULTURE

A. Wireless technologies

Wireless technologies have numerous applications in agriculture. One major usage is the simplification of closed-circuit television camera systems; the use of wireless communications eliminates the need for the installation of coaxial cables.

B. Global Positioning System

In agriculture, the use of the Global Positioning System provides benefits in geo-fencing, map-making and surveying. With the use of GPS, civilians can produce simple yet highly accurate digitized map without the help of a professional cartographer. In Kenya, for example, the solution to prevent an elephant bull from wandering into farms and destroying precious crops was to tag the elephant with a device that sends a text message when it crosses a geo-fence. Using the technology of SMS and GPS, the elephant can roam freely and the authorities are alerted whenever it is near the farm.

C. Smartphone mobile apps in Agriculture

Use of Mobile technologies as a tool of intervention in agriculture is increasingly popular. Smartphone penetration enhances the multi-dimensional positive impact on sustainable poverty reduction and identifies accessibility as the main challenge in harnessing the full potential in agricultural space. Reach of smart phone even in rural areas extended the ICT services beyond simple voice or text messages. Several Smartphone apps are available for agriculture, horticulture, animal husbandry and farm machinery.

VII. CONCLUSION

Agriculture is the most promising application area particularly in the developing countries like India. Usage of information technology in agriculture can change the criteria and scenario of decision making and farmers can yield in better and convenient way. Data mining plays a crucial and lead role for decision making on several issues related to agriculture field. It discusses about the role of data mining in the agriculture field and their related work in context to agriculture. It also discusses on different data mining applications in solving the different agricultural problems that has been reported[1].

In the survey conducted i introduced different data mining techniques and their functions which helped in producing a solution for all possible problems occurred. Besides this i introduced the main concept of e-agriculture that made improvement in the quality of agricultural data mining by means of modern ways and technologies. It also tried in empowering knowledge to farmers by leading to less unemployment. In conclusion, it is our opinion there is a lot of work to be done on this emerging and interesting research field. This is often comparatively considered a completely unique analysis field and it's expected to grow along with the future and produce more interesting concepts.

VIII. FUTURE WORK

Data mining not only gives us information about a particular task but there are chances where the obtained information can be changing according to the modern technologies emerging widely. Not only by these mentioned ways there are several other ways in which data can be mined and produced more and efficient data by implement a precise and a bit more easier way of introducing new concept. Several other differentiate algorithms can be used in order to gain higher standard applications. By doing proper study and collecting required data there can be a possibility of implementing wireless techniques without use of a Smartphone. The future work aimed at the analysis of the entire obtained data and will be converted to suitable strategies for improving the efficiency of the proposed algorithm.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to Ms.Greeshma Sunny and all other faculty members from department of computer science, St.joseph's college, Irinjalakuda for making the paper completion done successfully.

REFERENCES

- [1] D Ramesh, B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.
- [2] Raorane A.A., Kulkarni R.V, "Review- Role of Data Mining in Agriculture", International Journal of Computer Science and Information Technologies, Vol. 4 (2) , 2013, 270 – 272.
- [3] Mucherino, A., Papajorgji, P., & Pardalos, P. (2009), "Data mining in agriculture" (Vol. 34), Springer.
- [4] Han, J, Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques. Morgan kaufmann
- [5] Mr. Abhishek B. Mankar, Mr. Mayur S. Burange, "Data Mining - An Evolutionary View of Agriculture", International Journal of Application or Innovation in Engineering & Management (IJAEM), Volume 3, Issue 3, March 2014.
- [6] S.S.Baskar, L.Arockiam, V.Arul Kumar,L.Jeyasimman, "Brief Survey of Application of Data Mining Techniques to Agriculture", Agricultural Journal 5(2): 116:118,2010 ISSN:1816-9155
- [7] Raorane A.A and Kulkarni R.V, "Review- Roleof Data Mining in Agriculture", InternationalJournal of Computer Science andInformation Technologies, Vol. 4 (2) , 2013, 270 – 272
- [8] S.Veenadhar, Dr. Bharat Misra, Dr. CD Singh,"Data mining Techniques for PredictingCrop Productivity – A review article",International Journal of Computer Science andTechnology IJCST Vol. 2, Issue 1, March 2011
- [9] Sumitha Thankachan, dr S.Kirubakaran, Vol 3.May 2014 Issue 5
- [10] Hetal Patel, Dharmendra Patel "A brief survey of data mining technique applied to agricultural data".
- [11] MsKalpana,Dr.Shanti,Dr.ArumugamS Vol.3,No.8,Auguest 2014
- [12] www.eagriculture.org/sites/.../eagriculture%20in%20india.pdf
- [13] www.ijcsmc.com/docs/papers/may2014/v3i5201499a28.pdf
- [14] https://en.wikipedia.org/wiki/data_mining_in_agriculture
- [15] crpit.com/confpapers/CRPITV70Armstrong.pdf