IJRSET OCTOBER 2020 Volume 7, Issue 10 Pages: 1-8

ISSN 2394-739X

International Journal for Research in Science Engineering and Technology

A SURVEY ON DATAMINING IN AGRICULTURE

[1] HARSHINI. N, [2] Dr. M. RATHAMANI
[1, 2] Assistant professor,
[1, 2] Department of Computer Science,
[1, 2] NGM College,
[1, 2] Pollachi.

ABSTARCT- Agriculture assumes an essential part in the life of an economy. It is the backbone for creating nations like India as over 70% of populace relies upon agriculture. To expand crop production numerous components are mindful like soil, climate, downpour, fertilizers and pesticides. The Author has utilized soil boundaries to expand crop production since it is a fundamental key factor of agriculture. To keep up supplement levels in the soil if there should arise an occurrence of inadequacy, fertilizers are added to soil. The basic issue existing among the Indian ranchers is that they pick surmised measure of fertilizers and include them physically. Abundance or lacking expansion of manure can hurt the vegetation and diminish the yield. This paper gives audit of different data mining procedures utilized on agriculture soil dataset for compost suggestion. Basically I zeroed in on different soil boundaries like Fe, S, Zn, Cu, N and Ph esteem and so forth. In this review, we additionally depict some Agriculture issues that can be illuminated by utilizing data mining strategies.

Keywords – [K-means-Nearest Neighbour, Data Mining, Clustering, Classification, Artificial Neural network, support vector machine.]

1. INTRODUCTION

Data mining

Data mining is the way toward finding already unknown and possibly intriguing examples with regards to huge datasets. The dug data is utilized for speaking to as a model for expectation or classification. Datasets from the farming space have all the earmarks of being fundamentally more unpredictable than the datasets customarily utilized in machine learning. Data mining is for the most part sorted as clear and predictive data mining. Be that as it may, in the agriculture area, predictive data mining is primarily utilized. An assortment of data mining procedures are utilized in various fields to improve or pick up knowledge. Presently a day the specialists, data experts and researchers has more focused on how mining and machine learning strategies are utilized to examine different soil profiles to improve the field of agriculture . Crop Productivity Mr. Narsi Reddy Gayam expressed in his exploration learning A investigation of crop yield distribution and crop insurance‖ which takes the info data from INDIA relating sugarcane and Soybean. After his exploration

he closed crop yield are not ordinarily circulated.

Use of DM Techniques in Agriculture Dr. Bharat Misra, et al., watched the examination concentrates on use of data mining strategies in the field of agriculture. A portion of the strategies, for example, ID3 calculations, the k-means, and the k-nearest neighbor, artificial neural networks and support vector machines applied in the field of agriculture were introduced. Data mining in application in agriculture is a generally new methodology for estimating/foreseeing of agrarian crop/creature the board.

DATA MINING IN AGRICULTURE

Extricating significance full subtleties from a lot of data which can be in various structures or terms is the technique known as data mining. These sorts of terms incorporate binary content, numbers, and messages as purchaser data. Data mining is likewise Known knowledge revelation in Databases (KDD). There are number of significant areas where data mining is broadly utilized, for example, Future medical services, Education, Customer Segmentation, Fraud discovery, Financial Banking and Agriculture.

Data mining in agriculture is an exceptionally flow research theme in light of the fact that farming area needs more support for its improvement in creating nations like India. As Mahatma Gandhi stated, "India lives in towns and agriculture is the spirit of Indian economy".

In Indian agriculture, the volume of data is gigantic. The data when become data is exceptionally valuable for some reasons. The customary and conventional arrangement of data examination in agriculture is simply subject to insights. Data mining is an advanced data investigation procedure. It has wide scope of utilizations in the field of agriculture.

In this examination, uses of the data mining procedures in the area of agriculture and its united areas are contemplated. Various procedures of data mining have been utilized in this field. However, there are heaps of strategies accessible in the data mining, not many systems, for example, K-means approach, K-nearest neighbor, Bi-clustering are well known presently relies upon the idea of the data

Rural data expands step by step. So there is a need to change this tremendous data into innovations and make them accessible to the ranchers. Still suggestions for ranchers depend on coordinated collaboration among ranchers and specialists and various specialists have assortment of proposals. So Recommendation can be given to ranchers utilizing past farming data help of data mining ideas and they can get streamlined outcomes from recommender. The inspiration driving this work is to investigate data mining strategies, which are appropriate for taking care of complex agrarian issues. Analysts plan and create applications for tackling complex farming issues utilizing data mining like Yield Data Prediction, Soil Mapping and Classification, Fertilizer and Pesticide Management, Grading and arranging of agriculture item and so on.

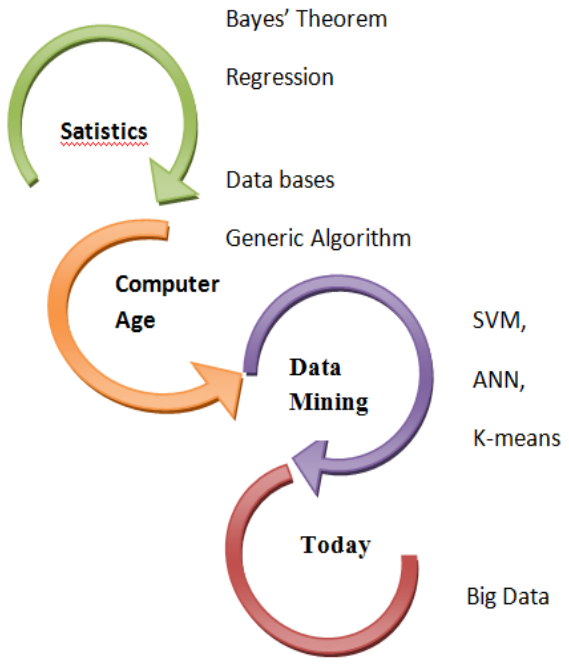
Agriculture is subject to certain variables like soil, climate, cultivation, irrigation, fertilizers, temperature, rainfall, harvesting, pesticide weeds and so forth. Numerous businesses utilize agrarian items as crude material, livestock, food, creature feed, compound, poultry, compost, pesticides, seed and paper.

An exact gauge of crop production and risk enables these organizations in arranging flexibly to chain choice like production booking. Business, for example, seed, compost, agrochemical and agrarian machinery enterprises plan production and marketing exercises dependent on crop production gauges [1, 2].

There are 2 components which are useful for the ranchers and the administration in dynamic specifically,

1. It helps ranchers in furnishing the verifiable crop yield record with an estimate diminishing the risk the board.

2. It helps the legislature in making crop insurance strategies and arrangements for flexibly chain activity.



Data mining in Agriculture

Benefits of Data Mining in Agriculture

Using data mining in agriculture benefits farmers, government, and so on from various perspectives. One of the manners in which that utilizing data mining is valuable in agriculture is "possibility to examine concealed examples in datasets in agricultural domain. These examples can be utilized for diagnosing crop condition, prognosing market improvement, checking client dissolvability" at the end of the day, it can assist farmers with telling around the time that their crop would prosper, make expectations of when their item is bound to sell, and what item clients are purchasing more.

Agricultural establishments use data mining method and applications for various areas, for example agronomists use designs estimating development markers of plants, crop quality pointers, accomplishment of taken agro specialized measures and directors of agricultural associations focus on client fulfillment and financially ideal choices.

Data mining utilizes numerous applications to figure a few things that could occur with produce, crops, and so forth. For example, farmers use K – means group calculation so as to figure out which apples could be sold at a lower cost and which apples could be sold at a typical cost dependent on what amount harmed they have. In addition, during k – means group, the apples are separated and dependent on those apples that are awful, they are partitioned again to make sure which ones go available to be purchased and which ones don't. Moreover, so as to get which crops are utilized more and which ones are the nearest to that crop on the off chance that that is missing, we use Support Vector Machine (SVM).

2. LITERATURE SURVEY

Foreseen data mining in agriculture rising assessment field in crop yield examination. Various data mining strategies, for instance, K-Means, K-Nearest Neighbor (KNN), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are used for new application research in agriculture field. In this investigation, the techniques for foreseeing yield production of a gather are the point of convergence of center premium. Yield figure is huge in agriculture. The issue of yield conjecture can be settled by utilizing data mining systems.

1. Kazuo Yonekura, Hitoshi Hattori and Taiji Suzuki (2018) et.al proposed a short-term, i.e., under 60 minutes, local weather forecast strategy. In this technique, the creator proposes a deep learning architecture specifically developed for the short-term weather forecasting dependent on the dense weather station device. It consists of two folds: point prediction model and tensor prediction model. The point prediction model is helpful for forecasting exactly on the location of the dense weather station. The tensor prediction model interpolates the prediction of the point prediction model to cover entire range of locations around the interested zone. In request to offer a forecast at subjective any point around the weather stations, a tensor prediction model is proposed. The tensor learning model is input with the weather elements, geospatial information of the locales, and the forecast target point. Using the tensor prediction model, a forecasting map is generated. The accuracy of the tensor prediction model

increases as the number of weather stations considered in the model increases.

Merits

These short-term local predictions are valuable in fields like farming, construction, and others, to develop a precise working schedule and plan.

It is important to utilize dense data to obtain a precise prediction.

Demerits

In particular, the primary purpose of rain prediction is to detect the transition in physical situations. It experiences the case-imbalance problem. The cases for "weather-change" and "no-weather-change" are exceptionally imbalanced.

2. K. L. Ponce-Guevara, J. A. Palacios-Echeverr´ıa, E. Maya-Olalla, H. M. Dom´ınguez-Limaico, L. E. Suarez-Zambrano, P. D. Rosero-Montalvo, D. H. Peluffo-Ordo´nez, J. C. Alvarado-Perez (2017), et.al proposed Big Data and Data Mining techniques on vegetable crops data from a greenhouse by implementing the primary version of a software device, supposed GreenFarm-DM. Such an apparatus is pointed toward analyzing the factors that influence the development of the crops, and determine a predictive model of soil moisture. The prediction, in which, from the classification results obtained in the previous stage, the objective variable is predicted from information of a new arrangement of data. At the end of the day, the interface fabricates a predictive model to determine the conduct of soil moisture. After a research of the factors that affect the crop development within a greenhouse, the most important were determined, among them: Soil moisture, relative humidity, ambient temperature, lighting level and carbon dioxide, given that these factors directly influence the process of photosynthesis of plants and establishing correct qualities will achieve a superior nutrient uptake and obtain better natural products.

Merits

GreenFarmDM, which is an interface designed for data analysis in greenhouse crops. IT is intuitive and simple to-use for the horticulturist.

It likewise permitted the implementation of a user-friendly interface.

Demerits

The different classifications techniques aren't to be integrated into the GreenFarm-DM interface with the point of improve this development.

3. Priyanka Padalalu , Sonal Mahajan , Kartikee Dabir , Sushmita Mitkar and Deepali Javale (2017), et.al proposed thought is beneficial to the farmers to water the ranches efficiently using a robotized irrigation system dependent on soil temperature, moisture and pH. Respective sensors are utilized to find the soil water content level and dependent on this microcontroller drives the servo engine and pump. The subsequent watering of plants can be controlled using the aforementioned app. Depending upon the type of soil and crop, the fertilizers are proposed by applying Naïve Bayes algorithm on the database. The assessed amount of rain is predicted using weather forecasting using Web scraper and the crops are watered accordingly, i.e., is a substantial rainfall is predicted then the system will automatically reduce the water supplied to the crops and gives quick solution for the judicious use of water which has been a significant issue in developing countries like India.

Merits

This can help to control the excess water which is squandered.

There is minimal human intervention as the system takes into account the water content level in the soil and accordingly provides water. Hence humans need not constantly be on the field.

The supply of water can be controlled using an Android app. Hence, the rancher need not be present on the field to begin the water supply.

Demerits

The system will appraise the irrigation cost and check the reasonability of the technique.

4. Hasmita S, Fhira Nhita, Deni Saepudin, Annisa Aditsania (2019), et.al proposed Chili Commodity Price Forecasting in Bandung Regency using the Adaptive Synthetic Sampling (ADASYN) and K-Nearest Neighbor (KNN) Algorithms. The prediction of chili prices was carried out using the K-Nearest Neighbor (KNN) algorithm dependent on chili price data and weather data. The data obtained had imbalanced classes, so the Adaptive Synthetic (ADASYN) algorithm was utilized to overcome this issue.

Merits

Farmers will have the option to obtain the best yields from planting using the prediction consequences of chili prices provided.

The classification using KNN and ADASYN obtained 100% accuracy and a F1-Score of 100%.

ADASYN has better efficiency compared to other sampling algorithms because it is adaptive in the distribution update process and can conform to the characteristics of data distribution.

Demerits

One of the biggest issues with K-NN is to choose the optimal number of neighbors to be consider while classifying the new data entry.

5. Chutinan Trongtorkid, Part Pramokchon (2018) et.al proposed framework for Expert System for Diagnosis Mango Diseases Using Leaf Symptoms Analysis. This paper presents the knowledge-situated in a specialist framework for ordering plant disease which analysis from symptom on mango leaves. The knowledge-based acquired by data mining technique utilized choice tree model. The plant diagnosis application applies a knowledge base framework in type of rule-based model acquired by data mining technique. It present standard based model with leaf picture dataset. As the analysis result, this proposed knowledge-based got by data mining technique has effectively with an accuracy of ordering 3 sort of mango leaves.

Merits

Moreover, this proposed model can likewise grouped an ordinary leaves with a satisfactory degree of percent accuracy.

The proposed knowledge-base model can additionally arrange a contaminated anthracnose leaves with high exactness.

Demerits

It requires enormous extra room.

6. Priyadarshini Patil, Nagaratna Yaligar, Meena S M (2017) proposed the Comparision of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection utilizing Leaf Images. proposed a framework which identifies early blight and late blight diseases in potato crop at early stages utilizing leaf pictures. The authorhave utilized FCM clustering to portion disease influenced district. Surface highlights are separated from the disease influenced locale utilized for characterization. The exhibition of framework is discovered by estimating the accuracy of identifying disease influenced districts and grouping as diseased or typical leaf.

Merits

The execution of classifiers SVM, RF and ANN, ANN is best classifier for the framework with most noteworthy accuracy 92%, because of its capacity to demonstrate and learn complex continuous frameworks.

One significant preferred position of ANN is its adaptability without forcing any limitations on input factors, to demonstrate inconspicuous concealed connection between factors.

Demerits

The SVM doesn't perform well when the quantity of highlights is more noteworthy than the quantity of tests.

7. Velamakanni Sahithya, Brahmadevara Saivihari, Vellanki Krishna Vamsi,

Parvathreddy Sandeep Reddy and Karthigha Balamurugan (2019), et.al proposed GUI based Detection of Unhealthy Leaves using Image Processing Techniques. Consequently the work depicts the image processing techniques that are utilized to distinguish sound and unhealthy leaves in women finger plant. So as to oblige practical terrain limitations, commotion are remembered for the piece of preparing and testing data set. SVM and ANN are utilized to group leaves and thought about in their exhibition. This work can be stretched out to build up an 'application' kind of data age and sending 'sms' administrations to farmers and end clients. The Graphical User Interface (GUI) is made in this work as an agricultural application and the accuracy contrast between the classifiers for various diseases, for example, yellow vein mosaic, powdery mildew and leaf spot are calculated. The point of this work is to distinguish the unhealthy leaf of women finger plant by diagnosing the leaf image.

Merits

Results are given in GUI to simple comprehension.

Demerits

It turns out to be more perplexing if client needs to speak with the PC straightforwardly.

8. Abdul Hafiz Bin Abdul Wahab, Rahimi Zahari, Tiong Hoo Lim (2019), et.al proposed recognizing diseases in Chili Plants Using K-Means Segmented Support Vector Machine. In this strategy, an Artificial Intelligence based image processing calculation is proposed to identify diseases on a Chili plant using its leaves images. The proposed arrangement centers around using K-Means clustering calculation for image segmentation and analyzes distinctive Support Vector Machine (SVM) calculation for characterization. By and large, the calculation utilized can accurately distinguish between solid aspect of a Chili plant, the territory influenced by cucumber mosaic and just as the background. This can be improved via preparing more data set of the plant influenced by cucumber mosaic virus. The outcomes are arranged into background, sound and unhealthy (Cucumber Mosaic) and can separate among wellbeing and unhealthy plant.

Merits

A high level of accuracy is acquired for both order of background images and the solid aspect of the plant. Anyway just, 57.1% accuracy was gotten for grouping of cucumber mosaic.

Demerits

Large datasets are utilized for preparing the calculation, just certain highlights or data set are store in the prepared calculation.

9. Tanvi Mehra , Vinay Kumar, Pragya Gupta (2016), et.al proposed Maturity and disease detection in tomato using PC vision. The tomato crop is often tainted by a disease, where plant's leaves get secured with spots of colors dark earthy colored with purple outskirt and light dim focus; named as Septoria Leaf Spot. It makes the leaves turn yellow, however most harm happens because of loss of leaves by infection. In this strategy, tomato development based on color and fungal infection in the tomato leaves is resolved. At first thresholding calculation was performed to decide the development of tomato. To make the framework more summed up and self-adjusting a move to k-means clustering calculation is made. At long last a similar analysis of both the techniques was done to break down which strategy is more appropriate in various conditions**.**

Merits

It is the best technique as it is non-damaging, precise, and reliable and diminishes reliance on accessible labor.

The entire cycle utilized aides in isolating the ready for example red tomatoes from the rest part of the image.

Demerits

There are preferences of k-means clustering strategy yet it has a few hindrances too like the assurance of the k esteem as the group

size is fixed, the examination of bunch quality delivered is a troublesome task.
Moreover, it doesn't perform well with groups that are non-globular and assortment of beginning parcels can bring about variety in conclusive bunches.

10. Yudha Alif Auliya (2019), et.al proposed arrangement of Improve Hybrid Particle Swarm Optimization and K-Means by Random Injection for Land Clustering of Potato Plants. Potato crop productivity isn't ideal because of the shifted changeability and productivity of the dirt. Based on these issues it is important to bunch the potato planting land based fair and square of land reasonableness. The new methodology was utilized, to be specific mixture molecule swarm streamlining and K-Means (KCPSO). KCPSO is executed using the random injection idea to acquire better wellness esteems. Figuring of wellness esteem based on Silhouette Coefficient esteem. Based on the aftereffects of testing, the KCPSO engineering imported by random injection acquired the best wellness esteem.

Merits

The utilization of random injection at a specific emphasis span can expand the estimation of coefficent outline.
Testing the aftereffects of clustering is finished by contrasting the consequences of the KCPSO improve with random injection with master figurings.

Demerits

The most significant limitations of Simple k-means are: The client needs to indicate k (the quantity of groups) in the first place. K-Means can just deal with mathematical data.

CONCLUSION

Data mining assumes a critical function for decision making on a few issues identified with agriculture field. A few creators examined about the function of data mining and its applications in solving the diverse agricultural issues in their work. In this survey paper, it is examined about use of K-means calculation in soil classification and buildup areas of interest by shading pictures, detecting weeds in agriculture. This paper is valuable for explores to investigate the various creators' work and get information about data mining strategies and its applications in agriculture field. Investigation of our boundaries like soil, climate, and water offer effect on the agriculture. Versatile procedures can help minimize negative effects. In this survey paper, it is examined about use of K-means calculation in soil classification and buildup locales of interest by shading pictures, detecting weeds in agriculture. This paper is valuable for explores to investigate the various creators' work and get information about data mining strategies and its applications in agriculture field.

REFERENCES

[1]. Velamakanni Sahithya, Brahmadevara Saivihari, Vellanki Krishna Vamsi, Parvathreddy Sandeep Reddy and Karthigha Balamurugan (2019), “GUI based Detection of Unhealthy Leaves using Image Processing Techniques”,
DOI: 10.1109/ICCSP.2019.8697946,
Electronic ISBN: 978-1-5386-7595-3, IEEE.
[2]. Abdul Hafiz Bin Abdul Wahab, Rahimi Zahari, Tiong Hoo Lim (2019), “Detecting diseases in Chilli Plants Using K-Means Segmented Support Vector Machine, DOI: 10.1109/ICISPC.2019.8935722, Electronic ISBN: 978-1-7281-3663-9, IEEE.
[3]. Tanvi Mehra , Vinay Kumar, Pragya Gupta (2016), “ Maturity and disease detection in tomato using computer vision”, DOI: 10.1109/PDGC.2016.7913228,
Electronic ISBN: 978-1-5090-3669-1,IEEE.
[4]. Yudha Alif Auliya (2019), ” Improve Hybrid Particle Swarm Optimization and K-Means by Random Injection for Land Clustering of Potato Plants” ,
DOI: 10.1109/ICOMITEE.2019.8921207,
Electronic ISBN: 978-1-7281-3436-9, IEEE.
[5]. Chutinan Trongtorkid , Part Pramokchon (2018) ,” Expert System for Diagnosis Mango Diseases Using Leaf Symptoms Analysis”,
DOI: 10.1109/ICDAMT.2018.8376496,
Electronic ISBN: 978-1-5386-0573-8, IEEE.

[6]. Priyadarshini Patil, Nagaratna Yaligar, Meena S M (2017),“ Comparision of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection using Leaf Images”, DOI: 10.1109/ICCIC.2017.8524301, Electronic ISBN: 978-1-5090-6621-6,IEEE.
[7]. Priyanka Padalalu , Sonal Mahajan , Kartikee Dabir , Sushmita Mitkar and Deepali Javale (2017), “Smart Water Dripping System for Agriculture/ Farming ”, DOI: 10.1109/I2CT.2017.8226212, Electronic ISBN: 978-1-5090-4307-1, IEEE.
[8]. Hasmita S, Fhira Nhita, Deni Saepudin, Annisa Aditsania (2019) , “Chili Commodity Price Forecasting in Bandung Regency using the Adaptive Synthetic Sampling (ADASYN) and K-Nearest Neighbor (KNN) Algorithms”, DOI:10.1109/ICOIACT46704.2019.8938525, Electronic ISBN: 978-1-7281-1655-6, IEEE.
[9]. K. L. Ponce-Guevara, J. A. Palacios-Echeverr´ıa, E. Maya-Olalla, H. M. Dom´ınguez-Limaico, L. E. Suarez-Zambrano, P. D. Rosero-Montalvo, D. H. Peluffo-Ordo´nez, J. C. Alvarado-Perez (2017), GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial)”, DOI: 10.1109/ETCM.2017.8247519, Electronic ISBN: 978-1-5386-3894-1, IEEE.
[10]. Kazuo Yonekura, Hitoshi Hattori and Taiji Suzuki (2018), “Short-term local weather forecast using dense weather station by deep neural network”, 10.1109/BigData.2018.8622195, Electronic ISBN: 978-1-5386-5035-6, IEEE.
[11]. Jignasha M. Jethva, Nikhil Gondaliya, Vinita Shah (2018), A Review on Data Mining Techniques for Fertilizer Recommendation, ISSN: 2456-3307, IJSRCSEIT.
[12]. P Jaganathan, S.Vinothini, & P.Backialakshmi (2014), A Study of Data Mining Techniques to Agriculture, IJRIT.
[13]. Mythili. R, Pradeep Raj. D, Survey of Data Mining Techniques and Applications of New Techniques in Agriculture, ISSN (e): 2250-3021, ISSN (p): 2278-8719, (IOSRJEN).