Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 06, June-2019

Review on Yield Estimation

Dr. S. Mohideen Badushah, Mallikarjun H T, Jyothi Lakshmi. S, Syed Rabeya Ameer
Computer Science,
Alva's Institute of Engineering and Technology

***Abstract*:- For human survival, agriculture is important because it serves the basic necessity of human life. A well-known fact is that most of the Indian population (about 55 percent) depends on farming. Because of environmental and climatic variations, there exists an inefficiency in the crop pro- duction of India. In this precarious domain, the farmers and agricultural businessmen have to make numerous decisions every day and various factors affect complexities. An essential issue for the intention of agricultural planning is the precise estimation of yields for numerous plan- ning crops. In this context, achieving desired goals in agriculture-based crop yield has become a challenging task. There are different factors that have a direct impact on crop production and productivity of agricultural products. Predicting crop yields is merely only one of the key factors in farming practices. The user will feed certain key parameters into the system such as temper- ature, humidity, daily rainfall, soil type and presence of macro nutritions to verify the suitable crops which can be cultivable in the particular soil. The soil test cases will also be considered in this regard to validating the system. This paper focuses on the evaluation of agricultural data and discovering optimal parameters to gain maximum cultivation using data mining techniques such as PAM, CLARA, DBSCAN, and multiple linear regression.**

1.0. INTRODUCTION

Agriculture is the foundation of food security and is therefore important. In India, according to recent information, the majority of the population, i.e. above 55 percent, is dependent on agriculture. Agriculture is the field that allows farmers to grow ideal crops in line with the balance of the environment. In India, wheat, and rice together with sugarcane, potatoes, oilseeds, etc. are the major crops grown. India is currently ranked second in the world for farm output. As with the grow- ing economic sector, the farm's economic contribution is diminishing. India is the world's largest producer of many crops, including fresh fruits and vegetables, milk, tissue crops, and several other crops, including cas- tor oil seeds. As we know, agriculture is India's main economic sector that plays a major role in economic growth. India is second in wheat and rice production. In India, wheat and rice alongside sugarcane, potatoes, oilseeds, and so on are the major crops grown. Farmers also grow non-food items such as rubber, cotton, jute and so on. More than 70 percent of rural households depend on farming. The prediction of yield is a ma-jor agricultural issue. Every farmer wants to know how much yield he expects. Predicting yields in the past was carried out by considering the previous experience of a farmer on a particular crop. The data volume in Indian agriculture is enormous. The data is very use- ful for many purposes when it becomes information. This domain provides more than 60percent of the total population with employment and also has a contribu- tion to GDP (about 17 percent). For several reasons, most farmers do not receive the expected crop yield. The yield of agricultural crops depends primarily on the weather. Farmers need constant advice to predict future crop production and analyze that helps farmers maximize their crop production. Agriculture depends on different factors such as climate and economic fac- tors such as temperature, irrigation, cultivation, soil, falling rain, pesticide, and fertilizer. Historical crop yield information provides significant input to compa- nies engaged in this domain. These firms use agri- cultural products as raw materials, feed for animals, production of paper, and so on. An accurate estimate of crop production and risk helps these firms to plan supply chain decisions such as production scheduling. Businesses like seed, fertilizer, agrochemical and agri- cultural machinery are planning production and mar- keting activities based on estimates of crop produc- tion. Farmers ' experience has been the only way in the past days to predict crop yield. Technology penetration into the field of agriculture has led to activities such as yield estimation, crop health monitoring, etc. being automated. The prediction of crop yields has gener- ated considerable interest in the research community, as well as in agriculture and related organizations. Data mining is an important research field for agriculture. Some tools Data mining are more powerful for generat- ing rules from a large amount of data. We can use such tools for generating patterns or knowledge from a large amount of agriculture dataset. Generally, data mining is a technology that extracts the data and summarizes it into useful and accurate information. We proposed to implement one of the data mining tools for analyzing, extracting and predicting the agricultural information. We proposed to use the k-nearest neighbor technique which creates the clusters predict the data sets and pro- vide the summarized and accurate patterns from origi- nal data. Data Mining is widely used for problems with agriculture. Data Mining is used to analyze large data sets and to determine useful data sets classifications and patterns. The Data Mining process ' overall objective is to extract the information from a data set and transform it into a comprehensible structure for further use. Some data mining tools are stronger to derive rules from huge amounts of data. We can use these tools to produce patterns or knowledge from a large num- ber of farm datasets. Data mining is generally a tech- nology that extracts and summarizes data into useful and relevant information. Artificial intelligence, statis- tics, machine

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

learning, and database systems are used in data mining and data analytics techniques. Unsuper- vised and supervised methods are used in data mining. Clusters are formed using large data sets in unsuper-vised learning and based on data sets are performed in supervised learning classification. Data points' are examined in the clustering technique to group them into' clusters' by specific parameters. Compared to data points of different clusters, data points in the same cluster have less distance. The cluster's anal- ysis divides data into well-organized groups. These well-formed groups capture the natural structure of the data. This survey focuses on different methods used to predict crop yield. The methods used are cluster- ing techniques based on density, multiple linear regres- sion, Clustering large applications (CLARA), Petition- ing around Medoids (PAM) and clustering algorithm based on density called DBSCAN. These methods are used to categorize the various Karnataka districts that are producing similar crops. Other techniques of data mining are not discussed in this survey. As pointed out earlier, our focus is on the most extensively used agricultural-related techniques. In addition, this sur- vey is not expected to provide a complete overview of all the data mining techniques used in agriculture. As an instance, statistical based techniques (such as prin- cipal component analysis and regression models) and bi-clustering techniques include some applications in agriculture. Even so, since only the highest rated tech- niques are the focus of this survey, but they will not be taken into consideration in this study.

2.0. STEPS FOR IMPLIMENTATION 2.1.

2.1.1. K-Nearest neighbor technology:

An object is classified by the clear majority of its neighbors, the object being assigned among its clos- est neighbors to the most common class. Recognizing pattern, a nonparametric technique used during catego- rization and regression is the k-nearest neighbor's algo- rithm (k-NN). The input consists of the nearest training examples in the feature space in both cases[1].

We can implement a KNN model by following the below steps:

1. Load the data

2. Initialise the value of k

3. For getting the predicted class, iterate from 1 to total number of training data points

1. Calculate the distance between test data and each row of training data. Here we will use Euclidean dis- tance as our distance metric since it's the most popular method. The other metrics that can be used are Cheby- shev, cosine, etc.

2. Sort the calculated distances in ascending order based on distance values

3. Get top k rows from the sorted array

4. Get the most frequent class of these rows

5. Return the predicted class

[Figure 1 about here.]

2.1.2. Fuzzy Logic:

Fuzzy logic is a mathematical approach that is used to compute and standardize data based on ”degrees of truth” instead of giving true or false values. Fuzzy logic is a form of multi-valued logic in which any real num- ber between 0 and 1 inclusive could be the truth values of variables.It's used to handle the concept of partial truth, where the value of truth may differ between com- pletely true and completely false In Boolean logic, on the other hand, the value of truth of variables may only be integer values 0 or 1[1].

[Figure 2 about here.]

2.1.3. Geospatial analysis:

Geospatial analysis is an approach for applying statistical analysis and other analysis techniques to data which has a geographical or spatial aspect. Is the catalog, display and manipulation of photographs, GPS, satellite photography and historical data, described explicitly in terms of geographic coordinates or implicitly in terms of the street address, postal

code or forest stand documentation as added to geographical models[1].

2.1.4. Application of Data Mining techniques in Agriculture

These techniques are plausible, theoretically well-founded and perform well on more or less artificial test data sets, guess it depends on their ability to make sense of real-world data. Data mining techniques are often used to study soil characteristics. Independent unit analysis techniques for spatiotemporal data mining have also been applied to mine for patterns in weather data. Applied data warehousing and Online Analytical Processing (OLAP) technologies for the suitable use of agricultural data[2].

2.1.5. Application of DBSCAN:

DBSCAN is a base algorithm for permeability-based clustering containing vast amounts of data with noise and outliers. DBSCAN has two variables, Eps and MinPts. DBSCAN has two variables, Eps and MinPts. Conventional DBSCAN, furthermore, cannot produce optimal Eps value. The KNN plot is used to evaluate the epsilon significance where customer input is de- fined throughout the KNN plot (K value). Batchelor Wilkins clustering algorithm is applied to the database and attain the K value along with its respective clus- ter hubs in order to prevent the user identifying the K value as input to the KNN plot[4].

Algorithm Description

For specified values of the epsilon neighborhood epsilon and the minimum number of neighbors minpts required for a core point, the dbscan function implements DBSCAN as follows:

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

1. From the input data set X, select the first unlabeled observation x_1 as the current point, and initialize the first cluster label *C* to 1.

2. Find the set of points within the epsilon neighborhood epsilon of the current point. These points are the neighbors.

a. If the number of neighbors is less than minpts, then label the current point as a noise point (or an out- lier). Go to step 4.Otherwise, label the current point as a core point belonging to cluster *C*.

3. Iterate over each neighbor (new current point) and repeat step 2 until no new neighbors are found that can be labeled as belonging to the current cluster *C*.

4. Select the next unlabeled point in X as the current point, and increase the cluster count by 1.

5. Repeat steps 2–4 until all points in X are labeled.

2.1.6. Partition around medoids (PAM):

It is an algorithm based on partitioning. It breaks the incoming data into numbers of groups. It finds a set of objects which are centrally located called medoids. With the medoids, the closest data points can also be calculated and succeeded in making as clusters[4].

2.1.7. CLARA (clustering large applications):

It is designed by Kaufman and Rousseeuw to handle large datasets, and CLARA (clustering large applications) depends on sampling. Instead of discovering fair representation objects for the entire data set, CLARA draws asample of the set of data, applies PAM to the sample and discovers the sample medoids. CLARA draws different samples for stronger approximations and provides the best grouping as output. Here, the performance of the clustering is measured for accuracy based on an average divergence of all objects in the entire data set[4].

2.1.8. Crop Yield Prediction Using Machine Learning:

It is believed to have such a significant role and significant impact on accelerating crop yield by provid- ing the optimum condition for plant growth and reduc- ing yield gaps and crop damage and end up wasting. Distinct atmospheric conditions and varying harvest parameters i.e. Precipitation rate, minimum, average, maximum and most extreme temperature, cultivable area reference trim, evapotranspiration and year-round yield. Numerous seasonal, financial and genetic factors influence crop production, but unexpected changes in these factors natural consequence in a significant loss for farmers. These risks can be calculated when suit- able mathematical and quantitative model designs are applied to soil, weather and past yield data. Weeds and pests were the major crop destructive microbial agents and farmers need to be well informed in accessing the various data mining technologies to acquire knowledge on the applications of effective weed and pest control strategies and to handle crop damage mitigation tech- niques. Data mining plays a pivotal role in decision- making on distinct agricultural procedures concerns. The objective of data mining methods is to mine un- derstanding from an accessible data set and convert it into a clear and understandable format for some con- siderable application of the Agri process[3].

[Figure 3 about here.]

4.0. CONCLUSION

Farmers are not getting expected crop yield now a day. Farmers need the advice to predict and analyze future crop production in order to achieve the expected crop yield so that farmers can easily make a decision before any crop production. This helps farmers to in- crease the yield of crops. There is a growing number of applications of data mining techniques in agriculture and a growing amount of data currently available from many resources. This is a growing field of research and is anticipated to grow in the future. A lot of work to be done on this emerging and interesting research field. There is a marginal improvement in productivity as a result of technology penetration into agriculture. The innovations have led to an increase in new concepts including digital farming, smart farming, precision farming, etc. Analysis of agricultural soils, the discovery of hidden patterns using data set related to climate conditions and crop yield data was observed. Agriculture field activities are numerous like weather forecasting, soil quality assessment, seed selection, estimation of crop yield, etc. This survey stared at the specific activity, crop yield prediction, and recognized the major developments.

ACKNOWLEDGEMENTS

[1] Annual Crop Yield Prediction and Recommend Planting of Different Crops by Using Data Mining Technique Naushina Farheen Imam Shaikh1, Prof. R. V. Argiddi PG Student, Department of Computer Sci- ence and Engineering WIT, Solapur, Maharashtra, In- dia Asst. Professor, Department of Computer Science and Engineering WIT, Solapur, Maharashtra, India.

[2] Data mining Techniques for Predicting Crop Pro- ductivity A review article S.Veenadhari, 2Dr. Bharat Misra, 3Dr. CD Singh Mahatma Gandhi Gramodaya Vishwavidyalaya, Chitrakoot, Satna, India Central In- stitute of Agricultural Engineering, Bhopal, India.

[3] Agriculture Data Analytics in Crop Yield Esti- mation: A Critical Review B M Sagar, Cauvery N K Department of Information Science Engg, R V College of Engineering, Bengaluru, Karnataka, India

[4] Analysis of agriculture data using data Mining techniques: application of big data Jharna Majumdar*, Sneha Naraseeyappa and Shilpa [5]AnkalakiANALY- SIS Expert System for Agriculture Extension Sujai Das, Laxmikanta Nayak NIRJAFT (ICAR), 12 Regent Park, Kolkata - 700040, India. sujai_das@yahoo.com

[6] Prediction of Crop Yield using Regression Anal- ysis V. Sellam* and E. Poovammal Department of Computer Science and Engineering, SRM University, Kattankulathur – 603203, Tamil Nadu, India; sell- amveera@gmail.com, poovammals@gmail.com

[7] Third Generation Agricultural Support System Development Using Data Mining Jeysenthil.KMS1, Manikandan.T2, Murali.E3 Research Scholar, Dept. of Computing Science and

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

Engineering, VIT university, Chennai, Tamilnadu, In- dia2 Student, M.E CSE, Apollo Engineering College, Chennai, Tamilnadu, India 1 Assistant Professor, Dept of CSE, Apollo Engineering College, Chennai, Tamil- nadu, India3

[8] A survey of data mining techniques applied to agriculture A. Mucherino ÆPetraq Papajorgji ÆP. M. Pardalos Received: 5 December 2008/Revised: 1 May 2009/Accepted: 25 May 2009 Springer-Verlag 2009

List of Figures

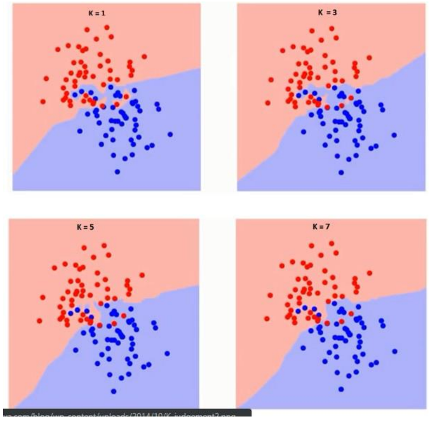

Figure 1. Application of K-NN

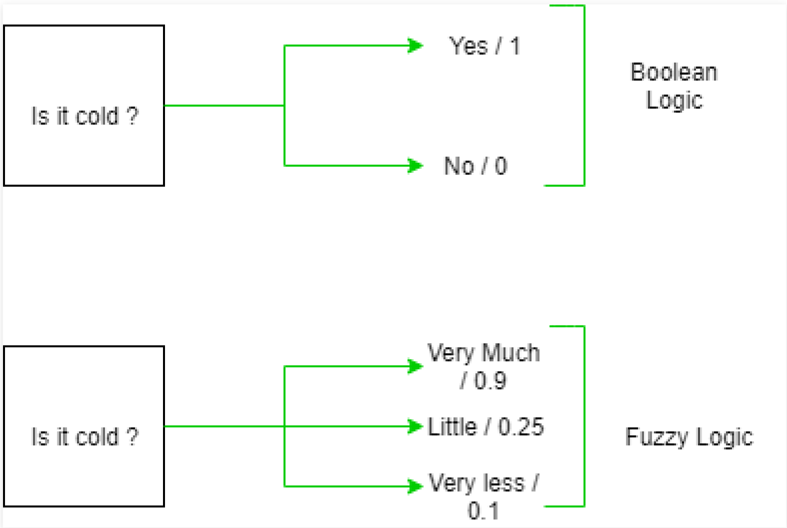

Figure 2. FUZZY LOGIC

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

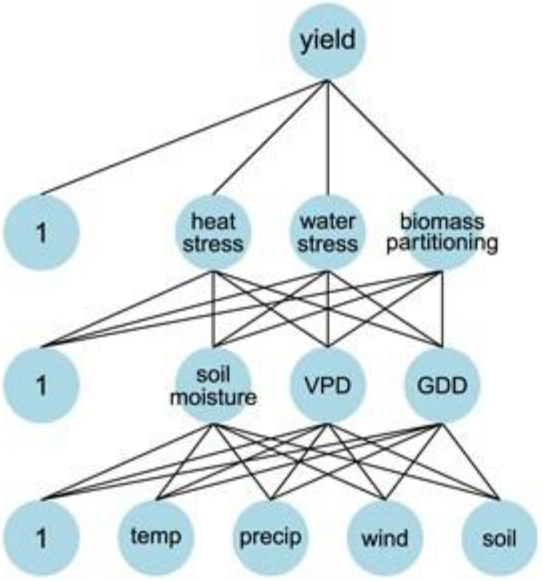

Figure 3. ML- Structure for crop prediction.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)