



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Special Issue 2, December 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569

A Comprehensive Analysis of Agriculture Data Set

Y.Shobana

Assistant Professor, National College, Trichy, Tamilnadu, India

ABSTRACT: India's agriculture is composed of many crops, with the foremost food staples being rice and wheat. India ranks second worldwide in farm outputs. The economy of a country depends on the agriculture. Agriculture is the primary source of livelihood for about 58% of India's population. The GDP growth of country is also depends on agriculture sector. It provides the employment of the nation's workforce. The development of crop depends on various factors such as climate condition, fertilization, water supply, land management. Data mining is an automated process that consists of searching large datasets for patterns humans might not spot. The benefits of data mining are applicable across industries, from supply chains to healthcare, advertising, marketing and also in agriculture sector. The Farmers in agriculture sectors face a lot of issues and difficulties due to the improper understanding and implementation of the activities to enhance their growth and productivity. As we have lots of raw data regarding agriculture, and there is a only need to analyse it properly and mine the data to yield the production rate. In order to achieve the goal we have to focus on proper methodology. Data mining is the most efficient tool to increase the growth rate of production. Data mining is an approach through which in an synchronized manner we can find a workable solution that will be beneficial to increase the growth.

KEYWORDS: Data mining, data analysis, data mining techniques, agriculture, agriculture sector.

I. INTRODUCTION

India's agriculture is composed of many crops, with the foremost food staples being rice and wheat. Indian farmers also grow pulses, potatoes, sugarcane, oilseeds, and such non-food items as cotton, tea, coffee, rubber, and jute. Gross Value Added by agriculture, forestry, and fishing was estimated at Rs. 19.48 lakh crore (US\$ 276.37 billion) in FY20. Share of agriculture and allied sectors in gross value added (GVA) of India at current prices stood at 17.8 % in FY20. Consumer spending in India will return to growth in 2021 post the pandemic-led contraction, expanding by as much as 6.6%. [1]. Despite the overwhelming size of the agricultural sector, however, yields per hectare of crops in India are generally low compared to international standards. Improper water management is another problem affecting India's agriculture. At a time of increasing water shortages and environmental crises, for example, the rice crop in India is allocated disproportionately high amounts of water. [2]. Due to increasing population and growth of industrialization, the land under cultivation is decreasing radically over the years, also agriculture sector is being greatly affected due to population control and climate change. Till date, only a few number of farmers are utilizing new procedures, apparatuses and methods of cultivation for better yield. Agricultural enterprises has the ability to collect and generate huge amount of data, which extracts the required data using automation. By using data mining, beneficial with relevant knowledge and patterns of data can be retrieved. Here data mining, which can be used to study and predict the future traits of agriculture. Data mining is a procedure of separating valuable and significant data from huge sets of data. In our paper, we have explained an outline of application of data mining in agrarian field. Agriculture has got a prime role in Indian economy. Though the share of agriculture in national income has come down, still it has a substantial share in GDP. agriculture occupies a central place in the Indian economy. Its

performance sets the pace of growth in the economy as a whole. It should, however, be noted that Indian agriculture is still in the state of backwardness, the per capita productivity in agriculture is less than in industry.[3] Agriculture is dependent on various factors such as cultivation, irrigation, rainfall harvesting, fertilizers, climate, soil, pesticide, weeds and other factors.[1] Companies in the agricultural sector use data mining to have an estimation of production in order to plan and implement supply chain strategies. Yield prediction is very important in agriculture. The enormous amount of information procured out of these procedures have unexplored potential for improving the effectiveness of the related sectors. Also Data mining discloses hidden information which agricultural management uses for improving its decisions. Data mining is also known as Knowledge Discovery in Data (KDD) and is divided in 2 categories which is Descriptive and Predictive data mining[2].

A Predictive model makes a prediction about values of data using known results found from different data while the Descriptive model identifies patterns or relationships in data. Unlike the predictive model, a descriptive model serves as a way to explore the properties. Predictive model data mining tasks include classification, prediction, regression and time series analysis. The Descriptive task encompasses methods such as Clustering, Summarizations, Association Rules, and Sequence analysis[4]

Data mining methods are :

- Association
- Classification
- Clustering Analysis
- Prediction
- Sequential Patterns or Pattern Tracking
- Decision Trees
- Outlier Analysis or Anomaly Analysis
- Neural Network

II. APPLICATIONS

Segregation level of water in fruits and vegetables

Micro-organisms in a healthy growing state may contain in excess of 80% water. They get this water from the food in which they grow. If the water is removed from the food it also will transfer out of the bacterial cell and multiplication will stop. Partial drying will be less effective than total drying, though for some micro-organisms partial drying may be quite sufficient to arrest bacterial growth and multiplication. Fruits and vegetables are the most utilized commodities among all horticultural crops. They are consumed raw, minimally processed, as well as processed, due to their nutrients and health-promoting compounds. With the growing population and changing diet habits, the production and processing of horticultural crops, especially fruits and vegetables, have increased very significantly to fulfill the increasing demands. Fruits and vegetables have a crucial role in our diet and human life, and therefore the demand for such important food commodities has increased very significantly as a result of the growing world population and the changing dietary habits[5]. Hence data mining technique play a major role in order to increase the productivity, the techniques are Statistical Technique, classification technique and cluster technique.

III. STATISTICAL TECHNIQUE

To access the various Statistical classification techniques, a study is carried out on data mining techniques such as Bayes Net, Naive Bayes, Naive Bayes updatable, J48 and Random forest. Naïve Bayes 100% classified the instances of the data set [1]. Another study was done to compare Bayesian method, fuzzy method and hybrids approaches to extract knowledge from a soil map for classification, and resulted that the learning accuracy of the Bayesian network is lower than that of the fuzzy membership functions method, but higher than Naïve Bayes. On the basis of t-test result, under the confidence level of 95%, the Bayesian network algorithm performs significantly worse

than the fuzzy method, but better than the Naïve. Bayes and the fuzzy method are significantly better than naïve bayes [2a][6].

Clustering technique :Hierarchical Clustering: using SPSS tool,Hierarchical clustering technique on herbal plants was done to groups the data, and to find out meaningful pattern like herbal plant has what type of habits,and the species exist in which location of India [3a]. In a study pest management framework model was developed using the combined techniques of hierarchical clustering and association rule mining [10].

Recursive noise Removal Techniques: Socio-economic factors are the social and economic experiences and realities that help to change one's attitude, personality, perception and life style. Examples of socio-economic factors are education, experience, age, income, wealth, occupation etc. Farmers' socio economy is a neglected area due to the emphasis on the level of agricultural productivity. The paper deals with the role of data mining techniques in the field of agriculture and related areas. Some of the most used data mining techniques such as classification; clustering and statistical techniques along with their applications in various domains of agriculture have been discussed. Data mining techniques in the field of agricultural decision making for forecasting agricultural production, quality of production and management are new research areas. This review is conducted to explore the application of data mining in the domains of agriculture for the purpose of finding some efficient technique to improve the socio-economy of the farmers in order to improve the country economy. Keywords: K-Means clustering, K-Nearest neighbor clustering, Support Vector Machine, Decision Tree.

Data mining is the task of discovering meaningful patterns from large data set. Data mining techniques can be categorized into three main classes: statistical techniques, classification techniques and clustering techniques. A. Statistical Techniques: There are various statistical concepts that are the basis of data mining such as point estimation, Bayes theorem, hypothesis testing, regression and correlation and graph based methods etc. B. Classification Techniques: Classification technique is supervised learning. These algorithms are based on decision tree, neural network, fuzzy set theory, rough set theory, genetic algorithm, k-nearest neighbor, rules based and grid based classification etc. C. Clustering Techniques: A process of grouping a set of physical or abstract object into a class of similar objects is called clustering [25]. Clustering Technique is unsupervised learning. Clustering is of two types hierarchical clustering and partition clustering. For example in farmers' data set with respect to their income three data clusters can be represented in the following manner [Figure 1]. Figure 1. Farmer data with respect to their income, showing 3 data clusters. Each data cluster center is represented with +. II.

IV. STATISTICAL TECHNIQUES

Testing of soil fertility can be done by statistical methods[23].A study is carry out to assess the various Statistical classification techniques of data mining such as Bayes Bayes Net, Bayes Naïve Bayes, Bayes Naïve Bayes updatable, J48 and Random forest. Naïve Bayes 100% classified the instances of the data set [1]. Another study was done to compare Bayesian method, fuzzy method and hybrids approaches to extract knowledge from a soil map for classification, and resulted that the learning accuracy of the Bayesian network is lower than that of the fuzzy membership functions method, but higher than Naïve Bayes. On the basis of t-test result, under the confidence level of 95%, the Bayesian network algorithm performs significantly worse than the fuzzy method, but better than the Naïve. Bayes and the fuzzy method are significantly better than Naïve Bayes [2].

V. CLUSTERING TECHNIQUES

A. Hierarchical Clustering: Hierarchical clustering technique on herbal plants was done, using SPSS tool, to groups the data, and to find out meaningful pattern like what type of habit of herbal plant species exist in which location of India [11]. In a study pest management framework model was developed using the combined techniques of hierarchical clustering and association rule mining [23]. B. Recursive noise Removal Techniques:



B. Recursive noise Removal Techniques:

Recursive Noise Removal Technique was used to find the conditions that lead the farmers to an excessive pesticide usage [15]

C. K-Means clustering: K-Means clustering was used to analyze the agricultural meteorology for the enhancement of crop yields and the reduction of crop losses [11], classifying soils in combination with GPS-based technologies [12] and to classify soils and plants [13]. In another study k-Means method was used to forecast pollution in the atmosphere [16]. K-means approach was used to analyze color images of fruits as they run on conveyor belts [16] and X-ray images of apples were used to monitor the presence of water cores [17].

VI. CLASSIFICATION TECHNIQUE

Decision tree: Decision tree is used to test whether the technique is able to determine the crop yield as a function of physical and chemical soil properties in order to test the reason for low yield [18]. Another study was done using decision tree classification to stimulate farmer's decision rule to catchment region [19].

J48 classifier: J48 Classifier Approach to Detect Characteristic of Bt Cotton base on Soil Micro Nutrient [20]

K-Nearest Neighbour

K-Nearest Neighbour is the simplest algorithms based on Supervised Learning technique. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-nearest neighbor was used for simulating daily precipitation and other weather variables [21]. In another study k nearest neighbor approach was used to evaluate forest inventories and to estimate forest variables analyzing satellite imagery [22].

VII. CONCLUSION

Agriculture is an initial need for every human being, the more we produce the more we survive. It also plays a major part in developing a countries economical growth. Data mining technology and its techniques also serve its part in increasing the productivity of crops , fruits and vegetables, by implicit and explicit support. Since data mining is a knowledge discovery data (KDD) and its mechanism is applied in examining fertility of soil, weather changing, water consistent ect. Farmers' decisions with regard to agriculture production and technology use are strongly influenced by their socio- economic environment. The various algorithms of data mining solve the complex problems of agriculture and provide the accurate results that help the formers for agriculture growth. As by the updation of data mining in future more advance tools and technolo-gies can be developed for predicting crop and more parameters can be added and tested for high rate of productivity.

REFERENCES

- [1] <https://www.ibef.org/industry/agriculture-india.aspx>
- [2] <https://www.nationsencyclopedia.com/economies/Asia-and-the-Pacific/India-AGRICULTURE.html>
- [3] <https://www.yourarticlelibrary.com/agriculture/importance-of-agriculture-in-the-indian-economy/40227>
- [4] https://www.researchgate.net/publication/267558504_Mining_Enrolment_Data_Using_Predictive_and_Descriptive_Ap
proaches
- [5] <https://onlinelibrary.wiley.com/doi/10.1111/1541-4337.12330>
- [6] https://www.researchgate.net/publication/338548344_Data_Mining_Technique_for_Agriculture_and_Related_Area
- [7] 2a--- [2] Qi Feng, "Fuzzyness vs. Probability in a Data Mining a. Application for Soil Classification", Seventh International
b. Conference on Fuzzy Systems and Knowledge Discovery, 2010 IEEE
- [8] https://www.researchgate.net/publication/338548344_Data_Mining_Technique_for_Agriculture_and_Related_Areas
- [9] 3a Rao N.S. and Das K.S "Classification of herbal gardens in a. India using Data mining", Publication of Little Lion
b. Scientific R & D", Journal of Theoretical and Applied
c. Information Technology, vol. 25, no 2, 2011



- [10] Tripathy A.K., Adinarayana J. and Sudharsan D., "Geospatial Data Mining for Agriculture Pest Management – a Framework", Geoinformatics, 2009, pp 1-6
- [11] Rajesh D., "Application of Spatial Data Mining for Agriculture", International Journal of Computer Application, Volume 15-No 2, February 2011.
- [12] Tripathy A.K., Adinarayana J. and Sudharsan D., "Geospatial Data Mining for Agriculture Pest Management – a Framework", Geoinformatics, 2009, pp 1
- [13] Mayer et al., 2004, "Intensified fuzzy clusters for classifying plant, soil and residue regions of interest from color images", Computers and Electronics in Agriculture, Volume 42, Issue 3, March 2004, PP 161-18
- [14] Jorquera H., Perez R., Cipriano A. and Acuna G. "Short term forecasting of Air Pollution episodes in Environmental Modeling", Zannetti, P. (ED). WIT Press, UK, 2001.
- [15] Abdullah A. Brobst S. and Pervaiz I., "Learning Dynamics of Pesticide Abuse through Data Mining", The Australasian Workshop on Data Mining and Web Intelligence, 2004, New Zealand.
- [16] Leemans V, Destain MF, "A real time grading method of apples based on features extracted from defects", J Food Eng 61: , 2004, 83-89.
- [17] Shahin MA, Tollner EW and McClendon RW, "Artificial Intelligence classifiers for sorting apples based on water core", J Agric Eng Res 79(3): 2001, 265-274.
- [18] Canteri M.G., Avila B.C., Santos E.L. dos, Sanches M.K., Kovalechyn D., Molin J.P. and Gimenez L.M., "Application of Data Mining in automatic description of yield behavior in agricultural areas, World congress on computers in agriculture and natural resources.
- [19] Ekasingh B., Ngamsomsuke K., Letcher R.A., Spate J., "A data mining approach to simulating farmers' crop choices for integrated water resource management", Elsevier Ltd., 2005
- [20] International Journal of Computer Trends and Technology (IJCTT) – volume 5 number 6 –Nov 2013, Youvrajsinh Chauhan #1, Jignesh Vania #2 #1 Department of Computer Engineering, LJIET Ahmedabad, India #2 Assistant Professor of L J Institute of Engineering and Technology Ahmedabad, India.
- [21] Rajagopalan b, Lall U., "A K-nearest neighbor simulator for daily precipitation and other weather variables", Water Resource. Res., 1999, 35:3089-3101.
- [22] Holmgren P. Thuresson T, "Satellite remote sensing for forestry planning: a review. Scand J for Res, 1998, 13(1):90-110.
- [23] Ramesh Babu Palepu, Rajesh Reddy Muley, "An analysis of agricultural soils by using Data Mining techniques", 2017 IJESCVo lume 7 Issue no 10



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details