DATA MINING TECHNIQUES FOR IDENTIFYING STRATEGIES FOR PRODUCTIVE AGRICULTURE SOLUTIONS

M. Sumi*, S. Manju Priya

Abstract

Agriculture is the inspiration of the country wide economy, and it performs a crucial position in its growth. Data mining permits farmers to become aware of doubtlessly exciting and unknown styles in huge extent of datasets. Data mining has been engaged with exclusive type of fields which includes medicine, brilliant markets and education. Data mining makes use of taking choices associated with numerous troubles in agricultural subject. Data mining permits farmers to become aware of doubtlessly exciting and unknown styles in huge extent of datasets. Due to farming's need on mined data sources, land and water resources, as well as experts, mining and agricultural are inextricably linked. Additionally, there are employing implications pertaining to areas where mining business acted as a basis to promote agrarian improvement. This collaboration seems to have a variety of effects. There is a warning sign that agri business is growing due to mining in certain areas while decreasing in other, according to the situation in the neighbourhood.

Keywords: Datamining, Agriculture, Association, Clustering, Classification, Regression, K-Means Approach, KNN, Biclustering

I INTRODUCTION

Detaining is a technique that employs Artificial Intelligence, machine learning, statistics and database system to look for patterns in large amount of data. It permits farmers with the resource of the usage of providing historic agricultural output statistics further to a projection, assisting in risk management. Data mining is identifying and visualizing the pattern in massive quantity of datasets. The purpose is defining as precise know-how from massive records set, examine the ones records and remodel the ones records to shape of human comprehensible for further use. When studying the records set keep in mind the distinct perspectives in awesome way and do not have any problem approach to examine the records. Descriptive and Predictive records mining are awesome instructions in records mining to classify the records. Descriptive records mining set out the records regular with latest homes and Predictive records mining permits to be waiting for future values considered the past results. The temperature, soil, water, plant, horticulture plants, medicinal plants, animals and other aspects of Indian agriculture are all quite unique.[1][2]. In India, rural productivities are uncertain since agriculture is copying with the challenge of converting with internal sources, which may be simultaneously reducing crop production.

II METHODOLOGY

The different kinds of algorithms are:

(i) Association: Association is an information mining task that determines the possibility of objects in an accumulation coexisting. Association rules express the relationships between occurring events. Association rules are frequently used to examine transaction changes.

(ii) Clustering: Identifies group of knowledge questions that are each comparable to one another in same way. They differ from members to other groups; group members are nearly identical to one another. In order to find high quality clustering with a low between-group similarity and high internal likeness, the goal of cluster analysis is to identify high quality bunches. Half of the

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

data is subject to bundling, which is similar to grouping. By distributing it to already characterized classes that are expresses during a target, other models distribute information. It does not use an objective. Bunching is used to locate data. If there are many cases and no obvious groups, grouping algorithm will be used to find characteristic groupings. It will also useful as an information preprocessing step in understanding homogenized bunches on which to build regulated models and also for detecting anomalies.[3]

(iii) **Classification:** It might be a capacity for data processing that distributes items during an assortment to particular classification or classes. The objective is to correctly estimate the target class for every data incidence. A group project starts with an information gathering initiative that is well known among the class assignments. Plans are specific and do not encourage planning. A numerical instead of an all-or-nothing aim would be indicated by consistently drifting purpose values. During a discriminating model using a numerical point, a relapse approach rather than a characterization algorithmic program is used. During the model preparation phase, the data characterization algorithm identifies link between the estimates of the indicator and the objectives. Characterization models are tested by comparing expected values to known target values. The information collection is used for model creation and other one is used for model inspection.

 Using the data provided by a collection of recognized samples, classification algorithms are used to categories unknown samples. Because the classification technique is frequently trained on how to execute its classification, this set is known as training set. In order to handle a particular classification problem, neural network and support vector machines, for instances, employ training sets to modify their parameters [4]. K-nearest neighbor is another classification method that does not require learning phase

• **K-Means Approach**

 K means is a data mining clustering method. The objective is to identify a partition of a record that contains similar data in a single group, given a record of unknown rank. A suitable spacing is reflected to degree of similarity between data samples, samples are closed to one another are deemed comparable. The k parameter in the k means algorithm is crucial. The cluster centers can be determined as the mean of every sample that is a part of cluster after the data has been divided into k clusters. Due to its proximity to all of the samples in the cluster and consequent similarity to all of them, the grouping core can be considered to be entire cluster representative. It follows that a group will have similar data if all of its samples are closer to its center than to a center of another group.In this way, the K means is a method moves the matching data samples from their original cluster to new cluster.
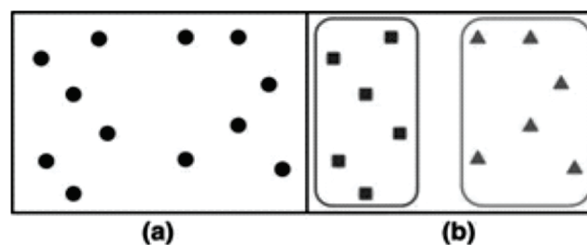

Fig 1: (a) the points are not assigned yet to any cluster; (b) points belonging to the same cluster are marked using the same symbol [5]

• **K Nearest Neighbor (KNN)**

 KNN approach is used for classification. To categorise samples with an unknown classification, a training set is used. Similar to k means technique, the similarities between samples are assessed using appropriate distance functions [5]. The parameter k describes how many comparable known samples were utilized to rate

an unknown sample. Find the k closest known samples by calculating distance between an unknown sample and every other sample in the training set. This approach fairly straight forward classification rule, but its implementation rather might be costly. It can be computationally demanding since it is calculating the distance of each unknown samples from all known samples. The KNN approach makes use of the data in the training set without extracting it.

KNN approach is used for classification. To categorise samples with an unknown classification, a training set is used. Similar to k means technique, the similarities between samples are assessed using appropriate distance functions [5]. The parameter k describes how many comparable known samples were utilized to rate an unknown sample. Find the k closest known samples by calculating distance between an unknown sample and every other sample in the training set. This approach fairly straight forward classification rule, but its implementation rather might be costly. It can be computationally demanding since it is calculating the distance of each unknown samples from all known samples. The KNN approach makes use of the data in the training set without extracting it.
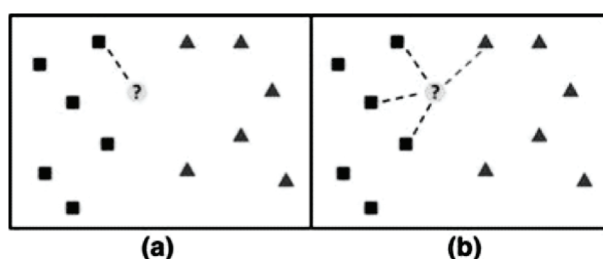

Fig. 2: The classification of the point by the symbol is determined by classification of closest neighbour. (a) When k=1 the unknown point is identified as being a member of the class denoted by squares. (b) When k=4 the unknown point is identified as being a member of the class denoted by square as well[5]

- **Biclustering**

Biclusteringa dataset is actually a classification technique (exploiting information from a training set). Therefore, in the following, the typical expressions used in techniques of classification such as sample classes and also clustering techniques such as division into groups are used.

A data set basically consists of samples that are constituted by a sequence of characteristics that are considered as relevant for the presentation of the samples. Biclustering aims to find simultaneous classifications of samples and their properties. A biclustering can be built by exploiting this training set, if the set is known

The corresponding partition in Biclusters is able to map sample subsets to feature subsets, revealing the features that drive the classification of the training set. This information can be used to perform the classification of samples that do not belong to the training set.

(iv) Regression: Regression is the theory behind the ability to predict variety. Once a regression task begins with an index, the goal esteems are known. The software makes an estimate of the objective as a component of the symptoms for each case with in the fictional data during the model-building process [6]. The connection between indicators and goals are described by a model, which may be then connected to a unique informative index that concern target values. The spread of data points used to measure the discrepancy between expected and actual values.

The Objective of this analysis is:
- To identify the important factors that affect crop output in agriculture by using the information mining technique.

- To promote crop utilizing data processing methods based on environmental characteristics.

III APPLICATIONS OF DATAMINING IN AGRICULTURE

Information technologies have found widespread application in many facets of human life, including agriculture, as a result of profound changes brought about by the modern era. Agriculture has been dubbed "IT Farming" due to the creation and use of new information technologies that enable global networking. Information technologies are progressively helping to solve agriculture challenges in a methodical way [6][7]. Having access to proper data enables the creation of reliable reports, such as those on the use of safety gear, the quantity of time a machine has spent working on a particular crop, or number of seasonal workers hired. It is also simpler to monitor the progress of the task and the information flow. Data Mining is necessary since the agricultural industry generates a wide range of information. Agricultural organisations can provide descriptive and predictive data through data mining to aid in decision making.

Agriculture is a developed industry, where information being added to steadily every day. A significant percentage of this information is presented in the form of written publications, with the majority coming from research on consumer provided data and knowledge that is not passed down through agricultural gene pool. These days, there is a great inclination to make the material available in electronic form, converting information to knowledge, which is not an easy operation. The subject of how effectively analyse agricultural data has grown in relevance due to the increase in agricultural enterprise pricing and growing need to control this cost. All agricultural business desire expert analysis of their agricultural data because it is a future and extremely expensive business. Agricultural businesses are extremely focused on obtaining information about the market for agricultural goods. The success of agriculture firms may depend on their ability to use database data to extract useful information for high quality crops.

Huge amounts of knowledge, as well as details of crops, customers and market are contained in agricultural information system. Useful information patterns may be discovered in this data by using Mining technique which may be then used for further analysis and report review. How to categorise a tonne of information is a crucial question. On the basis of similarities in data, automatic classification is completed. This kind of classification is only useful if the result is inheritable for the expert. The drawback of estimating manufacturing yield may be overcome by data processing methods [8]. As a result of availability of new detector data, all these information generate a collection of data that may be utilized to train methods for methods categorising upcoming production yields.

IV CONCLUSION

In order to make business decisions, agricultural firms and their management constantly work to find information in enormous data. Data Mining can help agricultural business by enhancing management and data analysis. Data mining requires certain technological and analytical techniques such as tracking, reporting and monitoring systems. When data mining is started, information gathering enters never-ending loop.

Users can directly access new and hidden patterns in data using data mining techniques, which enable them to create knowledge that can support agricultural decision making. Agricultural institutions use data mining methods and software for a range of functions. Agronomists utilize patterns to gauge growth of plant, quality of crop and the effectiveness of agro technical interventions.

Using data mining techniques in agricultural business, encourage the creation of conditions that allow for wise decision making.

REFERENCES

[1] S. O. Mezan, S. M. A. Absi, A. H. Jabbar, M. S. Roslan, and M. A. Agam, "Synthesis and characterization of enhanced silica nanoparticle (SiO2) prepared from rice husk ash immobilized of 3-(chloropropyl) triethoxysilanea," Materials Today Proceedings, vol. 42, pp. 2464–2468, 2021.

[2] V. Jagota, M. Luthra, J. Bhola, A. Sharma, and M. Shabaz, "A secure energy-aware game theory (SEGaT) mechanism for coordination in WSANs," International Journal of Swarm Intelligence Research, vol. 13, no. 2, pp. 1–16, 2022.

[3] S. F. Suhel, V. K. Shukla, S. Vyas, and V. P. Mishra, "Conversation to automation in banking through chatbot using artificial machine intelligence language," in Proceedings of the 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 611–618, IEEE, Noida, India, 2020 June.

[4] J. Bhola, S. Soni, and G. K. Cheema, "Recent trends for security applications in wireless sensor networks–A technical review," in Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 707–712, Bharati Vidyapeeth, New Delhi, India, 2019 March.

[5] R. Mythiliand . D. Pradeep Raj, "Survey of Data Mining Techniques and Applications of New Techniques in Agriculture," IOSR Journal of Engineering, vol. 3, no. 30, pp. 31–35, 2018.

[6] B. Kaushik, S. Amit, K. K. Shukla, and B. Rupankar, "Application and scope of data mining in agriculture," International Journal of Advanced Engineering Research and Science, vol. 3, no. 7, pp. 66–69, 2016.

[7] R. S. Kodeeshwari and K. Tamil Ilakkiya, "Different types of data mining techniques used in agriculture–a survey," SSRN Electronic Journal, vol. 4, 2017.

[8] Abdullah, A. and A. Hussain, 2006. Data Mining a New Pilot Agriculture Extension Data Warehouse. Journal of Research and Practice in Information Tehnology, 3 (3): 229-240.

[9] Cunningham, S. J. and G. Holmes, 1999. Developing Innovative Applications in Agriculture Using Data Mining. Department of Computer Science University of Waikato Hamilton, New Zealand

[10] https://www.researchgate.net/figure/a-The-1-NN-decision-rule-the-point-is-assigned-to-the-class-on-the-left-b-the-KNN_fig1_304826093