International Journal of Information Technology, Research and Applications (IJITRA)
Vol. 1, No. 1, May 2022, pp. 19~24
DOI: 10.5281/zenodo.6625831

Spatial Data Mining in Agriculture for Estimation of Crop Price

Deepa S.T[1]
[1] Associate Professor, Department of Computer Science, Shri Shankalal Sundarbai Shasun Jain College, Chennai, India.

Article Info

Article history:

Received April 20, 2022
Revised May 10, 2022
Accepted May 20, 2022

Keywords:

Support Vector Machine
Principle Component Analysis
Spatial Data mining
Cluster Analysis
Geospatial Assortments

ABSTRACT

Advances in computing and data storage have made it possible to access a tremendous amount of data. The difficulty has been to extract knowledge from this raw data; this has resulted in the development of new methodologies and techniques, such as data processing, that will link information knowledge to agricultural yield estimation. The goal of this study was to evaluate these novel data processing techniques and apply them to the database's various variables to see if any meaningful associations could be discovered. These advance forecasts, however, are merely that: estimates, not target estimates. Many subjective assessments are backed by many qualitative criteria in interpreting these estimates. As a result, there is a demand to establish statistically solid objective crop acreage and production estimates.

The study of geographical data is still in its infancy, and a precise approach for rule mining is required. The technique uses a quick algorithm to mine large data sets in a crude manner, enhancing the standard of mining in a reduced data set. The group to which every one of the excess items is most like is relegated, supported by the distance between the thing and consequently the bunch mean. Each bunch's mean is then recalculated. Emphasize the model capacity until it joins. The previously mentioned idea is utilized in agribusiness, where temperature and precipitation are utilized as the underlying geological information, trailed by agrarian examination.

We will examine the uses and methodologies of information mining in agriculture in this study. A few information handling procedures, like K-Means, K-Nearest Neighbor (KNN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM), are utilized for very current uses of data mining methods. Price prediction has become a significant agricultural problem that can only be solved with the use of accessible data. This challenge will be solved using data processing techniques. Various data processing approaches were tested on various data sets in order to solve this challenge.

Corresponding Author:

Deepa. S.T
Associate Professor
Department of Computer Science
Shri Shankarlal Sundarbai Shasun Jain College
Chennai, India
Email: deepatheodoredavid@gmail.com

1. INTRODUCTION:

Farming in India has a long and recognized history. India is as of now the world's second-greatest creator of agricultural things. Agribusiness' money related obligation to India's GDP is perpetually dropping as the's widespread based monetary progression continues.

Indian agriculture is known for its diversity, which is mostly due to differences in resource availability and climate, as well as topography, historical, institutional, and socioeconomic aspects. A lot of the variances stemming from natural variables have been accentuated by policies implemented inside the country and the sort of technology that has been accessible throughout time. Thus, the horticultural area's result execution has been lopsided, with enormous holes in efficiency improvement between various geographic destinations inside the country.

Agribusiness as a business depends exclusively on crop creation, which is impacted by an assortment of climatic, topographical, organic, political, and monetary elements that are essentially inconsequential. This blend of elements makes a gamble. For a fruitful rural activity and reliable food yield, powerful gamble the executives are basic.

Weather, diseases, and pests, as well as harvest operation design, all influence agricultural productivity. Effective management of those components is essential for estimating the likelihood of such an unpleasant situation and mitigating the consequences. For agricultural risk management choices, accurate and trustworthy information about historical crop yield is critical. Historical crop production data is also useful for supply chain operations in enterprises that use agricultural output as a source of raw materials. Rustic items are used as an intergradient in the collecting patterns of tamed creatures, food, animal feed, fabricated materials, poultry, fertilizer pesticides, seed, paper, and a grouping of various organizations. Seed, fertilizer, agrochemical, and cultivating equipment creators plan creation and promoting attempts considering harvest creation measures.

Spatial information handling is the method involved with separating fascinating examples from enormous geospatial assortments. It's the method involved with separating information, spatial connections, and other captivating examples from spatial datasets that aren't explicitly saved. In spatial association analysis, an unique mining optimization method known as progressive refinement is used. The technique enhances the standard of mining in a very trimmed data set by first mining huge data sets roughly with a fast algorithm. For every one of the excess things, the group to which it is most tantamount is designated, upheld by the dispersing between the article and subsequently the bunch mean. It then, at that point, ascertains another mean for each bunch. The standard capacity is iterated until it unites. The k-implies strategy is utilized to gain the temperature and precipitation as the underlying topographical information, and afterward examine the Agricultural meteorology for further developing harvest yields and diminishing yield misfortunes. The information articles can have many angles, but for ease, we'll pick two. The information objects are assembled or arranged by the enhancement of intra-class closeness and minimization of between class similarity standards. Each gathering will be seen as a subset of things from which rules can be inferred. All through this article, all classes and strategies adhere to the java coding style and naming standard. The capacity and refinement names are obvious. The report has been separated into a few areas to guarantee an unmistakable plan, understanding, and proficient coding.

The most common way of extricating significant and significant data from monstrous measures of information is known as information handling. In the field of horticulture, information handling is a significant examination theme. For any farmer, price prediction can be a difficult task because he is the one who must know how much his crops will cost. Accept we approach past information from which different value projections were recorded, and that these recorded value estimates are used to recognize future value figures.

2. DATA MINING TECHNIQUES:

Different data mining approaches have been utilised for a long time. Researchers have explored and commented on ten data processing techniques in depth. The most widely utilised data processing techniques in the agricultural area are presented in this study.

Data grouping and bunching calculations are two distinct sorts of information handling strategies. At the point when an assortment of recognizable examples gives information, characterization methods are utilized to arrange obscure examples. Two grouping strategies that are frequently used to order obscure examples are Support Vector Machines and Neural Networks. The K-Nearest Neighbor (KNN) method has no learning set, but it is the preparation set that is used for characterization, subsequently comparable examples ought to have comparable order all through this system.

The quantity of equivalent realized examples is shown utilizing the K-Nearest Neighbor boundary. Assuming no preparation set is given, the K-Nearest Neighbor calculation is utilized to segment a gathering of obscure information into bunches utilizing grouping methods. The K-Means procedure is perhaps the most widely utilized grouping calculation. In a bunch of information with obscure order, we will observe a parcel of the set where we have similar information coordinated in a similar group. The K boundary in the K-Means calculation indicates the number of groups the information ought to be partitioned into. The way that the focuses of groups can be determined involving the method for all examples in a bunch is the most convincing

justification for picking the K-Means approach. The delegate of the bunch will be assigned the group's center in light of the fact that the center is near all examples. Choosing parameter K, on the other hand, may be one of the K-Means algorithm's drawbacks. Another important factor to consider is the algorithm's computational cost. In rural and unified areas, different information handling procedures like Principal Component Analysis (PCA), Regression Models, and Bi bunching calculations are utilized.

3. APPLICATIONS:

Information has a variety of uses. Mining techniques, such as those linked with meteorological conditions and forecasts, are used in agriculture. For example, we will utilise the K-Means method to forecast air pollution. K-Nearest Neighbor will be utilized to reenact everyday precipitation and other climate factors, and SVMs will be utilized to examine conceivable weather conditions changes.

Sound recognition challenges can also benefit from data processing techniques. Birdsong and different clamors are ordinarily arranged utilizing SVMs. For assessing satellite pictures, K-Nearest Neighbor was utilized to assess backwoods credits and measure woodland inventories. Fake Neural Networks were used to classify eggs as fruitful, and Computer Vision was utilized to perceive breaks in eggs. SVMs were used to detect weeds and nitrogen stress in corn, as well as to classify pizza sauce spread. The K-Means technique is used to categorise soils using GPS-based technology. To classify soils and plants, the K-Means method was utilised. Previously, SVMs were employed to classify crops.

A Neural Network was used to recognize great and rotten ones, and X-beam pictures of apples were utilized to check for the presence of water centers. A supervised Bi-clustering technique was used to select out and see the features of a collection of wine fermentations to estimate the standard of recent fermentations. Taste sensors are used by ANNs to collect data from the fermentation process. SVMs also utilize sensors to detect milk.

In late many years, it has turned into an inexorably imperative piece of our day-to-day routines. Productivity gains can be acknowledged in essentially every business or administration nowadays, and this is particularly evident in agribusiness. A cutting-edge rancher harvests food, yet in addition gigantic measures of information. These are limited scale and exact information. Then again, having a ton of information might be both a gift and a revile. There is an assortment of information available that contains particulars on a specific resource. Here are sure soil and yield includes that rancher should exploit. This is a typical issue for which the expression "information handling" was created. The goal of data processing techniques is to identify important and interesting patterns or information within the data for the farmer.

Yield prediction is a regular specific issue that arises. A farmer who wants to know what quantity output he might expect as early as possible in the season. Farmers' long–term experience with certain yields, crops, and climates was previously utilised to forecast yields. This information might be accessible too, yet it is concealed inside the limited scale. Occasional information that can now be gathered with accuracy in a scope of seasons.

One of the most important requirements for agricultural development is for agricultural production to be upgraded and stabilised at a faster rate. In India, the chances of increasing the realm under any crop are practically nil until crop substitution or increased cropping intensity are restored. Moreover, many plans are contrived to boost the efficiency of different yields in different agro-environment districts, state divisions, credit foundations, seed/manure pesticide organizations, and numerous different accomplices because of the joined impact of many factors, for example, agro-climatic circumstances, asset blessing, innovation level, strategies embraced framework, social and monetary circumstances, and numerous different accomplices p Crop usefulness swings, then again, keep on tormenting the field, unleashing ruin.

Perhaps the main system did by government offices to screen the field's advancement and give protection to the field is assessing the result of different yields. The estimation procedure involves the departments of revenue, agriculture, economics, and statistics. Researchers and a variety of other organisations rely on the data supplied by government agencies. However, these are usually only available in aggregate form, and satellite photographs of crop slate at the taluka level are increasingly being utilised to estimate the realm, but productivity data must be obtained through crop cutting tests.

4. BASIC FACTS:

On account of the universal accessibility of tremendous measures of information and the subsequent pressing need to transform that information into important data and information, information mining has collected a great deal of consideration in the data business and in the public eye all in all as of late. The abilities and comprehension got are consistently utilized, from advertising exploration to misrepresentation identification to assembling control, emergency the executives, and science revelation. Information handling is by and large viewed as a coherent advance forward in the development of information innovation.

Information assortment and data set development, the board (counting information stockpiling and recovery, data set exchange handling, and refined information investigation), and data set exchange handling are altogether including that have arisen in the data set framework business.

5. SPATIAL DATA MINING:

Spatial information handling is the method involved with separating captivating and beforehand obscure however possibly important examples from huge spatial datasets. Getting intriguing and significant examples from spatial datasets is more troublesome than separating similar examples from customary numeric and straight-out information because of the intricacy of spatial information types, spatial connections, and spatial autocorrelation. The quick development of spatial information and broad utilization of spatial data sets underscore the significance of computerized spatial information disclosure. The intricacy of spatial information and inborn spatial connections block conventional information handling procedures for extricating geological examples.

These organizations work in the areas of biology and ecological administration, public wellbeing, transportation, innate science, the study of disease transmission, and climatology, to give some examples.

During an information assortment, affiliation rule digging searches for significant connections between objects. Numerous applications benefit from the disclosure of affiliation joins among a lot of information. The finding of normal thing sets from which solid affiliation rules of the structure A => B are developed, is the initial phase in mining affiliation rules. These principles likewise satisfy a base degree of reliability.

6. CLUSTER ANALYSIS:

Bunch Analysis brings the quest for comparative thing sets to a nearby. Bunching is the act of collection information into classes or groups so that articles inside a bunch have a serious level of comparability yet are altogether different from things in different bunches. Dissimilarities are estimated utilizing the characteristic qualities that portray the articles. The dividing procedure is one such bunching technique, in which an underlying arrangement of k parts is made, where k is the quantity of parcels to make, and afterward an iterative migration system is utilized to further develop apportioning by moving articles starting with one gathering then onto the next.

It is the most generally utilized centroid-based calculation, which takes the information boundary k and partitions an assortment of n things into k groups, bringing about high intra-bunch likeness yet low between bunch similitude. Closeness in gathering is assessed as to the bundle's mean, which can be considered as the gathering's centroid or focal point of gravity.

The k means calculation, which is used in this review, picks k items at irregular, every one of which at first addresses a bunch mean or focus. The cluster to which each of the remaining items is most comparable is assigned, supported by the spacing between the article and hence the cluster mean. Each cluster's mean is then recalculated. Iterate the criterion function until it converges.

The technique searches out the k allotments with the littlest squared blunder capacities. The procedure is in all actuality adaptable and proficient in handling enormous informational indexes on the grounds that the calculation's processing cost is O(nkt), where n is the all-out number of items, k is the quantity of bunches, and t is the quantity of emphases. At the local optimum, the method typically comes to a halt. The most well-known and extensively used centroid-based technique is K-means, which separates a collection of n items into k clusters with low intra-cluster similarity but high inter-cluster similarity using the input parameter k. In grouping, not entirely set in stone by the mean of the bunch's parts, which can be considered the bunch's centroid or focal point of gravity. The bunch to which every one of the leftover things is most equivalent is appointed, upheld by the separating between the article and henceforth the group mean. Each group's mean is then recalculated. Repeat the rule work until it meets.

7. K-MEANS ALGORITHM:

The algorithm for partitioning objects, in which the mid-point of each cluster is represented by an object within the cluster.

Input:

- k: Clusters count,
- D: Set-information that holds n content of the clusters.

Output: k clusters count.

8. RESULTS:

The method searches for k parts with the most minimal squared blunder capacities. Bunches are reduced mists that are appropriately secluded from each other, and they perform well. The procedure is actually adaptable and effective in handling huge datasets on the grounds that the calculation's computational expense is O(nkt), where n is the all-out number of items, k is the quantity of bunches, and t is the quantity of emphasess. The invention of association and sequential patterns is frequently useful as a beginning point for subsequent study in multidimensional analysis, making them a preferred way for grasping data.

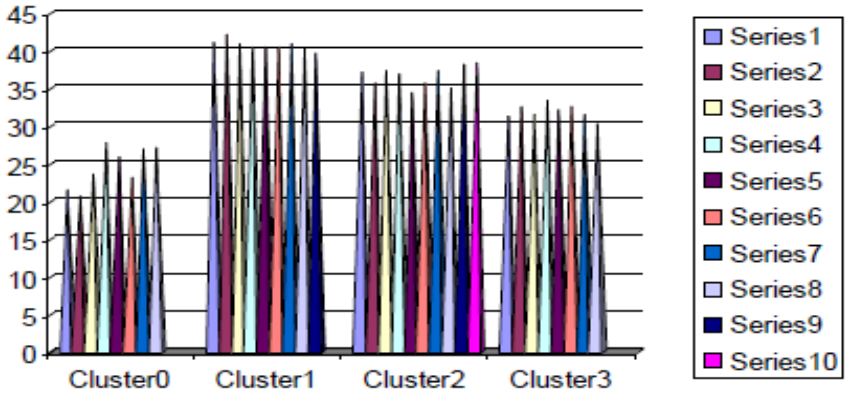

Fig 1 – Temperature Analysis

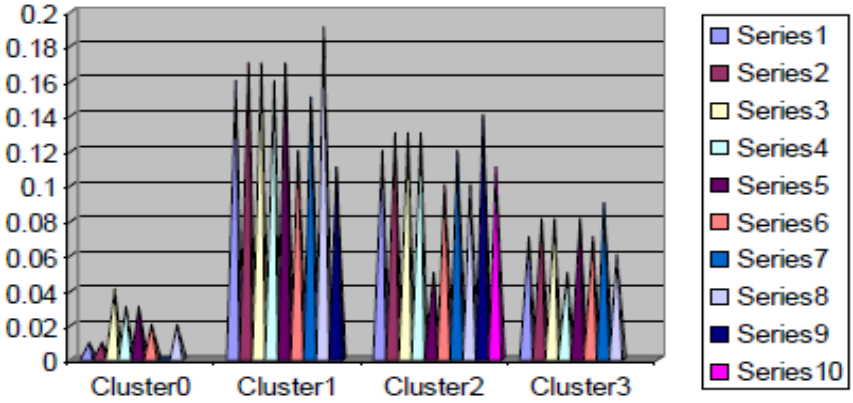

Fig 2 – Rainfall Clustering

9. FORECASTING:

The method involved with causing statements about occasions whose genuine result still can't seem to be seen is known as determining. a standard spot model likely could be assessment of some factor of interest at some characterized future date. Expectation is an expression that is comparable yet more broad. Both might look for counsel from formal measurable methodologies that utilization measurement, cross-sectional, or longitudinal information, or from less formal critical strategies. The expressions "gauge" and "guaging" are at times saved in hydrology for assessment of values at explicit future periods, while the expression "forecast" is utilized for more broad expectations like the times floods will happen over a significant stretch. Guaging and expectation are inseparably connected to hazard and vulnerability; it's not unexpected practice to show the level of vulnerability related with projections. Regardless, the information should be current to guarantee that the visualization is pretty much as exact as could really be expected.

10. FUTURE WORK & CONCLUSIONS:

A few information handling strategies were utilized in this review to appraise crop value examination with accessible information. While there are various different changes that might be made, applications that utilization the K-Means procedure just utilize the essential calculation. Some data processing tools have yet to be used to address agricultural concerns. Regression methods are also used to discover essential information from sets of knowledge connected to agriculture, for example.

Later on, a hereditary calculation based neural organization will be worked for value forecast to further develop exactness. Vegetable worth is anticipated utilizing a BP neural organization forecast model. As an example, we used MATLAB to simulate the market value of onions in Chennai.

As a result, an objective approach to crop forecasting prior to harvest is necessary. This necessitates the creation of more relevant forecast model(s) that outperform existing forecasting approaches. Objectivity and the ability to provide a measure of reliability that a standard forecast approach cannot supply are two of the advantages of the prediction. As a result, realistic pre-harvest agricultural yield forecasting approaches are necessary in India.

Mining association rules from spatial data processing could be a topic with a lot of potential applications. Knowledge mining methods are being studied. This paper outlines the knowledge clustering approach, which employs cluster analysis to identify patterns and rules. "Association Rule Mining Analysis" frequently appears to be a complex, difficult-to-understand concept that is only relevant to scholars and lecturers who wear thick glasses. The reality, on the other hand, is rather different. Regardless of whether we don't know about it, design investigation by means of affiliation rule mining is pervasive in numerous aspects of our day to day routines.

REFERENCES:

[1] M. J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li,. “Newalgorithms for fast discovery of association rules”,Proceedings of the Third International Conference onKnowledge Discovery and Data Mining, page283, 1997.

[2] .Jiawei Han, Member and Yongjian Fu, Member, “MiningMultiple-Level Association Rules in Large Databases”,ieee transactions on knowledge and data engineering, vol11, no.5, September/October, 2000.

[3] Sam Y. Sung, Member, IEEE Computer Society, Zhao Li,Chew L. Tan, and Peter A. Ng, “Forecasting AssociationRules Using Existing Data Sets”, ieee transactions onknowledge and data engineering, vol 15, no.6,November/December 2003.

[4] Chi-Farn Chen; Ching-Yueh Chang; Jiun-Bin Chen **“**Spatiialknowledge discovery using spatial dataminingmethod”,Geoscience and Remote Sensing Symposium, ieeeInternational Volume 8, Issue 25, Page(s): 5602 - 5605 July2005

[5] Eun-Jeong Son, In-Soo Kang, Tae-Wan Kim, Ki-JouneLi,“A Spatial Data Mining Method by Clustering Analysis”,Proceedings of the 6th International Symposium onAdvances in Geographic Information Systems, 1998.

[6] Georg Ruß Data Mining of Agricultural Yield Data: AComparison of Regression Models, ICDM'09, Leipzig,Germany, July 2009

[7] V. Ramesh and K. Ramr Classification of agricultural land soils: Adata mining approach” International Journal on Computer Scienceand Engineering (IJCSE) ISSN : 0975-3397 Vol. 3 No. 1 Jan 2011379

[8] Rainfall variability analysis and its impact on crop productivityIndian agriculture research journal 2002 29,33.,8) SPRS ArchivesXXXVI-8/W48 Workshop proceedings: Remote sensing support tocrop yield forecast and area estimates GENERALIZEDSOFTWARE TOOLS FOR CROP AREA ESTIMATES ANDYIELD FORECAST by Roberto Benedetti A, Remo CatenaroA,FedericaPiersimoni B

[9] “Risk in Agriculture: A study of crop yield distribution and cropinsurance” by Narsi Reddy Gayam Thesis (M. Eng. in Logistics)--Massachusetts Institute of Technology, Engineering SystemsDivision, 2006. Includes bibliographical references (leaves 52-53).

[10] Gazetteer of Kolhapur District (2001)

[11] J. Hartigan, Clustering Algorithms, John Wiles & Sons, New York, 1975.

[12] A. Mucherino, A. Urtubia, Consistent Bi clustering and Applications to Agriculture, IbaIConference Proceedings, Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop “Data Mining in Agriculture” (DMA10), Berlin, Germany, 105-113, 2010.

[13] Fagerlund S Bird species recognition using Support Vector Machines. EURASIP J Adv Signal Processing, Article ID 38637, p 8, 2007.

[14] Holmgren P, Thuresson T Satellite remote sensing for forestry planning: a review. Scand J For Res 13(1):90–110, 1998.

[15] Das KC, Evans MD Detecting fertility of hatching eggs using machine vision II: Neural Network classifiers. Trans ASAE 35(6):2035–2041, 1992.