

Retraction

Retracted: Identifying Smart Strategies for Effective Agriculture Solution Using Data Mining Techniques

Journal of Food Quality

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Journal of Food Quality. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] A. J. B. Suarez, B. Singh, F. H. Almkhtar, R. Kler, S. Vyas, and K. Kaliyaperumal, "Identifying Smart Strategies for Effective Agriculture Solution Using Data Mining Techniques," *Journal of Food Quality*, vol. 2022, Article ID 6600049, 9 pages, 2022.

Research Article

Identifying Smart Strategies for Effective Agriculture Solution Using Data Mining Techniques

Anthony Jesus Bustamante Suarez ¹, **Barjinder Singh**,² **Firas Husham Almukhtar**,³ **Rajnish Kler** ⁴, **Sonali Vyas** ⁵ and **Karthikeyan Kaliyaperumal** ⁶

¹San Luis Gonzaga National University of Ica, Ica, Peru

²Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab 144411, India

³Catholic University, in Erbil, Krg, Iraq

⁴Motilal Nehru College (Evening), University of Delhi, Delhi, India

⁵University of Petroleum and Energy Studies, Dehradun, India

⁶IT at IoT-HH Campus, Ambo University, Ambo, Ethiopia

Correspondence should be addressed to Sonali Vyas; svyas@ddn.upes.ac.in and Karthikeyan Kaliyaperumal; karthikeyan@ambou.edu.et

Received 13 January 2022; Accepted 21 February 2022; Published 13 April 2022

Academic Editor: Rijwan Khan

Copyright © 2022 Anthony Jesus Bustamante Suarez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Agricultural producers and enterprises face a dizzying array of decisions every day, and the many factors that influence them are incredibly complex. Agricultural planning relies heavily on accurately calculating the yields of the various crops that will be used. If you want realistic and successful solutions, data mining is an essential component. Researchers in this study are looking for ways to evaluate agricultural data and extract valuable information from the results in order to increase agricultural output. Use of the CART and random forest algorithms is a data mining technique that may be used to various datasets. It is possible to recognise the effects of various climatic and other factors on agricultural output using the MATLAB software and data mining methods, and a potential strategy is highlighted.

1. Introduction

Agriculture is a key business for the majority of countries. It is the principal food supply for the whole world's population. It does have big challenges such as providing more and better while simultaneously increasing sustainability via smart use of environmental assets, reducing environmental harm, and responding to climate change [1, 2]. Data mining in agriculture is a very modern research topic. Recent technologies are currently able to provide a lot of statistics on agricultural-related activities, which can then be studied in order to find significant knowledge. A related but not the equal term is precision agriculture. Mining and agriculture are straightforwardly interrelated through farming's reliance on mined data sources, land and water assets, and specialists [3, 4]. They are additionally by implication associated where mining firms have improved foundation in a way that

arranges agrarian advancement. The results of this cooperation seem assorted. There is a flag that agribusiness is expanding in a few zones as a result of mining and falling in others, subject on neighbourhood conditions [5].

It cannot be neglected that the shift from the traditional to modern farming systems is crucial. Smart agriculture is one solution for meeting the rising need for food while also contributing to sustainable development. The value of information is increasing in smart farming. Weather, soil, sickness, insects, seeds, fertilisers, and other elements all contribute considerably to the sector's economic and large growth [6, 7].

Agriculture mining: the meaning of agriculture mining remains same as data mining but due to the agriculture data, we give it a different name. Nowadays, there are various research departments and institutes working in this field. They perform a various experiments and analyze the output.

Then, the output of various experiments is compared, and if they are successful in their experiments, then, a similar activity chart is launched to market so that various farmers can get the benefit out of it. In agricultural mining, the researchers analyze the agricultural data of previous 50 years and obtain various trends. Furthermore, these trends are compared and effective predictions are drawn out. In this paper, a technique of information mining and learning revelation process in expansive, appropriated, and heterogeneous databases will be characterized [8]. After data collection, we will be performing some another process, which will produce some trends that can be used for various predictions required for maximizing profits for persons involved in agricultural occupation [9]. The data mining method is crucial in data analysis. Data mining is a computer technique that uses ideas from artificial intelligence, machine learning, statistics, and database systems to find patterns in large amounts of data. It helps farmers by providing historical agricultural output statistics as well as a projection, aiding in risk management. It works with the government to create crop insurance programmes and supply chain policies [10, 11]. The different types of algorithms used in data mining are as follows:

- (i) Association: association is an information mining errand that decides the likelihood of the coevent of things in an accumulation. The relations between cooccurring things are expressed as association rules. Association rules are regularly used to analyze deals exchanges.
- (ii) Clustering: clustering examination finds bunches of information questions that are similar in some sense to each other. The members of a bunch are more similar to each other than they resemble members of different bunches. The point of bunching investigation is to find high calibre bunches with the end goal that the between-group similarity is low and the intragroup resemblance is high. Bunching, similar to grouping, is utilized to part the information. Not at all like arrangement, has grouping models piece information into bunched that were not beforehand characterized. Order models piece information by dispensing it to already characterized classes, which are expressed in a target. Bunching models do not utilize an objective. Bunching is useful for finding information. In the event that there are various cases and no undeniable organizations together, grouping algorithms can be utilized to find characteristic groupings. Bunching can likewise help as a valuable information preprocessing advance to perceive homogeneous bunches on which to fabricate regulated models. Bunching can likewise be utilized for peculiarity recognition. Once the information has been divided in groups, you may find that a few cases do not reasonably well into any groups. These cases are irregularities or anomalies.
- (iii) Classification: classification is a data mining capability that distributes items in a collection to specific

classifications or classes. The goal of the arrangement is to accurately predict the objective class for each occurrence in the data. A grouping project begins with a data collecting phase in which the class assignments are known. Arrangements are specific and do not recommend arranging. Consistent drifting point values would indicate a numerical rather than an all-or-nothing goal. A relapse method, not a characterisation algorithm, is used in a prescient model with a numerical point. A characterisation algorithm identifies relationships between the indicators' estimations and the objective's estimations during the model development (preparation) phase. In an arrangement of test data, characterisation models are tested by contrasting expected esteems with known target esteems. The old data for an order endeavour is usually separated into two informational collections: one for model development and the other for model inspection.

- (iv) Regression: the ability to forecast a number is known as regression. The objective esteems are known when a regression work starts with an informational index. The estimation of the objective as a component of the indicators for each case in the fabricated information is surmised by a regression algorithm during the model form (preparing) procedure. These relationships between indicators and targets are defined in a model, which can subsequently be linked to a different informative index that hides target values. Regression models are tested by logging a variety of data points that quantify the difference between expected and actual values.

The main objectives of this research are stated below:

- (i) To identify the key components which effect the crop production in agriculture using the data mining technique
- (ii) To recommend crops on basis of environmental attributes using data mining techniques.

Section II discusses literature review about different techniques used for crop selection in agriculture.

2. Literature Review

Data mining techniques are essential for developing practical and viable solutions to this problem. Agriculture has long seemed to be a good fit for big data. Environmental considerations, soil variability, input volumes, combinations, and commodity pricing have all increased the need of farmers using data and seeking help when making critical agricultural decisions [12, 13]. This section provides an overview of studies on the use of data mining tools for agricultural decision-making. Artificial neural networks, Bayesian networks, and support vector machines are among the data mining approaches discussed in the paper [14]. The review has identified a number of effective strategies for recognizing the effects of various climatic and other

conditions on crop output. This review argues that more research is needed to understand how these strategies may be applied to complex agricultural datasets for crop yield prediction using GIS technologies that include seasonal and geographical elements [15, 16].

ANN interrelationships of linked variables can be used to build a model. These symbolically reflect the human brain's connected processing neurons or nodes that are utilized. A large number of input and output examples are utilized to progress a formula to uncover the relationship in order to construct a model prediction [17, 18]. Nonlinear relationships, which are overlooked by conventional prediction approaches, can be determined with little a priori information about the functional relationship. Bayesian Networks are a probabilistic technique for expressing beliefs and knowledge, which is particularly useful for systems with exceedingly complex structures and functional relationships [19,20]. In contrast to deterministic comparisons, BN leverages the probabilistic components of a structure to characterise the relationships between variables. SVMs are supervised machine learning methods that are relatively new. SVMs have been used in agriculture in a number of researches. Modeling of urban land use conversion was the application for SVM. The link between rural-urban land use change and a variety of characteristics was discovered in this study. By giving the structures contribution analysis for agricultural yield prediction, SVMs were also used to deliver visions into crop response patterns connected with climate circumstances.

In [21], the casing work of the pest administration framework utilizing information mining strategies is termed, which concentrates on giving authentic information, current and prescribed pest, pesticide information, and to be repeated pest models up to the cultivate level. In this work, an exertion has been made to indicate how geospatial information mining combined with agricultural counting irritation examination, pesticide, and climatological parameters are valuable for the advancement of pesticide utilization and better administration. The results will uncover energizing examples of agriculturist rehearses sideways with pesticide use elements both in the spatial and nonspatial way and can help to make out the clarifications for pest and pesticide abuse [5, 22].

The research paper [23] presents spatial information mining technique particularly decision tree algorithm applying to agriculture arrive reviewing. The clue is to pool spatial information mining/decision tree strategies with master framework strategies and apply them to shape a savvy agriculture arrive reviewing data framework. The creator execute the decision tree c4.5 algorithm and actualize with Mo2.0 and VC++6.0 to assemble agribusiness arrive evaluating master framework. Likewise, an investigation is offered to demonstrate the specific advantages of this system in tending to troubles in arrival evaluation, for example, missing area data and challenges in the quantitative investigation of components [8, 24, 25].

In [26], the authors offer a strategy for information mining and learning disclosure in huge, appropriated, and heterogeneous databases. Keeping in mind the end goal to

pick up conceivably energizing examples, connections, and standards from such vast and heterogeneous information accumulations, it is important that a procedure be built up to take profit from the suite of existing methodologies and apparatuses available for information mining and learning revelation in databases [27, 28].

In [29, 30], a standalone data mining tool is employed in cluster analysis or as a preprocessing step for other algorithms to achieve data dispersion. Clustering is another word for unsupervised learning. "Hidden patterns" are used in cluster analysis for machine learning.

In the field of agriculture, different forecasting approaches have been created and assessed by scholars all over the world. The following are a few examples of such studies.

An agricultural data from 1965 to 2009 in the Andhra Pradesh area of East Godavari is one such approach. The K means clustering method is used to divide the rainfall data into four groups. The process of modelling the linear connection between a dependent variable and one or more independent variables is known as multiple linear regression (MLR). Rainfall is the dependent variable, while year, sowing area, and productivity are the independent factors. The goal of this research is to find appropriate data models with high accuracy and generality in terms of yield prediction capabilities [31, 32].

From the literature review [33–36], we identify four key components which affect the agriculture crop which are stated below:

- (i) Component 1: Cluster Type-1 depends on the accompanying properties: rainfall, least temperature, greatest temperature, dampness, and daylight. These are ecological or climatic qualities considered for our examination. The level of similitude of the gathering of these traits ought to demonstrate unmistakable bunches for the chose areas.
- (ii) Component 2: Cluster Type-2 depends on the accompanying traits: soil pH and soil saltiness. As mentioned before, these biotic variables contribute to a great extent towards the forecast of the crops. The closeness of the estimations of these traits ought to likewise show independent bunches.
- (iii) Component 3: Cluster Type-3 depends on the irrigated zone. Bunching depending on the range characteristics for each region was considered in light of the fact that we can get different bunches in light of particular scopes of territories that were watered for each region.
- (iv) Component 4: Cluster Type-4 depends on the individual harvest yields of potato and wheat. This sort of grouping was considered so as to order the areas into isolate groups with comparable harvest yields and after examination of the outcomes, to see regardless of whether they show an example identified with impacts from the chosen properties.

Using the details and information gathered in Section II, a simple methodology is constructed for crop selection which is discussed in the next section.

It will solve a variety of agricultural problems. This is a novel strategy that holds a lot of potential for assisting firms in focusing on the most important information in their big data. Tools data mining may be used to foresee future trends and behaviours, making it possible to the capacity of a firm to make pre-emptive decisions based on information. The use of data mining has resulted in automated prospect analysis go beyond past event analysis by using retrospective approaches. This paper provides the following novel approach for crop selection in agriculture is as follows:

- (i) It uses type of soil, temperature, water resource, rainfall, and moisture as key components.
- (ii) To give strategically balanced crop selection, two data mining methods are combined namely classification and regression trees (CART) algorithm and random forest algorithm
- (iii) The goal of the research is to use online multilayer soil data and satellite imaging crop growth parameters to predict within-field variation in wheat yield.
- (iv) Using an unsupervised learning technique, supervised self-organizing maps capable of managing existing information from several soil and crop sensors were used.

3. Methodology for Crop Selection Using Data Mining Techniques

This section provides a brief approach regarding crop selection by farmers using data mining technique. The two possible techniques used to model the predictive system are classification and regression trees (CART) algorithm [37] and random forest algorithm [38, 39]. Before using these techniques, data collection and interrelation is performed as stated below.

3.1. Data Collection. For data collection, we have created a Google form. The Google form was shared with different universities students and also a place nearby by an area known as Sant Nagar. The participants gave the inputs to the questions asked in the form. The data which they have given is related to their hometown where their parents are living. The questions were asked about crops which they grow at their place. If there is anything that they did not understand, they asked their parents and then filled the form. Some of the students also filled the data twice as one crop for summer and one crop for winter which they grow in their hometowns. The survey form consists of following details as stated in Table 1.

3.2. Data Cleaning, Extraction, and Transformation. After the forms were submitted by the students and other people, the data were transformed into Google sheets. Later on, the data were visualized and the missing values were filled with the help of an expert. Where any possible solution was not available, then, that particular input was discarded. It was done because in later stages, it will not create any problem

TABLE 1: Data survey form details.

Details asked	Requirements
Name	Full name
Mobile number	Active mobile number
Major crop	Name of one major crop cultivated
Location	State/District/Village
Temperature range	Categories in four groups: (i) Less than 10°C (ii) Between 10–20°C (iii) Between 20 and 30°C (iv) Above 30°C
Soil type	Type of soil
Seed type	Categories in two: (i) Homemade (ii) Market buy

which will not be able to be solved at that particular time. Then, the data were downloaded in Excel sheet and csv file. This was done in order to fulfill the requirement of various data mining algorithms.

3.3. Proposed Integrated Data Mining Approach. This research proposed to integrate two data mining techniques, i.e., classification and regression trees (CART) Algorithm and random forest algorithm to get the best possible results in crop selection for a specific area and time. The proposed integration flow chart is shown in Figure 1.

Classification and regression trees (CART) algorithm and randomized forest algorithm are two techniques that might be utilized to describe the prediction system. A classification and regression trees (CART) is a machine learning prediction technique. It describes how the numbers of a target variable can be anticipated based on some other variables. It is a tree structure in which each fork is divided into predictor variables, and each node just at end has a forecast for the target attribute. The CART algorithm is a tree-based method which is at the heart of machine learning. It also serves as the foundation for more sophisticated machine learning techniques such as bagged decision forests, random forests, and boosted tree structure. The proposed approach has following steps:

- (i) Selection: data are selected from various databases.
- (ii) Preprocessing: incomplete, inconsistent data are completed.
- (iii) Transformation: transforming raw data into the required form.
- (iv) Data mining: after analyzing, data patterns are drawn.
- (v) Interpretation: patterns are interpreted to get knowledge.

CART stands for classification and regression trees, which are generally known as decision trees. Decision trees have been around for a long time and are useful in machine learning for predictive modelling. These trees are used to solve classification and prediction problems, as the name implies. Figure 2 shows the approach of the CART method.

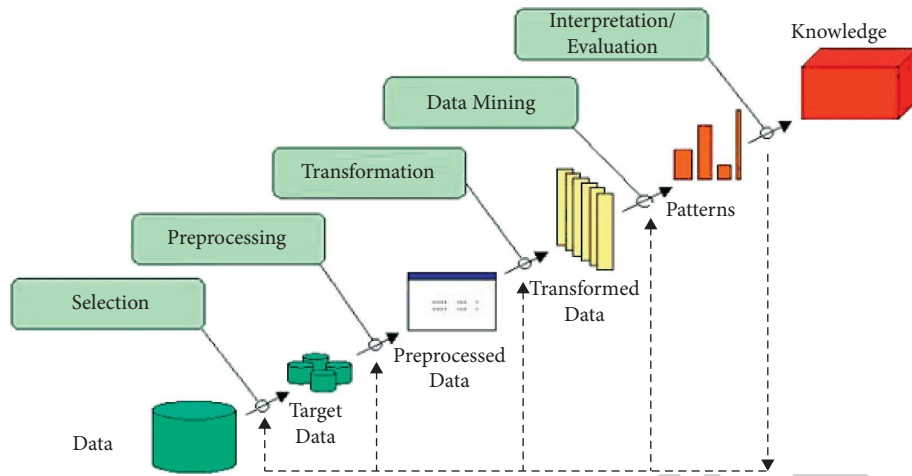


FIGURE 1: Proposed data mining approach.

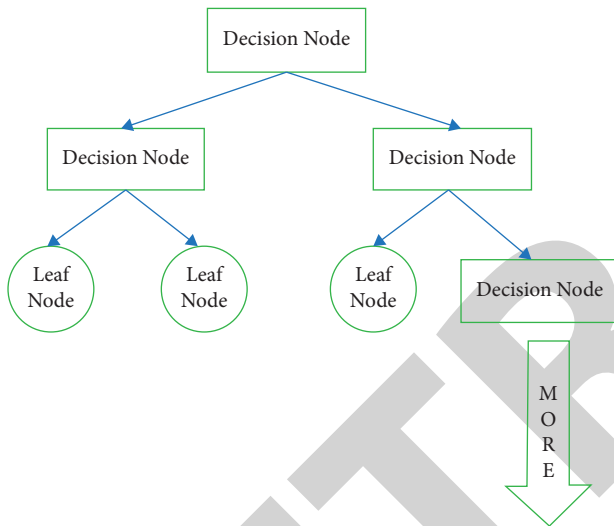


FIGURE 2: CART flow chart.

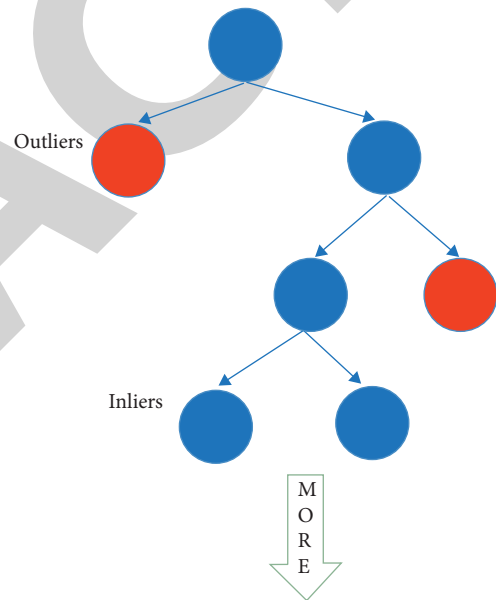


FIGURE 3: Isolation random forest approach.

It divides each of the data into subtrees at the microlevel so that the details can be extracted.

In a similar manner, the isolation forest “isolates” observations by selecting a feature at random and then selecting a split value between the feature’s maximum and minimum values at random. The number of splitting necessary to isolate a sample is equal to the path length from the root node to the terminating node because recursive partitioning can be represented by a tree structure. This path length is a measure of normalcy, and our decision function averaged over a forest of such random trees. Anomalies’ pathways are substantially shorter when random partitioning is used. As a result, when a forest of random trees produces shorter path lengths for specific samples, they are almost always outliers. Figure 3 shows the isolation random forest approach.

In both the technique, the tree formation is used, so it makes the comparison and decision making easily for this research paper.

Using these two techniques of data mining, the combined methodology flow chart is shown in Figure 4.

4. Result

Cultivation of a specific crop on land that does not match the crop’s minimal requirements will result in lesser-quality yields and lower revenue for the farmer. In agriculture, it is necessary to determine the quality of the soil. This provides information about the proportions of nutrients and minerals in the soil. Alkalinity, salinity, moisture content, and other elements influence the soil’s quality. Various types of soil are studied using data mining. Based on the fertility of the soil, soil data analyzers recommend the type of crop to be planted and harvested to maximise output. In this section, we will be discussing the results which are concluded after the methodology opted by us. The tools used in order to achieve the results are R Studio and MATLAB.

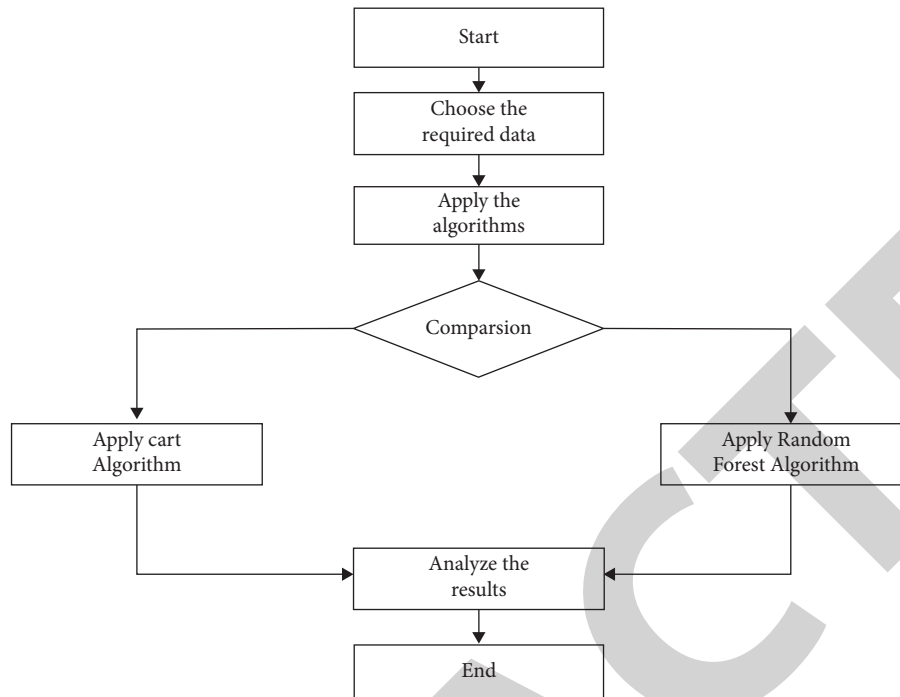


FIGURE 4: Proposed methodology approach.

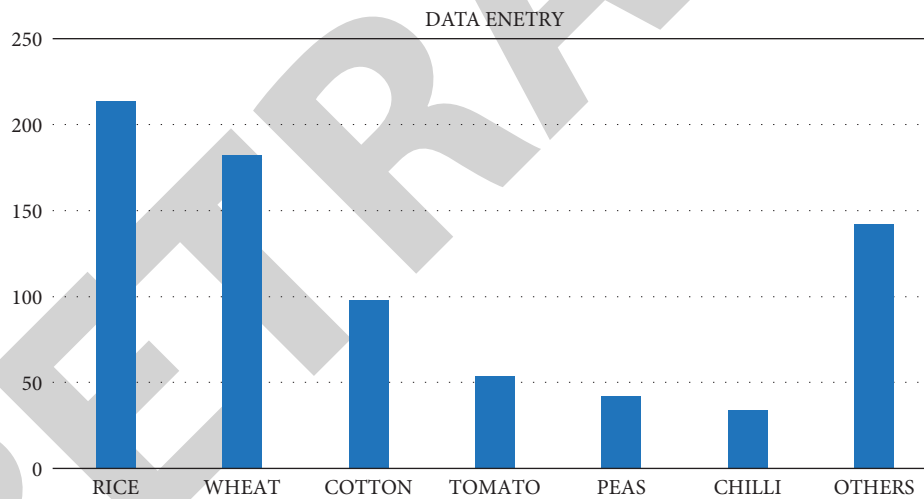


FIGURE 5: Survey details of crop.

A total of 766 survey data entries were taken to predict and justify the proposed methodology. The crops details are shown in Figure 5, which consist of rice, wheat, and others. Figure 6 shows the survey details of temperature. It shows the major portion have high temperature range. Figure 7 gives the details of type of soil. From the survey, it is clear that sandy loam soil is found the most. For the seed uses, more than 70% population uses the homemade seeds for crop cultivation.

The classification and regression trees (CART) algorithm results are shown in Table 2. It can be seen from the table that all have the same value. This shows that all have the same chance of growth. The isolation forest also shows the same results as shown in Table 3.

So, the data provide by the users give detailed information.

Using the data mining technique, it shows that the majority of the data suggest the users to go from the three crops, i.e., banana, bengal gram, and black gram. The comparison in Figure 8 shows that isolation forest is more accurate than CART. However, by combining both, we can predict more accurately which can benefit the farmers.

5. Discussion

In this paper, we have discussed a number of challenges in the agricultural sector and how data mining might assist in solving them. Many authors' works are discussed as well as

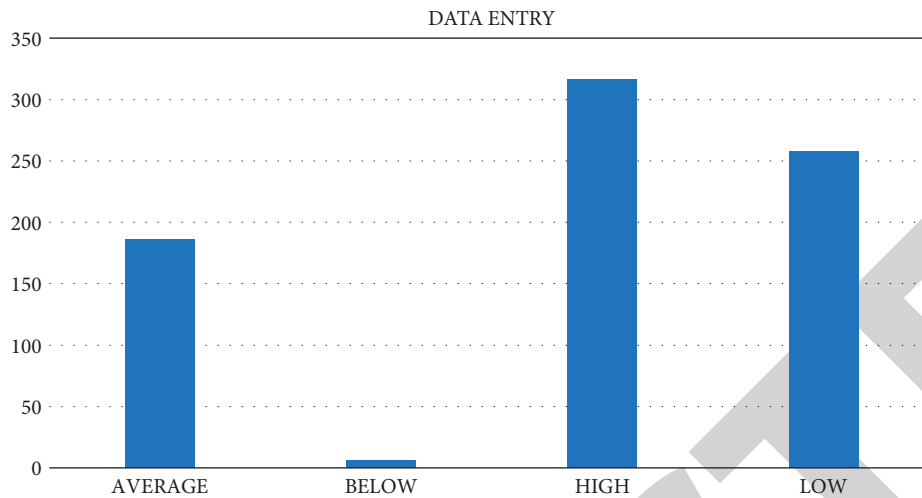


FIGURE 6: Survey details of temperature.

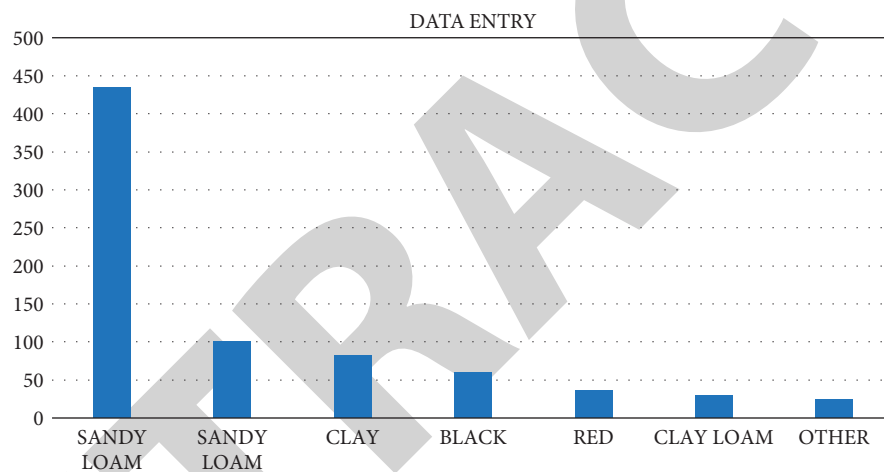


FIGURE 7: Survey details of type of soil.

TABLE 2: CART result.

Crop	Sensitivity	Specificity	Prevalence	PPV	NPV	DR	DP	Accuracy
Banana	0	1	0.005	Nan	0.994	0	0	0.5
Bengal gram	0	1	0.005	Nan	0.994	0	0	0.5
Black gram	0	1	0.005	Nan	0.994	0	0	0.5

TABLE 3: Isolation forest result.

Crop	Sensitivity	Specificity	Prevalence	PPV	NPV	DR	DP	Accuracy
Banana	1	1	0.005	1	1	0.005	0.005	1
Bengal gram	1	1	0.005	1	1	0.005	0.005	1
Black gram	1	1	0.005	1	1	0.005	0.005	1

the role of data mining in this subject. In order to make the best decisions possible, agricultural institutions use data mining technologies for a number of objectives, including problem prediction, disease diagnosis, and pesticide optimization. As a result, we can say that data mining has helped the agriculture industry.

In this case, data mining proved to be beneficial in making decisions about a number of agricultural difficulties. By increasing data analysis and administration, data mining can assist related organizations in achieving more benefits. Users can also use data mining to look for hidden patterns in data.

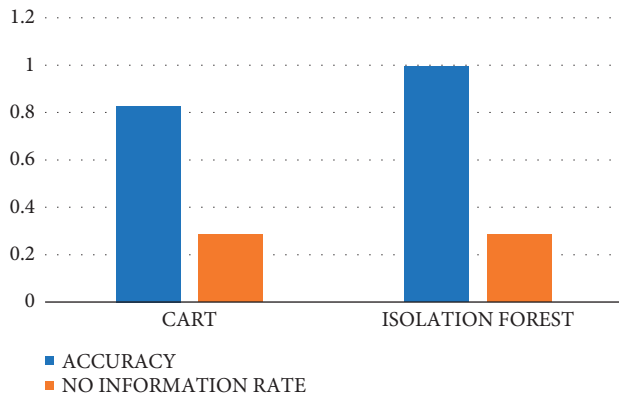


FIGURE 8: Comparison of data mining technique.

As a result of using the CART decision tree in our research, we were able to achieve an accuracy of 82.7 percent. This accuracy can be improved by incorporating some of the improvements made to the CART algorithm that we employed. We also used the random forest tree, which has a 99.46 accuracy rating. The random forest tree's accuracy is significantly superior to that of any other decision tree. A collection of trees is used in random forest tree, and the outcome is calculated at the end. As a result, when compared to other decision trees, the random forest tree produces more accurate results.

6. Conclusion

India is a country in which agriculture plays a vital role. In the success of the farmers, grows the nation. Agriculture mining will aid us in the study of diverse data related to agriculture. In this method, farmers may choose the proper seed depending on soil requirements to boost productivity and obtain profit out of such an approach. Thus, the farmers can plant the proper crop improving his production and also aggregate the overall productivity of the nation. Not only farmers, the other folks interested in other fields associated to agriculture such as market vendors and financiers can take profit from this proposed method. We will use several tendencies which are present in the prior data collection and that will assist us to deliver the efficient and more realistic outcomes for different elements. Furthermore, these outcomes can be employed by many experts who want to involve themselves in the subject of agriculture to reach their appropriate goals. As a result of using the CART decision tree in our research, we were able to achieve an accuracy of 82.7 percent. This accuracy can be improved by incorporating some of the improvements made to the CART algorithm that we employed. We also used the random forest tree, which has a 99.46 accuracy rating. The expert system's scope can be expanded. The inputs can be increased. Inputs include soil type and fertility. The climatic and environmental parameters can be supplied. Increasing input parameters increases output precision.

Data Availability

The data used are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

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) triethoxysilane," *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] 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.
- [4] R. S. Kodeeshwari and K. Tamil Ilakkiya, "Different types of data mining techniques used in agriculture—a survey," *SSRN Electronic Journal*, vol. 4, 2017.
- [5] N. Mirjankar and S. Hiremath, "Application of data mining in agriculture field," *International Journal of Computer Engineering and Applications*, vol. 10, no. 6, 2016.
- [6] 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.
- [7] 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.
- [8] H. Patel and D. Patel, "A brief survey of data mining techniques," *Applied to Agricultural Data International Journal of Computer Applications*, vol. 95, no. 9, 2014.
- [9] R. B. Palepu and R. R. Muley, "An analysis of agricultural soils by using Data Mining techniques," *International Journal of Electrical Systems and Control*, vol. 7, no. 10, 2017.
- [10] X. Huang, V. Jagota, E. Espinoza-Muñoz, and J. Flores-Albornoz, "Tourist hot spots prediction model based on optimized neural network algorithm," *International Journal of System Assurance Engineering and Management*, pp. 1–9, 2021.
- [11] G. Verma and R. Chakraborty, "A hybrid privacy preserving scheme using finger print detection in cloud environment," *Ingénierie des Systèmes d'Information*, vol. 24, no. 3, pp. 343–351, 2019.
- [12] S. N. H. Bukhari, A. Jain, E. Haq et al., "Machine learning-based ensemble model for zika virus T-cell epitope prediction," *Journal of Healthcare Engineering*, vol. 2021, Article ID 9591670, 10 pages, 2021.

- [13] S. Deshmukh, K. T. Rao, and M. Shabaz, "Collaborative learning based straggler prevention in large-scale distributed computing framework," *Security and Communication Networks*, vol. 2021, Article ID 8340925, 9 pages, 2021.
- [14] T. Alberto, X.-P. Burgos Artizzu, G. Pajares, and A. Ribeiro, "A vision-based hybrid classifier for weeds detection in precision agriculture through the bayesian and fuzzy k-means paradigms," *Advances in Soft Computing*, vol. 44.
- [15] C. C. Yang, S. O. Prasher, J. A. Landry, and H. S. Ramaswamy, "Development of an image processing system and a fuzzy algorithm for site-specific herbicide applications," *Precision Agriculture*, vol. 4, p. 518, 2003.
- [16] R. Mythili and . 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.
- [17] P. Shalin, "St. Josephs college: a survey on data mining techniques in agriculture," *International Journal of Engineering Research & Technology*, vol. 3, no. 30, 2015.
- [18] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.
- [20] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 955–974, 1998.
- [21] S. Busygin, O. A. Prokopyev, and P. M. Pardalos, "Feature selection for consistent biclustering via fractional 0-1 programming," *Journal of Combinatorial Optimization*, vol. 10, no. 1, pp. 7–21, 2005.
- [22] R. L. Brown, "Accelerated template matching using template trees grown by condensation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, no. 3, pp. 523–528, 1995.
- [23] K. Brudzewski, S. Osowski, and T. Markiewicz, "Classification of milk by means of an electronic nose and SVM neural network," *Sensors and Actuators B: Chemical*, vol. 98, no. 2-3, pp. 291–298, 2004.
- [24] G. Camps Valls, L. Gómez-Chova, J. Calpe-Maravilla, E. Soria-Olivas, J. D. Martín-Guerrero, and J. Moreno in *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis*, vol. 2652, pp. 134–141, Springer, Berlin, Heidelberg, 2003 June.
- [25] G. Castellano, A. M. Fanelli, and M. Pelillo, "An iterative pruning algorithm for feedforward neural networks," *IEEE Transactions on Neural Networks*, vol. 8, no. 3, pp. 519–531, 1997.
- [26] A. Chedad, D. Moshou, J. M. Aerts, A. Van Hirtum, H. Ramon, and D. Berckmans, "AP-animal production technology," *Journal of Agricultural Engineering Research*, vol. 79, no. 4, pp. 449–457, 2001.
- [27] M. N. Kumar, V. Jagota, and M. Shabaz, "Retrospection of the optimization model for designing the power train of a formula student race car," *Scientific Programming*, vol. 2021, Article ID 9465702, 9 pages, 2021.
- [28] S. Saralch, V. Jagota, D. Pathak, and V. Singh, "Response surface methodology based analysis of the impact of nanoclay addition on the wear resistance of polypropylene," *The European Physical Journal Applied Physics*, vol. 86, no. 1, Article ID 10401, 2019.
- [29] M. Sehgal and D. Bhargava, "Knowledge mining: an approach using comparison of data cleansing tools," *Journal of Information and Optimization Sciences*, vol. 39, no. 1, pp. 337–343, 2017.
- [30] R. Devkota, L. P. Pant, H. N. Gartaula et al., "Responsible agricultural mechanization innovation for the sustainable development of Nepal's hillside farming system," *Sustainability*, vol. 12, no. 1, p. 374, 2020.
- [31] H. Ait Issad, R. Aoudjit, and J. J. P. C. Rodrigues, "A comprehensive review of Data Mining techniques in smart agriculture," *Engineering in Agriculture, Environment and Food*, vol. 12, no. 4, pp. 511–525, 2019.
- [32] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [33] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [34] K. C. Das and M. D. Evans, "Detecting fertility of hatching eggs using machine vision II: neural network classifiers," *Transactions of the ASAE*, vol. 35, no. 6, pp. 2035–2041, 1992.
- [35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from Incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1–22, 1977.
- [36] V. Susheela Devi and M. N. Murty, "An incremental prototype set building technique," *Pattern Recognition*, vol. 35, no. 2, pp. 505–513, 2002.
- [37] C.-J. Du and D.-W. Sun, "Pizza sauce spread classification using colour vision and support vector machines," *Journal of Food Engineering*, vol. 66, no. 2, pp. 137–145, 2005.
- [38] R. Timofeev, *Classification and Regression Trees (CART) Theory and Applications*, pp. 1–40, Humboldt University, Berlin, Germany, 2004.
- [39] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, IEEE, NW Washington, DC, USA, 2008* December.