Journal of Advances in Science and Technology
Vol. 21, Issue No. 1, March-2024, ISSN 2230-9659

DOI: https://doi.org/10.29070/g2syer82

A comprehensive review of data mining in the agricultural sector in India

Rohit Kumar Raghuwanshi [1] * , Ravindra Kumar Tiwari [2]

1. Research Scholar, Department of Computer Application, LNCT University, Bhopal, Madhya Pradesh, India
rohitmca70@gmail.com ,
2. Associate Professor, Department of Computer Application, LNCT University, Bhopal, Madhya Pradesh, India

Abstract: India's agricultural production is second to none in the world. A large part of India's social and economic fabric is based on agriculture, the country's most populous industry. The agricultural sector is increasingly adopting data-driven technologies to enhance productivity, sustainability, and efficiency. Data mining techniques play a crucial role in analyzing vast agricultural datasets to extract meaningful patterns and insights. Nowadays, agricultural businesses can generate massive amounts of data, and it's crucial to use data mining tools to discover useful trends. With the use of data mining tools, we may uncover patterns in large amounts of data, which can then be utilised to aid farmers in crop planning. By providing a comprehensive overview of data models in agricultural data mining, this paper aims to guide researchers and practitioners in selecting appropriate methodologies for efficient and sustainable agricultural data management.

Keywords: Agriculture, Irrigation System, Data Mining, Techniques, Challenges

-------------------------- X --------------------------

INTRODUCTION

Presently, India's agricultural production is second to none in the world. A large part of India's social and economic fabric is based on agriculture, the country's most populous industry. Producing crops in agriculture is a one-of-a-kind industry that is very sensitive to macroeconomic and microclimate variables. Farming relies on a number of variables, including soil, weather, cultivation, irrigation, fertilisers, temperature, precipitation, pesticides, weeds, and harvesting. Companies in the industrial sector rely on accurate historical agricultural yield data for supply chain operations. Livestock, food, chemical, fertiliser, pesticides, seed, paper, and animal feed are all inputs into these sectors from the agricultural sector. These businesses may better plan supply chain decisions, such as production scheduling, with an accurate forecast of agricultural productivity and risk. Forecasts of crop yields inform the marketing and production strategies of companies that deal with seeds, fertilisers, agrochemicals, and agricultural machinery [Veenadhari 2011; Gleaso 1982].

IRRIGATION SYSTEMS

Over the course of a crop's growing period, it needs a specific amount of water at regular times. Irrigation is a useful tool for farmers in terms of crop growth. Artificially watering crops is known as irrigation. Particularly in regions with infrequent or unpredictable precipitation, this method is vital. For a plant to survive, water is essential for many reasons. To grasp the different concepts of water management in different climates, one must be familiar with the soil-water-plant interaction. Soil comes in several forms, including sandy, silty, clay, and so on. The benefits and drawbacks of each soil type are unique. As an example, sandy soil drains quite well. Still, the drain quickly removes the nutrients. In contrast, silty soil can hold onto water for a lot longer since the particles are so minute. Nevertheless, the soil's capacity to drain water is inadequate. Soil quality, including its ability to drain nutrients and water, is crucial for effective farming. As a result, knowing how to regulate the water needs of different types of soil requires an in-depth familiarity with soil characteristics.

AGRICULTURE IN INDIA

The Latin roots of the English term "agriculture" are "ager" and "cultura." The words Ager (land) and Cultura (cultivation) are synonymous. So, farming is the practice of cultivating land. For monetary gain, agriculture also includes the art and science of raising animals and crops. The practice of raising animals and crops using only the earth's natural resources is another name for it (Karuppannan et al., 2017). Agriculture not only employs a huge percentage of India's population but also produces food and raw resources, which are vital to the country's economic structure. As an example, the agricultural industry is strongly linked to India's overseas trade. The fundamental objective of Indian agriculture is to both increase crop yields and prevent the degradation and abuse of the soil. After years of struggling due to food shortages and recurrent droughts, India's agricultural sector has come a long way and is now a major exporter of agricultural commodities. Persistent attempts to utilise land and water resources for agriculture have made this possible.

The agricultural sector is vital to the global economy and provides essential goods and services to people all across the globe, including in India. The scarcity of irrigation water, however, is a severe problem for most agricultural land on Earth. Researchers have devised drip irrigation strategies to decrease water use in arid regions in such a case. There are no protocols for maximising water and energy use. Because of the increased necessity to physically visit and inspect farms on a regular basis with conventional drip irrigation, farmers now have significant overhead expenditures. An further benefit of data mining is the improvement it brings to decision-making, which is very helpful in many agricultural jobs. One of these processes is controlling the amount of water that crops receive by using the rules that are generated by association rule mining. Furthermore, wireless sensor networks are the state-of-the-art technology used in precision agriculture. Modern irrigation systems incorporate smart sensor networks to collect field values for effective plant watering.

CHALLENGES IN AGRICULTURE

The advent and widespread use of IT in farming has made digital data generation a reality. Among the many pieces of data stored digitally are records pertaining to crops, farmers, weather, soil, rainfall, and other crucial variables having a major impact on the agricultural sector [Raorane 2013]. The agricultural business can automate & streamline its processes with the help of an information system. Intelligent data gathering centres, online registration forms, online questionnaires, form submissions, etc. are only a few of the many tools and approaches available for data collection. These have made it possible to gather data through IT [Mr. Navtej Bhatt 2018].

In an effort to streamline the often-tedious and mistake-prone procedures of acquiring knowledge from empirical data, numerous methods for automatically learning about rules & relations from datasets have been created. These methods rely on their preparedness to construct genuine data since they are reliable, theoretically supported, and sufficient for most data [Tag Pang-Ning 2003].

Reducing the subjectivity of decision-making and providing fresh, useful agricultural knowledge are two benefits of integrating data mining into agricultural information systems. Agricultural organisations can benefit greatly from the information and experience provided by predictive models.

The agricultural prediction process consists of two stages: the learning stage & decision making stage. A huge dataset is reduced to a more manageable size during the learning phase. Data knowledge extraction involves four steps: selection, preprocessing, transformation, & evaluation. The procedure for gaining insight from unstructured data is illustrated in Figure 1.

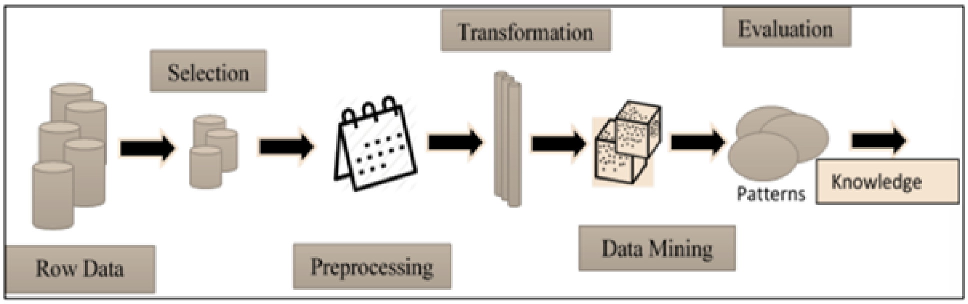

Figure 1: Utilise the raw data to extract information.

Compared to the original dataset, the newly created one has fewer objects and smaller features. At this stage, we can even come up with new guidelines that will help us make better decisions. If a new event occurs with an unknown outcome, the prediction algorithm will be able to use the updated dataset & rules to make a forecast. Algorithms for making predictions compare newly-discovered objects to those already included in the dataset. Assigning the same result to the new item as the matching object in the dataset happens if the two objects are a good fit.

In agriculture, the primary goal of using predictive data mining techniques is to create an easy-to-use decision-support system & predictive model that can reliably and precisely estimate future water demand and help irrigation offices refine their planning procedures. Predicting rainfall and agricultural output under specific weather & soil conditions is another use for this tool.

CURRENT PRACTICE TO ESTIMATE WATER DEMAND FOR IRRIGATION

When precipitation is insufficient to support crop development in semi-arid and dry regions, irrigation becomes crucial for agricultural output [FAO 1994]. Because of the unpredictability of rainfall & frequency of droughts, irrigation has emerged as the principal method in humid regions to boost and stabilise agricultural output, which is crucial for feeding a growing global population. The irrigation offices and farmers currently use a manual water management approach. Preserving rainwater and planning irrigation are key parts of the process.

- **Preservation of rainwater for future irrigation**

The Gujarati government is putting a lot of effort into collecting rainwater for use in irrigation in the future.

It has built large dams and checkdams to collect rainfall for use later on rivers. Dams store water for later use, which opens up new opportunities for irrigation, household uses, or manufacturing. Domestic water usage accounts for an average of 15% to 20% of the water supplied by municipalities. Employ a ratio of 10% to 15% in the industry. According to a survey by the International Water Management Institute (IWMI), evaporation accounts for an average loss of 5 to 10% of water. Irrigation is the use of the leftover water for farming. According to Mr. Navtej Bhatt (2014), the optimal time to water is between the months of January and June.

- **Demanding by farmer for irrigation**

Irrigation water is typically required by farmers ten to fifteen days before to seeding. Farmers' needs are determined by their level of expertise [Mr. Navtej Bhatt 2014]. The irrigation officer is responsible for scheduling the irrigation process once they have received the complete irrigation request.

- **Estimation of Crop Water Requirement**

A crop's water need is the amount of water it needs to grow normally in the field at a given period, independent of the source of that water. Both the plant's metabolic processes and its ET (evapotranspiration) needs, collectively referred to as consumptive usage (Cu), are major users of water. Here is the equation for Cu:

$$Cu = E + T + M$$

Where, Cu = Consumptive use of water E = Evaporation from the crop field T = Transpiration from the crop field M = Water utilized in metabolic activities by plant

The relation between crop production and ET should be better than that between crop production and irrigation. This is because irrigation may not account for all of the plant, since some of the irrigation water may go to drainage or surface runoff. ET, on the other hand, account for all of the water taken up by the plant and returned to the atmosphere as transpiration (T) as well as water lost by evaporation (E) from soil. Many authors have shown a strong relation of crop yield to ET. However, in field situation it is very difficult to separate E and T. Thus, it may be more practical for irrigation management purposes to use the relation of yield to ET rather than T.

Crop Water Requirement (CWR) encompasses consumptive consumption, application losses that cannot be avoided, and specific demands like land preparation, transplanting, leaching, and so on.

CWR = ET + Unavoidable application losses (A) + Special need (B

Water requirement is mainly supplemented by Irrigation (IR), Effective Rainfall (Re), stored residual soil moisture (ΔS) and contribution from shallow Ground water table (Ge). Thus, it can be expressed as shown in Table 2

Table 2 Crop Water Requirement demand and supply

Crop Water Requirement as a demand	Crop Water Requirement as a supply

CWR = ET + A + B	CWR = IR + Re + ΔS + Ge
And hence, IR = CWR – (Re + ΔS + Ge)	

- **Irrigation Scheduling**

It is necessary to prepare field irrigation schedules when the water and irrigation needs of the crop have been determined [Sharma 2012]. When planning an irrigation schedule, three factors must be taken into account [Raorane 2013]:

• The watering needs of crops per day

• The ground, and more specifically, the soil's capacity to retain water or total available moisture

• The optimal depth of the root zone

Irrigation scheduling is the term used for management techniques for allocation irrigation water supply both in time and quantity over an irrigated area. It allows a range of options, depending on the specific application the user is aiming at and the conditions and restrictions the field irrigation system impose. Two distinct types are referred to by the scheduling choice. Criteria for timing, which governs when to apply irrigation, & criteria for application, which dictate the amount of water to be applied per irrigation turn. The person in charge of irrigation prepares the schedule. During the scheduling process, they draw on their own expertise in addition to the different formulae. It usually takes around five to seven days for water to travel from upstream to the farm. Consequently, the irrigation officer is responsible for managing evapotranspiration, soil moisture, & seepage, all of which impact water transportation.

The level of expertise of the irrigation officer is going to play a significant role in the scheduling of irrigation. The anticipated time that water would reach its destination is defined by irrigation scheduling [Raol 2012]. Nevertheless, there are a number of factors, such as variations between anticipated and actual weather conditions and changes in agricultural practices, that can cause the water demands predicted at the start of the irrigation season to diverge from the actual water consumption.

A decision support system called "Coleambally IRIS" was developed by Mahmood A. Khan, Md. Zahidul Islam, and Mohsin Hafeez [2011] to predict the demand for irrigation water. On Spatio-Temporal Data, they had applied the data mining and preprocessing method.

Data from the weather, the water distribution system, & remote sensing photos were the three main sources from which they compiled the information. Their innovative pre-processing of the dataset, drawing on expertise in both irrigation and data mining, made it suitable for use in demand forecasting and pattern extraction. They used the decision tree technique to predict future water needs and created a web-based decision support system for researchers, farmers, and managers. One component of this DSS is the ability to foretell potential future water needs. The effectiveness of their pre-processing and prediction methods is supported by their experimental findings.

Data mining algorithms for rainfall forecasting have been developed by M. Kannan, S. Prabhakaran, and Ramachandran [2010]. Because water is so crucial to farming, it would be great to be able to anticipate when it will rain. This forecast is useful for farmers because there is a lot of unpredictability about the rain these days.

A review of data mining techniques' uses in farming was conducted by S.S. Baskar, L. Arockiam, in 2010. They demonstrated the potential utility of several methods, including Artificial Neural Networks, Decision Trees, Support Vector Machines, K-means, K-nearest, and Bayesian Networks, in the agriculture industry.

Drs. E. Kirubakaran and Sathianarayanan, C. Beulah Christalin Latha, and Sujni Paul [2010] used data mining techniques to build a.NET online service that can predict the weather and climate. An essential use of meteorology is weather forecasting. A huge obstacle to accurate weather forecasting is the fact that weather is a continuous, data-intensive, multifaceted, and dynamic process. They presented a new approach to data mining-based weather forecasting and service-oriented architecture development for weather information systems in their study.

One method to use software as a service is through web services. This software can be accessed from any location and on any device thanks to the development of service-oriented architecture.

Because agricultural data is mostly seasonal and unpredictable, Latika Sharma and Nitu Mehta [2010] try to highlight its distinctive computing requirements. On the topic of data mining techniques for agricultural knowledge management, they also offered some suggestions. Data mining techniques, which have traditionally been employed by corporations and businesses, can now be applied to the agricultural sector for data classification, categorisation, and forecasting and prediction. Soil characteristic assessment has already made an effort to apply data mining. Integrating computer science with agriculture in a multidisciplinary way will aid in forecasting and management.

Data mining approaches have been examined in relation to agriculture by Raorane A.A. and Kulkarni R.V. [2011]. An impartial approach to crop forecasting before harvest is required, they said. This necessitates the development of appropriate forecast model(s), which offer certain advantages over the conventional approach to forecasting.

Vidya Kumbhar, T.P. Singh [2016] says "The research and advancements in agriculture have made available huge amounts of data in different areas of agriculture. It is a great challenge to extract knowledge from data and this has led to methods and techniques such as decision support system that can bridge the knowledge gap. As per their view, in India, simulation-based techniques are widely applied in different areas of agriculture such as to increase crop yield, crop water requirements, on farm irrigation scheduling, and to study the impact of climatic parameters. ICT based advisory systems are also playing an important role in Indian scenario. In India, majority of the rural population lives in rain-fed regions, therefore, the challenge before Indian agriculture is to transform rain-fed farming into more sustainable and productive systems to better support the population dependent on it. The expert systems based on spatial database on agriculture will improve the performance on agriculture management which in turn will be helpful for sustainable agriculture management in India.

Data mining techniques for crop productivity prediction was reviewed by S. Veenadhari, Dr. Bharat Misra, and Dr. CD Singh in 2003. The purpose of this literature review was to compile and analyse existing research on data mining's potential uses in farming. Presented were some of the methods used in agriculture, including ID3 algorithms, k-means, k-nearest neighbour, artificial neural networks, and support vector machines. Data mining in agricultural applications is a novel strategy for agricultural crop and animal management forecasting and prediction, in their opinion. Effective crop forecasting and management can be achieved through the interdisciplinary integration of computer science and agriculture.

To enhance the prediction of irrigation water demand, Inmaculada Pulido-Calvo and Juan Carlos Gutierrez-Estrada [2018] created a soft-computing hybrid model. To estimate the corrections of forecasts derived from a univariate autoregressive neural network, this model utilised a fuzzy inference system. The optimal values of the fuzzy system's parameters were found using a genetic algorithm. Due to the importance of water demand data in scheduling pumping efforts, minimising operation costs of water distribution systems, evaluating the marginal value of irrigation water, & response level to different irrigation water rates, this hybrid model has shown to be a powerful tool that can be used to develop policies on irrigation water consumption with relatively small data requirements.

CONCLUSION

The integration of data mining in the agricultural sector has significantly improved decision-making, resource management, and productivity. A well-structured data model plays a crucial role in ensuring the efficient collection, storage, and analysis of vast and diverse agricultural datasets. Due to the unpredictability of crop yields, agriculture has become a high-risk enterprise for farmers across most of India. Obtaining lucrative prices in the market and the availability of water resources for cultivation are the key causes of risk. Future advancements in data modeling, coupled with innovations in machine learning and IoT, will drive more effective and sustainable agricultural practices. By understanding and implementing suitable data models, stakeholders can unlock the full potential of agricultural data mining, leading to improved efficiency, profitability, and environmental sustainability.

References

1. Mucherino, J. P. Petraq and M. P. Panos, "Data Mining in Agriculture", Springer, U.S.A., Vol. 34, 2009
2. Milovicet. Al. "Application of Data Mining in Agriculture", Bulgarian Journal of Agricultural Science, 21 (No 1) 2015, 26-34
3. Cravero, A., Pardo, S., Galeas, P., López Fenner, J., & Caniupán, M. (2022). Data type and data sources for agricultural big data and machine learning. Sustainability, 14(23), 16131.
4. Cruz, G. B. D., Gerardo, B. D., & Tanguilig III, B. T. (2014). Agricultural crops classification models based on PCA-GA implementation in data mining. International Journal of Modeling and Optimization, 4(5), 375.
5. Desai et.al. (2015): 'A design of a data warehouse and use of data mining techniques for analysis of risk factors affecting agriculture in India' IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS), Vol. 8(10). PP 81-88.

6. GotaMorota et. Al (2020) "Machine learning and data mining advance predictive big data analysis in precision animal agriculture",big data analytics and precision animal agriculture symposium:17 June.

7. Jyotshna Solanki, et. Al. "Different Techniques Used in Data Mining in Agriculture", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 5, May 2015 ISSN: 2277 128X

8. Khan, M., Islam, M. Z., Hafeez, M., "Irrigation Water Demand Forecasting – A Data Pre-Processing and Data Mining Approach based on Spatio-Temporal Data", In Proceedings of 9th Australasian Data Mining Conference (AusDm11). Ballarat, Australia., CRPIT, 121, pp. 183-194, 2011.

9. Masoudi, R., Mousavi, S. R., Rahimabadi, P. D., Panahi, M., & Rahmani, A. (2023). Assessing data mining algorithms to predict the quality of groundwater resources for determining irrigation hazard. Environmental monitoring and assessment, 195(2), 319.

10. Veenadhari S, Misra B, Singh CD. Data mining techniques for predicting crop productivity—A review article. In: IJCST. 2011; 2(1).

11. Yadav, S. A., Sahoo, B. M., Sharma, S., & Das, L. (2020, June). An analysis of data mining techniques to analyze the effect of weather on agriculture. In 2020 international conference on intelligent engineering and management (ICIEM) (pp. 29-32). IEEE.