Journal of Medicinal Plants and By-products (2024) 2: 247-256
https://jmpb.areeo.ac.ir/article_130763.html

Review Article

Data Mining Approach in the Agricultural Industry, Medicinal Plants (case study); A Review

Shahnaz Hatami[1*], Mohammad Hatami-B[2] and Danial Kahrizi[3]

[1] *Department of Industrial Engineering, Faculty of Industrial Engineering, System Management and Productivity, Arak Branch, Islamic Azad University, Arak, Iran*
[2] *Animal Science Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Alborz, Iran*
[3] *Agricultural Biotechnology Department, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran*

Article History

Received: 16 December 2023
Accepted: 01 February 2024

Keywords
Data mining
Medicinal plants
Classification
Clustering

***Corresponding author**
shanii.hatamii@gmail.com

ABSTRACT

In the realm of agriculture and natural resources, medicinal plants stand out as a valuable resource. In recent years, faced with challenges such as predicting climate changes, soil classification, land use, and identifying patterns, there is a growing need for optimal techniques with higher efficiency, particularly in the cultivation of medicinal plants. Therefore, this article introduces the application of data mining to analyze available data in the agriculture and natural resources areas, focusing specifically on the medicinal plant industry. The primary objective is to explore data mining techniques that can enhance various aspects of medicinal plant cultivation, addressing challenges related to climate predictions, soil classification, and optimizing production. The article concludes by presenting the most effective data analysis methods in this domain, accompanied by their corresponding algorithms. Additionally, the aforementioned research is a guide for those intending to investigate the applications of data mining methods are highlighted for increased productivity, encompassing areas such as predicting crop yield, forecasting weather conditions, rainfall patterns, seed and plant conditions, soil quality, and medicinal plant production. The summarization and analysis of the outcome indicated that implementing AI could improve the design and process engineering strategies in bioprocessing fields.

INTRODUCTION

Medicinal plants, renowned as primary resources for the development of economically and medically valuable products, have gained prominence and consistently provide a growing array of advantages to humanity [1]. According to current statistical data, there has been a gradual rise in attention towards medicinal plants in recent years. The utilization rate of medicinal plants in developing countries is notably high at 80%, and simultaneously, their market presence in developed Western countries, such as Europe, is steadily increasing [2- 4]. The domain of medicinal plants, traditionally associated with developing nations, is gradually expanding to include developed Western countries. The heightened focus on healthcare has led to an increasing demand for medicinal plants. Despite the involvement of millions of individuals globally in the cultivation, processing, and wild harvesting of medicinal plants, the available diversity remains insufficient to adequately meet the heightened demand [5]. The World Health Organization estimates that a substantial proportion, ranging from 70% to 95%, of the population in developing nations relies on traditional medicines, primarily derived from plant sources, as a primary healthcare resource [6]. Medicinal plants not only serve as a means of self-treatment but also constitute a significant source of revenue [7]. Furthermore, the utilization of medicinal herbs is increasing, particularly in the aftermath of the COVID-19 pandemic [8, 9], with global trade amounting to USD 138 billion in 2019 [10], characterized by an average annual growth of 2.4% in volume and 9.2% in value [11]. Consequently, there is a compelling need for actionable knowledge to address the potential threat to the sustainability of medicinal plants resulting from their escalating utilization [12].

This study investigates the effective application of machine learning algorithms in the agricultural sector, specifically focusing on data science principles. It introduces and scrutinizes both supervised and unsupervised machine learning methodologies, along with data mining techniques, examining their significance within the context of the medicinal plant industry. The text also deliberates on the utilization of diverse data mining approaches in this sector and offers recommendations for potential avenues of future research and applications.

Subsequent sections delve into the examination of climate changes and their repercussions on the agricultural industry, with a specific emphasis on medicinal plants. The imperative of employing data science to address these challenges is underscored. Section 2 elucidates the escalating significance of applying data science in the agricultural domain. Furthermore, it explores how mastery of machine learning algorithms can enhance this sector. In Section 3, emphasis is placed on machine learning methods (both unsupervised and supervised) as well as data mining techniques. The application of various data mining methods in the medicinal plants industry is detailed in Section 4, with further discussion in Section 5 summarizing suggestions for future research directions and applications.

Research Methodology

As the nature of research in n the application of data mining in medicinal plants and agriculture to confine to specific disciplines, the relevant materials are scattered across scholarly journals. Predominantly, artificial intelligence and knowledge discovery are the most common academic discipline for data mining research in agriculture and medicinal plants. As a result, the following online journal databases were searched to provide a comprehensive information of the academic literature on the application of data mining in medicinal plants and agriculture:

- IEEE Transaction;
- ECS Transactions;
- NRHA/LWT Database;
- IJAERS Journals;
- Agriculture, Environment and Food Journals; and
- WHO Database.

The literature search was based on the descriptor, " medicinal plants" ‘‘agriculture” and ‘‘data mining”, which originally produced approximately 500 articles. The full text of each article was reviewed to eliminate those that were not actually related to application of data mining techniques in medicinal plants. The selection criteria were as follows:

- Articles published in journals specializing in artificial intelligence and knowledge discovery were exclusively chosen, given their relevance as the most fitting platforms for research on data mining in medicinal plants and climate change, aligning with the focus of this review.
- Only articles clearly delineating the application of the mentioned data mining technique(s) in addressing agricultural issues were included in the final selection.
- Conference papers, masters and doctoral dissertations, textbooks and unpublished working papers were excluded, two sites are exceptions, as academics and practitioners alike most often use journals to acquire information and disseminate new findings. Thus, journals represent the highest level of research. [13].

Medicinal Plants

With the evolution of agricultural and natural resource policies globally, the entire agricultural sector has experienced rapid development [14]. In comparison to other industries, the agricultural sector is recognized for exhibiting lower levels of operational efficiency and constrained managerial authority, predominantly attributed to smaller farm sizes [15].

Moreover, these bioresources encounter significant uncertainty attributed to weather and environmental conditions, along with the volatile balance between food supply and demand, influenced by the growing and breeding times of crops and livestock [16]. Technological advancements in natural resources and agriculture can contribute to the broader benefit of society [17], and medicinal plants, as a subcategory within these domains, have demonstrated notable progress to date.

Medicinal plants represent a valuable resource capable of contributing to society's health, employment generation, and non-oil exports when scientifically identified, cultivated, developed, and utilized appropriately [18]. Over millennia, humans have selected medicinal plants and foods by evaluating their suitability and quality based on sensory characteristics. The use of medicinal plants as a healing resource has persisted as a fundamental healthcare practice for approximately 85% of the global population across diverse communities [19].

Preservation efforts must be directed towards both endemic and widespread medicinal plants, along with the associated knowledge about their usage. This is crucial as these plants could serve as renewable sources in the processing of foods and drugs [20].

The utilization of medicinal plants has traditionally been prevalent in developing countries, but it is increasingly becoming common in developed Western countries as well. With a heightened focus on health, there is a growing demand for medicinal plants. Despite millions of individuals globally being engaged in the cultivation, processing, and harvesting of these plants, the supply is still unable to meet the escalating demand [4]. The growth of medicinal plants is influenced by various environmental factors, including soil, climate, and terrain, leading to notable variations in active ingredients within the same medicinal plant [21]. Ensuring consistency in the chemical composition of medicinal plants is paramount for maintaining drug efficacy. Consequently, there is a pressing need to investigate and assess the quality of these plants. However, evaluating the quality of medicinal plants is intricate, involving multiple components that cannot be comprehensively characterized by a single analytical technique. This complexity renders quality evaluation a challenging task [22].

Over the past seventy years, the development of basic analytical techniques has progressed from reliance on sight, touch, and smell to the utilization of more sophisticated instrumentation. Some of the earliest analytical tools, such as the telescope and microscope, were devised to augment our sensory perception. As time has elapsed, our comprehension of the world has advanced from the exploration of new celestial bodies to the investigation of structural, chemical, and atomic levels of matter [23].

The advancement of computer systems and data management tools has facilitated the application of new techniques, allowing for the rapid and selective synthesis of information from the extensive array of instrumental and analytical data signals generated [24]. A pivotal factor that has propelled the analysis of medicinal plants, and analysis in general, in recent times is the capability to collect, assimilate, and utilize substantial volumes of data in a more meaningful and human-readable format. With advancements in modern analytical strategies, the analysis of multiple data sources has proven advantageous for comprehensively evaluating the quality of medicinal plants [25]. The properties of medicinal plants play a crucial role in determining the extent of the advantages derived from utilizing multi-source data to elucidate changes in chemical information. This is attributed to the collaborative nature of multi-source data, which furnishes valuable insights for a more profound understanding and characterization of the quality of medicinal plants [26].

Integrating data from multiple sources through data fusion strategies can enhance reliability, accuracy, and efficiency compared to relying on a single data source [27]. The analysis of medicinal plants is intricate due to the variety of data sources, but data pre-processing and feature extraction techniques can simplify the fusion process [28].

Due to the escalating demand for medicinal plants and their pivotal role in various industries, such as food, pharmaceuticals, and agriculture, researchers and the government are making concerted efforts to explore new techniques and increase plant production in this field. Consequently, automatic data collection technologies are increasingly being employed, leading to a growing volume of data collected in the agriculture and natural resources sectors.

This substantial amount of data, commonly referred to as big data, holds the potential to revolutionize decision-making in this field through the application of information technology and improve overall efficiency. Given the expanding global population, there exists an urgent need to enhance food production with limited land resources [14].

To address the challenges posed by increasing food demand and climate change, policymakers and industry leaders are seeking support from technological advancements such as IoT (Internet of Things), big data, analytics, and cloud computing [29].

Data Science

It is widely acknowledged that the advent of Big Data [30] has transformed the landscape of data generation, storage, and processing, to the extent that approximately 90% of the world's existing data has been generated in recent years [31]. Managing such extensive datasets can present challenges in terms of comprehension and analysis, necessitating the development of techniques to process this information. In response to this demand, novel tools have emerged under the umbrella term of Data Science [32].

Throughout the decades, the escalating requirements for provisioning agri-food products have exerted a considerable influence on global agricultural practices. Furthermore, shifts in human lifestyles, coupled with the continuous growth of the human population and urbanization, have directly shaped the dynamics of agri-food product production and consumption [33]. The economic significance of key plants and the limited availability of natural resources for agriculture have prompted plant producers and agricultural researchers to explore innovative approaches aimed at addressing the challenges associated with food scarcity.

Big data and data mining are becoming increasingly important in modern agriculture, offering numerous benefits that help optimize productivity and efficiency such as Improved prediction of crop yield and production, Expedited delivery, Real-time decision-making, Predictive maintenance, Energy efficiency, Data-driven decision-making, Optimized farming practices, Demand prediction, Innovative pricing programs, Food waste reduction, Cost savings, and business opportunities, Supply chain management. [34].

Furthermore, in some cases, big data can also create marvelous works in the form of the following:

1. **Ensuring food supply for a rapidly increasing population**

Big data offers farmers highly detailed information on various factors like rainfall patterns, water cycles, fertilizer requirements, and more. This information empowers farmers to make informed decisions such as which crops to plant for better profitability and when to harvest them. By making these smart decisions, farmers can optimize their farm yields and ensure better outcomes.

2. **Ethical use of pesticides**

The use of big data can aid farmers in managing this issue by providing recommendations on which pesticides to use, when to use them, and how much to apply. By closely monitoring their pesticide usage, farmers can comply with government regulations and avoid overusing chemicals in their food production process. Additionally, this results in increased profitability as crops are protected from destruction caused by weeds and insects.

3. **Improving the efficiency of farm machinery**

Some companies have incorporated sensors into their farming equipment and implemented big data applications that can assist in managing their fleet more effectively. For large farms, this level of monitoring can be crucial as it enables users to keep track of tractor availability, service due dates, and fuel refill alerts. This ultimately optimizes usage and ensures the longevity of farm equipment.

4. **Dealing with supply chain problems**

Tracking and optimizing delivery truck routes with big data can improve supply chain efficiency [35].

In the field of agriculture, farmers and agribusinesses face numerous decisions every day that involve intricate complexities and various factors that can influence them. Accurate yield estimation is vital for agricultural planning, as it involves numerous crops. To achieve practical and effective solutions for this problem, data mining techniques are necessary. Agriculture is an obvious target for big data, as environmental conditions, variability in soil, input levels, combinations, and commodity prices make it more relevant for farmers to use information and get help to make critical farming decisions [36].

In this regard, data mining technology [37] is used to manage, process, and predict big data in many fields of natural sciences [38]. Data mining is a well-established concept worldwide and many developed countries are already utilizing it in their systems and services [39]. It is currently one of the most rapidly advancing fields of study and involves extracting useful and meaningful information from large datasets [37]. It is the process of extracting valuable information from a large dataset and presenting it in a way that is easy for humans to comprehend. The concept of data mining dates back to the 1960s and it is built on the principles of machine learning, artificial intelligence, statistics, and probability [40].

Data mining refers to the process of discovering knowledge from data. It is typically divided into two categories. The first category is descriptive data mining, which involves describing the characteristics of data in the database. The second category is predictive data mining, which involves modeling available data and making predictions about future and experimental data [41], Currently, data mining prediction methods are widely used in the field of agriculture and natural resources, specifically for medicinal plants. These methods are utilized to solve yield estimation of crops [42], weather conditions, product income, and other relevant factors in the studied area [43]. The data mining process involves three essential steps: data extraction from various sources, data transfer to the database and multiple layers, and data analysis.

Knowing the variables associated with medicinal plants, such as planting, harvesting, harvest productivity, irrigation, rainfall, weather, soil, and extraction methods, plays a crucial role in the industry. Proper management of these variables and staying informed about new cultivation methods and devices are necessary for better performance and outcomes. To achieve maximum production and productivity, resources should be utilized efficiently [44].

In agriculture data mining starts with a hypothesis, and the results are then adjusted to fit that hypothesis. While standard data mining pays attention to patterns and trends in datasets, agricultural data mining focuses on the majority that stands out from patterns and trends [45]. Processing collected data automatically to support the decision-making process is a vital strategic issue. In the case of medicinal plants, data mining enables the discovery of patterns and practical methods to predict system behavior and desired relationships [46].

This study reviews data mining techniques and methods in natural systems, including medicinal plants. Through data mining, medicinal plant enterprises can provide predictive and descriptive information to support decision-making in crop, seed variety, fertilizer, pesticides, etc.

Several data mining processes have been proposed by researchers and practitioners, including KDD and CRISP-DM, which CRISP-DM is a reliable data mining model consisting of six phases [47].

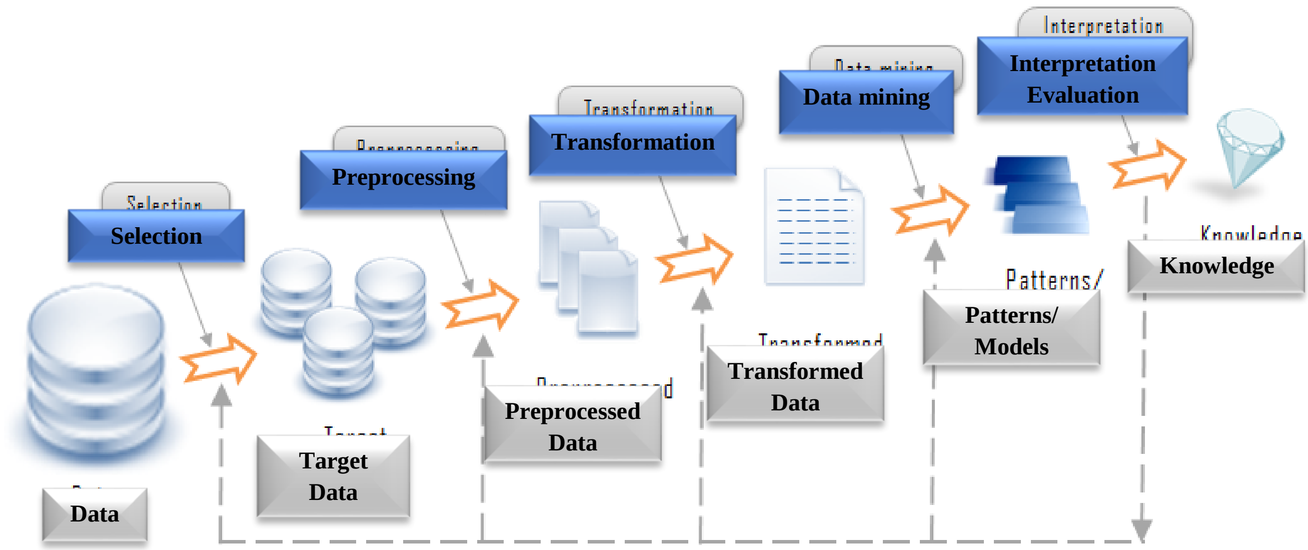

Fig. 1 Data Mining Process: Models, Process Steps & Challenges Involved

Machine Learning

With the rapid development of artificial intelligence, machine learning can be utilized to collect information on various medicinal plants and establish an ideal database. Moreover, leveraging machine learning to effectively extract available information from multi-source data can significantly improve the robustness and accuracy of the obtained results [48].

The swift advancement of artificial intelligence has significantly contributed to transforming the landscape of data processing and analysis. Within the realm of artificial intelligence, machine learning, functioning as a subset, depends on computer algorithms and models to acquire proficiency in intricate functions. [49]. Over the course of its development, it has transformed into a potent instrument for the statistical modeling and mining of data [50]. Hence, the necessity for strategies to avert partial data loss has emphasized the significance of utilizing machine learning algorithms in conjunction with multi-source data related to medicinal plants. Machine learning is acknowledged as a promising mechanism capable of improving the comprehensive efficiency of data fusion, owing to its exceptional computational and analytical capacities [51, 52].

Data Mining Techniques

Data science is an interdisciplinary field that aids in comprehending and analyzing large amounts of data to generate useful information. Various data mining activities may be applied to big biological data sets, such as classification, clustering, association rules, and regression mining, prediction [53]. Data mining in agriculture is a relatively novel research field. It is our opinion that efficient techniques can be developed and tailored for solving complex agricultural problems using data mining [54].

1. Classification

Classification is a technique used for discovering classes of unknown data. This technique involves using a set of reliable models to make predictions about new data, based on input ratios. In classification, the objective is to predict the desired value based on the given input variables. Various classification methods exist like Bayesian, decision trees, rule-based, neural networks, etc [55]. Algorithms used for classification can help analyze the relationship between features. A good algorithm is one that accurately predicts outcomes. In the cultivation of medicinal plants, it is possible to define different uses for the plants in the future pharmaceutical industry by identifying the different characteristics of available plant types and the percentage of their chemical compounds. This data can be classified into separate groups using data mining software.

2. Clustering

Clustering data mining is the process of putting together meaning-full or use-full similar objects into one group. It is a common technique for statistical data, machine learning, and computer science analysis. Clustering is a kind of unsupervised data mining technique that describes general working behavior, and pattern extraction and extracts useful information from electricity price time series [56], Clustering is very similar to classification, but it involves grouping chunks of data based on their similarities [57].

Algorithms used in cluster analysis have the ability to identify the various elements and chemicals present in herbal medicines. Clustering is a technique that is commonly employed to discuss water quality in agriculture and to improve the production of high-yield medicinal plants. Additionally, clustering and classification techniques are also useful to study land use in agriculture and medicinal plant growth [58].

3. Regression

Regression analysis, as a data mining process, is employed to identify and analyze relationships between variables, accounting for the influence of other factors. It serves to determine the probability of a specific variable, primarily functioning as a method for planning and modeling. For instance, regression analysis can be utilized to project the costs of medicinal plants, considering factors like availability, consumer demand, and competition. Its principal utility lies in providing an exact relationship between two or more variables within a given dataset [59].

Regression analysis, a statistical method, is utilized to predict quantities related to agricultural products, natural resources, and consumer demand for new products. This analysis can be conducted using both linear and non-linear regression prediction methods. For example, regression analysis facilitates the prediction of the market value of medicinal plants based on various factors such as production volume, regional location, and prices [60].

4. Association Rules

Association rule mining identifies significant associations and relationships within extensive datasets. This technique employs if-then statements, referred to as association rules, to illustrate the likelihood of interactions between data items within various types of databases. Association rule mining is frequently utilized to discern correlations in sales data or medical datasets and possesses diverse applications [61]. An exemplar application of association rule mining involves the examination of market portfolios for medicinal plants, customer segmentation, and the formulation of pertinent catalogs for advertising purposes.

The Application of Data Mining Methods in the Medicinal Plants Industry

Considering the importance of data in the agricultural industry and natural resources, especially medicinal plants in table 1, various data mining methods introduce the application of different algorithms in the medicinal plants industry.

Table 1 The Application of Data Mining Methods in the Medicinal Plants Industry

Technique	Method of analysis	Algorithms	Application
Classification	Prediction	Bayes networks Decision tree K-Nearest Neighbors Artificial neural network Support vector machine Hard sets Fuzzy Logic Genetic algorithms	Forecasting of climate change Prediction of soil fertility Remote sensing Analysis of correlations Soil classification Classification of crops land use Prediction of flowering and maturity in plants rainfall prediction prediction forecasting heat waves runoff simulation
Clustering	Descriptive	Hierarchy Division methods Density based methods Model-based clustering methods Network based methods Software computing (fuzzy, based on neural network) Clustering diagram Network data	Precipitation simulation Identify patterns Image analysis Weather analysis land use Estimation of soil parameters
Association rules	Prediction	Previous algorithm Classification Dynamic pruning Counting the set of dynamic factors Growth	Efficient storage management Prevention of disorders Index structures
Regression	Descriptive	Linear regression Non-linear regression Logical regression	Economic structures air pollution Medicinal plants market forecast Optimization

DISCUSSION

The field of agriculture and natural resources is the most crucial area for applying data in developed countries. Data mining is expected to play a significant role in Smart Agriculture, facilitating real-time data analysis with massive datasets [62]. The medicinal plant industry, vital for food and medicine, has been growing significantly and requires special attention in terms of data collection and analysis. The concept of big data in medicinal plants is not only related to the volume of data but also to the variety and speed of collected data. Big data is a fundamental concept for the future development of agriculture and natural resources, particularly in the medicinal plant industry, providing unprecedented capabilities through various tools and services to transform the current situation. Data mining, referring to the analysis of big data, enhances productivity by equipping existing systems, improving the management of seeds, poisons, and fertilizers, and predicting product yields through mathematical models and machine learning.

With the rapid development of artificial intelligence, machine learning can be employed to gather information on various medicinal plants and establish an ideal database. Furthermore, leveraging machine learning to effectively extract information from multiple sources can significantly enhance the robustness and accuracy of the results obtained [48]. Data mining is a crucial element in all databases used for data selection. Given the challenges faced by the medicinal plant industry in recent years, it is essential to explore various data mining techniques for decision-making. While big data and data mining in the medicinal plant industry are still in their early stages, this study aims to underscore the significance of data mining in agriculture and natural resources. The ultimate objective is to promote the use of data mining techniques in the field of medicinal plants. Furthermore, each of these techniques could be applied to a specific sector related to agriculture and natural resources, particularly medicinal plants, by closely examining them.

REFERENCES

1. Chen S.L., Yu H., Luo H.M., Wu Q., Li C.F. A Steinmetz. Conservation and sustainable use of medicinal plants: Problems, progress, and prospects. Chin. Med. 2016; 11.
2. He J., Yang B., M. Dong, et al. Crossing the roof of the world: Trade in medicinal plants from Nepal to China. J. Ethnopharmacol., 224. 2018;100-110.
3. Shen T., Yu H., Wang Y. Assessing the impacts of climate change and habitat suitability on the distribution and quality of medicinal plant using multiple information integration: Take Gentiana rigescens as an example. Ecol. Indic. 2021.; 123.
4. Hamilton A.C. Medicinal plants, conservation and livelihoods. Biodiversity & Conservation. 2004; 13(8), 1477–1517. https://doi.org/10.1023/B:BIOC.0000021333.23413.42
5. Molly Meri Robinson Robinson, Xiaorui Zhang. The World Medicines Situation 2011 Traditional Medicines: Global Situation, Issues and Challenges. World Health Organization, Geneva. 1-2. 2011.
6. Jean-Marc Fromentin, Marla R. Emery, John Donaldson, Marie-Claire Danner, Agnès Hallosserie, Daniel Kieling. Thematic assessment of the sustainable use of wild species of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat, Bonn, Germany. 2022. https://doi.org/10.5281/zenodo.6448567
7. Smith T., Majid F., Eckl V., Morton Reynolds C. Herbal supplement sales in US increase by record-breaking 17.3% in 2020. *Herbal Gram*. 2021; 131: 52–65.
8. Timoshyna A., Ke Z., Yang Y., Ling X., Leaman D. The invisible trade: Wild plants and you in the time of COVID-19. Traffic International. 2020.
9. WHO. WHO global report on traditional and complementary medicine 2019. World Health Organization, Geneva, Switzerland. WHO. 2019.
10. Industry Research Biz. Global herbal medicine market size, manufacturers, supply chain, sales channel and clients. IRB. 2020–2028.
11. Vasisht K., Sharma N., Karan M. Current perspective in the international trade of medicinal plants material: An update. Current Pharmaceutical Design. 2016; 22(27): 4288–4336. https://doi.org/10.2174/1381612822666160607070736
12. Kumar A., Kumar S., Komal R.N., Singh P. Role of traditional ethnobotanical knowledge and indigenous communities in achieving sustainable development goals. Sustainability. 2021; 13(6): 3062. https://doi.org/10.3390/su13063062
13. Nord J.H., Nord G.D. MIS research: Journal status and analysis. Information & Management. 1995; 29: 29–42.
14. Osinga S., Paudel D., Mouzakitis S., Athanasiadis I. Big data in agriculture: Between opportunity and solution. Agricultural Systems. 2022; 195: 103298. https://doi.org/10.1016/j.agsy.2021.103298
15. Rosenblatt P., Lasley P. Perspective on Farm Accident Statistics. The J. Rural Health. 2008; 7: 51–62. https://doi.org/10.1111/j.1748-0361.1991.tb00703.x
16. Kamble S., Gunasekaran A., Gawankar S. Achieving Sustainable Performance in a Data-driven Agriculture Supply Chain: A Review for Research and Applications. International Journal of Production Economics. 2019; 219: 179–194. https://doi.org/10.1016/j.ijpe.2019.05.022
17. Chavas J.-P., Chambers R., Pope R. Production Economics and Farm Management: A Century of Contributions. American Journal of Agricultural Economics. 2010; 92: 356–375. https://doi.org/10.1093/ajae/aaq004
18. Khate A., Sharma B. Medicinal Plant Classification Using Neural Network. 2023; 297–307. https://doi.org/10.1007/978-981-99-4362-3_28
19. Pešić M. Development of natural product drugs in a sustainable manner. Brief for United Nations Global Sustainable Development Report 2015. Available at: https://sustainabledevelopment.un.org/content/documents/6544118_Pesic_Development%20of%20natural%20product%20drugs%20in%20a%20%20sustainable%20manner.pdf. (Accessed August 15, 2018).
20. Cordell G.A. Sustainable medicines and global health care. Planta Med. 2011; 77(11):1129-38. doi: 10.1055/s-0030-1270731. Epub 2011 Feb 9. PMID: 21308611.
21. Applequist W., Brinckmann J.A., Cunningham A., Hart R., Heinrich M., Katerere D., Andel T. Scientists' Warning on Climate Change and Medicinal Plants. Planta Medica. 2019; 86. https://doi.org/10.1055/a-1041-3406
22. He M., Wu C., Li L., Zheng L., Tian T., Jiang L., Li Y., Teng F. Effects of Cavitation Jet Treatment on the Structure and Emulsification Properties of Oxidized Soy Protein Isolate. In Foods. 2021; 10(1). https://doi.org/10.3390/foods10010002
23. Fitzgerald M., Heinrich M., Booker A. Medicinal Plant Analysis: A Historical and Regional Discussion of Emergent Complex Techniques. Frontiers in Pharmacology. 2020; 10: 1480. https://doi.org/10.3389/fphar.2019.01480
24. Wang P., Yu Z. Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review. Journal of Pharmaceutical Analysis. 2015; 46. https://doi.org/10.1016/j.jpha.2015.04.001
25. Sanaeifar A., Li X., He Y., Huang Z., Zhan Z. A data fusion approach on confocal Raman microspectroscopy and electronic nose for quantitative evaluation of pesticide residue in tea. Biosystems Engineering. 2021; 210: 206–222. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2021.08.016

26. Azcarate S.M., Ríos-Reina R., Amigo J.M., Goicoechea H.C. Data handling in data fusion: Methodologies and applications. TrAC Trends in Analytical Chemistry. 2021; 143: 116355. https://doi.org/https://doi.org/10.1016/j.trac.2021.116355
27. Zhang P., Li T., Yuan Z., Luo C., Wang G., Liu J., Du S. A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data. Information Fusion. 2022;80:87–103. https://doi.org/https://doi.org/10.1016/j.inffus.2021.10.017
28. Zhou X., Li X., Zhao B., Chen X., Zhang Q. Discriminant analysis of vegetable oils by thermogravimetric-gas chromatography/mass spectrometry combined with data fusion and chemometrics without sample pretreatment. LWT. 2022; 161: 113403. https://doi.org/https://doi.org/10.1016/j.lwt.2022.113403
29. Hariharan U., Kotteswaran R., Pathak N. The Convergence of IoT with Big Data and Cloud Computin. 2020: 1–23. https://doi.org/10.1201/9781003054115-1
30. Khan N., Yaqoob I., Hashem I.A.T., Inayat Z., Mahmoud Ali W.K., Alam M., Shiraz M., Gani A. Big Data: Survey, Technologies, Opportunities, and Challenges. Scientific World J. 2014.

A. Galicia J.F., Torres F., Martínez-Álvarez Troncoso A. A novel spark-based multi-step forecasting algorithm for big data time series. Inf. Sci. 2018; 467: 800-818.

31. Aggarwal C.C. Data Mining: The Textbook. Springer Publishing Company, Incorporated. 2015.
32. Veronique Bellon-Maurel, Ludovic Brossard, Frédérick Garcia, Nathalie Mitton, Termier Alexandre. Getting the Most Out of Digital Technology to Contribute to the Transition to. Sustainable Agriculture and Food Systems. Université de Rennes. 2022.
33. https://www.n-ix.com/big-data-in-agriculture/
34. (talend.com/resources/big-data-agriculture).
35. Majumdar J., Naraseeyappa S., Ankalaki S. Analysis of agriculture data using data mining techniques: application of big data. Journal of Big Data. 2017; 4(1): 20. https://doi.org/10.1186/s40537-017-0077-4
36. Liao S.-H., Chu P.H., Hsiao pei-yuan. Review: Data mining techniques and applications - A decade review from 2000 to 2011. Expert Systems with Applications: An International J. 2012; 39: 11303–11311. https://doi.org/10.1016/j.eswa.2012.02.063
37. Eskandari S., Ravanbakhsh H., Ahangaran Y., Rezapour Z., Pourghasemi H. Effect of climate change on fire regimes in natural resources of northern Iran: investigation of spatiotemporal relationships using regression and data mining models. Natural Hazards. 2023; 119: 1–25. https://doi.org/10.1007/s11069-023-06133-4
38. Olson D. L. Data Mining BT - Encyclopedia of Optimization (C. A. Floudas & P. M. Pardalos (eds.). 2009; 600–607. Springer US. https://doi.org/10.1007/978-0-387-74759-0_108
39. Attewell P., Monaghan D.B., Kwong D. Data Mining for the Social Sciences (1st ed.). University of California Press. 2015. http://www.jstor.org/stable/10.1525/j.ctt13x1gcg
40. Hassani H., Saporta G., Silva E. Data Mining and Official Statistics: The Past, the Present and the Future. Big Data. 2014; 2: 34–43. https://doi.org/10.1089/big.2013.0038
41. Mulik S., More A. Future of Agriculture with Data Mining. 2023; 47: 32–39.
42. Aishwarya K., Jabbar M.A. Data Mining Analysis for Precision Agriculture: A Comprehensive Survey. ECS Transactions. 2022; 107: 17769–17781. https://doi.org/10.1149/10701.17769ecst
43. Bhagawati K., Sen A., Shukla K.K., Bhagawati R. Application and Scope of Data Mining in Agriculture. International Journal of Advanced Engineering Research and Science. 2016; 3(7).
44. Milovic B., Radojević V. Application of data mining in agriculture. 2015; 21: 26–34.
45. Coleman S. A Practical Guide to Data Mining for Business and Industry. 2014. https://doi.org/10.1002/9781118763704
46. Plotnikova V., Dumas M., Fredrik P. Milani. Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. Data & Knowledge Engineering. 2022; 139; 102013.
47. Zhang Y., Wang Y. Recent trends of machine learning applied to multi-source data of medicinal plants. Journal of Pharmaceutical Analysis. 2023. https://doi.org/https://doi.org/10.1016/j.jpha.2023.07.012
48. Taoufik N., Boumya W., Achak M., et al. The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. Sci. Total Environ. 2022; 807. https://doi.org/10.1016/j.scitotenv.2021.150554.
49. D.V. Nazarenko, P.V. Kharyuk, I.V. Oseledets, I.A. Rodin, O.A. Shpigun. Machine learning for LC–MS medicinal plants identification. Chemom. Intell. Lab. Syst., 156 (2016), pp. 174-180. https://doi.org/10.1016/j.chemolab.2016.06.003
50. Meng T., Jing X., Yan Zh., Pedrycz W. A survey on machine learning for data fusion. Inf. Fusion. 2020; 57: 115-129.
51. Magnus M., Virte H. Thienpont, L. Smeesters. Combining optical spectroscopy and machine learning to improve food classification. Food Contr. 2021; 130.
52. Krishnamoorthy M., Karthikeyan R. Competence of medicinal plant database using data mining algorithms for large biological databases. Measurement: Sensors. 2022;24:100420.

https://doi.org/https://doi.org/10.1016/j.measen.2022.100420

53. Mucherino A., Papajorgji P., Pardalos P. Data Mining in Agriculture. 2009; 34. https://doi.org/10.1007/978-0-387-88615-2
54. Beniwal S., Arora J. Classification and feature selection techniques in data mining. Inter J Engin. Res. and Technol. 2012;1.
55. Patel D., Modi R., Sarvakar K. A Comparative Study of Clustering Data Mining: Techniques and Res Challenges. 2014. https://doi.org/10.13140/2.1.1508.7042
56. Xu R., Wunsch D. Survey of Clustering Algorithms. Neural Networks, IEEE Transactions On. 2005; 16: 645–678. https://doi.org/10.1109/TNN.2005.845141
57. Megala S., Hemalatha D. A Novel Datamining Approach to Determine the Vanished Agricultural Land in Tamilnadu. International Journal of Computer Applications. 2011; 23. https://doi.org/10.5120/2869-3718
58. Rainsford C., Roddick J. Database Issues in Knowledge Discovery and Data Mining. Australasian J. of Inf. Systems. 1999; 6. https://doi.org/10.3127/ajis.v6i2.310
59. Yavuz E., Şahin M. Semiparametric Regression Models and Applicability in Agriculture. 2022; 5; 160–166. https://doi.org/10.47115/bsagriculture.1077101
60. Chen M.-S., Han J., Yu P. Data mining: An overview from a database perspective. Knowledge and Data Engineering, IEEE Transactions On. 1997; 8: 866–883. https://doi.org/10.1109/69.553155
61. Ait Issad H., Aoudjit R., Rodrigues J.J.P.C. A comprehensive review of Data Mining techniques in smart agriculture. Engineering in Agriculture, Environment and Food. 2019; 12(4): 511–525. https://doi.org/https://doi.org/10.1016/j.eaef.2019.11.003.