# Survey of Data Mining Techniques in Precision Agriculture
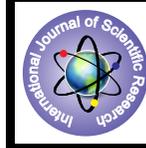
| **Tejas S. Mehta** | Ph.D. Student, Department of Computer Science, Dr. Babasaheb Ambedkar Open University, Gujarat, India |
|---|---|
| **Dr. Dhaval R. Kathiriya** | Director - Information Technology, Anand Agricultural University, Gujarat, India |

**ABSTRACT**    *Agriculture sector plays crucial role in our economy. More than 70 percentage of population depends on agriculture or agriculture practice. When farmers harvest their yield, they also produce agricultural raw data. This data can be further used in data mining techniques for better crop management. There are various data mining techniques have been developed and successfully utilized in agriculture data mining tasks. The goal of this survey is to provide a comprehensive review of different types of data mining techniques used in Precision Agriculture.*

## Introduction

Precision Agriculture, Information Technology and Data Mining are closely inter-connected for better crop management. Ruß, Kruse, Schneider, & Wagner (2008) stats that, it is possible to collect huge data from agriculture practice due to sensors and GPS (Global Positioning System), and can be used for data mining analysis. Shibusawa & Hache emphasizes that the GPS enables collection of geo-referenced data, while the GIS (Geographic information system) allow spatial analysis and visualization of interpolated maps. Application of GPS/GIS into agriculture has caused a revolution called precision agriculture (PA), where fields are managed at a detailed scale based on information and knowledge. McBratney, Whelan, & Ancev (2005) wrote in their research paper that besides decision-support question, there are six other issues: Appropriate criteria for economic assessment of PA; Insufficient recognition of temporal variation; Lack of whole-farm focus; Crop quality assessment methods; Product tracking and traceability; and Environmental auditing require urgent and ongoing attention by researchers to develop the PA concept to its full potential.

## Data Mining Techniques

Data mining is the process of using large data sets to infer important hidden knowledge. Jiawei, Micheline, & Jian (2012) show that data mining process is divided in seven steps: 1) Data cleaning; 2) Data integration; 3) Data selection; 4) Data transformation; 5) Data mining; 6) Pattern evaluation and 7) Knowledge presentation. Data mining is used for any kind of data including database data, data warehouse data or transactional data. There are numbers of techniques used for data mining:

**Association rule learning:** Association (or relation) is probably the better known and most familiar and straightforward data mining technique. Martin (2012) highlighted that simple correlation can be made to identify patterns between two or more items, often of the same type. Agrawal, Imieliński, & Swami (1993) introduced association rules for discovering regularities between products in large-scale transaction data recorded by (POS) systems in supermarket.

**Classification**: Classification algorithm that implements classification is known as a mathematical function or model or classifier, used to maps input data to a category. The derived model may be represented in various forms, such as classification rules, decision trees, mathematical formulas or neural networks (Jiawei, Micheline, & Jian, 2012). Phyu (2009) surveys classification techniques in Data Mining and conclude that classification methods are typically strong in modeling interactions.

**Clustering**: Clustering analyzes data objects without consulting class labels. The objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity (Jiawei, Micheline, & Jian, 2012). The broadly used clustering algorithms are Hierarchical, Partitioning, Density based, Grid based and Graph Based algorithm.

**Anomaly detection:** Anomaly detection is performed to see which record sticks out from the rest. Chandola, Banerjee, & Vipin (2009) have discussed different ways in which the problem of anomaly detection has been formulated in the literature, and have attempted to provide an overview of the huge literature on various techniques. They also classified anomalies into following three categories: 1) Point Anomalies; 2) Contextual Anomalies; and 3) Collective Anomalies.

**Regression:** Regression analysis models the relationship between one or more response variables and the predictors. Multivariate regression describes models that have more than one response variable. There are different families of regression functions and different ways of measuring the error. Support Vector Machines (SVM) is a powerful, state-of-the-art algorithm for linear and nonlinear regression. The Root Mean Squared Error and the Mean Absolute Error are statistics for evaluating the overall quality of a regression model.

## Data Mining in Precision Agriculture

Precision agriculture is modern farming practices. Precision agriculture helps farmers to produce more food and conserve soil for sustainable food production, results in a stable food supply (What is "precision agriculture" and why is it important?, 2015). Precision agriculture research has ability to define a Decision Support System (DSS) with the goal of optimizing returns on inputs while preserving resources. Precision agriculture can also be described as a data-driven approach to agriculture, which is strongly connected with a number of data mining algorithms (Ruß & Brenning, Data Mining in Precision Agriculture: Management of Spatial Information, 2010).

## Data Mining Applications in Precision Agriculture:

Data mining techniques applied for Precision Agriculture include Neural Networks, Fuzzy logic, Hierarchical Clustering, Density-based Clustering and Association rule Support Vector Machine (SVM). Researchers design and develop applications for solving complex agricultural problems using data mining like:

## Yield Data Prediction:

Ruß, Kruse, Schneider, & Wagner(2008) conducted experi-

ments using Matlab 2007b with different farm data sets. Their study contributes to finding and evaluating models of agricultural yield data using Neural Networks.

## Soil Mapping and Classification:

Soil mapping is the geo-encoding based process for classifying soil types and other soil properties in a given area. Moran & Bui (2002) concluded using spatial data mining techniques, that there is sufficient predictive capacity in the environmental correlation attributes representing geology, terrain, and soil / water / vegetation interactions to model a known soil map. Zhu, Hudson, Burt, Lubich, & Simonson (2001) had written about several issues affect the reliability and usefulness of traditional soil survey process and its products. They used fuzzy logic for their research. Armstrong, Diepeveen, & Maddern (2007) designed five study-cases to test the concept and methodology of data mining and to establish the accuracy of the Expectation-Maximization (EM) algorithm and Farthest-First clustering algorithms for agricultural soil profiles.

## Fertilizer and Pesticides management:

Better crop yield is highly affected on inputs of fertilizer and herbicides. Fertilizer not only increases crop yields, but also builds soil organic matter. Vijaykumar & Rajinikanth(2013) demonstrated that how Regression analysis and cluster analysis used for estimation of the influence of fertilizer nutrients consumption on the wheat crop yield. Researchers and Data scientist collect pest surveillance data and related agricultural operations regarding crops details, farming techniques, weather details. These data contain details of pest incidence, soil, climatic, agricultural practices and serve as repositories of information. Tripathy, Adinarayana, & Sudharsan (2009) attempted to express how data mining techniques like Hierarchical Clustering, Density-based Clustering and Association rule mining integrated with agriculture including pest scouting, pesticide and climatological parameters were useful for optimization of pesticide usage and better management.

## Grading and Sorting of agriculture product:

The overall appearance of fruit object is a combination of its color and its geometric attributes like shape, size and texture. The presence of defects can diminish the external quality. Automated fruit gradation plays an important role to increase the value of produces. Gill, Sandhu, & Singh( 2014) studied various data mining techniques like Neural Networks, Fuzzy Logic, Evolutionary Algorithms, Bayesian Network and Support Vector Machine for grading agriculture products like apple, mango, strawberry, cherries, orange and etc.

## Conclusion and Future Directions

There is huge database of agriculture is available for data mining techniques. There are a lot of opportunities for research on data mining in agriculture to develop practical applications for the enhancing of agriculture production. If more complex techniques implemented with high-performing computational systems for data mining, will surely solves difficult problems in agriculture sector.

## REFERENCE

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD international conference on Management of data (pp. 207-216 ). New York: ACM. | Armstrong, L., Diepeveen, D., & Maddern, R. (2007). The application of data mining techniques to characterize agricultural soil profiles. Sixth Australasian Data Mining Conference (AusDM 2007) (pp. 85-100). Australia: ACS. | Chandola, V., Banerjee, A., & Vipin, K. (2009). Anomaly detection: A survey. ACM Computing Surveys , 41 (3). | Gill, J., Sandhu, D. S., & Singh, D. (2014). A Review of Automatic Fruit Classification using Soft Computing Techniques. International Conference on Computer, Systems and Electronics Engineering (ICSCEE'2014), (pp. 91-98). Johannesburg (South Africa). | Jiawei, H., Micheline, K., & Jian, P. (2012). Data Mining : Concepts and Techniques (3rd ed.). USA: Morgan Kaufmann Publishers. | Martin, B. (2012). Data mining techniques. developer Works(IBM). | McBratney, A., Whelan, B., & Ancev, T. (2005). Future Directions of Precision Agriculture. Precision Agriculture , 4, 7-23. | Moran, C. J., & Bui, E. N. (2002). Spatial Data Mining for Enhanced Soil Map Modeling. International Journal on Geographical Information Science , 16 (6), 533-549. | Phyu, T. N. (2009). Survey of Classification Techniques in Data Mining. International MultiConference of Engineers and Computer Scientists. I. Hongkong: IMECS. | Ruß, G., & Brenning, A. (2010). Data Mining in Precision Agriculture: Management of Spatial Information. Computational Intelligence for Knowledge-Based Systems Design, 13th International Conference on Information Processing and Management of Uncertainty (pp. 350-359). Dortmund, Germany: IPMU. | Ruß, G., Kruse, R., Schneider, M., & Wagner, P. (2008). Data Mining with Neural Networks for Wheat Yield Prediction. 8th International Conference on Data Mining 2008 , 47-56. | Shibusawa, S., & Hache, C. (n.d.). Data Collection and analysis methods for data from field experiments. Systems Analysis and Modeling in Food and Agriculture . | Tripathy, A. K., Adinarayana, J., & Sudharsan, D. (2009). Geospatial Data Mining for Agriculture Pest Management - a Framework. 17th International Conference on Geoinformatics , 1 - 6. | Vijaykumar, A., & Rajinikanth, T. (December 2013). Estimation of the Influence of Fertilizer Nutrients Consumption on the Wheat Crop yield in India- a Data mining Approach. International Journal of Engineering and Advanced Technology (IJEAT) , 3 (2), 316-320. | What is "precision agriculture" and why is it important? (2015, February 27). Retrieved from https://soilsmatter.wordpress.com: https://soilsmatter.wordpress.com/2015/02/27/what-is-precision-agriculture-and-why-is-it-important/ | Zhu, A., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. Soil Science Society of America Journal , 65, 1463-1472. | |