

## Chapter 17

# Data Mining Techniques in Agricultural and Environmental Sciences

**Altannar Chinchuluun**  
*University of Florida, USA*

**Petros Xanthopoulos**  
*University of Florida, USA*

**Vera Tomaino**  
*University of Florida, USA & University Magna Græcia of Catanzaro, Italy*

**P.M. Pardalos**  
*University of Florida, USA*

### ABSTRACT

*Data mining techniques are largely used in different sectors of the economy and they increasingly are playing an important role in agriculture and environment-related areas. This paper aims to show our vision on the importance of knowing and efficiently using data mining and machine learning-related techniques for knowledge discovery in the field of agriculture and environment. Efforts for searching hidden patterns in data are not a recent phenomenon. History shows that extensive observations on data have helped discover empirical laws in different fields of research. Therefore, it is important to provide researchers in agriculture and environmental-related areas with the most advanced knowledge discovery techniques. Data mining is the process of extracting important and useful information from large sets of data. This information can be converted into useful knowledge that could help to better understand the problem in study and to better predict future developments. The paper presents the state of the art in data mining and knowledge discovery techniques and provides discussions for future directions.*

DOI: 10.4018/978-1-4666-0333-2.ch017

## INTRODUCTION

The problem of searching for patterns in data is a fundamental one and has a long and successful history. There are many examples in different research areas that extensive observations of data has led to discovering empirical laws. As an example, the attentiv astronomical observations undertaken by several astronoms allowed Kepler to discover the laws of planetary motion (Bishop, 2006).

Over the years, several techniques have been developed to discover hidden patterns in data and these efforts led to the creation of a rigourous discipline known as data mining or knowledge discovery. Data mining is the process of finding useful patterns or correlations amongst data. These patterns, associations, or relationships between data can provide information about the problem in study and information can then be transformed into knowledge. The idea of using information hidden into relationships amongst data inspired researchers to apply these techniques for predicting future trends (Mucherino, Papajorgi, & Pardalos, 2009). Data mining techniques are developed from mainly three areas: statistics, artificial intelligence and machine learning. Although the roots of data mining may seem different, but essentially they aim the same target: discover a relationship that more or less maps measurements in one part of a

data set to measurements in another, linked part of the data set (Pyle, 2003).

Regardless of the method used, the goal of data mining techniques is to split data in different categories, each of them representing some feature of interest the data may have (Mucherino et al., 2009). Thus, fundamental for the success of a data mining technique is the ability to group available data in disjoint categories, where each category contains data with similar properties. The similarity between data is usually measured using a distance function; similar data should belong to the same group or cluster. Therefore, the success of a data mining technique depends on the adequate definition of a suitable distance between data samples.

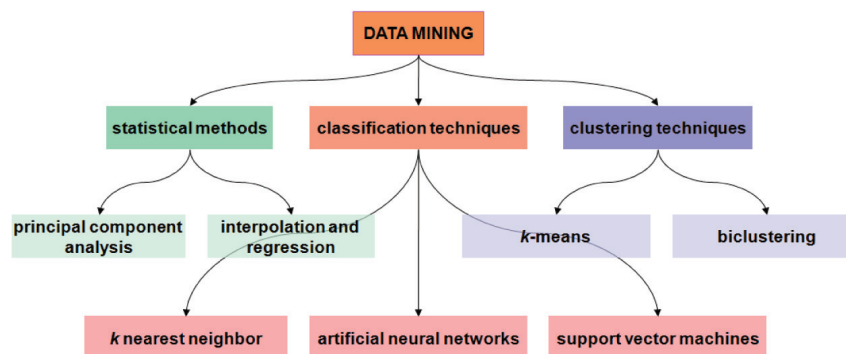
As the similarity between data samples is measured using a distance function, often it occurs that this distance needs to be optimal. Thus, many data mining techniques led to the formulation of a global optimization problem (Mucherino et al., 2009).

Data mining techniques can be grouped in three categories as shown in Figure 1.

## Statistical Methods

Statistical methods such as Principal Component Analysis (PCA) and regression techniques are commonly used as simple methods for finding

Figure 1. A schematic representation of the classification of the data mining techniques



13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/data-mining-techniques-agricultural-environmental/63768](http://www.igi-global.com/chapter/data-mining-techniques-agricultural-environmental/63768)

## Related Content

---

### Analysis of Terrestrial Vegetation Trends and Correlation Between Vegetation Indices and Climatic Factors

Dhanapriya M., Hiren P. Bhattand Vyas S. P. (2020). *Predicting, Monitoring, and Assessing Forest Fire Dangers and Risks* (pp. 233-257).

[www.irma-international.org/chapter/analysis-of-terrestrial-vegetation-trends-and-correlation-between-vegetation-indices-and-climatic-factors/240933](http://www.irma-international.org/chapter/analysis-of-terrestrial-vegetation-trends-and-correlation-between-vegetation-indices-and-climatic-factors/240933)

### Adopting Green ICT in Business

Subramanian Chitra (2011). *Handbook of Research on Green ICT: Technology, Business and Social Perspectives* (pp. 643-651).

[www.irma-international.org/chapter/adopting-green-ict-business/48462](http://www.irma-international.org/chapter/adopting-green-ict-business/48462)

### Towards Self Energy-Management and Sustainable Citizens' Engagement in Local Energy Efficiency Agenda

Francesco Scorza (2016). *International Journal of Agricultural and Environmental Information Systems* (pp. 44-53).

[www.irma-international.org/article/towards-self-energy-management-and-sustainable-citizens-engagement-in-local-energy-efficiency-agenda/153625](http://www.irma-international.org/article/towards-self-energy-management-and-sustainable-citizens-engagement-in-local-energy-efficiency-agenda/153625)

### Detection and Identification of Microbial Volatile Organic Compounds of the Green Mold Disease: MVOC Profile on Different Media

Dalma Radványi, András Geösel, Zsuzsa Jókai, Péter Fodorand Attila Gere (2020). *International Journal of Agricultural and Environmental Information Systems* (pp. 14-28).

[www.irma-international.org/article/detection-and-identification-of-microbial-volatile-organic-compounds-of-the-green-mold-disease/249689](http://www.irma-international.org/article/detection-and-identification-of-microbial-volatile-organic-compounds-of-the-green-mold-disease/249689)

### Prediction of Rice Yield via Stacked LSTM

Xiangyan Meng, Muyan Li and Qiufeng Wu (2020). *International Journal of Agricultural and Environmental Information Systems* (pp. 86-95).

[www.irma-international.org/article/prediction-of-rice-yield-via-stacked-lstm/244149](http://www.irma-international.org/article/prediction-of-rice-yield-via-stacked-lstm/244149)