

## Chapter 3

# Data Classification: Its Techniques and Big Data

**A. Sheik Abdullah**

*Thiagarajar College of Engineering, India*

**R. Suganya**

*Thiagarajar College of Engineering, India*

**S. Selvakumar**

*G. K. M. College of Engineering and Technology,  
India*

**S. Rajaram**

*Thiagarajar College of Engineering, India*

### ABSTRACT

*Classification is considered to be the one of the data analysis technique which can be used over many applications. Classification model predicts categorical continuous class labels. Clustering mainly deals with grouping of variables based upon similar characteristics. Classification models are experienced by comparing the predicted values to that of the known target values in a set of test data. Data classification has many applications in business modeling, marketing analysis, credit risk analysis; biomedical engineering and drug retort modeling. The extension of data analysis and classification makes the insight into big data with an exploration to processing and managing large data sets. This chapter deals with various techniques, methodologies that correspond to the classification problem in data analysis process and its methodological impacts to big data.*

### INTRODUCTION

Data is an abstract concept from which information and knowledge are derived. Raw unprocessed data often moves and crosses stage by stage for its exact representation and processed form of representation. Data is a collection of facts which is the representation of values and measurements.

Meanwhile information is referred to as processed data. It reveals the content or message through direct or indirect form of representation. Hence it is in a meaningful form of representation which can be easily conveyed and understood by the users. It resolves uncertainty and ambiguity.

DOI: 10.4018/978-1-5225-2031-3.ch003

## **Data Classification**

### **Qualities of Data**

The quality signifies the characteristics of data which are specifically suited for the data analysis process. The following characteristics represent the quality of a good data:

1. Accurate
2. Represented numerically
3. Relationship
4. Signified for definite purpose
5. Completeness
6. Clearly Understandable

### **Types of Data Elements**

At the start of every data analysis it is necessary to identify the type of data which can then be considered for analysis. The following represents the types of data elements which can be used up for the determination of the type of data.

#### **Continuous Data**

These are the type of data elements which are defined upon an interval scale. Examples include income of employees in an organization, sales of an enterprise and so on.

#### **Categorical Data**

These kinds of data elements are of three types:

1. **Ordinal Data:** The type of data elements which takes restricted set of values with meaningful ordering. Example includes the classification of age into young, middle age and old group.
2. **Nominal Data:** These are the type of data elements which takes restricted set of values with no any such meaningful ordering between them. Example includes profession of employees, marital status and so on.
3. **Binary Data:** These are the types of data elements that can take only two values. Examples include gender and employment status of an employee.

### **Data Standardization**

Data standardization is the mechanism of normalizing the data to a defined specified range. It provides the mechanism of coding the data to a smaller specified range. The following are the data normalization procedures used up for scaling the given variable.

16 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-classification/178096](http://www.igi-global.com/chapter/data-classification/178096)

## Related Content

---

### Logistics Optimisation: A Cyber Physical Model

Chuks Nnamdi Medohand Arnesh Telukdarie (2020). *International Journal of Business Analytics* (pp. 54-76).

[www.irma-international.org/article/logistics-optimisation/246342](http://www.irma-international.org/article/logistics-optimisation/246342)

### ERP On-Premise or On-Demand

Fan Zhaoand Elias T. Kirche (2018). *International Journal of Business Analytics* (pp. 1-16).

[www.irma-international.org/article/erp-on-premise-or-on-demand/201450](http://www.irma-international.org/article/erp-on-premise-or-on-demand/201450)

### Organizational Characteristics of Middle Managers' Deterioration as Sources of Organizational Decline

Masaru Karube, Toshihiko Katoand Tsuyoshi Numagami (2012). *Managing Dynamic Technology-Oriented Businesses: High-Tech Organizations and Workplaces* (pp. 209-227).

[www.irma-international.org/chapter/organizational-characteristics-middle-managers-deterioration/67437](http://www.irma-international.org/chapter/organizational-characteristics-middle-managers-deterioration/67437)

### Robust Supply Chain Risk Management

Amir H. Ansariipoorand Fernando S. Oliveira (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 2093-2103).

[www.irma-international.org/chapter/robust-supply-chain-risk-management/107396](http://www.irma-international.org/chapter/robust-supply-chain-risk-management/107396)

### Analysis of Dynamics Competitiveness by Using Strategic Groups Maps: Case of Furniture Industry

Hamed Aboutorab, Alireza Aslaniand Mohsen Nazari (2018). *International Journal of Business Analytics* (pp. 52-66).

[www.irma-international.org/article/analysis-of-dynamics-competitiveness-by-using-strategic-groups-maps/205643](http://www.irma-international.org/article/analysis-of-dynamics-competitiveness-by-using-strategic-groups-maps/205643)