| A Case Study |

Statistical and mathematical models in data mining

■ SUNIL KAWALE

Aurhor for
Correspondence :
SUNIL KAWALE
Department of Statistics,
Dr. Babasahed Ambedkar
Marathwada University,
AURANGABAD (M.S.)
INDIA

ABSTRACT : Data mining is common and widely used tools in computer world. Data mining is one of the most important phases of the knowledge discovery in database activity. Therefore there is an increasing need for training professionals to work as analysts or to interface with data mining. The Statistician can be handling very effectively the raw data by modeling in appropriate. The various techniques are used for data mining to get the exact information form the raw data. Here the some of important data mining model is discussed based on statistics and mathematics.

KEY WORDS : Data mining, Statistics and mathematical models, Data analysis.

HOW TO CITE THIS PAPER : Kawale, Sunil (2012). Statistical and mathematical models in data mining, *Internat. Res. J. agric. Eco. & Stat.,* **3** (2) : 359-362.

INTRODUCTION

Data collection was very expensive activity in time and resources before the advent of computer and internet. Data have become so plentiful that corporations have created data warehouse to store them and have hired statistician to analyze their information content and discovering hidden value in warehouse.

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scure databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

The foundations of data mining:

Data mining techniques are the result of a long process of research and product development. This evolution began when business data were first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining is ready for application in the business community because it is supported by technologies that are now sufficiently mature:

Definition of data mining:

Data mining is the process of exploration and analysis by automatic or semiautomatic means of large and analysis ties of data in order to discover meaningful patterns and rules – M.J. Berry and G.S. Linoff.

Data mining process:

The ultimate goal of data mining is prediction and predictive data mining is the most common type of data mining (Fig. 1). The process of data mining consisted of three stages:

Exploration :

It involves data preparation such as clearing data transportation, selecting subsets of records etc.

Models building and validation:

Considering various models and choosing the best one based on their predictive performance.

Development:

Choosing the model best one and applying it on to new data in order to predictions or estimate of the expected out comes models for data mining.