CHAPTER-8

DATA MINING TASKS
CHAPTER-8
DATA MINING TASKS

8.1 INTRODUCTION

The goal of any data mining effort can be divided in one of the following two types (Cha & Lweis, 2002:57): [22].

- Using data mining to generate descriptive models to solve problems.
- Using data mining to generate predictive models to solve problems.

The descriptive data mining tasks characterize the general properties of the data in the database, while predictive data mining tasks perform inference on the current data in order to make prediction. Descriptive data mining focus on finding patterns describing the data that can be interpreted by humans, and produces new, nontrivial information based on the available data set. Predictive data mining involves using some variables or fields in the data set to predict unknown or future values of other variables of interest, and produces the model of the system described by the given data set. The goal of predictive data mining is to produce a model that can be used to perform tasks such as classification, prediction or estimation, while the goal of descriptive data mining is to gain an understanding of the analysed system by uncovering patterns and relationships in large data sets.

The goal of a descriptive data mining model is therefore to discover patterns in the data and to understand the relationships between attributes represented by the data, while the goal of a predictive data mining model is to predict the future outcomes based on passed records.
with known answers. Further divide the data mining task of generating models into the following two approaches:

- Supervised or directed data mining modeling.
- Unsupervised or undirected data mining modeling.

The goal in supervised or directed data mining is to use the available data to build a model that describes one particular variable of interest in terms of the rest of the available data. The task is to explain the values of some particular field. The user selects the target field and directs the computer to determine how to estimate, classify or predict its value.

In unsupervised or undirected data mining however variable is sigled out as the target. The goals of predictive and descriptive data mining are achieved by using specific data mining techniques that fall within certain primary data mining tasks. The goal is rather to establish some relationship among all the variables in the data. The user asks the computer to identify patterns in the data that may be significant. Undirected modeling is used to explain those patters and relationships one they have been found.

### 8.2 CLASSIFICATION/PREDICTION

Classification involves the discovery of a predictive learning function that classifies a data item into one of several predefines classes. It involves examining the features of a newly presented object and assigning to it a predefined class. Define classification has a two-step process. First a model is built describing a predetermined set of data classes or concepts and secondly, the model is used for classification.

Prediction can be viewed as the construction and use of a model to assess the class of a unlabeled sample, or to assess the value or value
range of an attribute that a given sample is likely to have. According to, any of the techniques used for classification can be adapted for use in prediction by using training examples where the value of the variable to be predicted is already known, along with historical data for those examples. Typical business related questions that can be answered using classification or prediction tasks are:

- Which customers will buy?
- Which products will customer by?
- How much will customer buy?

8.3 ESTIMATION

While classification deals with discrete outcomes such as yes or no, debit card, home loan or vehicle financing, estimation deals with continuously valued outcomes. If some input data is available, estimation cab be used to come up with some unknown continuous variable such as income or height. In estimation, one wants to come up with a plausible value or a range of plausible values for the unknown parameters of a system. Typical examples of estimation and the business related questions that can be addressed by making use it include the following:

- How many children are in a family?
- Estimating a family’s total household income.
- Estimating the value of a piece of property.

Classification and estimation are often used together, as when data mining is used or predict who is likely to respond to a credit card balance transfer offer and also to estimate the size of the balance to be transferred. Estimation is grouped under predictive data mining tasks.
8.4 SEGMENTATION

Segmentation simply means making different offers to different markets segments; groups of people defined by some combination of demographic variables such as age, gender or income. Define the segmentation as a form of analysis used to for instance break down the visitors to a website into unique groups with individual behaviors. The grouping can then be used to make statistical projections, such as the potential amount of purchases they are likely to make. Typical business questions that can be answered using segmentation are:

What are the different types of visitors attracted to our website?

In which age groups do the listeners of a certain radio station fall into?

Segmentation is grouped under descriptive data mining tasks.

8.5 CLUSTERING

Clustering is the task of segmenting a diverse group into a number of similar subgroups or clusters. Clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Clustering is commonly used to search for unique groupings within a data set. The distinguishing factor between clustering and classification is that in clustering there are no predefined classes and no examples. The objects are grouped together based on self-similarity. Typical business question that can be answered using clustering are: What are the groupings hidden in your data. Which customer should be grouped together for target marketing purposes?

Clustering is grouped under descriptive data mining tasks.
8.6 DESCRIPTION AND VISUALISATION

The purpose of data mining is sometimes simply to describe what is going on in a complicated database in a way that increased our understanding of the people, products or processes that produced the data in the first place. They state that a good enough description of behavior will often suggest an explanation for it as well. One of the most powerful forms of descriptive data mining is data visualization. Although visualization is not always easy, the right picture can truly speak a thousand words since human beings are extremely practiced at extracting meaning from visual scenes. Visualization can be useful in providing a visual representation of the location and distribution of a company’s major clients on a map of a city or a province or even a country. Allowing the visualization of discovered patterns in various forms can help users with different backgrounds to