ATHENS UNIVERSITY
OF ECONOMIC AND BUSINESS
DEPARTMENT OF INFORMATICS

MSc IN INFORMATION SYSTEMS

MSc Thesis

«Using Data Mining to Support Marketing Business Analysis in Traditional and Internet Retailing»

George Prassas
M3990010

ATHENS, FEBRUARY 2001

MSc in INFORMATION SYSTEMS

MSc Thesis

«Using Data Mining to Support Marketing Business Analysis in Traditional and Internet Retailing»

George Prassas

M3990010

Supervisor Professor: Georgios I. Doukidis
Second Referee: Angeliki Poulimenakou

ATHENS UNIVERSITY OF ECONOMIC AND BUSINESS
DEPARTMENT OF INFORMATICS

ATHENS, FEBRUARY 2001

Contents

Contents

Executive summary

Marketing Business Analysis in Retailing has focused in the last three decades in resolving the interrelationships between the consumers' decision process and their purchase preferences. The marketing research community has used known statistical analysis methods in controlled environments to affirm hypotheses about the factors that relate to the decision process.

POS scanning technology in the 80s and the Internet in the 90s have boosted the data collection capabilities of the information systems. This increase in data required more intelligent methods for analysis. Advanced data processing algorithms have arisen from the area of machine learning. Knowledge discovery in databases (or data mining) is an area in computer science that is the evolution of machine learning and provides knowledge extraction techniques that can cope with enormous sets of data.

In this thesis, we identify possible data mining models that facilitate marketing business analysis in traditional and Internet retailing. We were based on the CRISP-DM methodology to achieve this goal and to structure the thesis. Firstly, we used marketing and retailing bibliography to identify the most value-adding business cases and then built the mining models that assist the marketing managers; in every mining model we pinpoint the knowledge extracted from the historical data and its value to the business case. Additionally, for every mining model we list the required data to be collected and the set of algorithms, which are appropriate for each case.

Extensions of this work have two axes of approach. The first would be to expand to other marketing business cases that this thesis has not covered and the second would be to validate the models that were proposed in the thesis by applying the models on real historical data.

Regarding further research, the concept of applying data mining to support Marketing Business Analysis can be evaluated in the new scenery of retailing that will be formed with the intrusion of "smart" devices and "smart" homes along with the mobile network expansion. These changes in retailing will undoubtedly differentiate the Marketing Business Analysis of this environment.

1 Introduction

Retailing is a market with a vast customer base worldwide; its behavior has underlain the research interest of many research communities during the century. The results of this research are part of our everyday interaction with the retail stores and all of us are the receptors of carefully designed, indirect signals that stimulate our purchasing intentions. Repeated purchasing in a high frequent base is the main characteristic of retailing and the repeated patronage phenomenon is the crux of the marketing research in this area.

Repeated patronage and generally repeated purchasing behavior has been the focus of the majority of marketing researchers along with identifying the factors and their relation to the purchasing decision process. This decision process has turned to be very complicated due to the plethora of correlated elements that characterize different dimensions like the product, the store, the customer, and the social evolutions. No general applicable model exists that explains these correlations. The marketing research has focused in identifying the relations between specific factors each time in controlled environments, using samples of customers.

The results of the marketing research in retailing is a large set of hypotheses affirmed on controlled environments and there is no solid way of measuring their success on real cases due to the large number of factors that may affect the outcomes of any marketing policy. Marketing performance measures is another important issue that attracted academic and managerial attention with an urgency and scope unprecedented in the field's history. Different trends have converged proving the multidimensionality of the problem.

Data collection and data processing were the main obstacles before the evolution of technology. The main methods of data processing used in the marketing literature for the last three decades have originated in statistics and econometrics; probably the most common is regression analysis. In the early 70s these methods were used to find the correlations between variables with the limitation of correlating only a few variables on small samples of data due to the lack of computational assistance and the limitation in data collection [Sevin, 1965][Goodman, 1970]. In the late 70s and early 80s, the database growth explosion combined with POS (point-of-sale) scanning technology has given a boost to data collection. In parallel, the broad expansion of computers has given cheap computational power that facilitated the statistic processing using the traditional methods [Guadagni and Little, 1983].

In the late 80s, more advanced techniques on data processing have arisen from the area of machine learning [Jain and Dubes, 1988][Fisher, 1987]. These techniques were used on a small scale in the marketing literature at that period. The main reasons were that mass marketing was adequate for the time being and the data collected in the databases were not in a flexible exploitable form. The second obstacle was overcome with the evolution to data warehouses, where data was collected not only for operational purposes but for managerial purposes as well. The early 90s have revealed the need for target marketing, an issue that was discussed in the marketing literature several years before. Machine learning has fed a new area of research that was named knowledge discovery in databases or data mining [Fayyad et al, 1996]. Data mining gained great amount of attention from the computer science community and has set new fundamentals in data exploitation. John Nibblet's modo summarizes the marketing need of the last decade: "We are drowning in information, but starving for knowledge". Data mining evolved data exploitation because it confined the need for direct human interference. Traditional methods required forming the hypotheses between the correlations of variables; data mining automatically and intelligently extracts these correlations without any hypotheses.

Although data mining had at last set the means of revealing the customer's behavior, the knowledge extracted would never be fully applicable without the Internet explosion. Traditional retailing has one major limitation: there can be only one store for all customers. Target marketing was limited to targeted mail campaigns and abstract target promotion campaigns. However, the crux of a marketing policy is inside the store, ranging from the product and category assortment to the promotion of every product.

Internet retailing has facilitated target marketing and set the fundamentals for one-to-one marketing. One-to-one marketing visions one different store for each customer that is designed to fulfill his/her personal and unique needs and preferences. Data mining is the technology that can facilitate the achievement of such a goal by dynamically extracting the rules that should govern each store.

Additionally, data collection has boosted again (since the POS scanning technology) by providing the possibility to monitor in detail every action of the customer. Without the need of a loyalty card, in Internet retailing the customer is always associated with the shopping cart and purchasing ceases to be anonymous; this allows focused analysis per customer, whilst in traditional retailing data processing was made for the whole of the customer base. Apart from the purchasing habits of the customers, in Internet retailing different kind of behavior can be monitored that reveals the purchase intentions and the preferences regarding other factors like store atmosphere. Also, part of the decision process can be recorded by monitoring what comparisons the customer makes before the final purchasing decision, the direct effects of promotions, impulse buying etc.

Similar situations with Internet retailing are also met in the evolving area of mobile commerce. We are all experiencing the rapid expansion of mobile telecommunications and their intrusion in everyday life, which is even faster than the expansion of the Internet. Mobile retailing is maybe the next step and undoubtedly an area worth researching. Additionally, the technology of smart homes and smart home devices that will recognize in precision the product life cycles for every household will give the third boost in data collection, since it will expand it outside the store and contain detailed information about the consuming behavior.

The future seems to be promising for even more radical and surprising evolutions.

1.1 Purpose and outline of the thesis

The purpose of the thesis is to identify possible data mining models that facilitate marketing business analysis in traditional and Internet retailing. We try to exploit all the available data in order to facilitate the process of one-to-one marketing. The goal is dual: Firstly, to identify the parts of the marketing business cases in the traditional retail environment that can be assisted by data mining and provide more reliable results without direct human interference. Secondly, to identify new marketing business cases due to the expanded data collection in Internet retailing, that can also be assisted by data mining techniques.

The literature that has been used in this thesis comes from three different areas: retailing, marketing and knowledge discovery. The thesis outline has been based in the data mining methodology CRISP-DM. CRISP-DM has not been built in a theoretical, academic manner working from technical principles. Since no such methodology exists and based on the fact that it is becoming the de facto standard for the industry, we used it as an outline. The purpose of this thesis is not to apply in detail CRISP-DM methodology, but to use it as a guide to recognize the key issues that arise by exploiting data mining technologies in traditional and Internet retailing.

The chapters are organized as follows:

Chapter 1 serves as the introduction of the thesis. Chapter 2 overviews data mining technology and describes in summary the CRISP-DM reference model, which includes a quick overview of phases, tasks and their outputs. Chapter 3 is the core of the thesis, using the phases of CRISP-DM as a framework to describe the business cases and the respective mining models used to support them, as long as the data associated with them. Finally, chapter 4 concludes with some general thoughts and suggestions regarding further research and development in this area.

2 Knowledge discovery in databases

2.1 The concept of knowledge discovery

In the last decade [Fayyad et al., 1996], we have seen an explosive growth in our capabilities to both generate and collect data. The widespread introduction of bar codes for almost all commercial products and the computerization of many businesses (e.g. credit cad purchases) have generated a flood of data. Advances in data storage technology, such as faster, higher capacity, and cheaper storage devices, better database management systems, and data warehousing technology, have allowed transforming this data deluge into "mountains" of stored data.

Representative examples are easy to find. In the business world, one of the largest databases in the world was created by Wal-Mart, the biggest U.S. retailer, which handles over 20 million transactions a day [Babcock, 1994]. Such volumes of data clearly overwhelm the traditional manual method of data analysis such as spreadsheets and ad-hoc queries. Those methods can create informative reports from data, but cannot analyze the contents of those reports to focus on important knowledge. The need for a new generation of techniques and tools with the ability to intelligently and automatically assist humans in analyzing the mountains of data for nuggets of useful knowledge. These techniques and tools are the subject of the emerging field of knowledge discovery in databases (KDD).

Historically the notion of finding useful patterns (or nuggets of knowledge) in raw data has been given various names, including knowledge discovery in databases, data mining, knowledge extraction, information discovery, information harvesting, data archeology and data pattern processing. The term *knowledge discovery in databases*, or KDD for short, was coined in 1989 to refer to the broad process of finding knowledge in data, and to emphasize the "high-level" application of particular data mining methods. The term *data mining* has been commonly used by statisticians, data analysts and the MIS (Management Information Systems) community, while KDD has been used by artificial intelligence and machine learning researchers.

KDD systems draw upon methods, algorithms and techniques from diverse fields like machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for expert systems and data visualization.

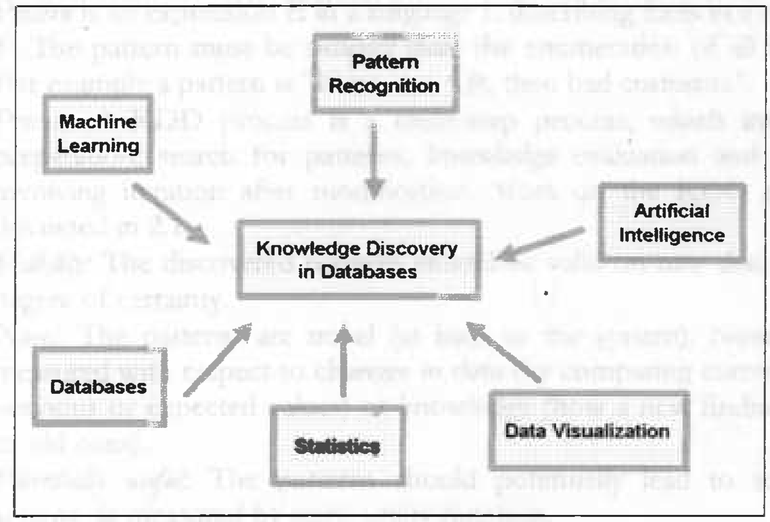


Figure 2.1 Associated fields with KDD

The fields of machine learning and pattern recognition overlap with KDD in the study of theories and algorithms for systems, which extract patterns, and models for data (mainly data mining methods). KDD focuses on the extension of these theories and algorithms to the problem of finding special patters in large sets of real-world data. KDD also has much in common with statistics, particularly exploratory data analysis (EDA). KDD systems often embed particular statistical procedures for modeling data and handling noise within an overall knowledge discovery framework.

Another related area is data warehousing, which refers to the MIS trend for collecting and cleaning transactional data and making them available for on-line retrieval. A popular approach for analysis of data warehouse is OLAP (on-line analytical processing) [Codd, 1993]. OLAP tools focus on providing multi-dimensional data analysis, which is superior to SQL (standard query language) in computing summaries and breakdowns along many dimensions.

Though there are many definitions for knowledge discovery in databases and data mining, most definitions reference the one given in [Fayyad et a., 1996].

Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data.

We examine in more detail the terms of the definition,

- *Data* is a set of facts (e.g. cases/records of a database). For example a transaction database is a set of transactions and a customer database is a set of each customer characteristics.
- *Pattern* is an expression E in a language L describing facts in a subset F_E of F. The pattern must be simpler than the enumeration of all facts in F_E. For example a pattern is "If income <$t, then bad customer".
- *Process*: A KDD process is a multi-step process, which involves data preparation, search for patterns, knowledge evaluation and refinement involving iteration after modification. More on the KDD process are discussed in 2.1.
- *Validity:* The discovered patterns should be valid on new data with some degree of certainty.
- *Novel*: The patterns are novel (at least to the system). Novelty can be measured with respect to changes in data (by comparing current values to previous or expected values) or knowledge (how a new finding is related to old ones).
- *Potentially useful*: The patterns should potentially lead to some useful actions, as measured by some utility function.
- *Ultimately understandable*: A goal of KDD is to make patterns understandable to humans in order to facilitate a better understanding of underlying data. Because of the difficulty of measuring, frequent substitutes are simplicity measures, which range from purely syntactic to the semantic.

> *Data mining* is a step in the KDD process consisting of particular data mining algorithms that, under some acceptable computational efficiency limitations, produces a particular enumeration of patterns E_j over F.

In the following sections we discuss the KDD process and several basic data mining techniques. The KDD process that is discussed in 2.2 is an evolution of the KDD process that was first introduced in [Fayyad et al., 1996], and is becoming a de facto standard for the industry.

2.2 The CRISP methodology

CRISP-DM was conceived in late 1996 by three "veterans" of the young and immature data mining market. DaimlerChrysler (then Daimler-Benz) was already experienced, ahead of most industrial and commercial organizations, in applying data mining in its business operations. SPSS (then ISL) had been providing services based on data mining since 1990 and had launched the first commercial data mining workbench – Clementine – in 1994. NCR, as part of its aim to deliver added value to its Teradata data warehouse customers, had established teams of data mining consultants and technology specialists to service its clients' requirements. At that time, early market interest in data mining was showing signs of exploding into widespread uptake. This was both exciting and terrifying. In these early years of the data mining market different approaches were being developed, without anyone knowing if the correct procedure was being followed. In order to stop this trial and error learning for every new adopter and make data mining a mature technology that can become part of the business processes, these three companies co-operated in the creation of a standard process model.

A year later a consortium was formed with the acronym (CRoss-Industry Standard Process for Data Mining), that obtained funding from the European Commission and begun to set out the initial ideas of the three companies. As CRISP-DM was intended to be industry-, tool- and application-neutral, input was needed from as wide a range as possible of practitioners and others (such as data warehouse vendors and management consultancies) with a vested interest in data mining. This was the primary reason for the creation of the CRISP-DM Special Interest Group ("The SIG", as it became known). It was launched by broadcasting an invitation to all interested parties to join the founders for a day-long workshop in order to share ideas openly discuss how to take CRISP-DM forward. The workshop surpassed all expectations. The need for a standard process model became clear and there was tremendous common ground in how everybody viewed the process of data mining, despite the demarcation of phases and terminology.

For the next two and a half years, CRISP-DM was developed and refined. It was tested on large-scale data mining projects at Mercedes-Benz and at OHRA (insurance sector partner). It was also integrated with commercial data mining tools. By the end of the EC-funded part of the project – mid-1999 – a good-quality draft of the process model was delivered. CRISP-DM 1.0 [Chapman et al., 2000] that was released one year later is much more complete and better presented, and does not differ radically from the first version. During the project the process model was still a work-in-progress since it was validated on a narrow set of projects. Over the past year, DaimlerChrysler had the opportunity to apply CRISP-DM to a wider range of applications. SPSS' and NCR's Professional Services groups have adopted CRISP-DM and used it successfully on numerous customer engagement covering many industries and business problems.

Throughout this time, service suppliers from outside the consortium have adopted CRISP-DM and repeatedly analysts refer to it as the de facto standard for the industry.

CRISP-DM has not been built in a theoretical, academic manner working from technical principles, nor did elite committees of gurus create it behind closed doors. Both these approaches to developing methodologies have been tried in the past, but have seldom led to practical, successful and widely–adopted standards. CRISP-DM succeeds because it is soundly based on the practical, real-world experience of how people do data mining projects.

2.2.1 Introduction - the CRISP-DM methodology

Hierarchical breakdown

The CRISP-DM data mining methodology is described in terms of a hierarchical process model, consisting of sets of tasks described at four levels of abstraction (from general to specific): phase, generic task, specialized task and process instance (see figure 2.2.1).

At the top level, the data mining process is organized into a number of phases; each phase consists of several second-level generic tasks. This second level is called generic, because it is intended to be general enough to cover all possible data mining situations. The generic tasks are intended to be as complete and stable as possible. Complete means covering both the whole process of data mining and all possible data mining applications. Stable means that the model should be valid for yet unforeseen developments like new modeling techniques.

The third level, the specialized task level, is the place to describe how actions in the generic tasks should be carried out in certain specific situations. For example, at the second level there might be a generic task called clean data. The third level describes how this task differed in different situations, such as cleaning numeric values versus cleaning categorical values or whether the problem type is clustering or predictive modeling.

The description of phases and tasks as discrete steps performed in a specific order represent an idealized sequence of events. In practice, many of the tasks can be performed in a different order and it will often be necessary to repeatedly backtrack to previous tasks and repeat certain actions. CRISP-DM's process model does not attempt to capture all of these possible routes through the data mining process because this would require an overly complex process model.

The fourth level, the process instance, is a record of the actions, decisions and results of an actual data mining engagement. A process instance is organized according to the tasks defined at the higher levels, but represents what actually happened in a particular engagement, rather than what happens in general.

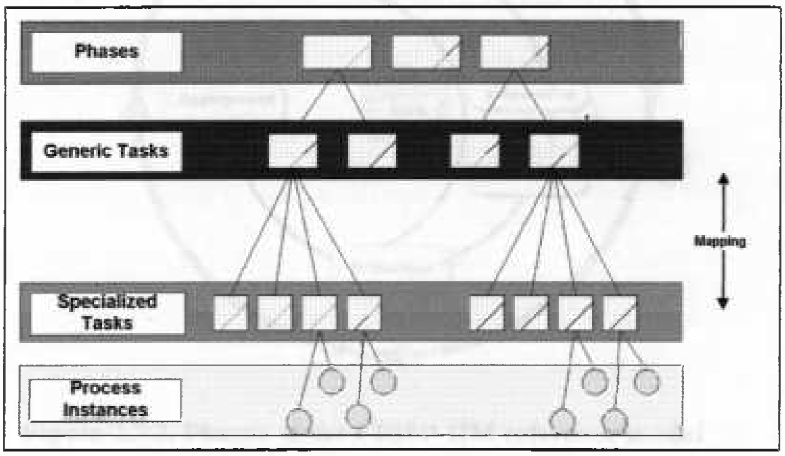


Figure 2.2.1: Four level breakdown of the CRISP-DM methodology

Mapping generic models to specialized models

The data mining context drives mapping between the generic and the specialized level in CRISP-DM. Four different dimensions of data mining contexts are distinguished:

- The **application domain** is the specific area in which the data mining project takes place. Examples are response modeling, customer retention etc.
- The **data mining problem type** describes the specific class(es) of objective(s) that the data mining project deals with. Examples are classification, segmentation, prediction etc.
- The **technical aspect** covers specific issues in data mining that describe different (technical) challenges that usually occur during data mining. Examples are outliers, decision trees etc.
- The **tool and technique** dimension specifies which data mining tool(s) and/or techniques are applied during the data mining project. Examples are Clementine, Intelligent Miner etc.

2.2.2 The CRISP-DM reference model

The CRISP-DM process model for data mining provides an overview of the life cycle of a data mining project. It contains the phases of a project, their respective tasks and relationships between these tasks. At this description level, it is not possible to identify all relationships. Essentially, relationships could exist between any data mining tasks depending on the goals, the background and interest of the user and most importantly on the data.

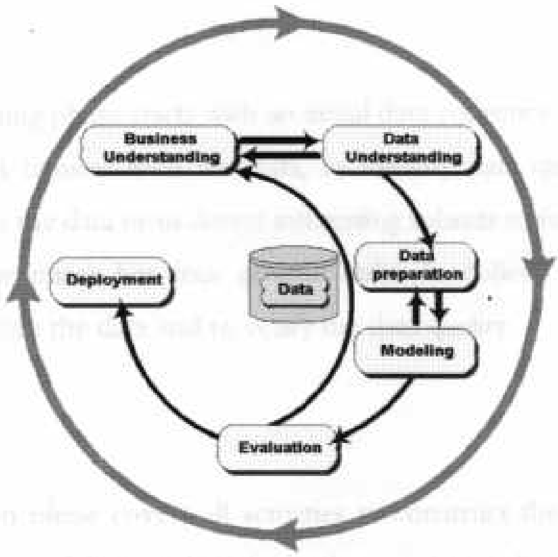


Figure 2.2.2. Phases of the CRISP-DM reference model

The life cycle of a data mining project consists of six phases. Figure 2.2.2 shows the phases of a data mining process. The sequence of the phases is not rigid. Moving back and forth between different phases is always required. It depends on the outcome of each phase, which phase or which particular task of a phase has to be performed next. The arrows indicate the most important and frequent dependencies between phases.

The outer circle in Figure 2.2.2 symbolizes the cyclic nature of data mining itself. Data mining is not over once a solution is deployed. The deployed solution can trigger new, often more focused business questions. Subsequent data mining processes benefit from the experiences of previous ones.

In the following, we briefly outline each phase:

Business understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives. This phase has four generic tasks: to determine the business objectives, to assess the situation (more detailed analysis of the first task), to determine the data mining goals and produce the project plan.

Data understanding

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information. This phase has four generic tasks: to collect the initial data, to describe the data, to explore the data and to verify the data quality.

Data preparation

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record and attribute selection as well as transformation and cleaning of data for modeling tools. This phase consists of five generic tasks: to select the data, to clean the data, to construct data, to integrate the data and to format the data.

Modeling

In this phase, various modeling techniques are selected and applied and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often necessary. This phase consists of four generic tasks: to select the modeling technique, to generate the test design, to build the model and to assess the model.

Evaluation

At this stage in the project the model (or models) built, appears to have high quality from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model and review the steps executed to construct the model to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached. This phase consists of three generic tasks: to evaluate the results, to review the process and to determine the next steps.

Deployment

Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. It often involves applying "live" models within an organization's decision making processes, for example in real-time personalization of Web pages or repeated scoring of marketing databases. However, depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. In many cases it is the customer, not the data analyst, who carries out the deployment steps. However, even if the analyst will not carry out the deployment effort it is important for the customer to understand up front what actions need to be carried out in order to actually make use of the created models. This phase consists of four generic tasks: to make the deployment plan, to plan monitoring and maintenance, to produce the final report and to review the project.

2.3 Data mining techniques

In this chapter we review the basic notions of the three most well known categories of data mining: clustering, predictive modeling and association rule mining. Chen et al. [1996] give a complete overview of the area of data mining for further reference.

2.3.1 Clustering

The process of grouping physical or abstract objects into classes of similar objects is called clustering or unsupervised classification. Clustering analysis helps construct meaningful partitioning of a large set of objects based on a "divide and conquer" methodology, which decomposes a large-scale system into smaller components to simplify design and implementation.

As a data mining task, data clustering identifies clusters, or densely populated regions, according to some distance measurement, in a large, multidimensional data set. Given a large set of multidimensional data points, the data space is usually not uniformly occupied by the data points. Data clustering identifies the sparse and crowded places, and hence discovers the overall distribution patterns of the data set (Figure 2.3.1). Data clustering has been studied in statistics [Fayyad et al., 1996], machine learning [Fisher, 1995] and data mining [Zhang et al., 1996] areas with different emphases.

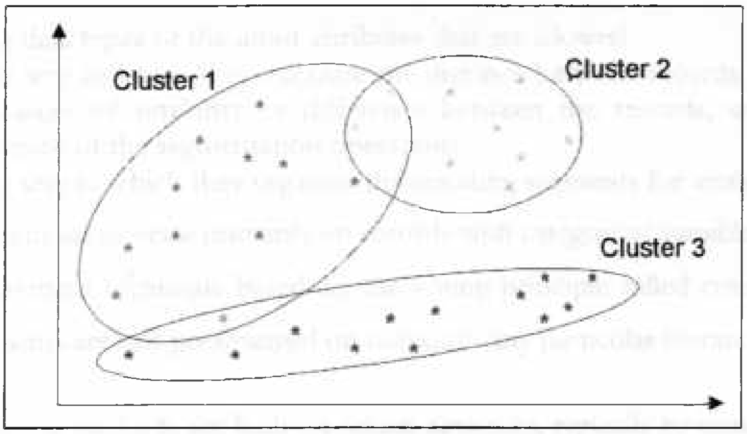


Figure 2.3.1 A simple clustering of a 2-dimensional data set

The goal of database clustering is to partition a database into clusters (segments) of similar records, that is, records that share a number of properties and so are considered to be homogeneous. In some literature the words clustering and segmentation are used interchangeably. Here, we use clustering to describe the data mining operation, and segments or clusters to describe the resulting groups of data records. By definition, two records in different segments are different in some way. The segments should have high internal (within segment) homogeneity and high external (between segment) heterogeneity.

Database clustering is typically done to discover homogeneous subpopulations in a customer database to improve the accuracy of the profiles. A subpopulation, which might be wealthy, older, males or urban, professional females, can be targeted for specialized treatment. Equally, as databases grow and are populated with diverse types of data, it is often necessary to partition them into collections of related records to obtain a summary of each database or before performing a data mining operation such as predictive modeling (classification).

A clustering algorithm can segment a database without any prompting from the user about the type of segments or even the number of segments it is expected to find in the database. Thus, any element of human bias or intuition is removed, and the true discovery nature of the mining can be leveraged. When an algorithm works in this way, the approach is called *unsupervised learning*.

Database clustering can be accomplished by using either demographic or neural clustering methods. The methods are distinguished by

- the data types of the input attributes that are allowed
- the way in which they calculate the distance between records (that is, the measure of similarity or difference between the records, which is the essence of the segmentation operation)
- the way in which they organize the resulting segments for analysis

Clustering methods operate primarily on records with categorical variables. They use a distance measurement technique based on the voting principle called condorect, and the resulting segments are not prearranged on output in any particular hierarchy.

Neural clustering methods are built on neural networks, typically by using Kohonen feature maps. Neural networks accept only numeric input, but categorical input is possible by first transforming the input variables into quantitative variables. The distance measurement technique is based on Euclidean distance, and the resulting segments are arranged in a hierarchy where the most similar segments are placed closest together.

Clustering differs from other data mining techniques in that its objective is generally far less precise than the objectives of predictive modeling or link analysis. As a result, clustering algorithms are sensitive to redundant and irrelevant features. This sensitivity can be alleviated by directing the algorithm to ignore a subset of the attributes that describe each instance or by assigning a weight factor to each variable.

Clustering supports such business applications as customer profiling or target marketing, cross selling, and customer retention. Clearly, this operation has broad, cross-industry applicability.

2.3.2 Predictive modeling

Another important application of data mining is the ability to perform predictive modeling in a huge amount of data. This is referred to as mining classification rules. Data classification is to classify a set of data based on their values in certain attributes.

Predictive modeling is akin to the human learning experience, where we use observations to form a model of the essential, underlying characteristics of some phenomenon. For example, in its early years, a young child observes several different examples of dogs and can then later in life use the essential characteristics of dogs to accurately identify (classify) new animals as dogs.

This predictive ability is critical in that it helps us make sound generalizations about the world around us and to fit new information into a general framework. In data mining, a predictive model is used to analyze an existing database to determine some essential characteristics about the data. Of course, the data must include complete, valid observations from which the model can learn how to make accurate predictions. The model must be told the correct answer to some already solved cases before it can start to make up its own mind about new observations. When an algorithm works in this way, the approach is called *supervised learning*. Physically, the model can be a set of *If-Then* rules in some proprietary format, a block of SQL, or a segment of C source code.

Figure 2.3.2 illustrates the predictive modeling approach. Here a bank is interested in automatically deciding whether future applicants will be given a loan or not. A predictive model has determined that only two variables are of interest: the client's income and the client's debt. The decision tree presents the analysis in an intuitive way.

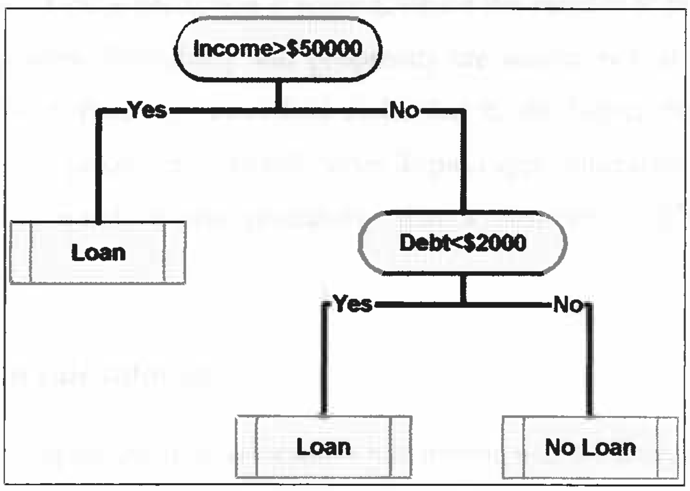


Figure 2.3.2. A decision tree representing the predictive modeling approach

Models are developed in two phases: training and testing. Training refers to building a new model by using historical data, and testing refers to trying out the model on new, previously unseen data to determine its accuracy and physical performance characteristics. Training is typically done on a large proportion of the total data available, whereas testing is done on some small percentage of the data that has been held out exclusively for this purpose. The predictive modeling approach has broad applicability across many industries. Typical business applications that it supports are customer retention management, credit approval, cross selling and target marketing.

There are two specializations of predictive modeling: classification and value prediction. Although both have the same basic objective, namely, to make an educated guess about some variable of interest, they can be distinguished by the nature of the variable being predicted.

With *classification*, a predictive model is used to establish a specific class for each record in a database. The class must be one from a finite set of possible, predetermined class values. The example in Figure 2.3.2 is a case in point. The variable of interest is the approval of a loan, and it has two possible values: "loan" and "no loan".

With *value prediction*, a predictive model is used to estimate a continuous numeric value that is associated with a database record. For example, a car retailer may want to predict the lifetime value of a new customer. A mining run on the historical data of present long-standing clients, including some agreed-upon measure of their financial worth produces a model that can estimate the likely lifetime value of new customers.

A specialization of value prediction is scoring, where the variable to be predicted is a probability or propensity. Probability and propensity are similar in that they are both indicators of likelihood. Both use an ordinal scale, that is, the higher the number, the more likely it is that the predicted event will occur. Typical applications are the prediction of the likelihood of fraud or the probability that a customer will respond to a promotional mailing.

2.3.3 Association rule mining

The prototypical application of association rule mining was the analysis of sales data of retailing organizations. Typical business decisions that the management of the supermarket has to make include what to put on sale, how to design coupons, how to place merchandise on shelve in order to maximize profits etc. Analysis of past transaction data is the most common approach in order to improve the quality of the decisions. Before the POS scanning, only global data about the cumulative sales during a specific period was available. Barcode technology has made it possible for retail organizations to collect and store massive amounts of sales data, referred to as *basket data.* Basket data store items purchased on a per-transaction basis. Basket data type transactions do not necessarily consist of items bought together at the same point of time; it may consist of items bought by a customer over a period of time [Fayyad et al, 1996].

Several organizations have collected massive amounts of basket data. These data are usually stored on tertiary storage and are very slowly migrated in database systems. One of the main reasons for the limited success of database systems in this area is that current database systems do not provide necessary functionality for a user interested in taking advantage of this information. The problem of mining association rules [Agrawal wt al, 1993] regards the extraction of association rules from databases; An example of association rule is the statement that 90% of transactions that purchase bread and butter also purchase milk. The antecedent of this rule consists of bread and butter and the consequent of milk alone. The number 90% is the confidence factor of the rule.

Association rule mining goals answering enhanced queries like the following:

- Find all rules that have "Diet Coke" as consequent. These rules may help plan what the store should do to boost the sale of Diet Coke.
- Find all rules that have "bagels" in the antecedent. These rules may help determine what products may be impacted if the store discontinues selling bagels.
- Find all rules that have "sausage" in the antecedent and "mustard" in the consequent. This query can be phrased alternatively as a request for the additional items that have to be sold together with sausage in order to make highly likely that mustard will also be sold.
- Find all rules relating items located on shelves A and B in the store. These rules may help shelf planning by determining if the sale of items on shelf A is related to the sale of items on shelf B.

We describe in short the formal model of *association rules.*

Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of literals, called items. Let D be a set of transactions, where each transaction T is an itemset such that $T \subseteq I$. In other words, $I = \{i_1, i_2, \ldots, i_m\}$ is a set of attributes over the binary domain $\{0,1\}$. A tuple of the database D is represented by identifying the attributes with value 1. Associated with each transaction is a unique identifier, called *TID*. A set of items $X \subset I$ is called an *itemset.* A transaction T contains an itemset X, if $X \subseteq T$. An *association rule* is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = 0$. The rule $X \Rightarrow Y$ holds in the transaction set D with *confidence c* if c% of transactions in D that contain X also contain Y. The rule $X \Rightarrow Y$ has *support* s in the transaction set D if s% of transactions in D contain $X \cup Y$.

Given a set of transactions D, the problem of mining association rules is to generate all association rules that have certain user-specified minimum support (called *minsup*) and confidence (called *minconf*). The most important thing of association rule mining is that it is interested in discovering all rules rather that verifying whether a particular rule holds.

The extraction of association rules is of sufficient support and confidence is decomposed into two subproblems [Agrawal et al, 1993]:

1. Find all combinations of items that have transaction support above minimum support. Call those combinations large itemsets and all other combinations small itemsets.

2. Use the large itemsets to generate the desired rules. If the itemsets ABCD and AB are large itemsets, the rule AB⇒CD holds if the ratio r = support(ABCD)/support(AB) is greater than *minconf*.

The first efficient algorithms to for solving the problem are Apriori and AprioriTID [Agrawal and Srikant, 1994]. These algorithms focus on the efficient retrieval of all the large item sets, which are all the item sets that have support larger than *minsup*.

Since the first formation of the problem, many other similar problems have appeared. Srikant and Agrawal [1995] introduced *generalized association rules*, which are extracted form a database of transactions, where each transaction consists of a set of items, and a taxonomy (*is-a* hierarchy) on the items. The associations are extracted at any level of the hierarchy. For example, given a taxonomy that says that jackets *is-a* outerwear *is-a* clothes, a rule "people who buy outerwear tend to buy shoes" might be inferred.

Srikant and Agrawal [1996] introduced *quantitative association rules* to cover the cases of tables that contain both quantitative and categorical data. An example of such an association might be '10% of married people between age 50 and 60 have at least two cars". This mining problem is referred to as the Quantitative Association Rules problem in comparison with the original, which is referred to as the Boolean Association Rules problem.

Another specific case of quantitative association rules is the *profile association rules* [Aggarwal et al, 1998]. Profile association rules are the ones in which the left hand side of the rule consists of customer profile information, such as age, salary, years of education and marital status. The right hand side of such rules consists of customer behaviour information, such as buying milk, diaper, beer etc.

Finally, the extraction of *sequential patterns* [Agrawal and Srikant, 1995] is another case of exploiting basket data and the solutions proposed depend on the theory of association rules. The problem refers to the extraction of sequences in the data, where a sequence is an ordered list of itemsets.

3 Applying CRISP-DM in retailing

In this chapter we are interested in identifying the basic issues arising, when applying CRISP-DM in retailing. The purpose of this chapter is to give insight to the extended opportunities that data mining offers in the retailing area. More specifically, we plan in the first phase to record business cases of traditional and Internet retailing that can be translated into data mining goals. In addition, the more technical phases will summarize the analyst's options to achieve the goals set in the "Business understanding" phase, by recording the data mining problem types and the available techniques. Regarding the data, we will record the available data and the data requirements that are needed to feed the data mining models in order to achieve the business objectives.

3.1 Business understanding

Starting with the *Business Understanding* phase of the CRISP-DM methodology, we will confine our analysis at this point to the first and third task (determine business objectives and data mining goals), since there is no real case to confront so as to list the details. In addition, these tasks cover the main objective of this thesis and the scope of the chapter. In order to determine the business objectives, we identify specific business cases in the retailing area, based on marketing and retailing literature. Most of the business cases that follow are based on issues discussed in the retailing and marketing research community.

3.1.1 Customer segmentation

Customer Relationship Management (CRM) is a key focus area today in marketing departments in many different industries like finance, telecommunications, utilities, insurance and retailing. Businesses in these industries have changed or are changing their marketing focus from a product-centric view to a customer-centric view. There are several reasons for this change in focus: increased competition for non-growing markets, a technology revolution enabling the consolidation of corporate data and access to new data sources, and a growing awareness that the primary assets of a business are its customers [Cabena et al, 1999].

In managing their customers, businesses recognize that all customers are not created equal and that they should focus their marketing efforts on retaining their best customers, increasing the profitability of their high-potential customers, spending less marketing dollars on their low-potential customers, and acquiring new high-potential customers at a lower cost. A customer segmentation based on their key characteristics is central to CRM and is used to derive strategic marketing campaigns.

Customer segmentation is the essence of the customer-centric view of marketing. As the level of granularity of segmentation increases, marketing becomes more targeted and therefore more effective. In traditional retailing a very high level of granularity is of no use, since there is confinement in exploiting such kind of knowledge. A physical store cannot be 'personalized' and must fulfill the needs of the majority. In addition, targeted marketing is feasible only at a high level. On the contrary, in Internet retailing one-to-one marketing can be achieved and every kind of knowledge regarding the customers can be exploited.

The criteria to segment the customers are obtained from marketing and retailing literature, with respect to the customers' value to the stakeholder. The ultimate goal is to identify the different clusters of customers according to their value to the retailer, and characterize them by analyzing their shopping behavior and their demographics to adjust the marketing campaign in order to increase their total value. In the following, we list and describe the measures that founded the business case.

> The *ultimate goal* of customer segmentation is to identify the different clusters of customers according to their value to the retailer, and characterize them by analyzing their shopping behavior and their demographics to adjust the marketing campaign in order to increase their total value.

Criteria of segmentation

We seek for the adding-value criteria of segmentation to the marketing literature of the past three decades. We consider that the measures of marketing performance are a wise choice of selecting criteria, since all of them represent the stakeholders' value.

In [Clark, 1999] the history of measuring the performance of marketing is reviewed. The early work in the field focused on developing extensive profitability analysis. The target was to relate financial outputs to marketing inputs [Sevin, 1965][Goodman, 1970].

Profit is considered to be the first measure of performance in a firm. Since then, the term **profitability** has survived through the bibliography. Many different definitions of profitability have been proposed, and efforts have been made to model the profitability with the parameters that affect it. We believe that profitability is a key parameter to the customer-centric approach, since it answers the question "*who is the most profitable customer?*" which is of great value to the retailer. Further work borrowed other measures from microeconomics and finance literature like sales revenue [Feder, 1965] and cash flow [Day and Fahey, 1988]. In paragraph 3.2, where we list the data to be collected, profitability is described in detail.

The 1980's brought an expanded conception of output measures that included non-financial measures. **Market share** attracted tremendous attention as an output variable in this period. It was considered to be a strong predictor of profitability [Buzzell and Gale, 1987], but later on the market share-profitability relationship has proven to be both controversial and complicated [Szymanski et al., 1993]. One year later Ailawadi et al. showed that market share and market growth are not good predictors of the Advertising and Promotion/Sales ratio (A&P/S ratio) [Ailawadi et al., 1994]. Market share is a variable that is associated strongly with many crucial parameters and a great deal of research has been done to reveal these associations. The inability of the models is a challenge for data mining. We consider market share as a worth-searching variable to include in the segmenting criteria. Market share answers the question "*Which groups of customers contribute to the total market share?*" and is described in more detail in paragraph 3.2.

Understanding the process of **customer satisfaction** formation has been a concern of many marketing researchers and practitioners. Halstead et al. [1994] review the theoretical determinants of customer satisfaction, which are parts of the answer to the questions "*Who is the satisfied customer?*" and "*What satisfies the customer?*". The expectancy-disconfirmation model [Oliver, 1980] is the most known and proposed that customer satisfaction is a positive function of customer expectations (prepurchase beliefs about anticipated product performance) and disconfirmation beliefs (postpurchase beliefs about the extent to which product performance met expectations). Later modifications to the model have added new predictors like product performance, service quality, affective responses, causal attributes and equity judgments. Overtime, two variables are considered to be of central importance to customer satisfaction: product performance and disconfirmation beliefs. Customer satisfaction also depends on other factors that must be taken into consideration such as the product familiarity and the nature of the product. Customer satisfaction measurement is discussed in 3.2.

The problems of the customer satisfaction measurement led the researchers to **customer loyalty**, which attracted increasing attention as a measure of good marketing. It constitutes an underlying objective for strategic market planning [Kotler, 1984]. Although most marketing research on loyalty was focused on frequently purchased packaged goods (brand loyalty), the loyalty is also important for industrial goods (vendor loyalty), services (service loyalty) and retail establishments (store loyalty). In retailing and Internet retailing, we are interested in measuring and increasing **brand loyalty** and **store loyalty**. The brand loyalty literature contains a plethora of measures [Jacoby and Chestnut, 1978], which are based on repeated patronage. Dick and Basu [1994] proposed a conceptual framework to illuminate the cognitive, affective and conative antecedents of customer loyalty as well as its consequences. They identified four specific conditions of loyalty: *loyalty*, *latent loyalty*, *spurious loyalty*, *no loyalty*. These different conditions require different marketing strategy. In order to identify the condition of the customer loyalty, relative attitude and repeat patronage must be measured. For more details in measurement, refer to paragraph 3.2.

Apart from these measures, which have been examined thoroughly in the marketing literature the last thirty years, and capture the "value" of the customers to the stakeholder, we can use some other variables to assist segmentation; these variables are simple and can be created directly from data. We use combinations of the same variables with different time periods, average values, and ratios. Such variables are:

- Revenue
- Tenure
- Number of products purchased by the customer the last month
- Average number of products per visit the last month
- Average Revenue on Saturdays
- Number of visits the last month
- Average number of visits per month

These are just examples of the plethora of combinations we can create from combining time (at any level), revenue, profit, number of products, number of visits, tenure and so on.

Least but not last is all the demographic information that is available for the customers. **Demographics** should be included in the customer segmentation process in a second phase. Demographic segmentation on its own doesn't have high stakeholder value, because demographics are not used as basic criteria for defining the marketing policy (in comparison with the other criteria mentioned). In order to exploit a demographic segmentation, the marketing manager must be based on general theories about the behavior of different demographic clusters ("males aged 30-35", "teenagers", "married women aged 40-45" etc). This approach doesn't use the information from the retailer's customers and doesn't have a clear and direct correlation with a possible increase in sales or profit.

On the other hand, it would be very useful to know the demographic clusters within the segments that differentiate the "profitable customers", "potential customers" and "unprofitable customers". Customer demographic data may not typically correlate with customer profitability, but creating demographic segments allows the marketer to create relevant advertising, select the appropriate marketing channel and identify campaigns within the strategic customer segment that is defined in the first phase.

To clarify the previous syllogism, let us assume that a retailing organization has both a high-profit and a low profit behavioral customer segment that have similar demographic sub-segments. Further analysis of the purchasing behavior of the two sub-segments will indicate the marketing policy required to be applied in the demographic sub-segment of the low-profit customers to increase their profitability.

Segmentation mining model

The mining model that solves the customer segmentation problem is a simple clustering model that segments a population to several clusters with high intra-cluster similarity and low inter-cluster similarity. The clustering process requires defining the criteria of clustering and can be divided in more than one phases regarding the criteria chosen, with respect to the weights the stakeholders set.

In summary, **behavioral segmentation** helps deriving strategic marketing initiative by using the variables that determine the stakeholder's value, such as customer profitability. We must note here that the basic segmentation criteria can be combinations of all criteria proposed in this paragraph, and this depends on the perceived weight of the marketers for each variable. **Demographical segmentation** should be done in a second phase, since demographics and secondary variables (variables that were not used for primary segmentation) within the behavioral segments define tactical marketing campaigns. For a more detailed view of a segmentation mining model refer to the fictitious example of the paragraph 3.1.2.

3.1.2 Promotion design

Promotion and sales promotion are two terms widely examined from the marketing research community. One proof of the previous notion is that Promotion stands for the last P in McCarthy's 4P formula for the marketing mix [McCarthy, 1960]: Product, Price, Place and Promotion. The research has focused on two main axes, the effects of different types of promotions (promotion effectiveness) and the success factors of the promotional activity.

For the purposes of this thesis we settle with the following definition of sales promotion: *Sales promotion* is any temporary incentive to buy a product [Hardy, 1986]. Hardy [1986] summarizes the decisions that the manager must make to determine the promotion strategy. The manager must decide the **brands** that should receive promotional support, the **sizes** to support, **amount and type** of incentives, **timing** and the **length of the promotion period**. In these we add the **target group of customers**, since Internet retailing allows us to organize targeted marketing policies.

Promotions can be made either from the manufacturer or the retailer. The first case is more common and generally the objectives differ slightly. The success of the promotion varies with the objectives that are set and this is the primary task of the promotion manager. These objectives will be the measures of the actual performance of the promotion. Some examples of objectives are given in tables 3.1.1 and 3.1.2. We note that the objectives are not mutually exclusive and a promotion might target many of them.

The business case of *promotion design* regards assisting the manager to make the decisions about the brand and size of the products, the amount and type of incentives, the timing, the length of the promotional period and the target groups in order to achieve the objectives set.

For each of the decisions, different objectives form different mining cases. We tackle with each objective separately and describe the data mining goals that must be set in each case.

Which brand to promote?

Products can be described (and analyzed) in several levels. The lower level is the SKU of the product, which is unique for a specific brand of a specific size and a specific promotion. A whole taxonomy is built over the SKUs regarding the promotion, the size, the brand, the product category, the manufacturer etc.

We note that all the notions discussed in this paragraph can be expanded for the brand-size level (or any level of the taxonomy), in order to identify the size of the brand to be promoted. Kumar and Divakar [1999] pinpoint the differences between the two levels and the importance of analyzing in the brand-size level. The analysis remains the same, since the only thing that changes is the use of the SKU's that define the brand-size level, as long as it is conducted top-down.

✓ increase short-term retail volume	✓ add new store listings
✓ increase long-term market share	✓ add listing of new sizes
✓ build trade inventories	✓ heighten salesforce interest
✓ load the consumer	✓ defend against new brand entry
✓ preempt the competition	✓ defend against competitive promotions
✓ increase consumer trial	

Table 3.1.1 Manufacturer's promotion objectives

The manufacturer's main perspective is narrow in comparison to the retailer's. Promotion targets exclusively the sales increase of the manufacturer's products and sales decrease of the competitive products.

✓ increase short-term retail volume	✓ increase consumer trial
✓ increase long-term retail volume	✓ defend against new brand entry
✓ attract consumers to store	✓ defend against competitive promotions
✓ reduce inventories	

Table 3.1.2 Retailer's promotion objectives

On the other hand, the retailer is more interested in increasing the total volume of each store and this is the primary function of promotion [Doyle and Gidengil, 1977][Walters and MacKenzie. 1988]. This is the reason why the retailer will promote the products that will increase the total volume of the store and this is the main difference with the manufacturer's perspective. Nevertheless, in the case of private label products the retailer acts like a manufacturer and the objectives of table 3.1.1 stand for the retailer also. The main goal can be achieved through the following promotions, which are examined further.

We differentiate the mining cases with respect to each objective.

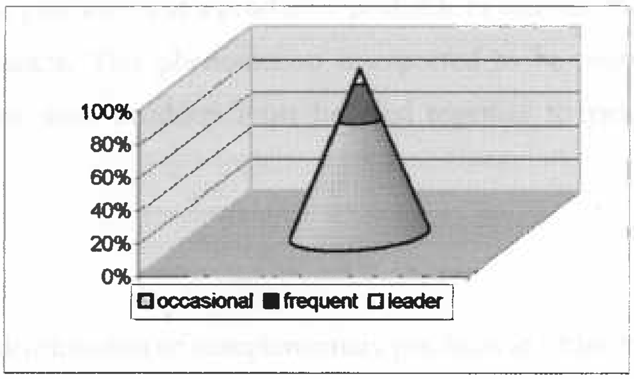
✓ *Promote products that attract customers to the store*

In retailing, there are some brands that can be considered popular. Their sales rate is very high in comparison with the other products. These products usually represent the 3-4% of the brands available and are referred to as *leader products*. Retailers use leader products as barkers to attract customers to the store. Leader products are promoted through gross discounts and constitute the retailer's flyers.

Since the leader products are the products that are found in most of the carts, customers will be attracted and visit the store (site) to purchase them. Of course this means that these customers will buy the rest of the products they are interested in from the same store; the rest of the products are the ones that cover the loss from the leader products (which in many cases are sold under their cost) and give profit to the retailer.

Mining model

The business goal is to identify the leader products. The data mining goal is to segment into three clusters the 1-itemsets. More specifically, we measure the purchasing frequency of each product, which is the support of the 1-itemsets. This is achieved easily either by conducting the first phase of the extraction of association rules (for the first loop only) or by advanced SQL queries to the database for each distinct product. Then we segment the supports of all products into three clusters. The segmentation of the supports is a simple two-dimensional clustering model with axes the products and their supports. The clustering can be even done manually by observing the support distribution.



✓ *Promote products with high margin profit*

There are two kinds of products with high-margin profit: the ones that were bought in low cost (due to a good deal with the manufacturer etc) and the private label products. Regarding private label products, the retailer can be considered as a manufacturer. This explains the existence of the three last objectives, which refer to brands, and competition comes from substitutional products and not from alternative retailers.

These products are easily identified and this task does not require advanced data analysis method. However, the decision for the other parameters of the promotion (type, timing, length of period, target group) is very crucial in the promotion of high-margin profits. These decisions can be assisted by data mining and are described in the analysis of the rest of the objectives of the current paragraph.

✓ *Promote products to increase complementary purchases*

A basic notion in retailing is that promotions also affect consumer purchasing patterns by stimulating purchases of non-promoted complements to the promoted products [Walters, 1991][Walters, 1988]. Product complements are products that are used in conjunction with another to satisfy a particular need [Henderson and Quandt, 1958]. In this case we will generalize the concept of complementary products, as products that have strong correlation because of their simultaneous occurrence in many carts. We hypothecate that this correlation implies a logical reason, which can be examined after the extraction of the correlations. The promotion of a product is probable to increase the sales of all the complementary products. This phenomenon is expected to be more intense for strict complements, since such products must be used together to yield another distinct product.

Mining model

The business case requires the identification of complementary products in order to decide which of them will be promoted, with target to increase the sales of their complements. The mining case that arises is the extraction of the large itemsets. We discuss a little further the case because the results might be misleading. It is possible that there are products that appear in the majority of the carts. Such products are the leader products, which where discussed earlier. Products with very high support can be misleading to the identification of the complementary products, since they will appear as complements to many products. These products should be excluded from the analysis.

Strict complementary products can be identified more confidently by the following heuristic: From the large itemsets extracted, search for products with supports near the support of the itemset that contains them. These products are bought together (from the large itemset) and are bought alone in little occasions (support comparison). The general heuristic for complements is to search for products that appear in large itemsets. From those it would be advisable to promote the product that appears in the left side of the rules with high confidence, since such rules imply that there is high probability to buy the rest of the products (the ones on the right side).

In this case data mining offers a more stable way of recognizing the complementary products, in comparison with the empirical rules that were used in research so far [Walters, 1991]. We must note that the knowledge of the expert is not eliminated; data mining simply shortens the space to be examined from the whole product hierarchy to a set of rules (or itemsets).

To which customers to promote each product?

The second, but probably the most crucial, decision to be made is to identify the target customers for each product. As we will see, the target customers depend on the objective of the promotion. We describe the mining model that answers this question based on a fictitious customer database to illustrate the results.

Mining model

We believe that there is need to segment the customers based on their brand loyalty. This segmentation should indicate at least the customers that are loyal to the product to be promoted, the ones that buy it occasionally and the ones that buy it rarely. Intermediate clusters are acceptable. The segmentation does not need to be based only on the brand loyalty, but other characteristics can be used as well (see paragraph 3.2), such as profitability. Another characteristic that might be useful in identifying the customers who will react positively to the promotion is the general attitude of the consumers to promotions or more specifically to promotions of products to the specified category. Such information will increase the confidence of labeling the potential customers, since occasional purchasing may be caused on several other reasons than proneness to promotional activity. Additionally, we need a segmentation of the customers based on their brand loyalty on products that have strong correlation with the product to be promoted. These criteria will identify customers that are loyal to all products, loyal to the one to be promoted but not to its complementary and reverse, etc. These possible clusters are illustrated in figure 3.1.1 for a fictitious example with three complementary products, one with high profit margin, one overstocked and one indifferent. The one with the high-profit margin was used in the clustering process. Based on the notion of loyalty, we characterize the clusters.

Then, we build a predictive model using as target attribute the label of each cluster based on demographic characteristics. We can have the same result by segmenting again each of the clusters based on the demographics characteristics (figure 3.1.1). The predictive model though is more credible for such a task, due to the nature of the algorithms on such data (more in paragraph 3.4). The example of figure 3.1.1 uses demographic clusters for illustrating reasons.

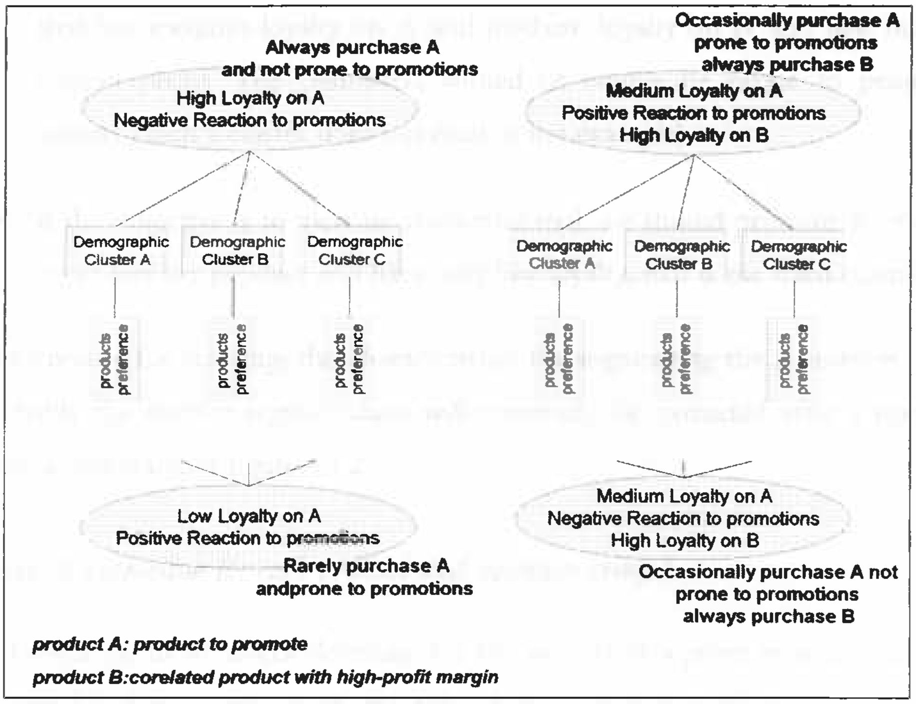


Figure 3.1.1. An example of clustering for determining target customers

First of all, we tag and comment each one of the clusters. The *first cluster* that always purchases the product A and is not prone to promotions will be excluded from the target group, since they are loyal to the product to be promoted. This cluster will be used as a guide to recognize the potential customers from the other clusters.

The *second cluster* that occasionally purchases A, is prone to promotions and always purchases B, is a cluster that contains potential customers to purchase A. The *fourth cluster* that occasionally purchases A, always purchases B but is not prone to promotions consists of customers that perhaps have other reasons for choosing between product A and a substitutional and are not affected by promotions. The *third cluster* consists of consumers that either purchase product A rarely or not at all.

Having the above clusters as base knowledge, the decision depends on the objective set. For example

- ✓ if the objective is to increase the short-term volume, we should target the second target group. To limit the size of the target group even more, we can search for those customers that have the same demographic characteristics as the ones of the first cluster or purchase other correlated products with product A.

- ✓ if the objective is to increase the short-term profit, we should target a cluster that has medium loyalty on A and medium loyalty on B, and that has a high margin profit. The customers should of course be prone to promotional activity (such a cluster does not exist in the example).

- ✓ if the objective is to increase consumer trial, we should promote to those that don't buy the product and have very low loyalty, that is the third cluster

The choices for selecting the characteristics for segmenting the customers are a lot and probably the correct segmentation will eventually be extracted after a number of iterations, as illustrated if figure 2.1.2.

What type of promotion for each product and customer group?

Although the most crucial decisions for the success of a promotion are the choices for the brand and the target group, the type of promotion that will be selected can play an important role for achieving the target set. The type of promotion regards the means that will be used to attract the customers to purchase the promoted product. The previous analysis has pointed out the customers that are probably interested in the product, but may need an additional incentive to make up the final decision. The customers, apart from their purchasing habits, may favorite specific types of promotions.

More specifically, some of the most known types of promotions are:

- Price discount
- Coupon discount
- Additional product quantity
- Bundled products
- Price Quantity discounts
- Gifts with purchasing

- In-store promotion
- Advertising

Many research approaches suggest that the promotion preferences characterize the customers. For example price-sensitive customers are more keen on price discounts [Mulhern and Padgett, 1995], whilst coupon-prone households are more educated, urban, less brand loyal and less store loyal [Bawa and Shoemaker, 1987]. In this area of research the more important issue is the consistency in the customers' purchase behavior across the different types of promotions and across the product classes, especially when the promotion-prone consumers are in most cases of low loyalty and stability to their decisions.

The previous observation indicates the need for identifying patterns of the priorities the customers set through their decisions, since no reliable model is applicable on specific cases. The business case sets the problem of associating consumer demographic profiles (the ones that characterize the target groups) with the preferences on the promotion types.

Mining model

We propose the following mining model as a solution to the above business case: the extraction of profile association rules [Aggarwal et al., 1998], considering as products (items) the promotion types of the products. The success of such a model premises the abstraction of noise like products with no promotion, which will probably bias the results. It would be advisable to use products of the same or similar products class, due to the inconsistency of behavior across the product classes.

In the case that the above model becomes too complicated in attaining reliable results, a simple voting schema could be used alternatively. From a research perspective the two models could give interesting comparative results.

Finally we must note that the decision on the selection of the means of promotion may be biased from the objective of the promotion. For example, price discounts promotions are used for protection from competitive brands [Wilcox et al., 1987], whilst bundled products are used to promote new products.

What is the correct Timing for promoting a specific brand to a target group?

Timing is a critical issue not so much for the success of the promotion itself, but for setting the objectives of the promotion. For example, if there is a prediction that the sales of a specific brand or product class will fall next month, this immediately requires promotion to prevent the sales reduction. However these causes must be checked from several perspectives (promotions can prevent sales reduction due to brand switching but not due to seasonal changes). Historical data can be used to predict sales and profit of the near future. These historical data should include, apart from aggregated purchase data, external sources of data, like promotions of competitive products and other external parameters that cause interstore sales displacement [Walters, 1991]. Using forecasting for planning periodic promotions is an issue that has received some attention by the research community [Mcintyre and Achabal, 1993].

Mining model

The mining model that arises is a simple predictive model that forecasts future parameters, which are the basis of setting the objectives, discussed earlier in the paragraph. Such parameters are the short-term retail volume, the long-term market share, the product demand, the product sales etc. The fact that timing participates in determining the objectives, led us to put it first in the proposed model for promotion design (figure 3.1.2).

What should be the Length of a specific promotion on a target group?

The length of promotion in traditional retailing is associated with achieving the objective set. However the length of promotion in Internet retailing has a new meaning, due to the additional information that is monitored for each customer. We focus on Internet retailing, where the case is interesting in being facilitated through data mining techniques.

The concept of the business case is dual: to stop the promotion when it succeeds and to change it when it is ignored. If, for example, the customer reacts positively to a price quantity promotion of a product that has a specific product cycle for the specific household, then the promotion should stop for the respective time, until the products are consumed. This business case implies a high level of personalization, which has no direct effects, but may have indirect to the customer satisfaction after the purchasing experience. The business case strengthens when monitoring is more detailed through smart homes, where the product life cycle for each household is measured with accuracy.

The particular problem is similar to the problem of dynamic recommendations [Prassas et al., 2001], which uses as feedback the reactions of the customers to the recommendations in order to determine future recommendations. The solution is based on a weighting schema that strengthens the weights of the recommendations when the customer appraises them positively and weakens them when they are ignored. The subject is still open for research.

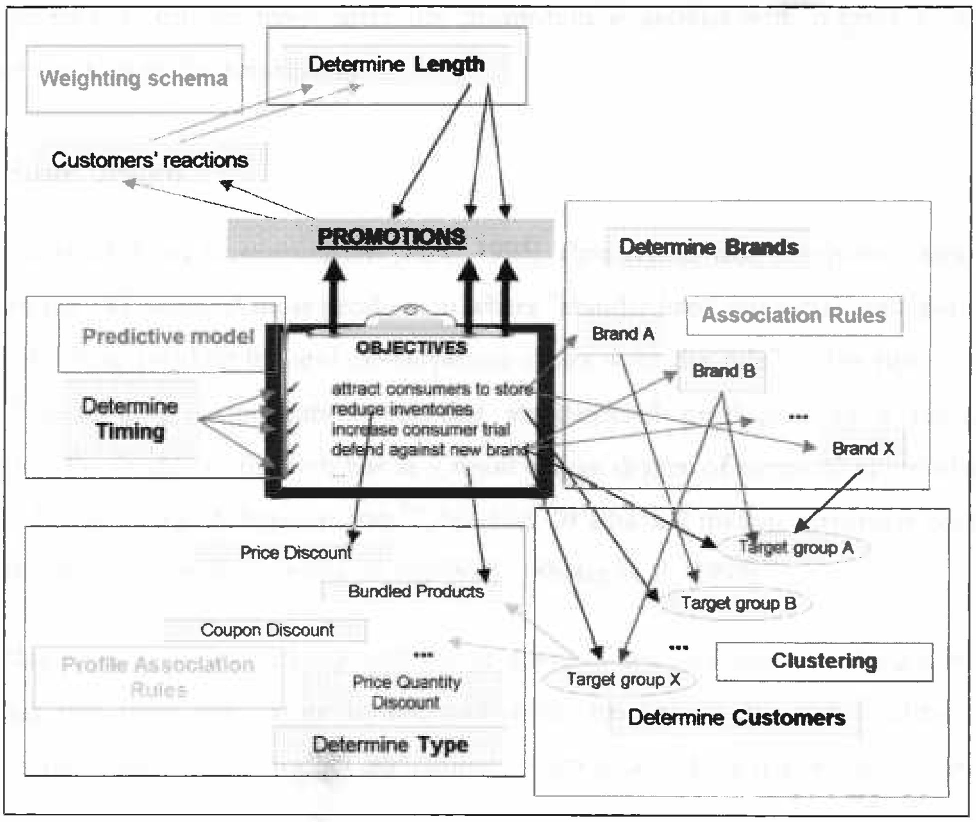


Figure 3.1.2. The Promotion Design Model facilitated by data mining

We illustrate the proposed model for the promotion design in figure 3.1.2. In each phase we pinpoint the data mining models that facilitates it. The primary task of the promotion manager is to set the objectives of the promotion campaign which are illustrated in the center of figure 3.1.2. Around the objectives five decisions must be made which are illustrated with the "determine" boxes: brands, customers, type, timing, length.

As we can see, Timing is the only decision that precedes setting the objectives as discussed earlier. This is represented with the red arrows that indicate a predictive model that predicts the parameters that form the objectives (volume, market share, sales etc.). The rest of the decisions depend on the objectives. The coloured arrows are accompanied by the data mining techniques that facilitate the extraction of the associations, as described in the paragraph. As illustrated and discussed, association rules can facilitate the determination of the brands to be promoted and the type of promotion to be used for each brand. Regarding the customers clustering is used. The black arrows represent associations that exist and are set a priori by the promotion manager; some of these associations were given as examples previously. The last decision about the length of the promotion can be made after the promotion is applied with respect to each customer reaction to the promotion.

3.1.3 Store design

In his book Mass Customization [Pine, 1993], Pine argues that companies need to shift from the old world of mass production where "standardized products, homogenous markets and long product life and development cycles were the rule" to the new world where "variety and customization supplant standardized products". In a broader perspective, Pine's idea to the web has as a result a large degree of personalization of the sites. Jeff Bezos, CEO of Amazon.com™, has said "If I have 2 million customers on the Web, I should have 2 million stores on the Web" [Shafer et al., 1999].

In this section, we are tackling with the problem of customizing store design. Store design has two main axes, store layout, and store atmosphere. We will confine our analysis to the category and product assortment, which is one of the major issues of store design.

We consider our business case of the problem of category assortment independent of the store layout that is used. There are consistent findings that in many situations consumers construct their preferences when faced with a specific purchase decision, rather than retrieve preformed evaluations of product features and alternatives [Bettman et al., 1998]. This is the reason why consumers are likely to evaluate the attractiveness of a product relative to the other options that are considered simultaneously. It is argued that the probability that a consumer will make a purchase is enhanced if the considered assortment is designed such that one option is clearly superior to another; additionally, the choice probability of a particular target option can be enhanced by adding an option that is clearly inferior relative to the target option, but not relative to other options in the considered product assortment [Simonson, 1999].

Category assortment precedes product assortment in the store design. Category managers firstly design the category hierarchy of the store and assign each product to a specific path of the hierarchy. Based on the hierarchy the assortment is designed. Product and category assortment are determined based on the empirical perception of category managers of associated products and categories. Products that belong to associated categories are placed in a near position, either on the same shelf, or on near shelves in the same aisle. Everyday products like dairy products are usually placed at the back end of the store to force the customer to walk through as many aisles as possible. Products that trigger impulse purchasing are placed in visible points, in stands near places were customers stand in queues like the checkout or the in-stores (delicatessen, butcher's shop). These examples of rules are the outcome of empirical research and hypotheses; there is no feedback from the shopping trips of the customers.

In Internet retailing the purchase sequence becomes more interesting since there is explicit feedback about the customers' shopping trips. We record the shopping trips of the consumers and try to capture the logic that underlies in sequential purchasing. Undoubtedly, sequential purchasing will be biased from the current category and product assortment, but the extraction of a pattern that always avoids certain aisles and repeats certain "paths" is knowledge that can be used to customize the store design for each customer separately.

At this point data gathering about the navigation of the customer through the aisles and the sequence of the products that are added in the cart is only possible in Internet Retailing. In traditional retailing, with the use of smart carts that scan each product, such recording will be feasible. Nevertheless, such knowledge cannot be used to a great extent, since the store design is unique and customization cannot be applied.

On the other hand, in Internet retailing such knowledge can be used to customize the category assortment to each customer's preferences. Such personalization limits the time consumed for searching, increases the purchasing probability and increases customer satisfaction of the purchase experience. Additionally, this knowledge can be exploited for up-selling purposes. For example, if a customer usually visits two aisles in sequence, the site could interpose between them an up-selling category of products. Such rules that will dynamically change from the feedback of reactions will eventually create a unique shop for each user.

Mining model

The problem stated in this section is a variation of the problem of mining sequential patterns [Agrawal and Srikant, 1995]. We consider two data collections: the one records the sequence of categories that the customers browse in each of their visits to the store, and the other one records the sequence of categories that they buy for the same visits. By mining these two data collections separately for each customer we can extract the usual shopping trips of the customers and their sequential purchasing patterns. Their shopping trips reveal their preferences (which products they are interested in evaluating for purchase). Their sequential purchase patterns reveal the preferred category assortment from the customer's perspective. A comparison with the shopping trips will reveal up-selling opportunities.

Combining the previous knowledge with the product life cycles, which can be extracted from the purchasing frequency, the shopping trips could be segmented into groups, regarding the shopping intentions. More specifically, most customers have period-dependent shopping lists, such as monthly shopping lists for stocking the household with non-everyday products, daily shopping lists for satisfying the daily needs etc. By tagging the shopping trips, the personalization level can become higher if the system could predict the shopping intentions of the customer. For example, if one month has already passed since the use of the monthly shopping list or the purchase of products with a monthly product life cycle, the system could adapt the store category assortment to the shopping trips that contain the categories of a monthly purchase.

3.2 Data understanding

Our scope in the data understanding phase is to describe in short the data that is needed to be fed in the models that were described in the previous phase. Beyond the description, which depicts the acquiring process, we comment on the critical points of each process. The data listed in the following paragraphs are not primary data (row data), but are derived from the primary data; these data is either secondary or aggregated primary data.

3.2.1 Profitability

Profitability is specified as ROI (return on investment), ROA (return on assessment), ROC (return on capital), ROS (return on sales) or ROE (return on equity) in the marketing bibliography [Szymanski et al., 1993]. To illustrate:

ROI = profits/investment = (profits/sales)×(sales/investment) = ROS × asset turnover ratio = ROS × (1/[investment/sales]) = ROS × (1/investment intensity)	ROE = profits/shareholder equity =(profits/[debt+equity])×([debt+equity]/equity) = (profits/total assets) × (total assets/equity) = ROI × financial leverage

Basically, ROS is one component of ROI. The measures are equal only when the investment to sales ratio equals one, which is rarely the case. The ROI illustrated regards the ROI of the firm, and therefore the *profits* and the *investment* refer to the total profits and total investment. However, when dealing with the business case of customer segmentation, the *profits* and *investment* must refer to each customer profits and investment.

If the marketing policy is not targeted and the case is applied for the first time, atomic investment has no meaning. So customer profitability can be calculated either by each customer's ROS, which is the customer's profits to customer's sales ratio, or by the ROI, where the customer's investment is the total investment divided by the number of customers (hypothesis for even distribution of investment on customers). In the case of targeted marketing already applied, there must be a previous segmentation of customers. Then, the customer's investment should be calculated as the ratio of total investment of the group to the number of members of the group. In Internet retailing, one-to-one marketing is feasible and atomic investment has a substance; it can be calculated as the sum of every dollar spent to attract the specific customer (special discounts, bundled offers etc).

3.2.2 Market share

There are two different measures that contribute to the variance in market share elasticities [Szymanski et al., 1993]. *Absolute market share* is the ratio of business' sales to total sales in the served market and *relative market share* is the ratio of a business' market share to the combined market share of its three largest competitors. Absolute measures of market share are preferred when specific industries are studied, because the sum constraint (the market shares of individual firms should sum to 100%) and the bound constraint (market shares of individual firms should be between zero and 100%) can be satisfied. Relative market share, on the other and, is preferred when cross-sectional data is pooled across industries, because the sum constraint and bound constraint cannot be satisfied. Further distinction could be made between absolute market share measures, based on unit sales and dollar sales.

In order to have market share as a cluster characteristic, when segmenting customers, in the clustering process the unit/dollar sales per customer must be included in the input data. The process should be done only for specific products and not for all because it would probably lead to a failure. The market share to be measured should regard either a specific product or a specific product family with strongly correlated products. A repeated procedure for different products would lead to extracting which products' high market share is associated with the other measures (for example profitability).

3.2.3 Customer satisfaction

Customer satisfaction's measurement requires explicit data input. Firstly, the attributes that determine the customer satisfaction must be identified. As it was mentioned in paragraph 3.1, customer satisfaction depends on many variables. These variables must be transformed into questions of a questionnaire [Rust and Zahoric, 1993][Halstead et al., 1994]. Each question will ask the customer to rate the variable involved on a pre-specified scale. This scale can be either numerical [Anderson and Sullivan, 1993] or semantic (e.g. very good, good, not strong etc) [Halstead et al., 1994]. In both cases satisfaction is defined as a function of the rates of the variables.

In the retail case we identify two cases of customer satisfaction that should be treated differently as proposed by Churchill and Surpenant [1982] and Anderson et al. [1997]: the satisfaction from the retail service and the products' satisfaction. The first one has as main stakeholder the retailer and secondary the supplier, whilst the second one reversely. A complete questionnaire that covers customer satisfaction in the area of retailing is out of the scope of this thesis; we will confine in highlighting the most important variables in our opinion, some of which have been used in the customer satisfaction measurement in other areas. Our suggestion is shown in the following sample questionnaire. The last column indicates whether the question regards traditional retailing (T), Internet retailing (I) or both (T/I).

Variable measured	Question	Range [anchored from ... to]	Apply in
Performance	Rate the friendliness of the employees	not friendly *to* friendly	T
Performance	Do you find all the products you are interested in?	all *to* none	T/I
Performance	Do you find easily the products of your interest?	always *to* never	T/I
Performance	How expensive do you find our products in general?	very cheap *to* very expensive	T/I
Disconfirmation	My shopping experience is	very bad *to* very good	T/I
Satisfaction	I like shopping in your store	strongly agree *to* strongly disagree	T/I
Repurchase intentions	I will shop from your store again	strongly agree *to* strongly disagree	T/I

Table 3.2.1 Demo questionnaire for measuring customer satisfaction

3.2.4 Customer loyalty

Customer loyalty value is an outcome of two dimensions: Repeat Patronage and Relative Attitude [Dick and Basu, 1994]. *Repeat Patronage* refers to the repeated purchase and *Relative Attitude* refers to the disposition of the customer towards the entity (brand, store, service etc). The main argument is that multiple purchase occasions might not have as cause the comparable attitudinal extremity towards other brands, and choice is based on situational factors like shelf positioning for example. Low relative attitude with high repeat patronage indicates "spurious loyalty", which must be treated differently from "loyalty". The cross-classification of the to dimensions is summarized in the following table.

		Repeat Patronage	
		High	Low
Relative Attitude	High	Loyalty	Latent Loyalty
	Low	Spurious Loyalty	No Loyalty

Table 3.2.2. Customer Loyalty values

Repeat patronage is a quantitative measure and until the conceptual framework of Dick and Basu [1994] it has attracted most of the attention of the research community. There are many different measures for measuring repurchasing; Jacoby and Chestnut [1978] review 53 measures. Here we discuss the BLOY variable [Guadagni and Little, 1983], which was designed to increase if a brand is recently purchased often and decrease if it is not recently purchased often, and is argued by Gedenk and Neslin [1999] to capture purchase event feedback; purchase event feedback is the effect of current purchases on future brand preference and is concerned with what consumers learn from the consumption experience.

The BLOY is computed as

$$BLOY_{iht} = \lambda \cdot BLOY_{ih,t-1} + (1-\lambda) \cdot PUR_{ih,t-1}$$

where

$BLOY_{iht}$ = Brand loyalty of household h for brand i at purchase occasion t
λ = Smoothing parameter
PUR_{iht} = Purchase indicator (1 if household h purchases brand i at purchase occasion t, 0 otherwise)

The first part of the equation describes the decay or carryover of BLOY from one purchase to the next. If a brand is not purchased, BLOY decreases by this decay factor. The second part of the equation indicates the size of the build-up of BLOY allocated to the brand purchased. Across brands BLOY sums to one. As a result, the purchase of a particular brand affects BLOY for that brand as well as for others. Therefore, BLOY can be interpreted as a relative measure. This makes cense because an increase in preference of one brand would naturally result in a decrease in preference for other brands.

Relative attitude depends on attitude strength and attitudinal differentiation [Dick and Basu, 1994]. *Attitude strength* refers to the positive or negative attitude of the customer towards the entity. *Attitudinal differentiation* refers to the positive or negative attitude of the customer towards the entity, on comparison with an alternative one. Strong attitude strength is expected to associate with high loyalty, but in rare case customers may patronize entities for which they have negative attitudes, because of a temporal situational exigency (e.g. luck of funds to afford a liked alternative, when consumption in the product class is mandatory). On the other hand, the level of perceived differentiation can reveal the existence of other situational factors. For example, this may be the case for consumers who have equally positive attitudes toward Coke® and Pepsi® with choice between them on shelf positioning. Cross-classifying two levels for each dimension lead to consider four conditions:

		Attitudinal Differentiation	
		No	Yes
Attitude Strength	Strong	Low Relative Attitude	Highest Relative Attitude
	Weak	Lowest Relative Attitude	High Relative Attitude

Table 3.2.3. Relative Attitude values

Attitude strength can be measured only through explicitly asking the customers (rates in x-scale). But as we can see from table 3.2.3 the high/low value of relative attitude depends only on attitudinal differentiation, which can be assessed by testing reactions on several situational factors. Such factors are: stock-outs of preferred brands (cannibalization), reduced prices and deals of competing brands, effective in-store promotions, shelf positioning (product assortment at an e-store). These factors cannot be extracted directly from transactional data.

Because of this difficulty, we propose measuring customer loyalty only by repeat patronage in the case of customer segmentation. In a second phase for each segment we should seek for the underlying factors that differentiate a "very high loyalty" customer from a "high loyalty" one. Such a further search has value for the Internet retailing environment, because the factors that affect the attitude strength can more easily be detected, due to the nature of the electronic environment.

3.2.5 Demographics

We consider as demographic information any information that directly characterizes the customer. In this paragraph we present a classification of some demographichs characteristics, which is a subset of the Consumer Profile Data Component, one of the Data Components of the UML analysis for the iMEDIA [Cyberce and ICOM, 2000]. Additional comments clarify the use of the data for the business cases described in chapter 2.

We note that the ways to obtain this information is out of the scope of this thesis. Additionally, we will not tackle with the issues regarding sensitivity and confidentiality of data. In this section we list some of the demographic characteristics, which we believe may be useful in a customer-centric marketing analysis.

Demographics Class

Demographics Class is used to represent the demographic information of the customer. The attributes of the Demographics Class are:

Demographics
✓ **Name**: The name of the customer.
✓ **Surname**: The surname of the customer.
✓ **Gender**: The gender of the customer.
✓ **Email**: The e-mail address of the customer.
✓ **Date Of Birth**: The day of birth of the customer
✓ **Income**: The income of the customer

Occupation Class

The Occupation class represents the occupation of the customer.

Occupation
✓ **Occupation**: The name of the occupation.
✓ **Occupation Description**: A short description of the occupation

Occupation Category

The occupation class is associated with the OccupationCategory class, which contains the following attributes:

Occupation Category
✓ **Occupation Category**: The name of the category
✓ **Occupation Category Description**: A description of the category.

The hierarchy of the occupation categories can be built separately, without the customers' effort.

Education

The Education class represents the education of the customer.

Education
✓ **Education Name**: The name of the education
✓ **Education Description**: A description of the education

Religion

The Religion class represents the religion of the customer.

Religion
✓ **Religion Name**: The name of the religion
✓ **Religion Description**: A description of the religion

Race

The Race class represents the race of the customer.

Race
✓ **Race Name**: The name of the race
✓ **Race Description**: A description of the race

Nationality

The Nationality class represents the nationality of the customer.

Nationality
✓ **Nationality Name**: The name of the nationality
✓ **Nationality Description**: A description of the nationality

Marital Status

The Marital Status class represents the marital status of the customer.

Marital Status
✓ **Marital Status**: The name of the marital status
✓ **Marital Status Description**: A description of the marital status

Age_Range Class

The Age_Range class indicates the age range that the customer belongs.

Age
✓ **Description**: brief description for the age range.
✓ **Age Range**: The actual age range.

Knowing the exact age of each consumer (from the date of birth), data mining algorithms can segment Age, regarding with the customers. However, a pre-segmentation can be used if the marketer has a-priori knowledge of it.

Income_Range

The Income_Range class represents the income range of the customer.

Income
✓ **Description**: brief description for the income range
✓ **Income Range**: the actual income range
✓ **Currency**: the currency of the income

Knowing the exact income of each consumer, data mining algorithms can segment Income, regarding with the customers. However, a pre-segmentation can be used if the marketer has a-priori knowledge of it.

Household Information Class

Household Information class is associated with the customer's profile and holds information about the household of the customer. Most information cannot be directly used, but should be stored for probable future use, by combining with other data sources. The attributes of the Household Information Class are:

Household
✓ **Num Of Children**: The number of children in the household .
✓ **Phone**: The telephone number of the household.
✓ **Connection Phone**: The telephone number for the Internet connection
✓ **Username**: The username that is used to have access to the Internet
✓ **Number Of Members**: The number of members in the household.
✓ **Occupancy ownership**: Indicates whether the user owns the house or s/he rents it.
✓ **Square Meters of House**: The square meters of the house
✓ **Number Of Vehicles**: The number of vehicles owned/used by household
✓ **Income**: The total income in the house

Vehicles Class

The Vehicle class holds information for the vehicles that a household/customer may have.

Vehicle
✓ **Description**: brief description about the vehicles.
✓ **Car ownership**: the car's proprietary indicator.
✓ **Car Price Range**: Range of the car's price.
✓ **Motorcycle ownership**: the motorcycle's proprietary indicator
✓ **Motrocycle Price Range**: Range of the motorcycle's price range.

Child Class

The Child class holds information about the children in the family.

Child
✓ **Residence**: Indicates the existence of children in the house.
✓ **Date Of Birth**: The date of birth of the child.
✓ **Gender**: The gender of the child

3.2.6 Navigation data

Navigation information can be recorded only in Internet Retailing or other electronic environment, where the customer interaction with the system can be monitored, e.g. interactive television. Even when advanced technology will be used in traditional retailing, like smart carts and smart shelves, the navigation behavior that can be recorded in the Internet is more detailed. The on-line customer browses through the hundreds of pages, devotes different time in each, reacts to promotions/recommendations/advertisements and generally interacts with the store. The mouse is the eye and the hand of the online customer. If something attracts the attention the mouse usually moves over it (eye movement); when the customer decides for further, more thorough examination, he/she clicks on the product (hand). These reactions are the footprint of their purchase logic that takes place and determine the decision between the available options.

We summarize in the following table the navigation behavior that can be recorded in an online store.

Browsing preferences
✓ sequence of products added in the cart ✓ sequence of categories added in the cart ✓ sequence of products browsed independently of purchase ✓ sequence of categories browsed independently of purchase ✓ products that are contained in shopping lists ✓ use of browsing means per product/category
Purchase decision
✓ time spent in reading product specifications ✓ product assortment of the category for the products that were purchased ✓ time required for deciding which product to add to cart ✓ the products clicked in a product assortment ✓ the products that the mouse rests for the first time ✓ movement of mouse above products or prices ✓ products that were removed from the cart ✓ products that their quantity was updated during the decision process
Reaction to promotions
✓ existence of promoted products in a product assortment ✓ ignorance of promotions ✓ interest in promotions independently of purchase (clicks) ✓ movement of mouse over promotions or banners ✓ time the mouse rests over a promotion
General preferences
✓ total time spent in the store ✓ timestamp of purchases

Table 3.2.4Navigation data to be recorded

There are two major problems in recording navigation information: the majority of the pages visited are dynamic and the information to be recorded is vast. The first problem does not allow saving the referenced product assortment or the promotions/ads that exist in a page and the second is infeasible due to the transactions with the database. The current bandwidths and transmission rates used for the connection to the Internet don't allow the record of information like mouse movement.

The solution to the first problem is the creation of XML files. For every page that is created dynamically, an XML file can be created simultaneously that holds the product assortment, or any other information like promotions, recommendations, banners, position of them, colors, page design. XML is a technology that is already used as an intermediary for the creation of dynamic pages to avoid transactions with the database.

In this way for every session the system can treat the site as consisting of static pages (the XML files). The second level of recording regards the sequence of browsing (clicking) between these pages with their timestamps. The timestamps can be used a posteriori for generating the time periods spent on each page. The sequence of clicks can be recorded as a list of pointers to the XML files that represent the dynamic pages.

Capturing the movement events of the mouse may be inhibitory to the tolerance of the overhead in transmitting the net information. For the time being such recording is impossible, but the evolution in telecommunications will soon solve the transmission rates limitations. Nevertheless, the technology that would be used is client site programming in order to eliminate the noise and post back to the server only the useful information.

Saving information using XML provides a solid and structured way of storing this kind of information that is very randomly generated. This process facilitates the data mining process that can be customized to use as input source the XML files and does not require database storage.

Capturing the navigation information in such detail is probably the only way of having the only reliable source of data that can reveal the purchasing process decision.

3.2.7 Purchase data

Purchase data refer to storing all the transactions per customer. In traditional retailing the primary purchase data is the POS scanning data; in Internet retailing the primary purchase data is the data from the post order form. This data in their primary form is not mineable. Even for the market basket analysis data preparation is required to lead to quality results.

The purchase data are aggregated in order to represent summarized information that is value-adding in comparison to the primary data. Examples of additional information that characterizes the baskets and the customers' purchases are:

- Revenue
- Number of products purchased by the customer the last month
- Average number of products per visit the last month
- Average Revenue on Saturdays
- Number of visits the last month
- Average number of visits per month

This information can be combined with time periods and can create a large number of variables that can participate in segmentation problems.

3.3 Data preparation

For the data miner, all objects consist of measurements of features. It is the group of features that are taken as the defining characteristics of the objects, and actual instance measurements of the values of the features are considered to represent instances of the object. Measurements are described as consisting of two components: the actual absolute perfect value, plus distortion. The distortion is often referred to as error. All kinds of errors are considered 'noise' to the data mining process and must be abstracted. In this section, we will summarize the issues of building mineable data representations [Pyle, 1999].

Data in the real world is dirty:

- it is *incomplete*: it usually lacks attribute values, certain attributes of interest, or contains only aggregate data
- it is *noisy*: it is common to contain errors or outliers
- it is *inconsistent*: it may contain discrepancies in codes or names

The indubitable fact is that in order to mine quality results, quality data is needed. Any bias in the data may lead to unreliable and misleading knowledge.

A well-accepted multidimensional view of quality data is shown in Table 3.3.1. Quality data should be accurate (valid values), complete (no missing values), consistent (valid physical and conceptual constraints), have timeliness (up to date values), believable, value adding, interpretable (comprehensible values) and accessible. The first four characteristics (accuracy, completeness, consistency and timeliness) ensure the reliability of the results and the second four (believability, value adding, interpretability and accessibility) facilitate the deployment of the extracted knowledge.

✓ Accuracy
✓ Completeness
✓ Consistency
✓ Timeliness
✓ Believability
✓ Value added
✓ Interpretability
✓ Accessibility

Table 3.3.1

We now summarize the major tasks in data preprocessing:

3.3.1 Data cleaning

> *Data cleaning* is the task of filling in missing values, smoothing noisy data, identifying and removing outliers, and resolving inconsistencies.

Filling in missing values targets the completeness of the data. Data is not always available since many tuples have no recorded value for several attributes. The missing data may due to several reasons like equipment malfunction, inconsistency with other recorded data and thus deleted, not entered data due to misunderstanding, perception of unimportance at the time of entry. At most cases missing data need to be inferred [Pyle, 1999].

There are many ways in handling missing data, most of them heuristic rules. One first solution is ignoring the tuples with the missing values. This rule is usually used when the class label is missing; it is not effective though, when the percentage of missing values per attribute varies considerably. The other solution is filling in the missing values. The most common heuristics are using a global constant to fill in the missing value, using the attribute mean to fill in the missing value, using the attribute mean for all samples belonging to the same class to fill in the missing value, or using the most probable value to fill in the missing value (Bayesian formula or decision tree).

Noisy data regard the errors or variances in a measured variable. *Smoothing noisy data* targets accuracy of the data. Incorrect attribute values may due to several reasons like faulty data collection instruments, data entry problems, data transmission problems, technology limitation, and inconsistency in naming convention. Similar problems are duplicate records, incomplete data and inconsistent data. *Outliers* are another significant problem that bias the data and disorientates the analysis.

The main methods that are used to tackle with these kind of problems are binning methods that smooth the data by the bin characteristics (median, boundaries etc), clustering methods that detect and remove outliers and regression analysis that smoothes data by fitting the data into regression functions.

3.3.2 Data integration

> *Data integration* is the task of combining data from multiple sources into a coherent schema.

The main problem of data integration is the schema integration, which requires integrating metadata from different sources. The secondary problems that arise regard the entity identification problem, the difficulty of identifying real world entities from multiple data sources and detecting and resolving data value conflicts. Unfortunately, for the same real world entity, attribute values from different sources are different. Possible reasons of this fact are different representations and different scales [Pyle, 1999].

Handling redundant data is another important issue of data integration. Redundant data occur often when integrating from multiple databases; the same attribute may have different names in different databases or one attribute may be a "derived" attribute in another table. Redundant data can be detected by correlational analysis. The careful integration of the data from multiple sources may help reducing, or avoiding, redundancies and inconsistencies and thus improving mining speed and quality.

3.3.3 Data transformation

> *Data transformation* is the task of normalizing, aggregating, and generalizing the data.

Normalization of data covers two key topics: normalizing the range of a variable and normalizing the distribution of a variable. We must note that neither of these normalization methods have anything in common with putting the data in normal-form tables (database normalization) [Pyle, 1999].

It is often convenient to use a specific state space, a common range for the variables. This need is covered with range normalization, which takes values that span one range and represents them in another range. The most common range is the 0-1 limits. Although a range is normalized in this scale, it is not certain that the distribution of values has been altered. Distribution normalization ensures the preservation of the balances between the values.

Aggregation of data regards the summarization of data in higher levels. Usually data is collected in detail at the lowest level. Such data are not mineable and must be aggregated in order to reveal interesting information. On-Line Analytical Processing (OLAP) is the technology of aggregating the data by creating data cube constructions.

Generalization of data regards concept hierarchy climbing. Concept hierarchies can be built over the data in order to extract knowledge in higher levels of granularity. For example the category hierarchy in retailing is the generalized concept hierarchy of the products database.

At most times derived attributes have to be constructed that represent stakeholder values and don't exist in the primary form of data. *Attribute construction* regards constructing new attributes from the given ones.

3.3.4 Data reduction

> *Data reduction* is the task of obtaining reduced representation in volume but producing the same or similar analytical results.

Warehouses can store terabytes of data. Complex data analysis like data mining may take a very long time to run on the complete data set. This fact is the primary need for data reduction. With data reduction a reduced representation of the data set is obtained, that is much smaller in volume but yet produces the same (or almost the same) analytical results [Pyle, 1999].

Some common data reduction strategies are data cube aggregation, dimensionality reduction, numerosity reduction, concept hierarchy generation, and discretization. Data reduction methods originate in sampling theory and statistics.

3.4 Modeling

The first step of modeling is selecting the actual modeling technique that is to be used to solve the mining case set in the first phase. It should not be forgotten that not all tools and techniques are applicable to each and every task. For certain problems, only some techniques are appropriate. From among these tools and techniques there are "Political Requirements" and other constraints, which further limit the choice available to the miner. It may be that only one tool or technique is available to solve the problem in hand – and even then the tool may not be the absolutely technical best for the problem in hand (figure 3.4.1).

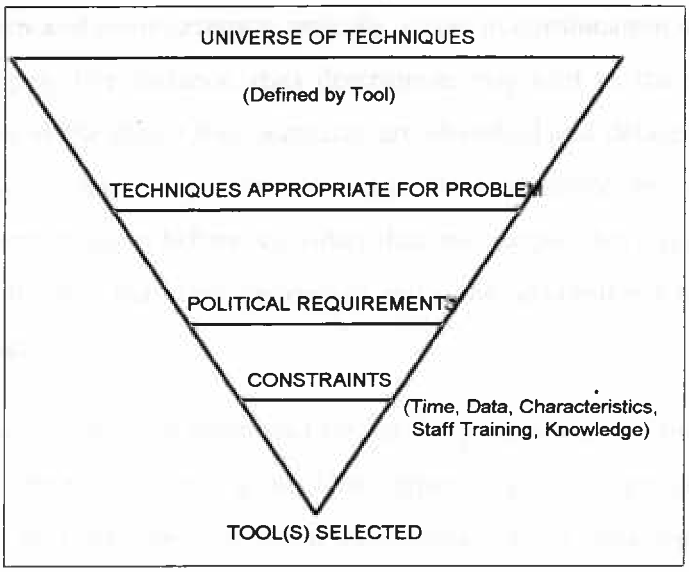


Figure 3.4.1 Selection of techniques

Since we are not dealing with a particular case, we don't confront a problem with all its parameters and therefore we cannot select the appropriate techniques. Instead, we list some of the techniques that usually map to specific problem types.

3.4.1 Data description and summarization

Data Description and Summarization aims at the concise description of characteristics of the data, typically in elementary and aggregated form. This gives the user an overview of the structure of the data. Sometimes, data description and summarization alone can be an objective of a data mining project. For instance, a retailer might be interested in the turnover of all outlets broken down by categories. Changes and differences to a previous period could be summarized and highlighted. This kind of problem would be at the lower end of the scale of data mining problems [Chapman et al, 2000].

However, in almost all data mining projects data description and summarization is a sub goal in the process, typically at early stages. At the beginning of a data mining process, the user often knows neither the precise goal of the analysis nor the precise nature of the data. Initial exploratory data analysis can help understand the nature of the data and to find potential hypotheses for hidden information. Simple descriptive statistical and visualization techniques provide first insights into the data. For example, the distribution of customer age and their living areas gives hints about which parts of a customer group need to be addressed by further marketing strategies.

Data description and summarization typically occurs in combination with other data mining problem types. For instance, data description may lead to the postulation of interesting segments in the data. Once segments are identified and defined a description and summarization of these segments is useful. It is advisable to carry out data description and summarization before any other data mining problem type is addressed; this is reflected by the fact that data description and summarization is a task in the data understanding phase.

Summarization also plays an important role in the presentation of final results. The outcomes of the other data mining problem types (e.g., concept descriptions or prediction models) may also be considered summarizations of data, but on a higher conceptual level.

Appropriate techniques:

- Many reporting systems
- Statistical packages
- OLAP systems

These techniques can cover data description and summarization but do usually not provide any methods to perform more advanced modeling. If data description and summarization is considered a stand alone problem type and no further modeling is required, these tools are also appropriate to carry out data mining engagements.

3.4.2 Segmentation

The *segmentation* data mining problem aims at the separation of the data into interesting and meaningful subgroups or classes. All members of a subgroup share common characteristics. For instance, in shopping basket analysis one could define segments of baskets depending on the items they contain [Chapman et al, 2000].

Segmentation can be performed manually or (semi-) automatically. The analyst can hypothesize certain subgroups as relevant for the business question based on prior knowledge or based on the outcome of data description and summarization. However, there are also automatic clustering techniques that can detect previously unsuspected and hidden structures in data that allow segmentation.

Segmentation can be a data mining problem type of its own. Then the detection of segments would be the main purpose of data mining. For example, all addresses in zip code areas with higher than average age and income might be selected for mailing advertisements on home nursing insurance.

Often, however, very often segmentation is a step towards solving other problem types. Then, the purpose can be to keep the size of the data manageable or to find homogeneous data subsets that are easier to analyze. Typically, in large datasets various influences overlay each other and obscure the interesting patterns. Then, appropriate segmentation makes the task easier. For instance, analyzing dependencies between items in millions of shopping baskets is very hard. It is much easier (and more meaningful, typically) to identify dependencies in interesting segments of shopping baskets, for instance high-value baskets, baskets containing convenience goods or baskets from a particular day or time.

Appropriate techniques:

- Clustering techniques.
- Neural nets.
- Visualization.

These techniques can be applied to the customer segmentation business case that was described in 3.1.1. Neural nets are applicable only in the case of numerical criteria of segmentation; in the case of categorical criteria other clustering techniques must be used. Visualization illustrates the results of the clustering process.

3.4.3 Concept descriptions

Concept description aims at an understandable description of concepts or classes. The purpose is not to develop complete models with high prediction accuracy, but to gain insights. For instance, a company may be interested in learning more about their loyal and disloyal customers. From a concept description of these concepts (loyal and disloyal customers) the company might infer what could be done to keep customers loyal or to transform disloyal customers to loyal customers [Chapman et al, 2000].

Concept description has a close connection to both segmentation and classification. Segmentation may lead to an enumeration of objects belonging to a concept or class without any understandable description. Typically, there is segmentation before concept description is performed. Some techniques, for example conceptual clustering techniques, perform segmentation and concept description at the same time.

Concept descriptions can also be used for classification purposes. On the other hand, some classification techniques produce understandable classification models, which can then be considered as concept descriptions. The important distinction is that classification aims to be complete in some sense. The classification model needs to apply to all cases in the selected population. On the other hand, concept descriptions need not be complete. It is sufficient if they describe important parts of the concepts or classes. In the example above, it may be sufficient to get concept descriptions of those customers who are clearly loyal.

Appropriate techniques:

- Rule induction methods.
- Conceptual clustering.

Concept description can be applied in the customer segmentation business case and help labeling the clusters that are the result of the clustering process.

3.4.4 Classification

Classification assumes that there is a set of objects – characterized by some attributes or features – which belong to different classes. The class label is a discrete (symbolic) value and is known for each object. The objective is to build classification models (sometimes called classifiers), which assign the correct class label to previously unseen and unlabeled objects. Classification models are mostly used for predictive modeling. The class labels can be given in advance, for instance defined by the user or derived from segmentation.

Classification is one of the most important data mining problem types that occur in a wide range of various applications. Many data mining problems can be transformed to classification problems. For example, credit scoring tries to assess the credit risk of a new customer. This can be transformed to a classification problem by creating two classes, good and bad customers. A classification model can be generated from existing customer data and their credit behavior. This classification model can then be used to assign a new potential customer to one of the two classes and hence accept or reject him.

Classification has connections to almost all other problem types. Prediction problems can be transformed to classification problems by binning continuous class labels, since binning techniques allow transforming continuous ranges into discrete intervals. These discrete intervals are then used as class labels rather than the exact numerical values and hence lead to a classification problem. Some classification techniques produce understandable class or concept descriptions. There is also a connection to dependency analysis because classification models typically exploit and elucidate dependencies between attributes.

Segmentation can either provide the class labels or restrict the dataset such that good classification models can be built.

It is useful to analyze deviations before a classification model is built. Deviations and outliers can obscure the patterns that would allow a good classification model. On the other hand, a classification model can also be used to identify deviations and other problems with the data.

Appropriate techniques:

- Discriminant analysis.
- Rule induction methods.
- Decision tree learning.
- Neural nets.
- K Nearest Neighbor.
- Case-based reasoning.
- Genetic algorithms.

3.4.5 Prediction

Another important problem type that occurs in a wide range of applications is *prediction*. Prediction is very similar to classification. The only difference is that in prediction the target attribute (class) is not a qualitative discrete attribute but a continuous one. The aim of prediction is to find the numerical value of the target attribute for unseen objects. In the literature, this problem type is sometimes called regression. If prediction deals with time series data then it is often called forecasting.

Appropriate techniques:

- Regression analysis.
- Regression trees.
- Neural nets.
- K Nearest Neighbor.
- Box-Jenkins methods.
- Genetic algorithms.

Prediction is proposed to assist the decision of the objectives of the promotion campaign in the promotion design business case, by predicting future sales, market share, brand volume.

3.4.6 Dependency analysis

Dependency analysis consists of finding a model that describes significant dependencies (or associations) between data items or events. Dependencies can be used to predict the value of a data item given information on other data items. Although dependencies can be used for predictive modeling, they are mostly used for understanding. Dependencies can be strict or probabilistic.

Associations are a special case of dependencies, which have recently become very popular. Associations describe affinities of data items (i.e., data items or events which frequently occur together). A typical application scenario for associations is the analysis of shopping baskets (*market basket analysis*). There, a rule like "in 30 percent of all purchases, beer and peanuts have been bought together" is a typical example for an association.

Dependency analysis has close connections to prediction and classification, where dependencies are implicitly used for the formulation of predictive models. There is also a connection to concept descriptions, which often highlight dependencies.

In applications, dependency analysis often co-occurs with segmentation. In large datasets, dependencies are seldom significant because many influences overlay each other. In such cases it is advisable to perform a dependency analysis on more homogeneous segments of the data.

Sequential patterns are a special kind of dependencies where the order of events is considered. In the shopping basket domain, associations describe dependencies between items at a given time. Sequential patterns describe shopping patterns of one particular customer or a group of customers over time.

Appropriate Techniques:

- Correlation analysis.
- Regression analysis.
- Association rules.
- Bayesian networks.

- Inductive Logic Programming.
- Visualization techniques.

Dependency analysis can be applied both in the promotion design and the store design business case. In the first it offers assistance on the decision about the brand to be promoted and the type of the promotion. In the second it extracts the sequences of purchasing for each customer.

Testing and Validating

After selecting the techniques but prior to building the model, a procedure must be defined that will test the model's quality and validity. For example, in supervised data mining tasks such as classification, it is common to use error rates as quality measures for data mining models. Therefore the test design specifies that the dataset should be separated into training and test set, the model is built on the training set and its quality estimated on the test set.

Building the model regards detailed parameter setting, which differs from technique to technique. Each model is assessed on the test procedure defined earlier according to the test strategy selected (Train and Test, Cross-validation etc). We note that this step of assessment checks the validity of the mining process, whilst the next phase assesses the degree to which the model meets the business objectives.

3.5 Evaluation

Previous evaluation steps dealt with factors such as the accuracy and generality of the model. This step assesses the degree to which the model meets the business objectives and seeks to determine if there is some business reason why this model is deficient. It compares results with the evaluation criteria defined at the start of the project.

Another option of evaluation is to test the model(s) on test applications in the real application if time and budget constraints permit. Moreover, evaluation also assesses other data mining results generated. Data mining results cover models which are necessarily related to the original business objectives and all other findings which are not necessarily related to the original business objectives but might also unveil additional challenges, information or hints for future directions.

After the model assessment with respect to business success criteria, the resultant models should appear to be satisfactory and to satisfy business needs. It is then appropriate to make a more thorough review of the data mining engagement in order to determine if there is any important factor or task that has somehow been overlooked.

According to the assessment results and the process review, the next steps must be determined. The project members need to decide whether to finish this project and move onto deployment or whether to initiate further iterations or whether to set up new data mining projects.

3.6 Deployment

In the deployment phase a strategy is concluded on deploying the evaluation results. The strategy depends both on the business goals and the results. The strategy should include the ways to propagate the knowledge, the ways to measure the benefits from the application of the results and identify possible problems in the deployment of the data mining results.

Deployment phase also includes plans for monitoring and maintenance in case the results become part of the day-to-day business along with a final report that summarizes the whole project and records problems and results.

4 Conclusions

Marketing analysis is undoubtedly essential to building the strategy of any kind of business. Unfortunately, there are no "golden" rules of marketing that guarantee business success. Every business has its unique characteristics that differentiate the marketing strategy. Marketing in Retailing is an area that attracted attention from the research community and one of its main scopes was and is to clarify the factors that affect the decision process of the customers. Until recently, the majority of the marketing decisions were based on empirical results and the analysis was restricted in controlled environments that do not reflect reality.

The goal of this thesis was to facilitate the marketing processes and offer a more stable and reliable tool to base the marketing decisions. We used data mining techniques to exploit the historical data and reveal the hidden knowledge about the customers. We believe that illuminating the interrelationships between the customers with common behavior contributes to the effectiveness of marketing, since the marketing decisions cease to base on hypotheses, but are based on actual facts that exist and stand for the specific business.

In order to achieve this goal, we used the marketing and retailing literature to identify the most important marketing business cases in retailing. We have also based on the CRISP-DM methodology, a data mining methodology, to structure the thesis and clarify the steps of applying data mining to the retailing business cases. Finally, we expanded many cases in Internet Retailing, when this new environment offered additional information that could be exploited by knowledge discovery.

Extensions of this work have two axes of approach. The first would be to expand to other marketing business cases that this thesis has not covered. A very interesting issue is pricing of products that has already attracted a lot of attention, but not yet reached to a model to predict the "ideal" price. This business case is even more interesting in Internet Retailing since the one-to-one interface is a challenge for predicting the price for each customer. The second approach would be to validate the models that were proposed in the thesis. This validation requires real historical data to apply the models and post-evaluation to measure the effectiveness of the marketing decisions that were made based on the results of the mining models.

The emerging telecommunication and electronic commerce technologies and the automatic product identification will change retailing one more time, after the transformation due to POS scanning technology and the Internet. "Smart" homes with "smart" devices along with mobile networks expansion will probably alternate the scenery of the ways of buying in the next decade [MyGrocer, 2000]. This new environment differentiates marketing from traditional and Internet marketing and undoubtedly data mining will be used as a tool to exploit all the additional information.

References

Aggarwal C.Charu., Zheng Sun and Philip. S. Yu (1998). Online Algorithms for Finding Profile Association Rules. *Proceedings of the International Conference on Knowledge Discovery and Data Mining.*KDD-98, pp. 129-133, AAAI Press

Agrawal Deepak, Christopher Schorling, "Market Share Forecasting: An Empirical Comparison of Artificial Neural Networks and Multinomial Logit Model", *Journal of Retailing,* Vol. 72(4), 383-407

Agrawal Rakesh and Ramakrishnan Srikant (1994). "Fast Algorithms for Mining Association Rules". *In Proceedings of the 20th VLDB Conference,* Santiago, Chile, 487-499

Agrawal Rakesh and Ramakrishnan Srikant (1995). "Mining Sequential Patterns". In *Proceedings of the 11th International Conference on Data Engineering,* Taipei, Taiwan, March 1995

Agrawal Rakesh, Tomasz Imielinski and Arun Swami (1993). "Mining Association Rules between /sets of Items in Large Databases ", *Proceedings of the 1993 ACM SIGMOD Conference Washington DC,* USA, May 1993, 207-216

Ailawadi L. Kusum, Paul W. Farris, Mark E. Parry, (1994). "Share and Growth Are Not Good Predictors of the Advertising and Promotion/Sales Ratio", *Journal of Marketing,* Vol. 58, January 1994, 86-97

Ambler Tim and Flora Kokkinaki (1997). "Measures of Marketing Success". *Journal of Marketing Management,* 13, 665-678

Anderson W. Eugene and Mary W. Sullivan (1993). "The antecedents and Consequences of Customer Satisfaction for Firms". *Marketing Science.* 12, 2, 125-143.

Anderson W. Eugene, Claes Fornell and Donald R. Lehmann (1994). "Customer Satisfaction and Profitability: Findings from Sweden". *Journal of Marketing,* Vol. 58 (July 1994), 53-66

Anderson W. Eugene, Claes Fornell and Roland T. Rust (1997). "Customer Satisfaction, Productivity and Profitability: Differences Between Goods and Services". Marketing Science. Vol. 16, No. 2, 129-145.

Babcock C. (1994). *Parallel Processing Mines Retail data.* ComputerWorld, 6, September 26, 1994

Bawa Kapil and Robert W. Shoemaker (1987). "The Coupon-Prone Consumer: Some Findings Based on Purchase Behavior Across Product Classes". *Journal of Marketing*, Vol. 51 (October 1987), 99-110

Bettman R. James, Mary Frances Luce and John W. Payne (1998). "Constructive Consumer Choice Process", *Journal of Consumer Research*

Buzzell D. Robert and Bradley T. Gale (1987). *The PIMS Principles: Linking Strategy to Performance*, Free Press, New York

Cabena Peter, Hyun Hee Choi, Il Soo Kim, Shuichi Otsuka, Joerg Reinschmidt and Gary Saarenvirta (1999). Intelligent Miner for Data Applications Guide. International Technical Support Organization, IBM, March 1999.

Chapman Pete, Julian Clinton, Randy Kerber, Thomas Khabaza (SPSS), Thomas Reinartz (DaimlerChrysler), Colin Shearer (SPSS) and Rudiger Wirth (DaimlerChrysler) (2000). CRISP-DM 1.0. Step by step data mining guide. CRISP-DM Consortium, August 2000.

Chen Ming-Syan, Jiawei Han and Philip S. Yu (1996). "Data Mining: An Overview from a Database Perspective." IEEE Transactions on Knowledge and Data Engineering, Vol. 8, No. 6, 1996

Churchill A. Gilbert and Carol Surpenant (1982). "An Investigation into the Determinants of Customer Satisfaction". Journal of Marketing Research 19 (November), pp 491-504

Clark H. Bruce (1999). "Marketing Performance Measures: History and Interrelationships", Journal of Marketing Management, Vol. 15, 1999, pp. 711-732

Codd E.F. (1993). *Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate.* E.F. Codd and Associates.

Cyberce and ICOM (2000). Data Components Design. Deliverable D2.3. iMEDIA

Day S. George and Liam Fahey (1988). "Valuing Market Strategies", Journal of Marketing, 52, 3, 45-57.

Dick S. Alan and Kunal Basu (1994). "Customer Loyalty: Toward an Integrated Conceptual Framework". Journal of the Academy of Marketing Science. Vol. 22, No. 2, pp. 99-113.

Doyle Peter and Zeki Gidengil (1977). "A Review of In-Store Experiments". *Journal of Retailing*, 53 (Summer), 47-62

East Robert, Kathy Hammond, Patricia Harris and Wendy Lomax (2000), "First-Store Loyalty and Retention", *Journal of Marketing Management*, Vol 16, 2000, 307-325

Ernst & Young (2000). Global Online Retailing. An Erst & Young Special Report. Ernst & Young LLP

Fayyad M. Usama, Gregory Piatesky-Shapiro, Padhraic Smyth and Ramasamy Uthurusamy (1996). Advances in Knowledge Discovery and Data Mining. AIII Press/The MIT Press.

Feder A. Robert (1965). "How to Measure Marketing Performance". Harvard Business Review, 43, May-June, 132-142

Fisher D. (1987). "Knowledge acquisition via incremental conceptual clustering". *Machine Learning*, 2:139-172Fisher D. (1995). "Optimization and simplification of hierarchical clusterings". *In Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining* (KDD'95), pages 118-123, Montreal, Canada, August.

Gedenk Karen and Scott A. Neslin (1999), "The Role of Retail Promotion in Determining Future Brand Loyalty: Its Effect on Purchase Event Feedback", *Journal of Retailing*, Vol. 75, No. 4, 1999, 433-459

Goodman R. Sam (1970). *Techniques of profitability analysis*. Wiley-Interscience, New York

Guadagni M. Peter and John D. Little (1983). "A Logit Model of Brand Choice Calibrated on Scanner Data". *Marketing Science*, 2 (Summer), 203-238.

Halstead Diane, David Hartman and Sandra L. Schmidt (1994). "Multisource Effects on the Satisfaction Formation Process". *Journal of the Academy of Marketing Science*. Vol. 22, No.2, pp. 114-129.

Hardy G. Kenneth (1986). "Key Success Factors for Manufacturers' Sales Promotions in Package Goods". *Journal of Marketing*. Vol. 50 (July 1986), 13-23

Henderson B. James and Richard Quandt (1958). *Micro Economics Theory: A Mathematical Approach*. New York: McGraw-Hill Book Company

Jacoby Jacob and Robert W. Chestnut (1978). *Brand Loyalty Measurement and Management*. New York: Wiley.

Jain A.K. and R. C. Dubes (1988). *Algorithms for Clustering Data.* Printice Hall, 1988Kent J. Robert, Chris T. Allen(1994). "Competitive Interference Effects in Consumer Memory for Advertising: The Role of Brand Familiarity", *Journal of Marketing,* Vol. 58, July 1994, 97-105

Kotler Philip (1984). *Marketing Management: Analysis, Planning, and Control.* Englewood Cliffs, NL: Prentice-Hall

Kumar Pankaj and Suresh Divakar (1999). "Size Does Matter: Analyzing Brand-Size Competition Using Store Level Scanner Data". *Journal of Retailing,* Vol 75(1), pp. 59-76

Kumar V., Kiran Karande, Werner J. Reinartz, "The Impact of Internal and External Prices on Brand Choice: The Moderating Role of Contextual Variables", *Journal of Retailing,* Vol. 43(3), 1998, 401-426

Leone P. Robert, Srini S. Srinivasan, "Coupon Face Value: Its Impact on Coupon Redemptions, Brand Sales and Brand Profitability", *Journal of Retailing,* Vol. 72(3), 1996, 273-289

Macintosh Gerrard and Lawrence S. Lockshin (1997), "Retail relationships and store loyalty: A multi-level perspective", *International Journal of Research in Marketing,* Vol. 14, 1997, 487-497

McCarthy E. Jerome (1960). *Basic Marketing: A Managerial Approach.* Homewood, IL: Richard D. Irwin, Inc.

Mcintyre H. Shelby and Dale D. Achabal (1993). "Applying Case-Based Reasoning to Forecasting Retail Sales". *Journal of Retailing,* Vol. 69 (Winter 1993), No. 4, 372-398

Morgan S. Michael, Chekitan S. Dev, "An Empirical Study of Brand Switching for a Retail Service", *Journal of Retailing,* Vol. 70, No. 3, 1994, 267-282

Mulhern J. Francis and Daniel T. Padgett (1995). "The Relationshp Between Retail Price Promotions and Regular Price Purchases". *Journal of Marketing,* Vol. 59 (October 1995), 83-90

Oliver L. Richard (1980). "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions". Journal of Marketing Research 17 (November), pp. 460-469.

Pine B. Joseph (1993). Mass Customization. Harvard Business School Press. Boston, Massachussets.

Prassas George, Katherine Pramataris and Olga Papaemmanouil (2001). "Dynamic Recommendations in Internet Retailing". *Submitted in ECIS 2001*

Pyle Dorian (1999). Data Preparation for Data Mining. Morgan Kaufmann Publishers Inc.

Richardson S. Paul, Arun K. Jain, Alan Dick, "Household Store Brand Proneness: A Framework", *Journal of Retailing*, Vol. 72(2), 1996, 159-185

Rust T. Roland and Antony J. Zahorik (1993). "Customer Satisfaction, Customer Retention and Market Share". Journal of Retailing. Vol. 69, No 2, Summer 1993, 193-215.

Schafer J. Ben, Joseph. A Konstan and John Riedl (1999). "Recommender Systems in E-Commerce". *In ACM Conference on Electronic Commerce (EC-99)*, pages 158-166

Sevin H. Charles (1965). *Marketing Productivity Analysis*. McGraw-Hill, New York.

Simonson Itamar (1999). "The Effect of Product Assortment on Buyer Preferences". *Journal of Retailing*, Vol. 75(3), 347-370

Srikant Ramakrishnan and Rakesh Agrawal (1995). "Mining generalized association rules". *VLDB'95*, 407-419, Zurich, Switzerland, Sept. 1995.Srikant Ramakrishnan and Rakesh Agrawal (1996). "Mining quantitative association rules in large relational tables". *SIGMOD'96*, 1-12, Montreal, Canada.Szymanski M. David, Sundar G. Bharadwaj and P. Rajan Varadarajan (1993). "Ana Analysis of the Market Share-Profitability Relationship". *Journal of Marketing*, Vol. 57, July 1993, 1-18.

Walters G. Rockney (1988). "Retail Promotions and Retail Store Performance: A Test of Some Key Hypotheses". *Journal of Retailing*, 64 (Summer), 153-180

Walters, Rockney G. (1991), "Assessing the Impact of Retail Price Promotions on Product Substitution, Complementary Purchase, and Interstore Sales Displacement", *Journal of Marketing*, Vol. 44, April 1991, 17-28

Watershoot van Walter & Christophe Van den Bulte (1992). "The 4P Classification of the Marketing Mix Revisited". *Journal of Marketing*. Vol. 56 (October 1992), 83-93

Wilcox B. James, Roy D. Howell, Paul Kuzdrall and Robert Britney (1987). "Price Quantity Discounts: Some Implications for Buyers and Sellers". *Journal of Marketing*, Vol. 51 (July 1987), 60-70

Yu S. Philip (1999). "Data Mining and Personalization Technologies". *DASFAA 1999*, 6-13.

Zhang T., R. Ramakrishnan, and M. Livny (1996). "BIRCH: an efficient data clustering method for very large databases". *In Proceedings of ACM-SIGMOD International Conference Management of Data*, Montreal, Canada, June 1996.