

CHAPTER FOUR

Advanced Business Analysis

From Business Intelligence: Making Decisions through Data Analytics
By Jerzy Surma (A Business Expert Press Book)



© 2011 by Business Expert Press. All rights reserved.

Harvard Business Publishing distributes in digital form the individual chapters from a wide selection of books on business from publishers including Harvard Business Press and numerous other companies. To order copies or request permission to reproduce materials, call 1-800-545-7685 or go to <http://www.hbsp.harvard.edu>. No part of this publication may be reproduced, stored in a retrieval system, used in a spreadsheet, or transmitted in any form or by any means – electronic, mechanical, photocopying, recording, or otherwise – without the permission of Harvard Business Publishing, which is an affiliate of Harvard Business School.

This document is authorized for use only by Ching Yang at University of Wales Trinity Saint David until April 2014.

Copying or posting is an infringement of copyright. Permissions@hbsp.harvard.edu or 617.783.7860.

Do Not Copy or Post

CHAPTER 4

Advanced Business Analysis

4.1. Introduction

The main aim of the approaches previously discussed was to provide decision makers with suitable information. It is possible to identify this information if you precisely define the monitored key performance indicator (KPI) or if an analyst formulates adequate queries during the analysis. They usually ask and then answer the questions “What happened?” (reporting), “Why did it happen?” (analyzing), and “What is happening now?” (monitoring). But the most intriguing are the questions about the future: “What will happen?” (extrapolation and prediction). Additionally, we can often discover unexpected connections (regularities) in the given business area. This can be achieved thanks to data-mining methods, which stem from techniques derived from statistics and algorithms of machine learning.¹ Data mining is defined as “the nontrivial extraction of implicit, previously unknown, and potentially useful information (patterns) from data.”²

Data mining is often called knowledge discovery in databases. The most vital element of this knowledge discovery is the identification of patterns, which are then defined against the background of the structure. The structure is a global summary of the database. The pattern in this context describes the structure, which refers to a relatively small part of the data, that is, reduced (local) space.³

Usually, the result of data-mining algorithms is understood as knowledge (see Figure 4.1) of which we were not previously aware. Knowledge, according to its definition in an encyclopedia, is all credible information about reality and the ability to use it. Sticking precisely to this definition, and particularly to its second part, the result of data mining cannot be treated as knowledge. However, as the term is popularly understood, such an interpretation seems to be

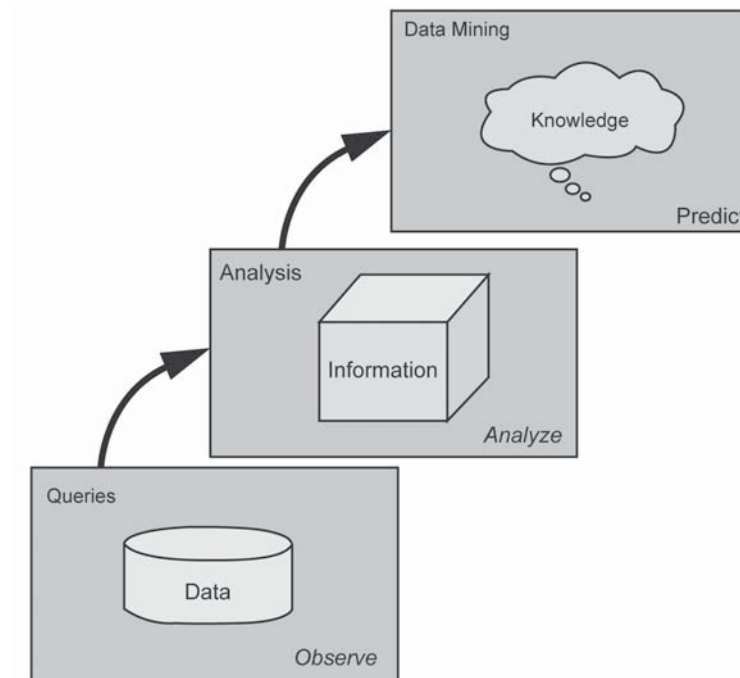


Figure 4.1. Data mining: from data to knowledge.

Source: Reinschmidt, Gottschalk, Kim, and Zweitering (1999).

acceptable because classic data analysis consists of answering a given question through data analysis. Advanced analysis enables you to go even further, which allows you to discover things that were not formerly known because there was no reason to ask about them. Therefore, you find out something new about a given area of reality and this can be regarded as discovering knowledge.

The most common applications of data mining in business are classification, prediction, cluster analysis, and mining association rules:⁴

1. *Classification* is assigning an object to a predefined category. Usually decision trees, neural nets, or k nearest neighbors, for instance, are used.
2. *Prediction (or estimation)* is similar to classification, but it assumes that the category to which you assign the object has a continuous set of values. Such prediction is usually used for forecasting, for which you often use regression decision trees, artificial neural nets, and standard statistical methods.

3. *Cluster analysis (or clustering)* consists of finding in the set of objects subsets (groups) of objects that have similar features (i.e., are highly homogenous). It is often called “taxonomy,” “cluster analysis,” or popularly in marketing “segmentation post hoc.”
4. *Mining association rules* involve searching sets (usually pairs) of objects that appear together in a given context. These objects are linked to each other (by association), which means that the presence of one object implies with a certain probability the presence of the object that is linked to it.

Later on, classification, cluster analysis, and association principle discovery will be discussed separately from prediction and estimation, which is commonly known from statistics.⁵

4.2. Classification

In business, decision trees are commonly used for classification.⁶ Their big advantage from a user’s point of view is that they provide an explicit and clear representation of knowledge. The central idea of decision trees, as understood by John Ross Quinlan,⁷ and the algorithm used by decision trees will be discussed by looking at an example of a basic system that supports credit decisions (loan applications) in a retail bank.⁸

The decision to grant credit is an example of classification that consists of putting a credit application described by a set of descriptive attributes into one of two categories: to grant the credit (YES) or to deny it (NO). To conduct classification, you have to define a classifier, which would enable you to categorize objects (i.e., credit applications) into one of at least two previously defined categories (classes; see Figure 4.2). A borrower is described by four descriptive attributes and one decision attribute that describes the decision made in the credit process (see Table 4.1). A demonstration classifier has the form of the decision tree (see Figure 4.3). As seen in Figure 4.3, in the internal nodes of the decision tree, there are selected attributes; branches represent values of a given attribute, while leaf nodes (terminal nodes) stand for the classification decision. Therefore, this tree represents the following knowledge about classification of credit applications expressed in the formula “if . . . then . . .”:

Table 4.1. The Description of Attributes for Classification of the Credit Application

Kind of attribute	Attribute	Set of values
Descriptive	Risk group	{low, medium, high}
Descriptive	Education	{primary, secondary, high}
Descriptive	Income forecast	{optimistic, pessimistic}
Descriptive	Guarantee	{full, partial, no}
Decision	To grant a credit	{YES, NO}

Source: author.

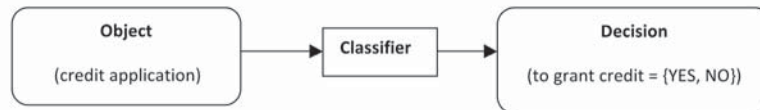


Figure 4.2. Classification of new objects.

Source: author.

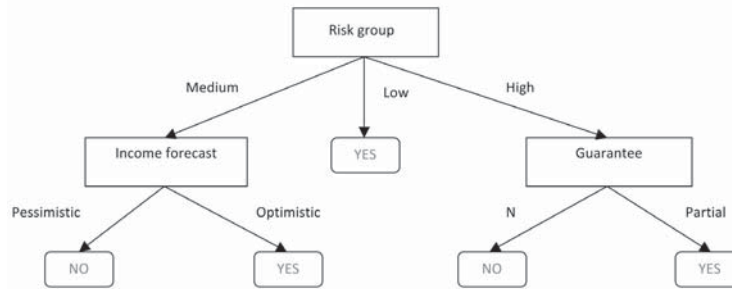


Figure 4.3. Classifier: Decision tree for the classification of a credit application.

Source: author.

- If Risk Group = medium and Income Forecast = pessimistic, then Grant a Credit = NO
- If Risk Group = medium and Income Forecast = optimistic, then Grant a Credit = YES
- If Risk Group = low, then Grant a Credit = YES
- If Risk Group = high and Guarantee = no, then Grant a Credit = NO
- If Risk Group = high and Guarantee = partial, then Grant a Credit = YES

Usually, this set of rules is supplemented by a default rule, which applies if none of the classifier rules is met by a categorized object. Let us now consider how you construct a classifier. The fundamental idea consists of learning (generating a decision tree) on the basis of learning examples (learning set; see Figure 4.4). Learning examples are examples of correctly categorized (classified) objects. For our example in Table 4.2, 14 examples from the history of credits granted by a bank are shown. Each example has a concrete value for four attributes that describe a credit application and a clear-cut classification decision (see the definition of attributes in Table 4.1). For every example, “YES” means that the credit was properly repaid, and “NO” signifies that there were some setbacks (including, among others, lack of repayment) and the borrower should not have been given credit approval.

In the described example, the learning system from Figure 4.4 will be represented by algorithm ID3.⁹ It is one of the most common algorithms for building decision trees based on top-down induction. It works as follows:

1. Name A the best attribute (according to the information gain test) for the next node (during the first iteration, it is a root) of the tree.
2. For each value of A attribute, create a branch with the next node of the tree.
3. Divide each set of teaching examples into subsets for each value of the Attribute A.
4. If all examples have the same decision, create a leaf node and stop. If not, then start again (from step 1) for each of the newly created nodes.

The key step of algorithm ID3 is the first step, which consists of specifying an attribute that will appear in the node of the tree. Intuitively,

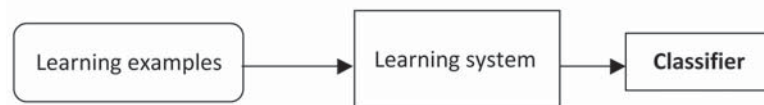


Figure 4.4. Learning the classifier from learning examples.

Source: author.

Table 4.2. Learning Examples to Build a Classifier for Credit Applications

Example	Risk group	Education	Income forecast	Guarantee	To grant credit
P1	medium	high	pessimistic	partial	NO
P2	medium	high	pessimistic	no	NO
P3	low	high	pessimistic	partial	YES
P4	high	secondary	pessimistic	partial	YES
P5	high	primary	optimistic	partial	YES
P6	high	primary	optimistic	no	NO
P7	low	primary	optimistic	no	YES
P8	medium	secondary	pessimistic	partial	NO
P9	medium	primary	optimistic	partial	YES
P10	high	secondary	optimistic	partial	YES
P11	medium	secondary	optimistic	no	YES
P12	low	secondary	pessimistic	partial	YES
P13	low	high	optimistic	partial	YES
P14	high	secondary	pessimistic	no	NO

Source: author.

you should choose an attribute whose values will divide a set of examples into maximum homogenous subsets in the context of being assigned to decision classes.

Figure 4.5 shows how the third step of algorithm ID3 would divide all learning examples into subsets for each of descriptive attribute. As you can see, the attribute “Risk Group” seems to divide the best elements of the set of examples into homogenous sets, and for the value “low,” the set is almost entirely homogenous and has the value “YES” for the decision attribute. In algorithm ID3, information gain serves as an attribute selection method:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \text{Entropy}(S|A),$$

where $\text{Gain}(S,A)$ is the expected reduction in entropy caused by knowing the value of attribute A , $\text{Entropy}(S)$ is the entropy connected with the classification of the set of examples S , and $\text{Entropy}(S|A)$ is the entropy of the division of the set of examples S according to the attribute A .

The lower the value of $\text{Entropy}(S|A)$, the greater the organization (purity)—that is, the homogeneity of the classification for examples

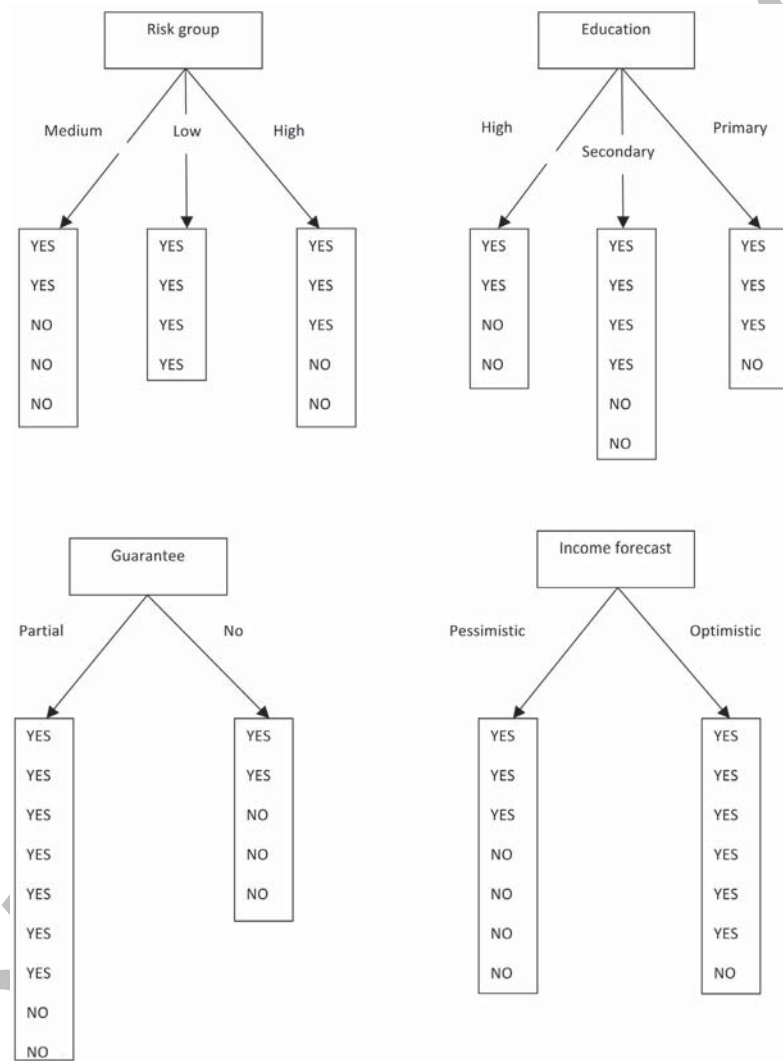


Figure 4.5. Division of set of examples according to selected attribute.

divided into subsets. For our example, suitable information gain has the highest value for the attribute “Risk Group”:¹⁰

- Gain (S, Risk group) = 0.246 bit
- Gain (S, Education) = 0.029 bit
- Gain (S, Income forecast) = 0.151 bit
- Gain (S, Guarantee) = 0.048 bit

Therefore, this attribute should appear first in the root of the tree (see Figure 4.3). You should also notice that the algorithm did not use all information included in the set of learning examples. The attribute “Education” turned out to be of no use while generating the tree, and as a result, it has no influence on the decision made about granting credit.

This classifier in the form of the decision tree for making credit authorizations is obviously purely illustrative. In reality, the whole process of constructing a credible classifier—including, for instance, choice of attributes, preparation of a set of learning examples, parameterization of the chosen algorithm, and finally its testing—is a challenging task that demands considerable work and experience.¹¹

4.2.1. Limitations

To construct a classifier on the basis of a set of examples, you need to solve many problems that are common for the majority of data-mining algorithms. However, if you are aware of these limitations, you should have reasonable expectations regarding their possible applications and the quality of the knowledge generated by them. This awareness also gives methodological guidelines concerning learning and testing techniques. The main problems are connected with induction, history, updating, and overfitting.

4.2.1.1. Induction Problem

Learning from examples is inductive reasoning. You make credible generalizations on the basis of a limited set of observations. This is the fundamental limitation of inductive conclusions: You are unable to prove that a given inductive conclusion is true. However, you can falsify it unambiguously.¹² Inductive reasoning maintains falsity, which means that if facts that are subject to generalizations are not true, inductive conclusions are also not true, but true facts do not necessarily bring you to true conclusions. Therefore, by definition you should treat every inductive conclusion as uncertain.

4.2.1.2. History and Updating

The process of generating a classifier involves learning on the basis of historical observations. It seems to be obvious that the fragment of reality described by this classifier is subject to changes in time. This is a key issue, particularly in business applications, in which the classifier may have become outdated even a relatively short time after the learning process (e.g., as a result of rapid changes in the market). This is then another fundamental factor that influences the uncertainty of classification. In fact, the classifier needs to be updated on the basis of new observations, whereas the more obsolete examples should be forgotten. There are some techniques to construct an incremental decision tree that take into account new learning examples without the necessity of building a whole new tree from zero.

4.2.1.3. Overfitting

A standard verification of classifiers is based on the examples from the testing set. The classic phenomenon that results from too deep learning based on learning set it is called overfitting. A tree into which too many classifications fit and that reflects the learning examples perfectly is usually highly complex and has very little ability to generalize. An illustration of this is the graph in Figure 4.6. The X -axis represents the complexity of the tree reflected in the number of nodes, while the Y -axis represents the accuracy of the classification calculated by the quotient of accurately fitted examples (numerator) and the number of all examples (denominator). The continuous graph represents classification accuracy on trees that have learned and been tested on the same series of examples (learning set). As you can see, it is an increasing function that almost reaches 1. It is more important to analyze the graph that is marked with a dotted line; the tree is based here on the set of learning examples (learning set) and tested on a separate set of test examples (test set). When the tree is a certain size (approximately 20 nodes), more complex trees are more closely adjusted to the learning set (see the continuous graph) and generalization is more and more difficult (see the dotted line). According to Tom Mitchel,¹³ overfitting is a problem not only of decision trees but also of other algorithms in machine learning. Here this problem is resolved by reducing the size

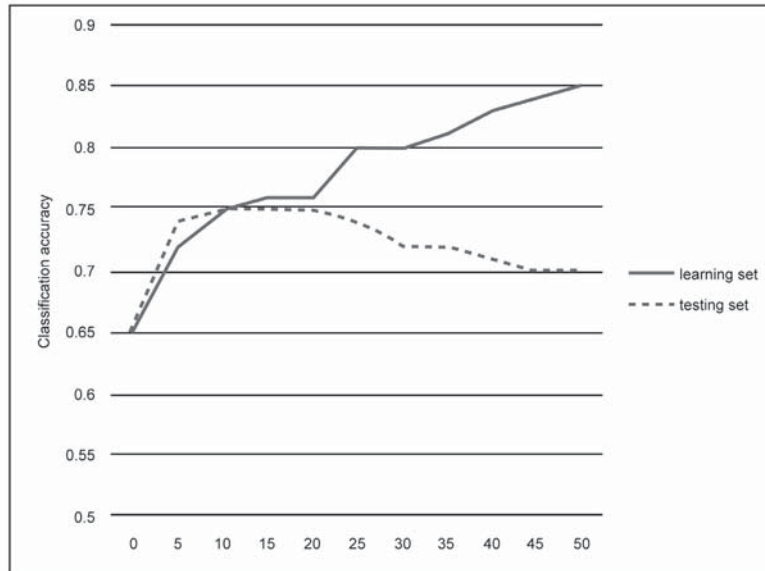


Figure 4.6. Overfitting during learning of decision trees.

Source: Based on Mitchell (1997).

of the tree by pruning: You remove some less-important fragments (sub-trees) from the complete decision tree.

Case Study 4.1 shows the interpretation of outcomes generated by using the decision tree approach.

Case Study 4.1

One of the main strategic initiatives achieved by the implementation of the strategy was a rapid growth in the market share of the company. This was realized by improving the selection procedures for locations of new stores and rapidly opening new stores in the chain.

As this initiative was being carried out, it was observed that the new locations of stores in most cases were selected intuitively without recourse to information about the functioning of other chain stores in similar locations and without access to demographic and business data. The effects of business intelligence (BI) implementation to date, technological tools (the data warehouse), and the desire to take advantage of information technology (IT) assets convinced the board of

directors to continue using advanced analytical methods to support the decision-making process in the choice of new locations.

Previously, the decision about opening a new store and its location was made by the member of the board responsible for chain development on the basis of a specialist's recommendation. This recommendation was based on

1. confirmation of the required legal and formal status for a new store,
2. a location report including factual and photo documentation and the situational plan describing the business environment of the store,
3. the specialist's intuitive evaluation based on his long-standing experience.

The board of directors decided to experimentally apply data-mining methods to find out whether they could use them to support the traditional procedures for selection of new locations. This was a classification task that included the following considerations:

- Descriptive attributes included information about the location of the store, such as total floor space, number of citizens within a 1000-meter radius, number of passersby, competitors within a 500-meter radius, number of parking places, description of the business environment, description of location, and so on. Geomarketing data purchased from an external company were also taken into account, including information about saturation of the target group around the store and suitable demographic and socioeconomic data.
- The decision about the desired profitability of the store was made and divided into three types:
 - *League 1*. Profitability considerably above the desired level of profitability
 - *League 2*. Profitability within certain limits (target values)
 - *League 3*. Profitability close to the desired level of profitability or below it

After looking at the aforementioned data, stores were classified based on the following type of rules:

*If certain location, then store of certain type =
{1 league, 2 league, 3 league}.*

Thanks to such a set of rules, the board of directors was able to predict how the store would potentially operate in a given location. The board of directors decided to construct a classifier in the form of a decision tree. To create a learning and test set, information about the locations and profitability of stores in cities with a population above 50,000 and with a similar demographic and economic profile was used. Consequently, a 300-element set of examples was constructed, which for better differentiation consisted of a comparable number of stores from only leagues 1 and 3. Of this set, 200 elements (learning sets) were used to generate the classification rules. The generated tree was tested on the remaining 100 elements of the set and included 50 examples from league 1 and exactly the same number from league 3 (test set). The results are:

Test examples (set)	Classified as league 1	Classified as league 3
Fifty examples from league 1	40	10
Fifty examples from league 3	20	30

The classification accuracy¹⁴ was to 0.7. This result proved to be satisfactory. However, the problem of incorrect classifications was brought to light. The classifier classified 20 examples from league 3 as league 1, and 10 league 1 stores were classified as league 3. Such a situation carries with it substantial business costs both if the company opts for YES for an unsuitable location or if the company opts for NO for a suitable location, which would result in lost sales. In the context of this analysis of the initial results, the board of directors decided to continue working to improve the quality of the classifier before it is applied to support real decisions about locations.

4.3. Cluster Analysis

Cluster analysis consists of finding in a set of objects subsets (groups) that have common attributes (features). The cluster analysis algorithm aims to divide a set of objects into subsets in which inter-class similarity is maximized and intraclass similarity is minimized. Whereas in a classification task you have to classify objects, in clustering your task is to find classes into which objects can be divided. That is why learning classification principles are called *supervised learning* and the identification of groups of objects is called *unsupervised learning*.

The key issue in cluster analysis is finding similarity between objects. If objects are described by quantitative attributes (sales amount, financial liquidity, etc.), you usually employ similarity measures based on the notion of distance and the Euclidean measure for calculations. It is more difficult to calculate similarity if the attributes are qualitative (e.g., attributes in Table 4.1). In such situations, you use techniques to transform qualitative attributes into quantitative ones so that you can finally use measures based on the notion of the distance between objects in space whose dimensions are described by attributes.¹⁵

One of the most popular approaches to cluster analysis is the k -means algorithm.¹⁶ It works as follows:

1. Choose k observations by random (number k is given a priori) as one-element classes, which will serve as the cluster center (center of gravity).
2. Assign each observation to the nearest class (the most similar cluster center).
3. Calculate the new cluster center for every k class.
4. If the cluster center has not changed significantly, you should stop here. If not, go back to step 2.

In Figure 4.7, the k -means algorithm is shown in the space of two attributes, where $k = 2$. The circled points represent the cluster center. In practice, the algorithm finishes when, for all groups, objects assigned to them remain in them without changing their membership.

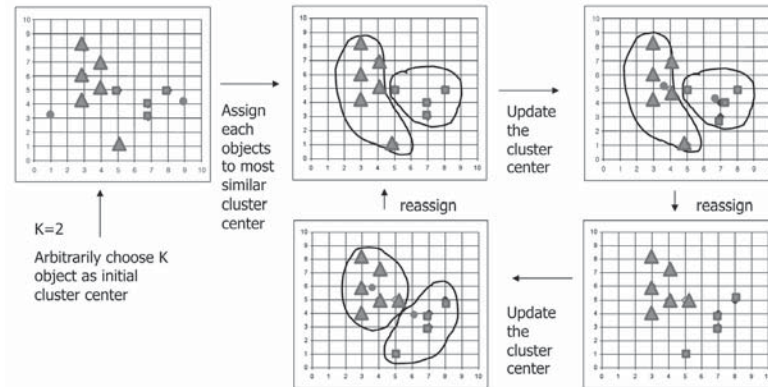


Figure 4.7. Example of *k*-means algorithm.

Source: Yang (2008).

Clustering is commonly used in marketing research. In the classic approach, market segments are given a priori in an arbitrary manner. Consequently, their further analysis depends on the accuracy of this choice. In the case of the cluster analysis algorithm, the segments are automatically discovered by the algorithm and reflect real connections between customer groups, which are often based on latent (implicit) attributes—this is called *post hoc segmentation*.¹⁷ A review of selected business applications of cluster analysis methods is shown later, in Tables 4.5 and 4.7.

4.4. Association Rules

Mining association rules consist of searching groups of objects that appear together in certain contexts. This task is performed by means of association rules analysis.¹⁸ Market basket analysis, in which you want to find out which products are most frequently sold together in one basket, is as a classic example.¹⁹ Let us consider a simple example of association rules analysis by looking at the following example of four market baskets. Each line in Table 4.3 represents a market basket registered on the receipts in a cash register.²⁰ You can use these receipts to analyze the frequency of certain pairs of products. In Table 4.4, the frequency of the simultaneous presence of two products for all possible pairs is shown. Table 4.4 gives the following information about sales:

Table 4.3. Example of a Set of Receipts

Receipt	Product names
1	Juice, water
2	Milk, juice, bread
3	Juice, beer
4	Juice, beer, water

Source: author.

Table 4.4. Frequency of the Simultaneous Presence of Pairs of Products in the Market Basket

	Juice	Bread	Milk	Water	Beer
Juice	4	1	1	2	2
Bread	1	1	1	0	0
Milk	1	1	1	0	0
Water	2	0	0	2	1
Beer	2	0	0	1	2

Source: author.

- Sales of juice are very good because this product appears in every single basket.
- Juice is relatively often sold together with water and beer.
- Milk is never sold together with beer or water.

Such conclusions would be well grounded if the sample of the market basket had an adequately high statistical significance. The essence of mining association rules during market basket research lies in formulating principles such as “if one puts product A into a basket, then it is highly plausible that he or she will put product B into a basket as well.” The accuracy of such a principle in the context of a conviction that such an event took place is described by two parameters: support and confidence:²¹

- *Support* describes how often products A and B appear together in the basket against all transactions.
- *Confidence* describes the conditional probability of putting product B into the basket if product A has been put there earlier.

Let us, for instance, consider the following rule: “*If* there is juice in a basket, *then* water is also in a basket.” For this principle with shopping, as in Table 4.3,

- support = (number of baskets with juice and water) ÷ (number of all baskets) = $2 \div 4$ (50%),
- confidence = (number of baskets with juice and water) ÷ (number of all baskets with juice) = $2 \div 4$ (50%).

Let us also consider the opposite principle to understand things more clearly: “*If* there is water in a basket, *then* juice is also in a basket.” Notice that support for this rule will be exactly the same, but confidence will be higher:

- confidence = (number of baskets with juice and water) ÷ (number of all baskets with water) = $2 \div 2$ (100%).

This means that (for a given sample of baskets) putting water into a basket always implies the purchase of juice.²² Figure 4.8 shows all possible combinations of rules and their parameters: support and confidence.

	If	Then	Support (%)	Confidence (%)
1	beer	juice	50.000	100.000
2	juice	beer	50.000	50.000
3	beer	juice, water	25.000	50.000
4	water	juice, beer	25.000	50.000
5	water, beer	juice	25.000	100.000
6	juice	water, beer	25.000	25.000
7	juice, beer	water	25.000	50.000
8	juice, water	beer	25.000	50.000
9	bread	juice	25.000	100.000
10	juice	bread	25.000	25.000
11	bread	juice, milk	25.000	100.000
12	milk	juice, bread	25.000	100.000
13	milk, bread	juice	25.000	100.000
14	juice	milk, bread	25.000	25.000
15	juice, bread	milk	25.000	100.000
16	juice, milk	bread	25.000	100.000
17	milk	juice	25.000	100.000
18	juice	milk	25.000	25.000
19	water	juice	50.000	100.000
20	juice	water	50.000	50.000
21	beer	water	25.000	50.000
22	water	beer	25.000	50.000
23	bread	milk	25.000	100.000
24	milk	bread	25.000	100.000

Figure 4.8. Example of association principles.

Source: author.

Association rules can be presented graphically, as shown in Figure 4.9. Each product is represented by a circle, and its size corresponds to the frequency of the presence of a given product in the baskets. The vector between the two circles represents the association rule, and its thickness is connected with the parameters of the evaluation of the rule.

Knowledge about connections between sales of products can be used for

- arrangement of products on store shelves,
- promotional packages,
- recommendation of upselling,
- comparative research of stores and taking into account the factor of time in carrying out research (mining sequence patterns).²³

Obviously, these applications go beyond the environment of chain stores and can be useful in other business areas.²⁴ It is particularly worth stressing that they can support upselling: You can analyze a customer basket in time and consequently suggest rational solutions.

4.4.1. Limitations

There are two vital limitations of both cluster analysis and association rules analysis that are of great importance for other data-mining methods as well: the problems of interpretation and computational complexity.

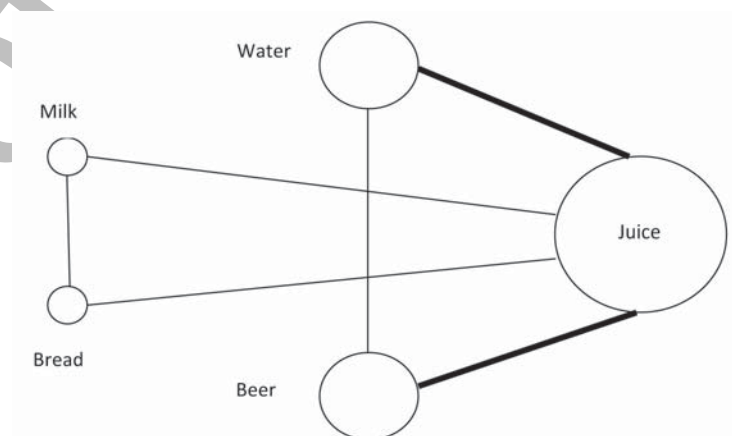


Figure 4.9. Example of graphical representation of association rules.

4.4.1.1. The Problem of Interpretation

In general, rules derived from association rules algorithms only slightly contribute to knowledge that was not previously known and would simultaneously be useful for business purposes. In most cases, those rules are trivial and do not go beyond our current knowledge. There is also a separate group of rules in which interpretation is problematic in business areas. These rules are therefore usually omitted because of their incidental and unclear character. Similarly, the interpretation of classification rules is also dubious. Another problem is business interpretation of segments generated automatically by cluster analysis algorithms. The proper identification (interpretation) of important segments and the rejection of those that are not useful in business require the analyst to have a profound knowledge of the market. In fact, it is a problem that concerns other data-mining algorithms and is a real limitation to its business applications.

4.4.1.2. The Problem of Computational Complexity

You can easily imagine how much data must be processed by the algorithm based on receipts from a chain of stores. There are millions of records even in a relatively short time period. The number of product indexes in an average chain usually reaches a few thousand. The issue becomes even more complicated if you take into account not pairs of products but three or more product items. The complexity of the analysis increases sharply, which carries with it unacceptable times of generating the rules.²⁵ This problem applies to most algorithms for discovering knowledge and is one of the basic limitations of these approaches.

4.5. Other Methods

The aforementioned techniques are only a selection of the most popular ones in business applications. There is a wide spectrum of other approaches that derive mainly from statistics. An extremely interesting group of techniques are those inspired by the computer simulation of natural phenomena. The human brain has been an inspiration for a

long time, and within works on artificial intelligence, numerous computer simulations of neural networks²⁶ have been carried out. Although this work has so far been unsuccessful in simulating human thinking, it has been tremendously effective in solving classification and prediction tasks. Another technique derived from man's observations of nature is the genetic algorithm, which is based on research into theories of evolution and genetics.²⁷ This approach worked out a lot of concrete solutions, which are in most cases applied to resolving optimization problems. However, this is rarely used in business. One of the key elements in proving the reliability of data-mining analyses is the possibility of explaining (clarifying) suggested answers (e.g., by showing a path of reasoning in a decision tree to justify credit decisions). Both artificial neural networks and genetic algorithms do not justify explicitly the solutions they generate and are a sort of black box in the process of supporting decisions.

4.6. Review of Selected BI Applications in Terms of Data Mining

The following is a review of selected BI applications of data-mining methods for main and controlling processes:

1. Main processes
 - Sales, marketing, and service (see Table 4.5)
 - Internal and external logistics (see Table 4.6)
 - Operations and production (see Table 4.6)
2. Controlling processes
 - Finance (see Table 4.7)

Case Study 4.2 shows the results of a pilot study using a market basket analysis. The subsequent chapter shows how BI systems can be used to analyze human behavior.

Table 4.5. Review of Selected BI Applications (Advanced Analysis) in Sales, Marketing, and Service

Name	Description	Tools
Analytical customer relationship management (CRM)	Data analysis system based on the registration of events concerning a relationship with a customer, that is, analysis of customer preferences and behavior in order to generate accurate decisions about upselling, suggesting suitable promotions, and buying new products. Can be achieved by accurate segmentation of the market and the calculation of customer live time value.	OLAP, data mining (classification, clustering, association rules), statistics (forecasting methods)
Churn vs. retention	Analysis of reasons for recent customer abandonments and finding plausibility of current customers' ones. Commonly used by, for example, telecommunications operators in order to devise suitable marketing campaigns or improve relationship with customers who are likely to abandon.	OLAP, data mining (classification and clustering)
Conjoint analysis	Analysis of customer preferences in the context of the influence of different product qualities (quality, price, functionality, service, etc.) on a decision to buy a given product.	Econometrics (analysis of variance)
Chi-square automatic interaction detection (CHAID) analysis	Segmentation of customers according to influence of certain descriptors (age, education, gender, etc.) on possibility of repeat choice of the product of the same brand.	Econometrics (automatic interaction detection model)
Recency, frequency, monetary (RFM) analysis	Segmentation of customers based on customer behavior in the context of recency (time from recent purchase), frequency (number of purchased products over certain time), and monetary value (value of purchased products over certain time). On the basis of such, an analysis of upselling and cross-selling strategies are devised.	OLAP, data mining (clustering)
Response model	Forecasting who in a given group of customers will respond to an offer on the basis of analysis of history of behavior of a similar group of customers. Used to in marketing campaign management.	Econometrics (regression analysis, prediction), data mining (classification)

Source: author.

Table 4.6. Review of Selected BI Applications in Logistics and Production

Area	Name	Description	Tools
Logistics	Demand planning	Using historic sales data for forecasting. It is often used by chain of stores because it enables you to discover which products are seasonal (etc.) and as a result to manage supply chain effectively.	Data mining (regression trees), statistics (forecasting methods)
Logistics	Optimization of trade routes	Mapping out of optimal route according to which you can reach in a certain order destination with minimal costs.	Operational research (transportation problem)
Production	Operations management	Supporting decisions concerning amount of resources you should allocate to certain activities ordered to perform a given task.	Operational research (techniques for solving resource allocation problems)

Source: author.

Table 4.7. Review of Selected BI Applications (Advanced Analysis) in Finance

Name	Description	Tools
Bankruptcy prediction	Use of financial analysis of an enterprise to construct early warning systems in the context of possible bankruptcy.	Data mining (classification algorithm, for instance discrimination function analysis)
Fraud detection	Used, for example, in banking; it is the analysis of different frauds in financial transactions to find patterns of criminal activities. By monitoring of current transactions and use of such patterns, you can locate current fraud attempts.	OLAP, data mining (classification and clustering)
Risk modeling	Analysis of financial risk in services and products where payment is postponed. A classic example is the analysis of credit ratings.	OLAP, data mining (classification and clustering)

Source: author.

Case Study 4.2

Within the scope of strategic initiatives, ALFA's board of directors concluded that it was extremely important to improve the chain's sales volume. One of the recommended methods for improving sales was to optimize the arrangement of products on the shelves to increase sales. To perform this task, the board decided to employ a market basket analysis and, after generating the association rules, modify the arrangement of products in its stores.

To verify the validity of this approach before applying it in the whole chain, a pilot study was carried out. The sample was 50 stores in which all receipts from a whole month were registered. The algorithm was parameterized in such a way (adequate values for support and confidence) that it generated rules only for pairs of products. The study resulted in a set of rules, which were divided into three groups:

1. *Trivial rules* (approx. 60%). Self-evident knowledge, usually previously known and used
2. *Incomprehensible rules* (approx. 30%). Knowledge that is difficult to interpret, not justified in business, or of no use in practice
3. *Useful rules* (approx. 10%). Vital and previously unknown knowledge that can be directly used for new arrangement of products

On the basis of the useful rules, modifications were introduced in some stores (in which the volume of sales was being monitored) and compared with test stores (in which products were not rearranged). Apart from the new arrangement of goods, salespeople were given special instructions on how, on the basis of the customer market basket, to convince customers to take advantage of upselling of products located near the checkout. After 3 months, the result of this comparison study convinced the board to make a systematic market basket analysis for all its stores and then implement the results across the whole chain.

Recommended Literature

Introduction to Data Mining

- Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). San Francisco, CA: Morgan Kaufmann.
- Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Amsterdam, Netherlands: Academic Press.

Classification

- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification*. New York, NY: Wiley-Interscience.
- Quinlan, J. (1993). *C4.5: Programs for machine learning*. San Francisco, CA: Morgan Kaufmann.

Cluster Analysis

- Anderberg, M. (1973). *Cluster analysis for applications*. Amsterdam, Netherlands: Academic Press.

Association Rules

- Adamo, J. (2000). *Data mining for association rules and sequential patterns: Sequential and parallel algorithms*. Berlin, Germany: Springer Verlag.

Neural Nets

- Hertz, J. A., Krogh, A. S., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Reading, MA: Addison-Wesley.

Genetic Algorithms

- Michalewicz, Z. (1996). *Genetic algorithms + data structure = evolution programs*. Berlin, Germany: Springer Verlag.

Data Mining in Business Applications

- Berry, M. J., & Linoff, G. S. (1999). *Mastering data mining: The art and science of customer relationship management*. New York, NY: Wiley.
- Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: For marketing, sales, and customer relationship management*. Indianapolis, IN: Wiley.
- Rud, O. (2000). *Data mining cookbook: Modeling data for marketing, risk and customer relationship management*. New York, NY: Wiley.

Case Study (Market Basket Analysis for a Chain Store)

- Corinne, B. (2001). *Mining your own business in retail using DB2 intelligent miner or data*. IBM Form No. SG24-6271-00. IBM Redbooks.

Internet Resources

- KD Nuggets. <http://www.kdnuggets.com/>
- Data Mining Group. <http://www.dmg.org/>
- Index of Machine Learning Courses. <http://www.cs.iastate.edu/~honavar/Courses/cs673/machine-learning-courses.html>
- Tom Mitchell Machine Learning textbook. <http://www.cs.cmu.edu/~tom/mlbook.html>
- Machine Learning Repository. <http://archive.ics.uci.edu/ml>
- Statistical Aspects of Data Mining (lecture taught by prof. Rajan Patel). <http://www.youtube.com/watch?v=zRsMEl6PHhM>
- Cross Industry Standard Process for Data Mining. <http://www.crisp-dm.org/>