



**T.Y.B.SC (I.T)**  
**SEMESTER - VI (CBCS)**

**BUSINESS INTELLIGENCE**

**SUBJECT CODE: USIT603**

**Dr. Suhas Pednekar**

Vice Chancellor

University of Mumbai, Mumbai

**Prof. Ravindra D. Kulkarni**

Pro Vice-Chancellor,  
University of Mumbai

**Prof. Prakash Mahanwar**

Director,  
IDOL, University of Mumbai

**Programme Co-ordinator**

**: Shri Mandar Bhanushe**

Head, Faculty of Science and Technology IDOL  
University of Mumbai – 400098

**Course Co-ordinator**

**: Gouri S. Sawant**

Assistant Professor B.Sc.I.T, IDOL  
University of Mumbai- 400098

**Course Writers**

**: Farzin Rais Qureshi**

Assistant Professor,  
Elphinstone College M.G Road, Fort,  
Mumbai

**: Prof Shahid Ansari**

Assistant Professor,  
Mumbai.

**: Dr Vinayak Pujari**

Assistant Professor,  
D.Y.Patil College of Engineering and Technology  
Kolhapur.

**: Dr Rakhee Yadav**

Assistant Professor,  
S. K Somaiya college,  
Somaiya Vidyavihar University, Mumbai

**: Dr. Rajeshri Shinkar**

Assistant Professor,  
SIES (NERUL) College of Arts, Science &  
Commerce, Navi Mumbai.

May 2023, Print - 1

**Published by : Director,**

Institute of Distance and Open Learning,

University of Mumbai,

Vidyanagari,Mumbai - 400 098.

**DTP composed and Printed by: Mumbai University Press**

# CONTENTS

<b>Chapter No.</b>	<b>Title</b>	<b>Page No.</b>
<b>Unit 1</b>		
1.	Business Intelligence.....	01
2.	Decision Support Systems.....	13
<b>Unit 2</b>		
3.	Mathematical Models for Decision Making.....	30
4.	Data Mining.....	39
5.	Data Preparation .....	51
<b>Unit 3</b>		
6.	Classification .....	63
7.	Clustering .....	73
<b>Unit 4</b>		
8.	Marketing, Logistic and production models .....	83
9.	Data envelopment analysis.....	116
<b>Unit 5</b>		
10.	Knowledge Mangament .....	126
11.	Artificial Intelligence & Expert Systems .....	142

**T.Y.B.SC (I.T)**  
**SEMESTER - VI**

**BUSINESS INTELLIGENCE**

<b>Unit</b>	<b>Details</b>	<b>Lectures</b>
<b>I</b>	<p><b>Business intelligence:</b> Effective and timely decisions, Data, information and knowledge, The role of mathematical models, Business intelligence architectures, Ethics and business intelligence</p> <p><b>Decision support systems:</b> Definition of system, Representation of the decision-making process, Evolution of information systems, Definition of decision support system, Development of a decision support system</p>	12
<b>II</b>	<p><b>Mathematical models for decision making:</b> Structure of mathematical models, Development of a model, Classes of models</p> <p><b>Data mining:</b> Definition of data mining, Representation of input data , Data mining process, Analysis methodologies</p> <p><b>Data preparation:</b> Data validation, Data transformation, Data reduction</p>	12
<b>III</b>	<p><b>Classification:</b> Classification problems, Evaluation of classification models, Bayesian methods, Logistic regression, Neural networks, Support vector machines</p> <p><b>Clustering:</b> Clustering methods, Partition methods, Hierarchical methods, Evaluation of clustering models</p>	12
<b>IV</b>	<p><b>Business intelligence applications:</b></p> <p><b>Marketing models:</b> Relational marketing, Sales force management,</p> <p><b>Logistic and production models:</b> Supply chain optimization, Optimization models for logistics planning, Revenue management systems.</p> <p><b>Data development analysis:</b> Efficiency measures, Efficient frontier, The CCR model, Identification of good operating practices</p>	12
<b>V</b>	<p><b>Knowledge Management:</b> Introduction to Knowledge Management, Organizational Learning and Transformation, Knowledge Management Activities, Approaches to Knowledge Management, Information Technology (IT) In Knowledge Management, Knowledge Management Systems Implementation, Roles of People in Knowledge Management</p> <p><b>Artificial Intelligence and Expert Systems:</b></p> <p>Concepts and Definitions of Artificial Intelligence, Artificial Intelligence Versus Natural Intelligence, Basic Concepts of Expert Systems, Applications of Expert Systems, Structure of Expert Systems, Knowledge Engineering, Development of Expert Systems</p>	12

<b>Books and References:</b>					
<b>Sr. No.</b>	<b>Title</b>	<b>Author/s</b>	<b>Publisher</b>	<b>Edition</b>	<b>Year</b>
1.	Business Intelligence: Data Mining and Optimization for Decision Making	Carlo Vercellis	Wiley	First	2009
2.	Decision support and Business Intelligence Systems	Efraim Turban, Ramesh Sharda, Dursun Delen	Pearson	Ninth	2011
3.	Fundamental of Business Intelligence	Grossmann W, Rinderle-Ma	Springer	First	2015



# 1

## BUSINESS INTELLIGENCE

### Unit Structure

- 1.0 Objective
- 1.1 Introduction
- 1.2 An Overview
- 1.3 Effective and timely decisions
- 1.4 Data, Information and Knowledge
- 1.5 The Role of Mathematical Models
- 1.6 Business Intelligence Architectures
- 1.7 Cycle of a Business Intelligence Analysis
- 1.8 Development of a Business Intelligence System
- 1.9 Ethics and Business Intelligence
- 1.10 Summary
- 1.11 List of References
- 1.12 Unit End Exercise

---

### 1.0 OBJECTIVE

---

- To learn about Business Intelligence
- To learn about how to take effective and timely decisions in an organization.
- To learn how to extract knowledge from data and information.
- To learn how to draw conclusions, predictions and take futuristic actions.
- To learn the architecture of BI system.

---

### 1.1 INTRODUCTION

---

- ***Business intelligence*** may be defined as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making processes.
- BI is concerned with the representation and organization of the decision-making process, and thus with the field of decision theory; with collecting and storing the data intended to facilitate the decision-making process.

- We can say that business intelligence systems tend to promote a scientific and rational approach to managing enterprises and complex organizations.

---

## 1.2 AN OVERVIEW

---

- ***Business intelligence*** may be defined as a set of mathematical models and analysis methodologies that systematically exploit the available data to retrieve information and knowledge useful in supporting complex decision-making processes.
- A business intelligence environment offers decision makers information and knowledge derived from data processing, through the application of mathematical models and algorithms.

---

## 1.3 EFFECTIVE AND TIMELY DECISIONS

---

- The main purpose of business intelligence systems is to provide knowledge workers with tools and methodologies that allow them to make effective and timely decisions.
- In complex organizations, public or private, decisions are made on a continual basis. Such decisions may be critical, have long- or short-term effects and involve people and roles at various hierarchical levels.
- The ability of these knowledge workers to make decisions, both as individuals and as a community, is one of the primary factors that influence the performance and competitive strength of a given organization

### **Effective Decisions**

- The application of analytical methods allows decision makers to rely on information and knowledge which are more dependable.
- As a result, they can make better decisions and devise action plans that allow their objectives to be reached in a more effective way.

### **Timely Decisions**

- Enterprises operate in economic environments characterized by growing levels of competition and high dynamism.
- Therefore, the ability to rapidly react to the actions of competitors and to new market conditions is a critical factor in the success or even the survival of a company.

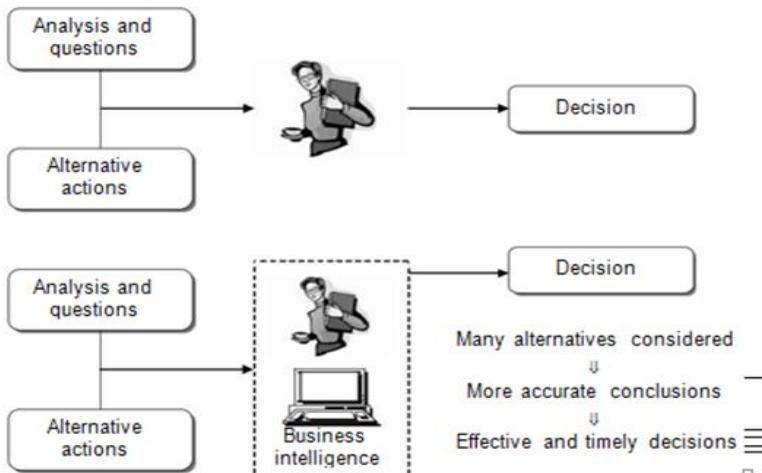


Figure 1.1 Benefits of a business intelligence system

## 1.4 DATA, INFORMATION AND KNOWLEDGE

- Large amount of data has been accumulated within the information systems of public and private organizations.
- These data originate from internal transactions of an administrative, logistical and commercial nature and partly from external sources.
- However, even if they have been gathered and stored in a systematic and structured way, these data cannot be used directly for decision-making purposes.
- They need to be processed by means of appropriate extraction tools and analytical methods capable of transforming them into information and knowledge that can be subsequently used by decision makers.

### Data

- Generally, data represent a structured codification of single primary entities, as well as of transactions involving two or more primary entities.
- For example, for a retailer data refer to primary entities such as customers, points of sale and items, while sales receipts represent the commercial transactions.

### Information

- Information is the outcome of extraction and processing activities carried out on data, and it appears meaningful for those who receive it in a specific domain.
- For example, to the sales manager of a retail company, the proportion of sales receipts in the amount of over 100 per week, or the number of customers holding a loyalty card who have reduced by more than 50% the monthly amount spent in the last three months, represent meaningful pieces of information that can be extracted from raw stored data.

## Knowledge

- Information is transformed into knowledge when it is used to make decisions and develop the corresponding actions.
- Therefore, we can think of knowledge as consisting of information put to work into a specific domain, enhanced by the experience and competence of decision makers in tackling and solving complex problems.
- For a retail company, a sales analysis may detect that a group of customers, living in an area where a competitor has recently opened a new point of sale, have reduced their usual amount of business.

## 1.5 THE ROLE OF MATHEMATICAL MODELS

- A business intelligence system provides decision makers with information and knowledge extracted from data, through the application of mathematical models and algorithms.
- In some instances, this activity may reduce to calculations of totals and percentages, graphically represented by simple histograms.
- In general terms, the adoption of a business intelligence system tends to promote a scientific and rational approach to the management of enterprises and complex organizations.
- Classical scientific disciplines, such as physics, have always resorted to mathematical models for the abstract representation of real systems.

The rational approach typical of a business intelligence analysis can be summarized systematically in the following main characteristics.

- First, the objectives of the analysis are identified and the performance indicators that will be used to evaluate alternative options are defined.
- Mathematical models are then developed by exploiting the relationships among system control variables, parameters and evaluation metrics.

Finally, what-if analyses are conducted to evaluate the effects on the performance decided by variations in the control variables and changes in the parameters.

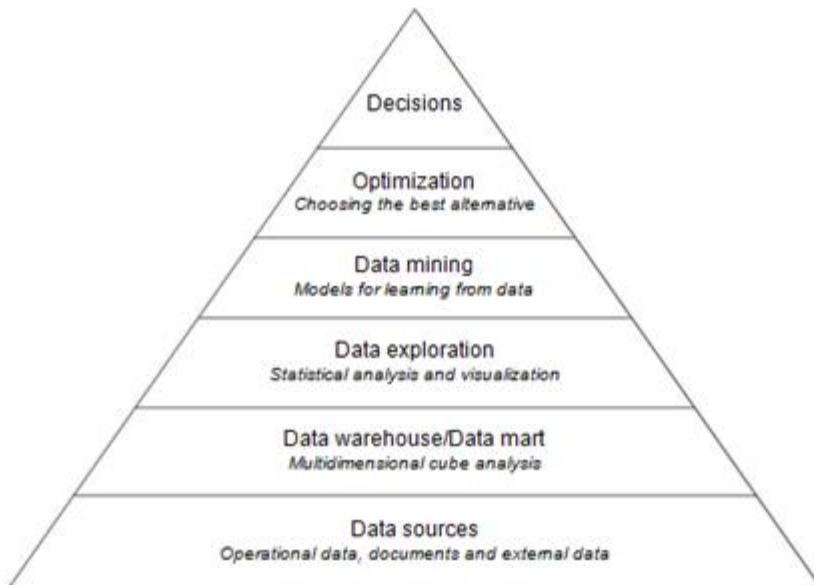
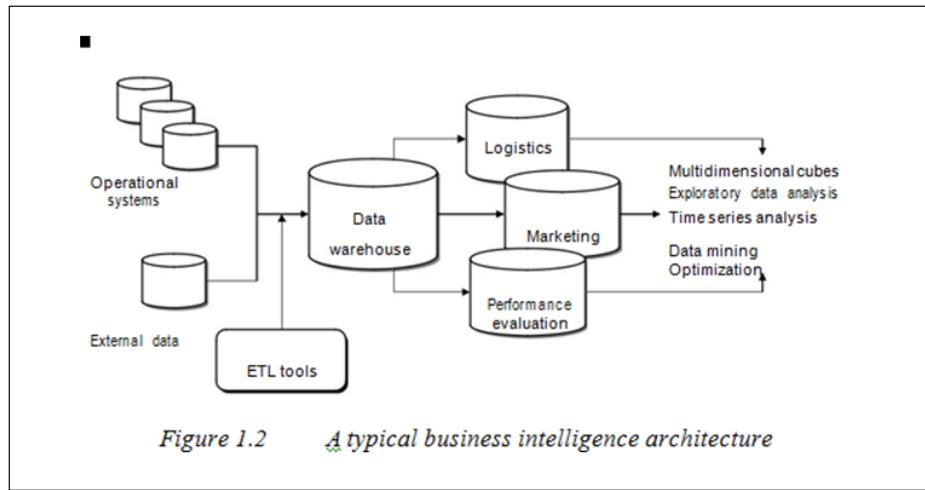
## Advantages

- The primary objective is to enhance the effectiveness of the decision-making process.
- The adoption of mathematical models also affords other advantages, which can be appreciated particularly in the long term.

- First, the development of an abstract model forces decision makers to focus on the main features of the analysed domain, thus inducing a deeper understanding of the phenomenon under investigation.
- The knowledge about the domain acquired when building a mathematical model can be more easily transferred in the long run to other individuals within the same organization, thus allowing a sharper preservation of knowledge in comparison to empirical decision-making processes.

## 1.6 BUSINESS INTELLIGENCE ARCHITECTURES

The architecture of a business intelligence system, shown in Figure, includes three major components.



*Figure 1.3 The main components of a business intelligence system*

## **Data sources.**

- In a first stage, it is necessary to gather and integrate the data stored in the various primary and secondary sources, which are heterogeneous in origin and type.
- The sources consist for the most part of data belonging to operational systems, but may also include unstructured documents, such as emails and data received from external providers.
- A major effort is required to unify and integrate the different data sources

## **Data warehouses and data marts.**

- Using extraction and transformation tools known as extract, transform, load (ETL), the data originating from the different sources are stored in databases intended to support business intelligence analyses.
- These databases are usually referred to as data warehouses and data marts.

## **Business intelligence methodologies.**

- Data are finally extracted and used to feed mathematical models and analysis methodologies intended to support decision makers.
- In a business intelligence system, several decision support applications may be implemented.
  - multidimensional cube analysis;
  - exploratory data analysis;
  - time series analysis;
  - inductive learning models for data mining;
  - Optimization models.

## **Data exploration.**

- At the third level of the pyramid, we find the tools for performing a passive business intelligence analysis, which consist of query and reporting systems, as well as statistical methods.
- These are referred to as passive methodologies because decision makers are requested to generate prior hypotheses or define data extraction criteria, and then use the analysis toolsto find answers and confirm their original insight.

## **Data mining.**

- The fourth level includes active business intelligence methodologies, whose purpose is the extraction of information and knowledge from data.

- These include mathematical models for pattern recognition, machine learning and data mining techniques.
- Models of an active kind do not require decision makers to formulate any prior hypothesis to be later verified. Their purpose is instead to expand the decision makers' knowledge.

Business Intelligence

### Optimization.

- By moving up one level in the pyramid we find optimization models that allow us to determine the best solution out of a set of alternative actions, which is usually extensive and sometimes even infinite.

### Decisions.

- Finally, the top of the pyramid corresponds to the choice and the actual adoption of a specific decision and in some way represents the natural conclusion of the decision-making process.
- Even when business intelligence methodologies are available and successfully adopted, the choice of a decision pertains to the decision makers

## 1.7 CYCLE OF A BUSINESS INTELLIGENCE ANALYSIS

- Each business intelligence analysis follows its own path according to the application domain, the personal attitude of the decision makers and the available analytical methodologies.
- It is possible to identify an ideal cyclical path characterizing the evolution of a typical business intelligence analysis, as shown in the following figure.

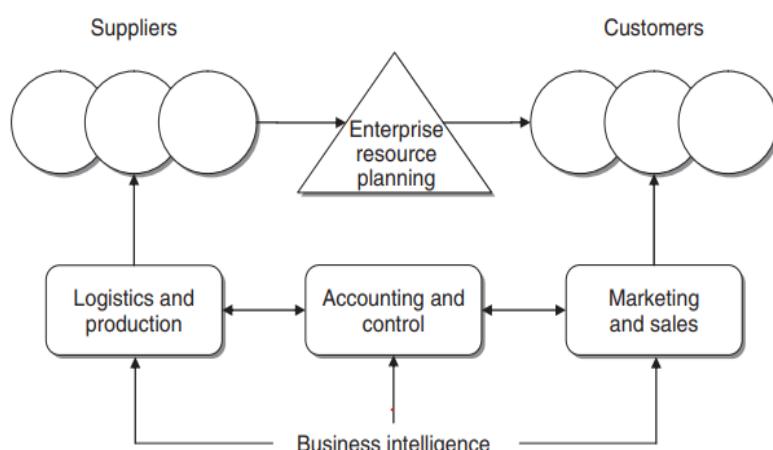
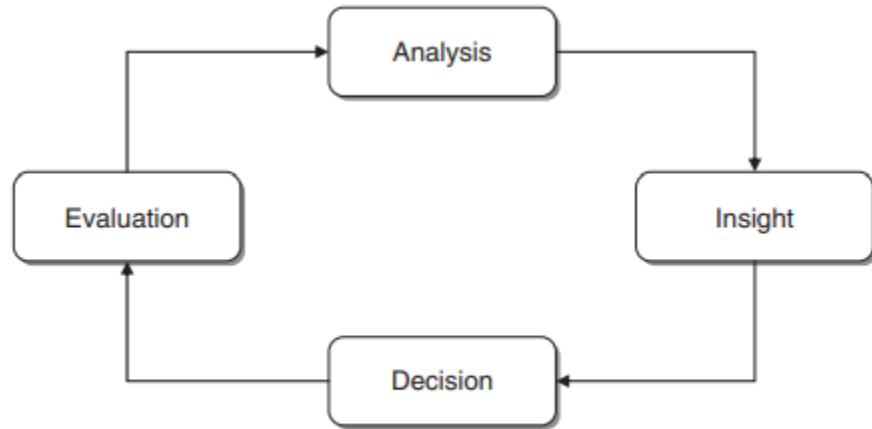


Figure 1.4 Departments of an enterprise concerned with business intelligence systems



*Figure 1.5 Cycle of a business intelligence analysis*

### **Analysis.**

- During the analysis phase, it is necessary to recognize and accurately spell out the problem at hand. Decision makers must then create a mental representation of the phenomenon being analyzed, by identifying the critical factors that are perceived as the most relevant.

### **Insight.**

- The second phase allows decision makers to better and more deeply understand the problem at hand.
- For instance, if the analysis carried out in the first phase shows that many customers are discontinuing an insurance policy upon yearly expiration, in the second phase it will be necessary to identify the profile and characteristics shared by such customers.
- The information obtained through the analysis phase is then transformed into knowledge during the insight phase.

### **Decision.**

- During the third phase, knowledge obtained as a result of the insight phase is converted into decisions and subsequently into actions.
- The availability of business intelligence methodologies allows the analysis and insight phases to be executed more rapidly so that more effective and timely decisions can be made that better suit the strategic priorities of a given organization. This leads to an overall reduction in the execution time of the analysis – decision – action – revision cycle, and thus to a decision-making process of better quality.

### **Evaluation.**

- Finally, the fourth phase of the business intelligence cycle involves performance measurement and evaluation.

- Extensive metrics should then be devised that are not exclusively limited to the financial aspects but also consider the major performance indicators defined for the different company departments.

## 1.8 DEVELOPMENT OF A BUSINESS INTELLIGENCE SYSTEM

- The development of a business intelligence system can be assimilated to a project, with a specific final objective, expected development times and costs, and the usage and coordination of the resources needed to perform planned activities.
- Figure shows the typical development cycle of a business intelligence architecture

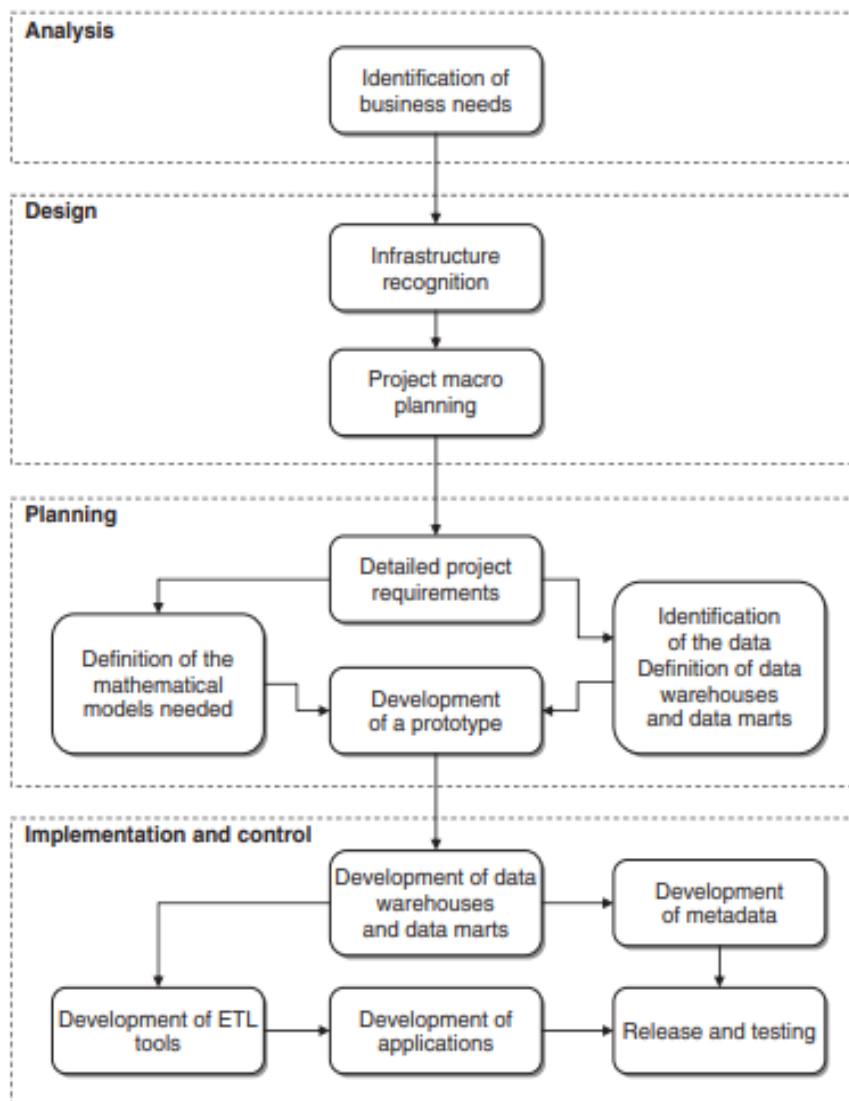


Figure 1.6 Phases in the development of a business intelligence system

**Analysis.**

- During the first phase, the needs of the organization relative to the development of a business intelligence system should be carefully identified.

**Design.**

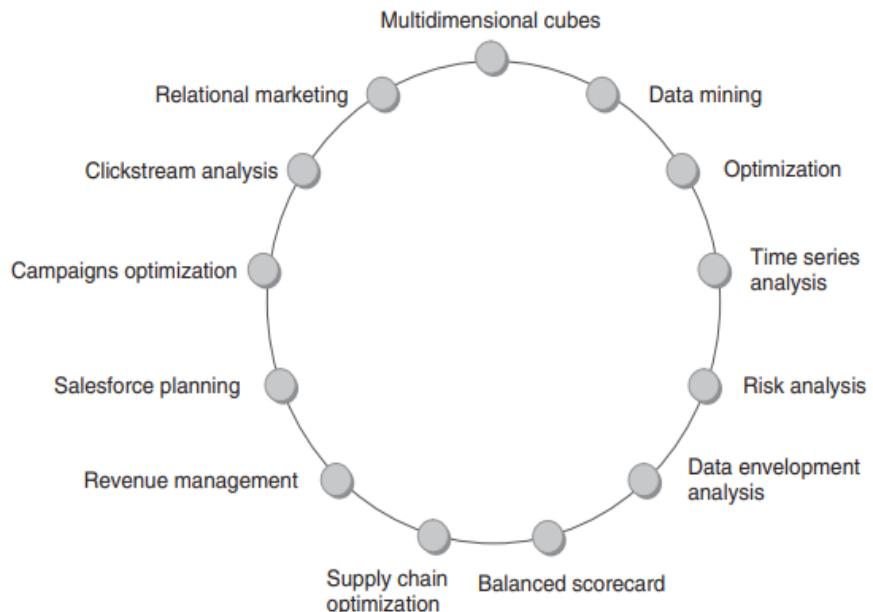
- The second phase includes two sub-phases and is aimed at deriving a provisional plan of the overall architecture, considering any development soon and the evolution of the system in the midterm.

**Planning.**

- The planning stage includes a sub-phase where the functions of the business intelligence system are defined and described in greater detail.
- Subsequently, existing data as well as other data that might be retrieved externally are assessed.

**Implementation and Control.**

- The last phase consists of five main sub-phases. First, the data warehouse and each specific data mart are developed. These represent the information infrastructures that will feed the business intelligence system.
- Figure provides an overview of the main methodologies that may be included in a business intelligence system, most of which will be described



*Figure 1.7 Portfolio of available methodologies in a business intelligence system*

## 1.9 ETHICS AND BUSINESS INTELLIGENCE

---

Business Intelligence

- The term ‘ethics’ defines the standards that bear on right and wrong issues of society.
- Business ethics is thus a set of professional standards, which emphasize principles of honesty and duty to the business and the general public.
- The adoption of business intelligence methodologies, data mining methods and decision support systems raises some ethical problems that should not be over-looked.
- Indeed, the progress toward the information and knowledge society opens countless opportunities but may also generate distortions and risks which should be prevented and avoided by using adequate control rules and mechanisms.

The other significant principles included in business ethics are:

- Fairness
- Integrity
- Commitment to agreements
- Broad-mindedness
- Considerateness
- Importance given to human esteem and self-respect
- Responsible citizenship
- Attempt to excel
- Accountability

---

## 1.10 SUMMARY

---

- The summarization of this chapter will allow us to understand about Business Intelligence.
- Also, how to deal with data and information and to convert it in knowledgeable data to be stored in business intelligent systems to be used by knowledge workers.
- To understand the development of BI systems and its implementation.

---

## 1.11 LIST OF REFERENCES

---

- Business Intelligence by Carlo Vercellis – Willey Publication (2009).
- Decision Support and Business Intelligence Systems by Efraim Turban, Ramesh Sharda, Dursun Delen – Pearson Publication (2011).
- Fundamental of Business Intelligence by Grossmann W, Rinderle-Ma – Springer (2015).

---

## 1.12 UNIT END EXERCISE

---

### Answer the following questions

- 1) What is business intelligence? Explain effective and timely decisions.
- 2) Write a short note on Data, Information and knowledge w.r.t BI.
- 3) Explain mathematical models and its advantages.
- 4) Explain business intelligence architecture with the help of a diagram.
- 5) Explain development of business intelligence system.
- 6) Explain Cycle of a business intelligence analysis.



# DECISION SUPPORT SYSTEMS

## Unit Structure

- 2.0 Objective
- 2.1 Introduction
- 2.2 An Overview
- 2.3 Definition of a System
- 2.4 Representation of the Decision-making process
  - 2.4.1 Rationality and problem solving
- 2.5 The decision-making process
- 2.6 Types of decisions
- 2.7 Approaches to the decision-making process
- 2.8 Decision Support System
- 2.9 Development of a Decision Support System
- 2.10 Summary
- 2.11 List of References
- 2.12 Unit End Exercise

---

## 2.0 OBJECTIVE

---

- To learn about Business Intelligence and how to take effective and timely decisions in an organization.
- To learn how to extract knowledge from data and information to take effective decisions.
- To learn how to draw conclusions, predictions and take futuristic actions.
- To learn the architecture of DSS system.

---

## 2.1 INTRODUCTION

---

- ***Business intelligence*** may be defined as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for **complex decision**.
- A *decision support system* (DSS) is an **interactive computer-based application** that combines **data** and **mathematical models** to help decision makers solve complex problems faced in managing the public and private enterprises and organizations.
- As described in Chapter 1, the analysis tools provided by business intelligence architecture can be regarded as DSSs capable of transforming data into information and knowledge helpful to decision makers.

- In this respect, DSSs are a basic component in the development of business intelligence architecture.

---

## 2.2 AN OVERVIEW

---

- In this chapter we will first discuss the structure of the decision-making process.
- Further on, the evolution of information systems will be briefly sketched.
- We will then define DSSs, outlining the major advantages and pointing out the critical success factors relative to their introduction.
- Finally, the development phases of a DSS project will be described, addressing the most relevant issues concerning its implementation.
- A *decision support system* (DSS) is an **interactive computer-based application** that combines **data** and **mathematical models** to help decision makers solve complex problems faced in managing the public and private enterprises and organizations.

---

## 2.3 DEFINITION OF A SYSTEM

---

- The term **system** is often used in everyday language: for instance, we refer to the solar system, the nervous system, or the justice system.
- The entities that we denominate **systems** share a common characteristic, which we will adopt as an abstract definition of the notion of system: each of them is made up of a set of components that are in some way connected to each other to provide a single collective result and a common purpose.
- Every system is characterized by boundaries that separate its internal components from the external environment.
- A system is said to be **open** if its boundaries can be crossed in both directions by flows of materials and information.
- When such flows are lacking, the system is said to be **closed**.
- In general terms, any given system receives specific input flows, carries out an internal transformation process and generates observable output flows.
- Figure 2.1 shows the structure that we will use as a reference to describe the concept of system.
- A system receives a set of *input* flows and returns a set of *output* flows through a *transformation* process regulated by *internal conditions* and *external conditions*.

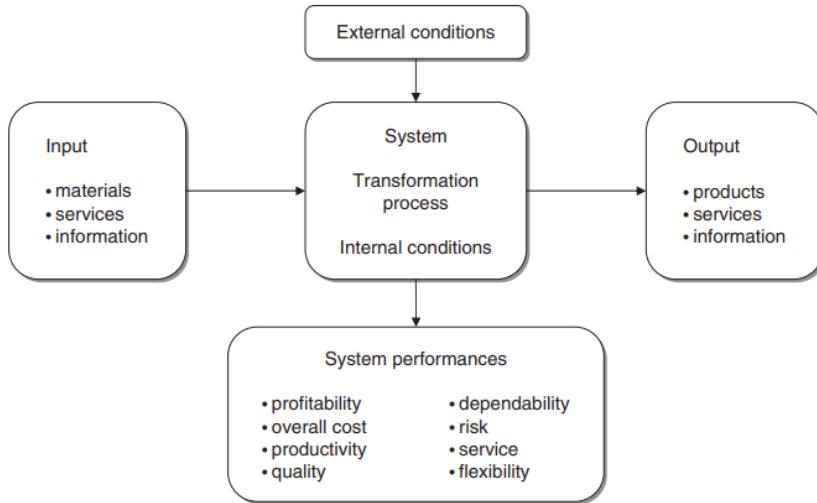


Figure 2.1 Abstract representation of a system

- A system will often incorporate a *feedback* mechanism.
- Feedback occurs when a system component generates an output flow that is fed back into the system itself as an input flow, possibly as a result of a further transformation.
- Systems that can modify their own output flows based on feedback are called

### ***Closed cycle systems.***

- For example, the closed cycle system outlined in Figure

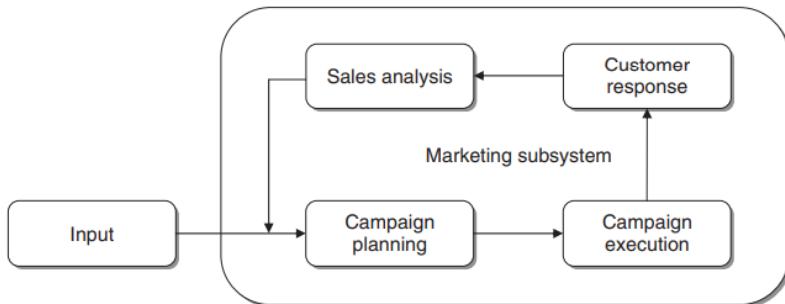


Figure 2.2 A closed cycle marketing system with feedback effects

- In connection with a decision-making process, whose structure will be described in the next section, it is often necessary to assess the performance of a system.
- For this purpose, it is appropriate to categorize the evaluation metrics into two main classes: *effectiveness* and *efficiency*.

## Effectiveness

- Effectiveness measurements express the level of conformity of a given system to the objectives for which it was designed.
- The associated performance indicators are therefore linked to the system output flows, such as production volumes, weekly sales and yield per share.

## Efficiency

- Efficiency measurements highlight the relationship between input flows used by the system and the corresponding output flows.
- Efficiency measurements are therefore associated with the quality of the transformation process. For example, they might express the number of resources needed to achieve a given sales volume.
- Effectiveness metrics indicate whether the *right* action is being carried out or not, while efficiency metrics show whether the action is being carried out in the *best* possible way or not.
- In order to build effective DSSs, we first need to describe in general terms how a decision-making process is articulated.

### 2.4.1 Rationality and Problem Solving

- A **decision** is a choice from multiple alternatives, usually made with a fair degree of rationality.
- In this section, we will focus on decisions made by knowledge workers in public and private enterprises and organizations.
- These decisions may concern the development of a strategic plan and imply therefore substantial investment choices, the definition of marketing initiatives and related sales.
- The decision-making process is part of a broader subject usually referred to as *problem solving*, which refers to the process through which individuals try to bridge the gap between the current operating conditions of a system (*as is*) and the supposedly better conditions to be achieved in the future (*to be*).
- Figure outlines the structure of the problem-solving process.

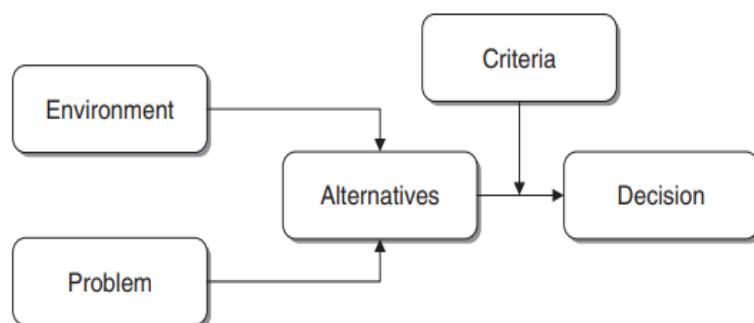


Figure 2.3 Logical flow of a problem-solving process

- *Criteria* are the measurements of effectiveness of the various alternatives and correspond to the different kinds of system performance.
- A *rational* approach to decision making implies that the option fulfilling the best performance criteria is selected out of all possible alternatives.
- Besides economic criteria, which tend to prevail in the decision-making process within companies, it is however possible to identify other factors influencing a rational choice.
- **Economic** - Economic factors are the most influential in decision-making processes and are often aimed at the minimization of costs or the maximization of profits. For example, an annual logistic plan may be preferred over alternative plans if it achieves a reduction in total costs.
- **Technical** - Options that are not technically feasible must be discarded. For instance, a production plan that exceeds the maximum capacity of a plant cannot be regarded as a feasible option.
- **Legal** - Legal rationality implies that before adopting any choice the decision makers should verify whether it is compatible with the legislation in force within the application domain.
- **Ethical** - Besides being compliant with the law, a decision should abide by the ethical principles and social rules of the community to which the system belongs.
- **Procedural** - A decision may be considered ideal from an economic, legal and social standpoint, but it may be unworkable due to cultural limitations of the organization in terms of prevailing procedures and common practice.
- **Political** - The decision maker must also assess the political consequences of a specific decision among individuals, departments and organizations.

---

## 2.5 THE DECISION-MAKING PROCESS

---

- A compelling representation of the decision-making process was proposed in the early 1960s, and remains today a major methodological reference.

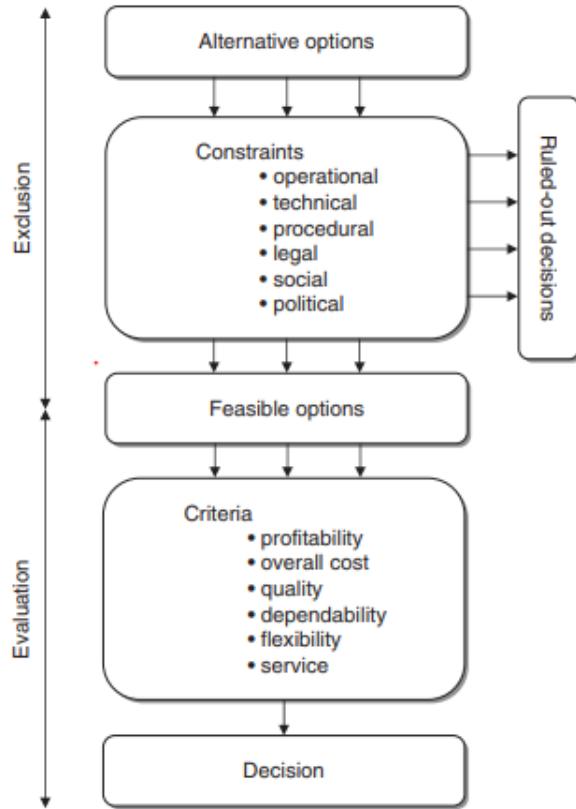


Figure 2.4 Logical structure of the decision-making process

- The model includes three phases, termed **intelligence, design and choice**.
- Figure shows an extended version of the original scheme, which results from the inclusion of two additional phases, namely **implementation and control**.

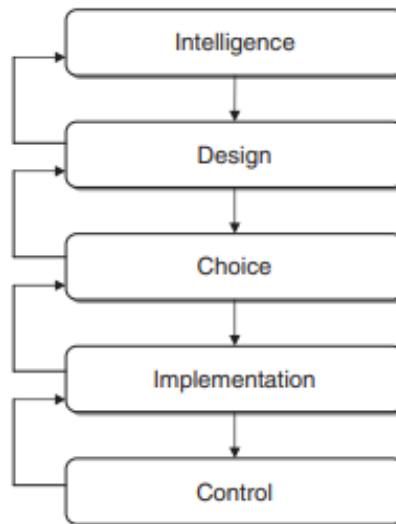


Figure 2.5 Phases of the decision-making process

- In the intelligence phase the task of the decision maker is to identify, circumscribe and explicitly define the problem that emerges in the system under study.
- The analysis of the context and all the available information may allow decision makers to quickly grasp the signals and symptoms pointing to a corrective action to improve the system performance

***Design***

- In the *design* phase, identification of problem should be developed and planned.
- At this level, the experience and creativity of the decision makers play a critical role, as they are asked to devise viable solutions that ultimately allow the intended purpose to be achieved.
- Decision makers can make an explicit enumeration of the alternatives to identify the best solution.

***Choice***

- Once the alternative actions have been identified, it is necessary to evaluate them based on the performance criteria deemed.
- Mathematical models and the corresponding solution methods usually play a valuable role during the *choice* phase.
- For example, optimization models and methods allow the best solution to be found in very complex solutions

***Implementation***

- When the best alternative has been selected by the decision maker, it is transformed into actions by means of an *implementation* plan.
- This involves assigning responsibilities and roles to all those involved into the actionplan.

***Control***

- Once the action has been implemented, it is finally necessary to verify and check that the original expectations have been satisfied and the effects of the action match the original intentions.
- In particular, the differences between the values of the performance indicators identified in the choice phase and the values observed at the end of the implementation plan should be measured.
- The results of these evaluations translate into experience and information, which are then transferred into the data warehouse to be used during subsequent decision-making processes.

The most relevant aspects characterizing a decision-making process can be briefly summarized as follows.

- Decisions are often devised by a group of individuals instead of a single decision maker.
- The number of alternative actions may be very high, and sometimes unlimited.
- The effects of a given decision usually appear later, not immediately.
- The decisions made within a public or private enterprise or organization are often interconnected and determine broad effects.
- During the decision-making process knowledge workers are asked to access data and information, and work on them based on a conceptual and analytical framework.
- Feedback plays an important role in providing information and knowledge for future decision-making processes within a given organization.
- In most instances, the decision-making process has multiple goals, with different performance indicators, that might also conflict with one another.
- Experiments carried out in a real-world system, according to a *trial-and-error* scheme, are too costly and risky to be of practical use for decision making.

---

## 2.6 TYPES OF DECISIONS

---

According to their nature, decisions can be classified as ***structured***, ***unstructured*** or ***semi-structured***.

### ***Structured decisions***

- A decision is structured if it is based on a well-defined and recurring decision-making procedure.
- In most cases structured decisions can be traced back to an algorithm, which may be explicit for decision makers and are therefore better suited for automation.
- More specifically, we have a structured decision if input flows, output flows and the transformations performed by the system can be clearly described in the three phases of intelligence, design and choice.

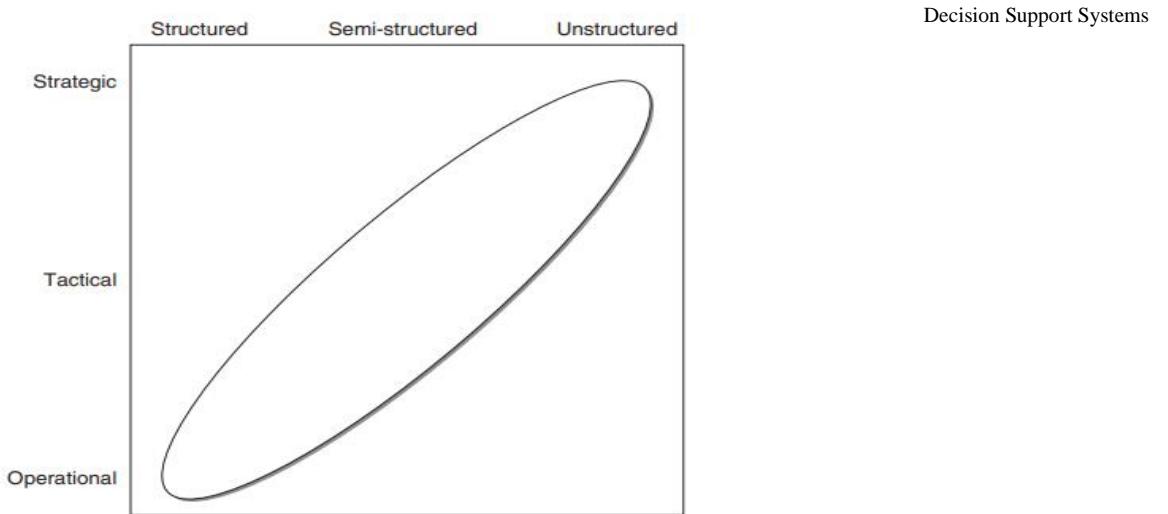


Figure 2.6 A taxonomy of decisions

### **Unstructured decisions**

- A decision is said to be unstructured if the three phases of intelligence, design and choice are also unstructured.
- This means that for each phase there is at least one element in the system (input flows, output flows and the transformation processes) that cannot be described in detail and reduced to a predefined sequence of steps.
- Such an event may occur when a decision-making process is implemented for the first time.

### **Semi-structured decisions**

- A decision is semi-structured when some phases are structured, and others are not.
- Most decisions faced by knowledge workers in managing public or private enterprises or organizations are semi-structured.
- Hence, they can take advantage of DSSs and a business intelligence environment primarily in two ways.
- For the unstructured phases of the decision-making process, business intelligence tools may offer a passive type of support.
- For the structured phases it is possible to provide an active form of support through mathematical models and algorithms that allow significant parts of the decision-making process to be automated.

### **Strategic Decisions**

- Decisions are strategic when they affect the entire organization or at least a substantial part of it for a long period of time.

- Strategic decisions strongly influence the general objectives and policies of an enterprise.
- Strategic decisions are taken at a higher organizational level, usually by the company top management.

### Tactical Decisions.

- Tactical decisions affect only parts of an enterprise and are usually restricted to a single department.
- Tactical decisions place themselves within the context determined by strategic decisions.
- In a company hierarchy, tactical decisions are made by middle managers, such as the heads of the company departments.

### Operational Decisions.

- Operational decisions refer to specific activities carried out within an organization and have a modest impact on the future.
- Operational decisions are framed within the elements and conditions determined by strategic and tactical decisions.
- Therefore, they are usually made at a lower organizational level.

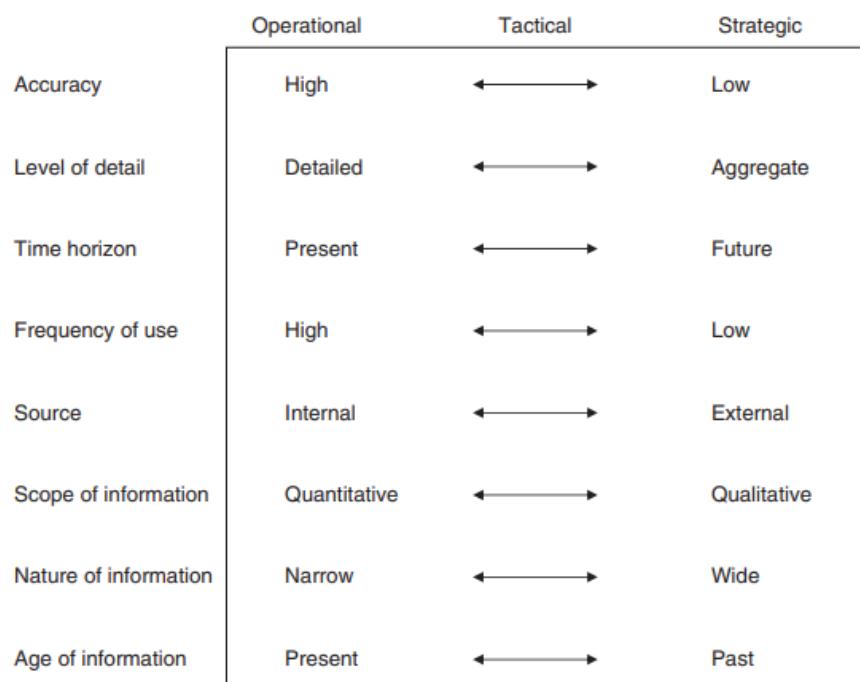


Figure 2.7 Characteristics of the information in terms of the scope of decisions

## 2.7 APPROACHES TO THE DECISION-MAKING PROCESS

---

There are two distinct approaches are used for decision making process.

- **Rational approach and**
- **Political-organizational approach.**

### **Rational approach**

- When a rational approach is followed, a decision maker considers major factors, such as economic, technical, legal, ethical, procedural and political, also establishing the criteria of evaluation to assess different options and then select the best decision.
- In this context, a DSS may help both in a passive way, through timely and versatile access to information, and in an active way, using mathematical models for decision making.

### **Political-organizational approach**

- When a political-organizational approach is pursued, a decision maker proceeds in amore instinctual and less systematic way.
- Decisions are not based on clearly defined alternatives and selection criteria.
- A DSS can only help in a passive way, providing timely and versatile access toinformation.
- It might also be useful during discussions and negotiations in those decision-making processes that involve multiple actors, such as managers operating in different departments.

Within the rational approach we can further distinguish between two alternative ways in which the actual decision-making process influences decisions:

- **Absolute rationality**
- **Bounded rationality**

### **Absolute Rationality**

- The term ‘absolute rationality’ refers to a decision-making process for which multiple performance indicators can be reduced to a single criterion, which therefore naturally lends itself to an optimization model.
- From a methodological perspective, this implies that multi-objective optimization problem is transformed into a single-objective problem by expressing all the relevant factors in a common measurement unit that allows the heterogeneous objectives to be added together.

### **Bounded rationality**

- Bounded rationality occurs whenever it is not possible to meaningfully reduce multiple criteria into a single objective, so that

the decision maker considers an option to be satisfactory when the corresponding performance indicators fall above or below prefixed threshold values.

- For instance, a production plan is acceptable if its cost is sufficiently low, the stock quantities are within a given threshold, and the service time is below customers' expectations.

## 2.8 DECISION SUPPORT SYSTEM

- A decision support system has been defined as an inter-active computer system helping decision makers to combine data and models to solve semi-structured and unstructured problems.
- This definition entails the three main elements of a DSS shown in, a database, a repository of mathematical models and a module for handling the dialogue between the system and the users.

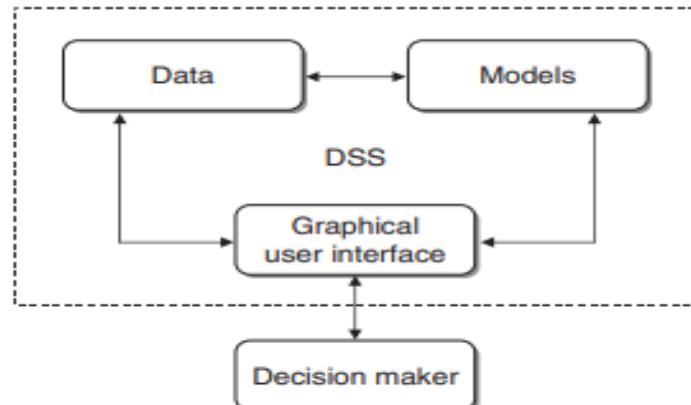


Figure 2.8 Structure of a decision support system

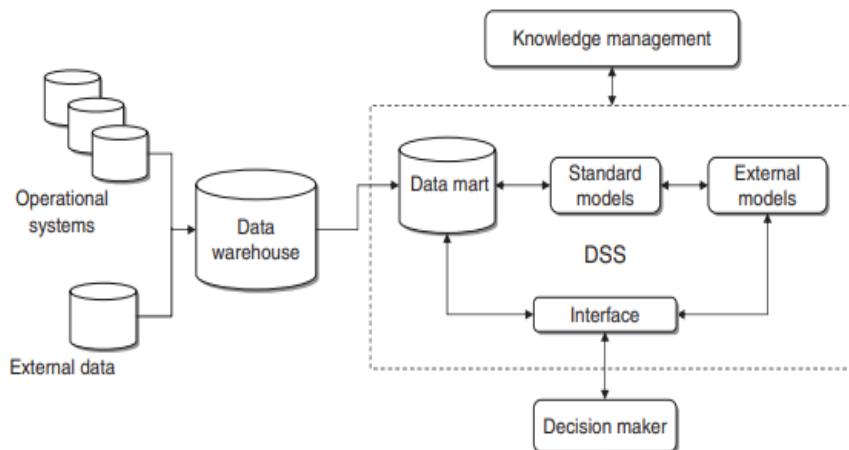


Figure 2.9 Extended structure of a decision support system

### **Effectiveness**

- Decision support systems should help knowledge workers to reach more effective decisions.
- Note that this does not necessarily imply an increased efficiency in the decision-making process.
- In fact, the adoption of a DSS may entail a more accurate analysis and therefore require a greater time investment by decision makers.
- However, the greater effort required will usually result in better decisions.

### **Mathematical models**

- In order to achieve more effective decisions, a DSS makes use of mathematical models, borrowed from disciplines such as operations research and statistics, which are applied to the data contained in data warehouses and data marts.
- The use of analytical models to transform data into knowledge and provide active support is the main characteristic that sets apart a DSS from a simple information system.

### **Integration in the decision-making process**

- A DSS should provide help for different kinds of knowledge workers, within the same application domain, particularly in respect of semi-structured and unstructured decision processes, both of an individual and a collective nature.
- Further, a DSS is intended for decision-making processes that are strategic, tactical and operational in scope.

### **Organizational role**

- In many situations the users of a DSS operate at different hierarchical levels within an enterprise, and a DSS tends to encourage communication between the various parts of an organization.
- By providing support for sequential and interdependent decision processes, a DSS can keep track of the analysis and the information that led to a specific decision.

### **Flexibility**

- A DSS must be flexible and adaptable in order to incorporate the changes required to reflect modifications in the environment or in the decision-making process.

- Moreover, it should be easy to use, with user-friendly and intuitive interaction methods and high-quality graphics for presenting the information extracted or generated.
- It is becoming increasingly common for DSSs to feature a web-browser interface to communicate with users.

### **Data management**

- The data management module includes a database designed to contain the data required by the decision-making processes to which the DSS is addressed.
- In most applications the database is a data mart.

### **Model management**

- The model management module provides end users with a collection of mathematical models derived from operations research, statistics and financial analysis.
- These are usually relatively simple models that allow analytical investigations to be carried out that are very helpful during the decision-making process

### **Interactions**

- In most applications, knowledge workers use a DSS interactively to carry out their analyses.

The module responsible for these interactions is expected to receive input data from users in the easiest and most intuitive way, usually through the graphic interface of a web browser.

### **Knowledge management**

- The knowledge management module is also interconnected with the company knowledge management integrated system.
- It allows decision makers to draw on the various forms of collective knowledge, usually unstructured, that represents the *corporate culture*.

### **Advantages of DSS:**

- An increase in the number of alternatives or options considered.
- An increase in the number of effective decisions devised.
- A greater awareness and a deeper understanding of the domain analyzed and the problems investigated.
- The possibility of executing scenario and what-if analyses by varying the hypotheses and parameters of the mathematical models.

- An improved ability to react promptly to unexpected events and unforeseen situations.
- A value-added exploitation of the available data.
- An improved communication and coordination among the individuals and the organizational departments.
- More effective development of teamwork.
- A greater reliability of the control mechanisms, due to the increased intelligibility of the decision process

## 2.9 DEVELOPMENT OF A DECISION SUPPORT SYSTEM

- Figure below shows the major steps in the development of a DSS.
- The logical flow of the activities is shown by the solid arrows.
- The dotted arrows in the opposite direction indicate revisions of one or more phases that might become necessary during the development of the system, through a feedback mechanism.

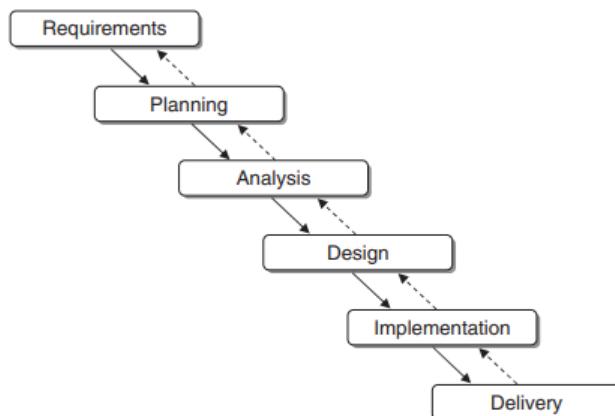


Figure 2.10 Phases in the development of a decision support system

### Planning

- The main purpose of the planning phase is to understand the needs and opportunities and to translate them into a project and later into a successful DSS.
- Planning usually involves a *feasibility study* to address the question: *Why do we wish to develop a DSS?*
- During the feasibility analysis, general and specific objectives of the system, recipients, possible benefits, execution times and costs are laid down.

## Analysis

- In the analysis phase, it is necessary to define in detail the functions of the DSS to be developed and achieved during the feasibility study.
- A response should therefore be given to the following question: *What should the DSS accomplish, and who will use it, when and how?*
- To provide an answer, it is necessary to analyze the decision processes to be supported, to try to thoroughly understand all interrelations existing between the problems addressed and the surrounding environment.

## Design

- During the design phase the main question is: *How will the DSS work?* The entire architecture of the system is therefore defined, through the identification of the hardware technology platforms, the network structure, the software tools to develop the applications and the specific database to be used.

## Implementation

- Once the specifications have been laid down, it is time for implementation, testing and the actual installation, when the DSS is rolled out and put to work.
- Any problems faced in this last phase can be traced back to project management methods.
- Effects should be monitored using *change management* techniques, making sure that no one feels excluded from the organizational innovation process and rejects the DSS.
- Sometimes a project may not come to a successful conclusion, may not succeed in fulfilling expectations, or may even turn out to be a complete failure.
- However, there are ways to reduce the risk of failure.
- The most significant of these is based on the use of rapid prototyping development where, instead of implementing the system, the approach is to identify a sequence of autonomous subsystems, of limited capabilities, and develop these subsystems step by step until the final stage is reached corresponding to the fully developed DSS.
- **Rapid prototyping** development offers clear advantages. Each subsystem can be developed more quickly and therefore is more readily available.

- The summarization of this chapter will allow us to understand about Business Intelligence and Decision Support Systems.
- With Business Intelligence and DSS Systems we can learn how to take effective and timely decisions in an organization.
- To learn how to extract knowledge from data and information to take effective decisions.
- To learn how to draw conclusions, predictions and take futuristic actions.
- To learn the architecture of DSS system.

---

## 2.11 LIST OF REFERENCES

---

- Business Intelligence by Carlo Vercellis – Willey Publication (2009).
- Decision Support and Business Intelligence Systems by Efraim Turban, Ramesh Sharda, Dursun Delen – Pearson Publication (2011).
- Fundamental of Business Intelligence by Grossmann W, Rinderle-Ma – Springer (2015).

---

## 2.12 UNIT END EXERCISE

---

### Answer the following questions

- 1) What is a System? Explain Open and Closed Systems with the help of a suitable diagram.
- 2) Explain decision and problem solving. List and explain various factors affecting decision making process.
- 3) Explain in brief the decision-making process with the help of diagram.
- 4) Explain in brief different types of decisions.
- 5) Explain in brief Decision Support System with the help of a diagram.
- 6) Explain the development of Decision Support System



## **MATHEMATICAL MODELS FOR DECISION MAKING**

### **Unit Structure**

- 3.0 Objectives
- 3.1 Mathematical Models for decision making
- 3.2 Structure of mathematical
- 3.3 Types of Mathematical Models
- 3.4 Development of a model
- 3.5 Classes of Models
- 3.6 Questions
- 3.7 Summary
- 3.8 Reference

---

### **3.0 OBJECTIVE**

---

After going through this unit, you will be able to,

- In this unit we will focus on the main characteristics shared by different mathematical models embedded into business intelligence systems.
- Role of a mathematical model in business intelligence.
- The structure of the mathematical model.
- The development phases of mathematical models.
- We also discuss data mining and its analysis methods.
- Data preparation, noise reduction, and data validation for the business intelligence model are also done in this unit.
- Data mining analysis methods to get information from raw data.

---

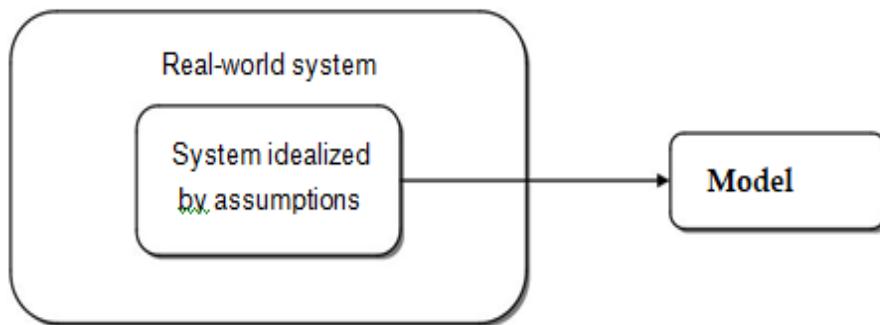
### **3.1 MATHEMATICAL MODELS FOR DECISION MAKING**

---

- Role of a mathematical model in business intelligence is to understand the function of systems.
- Explanation: The mathematical model is an abstract model which uses language to describe the behavior of the system.
- The mathematical modeling is classified into a black box or a white box model

## 3.2 STRUCTURE OF MATHEMATICAL MODELS

- Mathematical models have been developed and used in many application domains, from physics to architecture, engineering to economics.
- The models adopted in the various contexts differ substantially in terms of their mathematical structure.
- However, it is possible to identify a few fundamental features shared by most models.
- A model is a selective abstraction of a real system. In other words, a model is designed to analyze and understand from an abstract point of view the operating behavior of a real system, regarding which it only includes those elements deemed relevant for the investigation carried out.
- W.R.T. Einstein on the development of a model: ‘Everything should be made as simple as possible, but not simpler.’



**Figure 3.1** A model is a selective abstraction of reality

From Reference Book Business\_Intelligence\_-\_Carlo\_Vercellis

- Scientific and technological development has turned to mathematical models of various types for the abstract representation of real systems.

## 3.3 TYPES OF MATHEMATICAL MODELS

Models can be divided into the following.

### Iconic.

- An iconic model is a material representation of a real system, whose behavior is imitated for the analysis. A miniaturized model of a new city neighborhood is an example of an iconic model.

**Analogical.**

- An analogical model is also a material representation, although it imitates the real behavior by analogy rather than by replication. A wind tunnel built to investigate the aerodynamic properties of a motor vehicle is an example of an analogical model intended to represent the actual progression of a vehicle on the road.

**Symbolic.**

- A symbolic model, such as a mathematical model, is an abstract representation of a real system. It is intended to describe the behavior of the system through a series of symbolic variables, numerical parameters, and mathematical relationships.

**Stochastic.**

- In a stochastic model some input information represents random events and is therefore characterized by a probability distribution, which in turn can be assigned or unknown.

**Deterministic.**

- A model is called deterministic when all input data are supposed to be known *a priori* and with certainty.
- Since this assumption is rarely fulfilled in real systems, one resort to deterministic models when the problem at hand is sufficiently complex, and any stochastic elements are of limited relevance.
- Notice, however, that even for deterministic models the hypothesis of knowing the data with certainty may be relaxed.
- Sensitivity and scenario analyses, as well as what-if analysis, allow one to assess the robustness of optimal decisions to variations in the input parameters.

**Static.**

- Static models consider a given system and the related decision-making process within one single temporal stage.

---

**3.4 DEVELOPMENT OF A MODEL**

---

- It is possible to break down the development of a mathematical model for decision-making into four primary phases, shown in Figure below.
- The figure also includes a *feedback* mechanism that considers the possibility of changes and revisions of the model.

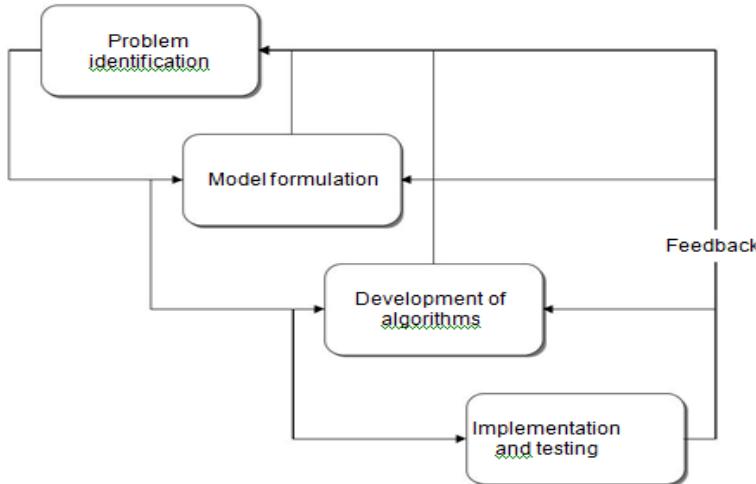


Figure 4.2 Phases in the development of mathematical models for decision making

From Reference Book Business\_Intelligence\_-\_Carlo\_Vercellis

## Problem identification

- First, the problem at hand must be correctly identified. The observed critical symptoms must be analyzed and interpreted to formulate hypotheses for investigation.

**A supposition or proposed explanation made based on limited evidence as a starting point for further investigation.**

- For example, too high a stock level, corresponding to an excessive stock turnover rate, may represent a symptom for a company manufacturing consumable goods.
- It is, therefore, necessary to understand what caused the problem, based on the opinion of the production managers.
- In this case, an ineffective production plan may be the cause of the stock accumulation.

## Model formulation

- Once the problem to be analyzed has been properly identified, effort should be directed toward defining an appropriate mathematical model to represent the system.
- Several factors affect and influence the choice of model, such as the *time horizon*, the *decision variables*, the *evaluation criteria*, the *numerical parameters*, and the *mathematical relationships*.

## Time Horizon.

Usually, a model includes a temporal dimension. For example, to formulate a tactical production plan over the medium term it is necessary to specify the production rate for each week in a year.

## **Evaluation criteria.**

Appropriate measurable performance indicators should be defined to establish a criterion for evaluating and comparing the alternative decisions.

These indicators may assume various forms in each different application, and may include the following factors:

- monetary costs and payoffs;
- effectiveness and level of service;
- quality of products and services;
- flexibility of the operating conditions;

## **Decision variables.**

Symbolic variables representing alternative decisions should then be defined. For example, if a problem consists of the formulation of a tactical production plan over the medium term, decision variables should express production volumes for each product, for each process, and for each period of the planning horizon.

## **Numerical parameters.**

It is also necessary to accurately identify and estimate all numerical parameters required by the model. In the production planning example, the available capacity should be known in advance for each process, as well as the capacity absorption coefficients for each combination of products and processes.

## **Mathematical relationships.**

The final step in the formulation of a model is the identification of mathematical relationships among the decision variables, the numerical parameters, and the performance indicators defined during the previous phases.

## **Development of Algorithms**

- Once a mathematical model has been defined, one will naturally wish to proceed with its solution to assess decisions and select the best alternative.
- In other words, a solution algorithm should be identified and a software tool that incorporates the solution method should be developed or acquired.
- An analyst in charge of the model formulation should possess a thorough knowledge of current solution methods and their characteristics.

- When a model is fully developed, then it is finally implemented, tested, and utilized in the application domain.
- It is also necessary that the correctness of the data and the numerical parameters entered in the model be assessed.
- These data usually come from a data warehouse or a data mart previously set up.
- Once the first numerical results have been obtained using the solution procedure devised, the model must be validated by submitting its conclusions to the opinion of decision-makers and other experts in the application domain. Several factors should be considered at this stage:
  - the plausibility and likelihood of the conclusions achieved;
  - the consistency of the results at extreme values of the numerical parameters;
  - the stability of the results when minor changes in the input parameters are introduced.

---

### 3.5 CLASSES OF MODELS

---

- There are several classes of mathematical models for decision-making, which in turn can be solved by several alternative solution techniques.
- Each model class is better suited to represent certain types of decision-making processes.
- In this section we will cover the main categories of mathematical models for decision-making, including:
  - Predictive models;
  - Pattern recognition and learning models;
  - Optimization models;
  - Project management models;
  - Risk analysis models;
  - Waiting, line models.

#### Predictive Models

- *Predictive* models play a primary role in business intelligence systems since they are logically placed upstream concerning other mathematical models and, more generally, to the whole decision-making process.
- Predictions allow input information to be fed into different decision-making processes, arising in strategy, research and development, administration and control, marketing, production, and logistics.

- Basically, all departmental functions of an enterprise make some use of predictive information to develop decision-making.

### **Pattern recognition and machine learning models**

- The purpose of pattern recognition and learning theory is to understand the mechanisms that regulate the development of intelligence, understood as the ability to extract knowledge from experience to apply it in the future.
- Mathematical learning models can be used to develop efficient algorithms that can perform such tasks.
- This has led to intelligent machines capable of learning from past observations and deriving new rules for the future, just like the human mind can do with great effectiveness due to the sophisticated mechanisms developed and fine-tuned during evolution.
- Mathematical learning models have two primary objectives.
- The purpose of *interpretation* models is to identify regular patterns in the data and to express them through easily understandable rules and criteria.

*Prediction* models help to forecast the value that a given random variable will assume in the future, based on the values of some variables associated with the entities of a database.

### **Optimization Models**

- Many decision-making processes faced by companies or complex organizations can be cast according to the following framework:
- given the problem at hand, the decision maker defines a set of *feasible* decisions and establishes a criterion for the evaluation and comparison of choices, such as monetary costs or payoffs.
- At this point, the decision maker must identify the *optimal* decision according to the evaluation criterion defined, that is, the choice corresponding to the minimum cost or the maximum payoff.
- In general, optimization models arise naturally in decision-making processes where a set of limited resources must be allocated most effectively to different entities.
- These resources may be personnel, production processes, raw materials, components, or financial factors.

### **Project management models**

- A *project* is a complex set of interrelated activities carried out with a specific goal, which may represent an industrial plant, a building, an information system, a new product, or a new organizational structure, depending on the different application domains.

- The execution of the project requires a planning and control process for the interdependent activities as well as the human, technical, and financial resources necessary to achieve the final goal.
- *Project management* methods are based on the contributions of various disciplines, such as business organization, behavioral psychology, and operations research.

### Risk analysis models

- Some decision problems can be described according to the following conceptual paradigm: the decision maker is required to choose among several available alternatives, having uncertain information regarding the effects that these options may have in the future
- For example, assume that senior management wishes to evaluate different alternatives to increase the company's production capacity. On the one hand, the company may build a new plant providing a high operating efficiency and requiring a high investment cost. On the other hand, it may expand an existing plant with a lower investment but with higher operating costs.

### Waiting for line models

- The purpose of *waiting for line* theory is to investigate congestion phenomena occurring when the demand for and provision of a service are stochastic in nature.
- If the arrival times of the customers and the duration of the service are not known beforehand in a deterministic way, conflicts may arise between customers in the use of limited shared resources. Therefore, some customers are forced to wait in line.
- A waiting line system is made up of three main components: a *source* generating a stochastic process in which entities, also referred to as customers, arrive at a given location to obtain a service; a set of *resources* providing the service; a *waiting area* able to receive the entities whose requests cannot immediately be satisfied.
- Waiting line models allow the performance of a system to be evaluated once its structure has been defined, and therefore are mostly useful within the system design phase.

---

## 3.6 QUESTIONS

---

### Descriptive Question

#### Chapter 3

- 1) Define the mathematical model? Explain the structure of the mathematical model.
- 2) Define the mathematical model? Explain its type.

- 3) Draw and explain the development process of the model.
- 4) List all the classes of the model and explain the following
  - a) Predictive models;
  - b) Pattern recognition and learning models;
  - c) Optimization models;
- 5) Explain the following classes of models used in business intelligence
  - a) Project management models;
  - b) Risk analysis models;
  - c) Waiting line models.

---

### 3.7 SUMMARY

---

A mathematical model is a description of a system using mathematical concepts and language. The process of developing a mathematical model is termed mathematical modeling.

Mathematical modeling is the process of describing a real-world problem in mathematical terms, usually in the form of equations, and then using these equations both to help understand the original problem and to discover new features about the problem.

---

### 3.8 REFERENCE

---

#### Books

- 1) Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis Wiley
- 2) Fundamental of Business Intelligence Grossmann W, Rinderle-Ma Springer

#### Websites

<https://technologyadvice.com/business-intelligence/>

<https://www.oracle.com/business-analytics/business-intelligence/>



# DATA MINING

## Unit Structure

- 4.0 Introduction
- 4.1 Data Mining
- 4.2 Definition of data mining
  - 4.2.1 Data Interpretation and Prediction
  - 4.2.2 Applications of Data Mining
- 4.3 Representation of input data
- 4.4 Data mining process
- 4.5 Analysis methodologies
  - 4.5.1 Clustering in Data Mining
  - 4.5.2 Association in Data Mining
  - 4.5.3 Data Cleansing
  - 4.5.4 Data Visualization
  - 4.5.5 Classification
  - 4.5.6 Machine Learning
- 4.6 Questions
- 4.7 Summary
- 4.8 Reference

---

## 4.0 INTRODUCTION

---

The term *data mining* indicates the process of exploration and analysis of a dataset, usually of large size, to find regular patterns, extract relevant knowledge, and obtain meaningful recurring rules.

OR

Data mining activities constitute an iterative process aimed at the analysis of large databases, to extract information and knowledge that may prove accurate and potentially useful for knowledge workers engaged in decision-making and problem-solving.

---

## 4.2 DEFINITION OF DATA MINING

---

The term *data mining* refers therefore to the overall process consisting of data gathering and analysis, development of inductive learning models, and adoption of practical decisions and consequent actions based on the knowledge acquired.

The data mining process is based on inductive learning methods, whose main purpose is to derive general rules starting from a set of available examples.

#### **4.2.1 Data Interpretation and Prediction**

Data mining activities can be subdivided into two major investigation streams, according to the main purpose of the analysis: *interpretation* and *prediction*.

##### **Interpretation.**

The purpose of interpretation is to identify regular patterns in the data and to express them through rules and criteria that can be easily understood by experts in the application domain.

The rules generated must be original and non-trivial to increase the level of knowledge and understanding of the system of interest.

For example, for a company in the retail industry, it might be advantageous to cluster those customers who have taken out loyalty cards according to their purchasing profile.

##### **Prediction.**

The purpose of prediction is to anticipate the value that a random variable will assume in the future or to estimate the likelihood of future events.

In a context, a retail company might predict the sales of a given product during the subsequent weeks or months.

Most data mining techniques derive their predictions from the value of a set of variables associated with the entities in a database.

#### **4.2.2 Applications of Data Mining**

Data mining methodologies can be applied to a variety of domains as follows.

##### **Relational marketing.**

Data mining applications in the field of relational marketing, have significantly contributed to the increase in the popularity of these methodologies. Some relevant applications within relational marketing are:

- identification of customer segments that are most likely to respond to targeted marketing campaigns, such as *cross-selling* and *up-selling*;
- prediction of the rate of positive responses to marketing campaigns;
- interpretation and understanding of the buying behavior of the customers;
- analysis of the products jointly purchased by customers, known as *market basket analysis*.

Fraud detection is another relevant field of application of data mining. Fraud may affect different industries such as telephony, insurance (false claims), and banking (illegal use of credit cards and bank checks; illegal monetary transactions).

**Risk evaluation.**

The purpose of risk analysis is to estimate the risk connected with future decisions. For example, using the past observations available, a bank may develop a predictive model to establish if it is appropriate to grant a monetary loan or a home loan, based on the characteristics of the applicant.

**Text mining.**

Data mining can be applied to different kinds of texts, which represent unstructured data, to classify articles, books, documents, emails, and web pages. Examples are web search engines or the automatic classification of press releases for storing purposes. Other text mining applications include the generation of filters for email messages and newsgroups.

**Image recognition.**

The treatment and classification of digital images, both static and dynamic, is an exciting subject both for its theoretical interest and the great number of applications it offers.

**Web mining.** Web mining applications, which will be briefly considered in section 13.1.9, are intended for the analysis of so-called *clickstreams* – the sequences of pages visited, and the choices made by a web surfer.

**Medical diagnosis.**

Learning models are an invaluable tool within the medical field for the early detection of diseases using clinical test results. Image analysis for diagnostic purposes is another field of investigation that is currently burgeoning.

---

## 4.3 REPRESENTATION OF INPUT DATA

---

In most cases, the input to a data mining analysis takes the form of a two-dimensional table, called a *dataset* irrespective of the actual logic and material representation adapted to store the information in files, databases, data warehouses, and data marts used as data sources.

The rows in the dataset correspond to the *observations* recorded in the past and are also called *examples*, *cases*, *instances*, or *records*.

The columns represent the information available for each observation and are termed *attributes*, *variables*, *characteristics*, or *features*.

The attributes contained in a dataset can be categorized as *categorical* or *numerical*, or *numerical*, depending on the type of values they take on.

**Categorical.**

Categorical attributes assume a finite number of distinct values, in most cases limited to less than a hundred, representing a qualitative property of an entity to which they refer.

**Numerical.** Numerical attributes assume a finite or infinite number of values and lend themselves to subtraction or division operations.

**Counts.** Counts are categorical attributes about which a specific property can be true or false.

**Nominal.** Nominal attributes are categorical attributes without a natural ordering.

**Ordinal.** Ordinal attributes, such as education level, are categorical attributes that lend themselves to a natural ordering but for which it makes no sense to calculate differences or ratios between the values.

**Discrete.** Discrete attributes are numerical attributes that assume a finite number or a countable infinity of values.

**Continuous.** Continuous attributes are numerical attributes that assume an uncountable infinity of values.

To represent a **generic dataset  $D$** , we will denote by  $m$  the **number of observations**, or **rows**, in the two-dimensional table containing the data and by  $n$  the **number of attributes**, or **columns**.

Furthermore, we will denote by

$$\mathbf{X} = [x_{ij}], \quad i \in M = \{1, 2, \dots, m\}, j \in N = \{1, 2, \dots, n\}, \quad (5.1)$$

the matrix of dimensions  $m \times n$  that corresponds to the entries in the dataset  $D$ . We will write

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in}) \quad (5.2)$$

$$\mathbf{a}_j = (x_{1j}, x_{2j}, \dots, x_{mj}) \quad (5.3)$$

for the  $n$ -dimensional row vector associated with the  $i$ th record of the dataset and the  $m$ -dimensional column vector representing the  $j$ th attribute in  $D$ , respectively.

## 4.4 DATA MINING PROCESS

The figure below shows the main phases of a generic data mining process.



**Figure 4.1** Data mining process

Data mining analyses are carried out in specific application domains and are intended to provide decision-makers with useful knowledge.

**Data gathering and integration.**

Once the objectives of the investigation have been identified, the gathering of data begins.

Data may come from different sources and therefore may require integration.

Data sources may be internal, external, or a combination of the two.

**Exploratory analysis.**

In the third phase of the data mining process, a preliminary analysis of the data is carried out to get acquainted with the available information and carry out *data cleansing*.

Usually, the data stored in a data warehouse are processed at loading time in such a way as to remove any *syntactical* inconsistencies.

For example, dates of birth that fall outside admissible ranges and negative sales charges are detected and corrected.

In the data mining process, data cleansing occurs at a *semantic* level.

**Attribute Selection.**

In the subsequent phase, the relevance of the different attributes is evaluated about the goals of the analysis.

Attributes that prove to be of little use are removed, to cleanse irrelevant information from the dataset.

New attributes obtained from the original variables through appropriate transformations are included in the dataset.

**Model development and validation.**

Once a high-quality high-quality dataset has been assembled with newly defined attributes, pattern recognition, and predictive models can be developed.

Usually, the *training* of the models is carried out using a sample of records extracted from the original dataset.

Then, the predictive accuracy of each model generated can be assessed using the rest of the data.

**Prediction and interpretation.**

The model selected among those generated during the development phase should be implemented and used to achieve the goals that were originally identified.

Moreover, it should be incorporated into the procedures supporting decision-making processes so that knowledge workers may be able to use it

to draw predictions and acquire a more in-depth knowledge of the phenomenon of interest.

The data mining process includes feedback cycles, represented by the dotted arrows in Figure, which may indicate a return to some previous phase, depending on the outcome of the subsequent phases.

## 4.5 ANALYSIS METHODOLOGIES

---

Data mining activities can be subdivided into a few major categories, based on the tasks and the objectives of the analysis.

They are *supervised* and *unsupervised* learning processes.

### **Supervised learning.**

In a supervised (or *direct*) learning analysis, a target attribute represents the class to which each record belongs as a second example of the supervised perspective, considering investment management.

A company wishes to predict the balance sheet of its customers based on their demographic characteristics and past investment transactions.

Supervised learning processes are therefore oriented toward prediction and interpretation concerning a target attribute.

### **Unsupervised learning.**

Unsupervised (or *indirect*) learning analyses are not guided by a target attribute.

Therefore, data mining tasks in this case are aimed at discovering recurring patterns in the dataset.

As an example, consider an investment management company wishing to identify clusters of customers who exhibit homogeneous investment behavior, based on data on past transactions.

In most unsupervised learning analyses, one is interested in identifying *clusters* of records that are similar within each cluster and different from members of other clusters.

**Clustering is a technique used to represent data visually — such as in graphs that show buying trends or sales demographics for a particular product.**

### **4.5.1 Clustering in Data Mining**

#### **What Is Clustering in Data Mining?**

Clustering refers to the process of grouping a series of different data points based on their characteristics. By doing so, data miners can seamlessly divide the data into subsets, allowing for more informed decisions in terms of broad demographics (such as consumers or users) and their respective behaviors.

Partitioning method: This involves dividing a data set into a group of specific clusters for the cluster for evaluation based on the criteria of each cluster. In this method, data points belong to just one group or cluster.

Hierarchical method: With the hierarchical method, data points are a single cluster, which is grouped based on similarities. These newly created clusters can then be analyzed separately from each other.

Density-based method: A machine learning method where data points plotted together are further analyzed, but data points by themselves are labeled “noise” and discarded.

Grid-based method: This involves dividing data into cells on a grid, which then can be clustered by individual cells rather than by the entire database. As a result, grid-based clustering has a fast-processing time.

Model-based method: In this method, models are created for each data cluster to locate the best data to fit that model.

### Examples of Clustering in Business

Clustering helps businesses manage their data more effectively. For example, retailers can use clustering models to determine which customers buy products, on which days, and with what frequency. This can help retailers target products and services to customers in a specific demographic or region.

Clustering can help grocery stores group products by a variety of characteristics (brand, size, cost, flavor, etc.) and better understand their sales tendencies. It can also help car insurance companies that want to identify a set of customers who typically have high annual claims to price policies more effectively. In addition, banks and financial institutions might use clustering to better understand how customers use in-person versus virtual services to better plan branch hours and staffing.

### 4.5.2 Association in Data Mining

Association rules are used to find correlations, or associations, between points in a data set.

#### What Is Association in Data Mining?

Data miners use association to discover unique or interesting relationships between variables in databases. The association is often employed to help companies determine marketing research and strategy.

#### Methods for Data Mining Association

Two primary approaches using association in data mining are the single-dimensional and multi-dimensional methods.

Single-dimensional association: This involves looking for one repeating instance of a data point or attribute. For instance, a retailer might search its database for the instances a particular product was purchased.

Multi-dimensional association: This involves looking for more than one data point in a data set. That same retailer might want to know more information than what a customer purchased — such as their age, method of purchase (cash or credit card), or age.

### Examples of Association in Business

The analysis of impromptu shopping behavior is an example of association — that is, retailers notice in data studies that parents shopping for childcare supplies are more likely to purchase specialty food or beverage items for themselves during the same trip. These purchases can be analyzed through statistical association.

Association analysis carries many other uses in business. For retailers, it's particularly helpful in making purchasing suggestions. For example, if a customer buys a smartphone, tablet, or video game device, association analysis can recommend related items like cables, applicable software, and protective cases.

Additionally, the association is used by the government to employ census data and plan for public services; it is also used by doctors to diagnose various illnesses and conditions more effectively.

### 4.5.3 Data cleaning

Data cleaning is the process of preparing data to be mined.

#### What Is Data Cleaning in Data Mining?

Data cleaning involves organizing data, eliminating duplicate or corrupted data, and filling in any null values. When this process is complete, the most useful information can be harvested for analysis.

#### Methods for Data Cleaning

Verifying the data: This involves checking that each data point in the data set is in the proper format (e.g. telephone numbers, social security numbers).

Converting data types: This ensures data is uniform across the data set. For instance, numeric variables only contain numbers, while string variables can contain letters, numbers, and characters.

Removing irrelevant data: This clears useless or inapplicable data so full emphasis can be placed on necessary data points.

Eliminating duplicate data points: This helps speed up the mining process by boosting efficiency and reducing errors.

Removing errors: This eliminates typing mistakes, spelling errors, and input errors that could negatively affect analysis outcomes.

Data Mining

Completing missing values: This provides an estimated value for all data and reduces missing values, which can lead to skewed or incorrect results.

### Examples of Data Cleaning in Business

According to Experian, 95 percent of businesses say they have been impacted by poor data quality. Working with incorrect data wastes time and resources increases analysis costs (because models need to be repeated), and often leads to faulty analytics.

## 4.5.4 Data Visualization

Data visualization is the translation of data into graphic form to illustrate its meaning to business stakeholders.

### What Is Data Visualization in Data Mining?

Data can be presented in visual ways through charts, graphs, maps, diagrams, and more. This is a primary way in which data scientists display their findings.

### Methods for Data Visualization

Many methods exist for representing data visually. Here are a few:

Comparison charts: Charts and tables express relationships in the data, such as monthly product sales over one year.

Maps: Data maps are used to visualize data about specific geographic locations. Through maps, data can be used to show population density and changes; compare populations of neighboring states, counties, and countries; detect how populations are spread over geographic regions; and compare characteristics in one region to those in other regions.

Heat maps: This is a popular visualization technique that represents data through different colors and shading to indicate patterns and ranges in the data. It can be used to track everything from a region's temperature changes to its food and pop culture trends.

Density plots: These visualizations track data over a period of period, creating what can look like a mountain range. Density plots make it easy to represent occurrences of single events over time (e.g., month, year, decade).

Histograms: These are like like density plots but are represented by bars on a graph instead of a linear form.

Network diagrams: These diagrams show how data points relate to each other by using a series of lines (or links) to connect objects.

Scatter plots: These graphs represent data point relationships on a two-variable axis. Scatter plots can be used to compare unique variables such as

a country's life expectancy or the amount of money spent on healthcare annually.

**Word clouds:** These graphics are used to highlight specific word or phrase instances appearing in a body of text; the larger the word's size in the cloud, the more frequent its use.

### Examples of Data Visualization in Business

Representing data visually is an important skill because it makes data readily understandable to executives, clients, and customers. According to Markets and Markets, the market size for global data visualization tools is expected to nearly double (to \$10.2 billion) by 2026.

Companies can make faster, more informed decisions when presented with data that is easy to understand and interpret. Today, this is typically accomplished through effective, visually accessible mediums such as graphs, 3D models, and even augmented reality. As a result, it's a good idea for aspiring data professionals to consider learning such skills through a data science and visualization boot camp.

#### **4.5.5 Classification**

Classification is a fundamental technique in data mining and can be applied to nearly every industry. It is a process in which data points from large data sets are assigned to categories based on how they're being used.

##### What Is Classification in Data Mining?

In data mining, classification is a form of clustering — that is, it is useful for extracting comparable points of data for comparative analysis. Classification is also used to designate broad groups within a demographic, target audience, or user base through which businesses can gain stronger insights.

##### Methods for Data Mining Classification

**Logistic regression:** This algorithm attempts to show the probability of a specific outcome within two possible results. For example, an email service can use logistic regression to predict whether an email is spam.

**Decision trees:** Once data is classified, follow-up questions can be asked, and the results diagrammed into a chart called a decision tree. For example, if a computer company wants to predict the likelihood of laptop purchases, it may ask, Is the potential buyer a student? The data is classified into "Yes" and "No" decision trees, with other questions to be asked afterward in a similar fashion.

**K-nearest neighbors (KNN):** This is an algorithm that tries to identify an unknown object by comparing it to others. For instance, grocery chains might use the K-nearest neighbors' algorithm to decide whether to include a sushi or hot meals station in their new store layout based on consumer habits in the local marketplace.

Naive Bayes: Based on the Bayes Theorem of Probability, this algorithm uses historical data to predict whether similar events will occur based on a different set of data.

Data Mining

Support Vector Machine (SVM): This machine learning algorithm is often used to define the line that best divides a data set into two classes. An SVM can help classify images and is used in facial and handwriting recognition software.

#### **4.5.6 Machine Learning**

Machine learning is the process by which computers use algorithms to learn on their own. An increasingly relevant part of modern technology, machine learning makes computers “smarter” by teaching them how to perform tasks based on the data they have gathered.

What Is Machine Learning in Data Mining?

In data mining, machine learning’s applications are vast. Machine learning and data mining fall under the umbrella of data science but aren’t interchangeable terms. For instance, computers perform data mining as part of their machine-learning functions.

Methods for Machine Learning

Supervised learning: In this method, algorithms train machines to learn using pre-labeled data with correct values, which the machines then classify on their own. It’s called supervised because the process trains (or “supervises”) computers to classify data and predict outcomes. Supervised machine learning is used in data mining classification.

Unsupervised learning: When computers handle unlabeled data, they engage in unsupervised learning. In this case, the computer classifies the data itself and then looks for patterns on its own. Unsupervised models are used to perform clustering and association.

Semi-supervised learning: Semi-supervised learning uses a combination of labeled and unlabeled data, making it a hybrid of the above models.

Reinforcement learning: This is a more layered process in which computers learn to make decisions based on examining data in a specific environment. For example, a computer might learn to play chess by examining data from thousands of games played online.

---

## **4.6 QUESTIONS**

---

### **Chapter 4**

1. What is Data mining? Explain its various applications.
2. Explain the Data Mining process in detail.
3. Difference between supervised and unsupervised learning.
4. Discuss different analysis methodologies for data mining (such as a supervised and unsupervised technique for data).
5. State and explain a few applications of data mining.

---

## 4.7 SUMMARY

---

- Data mining is a big area of data sciences, which aims to discover patterns and features in data, often large data sets. It includes regression, classification, clustering, detection of anomalies, and others. It also includes preprocessing, validation, summarization, and ultimately making sense of the data sets.
- Data Summarization is a simple term for a short conclusion of a big theory or a paragraph. This is something where you write the code and, in the end, you declare the result the result in the form of summarizing data. Data summarization has great importance in the data mining

---

## 4.8 REFERENCE

---

### Books

- 1) Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis Wiley
- 2) Fundamental of Business Intelligence Grossmann W, Rinderle-Ma Springer

### Websites

<https://technologyadvice.com/business-intelligence/>

<https://www.oracle.com/business-analytics/business-intelligence/>



# DATA PREPARATION

## Unit Structure

- 5.0 Introduction
- 5.1 Data Preparation
- 5.2 Data validation
  - 5.2.1 Incomplete Data
  - 5.2.2 Data Affected by Noise
- 5.3 Data transformation
  - 5.3.1 Standardization
  - 5.3.2 Feature Extraction
- 5.4 Data Reduction
  - 5.4.1 Sampling
  - 5.4.2 Feature selection
  - 5.4.3 Search Scheme
  - 5.4.4 Principal component analysis
  - 5.4.5 Data discretization
- 5.5 Questions
- 5.6 Summary
- 5.7 Reference

---

## 5.0 INTRODUCTION

---

Business intelligence systems and mathematical models for decision-making can achieve accurate and effective results only when the input data are highly reliable. However, the data extracted from the available primary sources and gathered into a data mart may have several anomalies which analysts must identify and correct.

This chapter deals with the activities involved in the creation of a high-quality dataset for subsequent use in business intelligence and data mining analysis. Several techniques can be employed to reach this goal: data validation, to identify and remove anomalies and inconsistencies; data integration and transformation, to improve the accuracy and efficiency of learning algorithms; data size reduction and discretization, to obtain a dataset with a lower number of attributes and records but which is as informative as the original dataset

## 5.1 DATA PREPARATION

---

Data preparation is the process of gathering, combining, structuring, and organizing data so it can be used in business intelligence (BI), analytics, and data visualization applications

## 5.2 DATA VALIDATION

---

The quality of input data may prove unsatisfactory due to incompleteness, noise, and inconsistency.

### 5.2.1 Incomplete Data

#### **Incompleteness.**

Some records may contain missing values corresponding to one or more attributes, and there may be a variety of reasons for this. It may be that some data were not recorded at the source systematically, or that they were not available when the transactions associated with a record took place. In other instances, data may be missing because of malfunctioning recording devices. It is also possible that some data were deliberately removed during previous stages of the gathering process because they were deemed incorrect. Incompleteness may also derive from a failure to transfer data from the operational databases to a data mart used for a specific business intelligence analysis.

#### **Noise.**

Data may contain erroneous or anomalous values, which are usually referred to as outliers. Other possible causes of noise are to be sought in malfunctioning devices for data measurement, recording, and transmission. The presence of data expressed in heterogeneous measurement units, which therefore require conversion, may in turn cause anomalies and inaccuracies.

#### **Inconsistency.**

Sometimes data contain discrepancies due to changes in the coding system used for their representation and therefore may appear inconsistent. For example, the coding of the products manufactured by a company may be subject to a revision taking effect on a given date, without the data recorded in previous periods being subject to the necessary transformations to adapt them to the revised encoding scheme.

The purpose of data validation techniques is to identify and implement corrective actions in case of incomplete and inconsistent data or data affected by noise.

To partially correct incomplete data one may adopt several techniques.

#### **Elimination.**

It is possible to discard all records for which the values of one or more attributes are missing. In the case of a supervised data mining analysis, it is

essential to eliminate a record if the value of the target attribute is missing. A policy based on the systematic elimination of records may be ineffective when the distribution of missing values varies irregularly across the different attributes, since one may run the risk of incurring a substantial loss of information.

Inspection.

Alternatively, one may opt for an inspection of each missing value, carried out by experts in the application domain, to obtain recommendations on possible substitute values. This approach suffers from a high degree of arbitrariness and subjectivity and is rather burdensome and time-consuming for large datasets. On the other hand, experience indicates that it is one of the most accurate corrective actions if skilfully exercised.

Identification.

As a third possibility, a conventional value might be used to encode and identify missing values, making it unnecessary to remove entire records from the given dataset. For example, for a continuous attribute that assumes only positive values it is possible to assign the value  $\{-1\}$  to all missing data. By the same token, for a categorical attribute one might replace missing values with a new value that differs from all those assumed by the attribute.

Substitution.

Several criteria exist for the automatic replacement of missing data, although most of them appear somehow arbitrary. For instance, missing values of an attribute may be replaced with the mean of the attribute calculated for the remaining observations. This technique can only be applied to numerical attributes, but it will be ineffective in the case of an asymmetric distribution of values. In a supervised analysis it is also possible to replace missing values by calculating the mean of the attribute only for those records having the same target class. Finally, the maximum likelihood value, estimated using regression models or Bayesian methods, can be used as a replacement for missing values. However, estimate procedures can become rather complex and time-consuming for a large dataset with a high percentage of missing data.

### 5.2.2 Data Affected by Noise

The term noise refers to a random perturbation within the values of a numerical attribute, usually resulting in noticeable anomalies. First, the outliers in a dataset need to be identified, so that subsequently either they can be corrected and regularized or entire records containing them are eliminated. In this section, we will describe a few simple techniques for identifying and regularizing data affected by noise.

The easiest way to identify outliers is based on the statistical concept of dispersion. The sample mean  $\mu^- j$  and the sample variance  $\sigma^- 2 j$  of the numerical attribute  $aj$  are calculated. If the attribute follows a distribution

that is not too far from normal, the values falling outside an appropriate interval centered around the mean value  $\bar{\mu}_j$  are identified as outliers, by the central limit theorem. More precisely, with a confidence of **100(1 -  $\alpha$ )%** (**approximately 96% for  $\alpha = 0.05$** ), it is possible to consider as outliers those values that fall outside the interval

$$(\bar{\mu}_j - za/2\sigma_j, \bar{\mu}_j + za/2\sigma_j),$$

where  $za/2$  is the  $\alpha/2$  quantile of the standard normal distribution. This technique is simple to use, although it has the drawback of relying on the critical assumption that the distribution of the values of the attribute is bell-shaped and roughly normal. However, by applying Chebyshev's theorem, described in Chapter 7, it is possible to obtain analogous bounds independent of the distribution, with intervals that are only slightly less stringent. Once the outliers have been identified, it is possible to correct them with values that are deemed more plausible or to remove an entire record containing them.

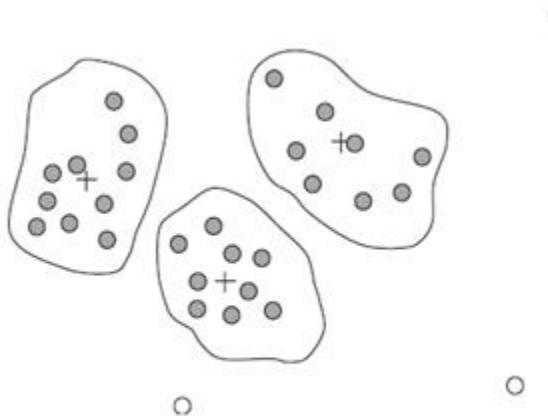


Figure 5.1 Identification of outliers using cluster analysis

From Reference Book Business\_Intelligence\_-\_Carlo\_Vercellis

An alternative technique, illustrated in Figure 6.1, is based on the distance between observations and the use of clustering methods. Once the clusters have been identified, representing sets of records having a mutual distance that is less than the distance from the records included in other groups, the observations that are not placed in any of the clusters are identified as outliers.

Clustering techniques offer the advantage of simultaneously considering several attributes, while methods based on dispersion can only consider every single attribute separately.

### 5.3 DATA TRANSFORMATION

In most data mining analyses, it is appropriate to apply a few transformations to the dataset to improve the accuracy of the learning models subsequently developed. Indeed, outlier correction techniques are

examples of transformations of the original data that facilitate subsequent learning phases. The principal component method, described in Section 6.3.3, can also be regarded as a data transformation process.

Data Preparation

### 5.3.1 Standardization

Most learning models benefit from a preventive standardization of the data, also called normalization. The most popular standardization techniques include the decimal scaling method, the min-max method, and the z-index method.

**Decimal scaling.** Decimal scaling is based on the transformation

$$\hat{x}_{ij} = (x_{ij} / 10^h)$$

where  $h$  is a given parameter that determines the scaling intensity. In practice, decimal scaling corresponds to shifting the decimal point by  $h$  positions toward the left. In general,  $h$  is fixed at a value that gives transformed values in the range  $[-1, 1]$ .

**Min-max.** Min-max standardization is achieved through the transformation

$$\hat{x}_{ij} = (x_{ij} - x_{\min,j}) / (x_{\max,j} - x_{\min,j}) * (\hat{x}_{\max} - \hat{x}_{\min}) + \hat{x}_{\min,j}$$

Where

$$x_{\min,j} = \min(x_{ij}), \quad x_{\max,j} = \max(x_{ij})$$

**z -index.** z-index based standardization uses the transformation

$$\hat{x}_{ij} = x_{ij} - \bar{\mu}_i / \bar{\sigma}_j$$

where  $\bar{\mu}_j$  and  $\bar{\sigma}_j$  are respectively the sample mean and sample standard deviation of the attribute  $a_j$ . If the distribution of values of the attribute  $a_j$  is roughly normal, the z-index-based transformation generates values that are almost certainly within the range  $(-3, 3)$ .

### 5.3.2 Feature Extraction

Standardization techniques aim to replace the values of an attribute with values obtained through an appropriate transformation. However, there are situations in which more complex transformations are used to generate new attributes that represent a set of additional columns in the matrix  $X$  representing the dataset  $D$ . Transformations of this kind are usually referred to as feature extraction. For example, suppose that a set of attributes indicate the spending of each customer over consecutive time intervals. It is then possible to define new variables capable of capturing the trends in the data

through differences or ratios between spending amounts of contiguous periods.

In other instances, the transformations may take even more complex forms, such as Fourier transforms, wavelets, and kernel functions. The use of such methods will be explained within the classification methods called support vector machines.

Attribute extraction may also consist of the creation of new variables that summarize within themselves the relevant information contained in a subset of the original attributes. For example, in the context of image recognition, one is often interested in identifying the existence of a face within a digitalized photograph. There are different indicators intended for the synthesis of each piece of information contained in a group of adjacent pixels, which makes it easier for classification algorithms to detect faces.

---

## 5.4 DATA REDUCTION

---

When dealing with a small dataset, the transformations described above are usually adequate to prepare input data for a data mining analysis. However, when facing a large dataset, it is also appropriate to reduce its size, to make learning algorithms more efficient, without sacrificing the quality of the results obtained.

There are three main criteria to determine whether a data reduction technique should be used: efficiency, accuracy, and simplicity of the models generated.

### **Efficiency.**

The application of learning algorithms to a dataset smaller than the original one usually means a shorter computation time. If the complexity of the algorithm is a superlinear function, as is the case for most known methods, the improvement in efficiency resulting from a reduction in the dataset size may be dramatic. As described in Chapter 5, within the data mining process it is customary to run several alternative learning algorithms to identify the most accurate model. Therefore, a reduction in processing times allows the analyses to be carried out more quickly.

### **Accuracy.**

In most applications, the accuracy of the models generated represents a critical success factor, and it is therefore the main criterion followed to select one class of learning methods over another. Therefore, data reduction techniques should not significantly compromise the accuracy of the model generated. As shown below, it may also be the case that some data reduction techniques, based on attribute selection, will lead to models with a higher generalization capability on future records.

In some data mining applications, concerned more with interpretation than with prediction, the models generated must be easily translated into simple rules that can be understood by experts in the application domain. As a trade-off for achieving simpler rules, decision-makers are sometimes willing to allow a slight decrease in accuracy. Data reduction often represents an effective technique for deriving models that are more easily interpretable.

#### Note

Data reduction can be pursued in three distinct directions, described below: a reduction in the number of observations through sampling, a reduction in the number of attributes through selection and projection, and a reduction in the number of values through discretization and aggregation.

#### **5.4.1 Sampling**

A further reduction in the size of the original dataset can be achieved by extracting a sample of observations that is significant from a statistical standpoint. This type of reduction is based on classical inferential reasoning. It is therefore necessary to determine the size of the sample that guarantees the level of accuracy required by the subsequent learning algorithms and to define an adequate sampling procedure. Sampling may be simple or stratified depending on whether one wishes to preserve in the sample the percentages of the original dataset concerning a categorical attribute that is considered critical.

Generally speaking, a sample comprising a few thousand observations is adequate to train most learning models. It is also useful to set up several independent samples, each of a predetermined size, to which learning algorithms should be applied. In this way, computation times increase linearly with the number of samples determined, and it is possible to compare the different models generated, in order to assess the robustness of each model and the quality of the knowledge extracted from data against the random fluctuations existing in the sample. It is obvious that the conclusions obtained can be regarded as robust when the models and the rules generated remain relatively stable as the sample set used for training varies.

#### **5.4.2 Feature selection**

The purpose of feature selection, also called feature reduction, is to eliminate from the dataset a subset of variables that are not deemed relevant for the data mining activities. One of the most critical aspects in a learning process is the choice of the combination of predictive variables more suited to accurately explain the investigated phenomenon.

Feature reduction has several potential advantages. Due to the presence of fewer columns, learning algorithms can be run more quickly on the reduced dataset than on the original one. Moreover, the models generated after the

elimination from the dataset of uninfluential attributes are often more accurate and easier to understand.

Feature selection methods can be classified into three main categories: filter methods, wrapper methods, and embedded methods.

### **Filter methods.**

Filter methods select the relevant attributes before moving on to the subsequent learning phase and phase and are therefore independent of the specific algorithm being used. The attributes deemed most significant are selected for learning, while the rest are excluded. Several alternative statistical metrics have been proposed to assess the predictive capability and relevance of a group of attributes. Generally, these are monotone metrics in that their value increases or decreases according to the number of attributes considered. The simplest filter method to apply for supervised learning involves the assessment of every single attribute based on its level of correlation with the target. Consequently, this led to the selection of the attributes that appear mostly correlated with the target.

### **Wrapper methods.**

If the purpose of the data mining investigation is classification or regression, and consequently performances are assessed mainly in terms of accuracy, the selection of predictive variables should be based not only on the level of relevance of every single attribute but also on the specific learning algorithm being utilized. Wrapper methods can meet this need since they assess a group of variables using the same classification or regression algorithm used to predict the value of the target variable. Each time, the algorithm uses a different subset of attributes for learning, identified by a search engine that works on the entire set of all possible combinations of variables, and selects the set of attributes that guarantees the best result in terms of accuracy. Wrapper methods are usually burdensome from a computational standpoint since the assessment of every possible combination identified by the search engine requires one to deal with the entire training phase of the learning algorithm. An example of the use of a wrapper method for attribute selection is given in Section 8.5 in the context of multiple linear regression models.

### **Embedded methods.**

For the embedded methods, the attribute selection process lies inside the learning algorithm, so that the selection of the optimal set of attributes is directly made during the phase of model generation. At each tree node, they use an evaluation function that estimates the predictive value of a single attribute or a linear combination of variables. In this way, the relevant attributes are automatically selected, selected, and they determine the rule for splitting the records in the corresponding node.

#### **5.4.3 Search Scheme**

Three distinct myopic search schemes can be followed: forward, backward, and forward–backward search.

**Forward.** According to the forward search scheme, also referred to as bottom-up search, the exploration starts with an empty set of attributes and subsequently introduces the attributes one at a time based on the ranking induced by the relevance indicator. The algorithm stops when the relevance index of all the attributes still excluded is lower than a prefixed threshold.

**Backward.** The backward search scheme, also referred to as top-down search, begins the exploration by selecting all the attributes and then eliminates them one at a time based on the preferred relevance indicator. The algorithm stops when the relevance index of all the attributes still included in the model is higher than a prefixed threshold.

**Forward–backward.** The forward–backward method represents a trade-off between the previous schemes, in the sense that at each step the best attribute among those excluded is introduced and the worst attribute among those included is eliminated. Also, in this case, threshold values for the included and excluded attributes determine the stopping criterion.

#### 5.4.4 Principal component analysis

*Principal component analysis* (PCA) is the most widely known technique of attribute reduction using projection. The purpose of this method is to obtain a projective transformation that replaces a subset of the original numerical attributes with a lower number of new attributes obtained as their linear combination, without this change causing a loss of information. Experience shows that a transformation of the attributes may lead in many instances to better accuracy in the learning models subsequently developed.

Before applying the principal component method, it is expedient to standardize the data, to obtain for all the attributes the same range of values, usually represented by the interval  $[-1, 1]$ . Moreover, the mean of each attribute  $a_j$  is made equal to 0 by applying the transformation

$$X_{ij} = X_{ij} - 1/m \sum X_{ij}$$

Let  $X$  denote the matrix resulting from applying the transformation (5.6) to the original data and let  $V = XX'$  be the covariance matrix of the attributes (for a definition of the covariance and variance matrices If the correlation matrix is used to develop the principal component analysis method instead of the covariance matrix, the transformation (5.6) is not required.

Starting from the  $n$  attributes in the original dataset, represented by the matrix  $X$ , the principal component method derives  $n$  orthogonal vectors, namely the principal components, which constitute a new basis of the space  $R^n$ . Principal components are better suited than the original attributes to explain fluctuations in the data, in the sense that usually a subset consisting of  $q$  principal components, with  $q < n$ , has an information content that is almost equivalent to that of the original dataset. Therefore, the original data are projected into a lower-dimensional space of dimension  $q$  having the same explanatory capability.

Principal components are generated in sequence using an iterative algorithm. The first component is determined by solving an appropriate optimization problem, to explain the highest percentage of variation in the data. At each iteration, the next principal component is selected, among those vectors that are orthogonal to all components already determined, as the one which explains the maximum percentage of variance not yet explained by the previously generated components. At the end of the procedure, the principal components are ranked in non-increasing order concerning the amount of variance that they can explain.

Let  $p_j$ ,  $j \in N$ , denote the  $n$  principal components, each of them being obtained as a linear combination  $p_j = Xw_j$  of the available attributes, where the weights  $w_j$  must be determined. The projection of a generic example  $x_i$  in the direction of the weights vector  $w_j$  is given by  $w_j x_i$ . It can easily be seen that its variance is given by

and applying the Karush–Kuhn–Tucker conditions, the solution of the maximization problem reduces to the solution of the system

$$\frac{\partial L(w_1, \lambda_1)}{\partial w_1} = 2Vw_1 - 2\lambda_1 w_1 = 0, \quad (6.10)$$

$$\frac{\partial L(w_1, \lambda_1)}{\partial \lambda_1} = 1 - w_1' w_1 = 0, \quad (6.11)$$

which can be rewritten as

$$(V - \lambda_1 I)w_1 = 0, \quad (6.12)$$

subject to the unit norm condition  $w_1' w_1 = 1$ , where  $I$  is the identity matrix. The solution of the maximization problem is therefore given by  $w_1 = u_1$ , where  $u_1$  is the eigenvector of unit norm associated with the maximum eigenvalue  $\lambda_1$  of the covariance matrix  $V$ . Since the variance we wish to maximize is given by

$$w_1' V w_1 = \lambda_1 w_1' w_1 = \lambda_1 u_1' u_1 = \lambda_1, \quad (6.13)$$

From Reference Book Business\_Intelligence\_-\_Carlo\_Vercellis

#### 5.4.5 Data discretization

The general purpose of data reduction methods is to obtain a decrease in the number of distinct values assumed by one or more attributes. Data discretization is the primary reduction method. On the one hand, it reduces continuous attributes to categorical attributes characterized by a limited number of distinct values. On the other hand, it aims to significantly reduce the number of distinct values assumed by the categorical attributes.

For instance, the weekly spending of a mobile phone customer is a continuous numerical value, which might be discretized into, say, five classes: low,  $[0 - 10)$  euros; medium-low,  $[10 - 20)$  euros; medium,  $[20 - 30)$  euros; medium-high,  $[30 - 40)$  euros; and high, over 40 euros.

As a further example applied to a categorical variable, consider the province of residence of each customer, and suppose it can assume a hundred distinct values. If instead of the province one uses the region of residence, the new attribute might take on twenty distinct values.

In both cases, the discretization process has brought about a reduction in the number of distinct values assumed by each attribute. The models that can be generated on the reduced dataset are likely to be more intuitive and less arbitrary. For instance, using a classification tree, it is possible to generate a rule of the form if spending is in the medium-low range, and if a customer resides in region A, then the probability of churning is higher than 0.85.

This is much more interpretable than the rule if spending is in the [12.21, 14.79] euro range, and if a customer resides in province B, then the probability of churning is higher than 0.85, which could have been generated for the original dataset.

The examples shown above suggest that discretization and reduction of the number of values taken by each attribute can improve the generalization capability of predictive models, thus making easier the interpretation of the rules obtained. Among the most popular discretization techniques are *subjective subdivision*, *subdivision into classes* and *hierarchical discretization*.

**Subjective subdivision.** Subjective subdivision is the most popular and intuitive method. Classes are defined based on the experience and judgment of experts in the application domain.

**Subdivision into classes.** Subdivision into categorical classes may be achieved in an automated way using the techniques described below. In particular, the subdivision can be based on classes of equal size or equal width.

### Hierarchical discretization.

The third type of discretization is based on hierarchical relationships between concepts and may be applied to categorical attributes, just as for the hierarchical relationships between provinces and regions. In general, given a hierarchical relationship of the one-to-many kind, it is possible to replace each value of an attribute with the corresponding value found at a higher level in the hierarchy of concepts.

### Subdivision into classes

The automated procedure of subdivision into classes consists of ordering in a non-decreasing way the values of the attribute  $\mathbf{a}_j$  and grouping them into a predetermined number  $K$  of contiguous classes. It is possible to form the classes of either different *size* or different *width*.

In the first case, the  $m$  observed values available for the attribute  $\mathbf{a}_j$  are distributed by placing  $_m/K$  or  $_m/K$  contiguous values in each class, so as to divide the  $m$  observed values almost equally among the  $K$  classes. In the second case, the range of total variation between the minimum value

and the maximum value taken by the attribute  $aj$  is subdivided into  $K$  contiguous intervals and the observed values are placed in the class corresponding to the interval where they fall. This second procedure is less effective if the distribution of the values significantly moves away from the uniform distribution. Once the  $K$  classes have been constructed, each observed value of  $aj$  is replaced by the average value of the corresponding class. As an alternative, instead of using the average value for regularization, it is possible to use the boundary value of the class that is closest to the original value taken by  $aj$ .

---

## 5.5 REVIEW QUESTIONS

---

### Descriptive Question

1. Explain the importance of Data validation.
2. What are Data transformation and Data discretization?
3. Write a note on data validation, data transformation, and data reduction.
4. Difference between structured and unstructured data?
5. Explain Data Reduction methods?

---

## 5.6 SUMMARY

---

Chapter 1 gives a basic introduction to the decision support system and how we can implement the DSS model in business intelligence.

This chapter 2 gives the introduction to data mining, analysis methods for data mining, and the development phases for the data mining model

Chapter 3 gives the basic of data preparation and how we remove noise data from the data sources. It also highlights the basic methods to validate the data

---

## 5.7 REFERENCE

---

- 1) Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis Wiley
- 2) Fundamental of Business Intelligence Grossmann W, Rinderle-Ma Springer

### Websites

<https://technologyadvice.com/business-intelligence/>

<https://www.oracle.com/business-analytics/business-intelligence/>



## 6

# CLASSIFICATION

## Unit Structure

- 6.0 Introduction
- 6.1 Classification Problems
- 6.1.1 Type of Classification Models
- 6.2 Evaluation of Classification Model
- 6.3 Bayesian Methods
  - 6.3.1 Naive Bayesian Classifiers
  - 6.3.2 Bayesian Networks
- 6.4 Logistic Regression
- 6.5 Neural Networks
  - 6.5.1 the Rosenblatt Perceptron
  - 6.5.3 Multi-level Feed-forward Networks
  - 6.5.4 Support Vector Machines
    - 6.5.4.1 Structural Risk Minimization
- 6.6 Summary
- 6.7 Unit End Questions
- 6.8 References

---

## 6.0 INTRODUCTION

---

Classification models are supervised learning approaches for predicting the value of a categorical target feature, in contrast to regression models, which deal with numerical characteristics. A collection of rules that allow the target class of incoming instances to be predicted given a set of past observations whose target class is known are created using classification models. Classification holds a significant place in learning theory due to its theoretical foundations and the wide range of applications it offers. The theoretical process of mimicking the inductive abilities of the human brain necessitates the development of algorithmic learning-based systems. On the other hand, categorization offers potential in a variety of application sectors. A few examples of practical problems that may be framed within the categorization paradigm are selecting the appropriate target market for a marketing campaign, fraud detection, picture identification, early illness diagnosis, text cataloguing, and spam email recognition.

## 6.1 CLASSIFICATION PROBLEMS

---

In a classification problem, a dataset  $D$  of  $m$  observations is characterized in terms of  $n$  explanatory qualities and a category target attribute. Predictive variables can be both categorized and numerical; they are also referred to as explanatory characteristics. The target property is also known as a class or label, whereas the observations are also called examples or instances. The target variable in classification models has fewer potential values than in regression models. We have a binary classification problem, namely, if there are only two classes, and a multiclass or multicategory classification problem, specifically, if there are more classes than two. Finding patterns in the connections between the explanatory variables used to identify samples that belong to the same class is the objective of a classification model. The class of cases for which simply the values of the explanatory attributes are known is predicted after the translation of these correlations into classification rules. The rules might take on a variety of forms depending on the type of model being used.

**Example 6.1** – Mobile phone industry retention. The binary classification issue in Example 6.1 on the study of customer loyalty in the mobile phone business has the goal attribute taking the value 1 if a client has ceased service and 0 otherwise. A classification model's goal is to extract general rules from the dataset's examples and then apply those rules to new instances for which the target value is unknown in order to categorize them. In this sense, a retention marketing campaign may be based on the classification model's ability to identify consumers who are likely to cancel their subscriptions.

**Example 6.2** – segmenting call center consumers by their demographics. Nowadays, a lot of industrial and service businesses have a contact center where clients may go to ask questions or report issues. It is helpful to categorize consumers based on the number of calls made to the call center in order to size the workforce and activities of a call center and to confirm the quality of the services provided. Through proper discretization of the numerical variable representing the number of calls, such as by setting, the desired attribute may be attained for example: class 0  $\equiv$  no calls, class 1  $\equiv$  1 call, class 2  $\equiv$  from 2 to 4 calls, class 4  $\equiv$  more than 4 calls. Once more, the properties of the consumers give the predicted attributes. As a result, the challenge of customer segmentation according to the volume of calls made to the call center involves many categories.

From a mathematical perspective, in a classification problem,  $m$  known examples are supplied, consisting of pairs  $(x_i, y_i)$ ,  $i \in M$ , where  $y_i \in H = \{v_1, v_2, \dots, v_H\}$  signifies the associated target class and  $x_i \in R^n$  is the vector of the values obtained by the  $n$  predictive attributes for the  $i^{\text{th}}$  case. A realization of the random variable  $X_i$ ,  $j \in N$ , which stands in for the attribute  $a_j$  in the dataset  $D$ , is in each component  $x_{ij}$  of the vector  $x_i$ . In a problem involving binary classification, one has  $H = 2$ , and the two classes may be denoted either by  $H = 0$  or  $H = 1$  without losing generality.

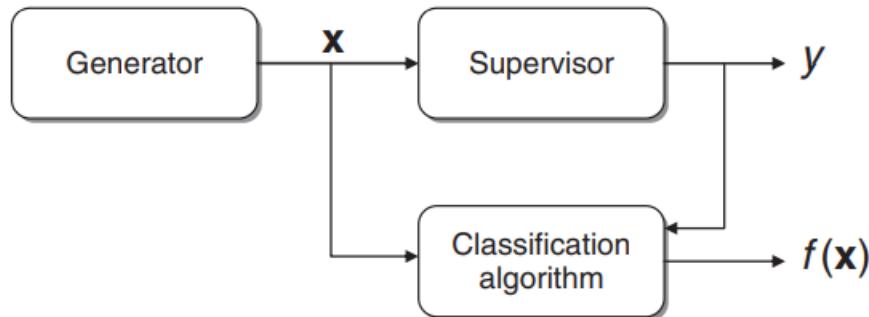
Let  $F$  be a class of functions  $f(x): \mathbb{R}^n \rightarrow H$  called hypotheses that represent hypothetical relationships of dependence between  $y_i$  and  $x_i$ . A classification problem consists of defining an appropriate hypothesis space  $F$  and an algorithm  $AF$  that identifies a function  $f^* \in F$  that can optimally describe the relationship between the predictive attributes and the target class. The joint probability distribution  $P_{x,y}(x,y)$  of the examples in the dataset  $D$ , defined over the space  $\mathbb{R}^n \times H$ , is generally unknown and most classification models are nonparametric, in the sense that they do not make any prior assumption on the form of the distribution  $P_{x,y}(x,y)$ .

The flow diagram shown in Figure 6.1 may clarify the probability assumptions concerning the three components of a classification problem: a generator of observations, a supervisor of the target class and a classification algorithm.

**Generator:** The generator's job is to generate random vectors  $x$  of instances based on an ambiguous probability distribution  $P_x(x)$ .

**Supervisor:** According to a conditional distribution, the supervisor returns the value of the target class for each vector  $x$  of examples  $P_{y|x}(y|x)$  which is also unknown.

**Algorithm:** A classification algorithm  $AF$ , also called a classifier, chooses a function  $f^* \in F$  in the hypothesis space to minimize a suitably defined loss function



**Figure 6.1:** Learning process for classification

As shown with broad reference to supervised learning techniques, classification models are applicable for both interpretation and prediction, much as regression models. While more complex models produce less understandable rules but typically produce more accurate predictions, simpler models typically produce clear classification rules that are simple to understand. The development of a classification model consists therefore of three main phases.

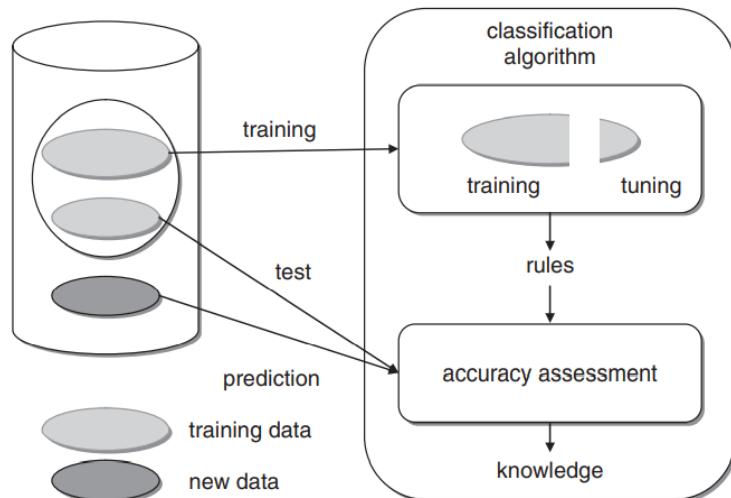
**Training phase:** During the training phase, the classification algorithm is applied to instances from a subset  $T$  of the dataset  $D$ , known as the training set, to produce classification rules that allow each observation  $x$  to be associated with the matching target class  $y$ .

**Test phase:** The rules developed during the training phase are utilized during the test phase to categorize the observations of  $D$  that were not part of the training set, for which the target class value is already known. To assess the accuracy of the classification model, the actual target class of each instance in the test set  $V = D - T$  is then compared with the class predicted by the classifier. The training set and the test set must be separate in order to prevent an overestimation of the model accuracy.

**Prediction phase:** The actual use of the classification model to categorize future additional data is represented by the prediction phase. Applying the rules created during the training phase to the explanatory variables that characterize the new instance yields a prediction.

### 6.1.1 Type of Classification Models

Before detailing the many types of classifiers in the next sections, it would be useful to present a taxonomy of classification models to place each unique algorithm in a broader perspective. We may divide categorization models into four primary categories.



**Figure 6.2:** Phases of the learning process for a classification algorithm

## 6.2 EVALUATION OF CLASSIFICATION MODEL

Before selecting the strategy that offers the greatest prediction accuracy, it is often wise to build alternative models for a classification analysis. One can get alternative models by using various approaches, such as classification trees, neural networks, Bayesian techniques, or support vector machines, as well as by altering the values of the pertinent parameters.

The following list of factors can be used to assess classification methods:

**Accuracy:** A classification model's accuracy must be assessed for two basic reasons. First, a model's precision is a measure of how well it can forecast the target class for upcoming data. It is also feasible to evaluate many

models based on their accuracy levels in order to choose the classifier that performs the best.

Classification

**Speed:** Some approaches can deal with more complex problems and have shorter computation times than others. However, a small-size training set obtained from numerous observations using random sampling schemes can be applied to classification methods with longer computation times. This approach is frequently used to achieve more precise categorization criteria.

**Robustness:** If the produced classification rules and the accompanying accuracy do not change considerably when the training set and test set are changed, and if the approach can manage missing data and outliers, the classification method is considered robust.

**Scalability:** The capacity of a classifier to learn from huge datasets is referred to as scalability, and it is inextricably linked to computing speed. The comments expressed in relation to sample strategies for data reduction, which frequently lead to rules having superior generalization capacity, thus also apply in this instance.

**Interpretability:** The rules produced by a classification analysis should be straightforward and easy enough for knowledge workers and application domain specialists to understand if the goal is to interpret as well as forecast.

---

## 6.3 BAYESIAN METHODS

---

The family of probabilistic classification models includes Bayesian approaches. Once the prior probability  $P(y)$  and the class conditional probabilities  $P(x|y)$  are known, they explicitly calculate the posterior probability  $P(y|x)$  that a given observation belongs to a specific target class using Bayes' theorem. Bayesian classifiers need the user to estimate the chance  $P(x|y)$  that a given observation may occur, provided it belongs to a certain class, in contrast to other methods presented in this chapter that are not reliant on probabilistic assumptions. A preliminary examination of the observations in the training set may thus be used to determine the learning phase of a Bayesian classifier and to predict the probability values needed to complete the classification assignment.

### 6.3.1 NAIVE BAYESIAN CLASSIFIERS

The foundation of naive Bayesian classifiers is the presumption that given the target class, the explanatory variables are conditionally independent. This hypothesis allows us to express the probability  $P(x|y)$  as

$$P(\mathbf{x}|y) = P(x_1|y) \times P(x_2|y) \times \cdots \times P(x_n|y) = \prod_{j=1}^n P(x_j|y).$$

Depending on the type of attribute being evaluated, it may be possible to estimate the probabilities  $P(x_j | y)$ ,  $j \in N$ , using instances from the training set.

### 6.3.2 BAYESIAN NETWORKS

By introducing some reticular hierarchical links through which it is possible to assign specific stochastic dependencies that application domain experts deem relevant, Bayesian networks, also known as belief networks, enable the hypothesis of conditional independence of the attributes to be relaxed. There are two key parts to a Bayesian network. The first is an acyclic directed graph where the arcs represent relationships of stochastic dependency, and the nodes represent the predictive variables. The variable  $X_j$  assigned to node  $a_j$  in the network is presumably dependent on the variables assigned to  $a_j$ 's predecessor nodes and conditionally independent of the variables assigned to  $a_j$ 's non-directly accessible nodes. A table with conditional probabilities allocated for each variable makes up the second part. The conditional distribution of  $P(X_j | C_j)$ , which is calculated using the relative frequencies in the dataset, is specifically shown in the table associated with the variable  $X_j$ .  $C_j$  is the set of explanatory variables linked to the predecessor nodes of node  $a_j$  in the network. In Bayesian networks, the computation simply reduces to those conditioning relationships determined by the precedence links present in the network, which reduces the complexity required to compute all possible combinations of predictor values for estimating the conditional probabilities, as already mentioned above.

---

### 6.4 LOGISTIC REGRESSION

---

By using an appropriate transformation, the technique of logistic regression may be used to turn binary classification issues into those of linear regression.

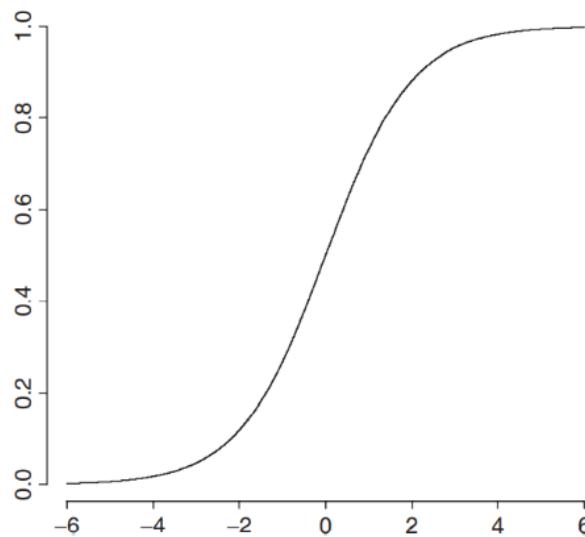
As in a binary classification problem, let's assume that the response variable  $y$  has values of 0 and 1. The logistic regression model asserts that a logistic function, denoted by, governs the posterior probability  $P(y|x)$  of the response variable conditional on the vector  $x$ .

$$P(y = 0|x) = \frac{1}{1 + e^{w'x}},$$

$$P(y = 1|x) = \frac{e^{w'x}}{1 + e^{w'x}}.$$

Here, we'll assume that the intercept has been added to the vector  $w$  and matrix  $X$ . Numerous statistical applications in the fields of economics and biology use the standard logistic function  $S(t)$ , also called the sigmoid function. It is defined as

$$S(t) = \frac{1}{1 + e^{-t}}.$$



**Figure 6.3:** Standard Logistic Function (Sigmoid)

The challenges associated with regression models, from which logistic regression models arise, are generally evident in logistic regression models. It is vital to go forward with attribute selection to prevent multicollinearity occurrences that put the significance of the regression results in jeopardy. Additionally, logistic regression models typically yield lower accuracy than other classifiers and demand more work during the model-development process. Finally, treating large datasets that have a lot of observations and attributes seems computationally challenging.

---

## 6.5 NEURAL NETWORKS

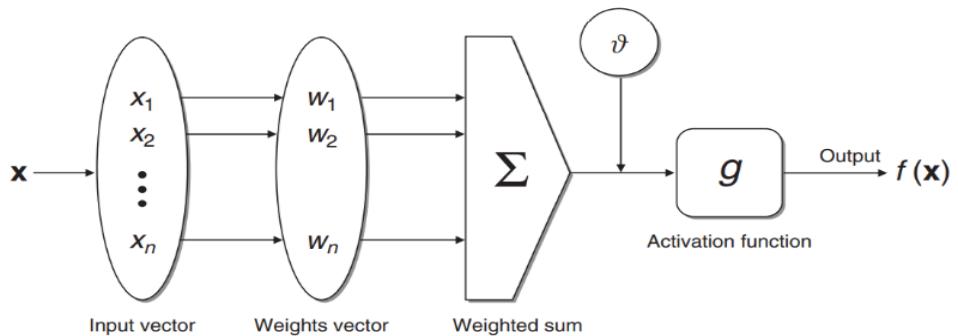
---

The purpose of neural networks is to mimic the behavior of neuron-based biological systems. Neural networks have been used for prediction since the 1950s, when the simplest models were put out. This includes classification as well as regression of continuous target characteristics. An oriented graph called a neural network is made up of nodes, which in a biological analogy correspond to neurons, and arcs, which stand in for dendrites and synapses. Each arc has a corresponding weight, and each node has a set activation function that is applied to the values it receives as input along the incoming arcs, with the arc weights considered. The training step is carried out by sequentially analyzing the observations present in the training set one after another and by altering the weights connected to the arcs at each iteration.

### 6.5.1 THE ROSENBLATT PERCEPTRON

The perceptron, shown in Figure below, is the simplest form of neural network and corresponds to a single neuron that receives as input the values  $(x_1, x_2, \dots, x_n)$  along the incoming connections, and returns an output value  $f(x)$ . The input values coincide with the values of the explanatory attributes, while the output value determines the prediction of the response variable  $y$ . Each of the  $n$  input connections is associated with a weight  $w_j$ . An activation function  $g$  and a constant  $\theta$ , called the distortion, are also assigned. Suppose

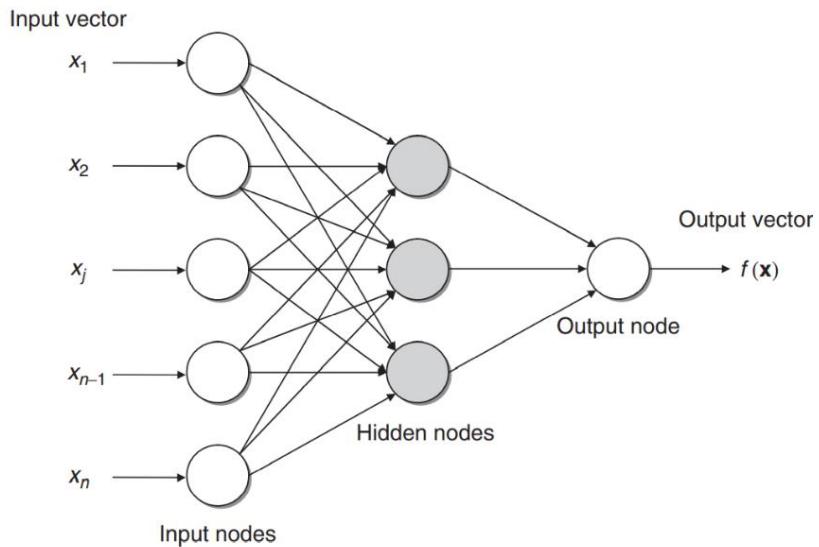
that the values of the weights and the distortion have already been determined during the training phase. The prediction for a new observation  $x$  is then derived by performing the following steps.



**Figure 6.4:** Operation of a single unit in a neural network

### 6.5.3 MULTI-LEVEL FEED-FORWARD NETWORKS

A multi-level feed-forward neural network is a more complex structure than the perceptron, since it includes the following components.



**Figure 6.5:** Neural Network

**Input nodes:** The values of the explanatory attributes for each observation are to be received as input by the input nodes. Typically, the number of explanatory variables is equal to the number of input nodes.

**Hidden nodes:** Inside the network, hidden nodes transform input values as instructed. Each node has connections to incoming arcs that come from input nodes or other hidden nodes, and it has connections to arcs that leave for output nodes or other hidden nodes.

**Output nodes:** Input or hidden node connections are sent to output nodes, which then send back an output value that matches the response variable's prediction. Most categorization issues have a single output node. The

network's nodes essentially behave as perceptron's since each node has a distortion coefficient and an activation function as well as certain weights that are assigned to the input arcs. In general, the activation function may assume forms that are more complex than the sign function  $\text{sgn}(\cdot)$ , such as a linear function, a sigmoid or a hyperbolic tangent. The backpropagation algorithm, which calculates the weights of all the arcs and the distortions at the nodes, employs reasoning not different from that of the single perceptron. The weights are initialized arbitrarily, for as by assigning their value to a collection of integers chosen at random.

#### 6.5.4 SUPPORT VECTOR MACHINES

A family of separation techniques for classification and regression called support vector machines was created in the framework of statistical learning theory. In numerous application fields, it has been demonstrated that they outperform conventional classifiers in terms of accuracy and that they are effectively scalable for big issues. The interpretation of the produced categorization rules is a further crucial aspect. A group of examples known as support vectors are found by support vector machines to be the most representative observations for each target class. They are somewhat more important than the other examples because they specify where the separating surface produced by the classifier is located in the attribute space.

##### 6.5.4.1 Structural risk minimization

As already observed, a classification algorithm AF defines an appropriate hypothesis space  $F$  and a function  $f^* \in F$  which optimally describes the relationship between the class value  $y$  and the vector of explanatory variables  $x$ . In order to describe the criteria for selecting the function  $f^*$ , let  $V(y, f(x))$  denote a loss function which measures the discrepancy between the values returned by the predictive function  $f(x)$  and the actual values of the class  $y$ . To select an optimal hypothesis  $f^* \in F$ , decision theory suggests minimizing the expected risk functional, defined as

$$R(f) = \frac{1}{2} \int V(y, f(x)) dP(x, y),$$

where  $P(x, y) = P_{x,y}(x, y)$  denotes the joint probability distribution over  $R^n \times H$  of the examples  $(x, y)$  from which the instances in the dataset  $D$  are assumed to be independently drawn. Since the distribution  $P(x, y)$  is generally unknown, in place of the expected risk one is naturally led to minimize the empirical risk over the training set  $T$ , defined as

$$R_{\text{emp}}(f) = \frac{1}{t} \sum_{i=1}^t V(y_i, f(x_i)).$$

---

## 6.6 SUMMARY

---

This chapter explains about the classification problems. How classification models are evaluated. The models are evaluated based on accuracy, speed, scalability, robustness and Interpretability. The classification models are Bayesian methods, Logistic regression, Neural networks and Support vector machines

---

## 6.7 UNIT END QUESTIONS

---

- 1) What is Classification? How is it evaluated?
- 2) Explain any two classification models in detail.
- 3) Discuss about multi-level feed-forward networks.

---

## 6.8 REFERENCES

---

- Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis, First Edition, Wiley
- Decision support and Business Intelligence Systems by Efraim Turban, Ramesh Sharda, Dursun Delen, ninth Edition, Pearson, 2011



# CLUSTERING

## Unit Structure

- 7.0 Introduction
- 7.1 Clustering Methods
  - 7.1.1 Taxonomy of Clustering Methods
  - 7.1.2 Affinity Measures
- 7.2 Partition Methods
  - 7.2.1 K-means Algorithm
  - 7.2.2 K-medoids Algorithm
- 7.3 Hierarchical Methods
- 7.4 Evaluation of Clustering Models
- 7.5 Summary
- 7.6 Unit End Questions
- 7.7 References

---

## 7.0 INTRODUCTION

---

Clustering techniques, which will be discussed in this chapter, are an example of the second class of unsupervised learning models. The goal of clustering methods is the identification of homogenous groupings of records known as clusters by specifying acceptable metrics and the induced ideas of distance and similarity between pairs of observations. The observations included in each cluster must be close to one another and remote from those found in other clusters, depending on the precise distance chosen. We discuss the key characteristics of clustering models at the beginning of this chapter. Following that, we will demonstrate the most common ways to measure the distance between pairs of observations in relation to the characteristics of the dataset's attributes. Then, partitioning techniques will be discussed with an emphasis on the K-means and K-medoids algorithms. In relation to the key metrics that describe the inhomogeneity among various clusters, we will finally illustrate both agglomerative and divisive hierarchical techniques. We will also go through several metrics for measuring the effectiveness of clustering methods.

---

## 7.1 CLUSTERING METHODS

---

The goal of clustering models is to partition the records of a dataset into clusters, which are homogenous groups of observations that are similar to one another and different from the observations contained in other groups. The human brain frequently uses a method of reasoning called affinity

grouping to organize objects. Additionally, because of this, clustering models have been used for a long time in a variety of fields, including social sciences, biology, astronomy, statistics, image recognition, handling digital data, marketing, and data mining. There are several uses for clustering models. The clusters produced may offer a useful understanding of the event in some applications of interest. For instance, categorizing clients based on their purchasing patterns may identify a cluster that corresponds to a market niche where it may be acceptable to focus marketing efforts for promotional purposes. In addition, a data mining project's preliminary phase can involve grouping data into clusters, which would be followed by using various approaches inside each cluster. In a retention study, a preliminary partition into clusters may be followed by the creation of unique classification models, with the goal of better identifying the clients with a high churn likelihood. To highlight outliers and find an observation that might stand in for an entire cluster on its own, grouping data into clusters may be useful during exploratory data analysis. This will help to reduce the size of the dataset. The following general criteria must be met by clustering methods:

**Flexibility:** Only numerical attributes can be used with some clustering techniques, and the distances between observations can be calculated using Euclidean metrics. The ability to analyze datasets with categorical features should be a feature of a flexible clustering technique, though. Euclidean metrics-based algorithms frequently produce spherical clusters and struggle to recognize more intricate geometrical patterns.

**Robustness:** The stability of the clusters produced about little variations in the values of each observation's attribute values is a sign of an algorithm's resilience. This characteristic makes sure that any noise that could be present in the data does not significantly impair the clustering procedure. Additionally, the clusters produced must be stable with respect to the dataset's observations' appearance order.

**Efficiency:** Because there are often quite a few observations in some applications, clustering algorithms must produce clusters quickly in order to ensure reasonable computation times for complex problems. When dealing with large datasets, one may also use the extraction of smaller samples to create clusters more quickly. This method inherently indicates a lesser resilience for the resulting clusters, though. In terms of the quantity of characteristics included in the dataset, clustering algorithms must also demonstrate their efficacy.

### 7.1.1 TAXONOMY OF CLUSTERING METHODS

Based on the reasoning behind how the clusters were created, clustering methods may be divided into four primary categories: partition methods, hierarchical methods, density-based approaches, and grid methods.

**Partition methods:** Develop a division of the provided dataset into a preset number  $K$  of non-empty subgroups using partition algorithms. They are applicable to datasets of small to medium size and are best suited for producing groups of a spherical or at most convex form.

**Hierarchical methods:** Based on a tree structure, hierarchical techniques perform several subdivisions into subsets and are distinguished by varying homogeneity thresholds inside each cluster and inhomogeneity thresholds across various clusters. Hierarchical algorithms do not demand that the number of clusters be predetermined, in contrast to partition techniques.

**Density-based methods:** Density-based approaches build clusters from the number of observations locally falling in a neighborhood of each observation, as opposed to the two preceding types of algorithms, which are based on the idea of distance between observations and between clusters. More specifically, a neighborhood with a particular diameter must contain numerous observations that must not be lower than a minimal threshold value for each record belonging to a certain cluster. Density-based techniques can locate non-convex clusters and efficiently separate any possible outliers.

**Grid methods:** By initially discretizing the space of the observations, grid techniques produce a grid structure made up of cells. Despite a decreased precision in the clusters produced, further clustering processes are developed with respect to the grid layout and often achieve reduced processing times. Regarding the techniques used to allocate the data to each individual cluster, there is a second distinction. Each observation can either be placed through superposition into many clusters or included entirely in one cluster. Additionally, fuzzy methods that assign observations to clusters with weights between 0 (the observation is completely unrelated to the cluster) and 1 (the observation exclusively belongs to the cluster) have been developed, with the additional requirement that the sum of the weights across all clusters be equal to 1. Finally, it is important to distinguish between full techniques of clustering, which place every observation in at least one cluster, and partial approaches, which could leave certain observations outside the clusters. The latter techniques are effective in locating outliers. The majority of clustering techniques are heuristic in nature, producing clusters of excellent quality but not always ideal size. Combinations would need to be looked at in an exhaustive way based on an extensive listing of the potential divisions of  $m$  observations into  $K$  clusters.

$$\frac{1}{K!} \sum_{h=1}^K \binom{K}{h} h^m.$$

Therefore, due to the exponential growth in computing time, it would be inapplicable even to small datasets. In terms of the difficulty of computation, clustering problems belong to the class of difficult (NP-hard) problems whenever  $K \geq 3$ .

### 7.1.2 AFFINITY MEASURES

The foundation of clustering models is often some metric of observational similarity. By specifying an acceptable concept of distance between each set of observations, this may frequently be achieved. In relation to the types of attributes being examined, we will review the most common metrics in this section. These qualities are listed below.

- Numerical attributes
- Binary attributes
- Nominal categorical attributes
- Ordinal categorical attributes
- Mixed composition attributes

## 7.2 PARTITION METHODS

Partitioning techniques divide a dataset D comprising m observations, each represented as a vector in n-dimensional space, into a number of non-empty subsets  $C = \{C_1, C_2, \dots, C_K\}$ , where  $K \leq m$ . The K number of clusters is often specified and assigned as an input to methods for partitioning data. In the sense that each observation only belongs to one cluster, the clusters produced by partition algorithms are often exhaustive and mutually exclusive. However, there are fuzzy partition methods that divide each observation into various clusters in accordance with a particular ratio. The initial assignment of the m available observations to the K clusters is the first step in the partitioning process. Then, they employ a reallocation technique repeatedly to move some observations to a different cluster and enhance the subdivision's overall quality. All the different criteria tend to indicate the degree of homogeneity of the observations belonging to the same cluster and their heterogeneity with respect to the records included in other clusters, even though alternative measures of the clustering quality might be utilized. Partition algorithms frequently reach their finish when no reallocation occurs within the same iteration and the subdivision appears stable in light of the chosen evaluation criterion. Therefore, partition methods are heuristic in nature because at each step they choose the option that initially seems to be most advantageous locally. This is typical of the class of so-called greedy methods. By doing this, at least for most of the datasets, there is no assurance that a globally optimal clustering will be achieved, simply that a decent subdivision will be. Two of the most well-known partition algorithms are the K-means and K-medoids methods, which will both be discussed in the paragraph that follows. They are effective clustering techniques that can identify spherical-shaped clusters.

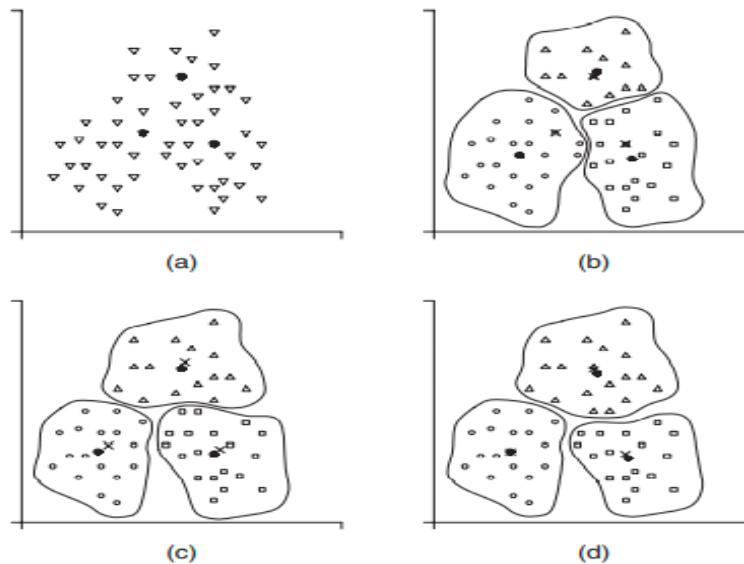
### 7.2.1 K-MEANS ALGORITHM

The K-means algorithm receives as input a dataset D, a number K of clusters to be generated and a function  $\text{dist}(x_i, x_k)$  that expresses the inhomogeneity.

between each pair of observations, or equivalently the matrix D of distances between observations. Given a cluster  $C_h$ ,  $h = 1, 2, \dots, K$ , the centroid of the cluster is defined as the point  $z_h$  having coordinates equal to the mean value of each attribute for the observations belonging to that cluster, that is,

$$z_{hj} = \frac{\sum_{x_i \in C_h} x_{ij}}{\text{card}\{C_h\}}.$$

- K observations are selected at random in D as the cluster centroids during the initialization phase.
- The cluster whose centroid is closest to the observation in terms of minimizing the distance from the record is iteratively allocated to each observation.
- If no observation is transferred from the previous iteration to a new cluster, the operation is finished.
- The method goes back to step 2 after computing the new centroid for each cluster as the mean of the values of the observations that belong to the cluster.



**Figure 7.1:** Application of K-means algorithm

The K-means method has been given several variations and expansions. Since the final subdivision into clusters is often strongly influenced by the initial assignment, it is helpful to generate a variety of initial random assignments, then derive a variety of clustering for each of them before selecting the best one. Since they take large numerical values in the quadratic objective function EQ, outliers can also have an impact on the outcome. Therefore, it is preferable to use the K-means method only after the outliers have been found and eliminated. Last but not least, it is feasible to use a posteriori technique to enhance the subdivision produced by the algorithm, such as by splitting one or more clusters and therefore raising the overall number K of clusters discovered. The number of clusters might, however, be decreased by merging two or more clusters into one.

## 7.2.2 K-MEDOIDS ALGORITHM

A variation of the K-means algorithm is known as partitioning around medoids or the K-medoids algorithm. To reduce the sensitivity of the partitions produced regarding the extreme values in the dataset, it is based on the usage of medoids rather than the means of the observations belonging to each cluster.

## 7.3 HIERARCHICAL METHODS

The foundation of hierarchical clustering techniques is a tree structure. They do not demand that the number of clusters be predetermined, unlike partition techniques. A dataset D comprising m observations and a matrix of distances  $\text{dist}(x_i, x_j)$  between all pairs of observations are therefore provided as its input. Most hierarchical algorithms use one of the following five possible methods to calculate the distance between two clusters: minimum distance, maximum distance, mean distance, distance between centroids, and Ward distance. Suppose that we wish to calculate the distance between two clusters  $C_h$  and  $C_f$  let  $z_h$  and  $z_f$  be the corresponding centroids.

**Minimum distance:** The single linkage criteria, also known as the minimal distance criterion, states that the dissimilarity between two clusters is determined by the least distance between all pairs of observations where one observation belongs to the first cluster and the other to the second cluster.

$$\text{dist}(C_h, C_f) = \min_{\substack{x_i \in C_h \\ x_k \in C_f}} \text{dist}(x_i, x_k).$$

**Maximum distance:** The greatest distance between all pairs of observations where one observation belongs to the first cluster and the other to the second cluster, also known as the full linkage criteria, is what determines how different two clusters are from one another.

$$\text{dist}(C_h, C_f) = \max_{\substack{x_i \in C_h \\ x_k \in C_f}} \text{dist}(x_i, x_k).$$

**Mean distance:** The mean of the distances between all pairs of observations belonging to the two clusters, or the mean distance criterion, quantifies the dissimilarity between two clusters,

$$\text{dist}(C_h, C_f) = \frac{\sum_{x_i \in C_h} \sum_{x_k \in C_f} \text{dist}(x_i, x_k)}{\text{card}\{C_h\} \text{card}\{C_f\}}.$$

**Distance between centroids:** The centroids representing the two clusters are separated by the centroids-based criteria, which establishes the dissimilarity between the two clusters that is,

$$\text{dist}(\mathbf{C}_h, \mathbf{C}_f) = \text{dist}(\mathbf{z}_h, \mathbf{z}_f).$$

The Ward distance criterion is a little more complicated than the criteria mentioned above since it is based on an examination of the variance of the Euclidean distances between the observations. In fact, the method must first determine the total squared distance between each pair of observations that makes up a cluster. The total variance is then calculated for each pair of clusters that might be combined at the present iteration as the sum of the two variances between the distances in each cluster, which were assessed in the first phase. The two clusters linked to the lowest overall variance are finally combined. The Ward distance-based techniques sometimes produce several clusters with only a few observations in each. Agglomerative and divisive methods are the two basic categories into which hierarchical approaches may be categorized.

## 7.4 EVALUATION OF CLUSTERING MODELS

---

As we've seen in earlier chapters, the evaluation of the predicted accuracy for supervised learning techniques like classification, regression, and time series analysis is a crucial step in the construction of a model and is based on a set of precise numerical metrics. The same is not true for unsupervised learning models or clustering techniques in general. Although the evaluation of an unsupervised model is less direct and intuitive in the absence of a target attribute, it is still possible to define reasonable measures of quality and significance for clustering techniques. It is initially required to confirm that the clusters produced match to an actual regular pattern in the data before evaluating a clustering approach. Therefore, it is appropriate to use additional clustering algorithms and to contrast the outcomes produced by various techniques. This makes it easy to assess the robustness of the number of discovered clusters in relation to the various strategies used. At a subsequent phase it is recommended to calculate some performance indicators. Let  $C = \{C_1, C_2, \dots, C_K\}$  be the set of  $K$  clusters generated. An indicator of homogeneity of the observations within each cluster  $C_h$  is given by the cohesion, defined as

$$\text{coh}(C_h) = \sum_{\substack{\mathbf{x}_i \in C_h \\ \mathbf{x}_k \in C_h}} \text{dist}(\mathbf{x}_i, \mathbf{x}_k).$$

The overall cohesion of the partition C can therefore be defined as

$$\text{coh}(\mathcal{C}) = \sum_{C_h \in \mathcal{C}} \text{coh}(C_h).$$

If one clustering has a lower overall cohesion, it is preferable to another in terms of homogeneity within each cluster. The spacing between two clusters is a sign of inhomogeneity between them, defined as

$$\text{sep}(C_h, C_f) = \sum_{\substack{\mathbf{x}_i \in C_h \\ \mathbf{x}_k \in C_f}} \text{dist}(\mathbf{x}_i, \mathbf{x}_k).$$

Again, the overall separation of the partition C can be defined as

$$\text{sep}(\mathcal{C}) = \sum_{\substack{C_h \in \mathcal{C} \\ C_f \in \mathcal{C}}} \text{sep}(C_h, C_f).$$

If one grouping has a higher overall separation than another, it is preferred in terms of inhomogeneity among all clusters. The silhouette coefficient, which combines cohesiveness and separation, provides another indication of the clustering quality. Three stages need be taken in order to determine the silhouette coefficient for a single observation  $x_i$ , as described in the procedure presented as:

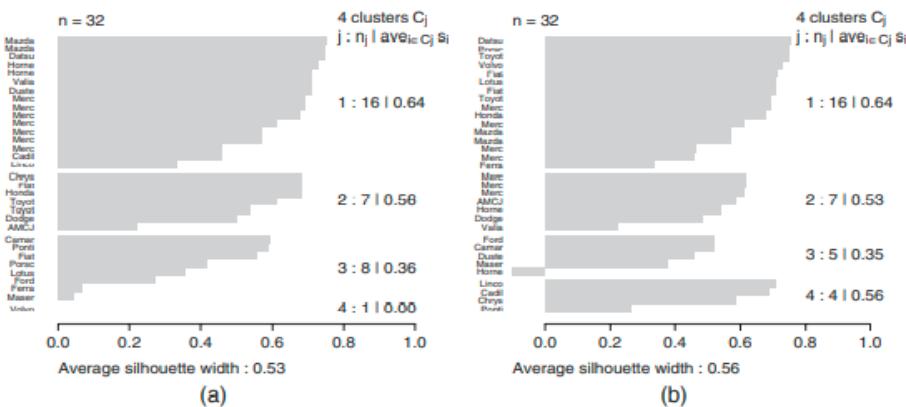
**Procedure:**

1. The mean distance  $u_i$  of  $x_i$  from all the remaining observations belonging to the same cluster is computed.
2. For each cluster  $C_f$  other than the cluster to which  $x_i$  belongs, the mean distance  $w_{if}$  between  $x_i$  and all the observations in  $C_f$  is calculated. The minimum  $v_i$  among the distances  $w_{if}$  is determined by varying the cluster  $C_f$ .
3. The silhouette coefficient of  $x_i$  is defined as

$$\text{silh}(\mathbf{x}_i) = \frac{v_i - u_i}{\max(u_i, v_i)}.$$

The silhouette coefficient varies between  $-1$  and  $1$ . If the value is negative, the membership of the observation  $x_i$  in its cluster is not well understood because the mean distance  $u_i$  of the observation  $x_i$  from the points of its cluster is greater than the minimum value  $v_i$  of the mean distances from the observations of the other clusters. This sees  $x_i$  undesirable. The silhouette coefficient should ideally be positive and  $u_i$  ought to be as near to  $0$ . Finally, it should be noted that the mean of the silhouette coefficients for all the observations in the dataset  $D$  may be used to compute the overall silhouette coefficient of a clustering. By using silhouette diagrams, which split the observations into clusters on the vertical axis and display the values of the silhouette coefficient for each cluster on the horizontal axis, silhouette coefficients may be visualized. In a silhouette diagram, in addition to the overall mean for the whole dataset, the mean value of the silhouette coefficient for each cluster is often also included.

Figures given below show the silhouette diagrams corresponding to different clustering. Figure 7.2(a) shows the silhouette diagram corresponding to a cut of four clusters in the dendrogram shown in Figure 7.2(a), obtained by applying a divisive hierarchical algorithm to the mtcars dataset using the mean Euclidean distance. Figure 7.2(b) shows the silhouette diagram corresponding to a cut of four clusters in the dendrogram shown in Figure 7.2(a), obtained by applying an agglomerative hierarchical algorithm to the mtcars dataset using the mean Euclidean distance. Finally, Figure 7.2(b) shows the silhouette diagram corresponding to the clustering with  $K = 4$  obtained by applying a medoids partitioning algorithm.



**Figure 7.2:** silhouette diagrams with four clusters for an agglomerative hierarchical algorithm, applied to the mtcars dataset (a) with the mean Euclidean distance, and (b) for a medoids partitioning algorithm

---

## 7.5 SUMMARY

This chapter explain about the clustering problems. The evaluation of the models are discussed throughout this chapter. There are four methods of clustering which are partition, hierarchical, density based and grid respectively. K-means and K medoids are examples of partition methods of clustering. Hierarchical methods can be subdivided into two main groups: agglomerative and divisive methods.

---

## 7.6 UNIT END QUESTIONS

---

- 1) What is Clustering? Explain its methods?
- 2) How clustering models are evaluated?
- 3) Explain K-means algorithms with steps and example.

---

## 7.7 REFERENCES

---

- Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis, First Edition, Wiley
- Decision support and Business Intelligence Systems by Efraim Turban, Ramesh Sharda, Dursun Delen, ninth Edition, Pearson, 2011



# MARKETING, LOGISTIC AND PRODUCTION MODELS

## Unit Structure

- 8.0. Objective
- 8.1. Relational marketing
  - 8.1.1 Motivations and objectives
  - 8.1.2 An environment for relational marketing analysis
  - 8.1.3 Lifetime value
  - 8.1.4 The effect of latency in predictive models
  - 8.1.5 Acquisition
  - 8.1.6 Retention
  - 8.1.7 Cross-selling and up-selling
  - 8.1.8 Market basket analysis
  - 8.1.9 Web mining
- 8.2 Salesforce management
  - 8.2.1 Decision processes in salesforce management
  - 8.2.2 Models for salesforce management
  - 8.2.3 Response functions
- 8.3. Logistic and production models
  - 8.3.1. Supply chain optimization
  - 8.3.2. Optimization models for logistics planning
  - 8.3.3. Tactical planning
  - 8.3.4. Multiple resources
  - 8.3.5. Backlogging
  - 8.3.5. Minimum lots and fixed costs
  - 8.3.7. Bill of materials
  - 8.3.8. Multiple Plants
  - 8.3.9. Revenue management systems
- 8.4. Summary
- 8.5 Questions
- 8.6 Reference

---

## 8.0. OBJECTIVE

---

This chapter would make you understand the following concepts:

- Relational marketing concept like an environment for relational marketing analysis, Lifetime value, The effect of latency in predictive models, etc..
- Salesforce management and its Decision processes in salesforce management, Models for salesforce management and Response functions
- Various Logistic and production models

Due to the simultaneous presence of numerous objectives and the vast array of potential course of action that can be taken as a result of combining the key objectives, marketing decision processes are known for their high level of complexity. choice options available to decision makers. Therefore, it should come as no surprise that many mathematical models for marketing have been successfully developed and applied in recent decades. The importance of mathematical models for marketing has been further strengthened by the availability of massive databases of sales transactions that provide accurate information on how customers make use of services or purchase products.

This chapter will primarily focus on two prominent topics in the field of marketing intelligence. The first theme is particularly broad and concerns the application of predictive models to support relational marketing strategies, whose purpose is to customize and strengthen the relationship between a company and its customers. After a brief introduction to relational marketing, we will describe the main streams of analysis that can be dealt with in this domain of application, indicating for each of them the classes of predictive models that are best suited to dealing with the problems considered. The subjects discussed in this context can be partly extended to the relationship between citizens and the public administration.

The second theme concerns salesforce management. First, we will provide an overview of the major decision-making processes emerging in the organization of a sales staff, highlighting also the role played by response functions. Then, we will illustrate some optimization models which aim to allocate a set of geographical territories to sales agents as well as planning the activities of sales agents. Finally, some business cases consisting of applied marketing models will be discussed.

---

### 8.1. RELATIONAL MARKETING

---

In order to fully understand the reasons why enterprises develop relational marketing initiatives, consider the following three examples: an insurance company that wishes to select the most promising market segment to target for a new type of policy; a mobile phone provider that wishes to

identify those customers with the highest probability of churning, that is, of discontinuing their service and taking out a new contract with a competitor, in order to develop targeted retention initiatives; a bank issuing credit cards that needs to identify a group of customers to whom a new savings management service should be offered. These situations share some common features: a company owning a massive database which describes the purchasing behaviour of its customers and the way they make use of services, wishes to extract from these data useful and accurate knowledge to develop targeted and effective marketing campaigns.

The aim of a relational marketing strategy is to initiate, strengthen, intensify and preserve over time the relationships between a company and its stakeholders, represented primarily by its customers, and involves the analysis, planning, execution and evaluation of the activities carried out to pursue these objectives.

Relational marketing became popular during the late 1990s as an approach to increasing customer satisfaction in order to achieve a sustainable competitive advantage. So far, most enterprises have taken at least the first steps in this direction, through a process of cultural change which directs greater attention toward customers, considering them as a formidable asset and one of the main sources of competitive advantage. A relational marketing approach has been followed in a first stage by service companies in the financial and telecommunications industries, and has later influenced industries such as consumer goods, finally reaching also manufacturing companies, from automotive and commercial vehicles to agricultural equipment's, traditionally more prone to a vision characterized by the centrality of products with respect to customers.

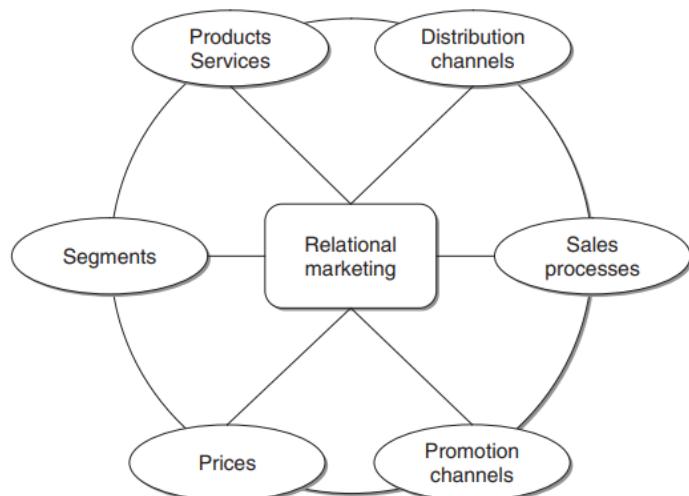
### **8.1.1 Motivations and objectives**

The reasons for the spread of relational marketing strategies are complex and interconnected. Some of them are listed below, although for additional information the reader is referred to the suggested references at the end of the chapter.

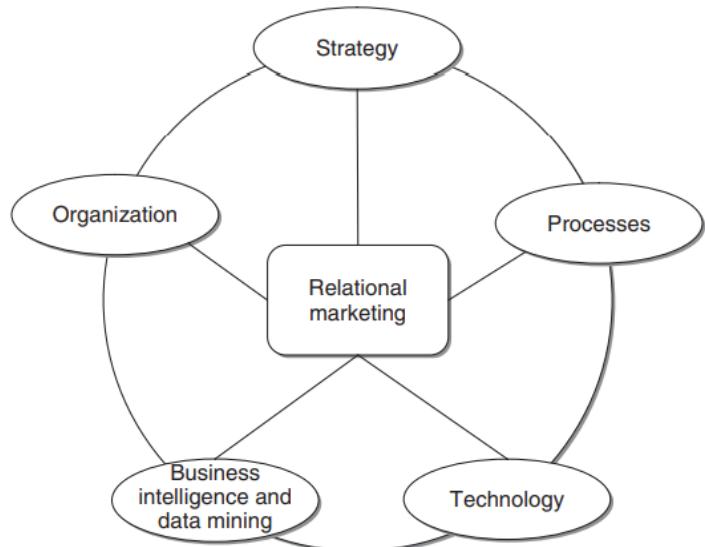
- The increasing concentration of companies in large enterprises and the resulting growth in the number of customers have led to greater complexity in the markets.
- Since the 1980s, the innovation— production— obsolescence cycle has progressively shortened, causing a growth in the number of customized options on the part of customers, and an acceleration of marketing activities by enterprises.
- The increased flow of information and the introduction of e-commerce have enabled global comparisons. Customers can use the Internet to compare features, prices and opinions on products and services offered by the various competitors.
- Customer loyalty has become more uncertain, primarily in the service industries, where often filling out an on-line form is all one must do to change service provider.

- In many industries a progressive commoditization of products and services is taking place, since their quality is perceived by consumers as equivalent, so that differentiation is mainly due to levels of service.
- The systematic gathering of sales transactions, largely automated in most businesses, has made available large amounts of data that can be transformed into knowledge and then into effective and targeted marketing actions.
- The number of competitors using advanced techniques for the analysis of marketing data has increased.

Relational marketing strategies revolve around the choices shown in Figure 8.1, which can be effectively summarized as formulating for each segment, ideally



**Figure 8.1 Decision-making options for a relational marketing strategy**



**Figure 8.2 Components of a relational marketing strategy**

for each customer, the appropriate offer through the most suitable channel, at the right time and at the best price.

Marketing, Logistic and production models

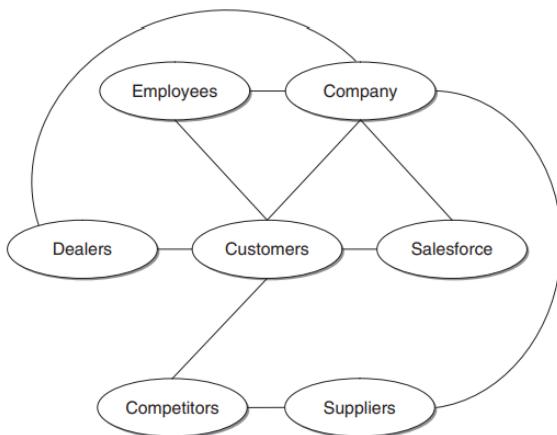
The ability to effectively exploit the information gathered on customers' behaviour represents today a powerful competitive weapon for an enterprise. A company capable of gathering, storing, analysing and understanding the huge amount of data on its customers can base its marketing actions on the knowledge extracted and achieve sustainable competitive advantages. Enterprises may profitably adopt relational marketing strategies to transform occasional contacts with their customers into highly customized long-term relationships. In this way, it is possible to achieve increased customer satisfaction and at the same time increased profits for the company, attaining a win-win relationship.

To obtain the desired advantages, a company should turn to relational marketing strategies by following a correct and careful approach. It is advisable to stress the distinction between a relational marketing vision and the software tools usually referred to as customer relationship management (CRM). As shown in Figure 8.2, relational marketing is not merely a collection of software applications, but rather a coherent project where the various company departments are called upon to cooperate and integrate the managerial culture and human resources, with a high impact on the organizational structures. It is then necessary to create within a company a true data culture, with the awareness that customer-related information should be enhanced through the adoption of business intelligence and data mining analytical tools.

Based on the investigation of cases of excellence, it can be said that a successful relational marketing strategy can be achieved through the development of a company-wide vision that puts customers at the center of the whole organization. Of course, this goal cannot be attained by exclusively relying on innovative computer technologies, which at most can be considered a relevant enabling factor.

The overlap between relational marketing strategies and CRM software led to a misunderstanding with several negative consequences. On one hand, the notion that substantial investments in CRM software applications were in themselves sufficient to generate a relational marketing strategy represents a dangerous simplification, which caused many project failures. On the other hand, the high cost of software applications has led many to believe that a viable approach to relational marketing was only possible for large companies in the service industries. This is a deceitful misconception: as a matter of fact, the essential components of relational marketing are a well-designed and correctly fed marketing data mart, a collection of business intelligence and data mining analytical tools, and, most of all, the cultural education of the decision makers. These tools will enable companies to carry out the required analyses and translate the knowledge acquired into targeted marketing actions.

The relationship system of an enterprise is not limited to the dyadic relationship with its customers, represented by individuals and companies that purchase the products and services offered, but also includes other actors, such as the employees, the suppliers and the sales network. For most relationships shown



**Figure 8.3 Network of relationships involved in a relational marketing strategy**

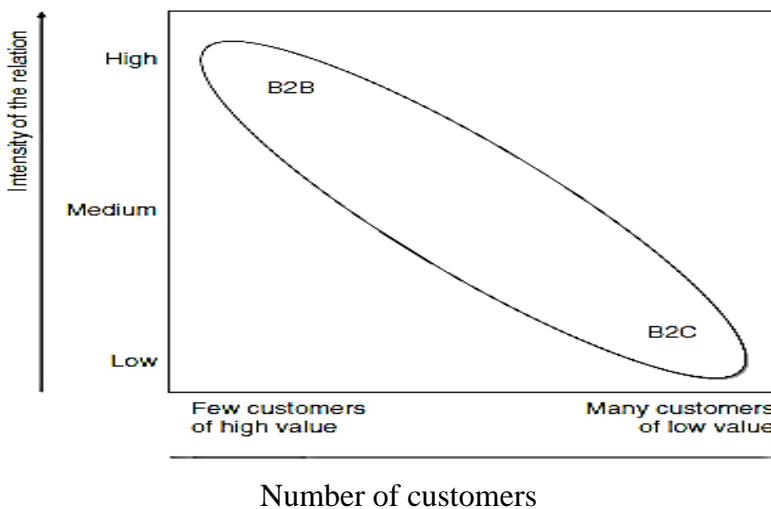
in Figure 8.3, a mutually beneficial exchange occurs between the different subjects involved. More generally, we can widen the boundaries of relational marketing systems to include the stakeholders of an enterprise. The relationship between an enterprise and its customers is sometimes mediated by the sales network, which in some instances can partially obstruct the visibility of the end customers.

Let us look at a few examples to better understand the implications of this issue. The manufacturers of consumer goods, available at the points of sale of large and small retailers, do not have direct information on the consumers purchasing their products. The manufacturers of goods covered by guarantees, such as electrical appliances or motor vehicles, have access to personal information on purchasers, even if they rarely also have access to information on the contacts of and promotional actions carried out by the network of dealers. Likewise, a savings management company usually places shares in its investment funds through a network of intermediaries, such as banks or agents, and often knows only the personal data of the subscribers. A pharmaceutical enterprise producing prescription drugs usually ignores the identity of the patients that use its drugs and medicinal products, even though promotional activities to influence consumers are carried out in some countries where the law permits.

It is not always easy for a company to obtain information on its end customers from dealers in the sales network and even from their agents. These may be reluctant to share the wealth of information for fear, rightly or wrongly, of compromising their role. In a relational marketing project specific initiatives should be devised to overcome these cultural and organizational barriers, usually through incentives and training courses.

The number of customers and their characteristics strongly influence the nature and intensity of the relationship with an enterprise, as shown in Figure 8.4. The relationships that might be established in a specific economic domain tend to lie on the diagonal shown in the figure. At one extreme, there are highly intense relationships existing between the company and a small number of customers of high individual value. Relationships of this type occur more frequently in *business-to-business* (B2B) activities, although they can also be found in other domains, such as private banking. The high value of each customer justifies the use of dedicated resources, usually consisting of sales agents and key account managers, to maintain and strengthen these more intense relationships. In situations of this kind, careful organization, and planning of the activities of sales agents is critical. Therefore, optimization models for *salesforce automation* (SFA), described in Section 8.2, can be useful in this context.

At the opposite extreme of the diagonal are the relationships typical of consumer goods and *business-to-consumer* (B2C) activities, for which a high number of low-value customers contact the company in an impersonal.



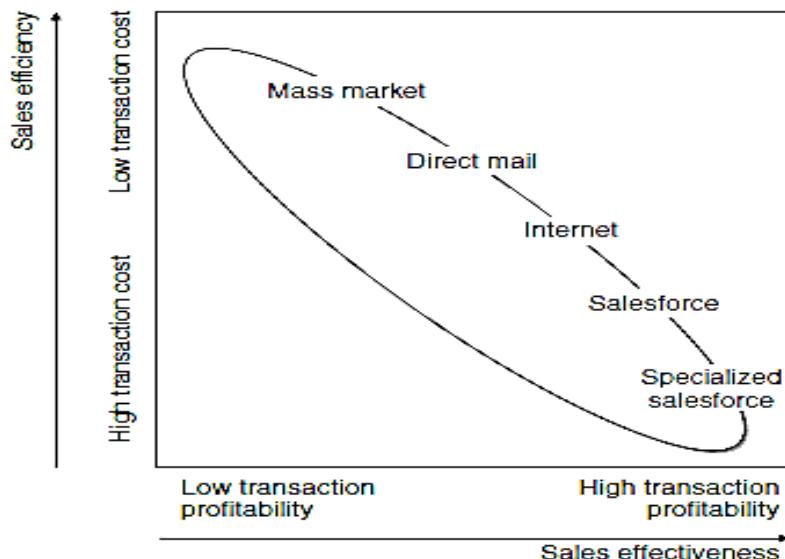
**Figure 8.4 Intensity of customer relationships as a function of number of customers**

way, through websites, call centers and points of sale. Data mining analyses for segmentation and profiling are particularly valuable especially in this context, characterized by many fragmented contacts and transactions. Relational marketing strategies, which are based on the knowledge extracted through data mining models, enable companies develop a targeted customization and differentiation of their products and/or services, including companies more prone toward a mass-market approach.

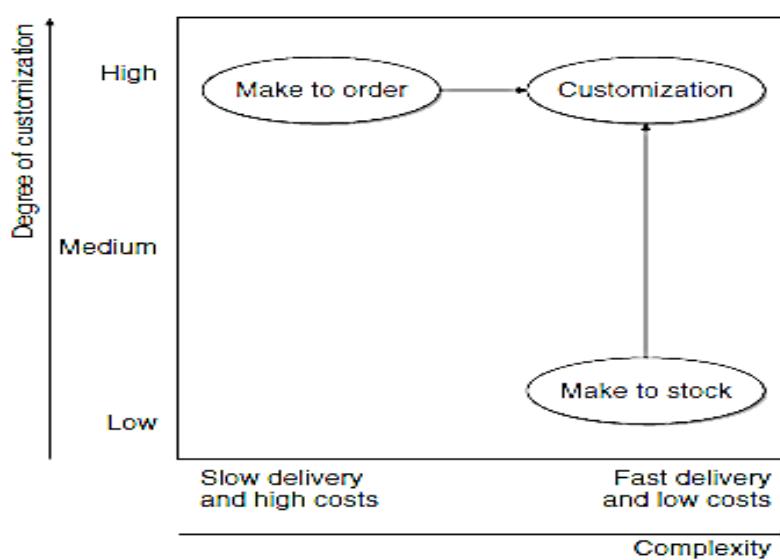
Figure 8.5 contrasts the cost of sales actions and the corresponding revenues. Where transactions earn a low revenue per unit, it is necessary to implement low-cost actions, as in the case of mass-marketing activities. Moving down along the diagonal in the figure, more evolved and intense relationships with the customers can be found. The relationships at the end

of the diagonal presuppose the action of a direct sales network and for the most part is typical of B2B relational contexts.

Figure 8.6 shows the ideal path that a company should follow to be able to offer customized products and services at low cost and in a brief time. On the one hand, companies operating in a mass market, well acquainted with fast delivery at low costs, must evolve in the direction of increased customization, by introducing more options and variants of products and services offered to the various market segments. Data mining analyses for relational marketing purposes are a powerful tool for identifying the segments to be targeted with customized products. On the other hand, the companies oriented toward make-to-order production must evolve in a direction that fosters reductions in both costs and delivery times, but without reducing the variety and the range of their products.



**Figure 8.5 Efficiency of sales actions as a function of their effectiveness**

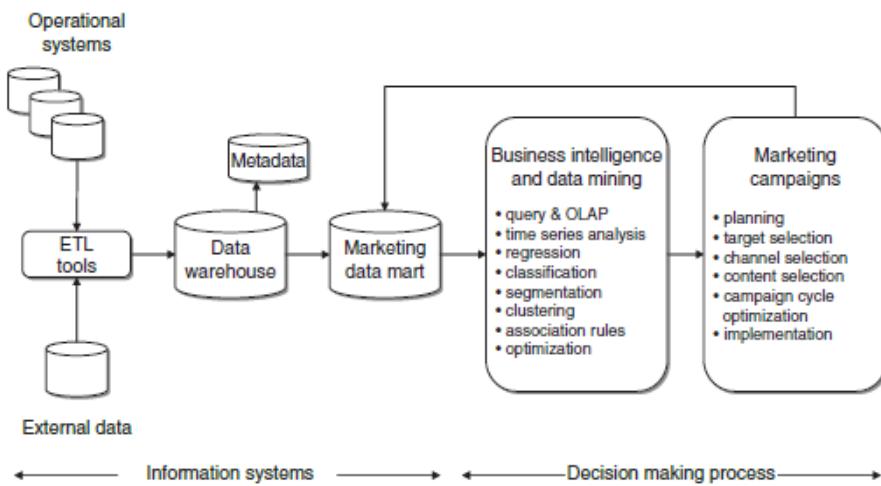


**Figure 8.6 Level of customization as a function of complexity of products and Services**

## 8.1.2 An environment for relational marketing analysis

Marketing, Logistic and production models

Figure 8.7 shows the main elements that make up an environment for relational marketing analysis. Information infrastructures include the company's data warehouse, obtained from the integration of the various internal and external data sources, and a marketing data mart that feeds business intelligence and data mining analyses for profiling potential and actual customers. Using pattern recognition and machine learning models as described in previous chapters, it is possible to derive different segmentations of the customer base, which are then used to design targeted and optimized marketing actions. A classification model can be used, for example, to generate a *scoring* system for customers according to their propensity to buy a service offered by a company, and to direct a cross-selling offer only toward those customers for whom a high probability of acceptance is predicted by the model, thus maximizing the overall *redemption* of the marketing actions.



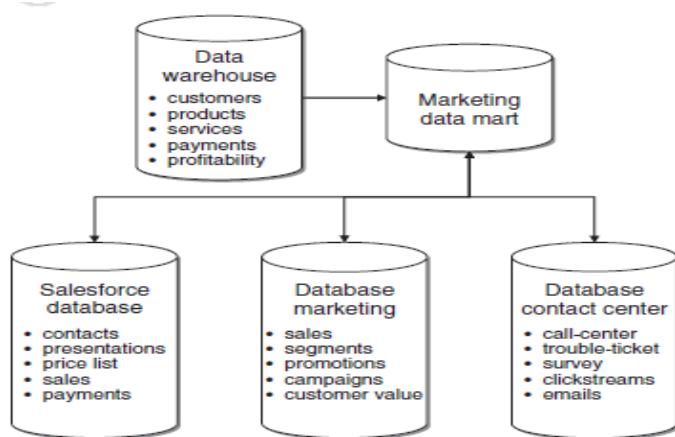
**Figure 8.7 Components of an environment for relational marketing analysis**

Effective management of frequent marketing campaign cycles is certainly a complex task that requires planning, for each segment of customers, the content of the actions and the communication channels, using the available human and financial resources. The corresponding decision-making process can be formally expressed by appropriate optimization models. The cycle of marketing activities terminates with the execution of the planned campaign, with the subsequent gathering of information on the results and the redemption among the recipients. The data collected are then fed into the marketing data mart for use in future data mining analyses. During the execution of each campaign, it is important to set up procedures for controlling and analyzing the results obtained. To assess the overall effectiveness of a campaign, it would be advisable to select a *control group* of customers, with characteristics like those of the campaign recipients, toward whom no action should be undertaken.

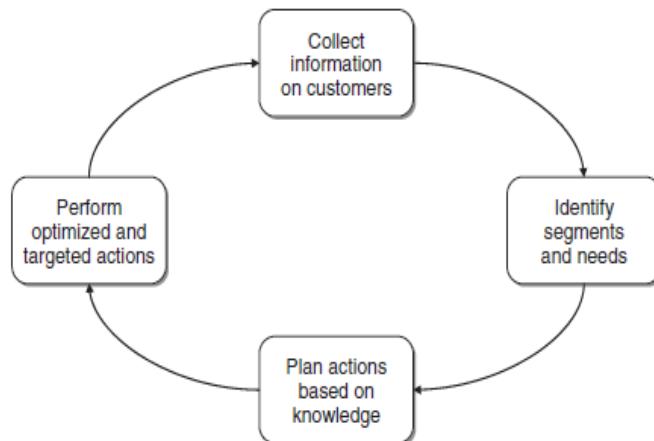
Figure 8.8 describes the main types of data stored in a data mart for relational marketing analyses. A company data warehouse provides

demographic and administrative information on each customer and the transactions carried out for purchasing products and using services. The marketing database contains data on initiatives carried out in the past, including previous campaigns and their results, promotions and advertising, and analyses of customer value. A further data source is the salesforce database, which provides information on established contacts, calls, and applicable sales conditions.

Finally, the *contact center* database provides access to data on customers' contacts with the call center, problems reported, sometimes called *trouble tickets*, and as shown in Figure 8.8, the available data are plentiful, providing an accurate representation of the behaviors and needs of the different customers, using inductive learning models. The main phases of a relational marketing analysis proceeds as shown in Figure 8.9. The first step is the exploration of the data available for each customer. Later, by using inductive learning models, it is possible to extract from those data the insights and the rules that allow market segments characterized by similar behaviors to be identified. Knowledge of customer profiles is used to design marketing actions which are then translated into promotional campaigns and generate in turn added information to be used during subsequent analyses.



**Figure 8.8 Types of data feeding a data mart for relational marketing analysis**

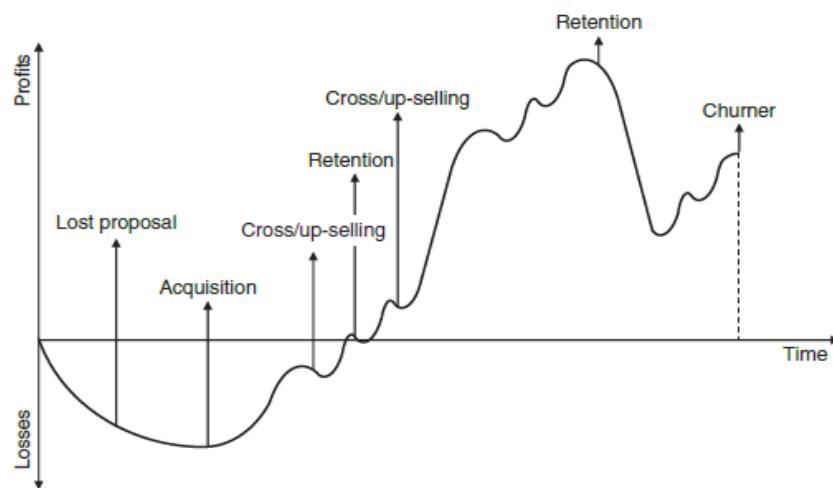


**Figure 8.9 Cycle of relational marketing analysis**

### 8.1.3 Lifetime value

Figure 8.10 shows the main stages during the customer *lifetime*, showing the cumulative value of a customer over time. The figure also illustrates the different actions that can be undertaken toward a customer by an enterprise. In the initial phase, an individual is a *prospect*, or potential customer, who has not yet begun to purchase the products or to use the services of the enterprise. Toward potential customers, *acquisition* actions are carried out, both directly (telephone contacts, emails, talks with sales agents) and indirectly (advertising, notices on the enterprise website). These actions incur a cost that can be assigned to each customer and determine an accumulated loss that lasts until a critical event in the relationship with a customer occurs: a prospect becomes a customer.

This event may take various forms in different situations: it may consist of a service subscription, the opening of a bank account, the first purchase at a retailer point of sale with the activation of a loyalty card. Before becoming a new customer, a prospect may receive from the enterprise repeated proposals aiming acquiring her custom, shown in the figure as *lost proposals*, which have a negative outcome. From the time of acquisition, each customer generates revenue, which produces a progressive rise along the curve of losses and cumulated profits. This phase, which corresponds to the maturity of the relationship with the enterprise, usually entails alternating *cross-selling*, *up-selling* and *retention* actions, to extend the duration and the profitability of the relationship to maximize the lifetime value of each customer. The last event in a customer lifetime is the interruption of the relationship. This may be *voluntary*, when a customer discontinues the services of an enterprise and switches to those of a competitor, *forced*, when for instance a customer does not comply with payment terms, or *unintentional*, when for example a customer changes her place of residence.



**Figure 8.10 Lifetime of a customer**

The progress of a customer lifetime highlights the main tasks of relational marketing. First, the purpose is to increase the ability to acquire new customers. Through the analysis of the available information for those

customers who in the past have purchased products or services, such as personal socio-demographic characteristics, purchased products, usage of services, previous contacts, and the comparison with the characteristics of those who have not taken up the offers of the enterprise, it is possible to identify the segments with the highest potential. This in turn allows the enterprise to optimize marketing campaigns, to increase the effectiveness of acquisition initiatives and to reduce the waste of resources due to offers addressed to unpromising market segments.

Furthermore, relational marketing strategies can improve the loyalty of customers, extending the duration of their relationship with the enterprise, and thus increasing the profitability. In this case, too, the comparative analysis of the characteristics of those who have remained loyal over time with respect to those who have switched to a competitor leads to predictions of the likelihood of churning for each customer. Retention actions can therefore be directed to the most relevant segments, represented by high-value customers with the highest risk of churning.

who are more likely to take up the offer of additional services and products (cross-selling), or of alternative services and products of a higher level and with a greater profitability for the enterprise (up-selling).

The tasks of acquisition, retention, cross-selling and up-selling, shown in Figure 8.11, are at the heart of relational marketing strategies and their aim is to maximize the profitability of customers during their lifetime. These analysis tasks, which will be described in the next sections, are clearly amenable to classification problems with a binary target class. Notice that attribute selection plays a critical role in this context, since the number of available explanatory variables is usually quite large and it is advisable for learning models to use a limited subset of predictive features, to generate meaningful and useful classification rules for the accurate segmentation of customers.

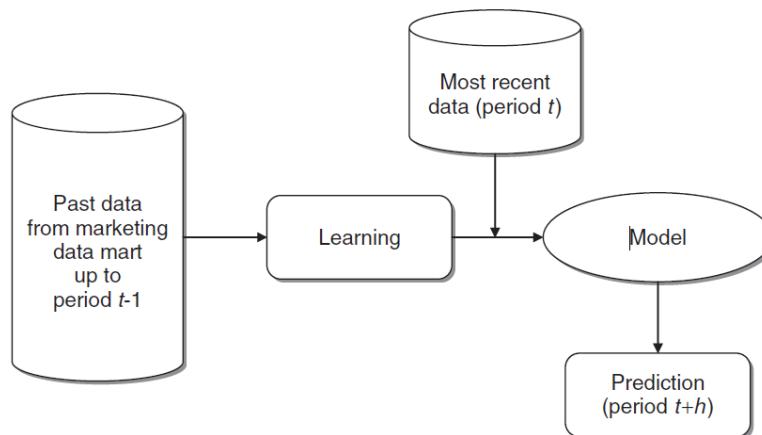
#### 8.1.4 The effect of latency in predictive models

Figure 8.12 illustrates the logic of development of a classification model for a relational marketing analysis, also considering the temporal dimension. Assume that  $t$  is the current time, and that we wish to derive an inductive learning model for a classification problem. For example, at the beginning of October a mobile telephone provider might want to develop a classification model to predict the probability of churning for its customers. The data mart contains the data for past periods, updated as far as period  $t - 1$ . In our example, it contains data up to and including September.

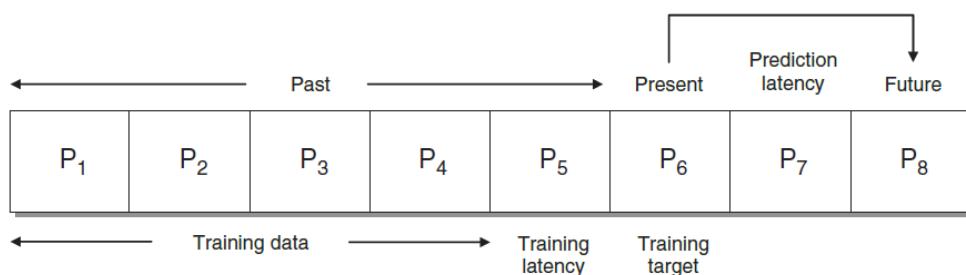
Furthermore, suppose that the company wishes to predict the probability of churning  $h$  months in advance, since in this way any retention action has a better chance of success. In our example, we wish to predict at the beginning of October the probability of churning in November, using data up to September. Notice that the data for period  $t$  cannot be used for the prediction, since they are clearly not available at the beginning of period.

To develop a classification model, we use the value of the target variable for the last known period  $t - 1$ , corresponding to the customers who churned in the month of September. It should be clear that for training and testing the model the explanatory variables for period  $t - 2$  should not be used, since in the training phase it is necessary to reproduce the same situation as will

be faced when using the model in the prediction stage. The target variable must be predicted  $h = 2$  periods in advance, and therefore there is an intermediate period of future data that are still unknown at time  $t$  (the month of October in our example). To reflect these dynamics, the training phase should be carried out without using August data. In general, the  $h - 1$  period corresponding to data still unknown during the prediction phase, and not used during the training phase, are referred to as the model's *latency*, as shown in Figure 8.8.



**Figure 8.11 Development and application flowchart for a predictive model**



**Figure 8.12 Latency of a predictive model**

### 8.1.5 Acquisition

Although retention plays a prominent role in relational marketing strategies, for many companies the *acquisition* of new customers also represents a critical factor for growth. The acquisition process requires the identification of new prospects, as they are potential customers who may be totally or partially unaware of the products and services offered by the company or did not possess in the past the characteristics to become

customers or were customers of competitors. It may also happen that some of the prospects were former customers who switched their custom to competitors, in which case much more information is generally available on them.

Once prospects have been identified, the enterprise should address acquisition campaigns to segments with a high potential profitability and a high probability of acquisition, to optimize the marketing resources. Traditional marketing techniques identify interesting segments using predefined profiling criteria, based on market polls and socio-demographic analyses, according to a top-down perspective. This approach can be successfully integrated or even replaced by a top-down segmentation logic which analyzes the data available in the data mart, as shown in Figure 8.8 (demographic information, contacts with prospects, use of products and services of competitors), and derives classification rules that characterize the most promising profiles for acquisition purposes. Also in this case, we are faced with a binary classification problem, which can be analyzed with the techniques described in previous Chapter.

### 8.1.6 Retention

The maturity stage reached by most products and services, and the subsequent saturation of their markets, have caused more severe competitive conditions. Therefore, the expansion of the customer base of an enterprise consists increasingly of switch mechanisms – the acquisition of customers at the expense of other companies. This phenomenon is particularly apparent in-service industries, such as telecommunications, banking, savings management, and insurance, although it also occurs in manufacturing, both for consumer goods and industrial products. For this reason, many companies invest significant amounts of resources in analyzing and characterizing the phenomenon of *attrition*, whereby customers switch from their company to a competitor.

There are also economic reasons for devoting substantial efforts to customer *retention*: indeed, it has been empirically observed that the cost of acquiring a new customer, or winning back a lost customer, is usually much higher – of the order of 5 to 9 times higher – than the cost of the marketing actions aimed at retaining customers considered at risk of churning. Furthermore, an action to win back a lost customer runs the risk of being too late and not achieving the desired result. In many instances, winning back a customer requires investments that do not generate a return.

One of the main difficulties in loyalty analysis is recognizing a churn. For subscription services there are unmistakable signals, such as a formal notice of withdrawal, while in other cases it is necessary to define adequate indicators that are correlated, a few periods in advance, with the actual churning. A customer who reduces by more than a given percentage her purchases at a selected point of sale using a loyalty card, or a customer who reduces below a given threshold the amount held in her checking account and the number of transactions, represent two examples of

disaffection indicators. They also highlight the difficulties involved in correctly defining the appropriate threshold values.

To optimize the marketing resources addressed to retention, it is therefore necessary to target efforts only toward high-value customers considered at risk of churning. To obtain a scoring system corresponding to the probability of churning for each customer, it is necessary to derive a segmentation based on the data on past instances of churning. Predicting the risk of churning requires analysis of records of transactions for each customer and identifying the attributes that are most relevant to accurately explaining the target variable. Again, we are faced with a binary classification problem. Once the customers with the highest risk of churning have been identified, a retention action can be directed toward them. The more accurately such action is targeted, the cheaper it is likely to be.

### **8.1.7 Cross-selling and up-selling**

Data mining models can also be used to support a relational marketing analysis aimed at identifying market segments with a higher propensity to purchase additional services or other products of a company. For example, a bank also offering insurance services may identify among its customers segments interested in purchasing a life insurance policy. In this case, demographic information on customers and their past transactions recorded in a data mart can be used as explanatory attributes to derive a classification model for predicting the target class, consisting in this example of a binary variable that indicates whether the customer accepted the offer or not.

The term cross-selling refers to the attempt to sell an additional product or service to an active customer, already involved in a long-lasting commercial relationship with the enterprise. By means of classification models, it is possible to identify the customers characterized by a high probability of accepting a cross-selling offer, starting from the information contained in the available attributes.

In other instances, it is possible to develop an up-selling initiative, by persuading a customer to purchase a higher-level product or service, richer in functions for the user and more profitable for the company, and therefore able to increase the lifetime value curve of a customer. For example, a bank issuing credit cards may offer customers holding a standard card an upgrade to a gold card, which is more profitable for the company, but also able to offer a series of complementary services and advantages to interested customers. In this case too, we are dealing with a binary classification problem, which requires construction of a model based on the training data of customers' demographic and operational attributes. The purpose of the model is to identify the most interesting segments, corresponding to customers who have taken up the gold service in the past, and who appear therefore more appreciative of the additional services offered by the gold card. The segments identified in this way represent the target of up-selling actions.

### 8.1.8 Market basket analysis

The purpose of *market basket analysis* is to gain insight from the purchases made by customers to extract useful knowledge to plan marketing actions. It is mostly used to analyze purchases in the retail industry and in e-commerce activities and is amenable to unsupervised learning problems. It may also be applied in other domains to analyze the purchases made using credit cards, the complementary services activated by mobile or fixed telephone customers, the policies or the checking accounts acquired by a same household.

The data used for this purpose mostly refer to purchase transactions and can be associated with the time dimension if the purchaser can be tracked through a loyalty card or the issue of an invoice. Each transaction consists of a list of purchased items. This list is called a *basket*, just like the baskets available at retail points of sale.

If transactions cannot be connected to one another, say because the purchaser is unknown, one may then apply association rules, described in next Chapter, to extract interesting correlations between the purchases of groups of items. The rules extracted in this way can then be used to support different decision-making processes, such as assigning the location of the items on the shelves, determining the layout of a point of sale, identifying which items should be included in promotional flyers, advertisements or coupons distributed to customers.

Clustering models, described in next Chapter, are also useful in determining homogeneous groups of items, once an incidence matrix  $\mathbf{X}$  has been created for the representation of the dataset, where the rows correspond to the transactions and the columns to the items. If customers are individually identified and traced, besides the above techniques it is also possible to develop further analyses that consider the time dimension of the purchases. For instance, one may generate sequential association rules, mentioned at the end of Chapter 11, or apply time series analysis, as described in previous Chapter.

### 8.1.9 Web mining

The web is a critical channel for the communication and promotion of a company's image. Moreover, e-commerce sites are important sales channels. Hence, it is natural to use *web mining* methods to analyze data on the activities carried out by the visitors to a website.

---

## 8.2 SALESFORCE MANAGEMENT

---

Most companies have a sales network and therefore rely on a substantial number of people employed in sales activities, who play a critical role in the profitability of the enterprise and in the implementation of a relational marketing strategy. The term *salesforce* is taken to mean the complete set of people and roles that are involved, with different tasks and responsibilities, in the sales process. A preliminary taxonomy of salesforces is based on the type of activity carried out, as indicated below.

**Residential.** *Residential* sales activities take place at one or more sites managed by a company supplying some products or services, where customers go to make their purchases. This category includes sales at retail outlets as well as wholesale trading centers and *cash-and-carry* shops.

**Mobile.** In *mobile* sales, agents of the supplying company go to the customers' homes or offices to promote their products and services and collect orders. Sales in this category occur mostly within B2B relationships, even though they can also be found in B2C contexts.

**Telephone.** *Telephone* sales are carried out through a series of contacts by telephone with prospective customers.

There are various problems connected with managing a mobile salesforce management, which will be the focus of this section. They can be subdivided into a few main categories:

- designing the sales network;
- planning the agents' activities;
- contact management;
- sales opportunity management;
- customer management;
- activity management;
- order management;
- area and territory management;
- support for the configuration of products and services;
- knowledge management about products and services.

Designing the sales network and planning the agents' activities involve decision-making tasks that may take advantage of the use of optimization models, such as those that will be described in the next sections. The remaining activities are operational in nature and may benefit from the use of software tools for *salesforce automation* (SFA), today widely implemented.

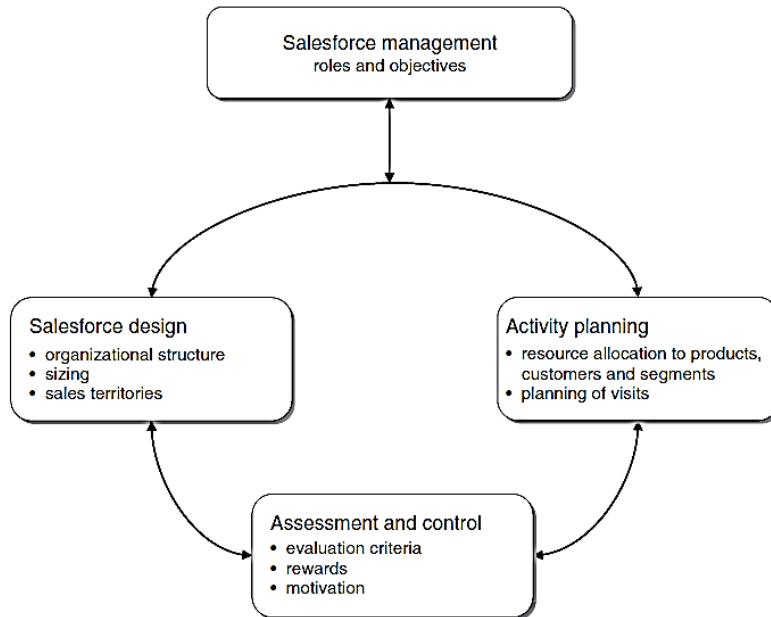
### 8.2.1 Decision processes in salesforce management

The design and management of a salesforce raise several decision-making problems, as shown in Figure 8.16. When successfully solved, they confer multiple advantages: maximization of profitability, increased effectiveness of sales actions, increased efficiency in the use of resources, and greater professional rewards for sales agents.

The decision processes described in Figure 8.16 should consider the strategic objectives of the company, with respect to other components of the marketing mix, and conform to the role assigned to the salesforce within the broader framework of a relational marketing strategy. The two-

way connections indicated in the figure suggest that the different components of the decision-making process interact with each other and with the general objectives of the marketing department.

In particular, the decision-making processes relative to salesforce management can be grouped into three categories: *design*, *planning* and *assessment*.



## 8.13 Decision processes in salesforce management

### Design

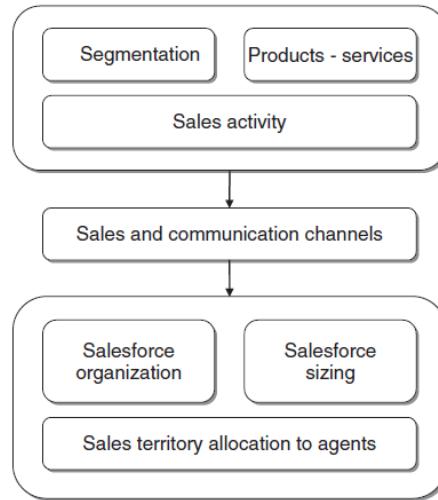
Salesforce design is dealt with during the start-up phase of a commercial activity or during subsequent restructuring phases, for example following the merger or acquisition of a group of companies.

As shown in Figure 8.16, the design phase is usually preceded by the creation of market segments through the application of data mining methods and by the articulation of the offer of products and services, which are in turn subdivided into homogeneous classes. Salesforce design includes three types of decisions.

**Organizational structure.** The *organizational structure* may take different forms, corresponding to hierarchical agglomerations of the agents by group of products, brand, or geographical area. In some situations, the structure may also be differentiated by markets. To determine the organizational structure, it is necessary to analyze the complexity of customers, products, and sales activities, and to decide whether and to what extent the agents should be specialized.

**Sizing.** Sales network *sizing* is a matter of working out the optimal number of agents that should operate within the selected structure, and depends on several factors, such as the number of customers and prospects, the desired level of sales area coverage, the estimated time for

each call and the agents' traveling time. One should bear in mind that a reduction in costs due to a decrease in the salesforce size is often followed by a reduction in sales and revenues. A better allocation of the existing salesforce, devised during the planning phase by means of optimization models, is usually more effective than a variation in size.



**Figure 8.14 Salesforce design process**

sales area coverage, the estimated time for each call and the agents' traveling time. One should bear in mind that a reduction in costs due to a decrease in the salesforce size is often followed by a reduction in sales and revenues. A better allocation of the existing salesforce, devised during the planning phase by means of optimization models, is usually more effective than a variation in size.

**Sales territories.** Designing a *sales territory* means grouping together the geographical areas into which a given region has been divided and assigning each territory to an agent. The design and assignment of sales territories should consider several factors, such as the sales potential of each geographical area, the time required to travel from one area to another and the total time each agent has available. The purpose of the assignment consists of determining a balanced situation between sales opportunities embedded in each territory, to avoid disparities among agents. The assignment of the geographical areas should be periodically reviewed since the sales potential balance in the various territories tends to vary over time.

Decisions concerning the design of the salesforce should consider decisions about salesforce planning, and this explains the two-way link between the two corresponding blocks in Figure 8.16

## Planning

Decision-making processes for *planning* purposes involve the assignment of sales resources, structured and sized during the design phase, to market entities.

Resources may correspond to the work time of agents or to the budget, while market entities consist of products, market segments, distribution channels and customers.

Allocation considers the time spent pitching the sale to each customer, the travel time and cost, and the effectiveness of the action for each product, service, or market segment. It is also possible to consider further ancillary activities carried out at the customers' sites, such as making suggestions that are conducive to future sales or explaining the technical and functional features of products and services. Salesforce planning can benefit from the use of optimization models, as explained below.

### 8.2.2 Models for salesforce management

In what follows we will describe some classes of optimization models for designing and planning the salesforce. These models are primarily intended for educational purposes, to familiarize readers with the reasoning behind specific aspects of a sales network, through the formulation of optimization models.

For the sake of clarity and conciseness, for each model we have limited the extensions to a single feature. Sales networks simultaneously possess more than one of the distinctive features previously described, and therefore the models developed in real-world applications, just like those described in the last section of the chapter, are more complex and result from a combination of distinctive characteristics.

Before proceeding, it is useful to introduce some notions common to the different models that will be described. Assume that a region is divided into  $J$  geographical sales areas, also called *sales coverage units*, and let  $J = \{1, 2, \dots, J\}$ . Areas must be aggregated into disjoint clusters, called *territories*, so that each area belongs to one single territory and is also connected to all the areas belonging to the same territory. The connection property implies that from each area it is possible to reach any other area of the same territory.

The time span is divided into  $T$  intervals of equal length, which usually correspond to weeks or months, indicated by the index  $t \in T = \{1, 2, \dots, T\}$ . Each territory is associated with a sales agent, located in one of the areas belonging to the territory, henceforth considered as her area of residence. The choice of the area of residence determines the time and cost of traveling to any other area in the same territory. Let  $I$  be the number of territories and therefore the number of agents that form the sales network, and let  $I = \{1, 2, \dots, I\}$ . In each area there are customers or prospects who can be visited by the agents as part of their promotions and sales activities. In some of the models that will be presented, customers or prospects are aggregated into segments, which are considered homogeneous with respect to the area of residence and to other characteristics, such as value, potential for development and purchasing

behaviors. Let  $H$  be the number of market entities, which in different models may represent either single customers or segments, and let  $H = \{1, 2, \dots, H\}$ . Let  $D_j$  be the set of customers, or segments of customers where necessary, located in area  $j$ .

Finally, assume that a given agent can promote and sell  $K$  products and services during the calls she makes on customers or prospects, and let  $K = \{1, 2, \dots, K\}$ .

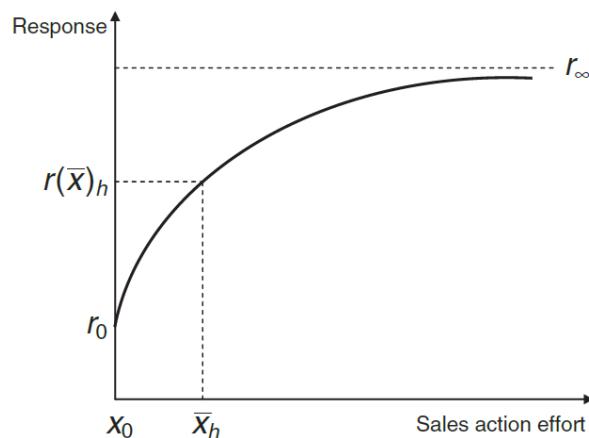
### 8.2.3 Response functions

Response functions play a key role in the formulation of models for designing and planning a sales network. In general terms, a response function describes the elasticity of sales in terms of the intensity of the sales actions and is a formal method to describe the complex relationship existing between sales actions and market reactions.

Sales to which the response function refers are expressed in product units or monetary units, such as revenues or margins. For the sake of uniformity, in the next sections response functions are assumed to be expressed as sales revenues. The intensity of a sales action can be related to different variables, such as the number of calls to a customer in each period, the number of mentions of a product in each period, and the time dedicated to each customer in each period.

In principle, it is possible to consider a response function in relation to each factor that is deemed critical to sales: the characteristics of customers and sales territories; the experience, education, and personal skills of the agents; promotions, prices, markdown policies operated by the company and the corresponding features for one or several competitors.

Figures 8.17 and 8.18 show two shapes of the response function, obtained by placing the sales of a product or service on the vertical axis and the intensity of the sales action of interest on the horizontal axis. To fix ideas, we will assume that the number  $x_h$  of calls that a specific agent makes to customer  $h$  in each period of the planning horizon is placed on the horizontal axis.



**Figure 8.15 A concave response function**

The concave response function shown in Figure 8.17 can be interpreted in the following way: as the number of calls increases, revenues grow at a decreasing rate approaching zero, according to the principle of decreasing marginal revenues. In general, a lack of sales actions toward a given customer does not imply a lack of sales, at least for a certain number of periods. This is an effect of the actions executed in previous periods that lasts over time. For this reason, the response function is greater than 0 at  $xh = 0$ .

The sigmoidal response function in Figure 8.18 reflects a different hypothesis of sales growth as a function of the actions carried out. The assumption made in this case is that the central interval of values on the horizontal axis corresponds to a higher rate of sales growth, while outside that area the growth rate is lower.

It is worth noting that each decision concerning the allocation of sales resources is based on a response function hypothesis, which is implicit and unaware in intuitive decision-making processes, while it is explicit and rigorous in mathematical models such as those presented below. Response functions can be estimated by considering two types of information.

On the one hand, one can use past available data regarding the intensity of the actions carried out and the corresponding sales, to develop a parametric regression model through variants of regression methods. On the other hand, interviews are carried out with agents and sales managers to obtain subjective information which is then incorporated into the procedure for calculating the response function.

We will now show by means of an example how the procedure for estimating the response function works. Let  $rh(xh)$  be the sales value for customer  $h$  associated with a number  $xh$  of calls during a given period. More generally, the variable that determines the response function  $r$  expresses the intensity of the sales action that has been carried out. A parametric form should first be selected in order to express the functional dependence. The following function, which may assume both concave and sigmoidal shapes by varying the parameters, can be used:

$$r_h(x_h) = r_0 + (r_\infty - r_0) \frac{x_h^\sigma}{\gamma + x_h^\sigma}.$$

The parameters in the expression  $rh(xh)$  have the following meaning:  $r_0$  represents the sales level that would be obtained at a sales action intensity equal to 0, as a prolonged effect of previous actions;  $r_\infty$  represents the maximum sales level, irrespective of the intensity of the sales action;  $\gamma$  and  $\sigma$  are two parameters to be estimated.

To obtain an estimate of the four parameters appearing in the expression for  $rh(xh)$  it is possible to proceed in two complementary ways. Past sales data can be used to set up a regression model and determine the values through the least squares method. In order to increase the value of the

opinions of the sales agents, it is also possible to ask agents and sales managers to estimate the value of the parameters  $r_0$  and  $r_\infty$ , as well as the values of the expected sales at other three critical points of the response function:  $r(-xh)$ , corresponding to the number of calls carried out at the time of the analysis,  $r(12 - xh)$  and  $r(32 - xh)$ , associated respectively with increasing and decreasing the number of calls by 50% with respect to the current value. Based on a subjective evaluation of the five response values derived through the procedure described above, an estimate by interpolation of the scale parameters  $\gamma$  and  $\sigma$  is then obtained.

#### 8.2.4 Sales territory design

*Sales territory design* involves allocating sales coverage units to individual agents to minimize a weighted sum of two terms, representing respectively the total distance between areas belonging to the same territory and the imbalance of sales opportunities for the agents.

Each region is subdivided into  $J$  geographical areas, which should then be clustered into  $I$  territories, whose total number has been determined beforehand.

A sales agent will be associated with each territory, and she should be in one of the sales coverage units, to be considered as her area of residence. It is further assumed that travel times within each area are negligible with respect to the corresponding travel times between a pair of distinct areas. Each area will be identified by the geographical coordinates  $(e_j, f_j)$  of one of its points, considered as representative of the entire sales coverage unit.

One might, for instance, choose the point whose coordinates are obtained as the average of the coordinates of all points belonging to that area. For each territory, let  $(e_i, f_i)$  denote the coordinates of the area where the agent associated with the territory resides.

$$d_{ij} = \sqrt{(e_i - e_j)^2 + (f_i - f_j)^2}; \quad (8.2)$$

Hence, the corresponding optimization problem can be formulated as

$$\begin{aligned} \min \quad & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} a_j d_{ij}^2 Y_{ij} + \beta \sum_{i \in \mathcal{I}} S_i, \\ \text{s.to} \quad & \sum_{j \in \mathcal{J}} a_j Y_{ij} - \frac{1}{I} \sum_{j \in \mathcal{J}} a_j \leq S_i, \quad i \in \mathcal{I}, \\ & \sum_{j \in \mathcal{J}} a_j Y_{ij} - \frac{1}{I} \sum_{j \in \mathcal{J}} a_j \geq -S_i, \quad i \in \mathcal{I}, \\ & \sum_{i \in \mathcal{I}} Y_{ij} = 1, \quad j \in \mathcal{J}, \\ & S_i \geq 0, \quad Y_{ij} \in \{0, 1\}, \quad i \in \mathcal{I}, j \in \mathcal{J}. \end{aligned}$$

The purpose of constraints (8.4) and (8.5) is to bound by means of variable  $S_i$  the absolute deviation between each territory sales opportunity and the

average sales opportunity, to make the assignment to territories more uniform with respect to sales opportunities, hence balancing the sales chances across the agents. Constraints (8.6) represent a multiple-choice condition imposed to guarantee that each sales coverage unit is exclusively assigned to one territory, and hence to one and only one agent.

Model (8.3) is a mixed binary optimization problem, which can be solved by a branch-and-bound method, truncated to limit the computing time and to achieve suboptimal solutions. Alternatively, an approximation algorithm can be devised for its *ad hoc* solution.

### 8.3. LOGISTIC AND PRODUCTION MODELS

---

In previous chapter we saw how the combination of relational marketing strategies with business intelligence and data mining models makes it possible to simultaneously increase revenues and reduce the costs of marketing actions, with an overall benefit for the profitability of an enterprise.

Besides acting on the marketing control levers, a manufacturing company can achieve further reductions in costs by improving its processes in another area that has received increasing attention in recent years: an effective supply chain management, understood as the logistic and production processes of a single enterprise as well as the network of companies composing the production chain of a given industry.

In this chapter we will focus on optimization models aimed at the integrated planning of the logistic chain from the perspective of a single company. We will begin with a qualitative description of the relevant processes within a logistic production system, by highlighting the major decisions that logistics managers have to face. The discussion will be confined to medium-term planning processes, which are concerned with some critical choices in the organization of the supply chain and can bring about substantial savings if appropriately optimized. We will then introduce some classes of optimization models, showing how the unique features of logistic production systems can be formally represented. Finally, we will discuss a few businesses case studies, with particular emphasis on a decision support system for supply chain optimization developed for a company in the food industry.

#### 8.3.1. Supply chain optimization

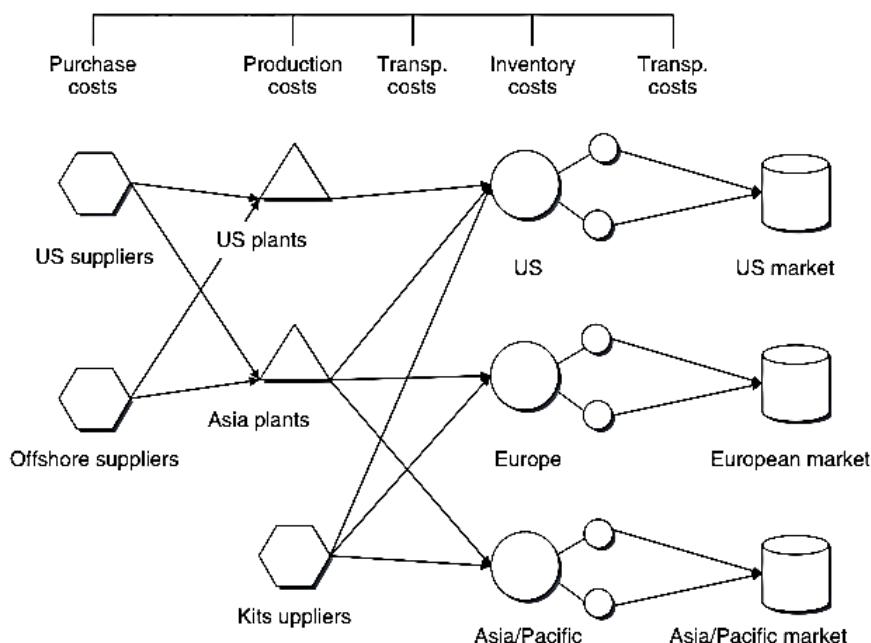
In a broad sense, a supply chain may be defined as a network of connected and interdependent organizational units that operate in a coordinated way to manage, control and improve the flow of materials and information originating from the suppliers and reaching the end customers, after going through the procurement, processing and distribution subsystems of a company, as shown in Figure 8.1.

The aim of the integrated planning and operations of the supply chain is to combine and evaluate from a systemic perspective the decisions made and

the actions undertaken within the various subprocesses that compose the logistic system of a company.

Many manufacturing companies, such as those operating in the consumer goods industry, have concentrated their efforts on the integrated operations of the supply chain, even to the point of incorporating parts of the logistic chain that are outside the company, both upstream and downstream.

The major purpose of an integrated logistic process is to minimize a function expressing the total cost, which comprises processing costs, transportation costs for procurement and distribution, inventory costs and equipment costs. Note that the optimization of the costs for each single phase does not imply that the minimum total cost of the entire logistic process has been achieved, so that a holistic perspective is required to attain a really optimized supply chain.



**Figure 8.16 An example of global supply chain**

The need to optimize the logistic chain, and therefore to have models and computerized tools for medium-term planning and for capacity analysis, is particularly critical in the face of the high complexity of current logistic systems, which operate in a dynamic and truly competitive environment. We are referring here to manufacturing companies that produce a vast array of products and that usually rely on a multicentric logistic system, distributed over several plants and markets, characterized by large investments in highly automated technology, by an intensive usage of the available production capacity and by short-order processing cycles. The features of the logistic system we have described reflect the profile of many enterprises operating in the consumer goods industry.

In the perspective outlined above, the aim of a medium-term planning process is therefore to devise an optimal logistic production plan, that is, a plan that can minimize the total cost, understood as the sum of

procurement, processing, storage, distribution costs and the penalty costs associated with the failure to achieve the predefined service level. However, to be implemented in practice, an optimal logistic production plan should also be feasible, that is, it should be able to meet the physical and logical constraints imposed by limits on the available production capacity, specific technological conditions, the structure of the bill of materials, the configuration of the logistic network, minimum production lots, as well as any other condition imposed by the decision makers in charge of the planning process.

Optimization models represent a powerful and versatile conceptual paradigm for analyzing and solving problems arising within integrated supply chain planning, and for developing the necessary software. Due to the complex interactions occurring between the different components of a logistic production system, other methods and tools intended to support the planning activity seem today inadequate, such as electronic spreadsheets, simulation systems and planning modules at infinite capacity included in enterprise resource planning software. Conversely, optimization models enable the development of realistic mathematical representations of a logistic production system, able to describe with reasonable accuracy the complex relationships among critical components of the logistic system, such as capacity, resources, plans, inventory, batch sizes, lead times and logistic flows, considering the various costs. Moreover, the evolution of information technologies and the latest developments in optimization algorithms mean that decision support systems based on optimization models for logistics planning can be efficiently developed.

### 8.3.2. Optimization models for logistics planning

In this section we will describe some optimization models that may be used to represent the most relevant features of logistic production systems. As already observed when introducing salesforce planning models in Chapter 8, for the sake of simplicity we have chosen to illustrate for each model a single feature of a logistic system. Readers should keep in mind that real-world logistic production systems feature simultaneously more than one of the elements considered, so that the models developed in applications, such as the business case studies presented in Section 8.4, will be more complex as they result from the combination of the different features.

Before proceeding with the description of specific models, it is useful to introduce some notation common to most models presented in this section. The logistic system includes  $I$  products, which will be denoted by the index  $i \in I = \{1, 2, \dots, I\}$ . The planning horizon is subdivided into  $T$  time intervals  $t \in T = \{1, 2, \dots, T\}$ , of equal length and usually corresponding to weeks or months.

The manufacturing process has at its disposal a set of critical resources shared among the assorted products and available in limited quantities. These resources may consist of production and assembly lines, to

manpower, to specific fixtures and tools required by manufacturing. The R critical resources considered in the logistic production system will be denoted by the index  $r \in R = \{1, 2, \dots, R\}$ . Whenever a single resource is relevant to the manufacturing process, the index  $r$  will be omitted for sake of simplicity.

### 8.3.3. Tactical planning

In its simplest form, the aim of tactical planning is to determine the production volumes for each product over the T periods included in the medium-term planning horizon in such a way as to satisfy the given demand and capacity limits for a single resource, and to minimize the total cost, defined as the sum of manufacturing production costs and inventory costs.

We therefore consider the decision variables

$P_{it}$  = units of product  $i$  to be manufactured in period  $t$ ,

$I_{it}$  = units of product  $i$  in inventory at the end of period  $t$ ,

and the parameters

$d_{it}$  = demand for product  $i$  in period  $t$ ,

$c_{it}$  = unit manufacturing cost for product  $i$  in period  $t$ ,

$h_{it}$  = unit inventory cost for product  $i$  in period  $t$ ,

$e_i$  = capacity absorption to manufacture a unit of product  $i$ ,

$b_t$  = capacity available in period  $t$ .

The resulting optimization problem is formulated as follows:

$$\begin{aligned} \min \quad & \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} (c_{it} P_{it} + h_{it} I_{it}) \\ \text{s.to} \quad & P_{it} + I_{i,t-1} - I_{it} = d_{it}, \quad i \in \mathcal{I}, t \in \mathcal{T}, \\ & \sum_{i \in \mathcal{I}} e_i P_{it} \leq b_t, \quad t \in \mathcal{T}, \\ & P_{it}, I_{it} \geq 0, \quad i \in \mathcal{I}, t \in \mathcal{T}. \end{aligned}$$

### 8.3.4. Multiple resources

If the manufacturing system requires R critical resources, a further extension of model (8.1) can be devised by considering multiple capacity constraints. The decision variables already included in model (8.1) remain unchanged, though it is necessary to consider the additional parameters

$b_{rt}$  = quantity of resource  $r$  available in period  $t$ ,

$e_{ir}$  = quantity of resource  $r$  absorbed to manufacture one unit of product  $i$ .

The resulting optimization problem is given by

$$\begin{aligned}
 \min \quad & \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} (c_{it} P_{it} + h_{it} I_{it}) \\
 \text{s.to} \quad & P_{it} + I_{i,t-1} - I_{it} = d_{it}, \quad i \in \mathcal{I}, t \in \mathcal{T}, \\
 & \sum_{i \in \mathcal{I}} e_{ir} P_{it} \leq b_{rt}, \quad r \in \mathcal{R}, t \in \mathcal{T}, \\
 & P_{it}, I_{it} \geq 0, \quad i \in \mathcal{I}, t \in \mathcal{T}.
 \end{aligned}$$

### 8.3.5. Backlogging

Another feature that needs to be modeled in some logistic systems is backlogging. The term backlog refers to the possibility that a portion of the demand due in each period may be satisfied in a subsequent period, incurring an additional penalty cost. Backlogs are a feature of production systems more likely to occur in B2B or make-to-order manufacturing contexts. In B2C industries, such as mass production consumer goods, on the other hand, one is more likely to find a variant of the backlog, known as lost sales, in which unfulfilled demand in a period cannot be transferred to a subsequent period and is lost. To model backlogging, it is necessary to introduce new decision variables

$B_{it}$  = units of demand for product  $i$  delayed in period  $t$ ,

and the parameters

$g_{it}$  = unit cost of delaying the demand for product  $i$  in period  $t$ .

The resulting optimization problem is

$$\begin{aligned}
 \min \quad & \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} (c_{it} P_{it} + h_{it} I_{it} + g_{it} B_{it}) \\
 \text{s.to} \quad & P_{it} + I_{i,t-1} - I_{it} + B_{it} - B_{i,t-1} = d_{it}, \quad i \in \mathcal{I}, t \in \mathcal{T}, \\
 & \sum_{i \in \mathcal{I}} e_i P_{it} \leq b_t, \quad t \in \mathcal{T}, \\
 & P_{it}, I_{it}, B_{it} \geq 0, \quad i \in \mathcal{I}, t \in \mathcal{T}.
 \end{aligned}$$

### 8.3.6. Minimum lots and fixed costs

A further feature often appearing in manufacturing systems is represented by minimum lot conditions: for technical or scale economy reasons, it is sometimes necessary that the production volume for one or more products be either equal to 0 (i.e., the product is not manufactured in a specific period) or not less than a given threshold value, the minimum lot.

To incorporate minimum lot conditions into the model, we define the binary decision variables

$$l_i = \text{minimum lot for product } i, \\ \gamma = \text{constant value larger than any producible volume for } i.$$

The optimization problem is now

$$\begin{aligned} \min \quad & \sum_{t \in T} \sum_{i \in \mathcal{I}} (c_{it} P_{it} + h_{it} I_{it}) \\ \text{s.to} \quad & P_{it} + I_{i,t-1} - I_{it} = d_{it}, \quad i \in \mathcal{I}, t \in T, \\ & \sum_{i \in \mathcal{I}} e_i P_{it} \leq b_t, \quad t \in T, \\ & P_{it} \geq l_i Y_{it}, \quad i \in \mathcal{I}, t \in T, \\ & P_{it} \leq \gamma Y_{it}, \quad i \in \mathcal{I}, t \in T, \\ & P_{it}, I_{it} \geq 0, Y_{it} \in \{0, 1\}, \quad i \in \mathcal{I}, t \in T. \end{aligned}$$

In all previous model formulations, we have implicitly assumed that production costs are proportional to production volumes. For some logistic systems, however, to manufacture a product, it may be necessary to set up a machine and incur a setup cost. However, such costs are required only if the production volume is strictly greater than zero, that is, if production of the product concerned is accomplished. A further parameter,

$$f_{it} = \text{unit setup cost for product } i \text{ in period } t,$$

is then assigned and the optimization problem becomes

$$\begin{aligned} \min \quad & \sum_{t \in T} \sum_{i \in \mathcal{I}} (c_{it} P_{it} + h_{it} I_{it} + f_{it} Y_{it}) \\ \text{s.to} \quad & P_{it} + I_{i,t-1} - I_{it} = d_{it}, \quad i \in \mathcal{I}, t \in T, \\ & \sum_{i \in \mathcal{I}} e_i P_{it} \leq b_t, \quad t \in T, \\ & P_{it} \leq \gamma Y_{it}, \quad i \in \mathcal{I}, t \in T, \\ & P_{it}, I_{it} \geq 0, Y_{it} \in \{0, 1\}, \quad i \in \mathcal{I}, t \in T. \end{aligned}$$

### 8.3.7. Bill of materials

A further extension of the basic planning model deals with the representation of products with a complex structure, described via the so-

called bill of materials, where end-items are made by components that in turn may include other components.

Formally, the following parameters are defined to describe the structure of the bill of materials:

$a_{ij}$  = units of product  $i$  directly required by one unit of product  $j$ ,

where the term product refers here to both end-items and components at various levels of the bill of materials. For each product  $i$  we assign an external demand  $d_{it}$  and an internal demand, the latter induced by the requirements of product  $I$  needed to manufacture the components or the end-items for which  $i$  represents a direct component. The external demand for components may originate from other plants of the same manufacturing company or from outside customers that also buy components.

The resulting optimization problem is formulated as

$$\begin{aligned}
 \min \quad & \sum_{t \in T} \sum_{i \in \mathcal{I}} (c_{it} P_{it} + h_{it} I_{it}) \\
 \text{s.to} \quad & P_{it} + I_{i,t-1} - I_{it} = d_{it} + \sum_{j \in \mathcal{I}, j \neq i} a_{ij} P_{jt}, \quad i \in \mathcal{I}, t \in T, \\
 & \sum_{i \in \mathcal{I}} e_i P_{it} \leq b_t, \quad t \in T, \\
 & P_{it}, I_{it} \geq 0, \quad i \in \mathcal{I}, t \in T.
 \end{aligned}$$

### 8.3.8. Multiple Plants

In this section it is assumed that a manufacturing company has a network of  $M$  production plants, located in geographically distinct sites, that manufacture a single product. The logistic system is responsible for supplying  $N$  peripheral depots, located in turn at distinct sites. Each production plant  $m \in M = \{1, 2, \dots, M\}$  is characterized by a maximum availability of product, denoted by  $s_m$ , while each plant  $n \in N = \{1, 2, \dots, N\}$  has a demand  $d_n$ . We further assume that a transportation cost  $c_{mn}$  is incurred by sending a unit of product from plant  $m$  to depot  $n$ , for each pair  $(m, n)$  of origins and destinations in the logistic network. The objective of the company is to determine an optimal logistic plan that satisfies at minimum cost the requests of the depots, without violating the maximum availability at the plants. It should be clear that the problem described arises frequently in logistic systems, at various levels in the logistic network (e.g., from suppliers to plants, from plants to warehouses or from warehouses to customers).

The decision variables needed to model the problem described represent the quantity to be transported for each plant-depot pair,

$$\begin{aligned}
 \min \quad & \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}} c_{mn} x_{mn} \\
 \text{s.to} \quad & \sum_{n \in \mathcal{N}} x_{mn} \leq s_m, \quad m \in \mathcal{M}, \\
 & \sum_{m \in \mathcal{M}} x_{mn} \geq d_n, \quad n \in \mathcal{N}, \\
 & x_{mn} \geq 0, \quad m \in \mathcal{M}, n \in \mathcal{N}.
 \end{aligned}$$

### 8.3.9. Revenue management systems

Revenue management is a managerial policy whose purpose is to maximize profits through an optimal balance between demand and supply. It is intended for marketing as well as logistic activities and has found growing interest in the service industry, particularly in the air transportation, tourism, and hotel sectors. More recently these methods have also begun to spread within the manufacturing and distribution industries.

The strong interest shown by such enterprises in the themes considered by revenue management should come as no surprise, if we consider the complexity and strategic relevance of decision-making processes concerning demand management, which are addressed by marketing and logistics managers. Consider, for example, the complex interactions among decisions on pricing, sales promotions, markdowns, mix definition and allocation to points of sale, in a highly dynamic and competitive context characterized by multiple sales channels and several alternative ways of contacting customers.

Despite the potential advantages that revenue management initiatives may offer for enterprises, there are certain difficulties that hamper the actual implementation of practical projects and actions aimed at adopting revenue management methodologies and tools. We can identify several explanations for the gap between intentions and initiatives undertaken. Certainly, the fear of implementation costs and uncertainty over the results that can be achieved play a key role, as happens for many innovation projects. Empirical investigations show, however, that the primary reason for prudence in adopting revenue management should be sought in the prerequisite conditions necessary to successfully start a revenue management project. There is an elevated level of interaction between revenue management and two other themes that we described earlier – optimization of the supply chain and relational marketing.

On the one hand, to apply revenue management methods and tools it is necessary to have an integrated and optimized logistic chain that guarantees the efficiency and responsiveness of the logistic flows. On the other hand, it is also necessary to possess a deep knowledge of the

customers and an accurate micro-segmentation of the market, achieved through data mining analytical models and therefore based on the analysis of the actual purchasing behaviors regularly recorded in the marketing data mart. Hence, to profitably adopt revenue management a company should be able to enhance and transform into knowledge, using business intelligence methodologies, the huge amount of information collected by means of automatic data gathering technologies.

---

## 8.4. SUMMARY

---

In this chapter we learn the different things about

**Relational marketing-** Marketing decision processes are characterized by a high level of complexity due to the simultaneous presence of multiple objectives and countless alternative actions resulting from the combination of the major choice options available to decision makers. Also concept regarding Motivations and objectives, An environment for relational marketing analysis, Lifetime value, The effect of latency in predictive models, etc.

**Salesforce management-** The term *salesforce* is taken to mean the complete set of people and roles that are involved, with different tasks and responsibilities, in the sales process.

**Logistic and production models-** Besides acting on the marketing control levers, a manufacturing company can achieve further reductions in costs by improving its processes in another area that has received increasing attention in recent years: an effective supply chain management, understood as the logistic and production processes of a single enterprise as well as the network of companies composing the production chain of a given industry.

---

## 8.5. QUESTIONS

---

1. What does relational Marketing explain in detail.
2. What are the Motivations and objectives of Relational marketing?
3. What is the effect of latency in predictive models.
4. What is the difference between Cross-selling and up-selling.
5. Explain the concept of Market basket analysis.
6. Write a brief note on Salesforce management.
7. Explain the Models for salesforce management
8. What is mean by Supply chain optimization.
9. Explain Optimization models for logistics planning.
10. Write a brief note on Backlogging.
11. What is the difference Minimum lots and fixed costs.
12. Explain Revenue management systems

---

## 8.6. REFERENCE

---

Marketing, Logistic and  
production models

1. Business Intelligence: Data Mining and Optimization for Decision Making, Carlo Vercellis Wiley First 2009
2. Decision support and Business Intelligence Systems, Efraim Turban, Ramesh Sharda, Dursun Delen, Pearson Ninth 2011
3. Fundamental of Business Intelligence Grossmann W, Rinderle-Ma Springer First 2015



# DATA ENVELOPMENT ANALYSIS

## Unit Structure

- 9.0. Objective
- 9.1 Efficiency measures
- 9.2 Efficient frontier
- 9.3 The CCR model
  - 9.3.1 Definition of target objectives
  - 9.3.2 Peer groups
- 9.4 Identification of good operating practices
  - 9.4.1 Cross-efficiency analysis
  - 9.4.2 Virtual inputs and virtual outputs
  - 9.4.3 Weight restrictions
- 9.5. Summary
- 9.6. Questions
- 9.7. Reference

---

## 9.0. OBJECTIVE

---

This chapter would make you understand the following concepts:

- Data envelopment analysis and how they amylase using the various measurements like Efficiency measures, Efficient frontier, and the CCR models in details.
- Identification of good operating practices like Cross-efficiency analysis, Virtual inputs and virtual outputs, Weight restrictions.

### Data envelopment analysis

The purpose of data envelopment analysis (DEA) is to compare the operating performance of a set of units such as companies, university departments, hospitals, bank branch offices, production plants, or transportation systems. In order for the comparison to be meaningful, the units being investigated must be homogeneous.

A unit's performance can be evaluated in a number of dimensions. For instance, one can use flexibility indicators to measure a system's capacity to respond to changes in requirements quickly and affordably, as well as quality indicators to estimate the rate of rejects resulting from manufacturing a set of products, to assess the activity of a production plant.

A productivity indicator that offers a measure of the efficiency that characterises the operating activity of the units being compared is the foundation of data envelopment analysis. This measure is based on the results obtained by each unit, which will be referred to as outputs, and on the resources utilized to achieve these results, which will be generically designated as inputs or production factors. If the units are bank branches, the inputs may be the number of cashiers, managers, or rooms utilised at each branch, while the outputs could be the number of active bank accounts, checks cashed, or loans raised. The number of active teaching courses and scientific papers created by each department's members may be taken into consideration as outputs if the units are university departments; The inputs might be things like the funding each department receives, the expense of instruction, the administrative staff, and the accessibility of offices and labs.

## 9.1 EFFICIENCY MEASURES

---

Decision-making units (DMUs) are the units that are being compared in data envelopment analysis because they have some degree of decisional autonomy. Let  $N = 1, 2, \dots, n$  be the collection of units being compared, with the assumption that we want to assess the effectiveness of  $n$  units.

If the units produce a single output using a single input only, the efficiency of the  $j^{\text{th}}$  decision-making unit  $DMU_j$ ,  $j \in N$ , is defined as

$$\theta_j = \frac{y_j}{x_j},$$

in which  $y_j$  is the output value produced by  $DMU_j$  and  $x_j$  the input value used.

The ratio between a weighted sum of the outputs and a weighted sum of the inputs is used to quantify the efficiency of  $DMU_j$  when the units use a variety of input parameters to produce multiple outputs of the inputs. Denote by  $H = \{1, 2, \dots, s\}$  the set of production factors and by  $K = \{1, 2, \dots, m\}$  the corresponding set of outputs. If  $x_{ij}$ ,  $i \in H$ , denotes the quantity of input  $i$  used by  $DMU_j$  and  $y_{rj}$ ,  $r \in K$ , the quantity of output  $r$  obtained, the efficiency of  $DMU_j$  is defined as

$$\theta_j = \frac{u_1 y_{1j} + u_2 y_{2j} + \dots + u_m y_{mj}}{v_1 x_{1j} + v_2 x_{2j} + \dots + v_s x_{sj}} = \frac{\sum_{r \in K} u_r y_{rj}}{\sum_{i \in H} v_i x_{ij}},$$

for weights  $u_1, u_2, \dots, u_m$  associated with the outputs and  $v_1, v_2, \dots, v_s$  assigned to the inputs.

In this second case, the efficiency of  $DMU_j$  depends strongly on the system of weights introduced. It becomes challenging to define a single structure of weights that might be shared and approved by all the assessed units because at different weights, the efficiency value may experience important variances. In order to avoid potential objections from the units

to a predetermined system of weights that may favour some DMUs over others, data envelopment analysis evaluates each unit's efficiency using the weights system that is best for the DMU itself—that is, the system that permits the efficiency value of the DMU to be maximised. The goal of data envelopment analysis is to identify the units that are efficient in absolute terms and those whose efficiency value mostly depends on the chosen system of weights through further investigations.

## 9.2 Efficient frontier

The link between the inputs used and the produced outputs is expressed by the efficient frontier, also known as the production function. It specifies the greatest number of outputs that can be produced from a particular set of inputs. It also expresses the bare minimal amount of inputs necessary to reach a specific output level. Thus, operating techniques that are technically efficient correspond to the efficient frontier.

A set of observations that represent the output level attained by using a particular combination of input production parameters can be used to empirically determine the efficient frontier. The observations in a data envelopment analysis relate to the units being assessed. The majority of parametric statistical techniques, such as those that compute regression curves, make some prior beliefs regarding the production function's structure. On the other hand, data envelopment analysis is nonparametric in nature since it makes no assumptions about the functional form of the efficient frontier. It just needs that, depending on their efficiency value, the units being compared are not positioned above the production function.

To further clarify the notion of efficient frontier, consider Example 9.1.

A possible alternative to the efficient frontier is the regression line that can be obtained based on the available observations, indicated in Figure 9.1 by a dashed line. The units that are above the regression line in this situation may be considered good, and the level of excellence of each unit might be described by how far it is from the line. However, it is important to emphasise the distinction between the efficient frontier achieved using data envelopment analysis and the prediction line generated using a regression model.

Table 9.1 Input and output values for the bank branches in Example 9.1

bank branch	staff size	accounts value	efficiency
A	3	2.5	0.733
B	2	1.0	0.500
C	5	2.7	0.540
D	3	3.0	1.000
E	7	5.0	0.714
F	5	2.3	0.460
G	4	3.2	0.700
H	5	4.5	0.900
I	6	4.5	0.633

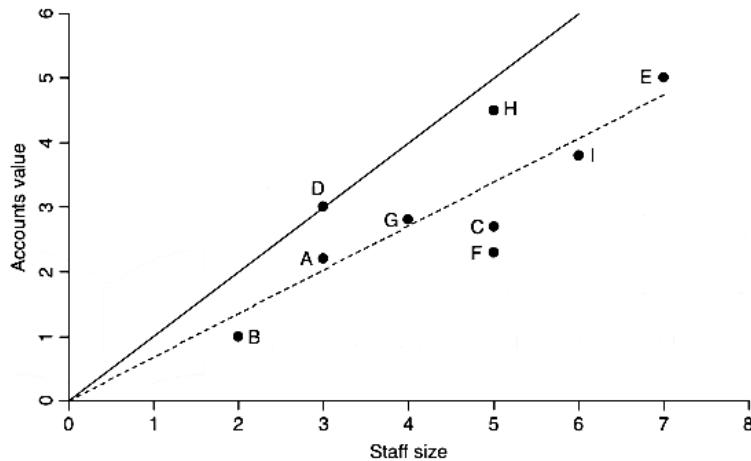


Figure 9.1 Evaluation of efficiency of bank branches regression line

Figure 9.1 Evaluation of efficiency of bank branches regression line reflects the average behavior of the units being compared, while the efficient frontier identifies the best behavior, and measures the inefficiency of a unit based on the distance from the frontier itself.

Notice also that the efficient frontier provides some indications for improving the performance of inefficient units. Indeed, it identifies for each input level the output level that can be achieved in conditions of efficiency. By the same token, it identifies for each output level the minimum level of input that should be used in conditions of efficiency. In particular, for each  $DMU_j$ ,  $j \in N$ , the input-oriented efficiency  $\theta_{Ij}$  can be defined as the ratio between the ideal input quantity  $x^*$  that should be used by the unit if it were efficient and the actually used quantity  $x_j$ :

Similarly, the output-oriented efficiency  $\theta_{Oj}$  is defined as the ratio between the quantity of output  $y_j$  actually produced by the unit and the ideal quantity  $y^*$  that it should produce in conditions of efficiency:

$$\theta_{Ij}^O = \frac{y_j}{y^*}.$$

The problem of making an inefficient unit efficient is then turned into one of devising a way by which the inefficient unit can be brought close to the efficient frontier. If the unit produces a single output only by using two inputs, the efficient frontier assumes the shape shown in Figure 9.2. In this case, the inefficiency of a given unit is evaluated by the length of the segment connecting the unit to the efficient frontier along the line passing through the origin of the axes.

For the example illustrated in Figure 9.2, the efficiency value of  $DMU_A$  is given by

$$\theta_A = \frac{\overline{OP}}{\overline{OA}},$$

where  $OP$  and  $OA$  represent the lengths of segments  $OP$  and  $OA$ , respectively. The inefficient unit may be made efficient by a displacement along segment  $OA$  that moves it onto the efficient frontier. Such displacement is tantamount to progressively decreasing the quantity of both inputs while keeping unchanged the quantity of output. In this case, the production possibility set is defined as the region delimited by the efficient frontier where the observed units being compared are found.

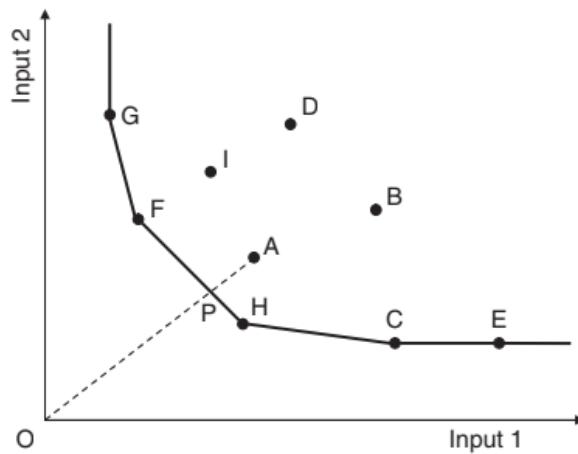


Figure 9.2 Efficient frontier with two inputs and one output

### 9.3 THE CCR MODEL

To select the best set of weights for a generic  $DMU_j$  using data envelopment analysis, a mathematical optimisation model whose decision variables are represented by the weights  $u_r$ ,  $r \in \mathcal{K}$ , and  $v_i$ ,  $i \in \mathcal{H}$ , connected to each output and input, must be solved. The Charnes-Cooper-Rhodes (CCR) model is arguably the most well-known formulation among the several that have been put out.

The CCR model formulated for  $DMU_j$  takes the form

$$\begin{aligned} \max \quad & \vartheta = \frac{\sum_{r \in \mathcal{K}} u_r y_{rj}}{\sum_{i \in \mathcal{H}} v_i x_{ij}}, \\ \text{s.to} \quad & \frac{\sum_{r \in \mathcal{K}} u_r y_{rj}}{\sum_{i \in \mathcal{H}} v_i x_{ij}} \leq 1, \quad j \in \mathcal{N}, \\ & u_r, v_i \geq 0, \quad r \in \mathcal{K}, i \in \mathcal{H}. \end{aligned}$$

The objective function involves the maximization of the efficiency measure for  $DMU_j$ . Constraints (9.7) require that the efficiency values of

all the units, calculated by means of the weights system for the unit being examined, be lower than one. Finally, conditions (9.8) guarantee that the weights associated with the inputs and the outputs are non-negative. In place of these conditions, sometimes the constraints  $u_r, v_i \geq \delta, r \in K, i \in H$  may be applied, where  $\delta > 0$ , preventing the unit from assigning a null weight to an input or output.

Model (9.6) can be linearized by requiring the weighted sum of the inputs to take a constant value, for example 1. This condition leads to an alternative optimization problem, the input-oriented CCR model, where the objective function consists of the maximization of the weighted sum of the outputs

$$\begin{aligned} \max \quad & \vartheta = \sum_{r \in K} u_r y_{rj}, \\ \text{s.to} \quad & \sum_{i \in H} v_i x_{ij} = 1, \\ & \sum_{r \in K} u_r y_{rj} - \sum_{i \in H} v_i x_{ij} \leq 0, \quad j \in N, \\ & u_r, v_i \geq 0, \quad r \in K, i \in H. \end{aligned}$$

Let  $\theta^*$  be the optimum value of the objective function corresponding to the optimal solution  $(v^*, u^*)$  of problem (9.9). If  $\theta^* = 1$  and there is at least one optimal solution  $(v, u)$  where  $v > 0$  and  $u > 0$  then DMUj is said to be efficient. One can get n systems of weights by solving a comparable optimization model for each of the n units being compared. The flexibility enjoyed by the units in choosing the weights represents an undisputed advantage, in that an inefficient unit cannot be blamed for being inefficient if the most advantageous system of weights reveals that it is inefficient. The question of whether a unit's efficiency value should be attributed to an actual high level of performance or only to an ideal choice of the weights structure must be answered for a unit that receives a score of = 1.

### Dual of the CCR model

For the input-oriented CCR model, the following dual problem, which lends itself to an interesting interpretation, can be formulated:

$$\begin{aligned} \min \quad & \vartheta, \\ \text{s.to} \quad & \sum_{j \in N} \lambda_j x_{ij} - \vartheta x_{ij} \leq 0, \quad i \in H, \\ & \sum_{j \in N} \lambda_j y_{rj} - y_{rj} \geq 0, \quad r \in K, \\ & \lambda_j \geq 0, \quad j \in N. \end{aligned}$$

Based on the optimum value of the variables  $\lambda_j^*$ ,  $j \in N$ , the aim of model (9.13) is to identify an ideal unit that lies on the efficient frontier and represents a term of comparison for  $DMU_j$ . The output of this unit must be at least equal to the output of  $DMU_j$ , and it must consume an amount of inputs that is a portion of the amount used by the unit being studied, according to model constraints (9.14) and (9.9). The ideal value of the dual variable is defined as the ratio between the input utilized by the ideal unit and the input absorbed by  $DMU_j$ .  $DMU_j$  is below the efficient frontier if  $\lambda_j^* = 1$ . This unit must use  $x_{ij}$ ,  $i \in H$  of each input in order to be effective.

### 9.3.1 Definition of target objectives

In real-world applications it is often desirable to set improvement objectives for inefficient units, in terms of both outputs produced and inputs utilized. In this regard, data envelopment analysis offers crucial advice by identifying the output and input levels at which a specific inefficient unit may become efficient. The efficiency score of a unit expresses the maximum proportion of the utilized inputs that the unit should use in conditions of efficiency, in order to guarantee its current output levels. As an alternative, the inverse of the efficiency score shows the multiplier that should be applied to a unit's current output levels in order to make the unit more efficient while maintaining the same level of productive inputs. Based on the efficiency values, data envelopment analysis therefore gives a measure for each unit being compared of the savings in inputs or the increases in outputs required for the unit to become efficient.

To determine the target values, it is possible to follow an input- or output-oriented strategy. The target values for inputs and outputs are provided by in the first instance, when the improvement objectives principally concern the resources consumed.

$$x_{ij}^{\text{target}} = \vartheta^* x_{ij} - s_i^{-*}, \quad i \in \mathcal{H},$$

$$y_{rj}^{\text{target}} = y_{rj} + s_r^{+*}, \quad r \in \mathcal{K}.$$

In the second case, target values for inputs and outputs are given by

$$x_{ij}^{\text{target}} = x_{ij} - \frac{s_i^{-*}}{\vartheta^*}, \quad i \in \mathcal{H},$$

$$y_{rj}^{\text{target}} = \frac{y_{rj} + s_r^{+*}}{\vartheta^*}, \quad r \in \mathcal{K}.$$

Other performance improvement strategies may be preferred over the proportional reduction in the quantities of inputs used or the proportional increase in the output quantities produced:

- priority order for the production factors – the target values for the inputs are set in such a way as to minimize the quantity used of the resources to which the highest priority has been assigned, without allowing variations in the level of other inputs or in the outputs produced;
- priority order for the outputs – the target values for the outputs are set in such a way as to maximize the quantity produced of the outputs to which highest priority has been assigned, without allowing variations in the level of other outputs or inputs used;
- preferences expressed by the decision makers with respect to a decrease in some inputs or an increase in specific outputs.

### 9.3.2 Peer groups

Data envelopment analysis identifies for each inefficient unit a set of excellent units, called a peer group, which includes those units that are efficient if evaluated with the optimal system of weights of an inefficient unit. The peer group, made up of DMUs which are characterized by operating methods similar to the inefficient unit being examined, is a realistic term of comparison which the unit should aim to imitate in order to improve its performance.

The units included in the peer group of a given unit  $DMU_j$  may be identified by the solution to model (9.9). Indeed, these correspond to the DMUs for which the first and the second member of constraints (9.11) are equal:

$$E_j = \left\{ j : \sum_{r \in \mathcal{K}} u_r^* y_{rj} = \sum_{i \in \mathcal{H}} v_i^* x_{ij} \right\}.$$

Alternatively, with respect to formulation (9.13), the peer group consists of those units whose variable  $\lambda_j$  in the optimal solution is strictly positive:

$$E_j = \{ j : \lambda_j^* > 0 \}.$$

Notice that within a peer group a few excellent units more than others may represent a reasonable term of comparison. The relative importance of a unit belonging to a peer group depends on the value of the corresponding variable  $\lambda_j$  in the optimal solution of the dual model.

The analysis of peer groups allows one to differentiate between really efficient units and apparently efficient units for which the choice of an optimal system of weights conceals some abnormal behavior. In order to draw this distinction, it is necessary to consider the efficient units and to evaluate how often each belongs to a peer group. One may reasonably expect that an efficient unit

often included in the peer groups uses for the evaluation of its own efficiency a robust weights structure. Conversely, if an efficient unit rarely represents a term of comparison, its own system of optimal weights may appear distorted, in the sense that it may implicitly reflect the specialization of the unit along a particular dimension of analysis.

---

## 9.4 IDENTIFICATION OF GOOD OPERATING PRACTICES

---

By identifying and sharing good operating practices, one may hope to achieve an improvement in the performance of all units being compared. The units that appear efficient according to data envelopment analysis certainly represent terms of comparison and examples to be imitated for the other units. However, among efficient units some more than others may represent a target to be reached in improving the efficiency.

The need to identify the efficient units, for the purpose of defining the best operating practices, stems from the principle itself on which data envelopment analysis is grounded, since it allows each unit to evaluate its own degree of efficiency by choosing the most advantageous structure of weights for inputs and outputs. In this way, a unit might appear efficient by purposely attributing

a non-negligible weight only to a limited subset of inputs and outputs. Further-more, those inputs and outputs that receive greater weights may be less critical than other factors more intimately connected to the primary activity performed by the units being analyzed. In order to identify good operating practices, it is therefore expedient to detect the units that are really efficient, that is, those

units whose efficiency score does not primarily depend on the system of weights selected. To differentiate these units, we may resort to a combination of different methods: cross-efficiency analysis, evaluation of virtual inputs and virtual outputs, and weight restrictions.

### 9.4.1 Cross-efficiency analysis

The analysis of cross-efficiency is based on the definition of the efficiency matrix, which provides information on the nature of the weights system adopted by the units for their own efficiency evaluation. The square efficiency matrix contains as many rows and columns as there are units being compared. The generic element  $\theta_{ij}$  of the matrix represents the efficiency of  $DMU_j$  evaluated

through the optimal weights structure for  $DMU_i$ , while the element  $\theta_{jj}$  provides the efficiency of  $DMU_j$  calculated using its own optimal weights. If  $DMU_j$  is efficient (i.e. if  $\theta_{jj} = 1$ ), although it exhibits a behavior specialized along a given dimension with respect to the other units, the efficiency values in the column corresponding to  $DMU_j$  will be less than 1.

Two quantities of interest can be derived from the efficiency matrix. The first represents the average efficiency of a unit with respect to the optimal weights systems for the different units, obtained as the average of the values in the  $j^{\text{th}}$  column. The second is the average efficiency of a unit measured applying its optimal system of weights to the other units. The latter is obtained by averaging the values in the row associated with the unit being examined. The difference between the efficiency score  $\theta_{jj}$  of  $\text{DMU}_j$  and the efficiency obtained as the average of the values in the  $j^{\text{th}}$  column provides an indication of how much the unit relies on a system of weights conforming with the one used by the other units in the evaluation process. If the difference between the two terms is significant,  $\text{DMU}_j$  may have chosen a structure of weights that is not shared by the other DMUs in order to privilege the dimensions of analysis on which it appears particularly efficient.

#### 9.4.2 Virtual inputs and virtual outputs

Virtual inputs and virtual outputs provide information on the relative importance that each unit attributes to each individual input and output, for the purpose of maximizing its own efficiency score. Thus, they allow the specific competencies of each unit to be identified, highlighting at the same time its weaknesses. The virtual inputs of a DMU are defined as the product of the inputs used by the unit and the corresponding optimal weights. Similarly, virtual outputs are given by the product of the outputs of the unit and the associated optimal weights. Inputs and outputs for which the unit shows high virtual scores provide an indication of the activities in which the unit being analyzed appears particularly efficient. Notice that model (9.9) admits in general multiple optimal solutions, corresponding to which it is possible to obtain different combinations of virtual inputs and virtual outputs. Two efficient units may yield high virtual values corresponding to different combinations of inputs and outputs, showing good operating practices in different contexts. In this case, it might be convenient for each unit to follow the principles and operating methods shown by the other, aiming at improving its own efficiency on a specific dimension.

#### 9.4.3 Weight restrictions

To separate the units that are really efficient from those whose efficiency score largely depends on the selected weights system, we may impose some restrictions on the value of the weights to be associated with inputs and outputs. In general, these restrictions translate into the definition of maximum thresholds for the weight of specific outputs or minimum thresholds for the weight of specific inputs. Notice that, despite possible restrictions on the weights, the units still enjoy a certain flexibility in the choice of multiplicative factors for inputs and outputs. For this reason, it may be useful to resort to the evaluation of virtual inputs and virtual outputs in order to identify the units with the most efficient operating practices with respect to the usage of a specific input resource or to the production of a given output.

---

## 9.5. SUMMARY

---

The purpose of data envelopment analysis (DEA) is to compare the operating performance of a set of units such as companies, university departments, hospitals, bank branch offices, production plants, or transportation systems.

The efficient frontier, also known as production function, expresses the relationship between the inputs utilized and the outputs produced.

Using data envelopment analysis, the choice of the optimal system of weights for a generic  $MU_j$  involves solving a mathematical optimization model whose decision variables are represented by the weights  $u_r$ ,  $r \in K$ , and  $v_i$ ,  $i \in H$ , associated with each output and input.

---

## 9.6. QUESTIONS

---

1. Write short note on Data envelopment analysis.
2. How we can measure Efficiency.
3. Write a short note on CCR model.
4. Write a Definition of target objectives.
5. What is mean by Cross-efficiency analysis.
6. Write a short note on Virtual inputs and virtual outputs.

---

## 9.7. REFERENCE

---

1. Business Intelligence: Data Mining and Optimization for Decision Making, Carlo Vercellis Wiley First 2009
2. Decision support and Business Intelligence Systems, Efraim Turban, Ramesh Sharda, Dursun Delen, Pearson Ninth 2011
3. Fundamental of Business Intelligence Grossmann W, Rinderle-Ma Springer First 2015



# KNOWLEDGE MANAGEMENT

## Unit Structure

- 10.0 Objectives
- 10.1 Introduction to Knowledge Management
  - 10.1.1 Define Knowledge Management
  - 10.1.2 What is Data, Information and Knowledge
- 10.2 The Knowledge Management Process
- 10.3 Organizational Learning and Transformation
  - 10.3.1 Organizational Learning
  - 10.3.2 Organizational Transformation
- 10.4 Power of Knowledge Management
  - 10.4.1 Approaches to Knowledge Management
- 10.5 Information Technology in Knowledge Management
  - 10.5.1 Knowledge Management System (KMS) Cycle
- 10.6 Knowledge Management Systems Implementation
- 10.7 Summary
- 10.8 Questions
- 10.9 References

---

## 10.0 OBJECTIVES:

---

After going through this unit, you will be able to understand:

- Define Knowledge Management
- What exactly a Data, Information and Knowledge is
- Detail study of Knowledge Management process
- Knowledge Management Systems Cycle
- Knowledge Management Systems Implementations.

---

## 10.1 INTRODUCTION TO KNOWLEDGE MANAGEMENT

---

Knowledge management is the process of capturing, distributing, and effectively using knowledge. Knowledge management has developed as a way to make sense of information collected with the help of business intelligence and utilize it for the best possible way in business expansion.

Knowledge management framework defines the knowledge gathering points, various techniques, tools, and data storing tools, techniques and analyzing mechanism.

The main purpose of Knowledge Management includes effective and efficient problem solving, dynamic learning, strategic planning and decision making.

Knowledge Management initiatives focus on identifying knowledge, shared it with formal manner and increasing its value for reuse.

### **10.1.1 Define Knowledge Management**

- The process of creating, using, sharing and managing the information and knowledge of an organization is known as Knowledge Management.
- It is an activity of multidisciplinary area where goal is by making the best use of knowledge to fulfill organizational objectives.
- In this process various enterprises will collect the information with the help of many methods and various tools which are available in market.
- Collected information analysis is done with the help of different techniques, these analyzed information is totally depends on resources, soft and hard copy of documents, people, and their skills.
- The main objective of Knowledge Management is how to improve the performance.
- Knowledge Management includes
  1. Individual knowledge, transfer into database.
  2. Filtration and separation of the most relevant knowledge.
  3. For providing easy access organization of knowledge is essential and later it provides to employees depends on their needs.
  4. Automation process for organizational knowledge and management is available in most of the organizations.
  5. Because of Automation process for storing, retrieval and sharing of databases have become very convenient.

### **10.1.2 What is data, Information and Knowledge**

Data: Facts, Figures and Measurements.

Information: It is processed data or organized data.

Time and accuracy are most crucial factor.

Knowledge: It is information that is contextual, relevant, and actionable and describe its purpose.

- Any organization can use the process of knowledge management which is universal.
- The resources (tools and techniques) can be unique and usage is depended on need and the environment of an organization.
- There are six basic steps involved in Knowledge Management with the help of various techniques and tools.
- Six basic steps are Collecting, Organizing, Summarizing, Analyzing, Synthesizing and Decision Making.
- While following these six steps sequentially the data transforms into knowledge.
- The main objective of knowledge management process is to transform the data into knowledge.

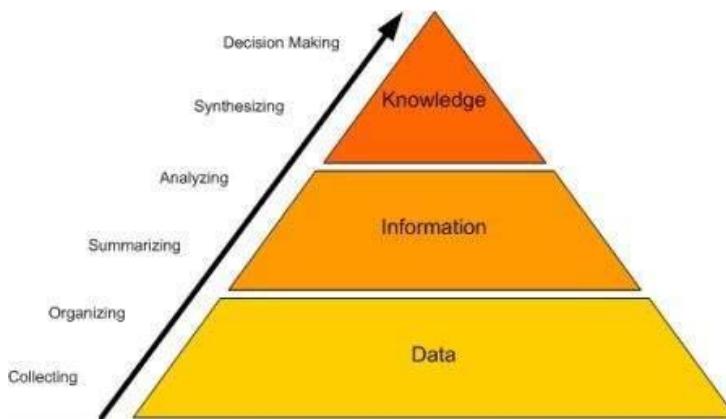


Figure:10.1 Knowledge Management Process

[Source: <https://www.semanticscholar.org>]

### Step 1. Collecting

- The key phase in the knowledge management process is this one. It is possible that the knowledge you produce will not be the most accurate if you gather inaccurate or irrelevant facts. As a result, decisions that are based on such information may also be incorrect.
- The methods and resources used for data collection are numerous. Data collecting should be a step in the knowledge management process, to start. People involved in the data collection process should appropriately document and follow these protocols.
- Specific data collecting points are specified by the data gathering process.
- The methods and instruments for data extraction are defined together with the data gathering points. For instance, the daily attendance report might be an online report that is instantly entered in the database, whereas the sales report might be a paper-based report that requires a data entry operator to manually input the data.

- In this step, data storage is defined together with data collection sites and data extraction methods. Most businesses today employ a software database application for this.

## **Step 2. Organizing**

- The data collected need to be organized. This organization usually happens based on certain rules. These rules are defined by the organization.
- As an example, all sales-related data can be filed together, and all staff-related data could be stored in the same database table. This type of organization helps to maintain data accurately within a database.
- If there is much data in the database, techniques such as 'normalization' can be used for organizing and reducing the duplication.
- This way, data is logically arranged and related to one another for easy retrieval. When data passes step 2, it becomes information.

## **Step 3. Summarizing**

- The information is summarized to take the essence of it. The lengthy information is presented in tabular or graphical format and stored appropriately.
- For summarizing, there are many tools that can be used such as software packages, charts (Pareto, cause-and-effect), and different techniques.

## **Step 4. Analyzing**

- The information analyzed to find the relationships, redundancies, and patterns.
- An expert team ought to be assigned for this purpose because the expertise from the team plays a significant role.
- There are reports created during the analysis of information.

## **Step 5. Synthesizing**

- In this stage the information converts into the knowledge. The results of research (usually the reports) are combined to derive numerous concepts and data.
- A pattern or behavior of one entity is applied to explain another, and collectively, the organization can have a set of data elements that may be used across the organization.
- This knowledge is then stored within the organizational knowledge base for more use.

## **Step 6. Decision Making**

- In this stage the knowledge is used for decision making.
- Example, once estimating a specific type of a project or a task, the knowledge related to previous estimation is used.

- This accelerates the estimation process and adds high accuracy. This is how the organizational knowledge management adds value and saves money overall.

### **10.3 ORGANIZATIONAL LEARNING AND TRANSFORMATION**

People are the holders of knowledge. The main aim or goal is to encourage them to not only search for it and improve it for applying it to improving internal processes, but to make them see the benefits of sharing it with the organizations. It is important for:

- To provide proper storage and sharing of knowledge systems.
- To give people autonomy in their jobs and find new ways to fulfill them.
- To empower them and continually train them.
- To give them adequate remuneration, to ensure their commitment.

The manager should always know decisions made by people can affect the entire organization.

Employees will share the knowledge they accumulate in their activities in the company with colleagues.

Only the disadvantage is to lose that talent to the competition, along with everything they have learned.

#### **10.3.1 Organizational Learning**

The learning organization is an organization characterized by a deep commitment to learning and education with the intention of continuous improvement.

This concept reviews several theories relating to the learning organization.

As an aspect of an organization, organizational learning is that the method of creating, retaining, and transferring data.

The organizational learning leads to an enhanced ability to react quickly to opportunities and threats.

#### **10.3.2 Organizational Transformation**

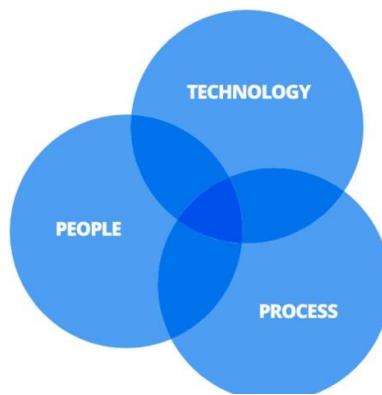


Figure 10.2 Organizational Transformation

- A change management business strategy called organizational transformation aims to shift your organization from its current condition to a desired future one. These change initiatives include initiatives aimed at improving the employee experience.
- The attitude of the employees, their perspectives as well as the culture of the organization undergoes a meaningful change.
- In short, It is about re-modelling an organization.

There are three key factors for managing organizational transformation along with the critical success factors for managing change at each stage.

Step 1) Break with the past

Step 2) Manage the present

Step 3) Invest in the future

### **Step 1) Break with the past**

- Introduce entrepreneurial outsiders with the targeted expertise onto the top management team.
- Break with your administrative heritage. Important steps are removal of blockers, rotation of managers, promotion of young managers with the help of organizational heritage, designing a suitable bonus or incentive system.
- The useful administrative heritage of the past will be continued where non useful processes in the past will be thrown away.
- It will vary from organization to organization. The traditional way is command and control management style to achieve more quickly implementation of change, whereas more democratic leadership style is in the form is appropriate to leverage example, customer relationships, a strong R&D department, or latest talent enthusiasm of organizational members for participating a new idea.

### **Step 2) Manage the present**

- Top-down approach of stage 1 where vary your leadership style as appropriate. It may require breaking with the past in some part of the organization, whereas other part of the organization can learn so it will help in empowerment to act.
- For streamlining business organization system reconfigure, divest, and integrate resources are important strategies whereas removing non-aligned employees to consolidating new acquisitions operationally and culturally.

- Empower the organization. Top management team should delegate to employees as well as motivating and enabling them to act.
- Enable the organization to engage in exploration of new strategies, encouraging innovations, trial, and experimentation for developing a culture.
- Creating new path or capabilities in terms of new products, services, processes for improvement in organizational model.
- With the help of these three stages, organization can establish new development pathways where enhancement in their strategic flexibility and react successfully to changes in the environment.

---

## **10.4 POWER OF KNOWLEDGE MANAGEMENT**

---

Knowledge Management Activities involve three activities

- Knowledge creation
- Knowledge sharing
- Knowledge seeking

### **1) Knowledge creation**

- To create new knowledge means existing knowledge will use for recreating the company and everyone in it in a nonstop process of personal and organizational self-renewal.
- For creation of new knowledge, the employees must create innovative ideas.
- Socialization, Externalization, Internalization, and combination are the various four modes for conversion of tacit knowledge to new knowledge with the help of social interaction and sharing experiences among organizational members.
- Externalization is the converting tacit knowledge to new explicit knowledge.
- Internalization is the converting tacit knowledge from explicit knowledge.
- External team and Internal team members of the organization refers to the creation of new explicit knowledge by merging, categorizing, and reclassifying existing explicit knowledge.

### **2) Knowledge sharing**

- Most of the organization, information and knowledge are not considered organizational resource to be shared but individual competitive weapons to be kept private.

- Organizational members may share personal knowledge with trepidation, they perceive that they are of less value if their knowledge is a part of the organization public domain.
- Knowledge sharing is the ideas insights, solutions, experience of individual person to another individual with the help of intermediary i.e., ICT based system or directly.

### 3) **Knowledge seeking**

- It is source or search for use of internal organizational knowledge.
- Individual may sometimes prefer to not reuse knowledge if they feel that their own performance review is based on the originality or creativity of their innovative ideas.

#### **10.4.1 Approaches to knowledge Management**

There are three fundamental approaches to knowledge management process

- Process approach to knowledge management
- Practice approach to knowledge management
- Hybrid approach to knowledge management

##### **1) Process approach to knowledge management**

- Through established controls, processes, and technology, the process approach to knowledge management aims to codify corporate knowledge.
- Organizations that use the process method may put in place explicit regulations dictating how information should be gathered, maintained, and shared across the whole business.

##### **2) Practice approach to knowledge management**

- It assumes that the organizational knowledge is tacit in nature and that formal controls, processes and various technologies are not suitable for transmitting the proper understanding.
- The main aim of this aspect is to build the social environment or communities of practice necessary to facilitate the sharing of tacit understanding.

##### **3) Hybrid approach to knowledge management**

- Most of the organization uses the hybrid model of the knowledge management.

## 10.5 INFORMATION TECHNOLOGY IN KNOWLEDGE MANAGEMENT

Information Technology plays a vital role in the management of knowledge. There are two functions of Information Technology in aspect of knowledge management namely retrieval and communication.

- Usage of knowledge and enhance the speed of knowledge transfer that enables the use of information technology.
- Information technology is more important for storage and retrieval the knowledgeable information.
- Whereas capturing, storing, and managing the tacit knowledge usually requires a various set of tools and techniques.
- E-content management system requires specialized tools and storage system that are part of collaborative computing system. This system is known as Knowledge repositories.

Study of Knowledge management technologies and web impact as follows:

Knowledge Management	Web Impact	Impact on the web
Communication	User friendly GUI system, improves communication tools convenient, fast access to knowledge and knowledgeable individuals.	Knowledge capture and share information is used many areas to improving information and communication technology.
Collaboration	It helps in collaboration between organization, customer and vendors.	Capturing and sharing of knowledge is useful in enhancement of collaboration, management of collaboration and technology.
Storage and Retrieval	Friendly GUI system for clients and server provide for efficient and effective storage and retrieval of knowledge.	Knowledge capture and share is utilized in improving data storage and retrieval systems, database management and knowledge repository technology.

Table 10.1 Knowledge Management

### 10.5.1 Knowledge Management System (KMS) Cycle

Knowledge management system cycle is the process of transforming information into knowledge within an organization which explains how knowledge is captured, processed, and distributed in an organization.

The Knowledge Management Cycle system consist of six different steps of functions:

Step 1) Creating Knowledge: Knowledge is created by sharing of ideas by people working in an organization which helps to leads better ideas were creating a valuable knowledge repository.

Step 2) Capturing Knowledge: Capture and acquisition of knowledge is one in which the knowledge created is collected in vast numbers and store it in a knowledge repository.

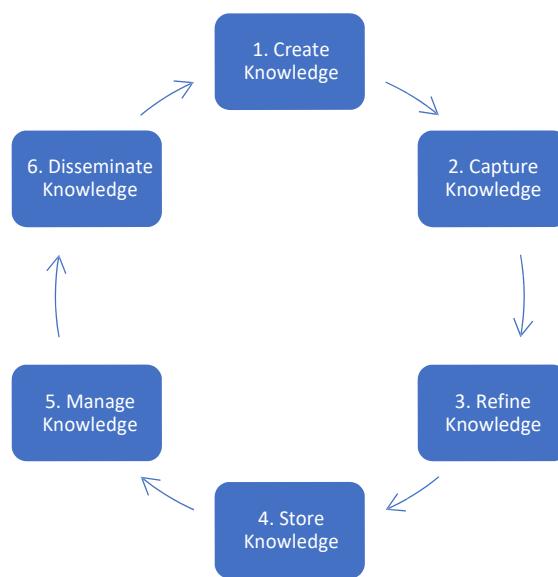
Step 3) Refine Knowledge: Refine is the next step to capturing of knowledge. The captured knowledge is organized using s framework or the knowledge model. The model shows the various elements of the knowledge and flows that are embedded inherently in the specific processes and culture of organization.

Step 4) Store Knowledge: Useful and processed knowledge must be stored in a reasonable format in a knowledge repository so that we can access it whenever need.

Step 5) Manage Knowledge: The organized knowledge is put in a such a way that it can be accessed, searched and manageable with problem solving by the users working in that organization.

Step 6) Disseminate Knowledge: The organization knowledge is put in such a way that it can be accessed, search and disseminated by the users working in the organization.

### 10.3 Knowledge Management Cycle



[Source: researchgate.net]

Following are the steps to Implement

Implementing a knowledge management program is no easy feat. Some familiar challenges are:

- Inability to recognize or articulate knowledge turning tacit knowledge into explicit knowledge.
- Geographical distance or language barriers in an international organization.
- Limitations of information and communication technologies.
- Loosely defined areas of expertise.
- Constantly changing business.
- Internal conflicts (e.g., professional territoriality)
- Lack of incentives or performance management goals.
- Poor training or mentoring programs.
- Cultural barriers.

To minimize the risk and maximizing the rewards with overcome the above challenges we can plan them appropriately with the help of following eight steps:

The early steps involve strategy, planning and requirements gathering while the later steps involve on execution and continual improvement.

### **Step 1) Establish knowledge management program objectives**

The ideal end state should be envisioned and expressed before choosing a tool, establishing a process, and creating workflows. Identify and record the business issues that must be resolved as well as the business drivers that will give the implementation impetus and justification to set the proper programme objectives.

Record both short- and long-term goals that support the business drivers and address the business concerns. Long-term objectives will help to build and explain the big picture, while short-term objectives should strive to validate that the programme is headed in the right direction.

### **Step 2) Prepare for Change**

Knowledge management is a cultural shift rather than only a technological one. Employees will need to reconsider how they disseminate the knowledge they acquire and have.

Companies' tendency to prioritize individual achievement is a typical barrier to greater knowledge sharing. This technique fosters a "knowledge is

power" mentality that runs counter to a culture that values knowledge sharing and knowledge creation.

Successfully implementing a new knowledge management program may require changes within the organization's norms and shared values; changes that some people might resist or even attempt to quash. To minimize the negative impact, prepare to manage cultural change.

Recruit knowledge management champions throughout the organization who will encourage knowledge sharing behaviors within their departments and provide valuable feedback to the implementation team.

### **Step 3) Define a High-Level process as a foundation**

Setting up a comprehensive knowledge management methodology is a crucial first step towards successful deployment. You can gradually construct and perfect detailed procedures via phases four, five, and six by starting with a high-level process. Remember that this conversation should include the individuals who will be the knowledge's contributors and users. Before moving on to step 7, the entire established procedure needs to be finalized and authorized (implementation).

Organizations will not achieve their knowledge management fully goals if they ignore or have a loose definition of the knowledge management process. At best, ad hoc methods will be used for knowledge identification, capture, classification, and dissemination. Knowledge strategy, creation, identification, classification, capture, validation, transfer, maintenance, archival, measurement, and reporting are examples of common knowledge management best practices to consider when creating your plan.

### **Step 4) Determine and Prioritize technological Needs**

In this step to evaluate the technologies that will improve and automate your knowledge management-related tasks. Based on the programme objectives established in step one, the process controls, and criteria you set in step three, and your programme objectives, you may identify and prioritize your needs for knowledge management technology.

The market for information management solutions is huge and diversified; it is crucial to be aware of the major vendors, comprehend the advantages and disadvantages of each technology, and choose how each solution could assist or obstruct you in achieving your goals.

Learn about the tools that employees are using today and what is and is not working for them. Don't rush into buying a modern technology before making sure that your current ones are still satisfying your needs. If there is widespread support and a need for improved computing and automation, you can also postpone making expensive technology decisions until the knowledge management programme is well under way.

You can evaluate the existing state of knowledge management within your business once you have identified your programme objectives, planned for cultural changes, defined a high-level procedure, and evaluated and prioritized your technology needs.

The five fundamental elements of knowledge management—people, procedures, technology, structure, and culture—should all be covered in the assessment. A typical assessment should give a broad overview of the present situation, the differences between it and the desired state, and suggestions for bridging those differences.

These recommendations will be the foundation for the roadmap in step no. six.

### **Step 6) Build a Knowledge Management Implementation Roadmap**

To create the knowledge management program's implementation roadmap now that you have the current-state evaluation in hand. Prior to moving forward, you should reaffirm senior leadership's commitment and support, as well as the availability of funds to launch and maintain the knowledge management programme. Your efforts are useless without these conditions. The assessment's indisputable proof of your company's inadequacies needs to increase the urgency level.

Gaining the backing of leadership and obtaining the cash you require will depend on your ability to overcome the obstacles. This technique can be described as a roadmap of connected projects, each of which fills in a particular assessment-identified need. The roadmap might show important dependencies and milestones over the course of months or years. A successful project roadmap will result in some immediate successes in the first stage, which will increase support for the next stage.

Continue to analyse and modify the roadmap over time considering the shifting business needs and economic conditions. Lessons from past projects that can be used in present and future initiatives will provide you new insights.

### **Step 7) Implementation**

Implementing a knowledge management program and maturing the overall effectiveness of your organization will require significant personnel resources and funding. Be prepared for the long haul, but make sure that you are making incremental advances and celebrate them. If there the value and benefits of the developing program are recognized, there should be little resistance to continue investing in knowledge management.

With that said, it is time for the rubber to meet the road. You know what the objectives are. You have properly mitigated cultural issues. You have got the processes and technologies that will enable and launch your knowledge management program. You know what the gaps are and have a roadmap to tell you how to address them.

## Step 8) Measure and Improve the Knowledge Management Program

How will you know if your investments in knowledge management are profitable? You will require a method for gauging your performance and comparing it to the expected outcomes. If possible, establish some baseline metrics to provide a snapshot of the organization's performance before the knowledge management programme is put in place. After deployment, trend the new findings against the previous results to determine whether performance has improved.

Create a balanced scorecard that includes measures for performance, quality, compliance, and value when choosing the right metrics to gauge your organization's success. The main goal of creating a knowledge management balanced scorecard is to gain important insight into what is and is not functioning. After that, you can take the appropriate steps to close compliance, performance, quality, and value gaps, enhancing the knowledge management program's overall effectiveness.

As you advance through each step of the roadmap, make sure you are realizing your short-term wins. Without them, your program may lose momentum and the support of key stakeholders.

---

## 10.7 SUMMARY

---

This chapter gives the details about:

**Knowledge Management:** Knowledge management is the process of creating, sharing, and managing the knowledge and information of an organization. The main aim of Knowledge Management includes transfer individual knowledge into databases. Separate and filter the most relevant knowledge, Organize the knowledge so that it will provide easy access for the user, to solve the problem efficiently and effectively. Data, Information and Knowledge differences, Knowledge Management process, Organizational Learning and Transformation, Various Knowledge Management Activities, Power of Knowledge Management, Information Technology in Knowledge Management, KMS Cycle, Various steps in Knowledge Management systems Implementation.

---

## 10.8 QUESTIONS:

---

Q1. What is meant by Data, Knowledge, and Information?

Q2 Define Knowledge Management.

Q3. What are the role of knowledge management?

Q4. Explain in detail KMS cycle.

Q5. Explain power of knowledge management in detail.

Q6. What is Organizational Learning?

Q7. What is Organizational Transformation?

---

## 10.9 REFERENCES:

---

1. Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis publisher Wiley 1<sup>st</sup> edition 2009
2. Decision support and Business Intelligence Systems by Efraim Turban, Ramesh Sharda, Dursun Delen, publisher Pearson 9<sup>th</sup> edition 2011
3. Fundamental of Business Intelligence by Grossmann W, Rinderle-Ma Publisher Springer 1<sup>st</sup> edition 2015.
4. [https://tutorialspoint.com/knowledge\\_management/models](https://tutorialspoint.com/knowledge_management/models)
5. <https://dlisnbu.ac.in/lesson/knowledge-processing-basics>
6. <https://edge.siriuscom.com/strategy/8-steps-to-implementing-a-knowledge-management-program>



## ARTIFICIAL INTELLIGENCE & EXPERT SYSTEMS

### Unit Structure

- 11.0 Objectives
- 11.1 Concepts & Definitions of AI
  - 11.1.1 Introduction of AI
  - 11.1.2 Characteristics of AI
  - 11.1.3 Applications of AI
- 11.2 AI verses Natural Intelligence
- 11.3 Concepts of Expert Systems
  - 11.3.1 Structure of Expert Systems
  - 11.3.2 Components of Expert Systems
- 11.4 Knowledge Engineering
  - 11.4.1 Knowledge Base
  - 11.4.2 Components of Knowledge Base
  - 11.4.3 Applications of Expert Systems
- 11.5 Development of Expert Systems
- 11.6 Benefits of Experts Systems
- 11.7 Summary
- 11.8 Questions
- 11.9 References

---

### 11.0 OBJECTIVES

---

After going through this unit, you will be able to:

- Concept of Artificial Intelligence
- Define Artificial Intelligence
- Understand characteristics of Artificial Intelligence
- Practical Applications of Artificial Intelligence
- Concepts of Expert Systems
- Knowledge Engineering
- Development of Expert Systems
- Various applications of Expert Systems

Artificial Intelligence is a branch of computer science, with the help of technology that mimics human intelligence to perform various tasks to solve problems.

### 11.1.1 Introduction of AI

#### Concepts

- Father of Artificial Intelligence: John McCarthy, it is “The science and engineering of making intelligent machines, especially intelligent computer programs.”
- Artificial Intelligence is a way of making computers like a controlled robot, which thinks intelligently, like a human think.
- Artificial Intelligence is the study of how human brain thinks, and how human learn, decide, and work while solving any problem, and then using the outcomes of that study on that basis the development of intelligence software and system generation takes place.
- With the help of AI machine which collects real time data and after processing of data it shows the various new patterns in it. Machine learns like a human being by observations and respond according to the experience.

For example:

Any smart watch initially we have to fed all the details like no. of steps per day, alarm for water drink, alarm for sitting at one place more time, sleeping time, dinner time and so on, Later machine will give the reminders of the similar things based on the fed data.

Artificial intelligence (AI) mimics the human intelligence by using algorithms to understand human goals or methods for achieving those goals. For achieving the objective, it develops the relationship between goal seeking, data processing and acquisition. There are 4 approaches of AI.

1. Acting humanly- Computer machine behaves exactly like a human being, it becomes difficult to distinguish between the two when using technologies like automated reasoning, machine learning, natural language processing.
2. Thinking humanly- Computer is capable of thinking exactly like a human and performing different tasks that need human intelligence, like driving a car. With the help of cognitive modelling approach is used for three different techniques namely, introspection, psychological assessment and brain imaging. Same technology is useful for psychology and healthcare to create realistic simulations when required.

3. Thinking rationally- Study of how human thinks use some principles that will help in creating a guideline of human behavior. A person is considered rational means reasonable, sensible and with good judgement). The computer thinks rationally and solve the different problems.
4. Acting Rationally- It is the study of how human behaviour, which relies on the rational agents. To optimize the expected value of its performance, actions are dependent on circumstances, environmental factors and available data. Engineering approach and black-box are typically used to get the desired output.

### **Definitions**

Artificial Intelligence is a type of technology as well as field of computer science for making a computer, a computer-controlled robot or a software which thinks intelligently like human.

OR

Artificial Intelligence is like a human intelligence which works in machines for creating different applications that understand, think, learn and behave like human being.

OR

Artificial Intelligence is a technique that enables the computers to mimic human intelligence. That machine performs on algorithms and data fed to it and gives the desired output.

#### **11.1.2 Characteristics of Artificial Intelligence**

Following are the top 3 main characteristics of framework that majorly contribute to Artificial Intelligence.

##### **1. Feature Engineering**

After collection of data, we need to process it first and then data are identified through the feature extraction process. The accuracy of the data set and features are always depending on the correctness of the data. If data is correct then the model always gives the good performance. Feature extraction is the process or technique to find the key features of the data.

**The feature extraction process includes: –**

- For minimization of entropy of the model in a system the primary classification heuristic is used. When a system of data being classified has been subdivided to the point where it can not be further subdivided, feature selection may be reused and used to another dataset, It is called as an algorithmic technique. Hence AI can maximize knowledge gain.

There are various algorithms used to select a subset of the features as per the importance in the model it is known as feature selection algorithms. The subset is selected for zero correlation between these features. By using Principal Component Analysis (PCA) the objective can be achieved.

By using feature engineering, it produces new features for the learning algorithms like supervised and unsupervised, by converting raw data and keeping the goal in mind for solving the problem with high accuracy.

## 2. Artificial Neural Networks

Artificial Neural Networks (ANN) also known as neural networks (NNs), where it is the collection of artificial neurons which are connected nodes like human brain cells. Each connection passes a signal from one neuron to the other neurons for processing the task. The output of each neuron gives a real number for a signal at a connection. The connections are known as edges. Neurons are divided into different layers for transformations with the help of algorithms. Usually, signals travel from various layers i.e., from first layer to the last layer many times.

In Neural network there are two types of layers, feedforward neural network and recurrent neural network. Feedforward neural network (FNN) is also known as acyclic where signal always travels from one direction to another. Examples are perceptron, radial basis network and multi-layer perceptron. In Recurrent neural network (RNN) has the small memories of previous input events and allows opinion as well.

There are various methods used for reducing the model size of AI for developing a neural network with good performance.

These techniques are clustered in the following 5 major categories:

1. **Pruning-** It is the method for identification as well as elimination of redundant connections in the neural network to reduce the size of the network so that it will perform well and will save the time.
2. **Quantization-** It is a method which compresses the model which will help in representation of values in bits.
3. **Low-rank factorization-** In this method, the model's tensors are decayed to create a short version being quite close to original tensors.
4. **Compact convolutional filters –** In this technique filters are used which are specially designed, that filters are useful for convolution where it reduces the number of parameters.

5. Knowledge distillation- In this method, it uses the full-size version of the model to treat it just like a small model to give the output quickly.

All the techniques or methods are independent on each other and we can use for the good performance.

Artificial Neural Network are used for solving the complex problems in real life situations with the help of finding the hidden relationship between patterns and predictions in various fields like marketing, finance, predicting rare events like fraud detection, or diagnosing harmful diseases.

For example, Alitheon proves the use of power of Artificial Neural Network in operational efficiency of commercial airlines and airports. The combination of Artificial Neural Network and Deep learning enhances the reliability of airport operations to automate redundant tasks of air-traffic control and perform all the processes well.

### **11.1.3 Applications of Artificial Intelligence**

Artificial Intelligence is the collection of concepts and ideas that are related to the development of intelligent system. These concepts and ideas may be developed in different areas and be applied to various domains.

Various domains and areas in intelligent systems can be given as follows:

Expert Systems

Natural Language Processing

Neural Networks

Robotics

### **Fuzzy Logic**

1. Expert Systems: It is an information system that is useful for human knowledge captured in a computer to solve the problems that requires human expertise and reasoning.
2. Natural Language Processing: It is a collection of technology that is useful to enable communication between computer and the user using native human language. It uses conversational type of interface for communication of human and the machine, while in traditional type of interface uses programming languages which consist of syntax and commands. NLP consist of two main subfields namely Natural language understanding and Natural language generation.
  - Natural language understanding: It is used to enable computers to understand human languages (semantics and syntax)
  - Natural language generation: It is used to enable computers to express or produce human languages.

Natural Language Processing is useful in text-mining systems, where processing of unstructured text documents processed successfully i.e., recognized, understood, and interpreted for acquisition of new knowledge.

3. Neural Network: It is also known as Neural Computing. It describes the mathematical models set that simulate the functions of a human brain.

There are many applications of Neural Networks in business like

- Language Translation: Computer programs that translates words or sentences from one language to another language automatically without any interpretation of human being.
- Game Playing: In this application of AI new strategies and heuristics are used to get good performance.

4. Robotics: It is sensory system which has vision system, and processing of signals with the help of AI technology is known as robots. Robot is an electromechanical AI based device which has programmed fed in to it for performing a specific task. It is also defined as a reprogrammable multifunctional manipulator designed to move parts, materials, tools, or specialized devices with the help of that to perform a specific task. Robot which has some kind of sensory apparatus is known as “intelligent robot”. Sensory apparatus like camera that collects the robots’ surroundings and its operations. It also allows to interpret the collected data or information and give the respond and try to adapt the changes as well. Example: humanoid HRP-2 robots (Promets), developed by the National Institute of Advanced Industrial Science & Technology.

5. Fuzzy Logic: It is a technique of processing of data with notations of logic, true/false statements. True and False value is get replaced by the degree of set of membership.

Sr.No.	Domains of AI	Example
1	<b>Expert Systems</b> <b>Examples – Flight-tracking systems, Clinical systems.</b>	
2	<b>Natural Language Processing</b> <b>Examples: Google Now feature, speech recognition, Automatic voice output.</b>	

3	<p><b>Neural Networks</b></p> <p><b>Examples –</b> Pattern recognition systems such as face recognition, character recognition, handwriting recognition.</p>	
4	<p><b>Robotics</b></p> <p><b>Examples –</b> Industrial robots for moving, spraying, painting, precision checking, drilling, cleaning, coating, carving, etc.</p>	
5	<p><b>Fuzzy Logic Systems</b></p> <p><b>Examples –</b> Consumer electronics, automobiles, etc.</p>	

**Table 11.1 Applications of AI**

## 11.2 ARTIFICIAL INTELLIGENCE VERSUS NATURAL INTELLIGENCE

- Natural Intelligence (NL) is the opposite of Artificial Intelligence. It is all the systems of control present in history.
- Natural intelligence perceive by various patterns whereas the artificial intelligence perceive by the set of rules and regulations and various ideas.
- Natural intelligence store and recall information by patterns, artificial intelligence do it by searching algorithms.
- Human Intelligence aims to adapt to new environments by combining diverse cognitive processes, Artificial Intelligence seeks to build computers that can mimic human behaviour and do human-like tasks. The human brain is analogue, whereas machines are digital.

For example, remember the number 50505050 is easy for store, recall and pattern understanding.

- Artificial intelligence can figure out the complete object even if some part of it missing or distorted, whereas the artificial intelligence cannot do it correctly.
- AI is developing with such an incredible speed, sometimes it seems magical. There is an option among researchers and developers that AI could grow so immensely strong that it would be difficult for humans to control.
- Natural intelligence created AI systems by infusing them with every type of intellect imaginable, a threat to which modern humans appear to be exposed.

Humans use the memory, reasoning, and processing capability of the brain, but AI-powered machines rely on data and specific instructions fed into the system. Human intelligence is rooted in learning from a variety of events and past experiences. It all comes down to using trial and error throughout one's life to learn from mistakes. Robots cannot reason, on the other hand, which is where artificial intelligence fails.

Human Intelligence is all about learning from various incidents and past experiences. It is about learning from mistakes made via trial-and-error approach throughout one's life. Intelligent thought and intelligent behavior lie at the core of Human Intelligence. However, Artificial Intelligence falls behind in this respect – machines cannot think.

They can learn from data and through continuous training, but they can never achieve the thought process unique to humans. While AI-powered systems can perform specific tasks quite well, it can take years for them to learn a completely different set of functions for a new application area.

Sr. No.	- Factor for Comparison	- Human Intelligence	- Artificial Intelligence
1	- Energy efficiency	- 25 watts human brain	- 2 watts for modern machine.
2	- Multitasking	- Human worker work on multiple responsibilities.	- The time needed to teach system on response is considerably high.
3	- Decision Making	- Humans can learn decision making from experienced scenarios.	- Even the most advanced robots can hardly compete in mobility with 6 years old child. And these results we have after 60

			years of research and development.
4	- Universal	- Humans usually learn how to manage hundreds of different skills during life.	- While consuming kilowatts of energy, this machine is usually designed for a few tasks,
5	- State	- Brains are Analogue	- Computers are Digital

**Table 11.2 Comparison of Human Intelligence and Artificial Intelligence**

### **11.3 CONCEPTS OF EXPERT SYSTEMS**

An expert system is a computer program that is designed to solve the complicated problems and to get decision making ability like a human expert. It performs this by extracting knowledge from its knowledge base using the reasoning and inference rules according to the user queries.

The first expert system (ES), which was the first effective use of artificial intelligence, was established in the year 1970 and is a subset of AI. By drawing on the knowledge that is kept in its knowledge base, it can solve even the most complicated problems like an expert. Like a human expert, the system aids in decision-making for complex issues by using both facts and heuristics. It is so named because it possesses in-depth knowledge of a certain field and can resolve any challenging issue in that field. These systems are created for a certain industry, like science, medical, etc.

The knowledge that an expert system has stored in its knowledge base determines how well it performs. The performance of the system increases as more knowledge is kept in the KB. The Google search box's recommendation of spelling problems is one of the typical examples of an Expert System.

#### **Characteristics of Expert Systems**

- High performance
- Understandable
- Reliable
- Highly responsive

The expert systems are capable of:

Advising

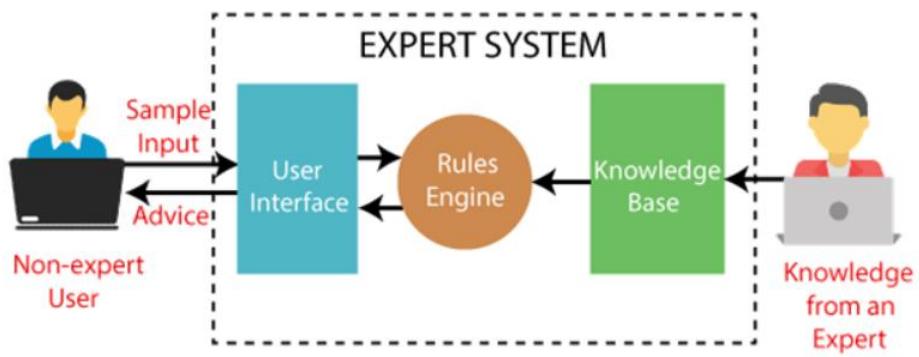
- Instructing and assisting human in decision making.
- Demonstrating
- Deriving a solution
- Diagnosing
- Explaining
- Interpreting input
- Predicting results
- Justifying the conclusion
- Suggesting alternative options to a problem

### In Capabilities of Expert Systems

- Substituting human decision makers
- Possessing human capabilities
- Producing accurate output for inadequate knowledge base
- Refining their own knowledge.

#### 11.3.1 Structure of Expert Systems

- An expert system is a set of programs that manipulate encoded knowledge to solve problems in a specialized domain that normally requires human expertise.
- An expert system's knowledge is obtained from expert sources and coded in a form suitable for the system to use in its inference or reasoning processes.
- The expert knowledge should be obtained from specialists or alternative sources of expertise, like texts, journal, articles and databases.
- This kind of knowledge typically needs a lot of training and skill in some specialized field like medicine, geology, system configuration, or engineering design.
- Once a sufficient body of expert knowledge has been acquired, it should be encoded in some kind, loaded into a base then tested, and refined regularly throughout the life of the system.



**Figure 11.1 Structure of Expert Systems**

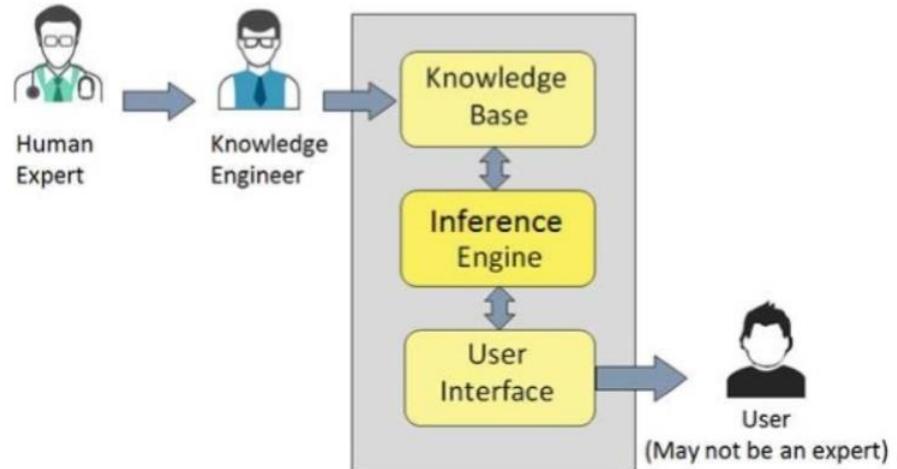
[Source: [www.javatpoint.com](http://www.javatpoint.com)]

### 11.3.2 Components of Expert Systems

The component of Expert System includes:

Expert System: A type of knowledge-based system that targets the specific knowledge of one or more domain experts. This computer system solves problems by emulating the specific processes of the expert.

- Knowledge Base: The actual knowledge stored as ontologies within a Knowledge Base Systems.
- Inference Engine: The “brain” of a KBS that uses logical assertions and conditions to solve problems and derive information
- User Interface: The front-end of a KBS where users interact with the system.



**Figure 11.2 Components of Expert Systems**

[Source: [Tutorialspoint](http://www.tutorialspoint.com)]

Knowledge is dynamic in nature. It is information in action.

### What is Knowledge?

**The knowledge base is an Expert System is a store of both, factual and heuristic knowledge.**

- Factual Knowledge – It is the information widely accepted by the Knowledge Engineers and scholars in the task domain.
- Heuristic Knowledge – It is about practice, accurate judgement, one's ability of evaluation and guessing.

### 11.4.1 Knowledge Base

It has high-caliber, domain-specific expertise.

To be intelligent, one must have knowledge. Any ES's ability to succeed mostly hinges on its ability to gather extremely precise and accurate knowledge.

### Knowledge Representation

It is the method used to organize and formalize the knowledge in the knowledge base. It is in the form of IF-THEN-ELSE rules.

### Knowledge Acquisition

- The success of any expert system majorly depends on the quality, completeness, and accuracy of the information stored in the knowledge base.
- The knowledge base is formed by readings from various experts, scholars, and the Knowledge Engineers. The knowledge engineer is a person with the qualities of empathy, quick learning, and case analyzing skills.
- He acquires information from subject expert by recording, interviewing, and observing him at work, etc. He then categorizes and organizes the information in a meaningful way, in the form of IF-THEN-ELSE rules, to be used by inference machine. The knowledge engineer also monitors the development of the Expert Systems.

### Inference Engine

Use of efficient procedures and rules by the Inference Engine is essential in deducting a correct, flawless solution.

In case of knowledge-based ES, the Inference Engine acquires and manipulates the knowledge from the knowledge base to arrive at a particular solution.

In case of rule based Expert Systems, it –

- Applies rules repeatedly to the facts, which are obtained from earlier rule application.
- Adds new knowledge into the knowledge base if required.
- Resolves rules conflict when multiple rules are applicable to a particular case.

To recommend a solution, the Inference Engine uses the following strategies –

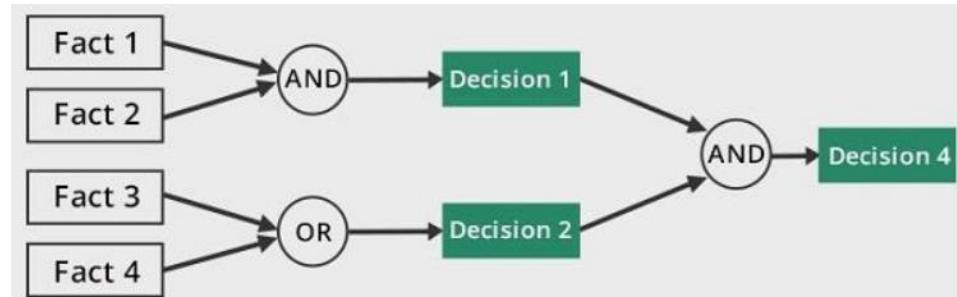
- Forward Chaining
- Backward Chaining

### Forward Chaining

It is a strategy of an expert system to answer the question, “What can happen next?”

Here, the Inference Engine follows the chain of conditions and derivations and finally deduces the outcome. It considers all the facts and rules and sorts them before concluding to a solution.

This strategy is followed for working on conclusion, result, or effect. For example, prediction of share market status as an effect of changes in interest rates.

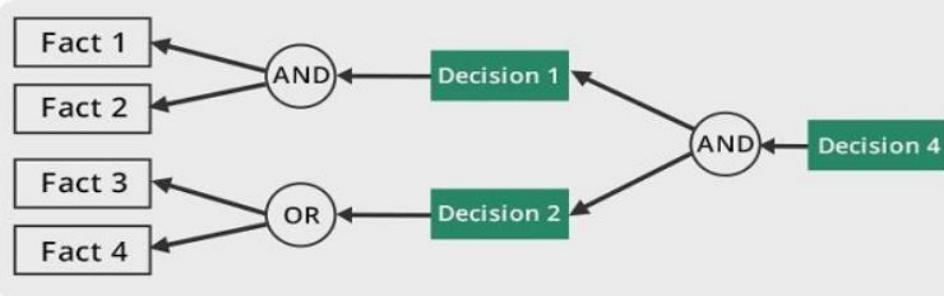


**Figure 11.3 Backward Chaining**

[Source: Tutorials Point]

With this strategy, an expert system finds out the answer to the question, “Why this happened?”

Based on what has already happened, the Inference Engine tries to find out which conditions could have happened in the past for this result. This strategy is followed for finding out cause or reason. For example, diagnosis of blood cancer in humans.



**Figure 11.4 User Interface**

[Source: Tutorial Point]

User interface provides interaction between user of the Expert Systems and the Expert System itself. It is generally Natural Language Processing to be used by the user who is well-versed in the task domain. The user of the Expert Systems need not be necessarily an expert in Artificial Intelligence.

It explains how the Expert System has arrived at a particular recommendation. The explanation may appear in the following forms –

Natural language displayed on screen.

Verbal narrations in natural language.

Listing of rule numbers displayed on the screen.

The user interface makes it easy to trace the credibility of the deductions.

### Requirements of Efficient ES User Interface

- It should help users to accomplish their goals in shortest possible way.
- It should be designed to work for user's existing or desired work practices.
- Its technology should be adaptable to user's requirements; not the other way round.
- It should make efficient use of user input.

### Expert Systems Limitations

No technology can offer easy and complete solution. Large systems are costly, require significant development time, and computer resources. Expert Systems have their limitations which include –

- Limitations of the technology
- Difficult knowledge acquisition
- Expert Systems are difficult to maintain
- High development costs

### 11.4.2 Components of Knowledge Base

#### What is Knowledge Base in an Expert System?

A type of Knowledge Base System called an expert system employs artificial intelligence to simulate human decision-making, access data from the underlying knowledge base, and retain knowledge. Early expert systems were designed to direct users toward a single, well-defined answer and did not accommodate numerous users. However, as the volume of stored data increased, expert systems expanded to support more complex knowledge types, to perform more complex problem-solving, and to support multiple users. The knowledge base in today's expert systems include data, information, and past experience.

Expert systems concentrate on the specialized, targeted knowledge of one or more domain experts and imitate their decision-making and procedures rather than collecting expertise from throughout an organization. Contrarily, general knowledge-based systems may cover a wider range of domains and be more heuristic-based.

#### There are three main components of a knowledge Based System:

**Knowledge Base:** The actual knowledge stored as ontologies in the system.

**Inference Engine:** The backend component of a KBS that applies logic rules (as assertions and conditions) to the knowledge base to derive answers from it. You can think of the inference engine as the “brain” of the KBS.

**User Interface:** The user-facing component that people interact with to find and extract knowledge stored in the system.

Regardless of the content stored, a Knowledge Base Systems should always aim to represent knowledge explicitly (as tools, data, and ontologies) rather than implicitly (computer code, vague human experience) - all for the benefit of the end user. Ultimately, however, a Knowledge Based System is still run by a computer.

### 11.4.3 Applications of Expert Systems

The application of an expert system can be found in almost all areas of business or government. They include areas such as –

- Different types of medical diagnosis like internal medicine, blood diseases and show on.
- Diagnosis of the complex electronic and electromechanical system.
- Diagnosis of a software development project.
- Planning experiment in biology, chemistry and molecular genetics.
- Forecasting crop damage.
- Diagnosis of the diesel-electric locomotive system.
- Identification of chemical compound structure.

- Scheduling of customer order, computer resources and various manufacturing task.
- Assessment of geologic structure from dip meter logs.
- Assessment of space structure through satellite and robot.
- The design of VLSI system.
- Teaching students specialize task.
- Assessment of log including civil case evaluation, product liability etc.

Application	Description
Design Domain	Automobile and Camera lens design.
Medical Domain	With the help of observed data, diagnosis system to deduce cause of disease.
Monitoring Systems	Comparing data continuously with observed system or with prescribed behavior such as leakage monitoring in long petroleum pipeline.
Process Control Systems	Controlling a physical process based on monitoring.
Knowledge Domain	Finding the faults in vehicles, computers.
Finance Or Commerce	Detection of possible fraud, suspicious transactions, stock marketing trading, cargo scheduling, Airline

**Table 11.3 Applications of Expert System**

## **11.5 DEVELOPMENT OF EXPERT SYSTEMS**

The process of ES development is iterative. Steps in developing the ES include –

### **Identify Problem Domain**

- The problem must be suitable for an expert system to solve it.
- Find the experts in task domain for the ES project.
- Establish cost-effectiveness of the system.

## Design the System

- Identify the ES Technology
- Know and establish the degree of integration with the other systems and databases.
- Realize how the concepts can represent the domain knowledge best.

## Develop the Prototype

From Knowledge Base: The knowledge engineer works to –

- Acquire domain knowledge from the expert.
- Represent it in the form of If-THEN-ELSE rules.

## Test and Refine the Prototype

- The knowledge engineer uses sample cases to test the prototype for any deficiencies in performance.
- End users test the prototypes of the ES.

## Develop and Complete the ES

- Test and ensure the interaction of the ES with all elements of its environment, including end users, databases, and other information systems.
- Document the ES project well.
- Train the user to use ES.

## Maintain the System

- Keep the knowledge base up to date by regular review and update.
- Cater for new interfaces with other information systems, as those systems evolve.

## 11.6 BENEFITS OF EXPERT SYSTEMS

- Availability – They are easily available due to mass production of software.
- Less Production Cost – Production cost is reasonable. This makes them affordable.
- Speed – They offer great speed. They reduce the amount of work an individual puts in.
- Less Error Rate – Error rate is low as compared to human errors.
- Reducing Risk – They can work in the environment dangerous to humans.
- Steady response – They work steadily without getting motionless, tensed or fatigued.

This chapter gives the details study of Artificial Intelligence, characteristics of Artificial Intelligence and applications of Artificial Intelligence. Explanation of difference between Artificial Intelligence verses Natural Intelligence. The study of Expert System along with different components, structure of Expert Systems and benefits of Expert Systems.

---

## 11.8 QUESTIONS:

---

- Q1. What is Artificial Intelligence?
- Q2. Give the difference between human intelligence and artificial intelligence.
- Q3. Explain basic concepts of expert systems.
- Q4. Explain characteristics of expert systems.
- Q5. Explain forward chaining and backward chaining.
- Q6. Explain components of expert system.
- Q7. Explain structure of expert system
- Q8. Explain applications of expert system.
- Q9. Explain benefits of expert system.

---

## 11.9 REFERENCES:

---

1. Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis publisher Wiley 1<sup>st</sup> edition 2009
2. Decision support and Business Intelligence Systems by Efraim Turban, Ramesh Sharda, Dursun Delen, publisher Pearson 9<sup>th</sup> edition 2011
3. Fundamental of Business Intelligence by Grossmann W, Rinderle-Ma Publisher Springer 1<sup>st</sup> edition 2015.
4. <https://www.javatpoint.com/expert-systems-in-artificial-intelligence>
5. <https://www.tutorialspoint.com>

