Advanced Analytics e Artificial Intelligence per il marketing

Chapter 1

Enhancement of marketing activities

Through a data-driven approach

Pearson

Evolution of the marketing concept

1 PRODUCTION ORIENTATION

The market is characterized by a scarcity of goods (demand > supply).

The strategy used by companies consists in production in a rapid, effective and efficient way, which favors the entry of new companies, decreasing the gap between supply and demand

1930

2 PRODUCT ORINETATION

The market is characterized by a smaller gap between supply and demand.

The strategy used by companies consists in creating a product or service that is better or different from the competitors.

However, being the inexperienced consumer, the company strategy focuses on the best way to sell the product compared to competitors

1950

3 SALES ORIENTATION

The market is focused on selling and is uninterested in understanding consumer needs.

The strategy used by companies consists in combining the study of production and sales processes with the study and segmentation of the market and with the identification of a specific target

1960

4 MARKETING ORIENTATION

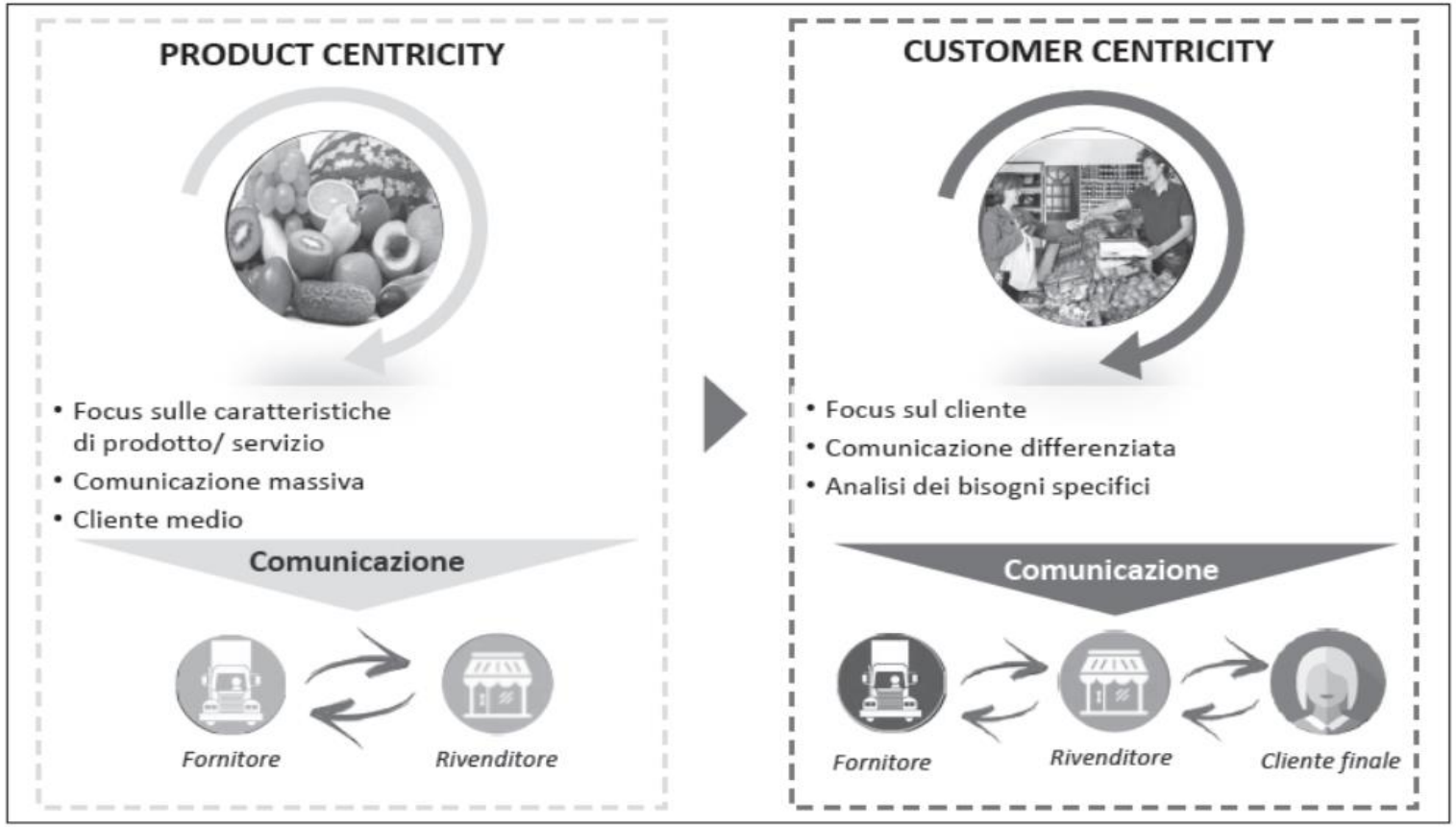
The market is characterized by saturated demand.

The strategy used by companies consists in creating tailor-made solutions for consumers with the aim of gaining trust and loyalty over time.

The analysis of the needs and requirements of customers thus becomes fundamental, with constant control of production and distribution costs

Evolution of the marketing concept

We highlight the evolution of the product-oriented vision - product centricity - with the involvement of only the supplier and reseller players in a customer-oriented vision - customer centric - which places the accent on his needs, his expectations and the mechanisms through which the company can establish solid relationships capable of adding value in the medium to long term.



How corporate marketing works: 3 interconnected areas

Process through which the most suitable strategy is determined to achieve the medium-long term corporate objectives.

Tools related to strategic marjeting: the systematic and continuous analysis of market needs for the development of increasingly competitive and innovative new products

Process through which short-term actions and plans are identified to achieve the set objectives

Use of techniques and tools for the collection and analysis of relevant information on the company's business, in order to provide support in defining the strategy

Strategic marketing

Strategic marketing translates into the need to carry out an analysis of the market and the competitive context, elements strongly correlated to the economic cycle and the socio-economic context in which the company is located.

Below are the most suitable tools for strategic analysis:

IL FRAMEWORK SWOT	5 FORCES OF PORTER	CUSTOMERS STUDY	PLACEMENT	KPI
Identification of four key components: - STRENGTH: strengths that distinguish the company and/or product/service on the market; - WEAKNESS: weaknesses of the company and/or of the product/service offered; - OPPORTUNITY: opportunities present on the market - THREAT: threats inherent in the corporate competitive context	Porter's five competitive forces seek to understand the company's competitive position through: - the threat of new entries - the threat of substitute products/services - the bargaining power of consumers - the bargaining power of suppliers - competitive rivalry	The study of current and potential customers (prospects) is aimed at identifying the reference target of the product/service and further segmenting it into profiled sub-targets (buying personas) characterized by different needs and interests	The study and analysis of the positioning of the brand in relation to the needs, values and expectations of the end customer, distinguishing itself from competitors	KPIs (Key Performance Indicators) are business indicators useful for assessing the consistency, efficiency and effectiveness between the strategic plan and the operational plan. The KPIs are to be monitored over time to implement any corrective measures to the strategy

The role of technology in the evolution of the marketing concept

The evolution of the concept of marketing also brings with it an evolution in the ways and means of communication of companies.

MARKETING 1.0

Companies focus on production and sales.
It uses one-way communication channels, from the company to the consumer, with the aim of communicating an undifferentiated message to a large audience of consumers with minimal financial expenditure.
Main means of communication: press, radio and TV.

MARKETING 2.0

In line with an ever growing interest in knowing the final consumer, communication becomes two-way with a one-to-one logic, supported by the diffusion of the mobile phone and the Internet.
Main means of communication: digital tools, telephone surveys, market research and data analysis.

MARKETING 3.0

Companies are focused not only on the consumer but also on their ability to contribute to creating a better world.
Corporate marketing focuses on three values (ESG): Environmental - impact on the environment and on the territory -, Social - initiatives with a social impact -, Governance - internal aspects of the company and its administration.

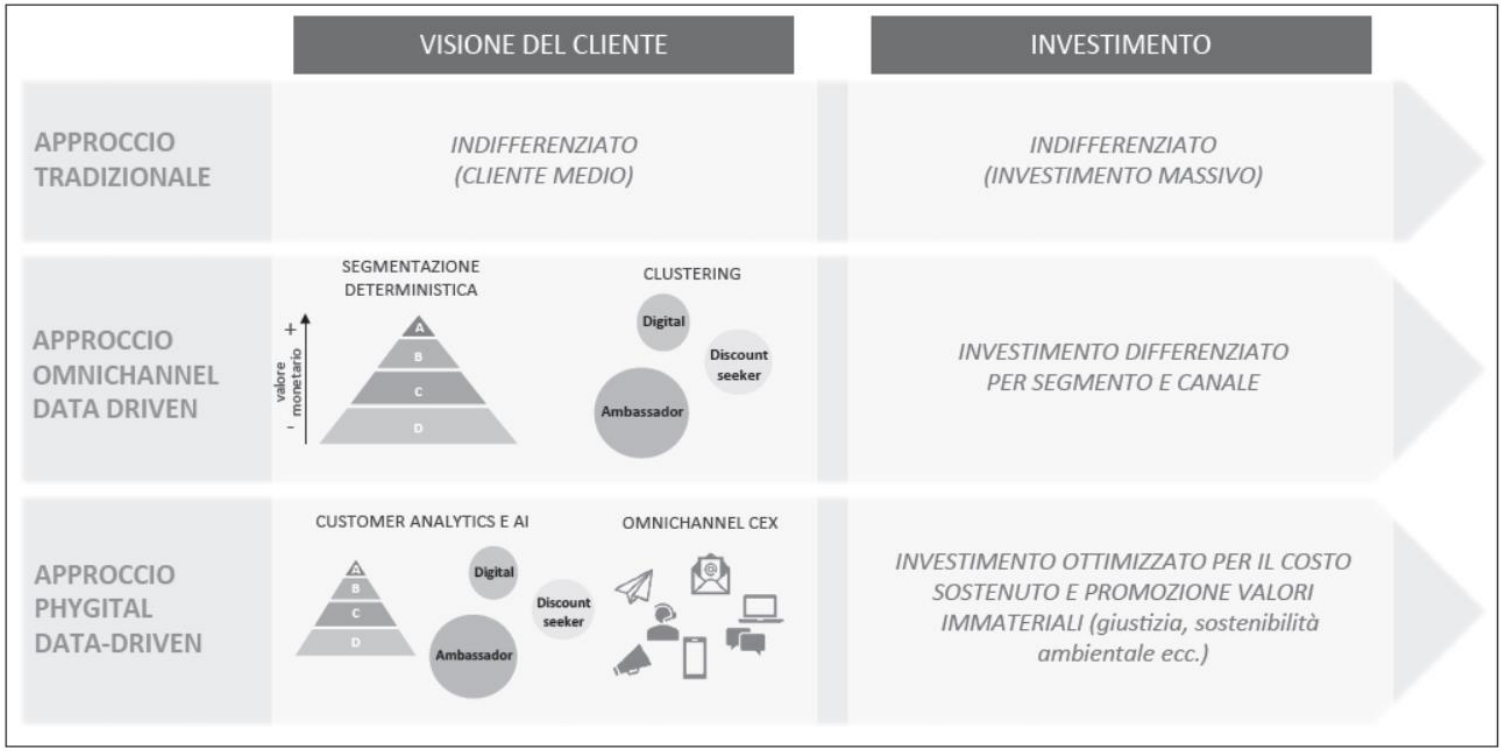
MARKETING 4.0

Better known as omnichannel marketing, supported by the spread of cloud computing and big data, digital transformation, the presence of generations X and Z and the interconnection of devices.
This approach transforms the customer jurney from the 4A model - Aware, Attitude, Act, Act again - to the 5A model - Aware, Appeal, Ask, Act, Advocate -

The role of technology in the evolution of the marketing concept

MARKETING 5.0

It follows the latest developments in technological progress, changes in consumer behavior and those of traditional business models that now merge physical and digital reality (phygital).
Thus "technology for humanity" develops, a technological development capable of satisfying the material needs of consumers, promoting the immaterial ones.
New tactics in use: data-driven marketing, predictive marketing, contextual marketing and augmented marketing



Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 2

TECHNOLOGY AS A MARKETING ENABLER FOR DATA ANALYSIS

Pearson

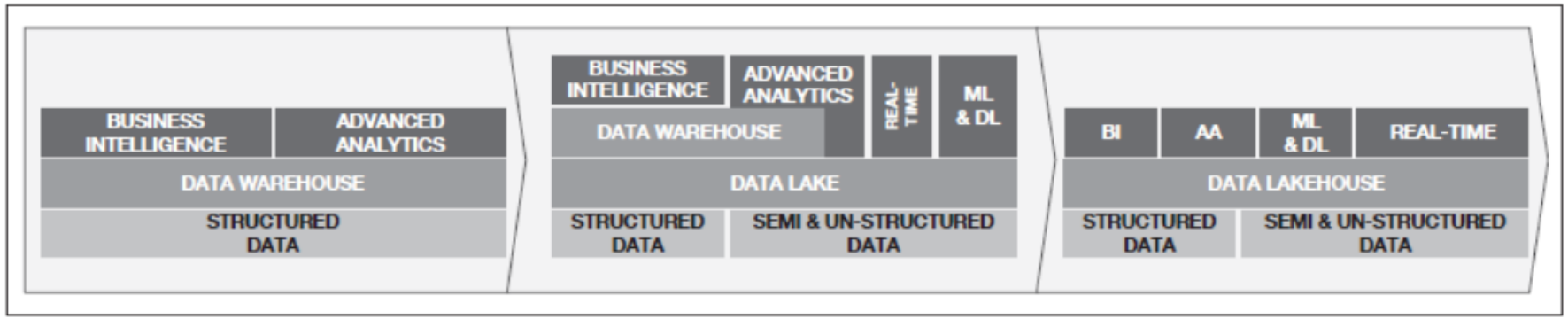
Intro

The main technologies used for the management, acquisition and exploitation of data in the company have been introduced starting from the simplest systems of MIS, DSS, the first client-server applications, up to the development of the first RDBMS in which the query language was introduced SQL data.

The exponential growth of the amount of data has led companies to develop new systems to standardize the data present in a fragmented and decentralized way within different vertical and centralized databases: this is how DWHs are born, which offer a single point of access and on a single application.

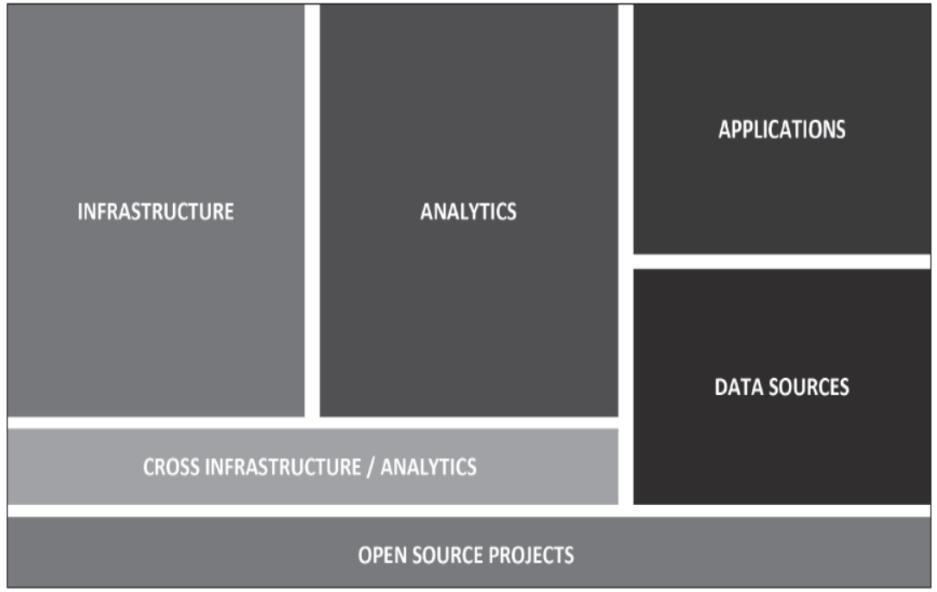
To cope with the problems of managing large amounts of data and their internal sharing and the growth of the presence of unstructured data, the MapReduce data processing framework, the HDFS file system and a YARN resource manager are introduced.
This is how datalakes are born, capable of storing structured and unstructured data, enabling complex and real-time use-cases.

The combination of data lake and data warehouse functions for Data visualization and BI activities has allowed the development of data lakehouses: a solution capable of providing the benefits of a DWH using a low-cost storage typical of data lakes, capable of support data analysis and the application of the main Machine Learning models, guaranteeing performance and reliability of a DWH, with the use of Data Visualization and BI tools in a single open-source environment.



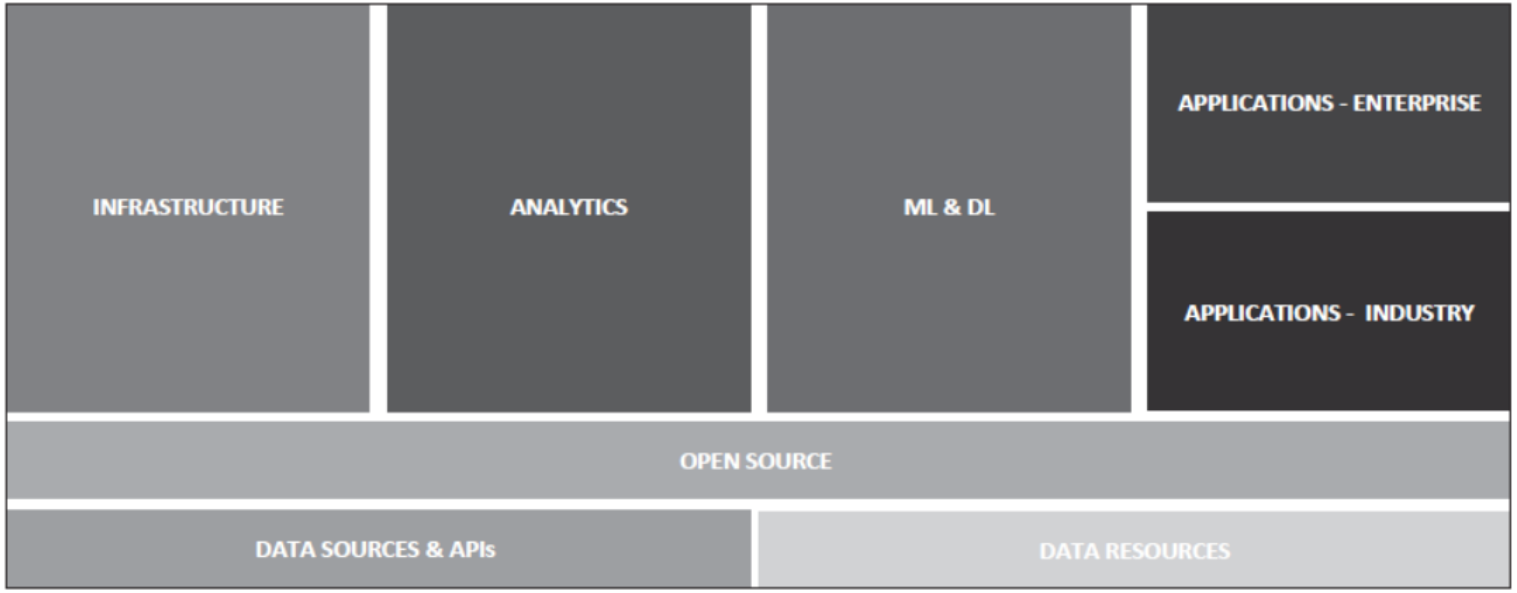
Data, Advanced Analytics e AI Landscape

Dal «Big Data Landscape» al «Machine Learning, AI and Data (MAD) Landscape»



Main components of the Big Data ecosystem

Main components of the Machine Learning, Deep Learning & Data ecosystem



Data, Advanced Analytics e AI Landscape

DATA SOURCES & APIs	External data sources divided into: Data Marketplace & Discovery, Financial & Economic Data, Air/space7Sea, People/Entites, Local Intelligence, Other. Phenomena such as data democtatization and the sharing economy have made the public use of data available.
DATA RESOURCES	- Data Services (Kaggle): provides public datasets, notebooks with analysis and models, customizable and free Jupyter Notebook environment; - Incubators&School (DataCamp): online platforms that provide free resources, courses, certifications on Data Science skills; - Research (Open AI, Google/Facebook Research): research and article publishing community
OPEN SOURCE	It includes: Frameworks, Format, Query/Data Flow, Data Access, Databases, Orchestration, Infrastructure, Data Ops, Streaming&Messaging, Stat Tools&Languages, ML Ops&Infra, AI/ML/DL, Search, Logging&Monitoring, Visualization, Collaboration, Security. The choice depends on: amount of data, type of data, access mode and frequency. The categories of Orchestration, Infrastructure, Data Ops, ML, Ops&Infra, Security allow you to coordinate pipelines, manage and track data transformation, guarantee data quality and delivery.
INFRASTRUCTURE	Includes: Storage, Hadoop, Data Lakes, DWH, Streaming/IN-Memory, RDBMS, NoSQL Database, NewSQL, Databases, Real-Time Databeses, Graph DBs, MPP DBs, ELT/ETL Data Transformation, Reverse ETL, Data Integration, Data Governance&Access, Privacy&Security, Data Observability, Data Quality, MGMT/Monitoring, Serverless, Custer SVCS. Architecture adoption can be on-cloud, on-promises or hybrid. Possible solutions for data storage and analysis: Azure Storage, Amazon S3 and Google Cloud Storage, Oracle, IMB DB2 or Microsoft SQL Server, MPP. Possible tools for system data acquisition and subsequent processing: Data Access, Data Integration, ETL, ELT, Data Transformation.

Data, Advanced Analytics e AI Landscape

ANLYTICS	It includes: BI Platform, Visualization, Data Analytics Platforms, Augmented Analytics, Data Catalog and Discovery, Metrics Store, Log Analytics, Query Engine, Search for the enhancement of data and therefore the transformation from raw data to business insights. Data Catalog and Discovery tools make it easier to search within datasets. Query Engine and Data Analytics are tools for building pipelines. BI and Data Visualization platforms allow you to analyze data through metrics, trends and insights.
MACHINE LEARNING & ARTIFICIAL INTELLIGENCE	Includes: Data Science Notebooks, Data Science Platforms, ML Platforms, Data Generation&Lebelling, Model Building, Feature Store, Deployment&Production, Model Monitoring&Observability, Computer Vision, Speech, NLP, Synthetic Media, Horizontal AI, GPU DBS & Cloud, AI Hardware.
APPLICATIONS – ENTERPRISE	Includes: Sales, Marketing-B2B, Marketing B2C, Customer Experience/Service, Human Capital, Legal, Partnerships, Regtech&Compliance, Finance, Automation&RPA, Security. Useful for managing a specific business area.
APPLICATIONS – INDUSTRY	Includes: Advertising, Education, Real Estate, Government&Intelligence, Commerce, Finance-Investing, Finance-Lending, Insurance, Healthcare, Life Sciences, Transportation, Agriculture, Industrial, Other.

Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 3

CLOUD ARCHITECTURES AND TECHNOLOGICAL SOLUTIONS FOR DIGITAL MARKETING

Pearson

Intro

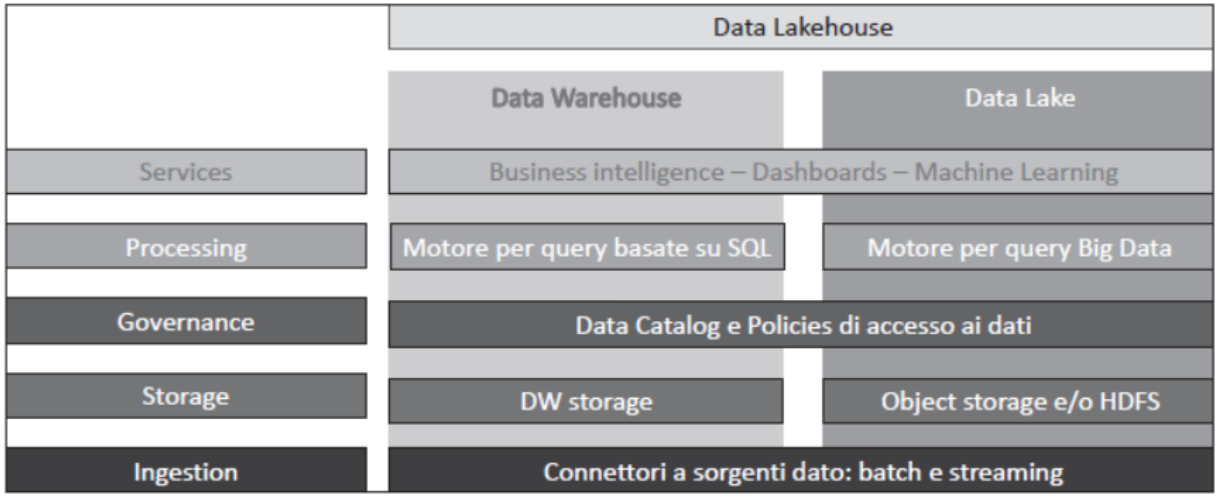
Cloud computing is defined as the possibility of offering computational services through the virtualization of hardware resources, thus creating the offer of a transversal service, transforming it into a utility.

There are four main providers: Google Cloud Platform, Amazon Web Services, Microsoft Azure, Databricks (the latter deals with cross-cloud PaaS services)

Comparison of data analysis platforms

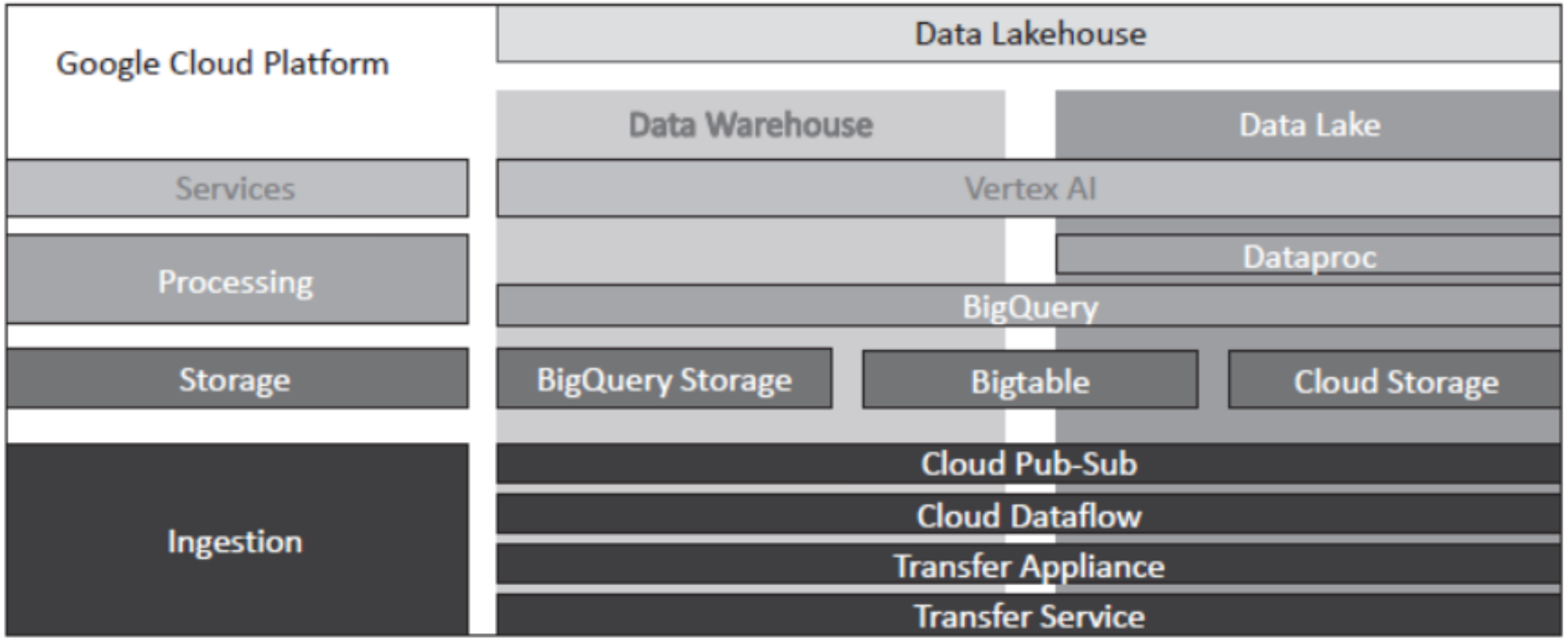
Caratteristiche	Database	Data Warehouse	Data Lake	Data Lakehouse
Struttura transazioni	OLTP (dati operazionali)	OLAP (dati storicizzati)	OLAP (dati storicizzati)	OLAP (dati storicizzati)
Formato dati	• Strutturato • Semi-strutturato	• Strutturato • Semi-strutturato	• Strutturato • Semi-strutturato • Non strutturato	• Strutturato • Semi-strutturato • Non strutturato
Consistenza	• Definizione Schema in scrittura • ACID • Formato basato su righe	• Definizione Schema in scrittura • ACID • Formato basato su colonne	• Definizione Schema in lettura • Dati distribuiti	• Definizione Schema in lettura • ACID • Dati distribuiti
Linguaggi per query	• SQL	• SQL	• SQL • JAVA • Python • Scala • R	• SQL • JAVA • Python • Scala • R
Layer applicativo supportato	• Applicazioni custom	• Business intelligence	• Machine learning	• Business intelligence • Machine learning
Figura professionale	• Sviluppatore full-stack	• Data analyst	• Data scientist • Data engineer • Sviluppatore full-stack	• Data analyst • Data scientist • Data engineer • Sviluppatore full-stack

Data analysis solutions with lambda architecture for streaming data processing:



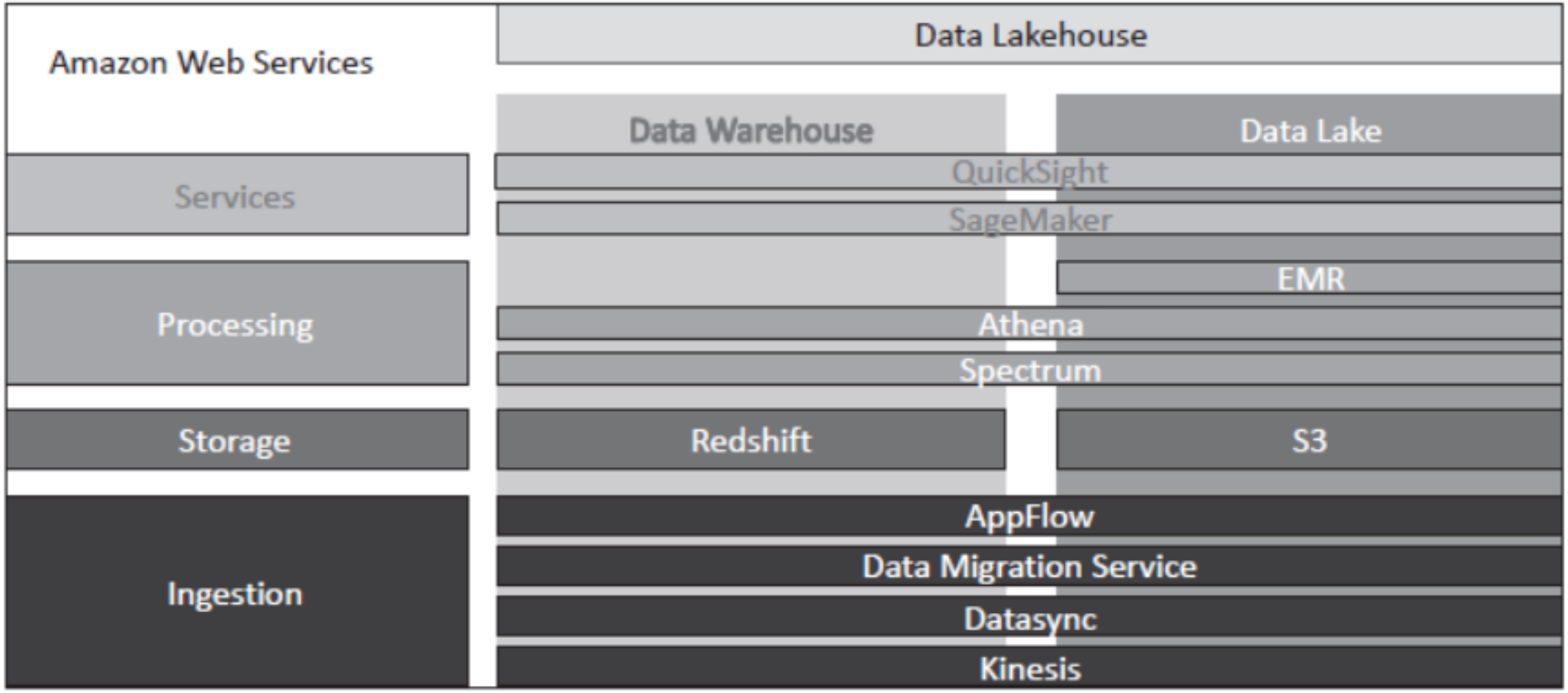
Google Cloud Platform

Data analysis solution proposed by Google Cloud Platform, schematized in a layered diagram:



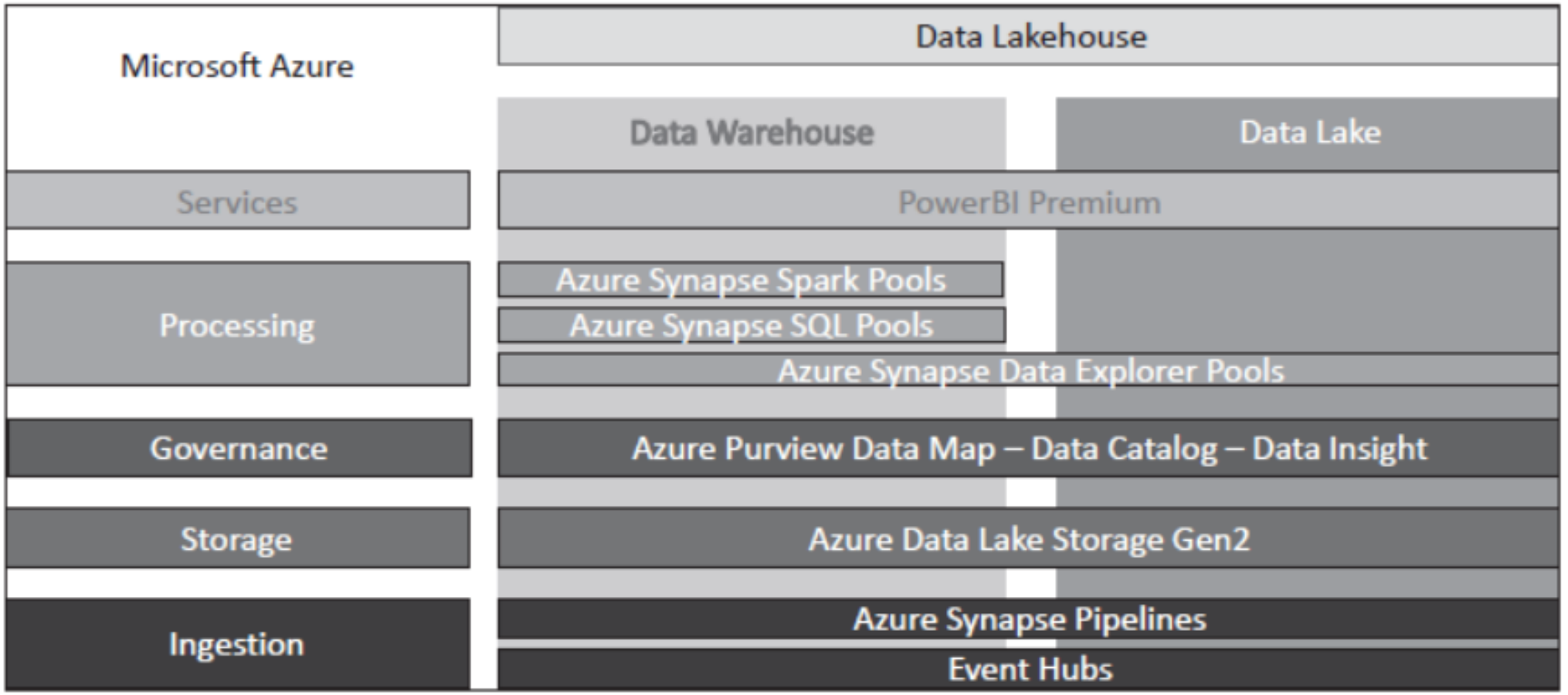
Amazon Web Services (AWS)

Data analysis solution proposed by AWS, schematized in a layer diagram:



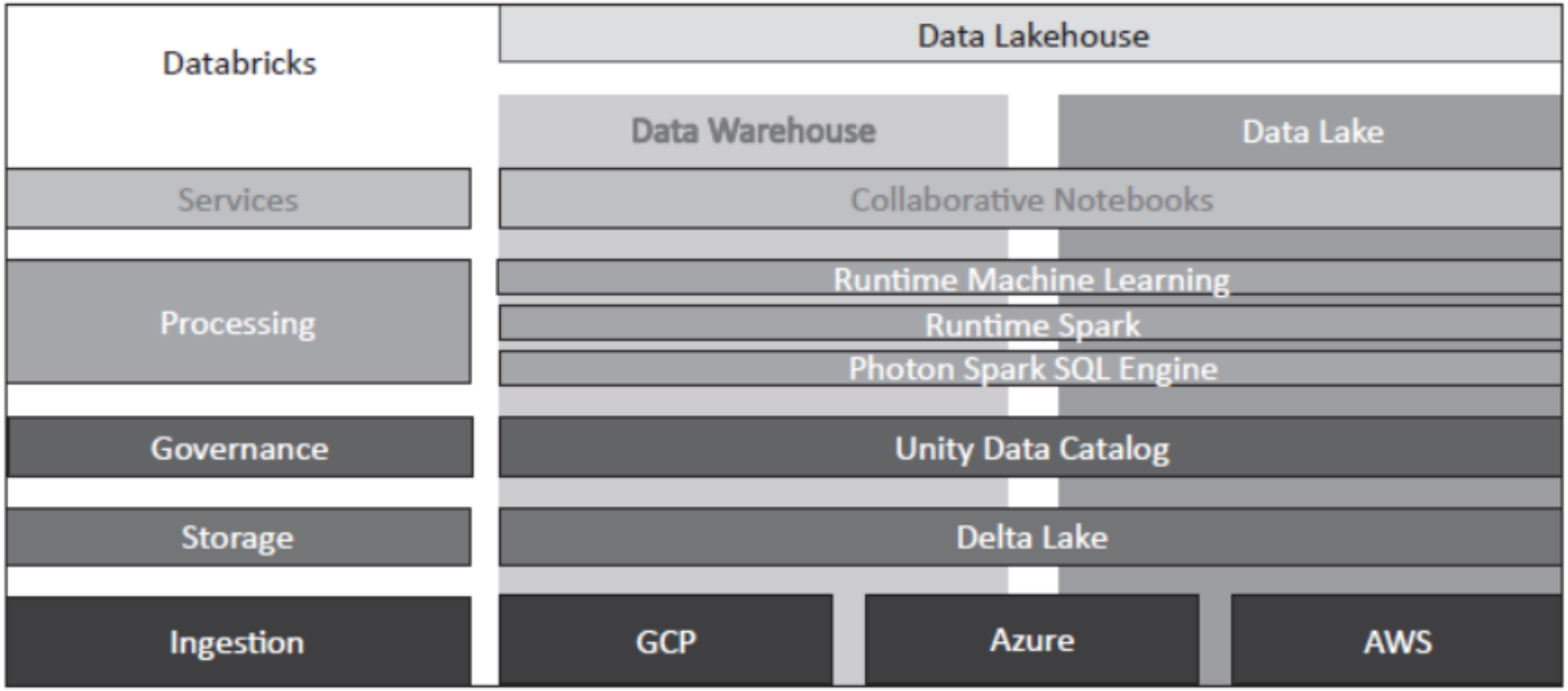
Microsoft Azure

Data analysis solution proposed by Microsoft Azure, schematized in a layer diagram:



Databricks

Data analysis solution proposed by Databricks, schematized in a layer diagram:



Advanced Analytics e Artificial Intelligence per il marketing

Chapter 4

Microservices in the cloud computing era for digital marketing

Pearson

General overview of microservices

Microservices indicate a programming paradigm in which the application is obtained as a composition of smaller and mutually independent services that communicate through a messaging system, often an API of HTTP resources.

The microservices architecture is built as a suite of services, ensuring dynamism, flexibility and scalability.

One of the strengths is the free choice of development toolchains: they can be developed using different frameworks, libraries and languages.
The problem of putting a single service into production and isolating it is obviated by the use of virtualization techniques:

VIRTUAL MACHINES

Computing resource that uses a software layer to run applications. One or more guest VMs can share the same physical server, called a host. A hypervisor takes care of extracting the interface to the host system's physical resources by providing guest systems with a virtual environment in which to operate.
The image of a VM is a single file in binary format containing the virtual hard disk which contains the operating system, installed applications and configuration files. The image is the fundamental unit through which an application can be put into production.

CONTAINER

A collection of one or more processes that maintain a separate execution context from the rest of the system. Multiple containers share the same host operating system. All the files and dependencies necessary to run an application are provided by an image built incrementally, adding the necessary dependencies to base images available in public repositories with a process that is automated.
The virtualization layer is provided by a container engine that takes care of reading and executing images. Among the most used engines we find the OCI: an image that adheres to the OCI Imagine Specification can be read and executed by any runtime that adheres to the OCI Runtime Specification. The most popular engine is Docker.

SERVERLESS CALCULATION

It represents a further abstraction step for virtualization techniques. A serverless technology supports runtimes such as Phyton, Java and Node.js. The developer is responsible for compacting the application in a ZIP package compatible with the runtime made available. The application can be executed on-demand in response to events that trigger its execution. The entire infrastructural management is entrusted to the serverless provider.

Orchestration

KUBERNETES	Open-source container orchestration platform. Tool for automating software development and management. Implement master-slave architecture by controlling compute and storage resources as objects while its components are managed through an API. The Control Plane has the role of master and consists of: etcd, scheduler, API server, controller manager. The "pod" + the elementary unit of computing resources and indicates a group of containers that share computing and network resources, as well as execution specifications. The containers of a pod are followed in the same context but the pod can be replicated and moved to other nodes. Each pod has an IP address.
APACHE MESOS	Implement master-slave architecture and use the Linux cgroups feature to isolate resources in cluster management. It consists of frameworks (e.g. Apache Aurora) that deal with task execution (as executor) and workload distribution (as scheduler). Apache Marathon takes care of the orchestration of the Mesos and Docker containers.
APACHE SERVICE MIX	Open-source project for implementing a distributed ESB. It is based on a SOA architecture model and includes Apache Active MQ, Camel Pache, Apache CXF and Apache Karaf services.
OPENSHIFT	Set of software containerization products. It is a PaaS platform and includes Docker images together with the Kubernetes platform and all its features. It is based on a Linux platform and is designed to support scaling of cloud-based applications.
KAAS	Cloud solutions that manage the orchestration through Kubernetes and take care of managing development, maintenance, hosting, updating, monitoring, automation support, load balancing and scalability. Main solutions: GCP KaaS, Amazon ECS, Amazon EKS, Azure AKS.

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 5

Introduction to the language of Python programming

Pearson

Intro

Python is an open source high-level object-oriented programming language.

Provides a program that can run other programs directly from their source code, without prior compilation.

Python programs are runnable out of the box.

Python is a general purpose language, which can be used as a scripting language, intended as a tool for coding system functionality as an alternative to shell scripts.

The Python language is highly portable, in the sense that most programs can be run on different platforms without making any changes to the code, and it is integrated with other programming languages such as C or C++.

The Python installation package (3.x versions) can be downloaded from the link: https://www.python.org/downloads/

Anaconda is a free distribution of Python for scientific computing and data analysis and is intended to facilitate the management of modules and additional libraries via the "conda" package manager.

A lighter version without a GUI is Miniconda.

The installation procedure is described at the link: https://docs.anaconda.com/anaconda/install/index.html

Python Interpreter

The Python interpreter is accessed via the interactive prompt or command line.

Just type python in a terminal and press Enter

```
$ python
Python 3.9.7 (default, Sep 10 2021, 14:59:43)
[GCC 11.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

The Python prompt is a REPL (read-evaluate-print-loop).

The user waits for the command to be executed and then displays the output which is printed on the screen.

To quit the Python interpreter, use the quit() command.

Python is a case sensitive language.

If you write Python code on text files: the command-line Python interpreter can be called on files with the .py extension.

Installing external modules

Modules or packages in the Python Package Index (PyPI) can be installed via pip, the package manager.

The modules are installed from the command line:

```
$ pip install numpy
```

Numpy is the main numerical computation library in Python.

If you are using the Anaconda or Miniconda package manager, the default apckage manager is conda, and the syntax for installing packages is similar to that of pip:

```
$ conda install numpy
```

The above installation is superfluous with Anaconda: the numpy library is installed by default.

Notebook

Ipython provides a REPL that backends Jupyter notebooks accessible via web browsers.
Ipython and Jupyter are installed together with Anaconda, or by following the instructions from the link: https://jupyter.org/install

The Jupyter interface is accessible via Anaconda Navigator or by typing Jupyter notebook from the command line:

```
$ jupyter notebook
[I 08:58:24.417 NotebookApp] Serving notebooks from local directory: /Users/
    catherine
[I 08:58:24.417 NotebookApp] 0 active kernels
[I 08:58:24.417 NotebookApp] The Jupyter Notebook is running at: http://
    localhost:8888/
[I 08:58:24.417 NotebookApp] Use Control-C to stop this server and shut down
    all kernels (twice to skip confirmation).
```

Jupyter interface

Jupyter editor view

Python: basic elements - Variables

Python prompts (or notebook cells) can be used as a calculator:

```
>>> 4*2
8
```

You can assign the result of an operation to a variable using the = symbol. Once the value has been assigned, it is possible to do any type of calculation:

```
>>> x = 4*2
>>> x
8
```

```
>>> x*3
24
```

The most correct way to print the value of an expression is via the built-in print() function:

```
>>> print(x)
8
```

The value of the variable « _ » is the output of the last operation performed, which coincides with the last value printed on the screen:

```
>>> print(_)
24
```

Python is a case sensitive language. Ex: if you call the Print function, an error is returned

```
>>> Print(_)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'Print' is not defined
```

Python: basic elements – Data type

Python is not a typed language. You can check the data type of a variable via the type() built-in function:

```
>>> type(x)
<class 'int'>
```

The variable x can be denoted as:
- Int : integer number
- Float : real number (floating point number)
- False or True: Boolean number with logical value 0 or 1

```
>>> type(x)
<class 'int'>
```

```
>>> x = 3.14
>>> type(x)
<class 'float'>
```

```
>>>print(3+2j, type(3+j2))
3+2j <class 'complex'>
>>> print(True, type(True))
True <class 'bool'>
>>> print(False, type(False))
False <class 'bool'>
```

Python:basic elements – Strings

Other built-in data types are strings: sequences of ASCII characters enclosed in single or double quotes. If double quotes are used, the single quote can be used as an element in the character sequence:

```
>>> string = "questa e' una stringa"
>>> string
"questa e' una stringa"
>>> type(string)
<class 'str'>
```

Strings are ordered sequences and therefore the characters within them can be indexed by integers. In Python, indices start at 0:

```
>>> string[0]
q
```

We can obtain a substring of consecutive elements using the slicing operator, which is used with the following syntax start : end , (start and end indicate the beginning and end of the sequence of characters to be extracted):

```
>>> string[0:6]
'questa'
```

The slicing operator start:end:step is similar and is used to extract a subset of non-consecutive elements: step indicates the difference between the indices corresponding to the characters to be extracted:

```
>>> string[0:6:2]
'qet'
```

Python: basic elements – Composite data types

Other types of data that fall within ordered sequences are lists and tuples, sequences of items separated by commas and delimited respectively by square brackets and round brackets:

```
>>> lista = [0,'bye', 3.14]
>>> type(lista)
<class 'list'>
```

```
>>> tupla = (0,'bye', 3.14)
>>> type(tupla)
<class 'tuple'>
```

Lists are a mutable data type, while tuples are an immutable data type. It is possible to modify only the values of a list through the assignment operator, but not those of the tuples:

```
>>> lista[1] = 7
>>> lista
[0, 7, 3.14]
>>> tupla[1] = 7
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: 'tuple' object does not support item assignment
```

Other data types are dictionaries: unordered sequences whose values are key-value pairs. Sequence of items separated by a comma and delimited by curly braces. Key-value pairs are separated by colons, following the key:value syntax.

The values can be accessed using the corresponding key in square brackets.

Dictionaries are mutable: you can change or add new items using the = operator:

```
release = {'Tenet': 2020,
           "Joker": 2019,
           "Aladdin": 2019,
           "The Hateful Eight": 2015}
```

→

```
>>> release['Joker']
2019
```

→

```
>>> release['Scarface'] = 1983
>>> release
{'Tenet': 2020, 'Joker': 2019, 'Aladdin': 2019, 'The Hateful Eight
    Scarface': 1983}
```

Python: instructions - Branching

The branching operator begins with the if keyword followed by a condition that must have Boolean values True or False and a colon; follows the block of code to be executed (if the condition is verified).

You can specify other conditions and other operations to be performed using the elif keyword.

If none of the conditions are true, the operation to be performed must be specified with else.

Else and elif have the same indentation as if. The branching closes by returning to the indentation level of if, without specifying any other keywords.

```
if condizione :
    <codice1>
elif condizione2: # opzionale
    <codice2>
else : # opzionale
    <codice3>
```

In the following example, branching prints two different strings depending on the value of x:

```
>>> x = 4
>>> if (x % 2 == 0):
        print("x e' pari")
    else:
        print("x e' dispari")
x e' pari
```

Python: instructions – For loop

The for loop is used to perform the same operation a fixed number of times, possibly on different inputs.
The for loop begins with the for keyword, followed by the name of a variable x that takes values from the specified list, and a colon.
The statements to be executed are at a higher indentation level and are completed when returning to the previous indentation level.

```
for x in lista:
    <codice da eseguire>
    ...
```

The following example prints a string.
The list is specified via the built-in function range(), which takes a positive integer as input and returns an iterator that can be thought of as the list of integers from 0 to the specified number.
Notice how the value of the iteration variable i changes with each iteration:

```
>>> for i in range(4): # i <- [0,1,2,3]
        print("Hello world iterazione",i)
Hello world iterazione 0
Hello world iterazione 1
Hello world iterazione 2
Hello world iterazione 3
```

Python: instructions – While loop

The while loop is used to execute a block of code an unknown number of times.
The loop begins with the while keyword followed by a condition and a colon.
The block of code is at a higher indentation level and the loop ends once it returns to the original indentation level.

```
while condizione:
    <codice da eseguire>
    ...
```

Take as an example a list of numbers of unknown length and content, and suppose you want to print their contents until an element greater than 7 is encountered, if any:

```
>>> numeri = [-7, 5, -10, 4, -1, 0, 3, 8, -9, -3, 7, 2]
>>> i=0
>>> while (i >= len(numeri)) or (numeri[i]<=7):
        print(numeri[i])
        i = i+1
-7
5
-10
4
-1
0
3
```

Python: functions

In Python, the syntax for defining a function is as follows:

```
def none_funzione(input1, ...): # gli argomenti in input sono opzionali
    ''' descrizione funzione
    inputs: ...
    outputs: ...
    ''' # optional
    <function body>
    ...
    <end function body>
    return output1, ... # gli argomenti in output sono opzionali
```

The definition of a function begins with the keyword def followed by the name of the function and a pair of parentheses holding the input arguments and a colon.

The rest of the function consists of the description enclosed in triple quotes and the body of the function, which contains the instructions that define the functionality. The function ends with the return keyword followed by the output arguments (if any).

Note that the content of the function must be at a higher level of indentation than the definition of the name.

The function concludes by returning to the indentation level of the def keyword.

```
def say_hello():
    print("Ciao!")
```

→

```
>>> say_hello():
Ciao!
```

```
def say_hello(saluto, nome = ''):
    if nome:
        stringa = saluto + ' ' + nome +'!'
    else:
        stringa = saluto + '!'
    print(stringa)
    return stringa
```

→

```
>>> stringa1 = say_hello('Ciao')
Ciao!
>>> stringa2 = say_hello('Ciao', nome = 'Federica')
Ciao Federica!
>>> print(stringa1, stringa2)
Ciao! Ciao Federica!
```

→

```
>>> stringa
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'stringa' is not defined
```

Python: Classes and objects

Classes and objects have the dual purpose of promoting the reuse of portions of code and making them easier to read and understand for other users.
Classes can be thought of as abstractions of entities and objects as members of a certain class.
Each object has variables, whose value varies from object to object, and methods, functions that allow you to modify or interact with the object.
The class definition begins with the keyword class, followed by the name of the class and a pair of parentheses containing the keyword object followed by a colon.
The class definition breaks down into the definition of various modules and is at a higher level of indentation than the name definition. Modules follow the syntax of functions.

```
class Nome_classe(object):
    def __init__(inputs...):
        <body module1>
    def modulo1(inputs...):
        <body module1>
    def modulo2(inputs...):
        <body module2>
```

```
class Persona(object):
    def __init__(self, nome, cognome, anni):
        self.nome = nome
        self.cognome = cognome
        self.anni = anni
    def stampa_nome(self):
        print('Nome:',self.nome,'Cognome:',self.cognome)
```

The init method takes as input the keyword self and the values of the object attributes necessary for the definition, which are represented by variables.
The values are associated with the corresponding attributes using the syntax self.attribute=value and it is the practice that attribute and value have the same name.
The second form print_name takes only the mandatory input self, through which the attribute values can be accessed.
An object of a class can be instantiated in a similar way to the function call and the output will be returned the instance which can be stored in a variable with the syntax object = nome_classe(inputs)

```
>>> alex = Persona('Alex', 'Burger',14)
>>> alex.stampa_nome()
Nome: Alex Cognome: Burger
>>> alex.anni
14
>>> alex.altezza = 170
>>> alex.altezza
170
```

Notice in the example how you can interact with objects, for example by adding new attributes at runtime.

Python: Virtual environments

Virtual environments are a tool that allows you to create spaces independent from the rest of the system where you can install different versions of Python and its modules.
Virtual environments can mainly be managed via conda or via the Python Standard Library venv package.

The following command displays a list of available environments: in the first case, the active environment is the basic one which is activated by default with the installation of Anaconda or Miniconda; in the second case, the instruction to create the py35 environment (Python 3.5) is presented:

```
(base) username:~$ conda info --envs
# conda environments:
#
base                  *  /home/username/miniconda3
```

```
(base) username:~$ conda create --name py35 python=3.5
```

Once the environment has been created, it can be activated, as shown in figure 1 below, and proceed with the installation of the desired packages via conda.
Figure 2 shows how it is sufficient to type conda deactive to exit the environment.

```
(base) username:~$ conda activate py35 python=3.5
(py35) username:~$
```

```
(py35) username:~$ conda deactivate
(base) username:~$
```

Using venv: it is advisable to create a folder where to store the various environments (fig.1); create and activate a my_env environment which involves creating a folder of the same name in the newly created Python_envs folder; activate the environment with the command source my_env/bin/activate (fig.2); proceed to install the packages via pip. To exit the environment type deactive (fig4).

```
username:~$ mkdir Python_envs
username:~$ cd Python_envs/
username:~/Python_envs$
```

→

```
username:~/Python_envs$ python -m venv my_env
username:~/Python_envs$ source my_env/bin/activate
(my_env) username:~/Python_envs$ which python
/home/username/Python_envs/my_env/bin/python
```

```
(my_env) username:~/Python_envs$ deactivate
username:~/Python_envs$
```

Python: forms

Python ships with a set of modules known as the Standard Library.
The Python Standard Library offers a wide variety of features, described at the link: https://docs.python.org/3/library/index.html
To use the functionality of the modules, you must first import them.
Below is an example of importing the math module, with two possible alternatives:

```
>>> import math
>>> math.sin(0)
0.0
>>> math.sin(math.pi)
1.2246467991473532e-16
```

```
>>> from math import *
>>> sin(0)
```

It is common to rename modules with a long name to a shorter name.
This can be done with the syntax import long_name as short_name:

```
>>> import numpy as np
>>> np.sin(0)
0.0
```

Python: modules – numerical computation with numpy

Numpy is the most used numeric library.
Numpy deals only with numeric data types and its main class is the array class, which can represent scalars, vectors, matrices or tensors.
The documentation is accessible at the link: https://numpy.org/doc/stable/

The numpy module allows you to perform the classic linear algebra operations.
For more sophisticated linear algebra operations you can access the functionality of numpy's linalg module.

In the following example we deviate a matrix A and a vector V:

```
>>> import numpy as np
>>> A = np.array([[-1,0,2],[-7,8,3]])
>>> print(A)
[[-1  0  2]
 [-7  8  3]]
>>> v = np.array([1,2,-1])
>>> print(v)
[ 1  2 -1]
```

Python: modules – statistical analysis with pandas

Pandas provides tools for reading datasets, redefined as dataframes.
Documentation available at the link: https://pandas.pydata.org/docs/

You can create a dataframe and initialize it with a dictionary where the values are lists of the same length.
Ex: we use the DataFrame head module to display the first five rows of the dataframe dt initialized below (fig.1).

Pandas is a library for the fundamental functions of Data Science. Two useful methods to get an overview of the dataframe (fig.2): the info() method informs about the size of the dataframe and the type of data contained in each column, while the describe() method provides statistical information.

```
>>> import pandas as pd
>>> df = pd.DataFrame({'Title' :["Tenet", "Joker",
                                 "The Hateful Eight", "Titanic",
                                 "The Matrix", "The Shining",
                                 "Dirty Dancing", "A Clockwork Orange"],
                        'Year' :[2020, 2019, 2015, 1997, 1999, 1980, 1987,
    1971],
                        'Director': ['Nolan', 'Phillips', 'Tarantino', '
    Cameron',
                                     'Wachowski','Kubrick', 'Ardolino','Kubrick
    ']})
>>> df.head()
                Title  Year   Director
0               Tenet  2020      Nolan
1               Joker  2019   Phillips
2  The Hateful Eight  2015  Tarantino
3             Titanic  1997    Cameron
4        The Matrix  1999  Wachowski
>>>
```

```
>>> df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Title     8 non-null      object
 1   Year      8 non-null      int64
 2   Director  8 non-null      object
dtypes: int64(1), object(2)
memory usage: 320.0+ bytes
>>> df.describe()
                Year
count      8.000000
mean    1998.500000
std       18.470053
min     1971.000000
25%     1985.250000
50%     1998.000000
75%     2016.000000
max     2020.000000
```

Python: modules – data visualization with matplotlib

Matplotlib provides tools for data visualization.
Documentation is available at the link: https://matplotlib.org/stable/index.html

Usually the *pyplot* module is imported with the *%matplottlib* inline instruction: this allows the display of the graphs in the same notebook card.

```
>>> import matplotlib.pyplot as plt
>>> %matplotlib inline # solo su notebook
```

The function to create 2d graphs is plot(), which takes as input two lists of equal length containing the abscissa and ordinate values.
The plot function draws a line that joins the points in the order in which they are specified. Ex: y = 2x

If you're using Python from the command line, the show command will create the graph in a separate window from the terminal. (fig1)

It is possible to modify the style of the graph which can be specified with a string in which the first character indicates the color and the second indicates the symbol to be used to mark the points (known as marker), and finally the style of the broken line is indicated. It is also possible to specify only part of the characteristics mentioned.
Finally, the linewidth and markersize keywords are used to increase the thickness of the linewidth and the size of the marker.
The savefig() function unloads the last figure created in the working directory with the name and extension specified in input (fig.2)

```
x = [1,4]
y = [2,8]
plt.plot(x,y)
plt.show()
```

```
>>> plt.figure()
>>> plt.plot(x,y,'ro--', linewidth = 3, markersize = 12)
>>> plt.savefig('figura1.png')
>>> plt.show()
```

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 6

Main methodological approaches in a customer centric key

Pearson

Data-driven marketing approach

The data-driven marketing approach consists in defining strategic corporate decisions with a scientific approach: companies deepen their knowledge of their customer base through the analysis, study and synthesis of the data collected and processed, generating insights to support/guide the strategic decisions with the ultimate goal of defining and developing a personalized proposal for the different customer segments, in order to maximize the profitability of the relationships along the entire customer life cycle.
Corporate value is maximized by offering the right product to the customer, at the right stage of the customer journey and according to the preferred channel and methods.
The data-driven approach improves business performance by 20%.

The interconnected pillars supporting data-driven marketings are:

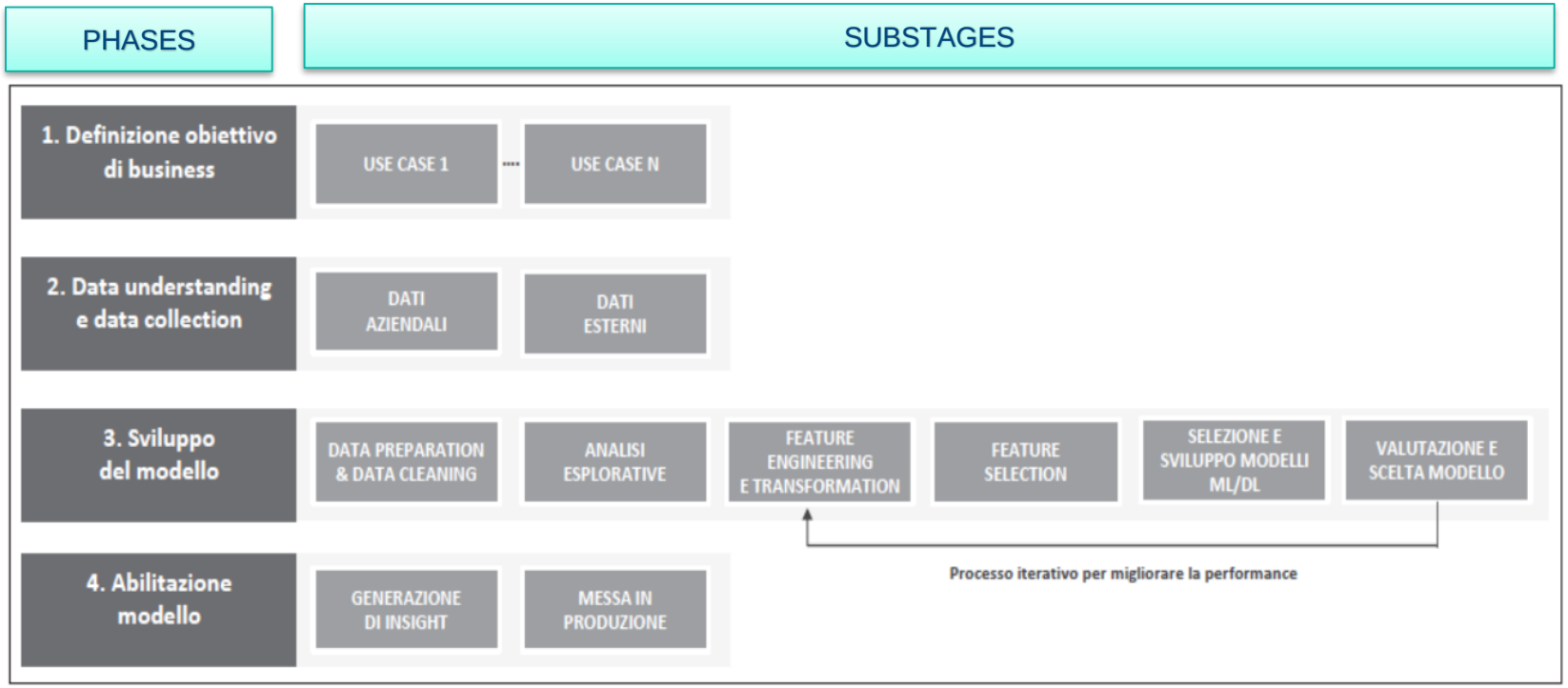
ANALYTICS	Statistical-mathematical analysis of the data and subsequent grouping into segments
ADVANCED ANALYTICS	Possibility of analyzing data and contents in autonomous or semi-autonomous mode with the aim of discovering relationships and correlations, developing forecast analyzes and recommendations
ARTIFICIAL INTELLIGENCE	Ability to use computers to understand human intelligence: The computer simulates human intelligence in solving a given problem. Artificial Intelligence (AI) encompasses two scientific sub-fields or models: - Machine Learning (ML): A set of methods that allow a software program to adapt and learn on its own. The computer is capable of making new and independent decisions, based on the study of input data; - Deep Learning (DL): consists of a class of machine learning algorithms capable of analyzing and learning from the effectiveness of its previous decisions, to continuously and constantly improve the quality of its future decisions, testing multiple options at a time.

Model development stages

The process of developing a model involves the collaboration of four players: data scientist, business owner, data engineer and data architect.

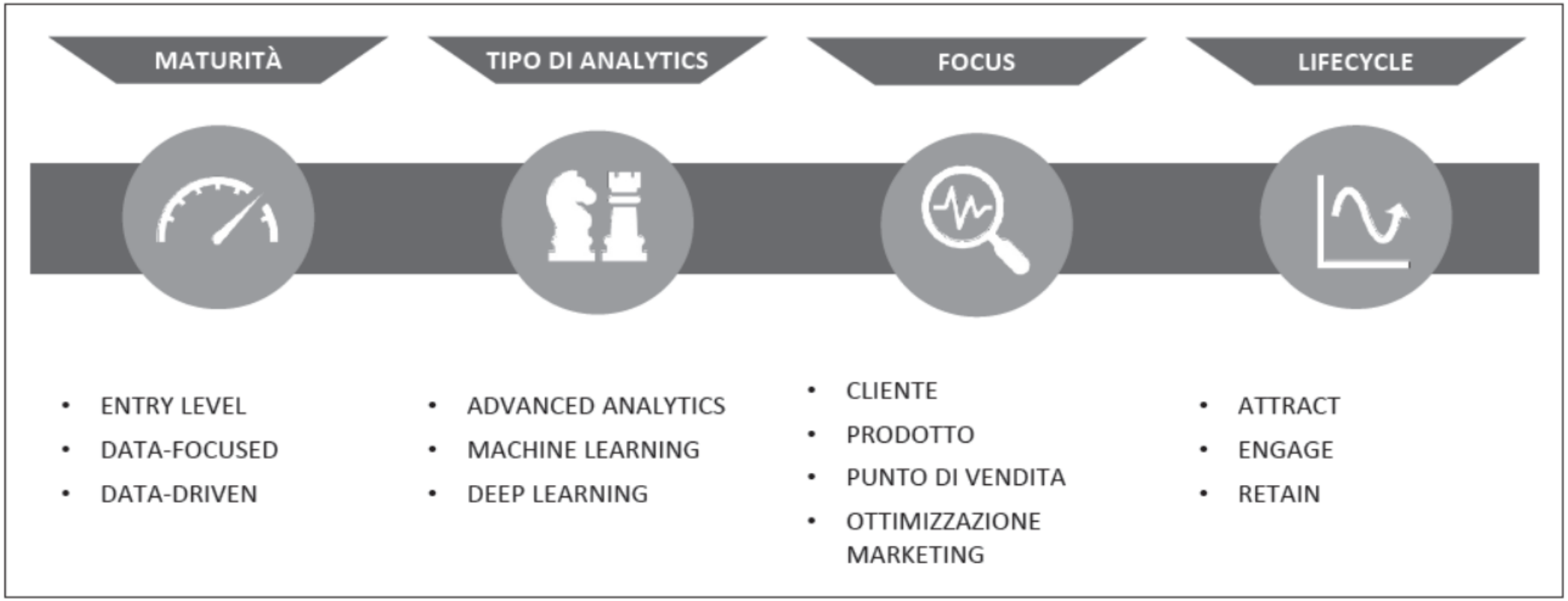
The development approach of a model is divided into 4 phases:

1. Identification of the objective of the analysis, of the tools to be used, of the type of output to be obtained and of the main business objectives
2. Mapping of internal and external data sources and the process of understanding the data and the needs they will have to satisfy
3. Definition of the components of the model, taking into account the complexity and specificity of each industry
4. Drafting of a summary documentation with the main insights that emerged from a strategic point of view and the technical details, with the description of all the decisions made during the development of the model.
   The testing of the code through unit-tests and UAT and the preparation of logging and monitoring systems is essential.



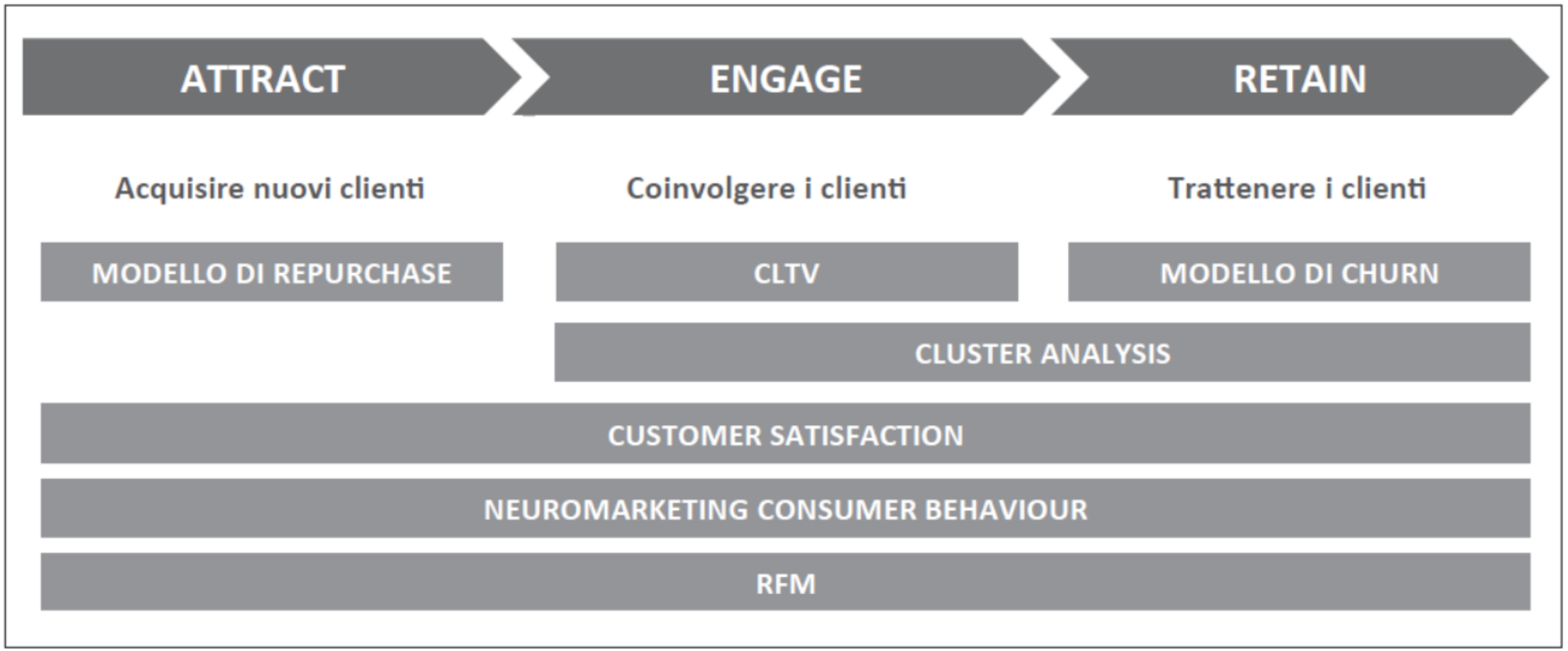
The 4 pillars at the basis of the development of a model

Below are the four fundamental pillars, and related types and phases, preparatory to the development of an Advanced Analytics or Artificial Intelligence algorithm



The main models in a customer centric key

The main customer models and algorithms are shown below, regardless of the company's maturity level towards a data-driven approach and the type of Analytics, useful for generating insights along the customer lifecycle



Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 7

DETERMINISTIC SEGMENTATION –

RFM

Pearson

Deterministic segmentation - RFM

The RFM algorithm is based on three components:
- Recency (R): length of time since the last purchase
- Frequency (F): number of purchases made in a given period of time
- Monetary (M): amount spent

By dividing each variable into categories, a series of segments are defined to which a probability of response to the commercial offer is given.
The goal is to take a snapshot of a company's customer base, identifying customer segments to which to propose specific marketing actions.

The methodological approach:

1. Definition of the time parameters T1 (reference period) and [T0 – T1] (analysis time interval)
2. Verify the correctness and distribution of the variables to define the number of optimal categories into which to divide the three dimensions identified
3. Perform the Cartesian product of the categories to identify the possible RFM segments, identified with an ordered triplet (R,F,M) which represents belonging to the category along the axes
4. Verification of segments that are not very penetrating and/or not very actionable and management of the trade-off, understood as the reduction of the number of categories or unification of segments
5. Creation of a summary indicator (score).

The main stages of development of the RFM algorithm

The phases and sub-phases necessary for the development of the RFM algorithm are shown below:

ESPLORAZIONE DATI E CALCOLO INDICATORI	SVILUPPO DELL'ALGORITMO	BUSINESS INSIGHT	INTEGRAZIONE NEI SISTEMI
• *Data understanding* e *data collection*. • *Data preparation* e *data cleaning*. • Analisi esplorative. • Definizione e sviluppo dei tre indicatori *Recency*, *Frequency* e *Monetary*. • Costruzione della *customer table* di analisi.	• Definizione e analisi delle soglie per ciascun indicatore. • Assegnazione di un eventuale peso a ciascun indicatore. • Combinazione dei valori ottenuti per definire un indicatore aggregato di sintesi RFM. • *Fine tuning* dell'algoritmo.	• Definizione degli obiettivi di marketing sui segmenti e ideazione di attività personalizzate. • Definizione di un piano di *contact strategy* per l'implementazione di azioni di marketing *data-driven*.	• Integrazione dell'algoritmo RFM nei sistemi aziendali.

Benefits and limitations of RFM

BENEFITS

- A priori choice of the dimensions on which to develop the analysis;
- Easy interpretation of the segments obtained;
- Model useful for entry-level companies for the development of an evolutionary roadmap in the data-driven field;
- Obtaining strategic insights on the customer base to increase loyalty and retention and to develop up-selling and cross-selling actions;
- Development of tailor-made marketing actions;
- Timely intervention on customers migrating to a lower value segment

LIMITATIONS

- The use of a limited number of variables penalizes the results compared to more advanced models;
- Recency, Frequency and Monetary are often correlated with each other, generating an overlapping effect during segment construction;
- It does not allow you to natively obtain a single summary indicator (score).

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 8

The behavioral segmentation–

cluster analysis

Pearson

The behavioral segmentation – Cluster Analysis

Cluster Analysis includes the set of all the algorithms through which observations can be divided into subgroups (clusters), in such a way that all observations within each cluster are similar to each other but massively heterogeneous from those contained in the others subgroups.

It is used to divide the customer base into clusters, in order to better understand the characteristics and needs of each group and to be able to convey tailor-made marketing activities.

Given a number n of observations, the possible subdivisions of these into k clusters of dimensionality x are equivalent to a number of combinations equal to $\frac{1}{K!}\sum_{x=1}^{k}(\frac{k}{x})x^n$

It is possible to categorize the Cluster Analysis algorithms according to the main cluster construction methods:

PARTITION METHODS	They subdivide the set of observations into a predetermined number of non-empty subgroups
HIERARCHICAL METHODS	They define a hierarchy within the subgroups using a tree structure and increasing similarity thresholds
OTHER METHODS	Graph method, grid method, density-based methods, neural methods etc

Similarity measures and PCA methodology (Principal Component Analysis)

It is necessary to define some similarity measures that depend on the types of variables used.

It is assumed to work with a dataset D made up of n observations and m variables which can be represented by a matrix with n rows and m columns, where each observation can be defined with a vector xi = (xi1, xi2, ..., xim)

Normalization:
Index of $\boldsymbol{Norm}_{(0,1)}$:

$$\boldsymbol{Norm}_{(0,1)} = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

METRICS FOR NUMERICAL VARIABLES

- Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2}$$

- *Mahnattan distance*:

$$d(x_i, x_j) = \sum_{k=1}^{m} |x_{ik} - x_{jk}|$$

- *Minkowski distance*:

$$d(x_i, x_j) = \sqrt[q]{\sum_{k=1}^{m} |x_{ik} - x_{jk}|^q}$$

METRICS FOR CATEGORIC VARIABLES

- Ordinal variables: the categorical values are reduced to numerical values belonging to the set of natural numbers. In this way it is possible to apply the measures of distance of the numerical variables
- Nominal variables: each variable can be represented by a number of binary attributes equal to the possible distinct values assumed by the variable itself. Thanks to this transformation into binary variables it is possible to calculate the Jaccard distance: $d(x_i, x_j) = \frac{q+r}{p+q+r}$

Following standardization it is important to simplify the analysis dataset by including only a minimal subset of relevant variables (that can be easily interpreted) through the PCA methodology.

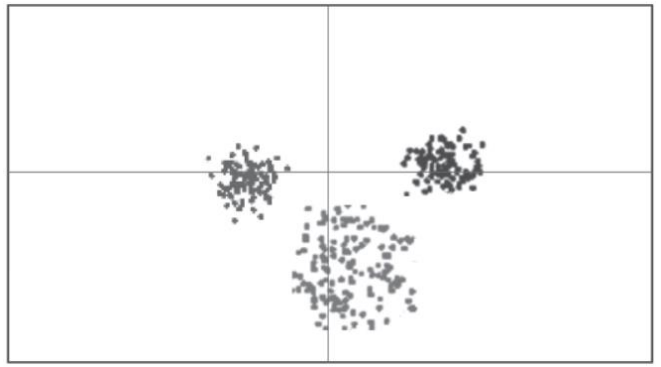
The goal of the PCA methodology is to create a linear transformation that allows replacing a set of M numerical variables with a smaller subset of K latent variables, without generating loss of information.

Partition methods

Partition methods apply an iterative procedure that reallocates the observations within the cluster in order to improve their representativeness. The best known of the partition methods is the k-means algorithm. The stages of development are as follows:

1. Initialization: a random choice is made of the observations within the dataset that constitute the centroids of the initial clusters in the current iteration;
2. Each observation is assigned to the cluster whose centroid is most similar to the observation itself;
3. In case no observations are assigned to a different cluster than in the previous iteration step, the algorithm terminates and converges;
4. Otherwise, the new centroid is calculated for each cluster as the average of the observations belonging to the cluster and we proceed with a new iteration, starting again from point 2.

This algorithm can be strongly influenced by any outliers present in the dataset, which risk having a great impact on the average value, thus moving the centroids from their optimal positions: it is advisable to proceed first with a data cleaning activity.



Graphical representation of three clusters in a two-dimensional space

The hierarchical methods

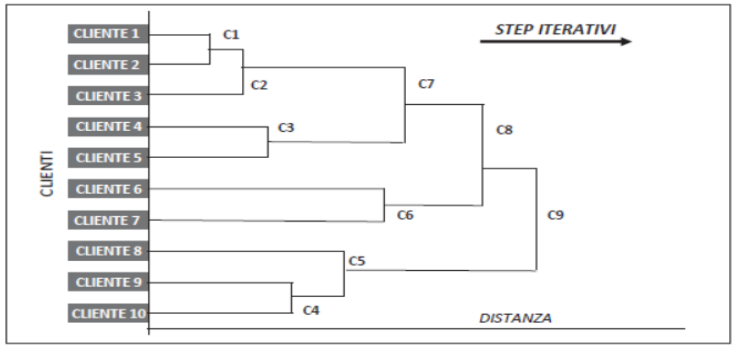
Hierarchical methods are necessary to overcome the problem of partition methods which require conducting tests as k varies and applying a posteriori evaluation methods to identify the solution that best meets business needs to define the number of k clusters a priori.

It is first necessary to define which distance between two clusters to apply between: average distance, maximum distance, distance between centroids, Ward distance.

Next you decide which hierarchical method to apply:

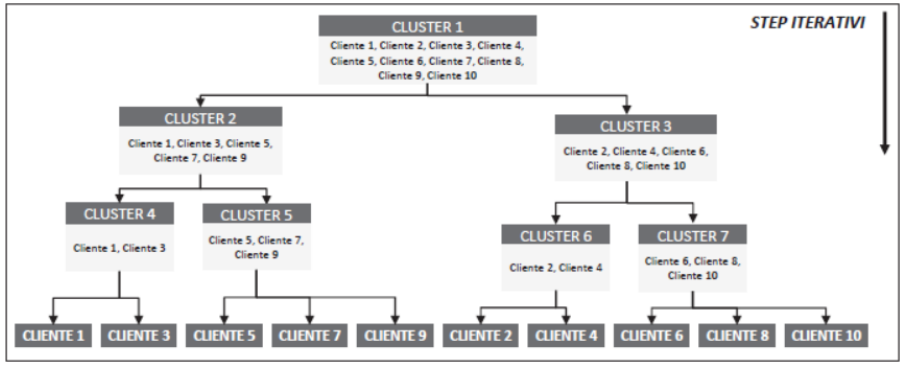
BOTTOM - UP

1. Initialization: as many clusters are created as there are total observations in the dataset;
2. The distance between the clusters is determined and the two clusters with the minimum distance are merged to create a new cluster;
3. The distance between the new cluster and the pre-existing clusters is calculated;



TOP - DOWN

1. Initialization: all observations are placed in a single cluster;
2. The possibility of further subdividing the clusters is evaluated to maximize the distances between them;
3. The procedure ends when a number of clusters equal to the number of observations in the dataset is obtained, i.e. when the clusters obtained are composed of a single observation, or when a specific stopping criterion is reached



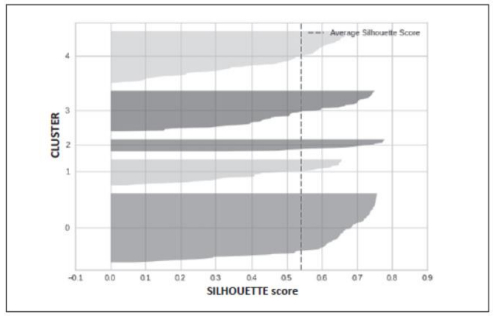
Evaluation methods

The evaluation of an unsupervised algorithm cannot take place on the basis of a target with which to make an objective comparison but must refer to the main objective of the algorithm itself: to identify groups of observations that are as similar as possible to each other based on certain predefined characteristics.

Evaluation methods:

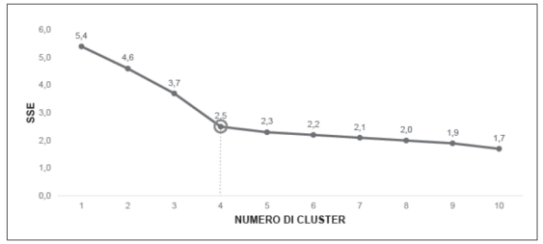
SOLHOUETTE'S COEFFICIENT

It measures how much an observation is similar to the others contained in its cluster (coefficient of cohesion) or how much it is dissimilar (coefficient of separation).
The Silhouette coefficient S_c is in the interval [-1,1]: if close to +1 the clusters are very distant, if equal to -1 it indicates a wrong assignment.



ELBOW METHOD

It presents the number of clusters on the abscissas and the SSE on the ordinates and shows how the quality of the segmentation improves as the number of clusters increases. The point identified by the "elbow" created by the curve represents the trade-off between the increase in complexity and the benefit obtained



PSEUDO – F STATISTIC (PSF)

Compare the ratio between the inter-cluster variance and the intra-cluster variance: : Pseudo-F = $\frac{(BSS)/(k-1)}{(WSS)/(n-k)}$.

A large value indicates separate cluster. The peaks in the graph allow to identify the ideal number of clusters.

The main development phases of the Cluster Analysis algorithm

The phases and sub-phases necessary for the development of the Cluster Analysis algorithm are shown below:

ESPLORAZIONE DATI E CALCOLO VARIABILI	SVILUPPO DELL'ALGORITMO	BUSINESS INSIGHT	INTEGRAZIONE NEI SISTEMI
• *Data understanding* e *data collection*. • *Data preparation* e *data cleaning*. • Analisi esplorative. • Definizione e calcolo delle principali variabili. • Costruzione della *customer table* di analisi.	• Standardizzazione/ normalizzazione delle variabili. • Analisi di correlazione, variabilità, analisi delle componenti principali (PCA). • Sviluppo di algoritmi gerarchici. • Sviluppo dell'algoritmo delle *k-means*. • Scelta della soluzione migliore in termini di performance.	• Interpretazione dei cluster. • Definizione degli obiettivi di marketing sui segmenti. • Definizione di un piano di *contact strategy* per l'implementazione di azioni di marketing data-driven.	• Automatizzazione e integrazione dell'algoritmo di clustering nei sistemi aziendali.

Benefits and limitations of Cluster Analysis

BENEFITS

- Obtaining strategic insights on the customer base that can be capitalized to increase loyalty, optimize acquisition, and develop up/cross-selling actions
- Take advantage of these insights to develop tailor-made marketing actions;
- Integrate insights into marketing automation;
- Understand which is the best acquisition and communication channel for each group;
- Compare the percentage distribution of the segments also from a geographical point of view

LIMITATIONS

- Difficulty in choosing an optimal set of variables;
- The need to resort to a small number of variables and to standardize them, losing information power and making it more difficult to interpret the clusters obtained;
- Use the results of the segmentation by proposing a specific offer to the customers of each group, without applying further filters afterwards in the selection of the target

Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 9

CHURN AND REPURCHASE MODELS

Pearson

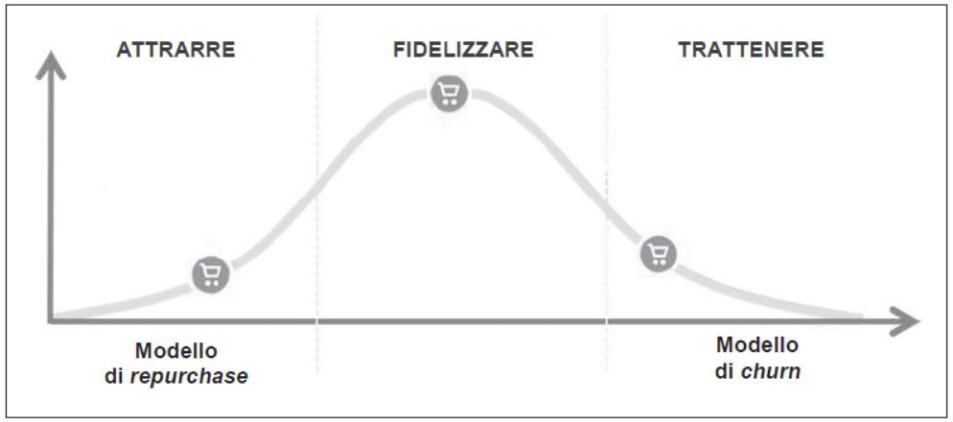
Churn and repurchase models

The churn and repurchase models study respectively the abandonment and the propensity to repurchase of new customers.
The purpose of the churn is to predict which customers are most likely to abandon and allow the company to intervene with appropriate retention initiatives.
It focuses on repeater customers (who have made more than one purchase) and analyzes the main characteristics of those who leave the company, interrupting any type of service or relationship with it.
In relation to the type of abandonment (total phenomenon or partial interruption) it is possible to divide the churn into scenarios and types.

The repurchase model focuses on one-shooter customers, i.e. customers who have made a first and only purchase, with the aim of estimating the probability of a repurchase.

SCENARIO	TIPO DI CHURN
CONTRATTUALE	VOLONTARIO
I clienti sottoscrivono un servizio che prevede una data di termine.	I clienti scelgono di dismettere l'utilizzo del prodotto o servizio.
NON CONTRATTUALE	INVOLONTARIO
I clienti sono liberi di acquistare o non acquistare in ogni momento.	I clienti sono spinti a dismettere l'utilizzo del prodotto o servizio per cause esterne.

Subdivision of the churn by scenario and type



Repurchase and churn models with respect to the customer lifecycle

Elements that affect customer return

The main factors influencing the purchase and return of customers are:

- Reference sector of the industry;
- customer experience
- Increase in omnichannel and touchpoints
- Structured onboarding process based on specific customer interests and needs;
- Evaluation of natural repurchase cycles

The methodological approach

1 **DEFINITION TARGET**

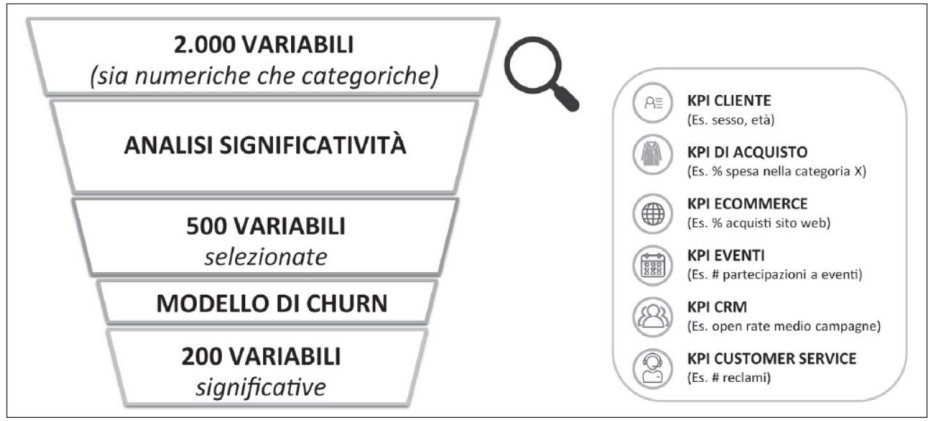
The first step for the development of the churn model (and also of the repurchase model) is the definition of the target. In the case of churn it is necessary:

- Evaluate which type and scenario of abandonment you want to deal with (consistent with the business model);
- Identify outlier values and evaluate potential exclusions of customer types;
- Develop exploratory analyzes prior to the creation of the model, to verify which indicators can be used as substitutes;
- Define an analysis interval and divide it into:
  - Observation period: analysis time window within which to calculate potential explanatory variables;
  - Latency period: any latency window within which to calculate the variables, update the scores, carry out marketing actions;
  - Churn period: time window within which to verify the customer's behavior to define the target variable

In the case of the repurchase model it is necessary:
- Identify new customers to define the target;
- Calculate the average distance between the first purchase and the next (inter-purchase distance);
- Analyze the repurchase curve and define the time frame to use

2 **EXPLANATORY VARIABLES**

We continue with the definition of the explanatory variables divided into
macro categories to include in the template: make your selection
Through correlation analysis and the main tests of significance



Example of selection funnel of explanatory variables

The methodological approach

3 BALANCING

We continue with the verification of the balance of the dataset through techniques:
- Undersampling: removal of some observations from the majority group;
- Oversampling: duplication of some minority group observations;
- Synthetic-data generation: algorithmic generation of new observations of the minority group with characteristics similar to those present

4 TRAIN-TEST CROSS-VALIDATION

Once the balancing assessments have been carried out, we continue by evaluating the need for split train-tests or cross-validation for training the Machine Learning models.
Below are the algorithms most commonly used to estimate the probability of abandonment or repurchase:

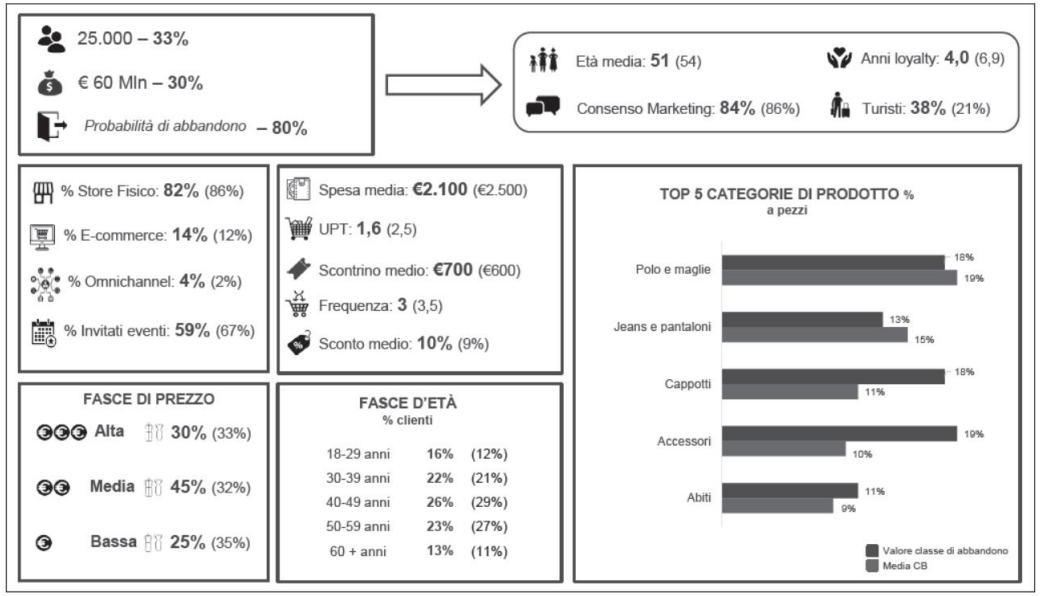
Performance measures and key KPIs

The main performance measures for churn and repurchase models are:

AUC	It is the area under the ROC curve that shows the positive flase indicator along the abscissa axis and the recall indicator along the ordinate axis and allows you to visually compare the performance of the different classification models. The closer the value is to unity, the more the model will perform.
ACCURACY	The accuracy of the predictions, given by the number of correct predictions on the total number of cases analysed
ACCURACY	Percentage of customers who are correctly recognized as positive by the model (true positive) out of all customers who abandoned in the analysis period

The figure shows the main KPIs relating to the customer group
with a high probability of abandonment.

The insights found and the characteristics of the groups allow for the construction of a tailor-made promotional activity plan, with ad hoc content, timing and methods of communication.



Benefits, limitations and evolutions of the churn model

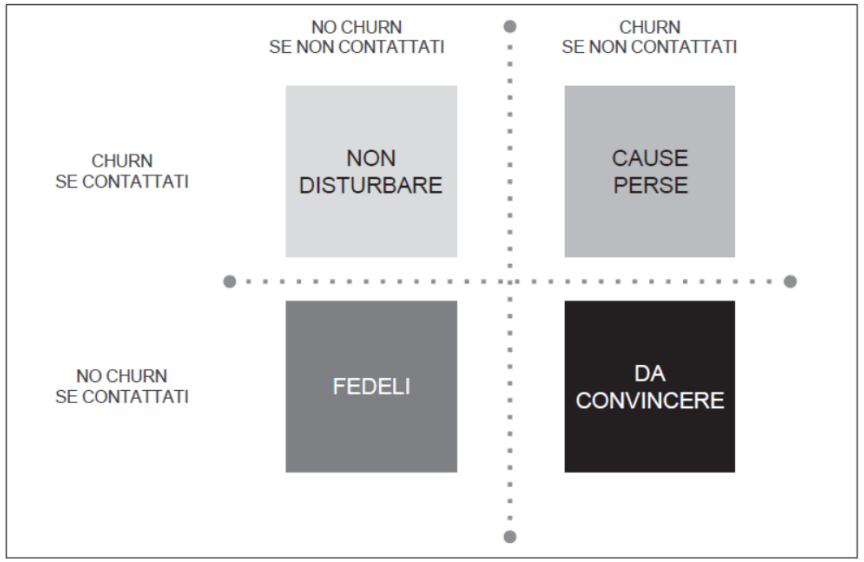
BENEFITS

- Knowing the reasons that drive a customer to abandonment allows you to implement data-driven retention strategies;
- The insights generated by the models can be integrated into customer care activities to customize targeted actions on specific customer targets;
- The model score can be used to diversify, prioritize and optimize your marketing contact plan

LIMITATIONS

- Difficulty of changing the outcome of the model through marketing actions. The model is developed on the actual customer churn, while the business need is to have a model that predicts which customers are willing to leave.

The uplift models are presented as an evolution of the approach to the churn problem: they divide the customer base into four categories along two dimensions, conditioned by the marketing actions carried out:



The repurchase model allows you to identify insights for the development of marketing actions to convert one-shooter customers into repeaters.
The limit is linked to the scarcity of available data, since only the first and only act of purchase is considered.

The evolutions are linked to data enrichment through the use of third-party data (Google Analytics, Adobe Analytics, ISTAT, IPSOS, SWG)

Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 10

CLTV MODEL

Pearson

CLTV model

The CLTV (customer lifetime value) model aims to predict the economic value of a customer throughout the lifecycle, from acquisition to possible abandonment, maximizing the relationship over time, focusing on enhancing the customer even in the long term.

Main applications:
- Evaluation of customer acquisition costs starting from segmentation by acquisition channel;
- Application linked to service models defined according to both the current and prospective value of customers;
- Identification of segments with the highest future value, study of needs in terms of product or service and creation of customer journeys dedicated to their development

Benefit: Represent the prospective value of each customer.

Limits: difficulty in finding historical data with enough depth to allow correct application in the long term;
difficulty in obtaining immediate indications on the marketing actions to be implemented.

Methodological approach

HISTORICAL APPROACH

Algorithmic formulation based on certain variables:
- CLTV amount, frequency and duration of the relationship: CLTV = rt = (mf)t
- CLTV amount, frequency, duration of the relationship and profit: CLTV = rpt = ((mf)p)t
- CLTV weighted by interest rate and churn: CLTV = rpt ($\frac{(1-c)}{(1+i)-(1-c)}$)

PREDICTIVE APPROACH - MACHINE LEARNING

Formulation using Machine Learning model:
1. Definition of the time horizon and of the customers whose tenure is >= the time horizon
2. Creation of model features based on mainly socio-demographic and behavioral variables
3. Data quality activities, exploratory analyzes and statistical tests to select, transform and verify the significance of features
4. Model testing via train-test or cross-validation
5. Evaluation of the ex-post model through: linear regression, decision tree, random forest, gradient boosting, neural network

PREDICTIVE APPROACH – ECONOMIC MODELS

Formulation using econometric models.
The distribution curves of some parameters are analyzed in an attempt to construct a probability distribution used as a model for estimating customers' CLTV.
Among the most used are the BTYD models which are based on two stochastic processes: a purchase frequency estimation model and a churn estimation model.
Examples of BTYD: Pareto model/NBD and Beta-Geometric model/NBD

Evaluation methods

The main evaluation methods of these models take place through backtesting, i.e. applying the model to customer tables built on historical data and evaluating its validity on the basis of the prediction error committed by the various models.

MEAN SQUARED ERROR: MSE = $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$

ROOT MEAN SQUARED ERROR: RMSE = $\sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}$

MEAN ABSOLUTE ERROR: MAE = $\frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$

MEAN ABSOLUTE PERCENTAGE ERROR: MAPE = $\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \hat{Y}_i|}{Y_i}$

Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 11

MEASUREMENT OF CUSTOMER SATISFACTION

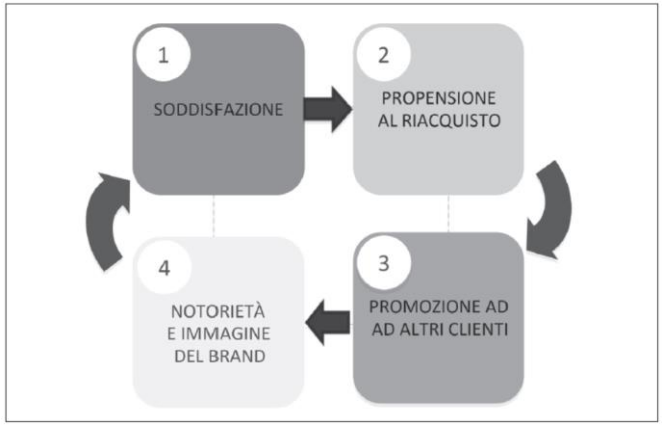
Pearson

Behavioral segmentation – Cluster Analysis

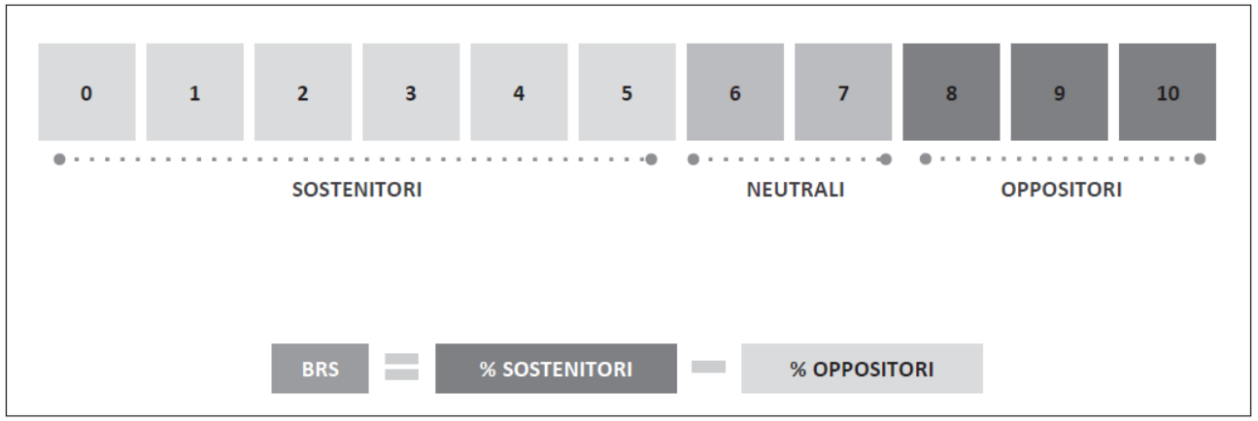
Customer Satisfaction is defined as the customer's opinion that emerges from the difference, negative or positive, between the perception of the service used or the product purchased and its expectations, which can be interpreted as expectations and desires.
The perception of customer satisfaction is subjective and depends on multiple factors, often representing a trade-off between the perceived benefits and the costs incurred for the product or service and the customer experience.
It becomes essential to satisfy the customer to generate trust and therefore loyalty, to have better economic results.

The main dimensions for measuring Customer Satisfaction through surveys are: tangible aspects, reliability, ability to respond, ability to reassure, empathy.

A satisfied customer, in addition to being more inclined to repurchase, is more likely to recommend the brand to other potential customers. The main brand recommendation indicator is the BRS (Brand Recommendation Score) and its measurement derives from a single question asked to the customer: «*On a scale from 0 (unlikely) to 10 (very likely) how much would you recommend a product or service of the brand to your acquaintances, friends and relatives?*»



Virtuous circle of CS



Calculation of BRS

The main stages of development of Customer Satisfaction

The phases and sub-phases necessary for the development of the Customer Satisfaction analysis are shown below:

MISURAZIONE CS	RACCOLTA DATI E VISUALIZZAZIONE	ROOT CAUSE ANALYSIS	INTEGRAZIONE NEI SISTEMI
• Misurazione della *customer satisfaction* tramite survey sui clienti.	• Raccolta dati. • Valutazione dei risultati: *data quality*, analisi descrittive e test di significatività.	• Identificazione delle problematiche tramite analisi multivariate per riconoscere quali sono i driver che impattano maggiormente sulla soddisfazione. • Interpretazione degli *insight* e definizione delle possibili azioni correttive.	• Implementazione delle azioni correttive all'interno dei sistemi e dei processi aziendali.

Ciclo continuativo di CS

Benefits and limits of Customer Satisfaction

BENEFITS	LIMITS
The systematic and capillary monitoring of the interactions between the company and its customers makes it possible to isolate any problems where they arise and to intervene promptly thanks to customer feedback.	The tendency of customers to be more inclined to complain and remember a negative experience than to provide positive feedback.

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 12

Neuromarketing and artificial intelligence

Pearson

Neuromarketing and Artifical Intelligence

To truly understand their customers, companies must integrate rational investigations and emotional investigations: the former measure what people say and help to discover the rational, conscious and verbal level; the latter allow you to complete the profile, measuring what people feel by highlighting the non-verbal, the subconscious and the more emotional aspects.

Companies capable of identifying emotions, qualifying them and applying them to support customer behavior are characterized by an unbeatable competitive advantage.

Neuromarketing bases its roots on:

AIDA MODEL	It investigates the purchase decision-making process, dividing it into four stages: - Awerness: the consumer becomes aware of the product - Interest: the consumer perceives the advantages deriving from the purchase - Desire: the consumer develops a preference for the brand or product - Action: the consumer makes the purchase
fMRI	It creates a map of the functions of various regions of the human brain, based on the variation in blood supply
ECG	Monitor cardiac activity and study the emotional impact generated by a stimulus
EEG	It monitors brain wave activity and measures the change in level of attention
RAGGIX & NEUROIMAGING	

Areas of application of neuromarketing

The four main areas of application of neuromarketing and the related benefits are shown below:

Aree di applicazione	ADVERTISING AREA	DIGITAL AREA: SITO WEB, APP E SOCIAL MEDIA	CUSTOMER EXPERIENCE	BRAND REPUTATION
Benefici	• Rilevare l'efficacia di un'adv in termini di attenzione, emozione e coinvolgimento cognitivo. • Migliorare l'efficacia delle campagne di comunicazione. • Valutare diverse alternative per una specifica pubblicità e scegliere la migliore da un punto di vista di coinvolgimento emotivo.	• Indagare il comportamento dell'utente mentre interagisce con sito web, app e social media. • Identificare le parti di un'immagine o un video che hanno registrato i maggiori livelli di coinvolgimento emotivo. • Identificare la risposta implicita ed esplicita dei tester a diversi formati, contenuti o argomenti trattati nei video/post.	• Misurare le reazioni inconsce che le persone provano mentre utilizzano i prodotti o usufruiscono dei servizi. • Scoprire l'interesse, l'attenzione e l'emozione del cliente per motivarlo all'acquisto e migliorare il processo di customer experience.	• Rilevare la percezione emotiva sui marchi che le misure esplicite non possono misurare. • Rivelare attributi ed emozioni associate al marchio e al logo. • Costruire mappe di percezione tra misurazioni esplicite (es. survey) e implicite derivanti da tecniche di misurazione di neuromarketing.

Benefits of neuromarketing and comparison with traditional marketing

BENEFITS
Understand and study the subconscious behavior of customers as they interact spontaneously;
Use reliable neurometric indicators through a highly innovative approach and a scientific model;
Capturing customers' instinctive responses free from any cognitive bias and rationalization process;
Learn from clients' subconscious thoughts and feelings that are otherwise inaccessible;
Improve your marketing strategies with an orientation towards innovation and the future;
Reduce the time between the data collection phase and the insight generation process compared to traditional survey or focus group techniques;
Find out which marketing stimuli are most effective in attracting, engaging and motivating a customer to stay loyal to the brand

Unlike the results achieved with traditional marketing research (focus groups, CATI, CAWI, CAPI) where the answers may contain a certain level of rational bias, neuromarketing makes it possible to scientifically decipher the customer's unconscious thoughts, obtaining information without of any kind of rational "filter".

Neuromarketing techniques and tools

Below are three types of neuromarketing analysis:

BRAIN IMAGING

Process that allows the anatomy and activity of the brain to be photographed.
In neuromarketing it is used to monitor the brain response of consumers following their exposure to physical and neurological stimuli.
EEG is used to detect changes in the level of attention of customers when they view different versions of an advertisement or compare multiple websites.
The fMRI is used to distinguish the decisions made following the activation of an area used for analytical thinking, therefore influenced by memory, from those deriving from the activation of an area that reflects a more primitive and natural emotional state

PHYSIOLOGICAL ANALYSIS

Analysis of physiological behaviors that clients unknowingly exhibit in response to a stimulus.
The ECG is used to record the variations of emotions experienced by a tester in response to situations of concentration, anxiety, happiness, embarrassment, fear. It combines with other techniques such as: Heart Rate Variability, Emotional Valence, Emotional Arousal.
GSR is used to monitor the emotional response of the testers, as an increase in sweat gland activity is generally proceeded by the activation of the sympathetic branch of the autonomic nervous system: exposure to strong emotions of the testers leads to an unconscious increase in sweating .

BEHAVIOR-MENTAL ANALYSIS

Used to study the behavior of customers during the purchase phase.
- The Eye-tracker records the movements, dilation and contraction of the pupil in response to a stimulus: it is used to understand which elements attract a customer's attention. Monitor where a tester is looking, how long they have been looking at a given spot, and any changes in pupil size that are affected by autonomic nervous system activation or deactivation. It therefore has a significant correlation with the person's emotional response to situations of aversion, pleasure, interest, etc.
- The IRT is a quick question-answer association test: it is used for benchmark analysis between brands

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 13

Methodological approaches for analyses

in product key

Pearson

Methodological approaches for analyses in product key

It is also essential to take into consideration the relevance that the product or service offered may have within the company environment, without forgetting the needs and requirements of the customer.
It is necessary to promote a joint analysis approach that considers both the customer and the product and/or service in the foreground.

The main algorithms and models in terms of product, with a customer centric view, are:

Algorithms for defining associative rules	Recommendation algorithms that aim to understand the dynamics and predict purchasing choices based on association rules that emerge from purchases made by the customer. At the basis there is the concept of conditioning probability. Main examples: market basket analysis (MBA) and sequence analysis.
Price sensitivity models	Models that evaluate the elasticity of demand for a product/service as a function of the price change. Main examples: regression model, price ladder method, Van Wastendorp model, Gabor-Granger method.
Propensity to purchase models	Machine Learning supervised models that allow you to identify the probability that a customer purchases a certain product, product category or service.

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 14

Algorithms for defining rules associative

Pearson

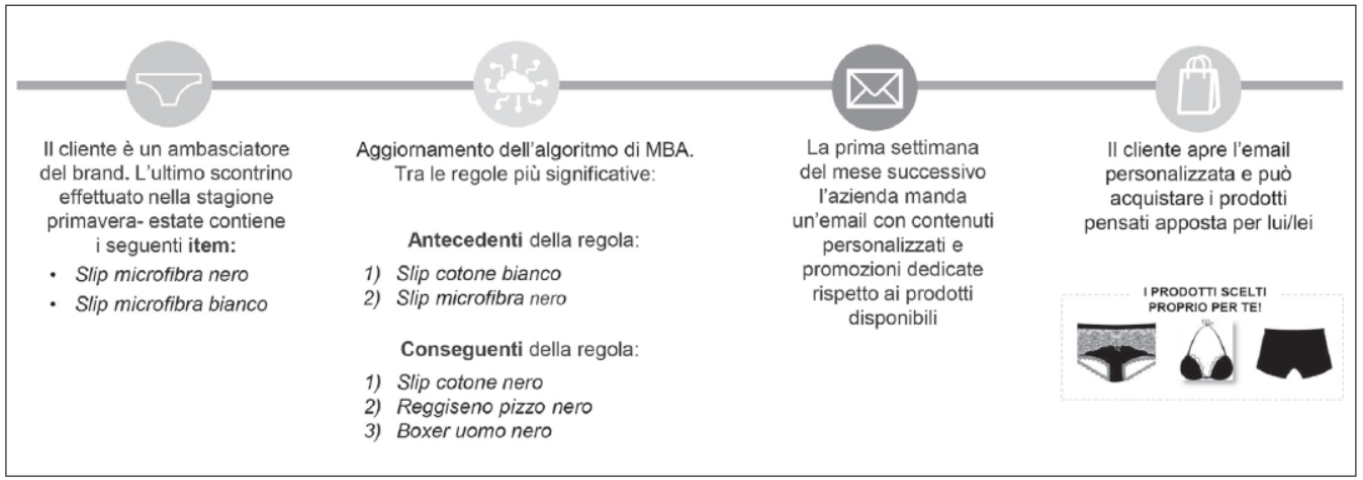
Algorithms for defining rules associative

Algorithms for defining associative rules between products or services make it possible to discover the hidden associations that exist between different sets of jointly purchased products and services.
In addition, they allow you to identify recurring behaviors among customers.

Associative rule: implication between two propositions of the type {X} => {Y}, such that if {X} is true then {Y} is also true.
We refer to probabilistic rules or rules that have a probability of occurrence p.

The MBA consists in analyzing the transactions made by customers to understand their purchase patterns by finding associations between the different products that the customer places in the cart.

Example of application in the marketing field: knowing the next purchase that the customer could make, a targeted discount is sent on a particular product to push him to prefer the company over its competitors.



Methodological approach

Key points of the methodological approach related to the association rules:

1. Consider a dataset D made up of n observations O, each of which made up of m binary variables indicating the presence or absence of object m within the single observation;
2. Define an associative rule as an implication X => Y such that X, Y $\subseteq$ 0 e X $\cap Y = \phi$, with X defined antecedent to the rule and Y consequent to the rule;
3. Definition of support and confidence of the associative rules: «support to the rule» is defined as the percentage of transitions within dataset D that contain both products X and Y; «confidence of the rule» is defined as the conditional probability p (X | Y), i.e. the percentage of observations containing X which also contains Y
4. Definition of business conditions that must not have too high a level of triviality and must be able to identify significant patterns
5. The choice of association rules is made by establishing a minimum support value and a minimum confidence value and a search is then conducted for the rules that exceed these criteria, called strong association rules. The search for strong rules is carried out in two steps:

    - Generation of frequent itemsets: means the generation of groups of products purchased together with a relative frequency higher than the minimum defined support. Apriori algorithms and FPGrowth algorithm apply;
    - Generation of strong rules starting from frequent itemsets

Another type of approach is represented by the sequence analysis: for the identification of the associations they also take into consideration the temporal order of the purchase, usually through the use of a timestamp or an incremental indicator.

Evaluation methods

Among the main evaluation methods we must consider those that are used for the construction of the rules themselves:

LIFT

Calculated as the confidence of the rule itself compared to the support of its consequent.
Lift values greater than unity indicate that the considered rule is more effective in predicting the relationship between different product patterns.
Conversely, lift values equal to unity indicate that the antecedent and consequent are independent.

$$Lift\ (X => Y) = \frac{p(X|Y)}{p(Y)}$$

LEVERAGE OF A RULE

Calculates the difference between the probability of co-occurrence of antecedent and consequent versus the probabilities in case of independence.
A Leverage value of 0 indicates independence, while a positive value indicates correlation.

$$Leverage\ (X => Y) = p(X,Y) - p(X)p(Y)$$

CONVICTION

It measures the product between the probability of the antecedent and the probability of lack of the consequent, compared to the probability of occurrence of the antecedent and consequent.

$$Conviction\ (X => Y) = p(X)p(\neg Y)p(X, \neg Y)$$

The main stages of development

The phases and sub-phases necessary for the definition of associative rules are shown below:

ESPLORAZIONE DATI E CREAZIONE DATAMART	APPLICAZIONE DELL'ALGORITMO	BUSINESS INSIGHT	INTEGRAZIONE NEI SISTEMI
• *Data quality* e *data exploration*. • Creazione del datamart di analisi.	• Scelta dell'algoritmo. • Applicazione dell'algoritmo. • Ottimizzazione e scelta dell'algoritmo e del set di parametri.	• Interpretazione dell'algoritmo di associazione: identificazione delle principali regole e validazione. • Presentazione dei principali risultati agli stakeholder.	• Automatizzazione dell'algoritmo. • Integrazione nei sistemi aziendali.

Benefits of algorithms

BENEFITS

- Observe, study and justify the correlations that lead to the purchase of certain baskets of goods, conditional on the previous purchase of others;
- Define, measure and monitor campaigns over time to increase both cross-selling and up-selling;
- A priori evaluate the introduction of a new product on the market;
- Design the best possible arrangement of the products on the shelf, in order to facilitate customer orientation within the store and the joint purchase of products;
- Analyze the consequences of removing obsolete or end-of-life products from the market.

Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 15

PRICE SENSITIVITY MODELS

Pearson

Price sensitivity models

Price sensitivity is defined as the analysis and measurement of how much changes in the price of a specific product or service affect the customers' willingness to purchase them, i.e. the quantities purchased.

$$\textit{Price elasticity of demand} = \frac{\%\ change\ in\ demand}{\%\ change\ in\ price}$$

Factors that influence a customer's price sensitivity are:

-- Type of product or service: products with least sensitivity are essential goods;
- Reference price: the competitive context and the availability of information in real time allow customers to quickly compare the prices offered by different companies;
- Uniqueness: customers are forced to suffer price increases if they do not have valid alternatives;
- Ease of exchange: the customer sometimes has difficulty evaluating the costs and services offered by competitors, or rather all the problems;
- Disposable income;
- Customer attitude

Methodological approach

Below are the most common methodological approaches to measure the price sensitivity of its customers:

MONADIC PROCE TESTING

It consists of an interview in which the customer is given various purchase price hypotheses for a specific good or service and the intention to purchase is recorded for each one.
It allows the company to quickly obtain initial feedback from current and potential customers.

CONJOINT ANALYSIS

Interviewed customers are shown a reduced set of alternative products and services and are asked to choose their favorite, without necessarily giving reasons.

LINEAR REGRESSION MODEL

It allows you to understand what is the percentage change in the quantities sold of a product/service with respect to a change in price. In this way it will be possible to identify elastic and inelastic products.
Despite the methodological and interpretative simplicity, they do not consider all the potential factors that can significantly impact the quantities sold of a specific product or service.

GABOR-GRANGER

Try to identify for each customer the maximum price they are willing to pay for the product or service offered. Gaber-Granger starts from the purchase response considering a randomly chosen price among those hypothesized and, depending on the response obtained, increases or decreases the proposed price until obtaining the highest price at which the customer is willing to buy. Once the highest price has been obtained, the intersection (optimum point) between the demand curve and the revenue curve is analysed.
The Gabor-Granger method is more suitable for products/services that have already been on the market for some time.
Its limitation lies in ignoring substitute products.

Methodological approach

VAN-WESTANDORP METHOD

Through a series of questions posed to the interviewees following a detailed description of the product offered, some key "psychological" prices that form in the minds of potential customers are identified.

Four basic questions are asked:

1. What price would be so low as to question the quality of the service?
2. At what price does the product begin to be considered a bargain?
3. At what price does the product start to seem expensive?
4. At what price is the product overpriced?

And the cumulative frequency distributions of the various responses are plotted and identified by their intersections:

- Marginal cost point (PME): intersection of "too expensive" and "cheap"
- Point of Marginal Economy (MPC): Intersection between "too cheap" and "expensive"
- Optimal price point (OPP): intersection of «too expensive» and «too cheap»
- Indifference price point (PPI): intersection of "expensive" and "cheap"

The acceptable price range is between the PMC and the PME.

The Van-Westandorp method is suitable for bringing new products to the market. The limitations of this method concern: (a) the assumption that customers know the product offered; (b) does not guarantee the intersection of the curve at the optimal point; (c) does not take into consideration the context of the purchase and the competitor.

NEUROMARKETING PRICING

Studies using fMRI have shown that identifying a price that is not appropriate to customer expectations leads to a brain response that approaches that of physical pain.

Research has highlighted how a preference for a product involves the activation of brain areas connected to the sensation of pleasure, while viewing prices deemed too high involves the activation of brain areas connected to the sensation of pain.

The emotional dynamics connected to pricing do not only concern the characteristics of the product and the brand, but also the context in which the purchase is made.

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 16

Propensity to purchase a product or a category of products

Pearson

Propensity to purchase

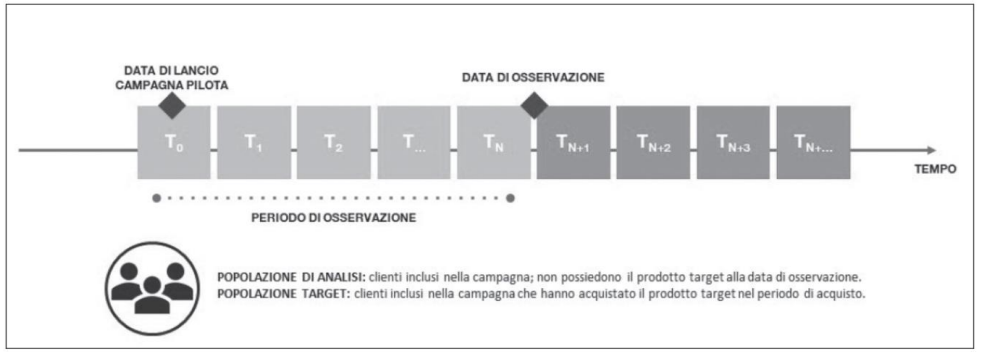
Propensity to purchase models make it possible to define which customers are most likely to purchase a specific product and on which to activate specific marketing campaigns, with the aim of increasing the redemption rate and cross-selling rates and upselling.

Cross-selling and up-selling campaigns aim to sell additional products to existing customers.

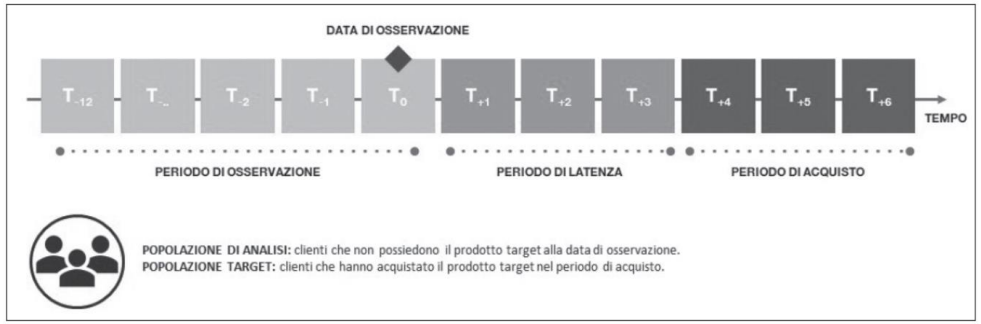
Through predictive models it is possible to estimate the probability of a purchase based on the patterns most associated with the acceptance of the offer.

The approaches used to identify probable buyers are as follows:

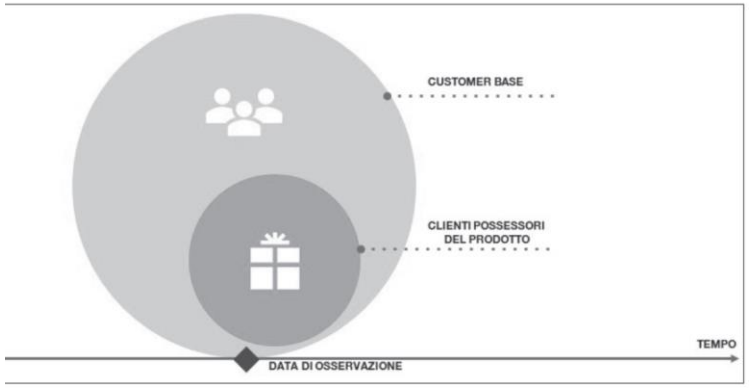
PILOT CAMPAIGN



PRODUCT UPTAKE



PROFILING OF ALREADY OWNERS



Methodological approach

Depending on the situation in which the company finds itself, the needs and availability of data, one of the proposed approaches will be preferred.

It starts from the definition and construction of the target variable and will continue with the selection of customers divided into the two training and validation samples.

It is important to dwell on how to use the main insights coming from the model for setting up personalized marketing campaigns
- *Definition of purchase probability classes:*
  - Up to the first tertile: customers with a low probability of buying
  - Between the first and second tertiles: customers with an average purchase probability
  - Beyond the second tertile: customers with a high probability of buying

If the customers have been correctly classified, attributing the right propensity score to each one, the redemption following the campaign will be significantly different in the three classes: it will have to increase as the assigned probability increases.
The company will be able to send different communications to each probability class, leaving aside a percentage between 5% and 10% of customers with the same characteristics as the target customers which will be used as a stratified control group for the main variables of interest .

Evolutions of the propensity model to purchase a product

Bid optimization is needed when customers are eligible for more than one product and the company needs to choose the best product to offer them.

It translates into the use of results from descriptive analyzes and propensity models.
The widely used strategy is the Next Best Offer (NBO) which should increase the customer's CLTV.

This approach requires the development of specific cross-selling and up-selling models, and purchase probability estimates are combined with the expected Net Present Value (NPV) for each product to estimate the propensity value of each offer (VPO).

VPO = Pr(product purchase) X NPV – (C_i)

C_i = cost of the offer = sum of the costs related to the promotion of the product
NPV = sum of NPV over n years divided by the number of initial customers

The best offer for each customer is the one that maximizes the estimated propensity value.

Customer ID	Prodotto proposto (da algoritmo MBA)	Propensione all'acquisto del prodotto	NPV previsto	Costo offerta	Valore propensione offerta	Offerta migliore (NBO)
1	Prodotto 1	0,06	1.000 €	2 €	58 €	
1	**Prodotto 2**	**0,14**	**800 €**	**2 €**	**110 €**	**X**
1	Prodotto 3	0,16	700 €	4 €	108€	
2	**Prodotto 1**	**0,12**	**1.000 €**	**2 €**	**118 €**	**X**
2	**Prodotto 2**	**0,14**	**800 €**	**2 €**	**110 €**	**X**
3	Prodotto 1	0,08	1.000 €	2 €	78 €	
3	**Prodotto 3**	**0,17**	**700 €**	**4 €**	**115 €**	**X**

Esempio di approccio della NBO

Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 17

THE OPTIMIZATION OF MARKETING CHANNELS AND COMMUNICATIONS

Pearson

Recognition patterns of objects within images

Computer Vision is also at the center of digital transformation: through advanced neural networks it is possible to recognize objects within images, simulating the cognitive behavior of the human brain, making image recognition and image detection processes similar to human ones.
Through CNN (convolutional neural networks), the algorithms for reading and interpreting images are able to recognize shapes and colors and follow moving objects.

Examples of application of image recognition are quality control (after the event) or in retail from a predictive perspective.

A further approach is social analytics: digital monitoring activities to collect and analyze the metrics of all social media on the Web.
Social analytics are developed in three phases:
- Identification of available data flows and data collection, considering the time frame, type of content, source and owner of the data;
- Descriptive analysis and development of Artificial Intelligence models (main tools: PostPickr, Hootsuite, Brandwatch, Spinklr);
- Interpretation of results and insights, specifically conversion rate, engagement score, brand reputation.

Another data source comes from neuromarketing research, whose main objectives are:
- Determine the most effective graphic elements in capturing the customer's attention;
- Analyze the emotional involvement, attention and cognitive effort recorded by the customer during the interaction with the brand or with the product/service;
- Identify the elements of a video or post that generated the most emotional engagement;
- Compare the effectiveness of two or more posts/videos and choose the one that generates the most engagement.

NPL e Sentiment Analysis

NPL

Natural language processing: Process of automatically processing information written or spoken in a given natural language by means of an electronic computer.
The data is processed automatically in order to obtain concise, useful and structured information.

SENTIMENT ANALYSIS

Application of the NPL. Speech processing software collects data from different platforms by taking into consideration portions of text and determining the tone, the intensity of opinion and the emotionality with which it is expressed.

Main applications:
- Understand what people think about a brand or product/service or marketing campaign;
- Track mentions online in real time;
- Track competitor mentions for benchmarking;
- Enrich your SWOT Analysis;
- Understand the brand's web reputation (meaning concrete actions that are reported in specific documents)

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 18

Introduction to computer vision

Pearson

Introduction to computer vision

Computer Vision consists of various techniques both classic and based on artificial intelligence for image processing and information extraction from them.

The main source of the digital imaging process is the electromagnetic spectrum (followed by ultrasound acoustic waves and electron beams).

SI defines a digital image as a matrix of values in which each pixel takes on a discrete value, which represents its intensity within a scale of values which can be two-dimensional (gray scale between black and white) or with three different RGB channels , obtaining a color image.

The main file formats in which the image is saved, also defining any type of compression, are jpeg, tiff and png. DICOM is the way images are saved along with their metadata.

The images are represented:
- In the spatial domain: it involves the discretization of a matrix in which the pixels represent its intensity. It is used to correct the appearance or extract information of interest. The underlying mathematical operation is convolution. An example of an approach to image processing is the Fourier transform;
- In the frequency domain: the wavelet transformation technique is used, i.e. the signal is broken down on the basis of small waves of different frequencies and time durations. The advantage is that you can process the same image at different resolutions.

Introduction to computer vision

Convolutional neural networks (CNNs) make image processing at different scales conditioned by filters defined in a data-driven way according to specific tasks.

The filters that make up the network structure have kernel values initialized randomly and modified during the training phase.

Main convolutional networks used over time: LeNet5, MNIST, CNN, VGGNet, GoogLeNet and ResNet.

The concept of autoencoder is then introduced in convolutional networks consisting of an encoder (compression channel) and a decoder (reconstruction channel).

A new type of architecture is subsequently developed, the GAN architecture, which is made up of a part defined as a generator, with the task of producing data as close as possible to the real data on which it was trained, and a second part defined as a descriminator, whose task is to make people understand how the generated data differ from the real ones.

The most recent architecture is the transformer which, through the "attention" mechanism, allows the model to weigh each part of the input image differently.

Application of Computer Vision

The main technological applications of Computer Vision are shown below:

- Classification of images by assigning them a specific label. It refers to images in which only one object appears and is analysed;
- Semantic segmentation: grouping the parts of an image that belong to the same object class. This is a form of pixel-level prediction;
- Object detection: classification and localization used to analyze the most realistic cases where there may be multiple objects in an image;
- Restoration of images in order to obtain a high quality image from a damaged input image;
- Detection of facial landmarks to detect the key points of reference on the times and follow them;
- Object Tracking: Take an initial set of object detections, create a unique ID for each of the initial detections, and then follow the objects as they move through the frames of a video;
- Image retrieval systems: aim to find images similar to a given image among a lot of image data.

Software and insights

Below are the main Deep Learning Python libraries for image processing:

PYTORCH	It allows you to develop models in Deep Learning. It allows you to add blocks in succession to obtain customized networks, or download the frameworks of networks that are already more consolidated through benchmarks on different datasets.
TORCHVISION	It is used for the manipulation of tensors, from resizing to cropping etc
FASTAI	It is a Pytorch wrapper that allows it to be used at a higher level but with a lower degree of customization
KERAS	It is a TensorFlow wrapper that allows for higher level programming
PYTORCH	It is the alternative to PyTorch with a different Python function dialect

Image processing libraries

The main Python libraries for building neural networks for image processing are listed below:

OPENCV	It is a C++ wrapper that allows image processing, from converting between different color spaces to cleaning operations
PILLOW	Image viewing library
SCIKIT-IMAGE	Similar to OpenCV, easier to use
ALBUMENTATION	It allows you to perform advanced operations on datasets for models based on neural networks. Provides features to increase image quality
MATPLOTLIB	It is used to render mathematical functions and digital images in graphic form

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 19

Language processing natural

Pearson

Processing techniques in text processing

The main terms and phases of the processing techniques used on the raw data before the application of the model in the context of textual processing are shown below:

- Corpus: collection of textual starting documents (emails, messages, comments);
- Terms: keywords that convey the information;
- Index: list of terms that appears only once;
- Reverse index: list of documents in which each term appears, identified for each term;
- Tokenization: method used to divide texts or sentences into tokens (collections of words) that carry useful information, removing characters that do not carry useful information;
- Stop words: meaningless words eliminated through a token filter operation;
- Stemming: identification of the root in compound words or conjugated verbs

Esempio 1. Mostriamo di seguito il codice per l'estrazione dei token con la libreria PySpark a partire da un *corpus* che supponiamo essere contenuto nella variabile `df`, che supponiamo essere un oggetto di tipo `DataFrame` della libreria PySpark.

```
1 from    pyspark.ml.feature   import   Tokenizer
2 tokenization=Tokenizer(inputCol=' review', outputCol='tokens')
3 tokenized_df=tokenization.transform(df)
```

In PySpark, si può utilizzare StopWordsRemover per la rimozione delle *stop words*.

```
1 from  pyspark.ml.feature import StopWordsRemover
2 stopword_removal=Stop WordsRemover(inputCol='tokens', outputCol='
      refined_tokens')
```

Si può costruire un DataFrame con i token rimasti, che costituiscono l'indice, attraverso il codice seguente.

```
1  refined_df=stopword_removal.transform(tokenized_df)
```

Esempio 2. Mostriamo di seguito il codice per l'estrazione dei token con la libreria NLTK a partire da un testo che supponiamo essere contenuto nella variabile `sample_text`, che supponiamo essere una stringa.

```
1 from nltk.tokenize import word_tokenize, sent_tokenize
2 sample_word_tokens = word_tokenize(sample_text)
3 sample_sent_tokens = sent_tokenize(sample_text)
```

In NLTK è disponibile una lista di *stop words* per la lingua inglese, che va però integrata con le varianti che inizialo per lettera maiuscola.

```
1 from nltk.corpus import stopwords
2 stop_words = [word.upper() for word in stopwords.words('english' )]
3 word_tokens = [ word for word in sample_word_tokens if word.upper()
      not in stop_words]
```

Il testo può poi essere filtrato tramite le tecniche di elaborazione delle liste Python. Di seguito è mostrato un esempio di come sia possibile implementare il modello *Bag of Words* (BoW).

```
1 import re
2 from nltk import collections
3 def bag_of_words (text):
4      _bag_of_words = [collections.Counter(re.findall(r'\w+',word))
       for word in text]
5      bag_of_words = sum (_bag_of_words.collections.Counter())
6      return bag_of_words
7 sample_word_tokens_bow = bag_of_words(text=sample_word_tokens)
```

Space-vector model

The key-tokens identified are to be translated into numerical language in order to apply an algorithm.
Through the word embeddings strategies, the terms of the index are mapped in a vector space: each term is associated with a non-zero multiple of that vector.
The terms are pairwise orthogonal.

The mathematical model of a text is a term-document matrix: a document is obtained as a linear combination with positive coefficients of the vectors corresponding to the terms appearing in it. Each document is associated with a vector of length equal to the length of the index.

Different approaches:
- Bag of Words: the presence of a term is indicated by 1 and does not take into account the frequency of the term (one-hot representation);
- Count Victorizer: Assign weight f_{ij} with i = number of occurrences of the term and j = document
- *TF-IDF*: exploits the concept of reverse index → $a_{ij} = f_{ij} \log(n/n_i)$

Esempio 3. In PySpark si può utilizzare `CountVectorizer` per contare le occorrenze dei termini dell'indice `refined_tokens` all'interno del *corpus*. I risultati vengono salvati nella nuova colonna `features` del dataframe.

```
1 from pyspark.ml.feature import CountVectorizer
2 count_vec=CountVectorizer(inputCol='refined_tokens', outputCol='features')
3 cv_df=count_vec.fit(refined_df).transform(refined_df)
```

Si può accedere all'indice attraverso il codice seguente.

```
1 count_vec.fit(refined_df).vocabulary
```

Utilizzando invece il modello TF-IDF, è possibile creare un dataframe a partire dagli stessi indici come segue.

```
1 from pyspark.ml.feature import HashingTF, IDF
2 hashing_vec=HashingTF(inputCol='refined_tokens', outputCol='tf_features')
3 hashing_df=hashing_vec.transform(refined_df)
4 tf_idf_vec=IDF(inputCol='tf_features', outputCol='tf_idf_features')
5 tf_idf_df=tf_idf_vec.fit(hashing_df).transform(hashing_df)
```

Co-occurrence matrix and similarity measures

The alternative to the term-document matrix is the symmetric co-occurrence matrix: it numerically represents the corpus through an m x m matrix, where m is the number of terms in the index.

It is a symmetric matrix: it means that $a_{ij} = a_{ji}$ where the element riamo a_{ij} indicates how many times the term i appears near the term j and k words are counted before and after the term in question.

To define the similarity between the two texts, the cosine similarity is applied, calculating a distance between the corresponding vectors

$$d_i^T \, d_j \, || \, d_i \, || \, {}_2^{-1} || \, d_j \, || \, {}_2^{-1}$$

Other measures used are the Jaccard or Manhattan distance.

Matrix methods

Here are some feature selection techniques:

TRUNCATED SVD

Singular value decomposition of the term-document matrix A

$$A_K := u_K \sum_k V^T, \; || A - A_k || = min_{rank(B)=k} || A - B ||$$

FATTORIZZAZIONE NON NEGATIVA

Factorization of the term-document matrix A into non-negative factors of rank k, i.e

$$A \approx WH, W \geq 0, H \geq 0$$

where W and H have dimensions m x k and k x n.
The non-negativity of the coefficients favors the interpretability of the model.

Esempio 4. Mostriamo com'è possibile calcolare la fattorizzazione SVD troncata di una matrice di input A con scikit-learn.

```
1 from sklearn.decomposition import NMF
2 model TruncatedSVD(n_components=5)
3 Ak = model.fit(A)
```

La fattorizzazione NMF invece può essere ottenuta come segue.

```
1 from sklearn.decomposition import NMF
2 model = NMF(n_components=2)
3 W = model.fit_transform(A)
4 H = model.components_
```

Qui *n_components* indica il numero di *features*, precedentemente indicato con *k*.

Methods based on neural networks

- Techniques based on text representation through projections on vector spaces do not reproduce a correct semantic and contextual understanding.
- Sequence embedding models, based on neural networks, try to learn the representation of a text one step at a time and can be used to predict the probability of a word in a given context.

WORD2VEC

Set of two templates used to produce word embeddings. CBOW is the first approach aimed at preventing the probability of a data taken $w^{((c)}$)
Starting from a context aprole window. The Skip-Gram architecture has the purpose of predicting in output the context words $w^{(c-k)}$ starting from a given term $w^{(c)}$ in input

GLOVE

Weighted least squares model that is trained on global co-occurrence between terms. We define the probability that word j is in the context of word i as P_{ij} := P(w_j| w_i) = X_{ij} / $\sum_k X_{ik}$
The model minimizes the following quadratic objective function : $\frac{1}{2}\sum_{i,j=1}^{m} f\ (P_{ij})\ (v_i^T\ v_j)$ - $\log(P_{ij})^2$

RECURRING NEURAL NETWORKS

Production of a model that considers all the words of the corpus. In the RNN architecture there are multiple hidden neural network layers, each of which has a certain number of neurons and performs matrix-vector multiplication on all inputs followed by a nonlinear operation. The cost function typically used to train the architecture is a cross-entropy function. The most common is the LSTM.

TRANSFORMER

GPT and BERT. Combining the transformer architecture with LSTM pre-training of the language model removed the need to train architectures from scratch for a specific task, satisfying almost any NLP benchmark, with significant margin of error.

Chatbot

Chatbots are prime examples of applications of natural language processing in digital marketing.

They are distinguished by type of intelligence realized in:
- Command-based chatbots: have a limited set of commands, use structured message prompts, and do not understand natural language. Database and rules of conduct are defined ex ante;
- AI-based chatbots: they have linguistic and dialogue management capabilities. Based on semantic engines or third-party AI platforms and linked to the evolution of Machine Learning.

They are distinguished according to the model of response to the questions of the human interlocutor in:
- Retrieval-based chatbot: Access a repository of predefined responses;
- Generative chatbot: translates from an input (question) to an output (answer) by adopting Deep Learning techniques.

They are distinguished according to the domain extension in:
- Open domain: you can ask anything on any topic;
- Closed domain: only specific questions relating to certain topics can be addressed.

Main functions:
- Customer care
- Entertainment
- Increase reach
- Offer something more
- Sell
- Increase productivity
- Get to know your customers better
- Recruiting
- Sentiment Analysis

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 20

Main approaches to the measurement of marketing activities

Pearson

Introduction and main methodological approaches

Often companies rely on correct metrics to measure their success, without assessing whether these metrics accurately reflect the economic value generated by specific marketing activities.

Understanding the effect of each marketing activity on brand perception, incremental sales and customer base growth is essential for proper corporate budget allocation, management and planning.

The three main methodological approaches for measuring marketing activities are:

- Main metrics for objective measurement of marketing activities: reference is made to the main evaluation KPIs;
- Attribution models: methodologies that allow the results of marketing activities to be attributed to specific touchpints with the customer, to identify which of these have had the greatest influence on the purchase path;
- Marketing Mix Models: Methods and Approaches for Developing Marketing Mix Modeling (MMM)

Advanced Analytics e Artificial Intelligence per il marketing

Chapter 21

Methods for measuring marketing activities

Pearson

Introduction and main metrics

The objective of each marketing activity is to carry out a specific action in order to impact their purchasing behavior, inducing a change in order to achieve certain business objectives.

The main objective of a company is to obtain an increase in the margin compared to the baseline (incremental margin) through the increase in revenues, therefore the increase in the conversion rate on the subscription of the services or on the purchase of the products, or through the decrease costs by reducing variable costs.

The main indicators that evaluate an incremental margin increase are ROI and time to break-even (payback time).

In order to have an a priori assessment of the impact of the initiative, an ex-ante business case is drawn up:
- The customer or prospect target is indicated;
- The engagement mechanic is identified;
- The unique selling preposition is defined.

The definition of these drivers is based on the analysis of the customer base and makes it possible to estimate the costs to be incurred and the expected revenues.

Ex-post it is essential to periodically monitor the progress of the initiative to verify that the assumptions made satisfy the expected results.

Evaluation methodologies

It is essential for the company to monitor the changes in its turnover considering:

- ROI: it must be positive, i.e. cover or, better, exceed the investment costs;
- Cost per sale: evaluate the cost of transformation from prospect to customer;
- Cost per lead: evaluate the conversion cost from user to prospect;
- Conversion rate for inbound channels: ratio between who fills out the contact form and who enters the company platform, newsletter subscriptions, page effectiveness;
- ADS: FacebookAds and GoogleAdWords;
- SEO: site indexing strategy with related keyword research;
- Brand awareness growth: by completing questionnaires;
- Construction of the control group: benchmark group with respect to the target;
- A/B Testing: evaluation of the effectiveness of two versions of campaigns

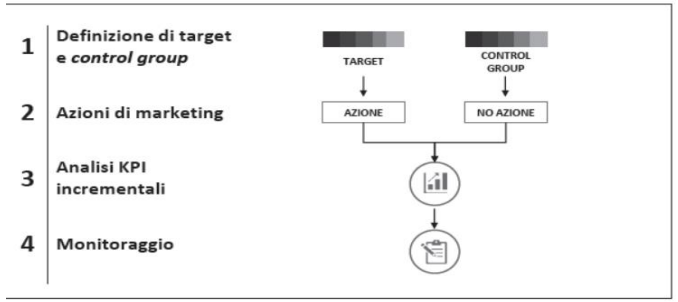
Control Group (CG) and A/B Testing

CONTROL GROUP

Set of customers similar to each other whose behavior is observed in the same conditions as the target without receiving the marketing action. The two groups are homogeneous: through a representative sampling stratification, they maintain the same proportions of the reference population. The study of differences in behavior reveals insights into the effectiveness of marketing activity.

It is also possible to proceed with the construction of a historical control group, made up of people who have participated in the past in the same exhalation by comparing them with the current target. It is less accurate as it is less representative of the population, not taking into account the moment in time and the context in which the action is performed.

Among the main additional sampling methods we find: simple random sampling and stratified sampling.



A/B TESTING

Different variants of marketing actions are proposed to the target, dividing it into two homogeneous groups, and the main KPIs are verified to identify the most effective variant.

It mainly applies to creativity, graphics, informative texts, promotional offers and calls-to-action.

The MBA methodology allows you to use an algorithm that dynamically allocates traffic between different versions (example of an online site).

Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 22

DIGITAL MARKETING ATTRIBUTION MODELS

Pearson

Introduction and goals

Attribution models allow you to measure the value of campaigns and understand which touchpoints have the greatest influence on the conversion journey of buyers.

Key concepts:

- Conversion: desired result/goal from the web or store;
- Conversion value: total cost of the action taken by the company;
- Touchpoints: touchpoints that the customer experiences in the purchasing process;
- Contributions: the different moments that contribute to a purchasing decision

The goals of the attribution model are:

- The maximization of volumes;
- Most budget efficient spending

In addition to the choice of the algorithm, it is important to understand the influence of the key characteristics of the path, such as:

- Length of user path: number of touchpoints;
- Information recency: location of touchpoints along the user journey (relative to the point of conversion);-
- Touchpoint granularity

Attribution models based on: rules, data and regression

RULES

Rule-based attribution is the "simple" approach that assigns values to each touchpoint based on a predefined formula. Are used:
- single-touch attribution: assigns 100% credit to only one touchpoint (usually the first)
- Linear attribution: assign equal credit to each touchpoint;
- Custom Attribution: Assign arbitrary credit to each touchpoint

DATA

Data-driven attribution uses an algorithm that leverages historical data to decide the value of a touchpoint within a given customer journey.
It is based on the estimated probabilities of conversion during the various touchpoints.
The algorithms used to calculate these estimates can include different approaches: Shapley's value, Markov models, other regression models.

REGRESSION

Attribution can also be modeled as a classification problem: each customer journey makes an observation with the binary conversion indicator y_i as the dependent variable and the logistic regression can be formulated as:

$$y_i = \begin{cases} 1, conversione \\ 0, altrimenti \end{cases} \qquad P(y_i = 1) = \Lambda(\beta_0 + \sum_{k=1}^{K} \beta_k C_{i,k})$$

Markov model and Shapley value

MARKOV MODEL

Probabilistic model that represents customer journeys as a graph whose nodes are the touchpoints or «layers» and the connecting edges are the observed transitions between the layers.
Through this model it is possible to:
Obtain the probability of success of a given customer journey given the history of all customer journeys;
Gauge the importance of each campaign by calculating what is known as the takedown effect.
Attribution value: $A = V \cdot (R_t / R_v)$

SHAPLEY VALUE

Shapley's value approach aims to fairly allocate each touchpoint contribution to the conversion through the following rules:

1. Marginal contribution;
2. Interchangeable players have the same value;
3. Fictitious players have zero value;
4. If a game has multiple parts, the cost must be broken down between those parts

The channel coalition does not take into account the sequential effect of the channels.
The results are built using existing, non-simulated customer journeys.

KPI per branding and performance

KPI di *branding*	KPI di *performance*
Brand awareness (per esempio, il logo dovrebbe essere riconosciuto dal 99% del pubblico di destinazione)	*Lead* (per esempio, 500 *lead*)
Conoscenza del marchio (per esempio, il 90% del pubblico di destinazione dovrebbe sapere che il prodotto è sano)	Transazioni (per esempio, 5.000 transazioni)
Emozioni del marchio (per esempio, il marchio dovrebbe evocare sentimenti positivi con il 70% del pubblico di destinazione)	Entrate (per esempio, €5.000.000 di entrate)
Posizionamento del marchio (per esempio, l'80% del pubblico di destinazione dovrebbe valutare il marchio come premium e innovativo)	
Atteggiamento del marchio (per esempio, il 70% del pubblico di destinazione dovrebbe avere un atteggiamento positivo nei confronti del marchio)	
Relazione con il marchio (per esempio, il 30% del pubblico di destinazione ammette di essere impegnato nel marchio)	

The main stages of development

The phases and sub-phases necessary for the definition of an attribution model are shown below

DATAMART DI ANALISI	SVILUPPO DEL MODELLO	BUSINESS INSIGHT	INTEGRAZIONE NEI SISTEMI
• *Data understanding* e *data quality*. • Valutazione dei touchpoint del *customer journey* e *data collection*. • Definizione del periodo temporale. • Creazione dei datamart di analisi e delle potenziali variabili con vista *customer centric*.	• Selezione e trasformazione delle variabili. • Analisi preliminare per determinare il set di modelli di attribuzione. • Sviluppo dei modelli di attribuzione. • Ottimizzazione e scelta del modello migliore.	• Interpretazione del modello. • Presentazione dei risultati.	• Automatizzazione del modello di attribuzione. • Integrazione del modello di attribuzione e creazione di un sistema di dashboard per il monitoraggio.

Advanced Analytics e Artificial Intelligence per il marketing

CHAPTER 23

MARKETING MIX MODELS

Pearson

Methodological approach

Marketing Mix Modeling (MMM) arises from the 4P marketing mix concept: it consists of statistical models built to be able to measure the effectiveness of communication spending using historical data in aggregate form, with the aim of analyzing sales trends in the time, according to advertising spending variables and in the various marketing channels.

The introduction of the digital channel has allowed the birth of digital attribution.

Data collection is necessary for the definition of the statistical methodology. Data is grouped by variables:

- target: the main target variable concerns sales, divided according to sales channels. Other types of target variables can be: purchase conversions, number of new customers in a period of time, volume of traffic on the site and subscriptions in the stores;
- Independent: a) average metric, for example: average investment, impressions, clicks, interections, gross rating point (GRP), share of voice (SOV);
  b) marketing: calendar of promotions or discounts; historical price trend per product; variable related to the introduction of new product lines; store opening/closing information; customer service KPIs;
  c) control: seasonality; weather forecast; competition; brand awareness etc

The general statistical representation of the MMM is represented as:: $Y = \alpha + \beta_i X_i + \varepsilon$

Response pattern to advertising

CURRENT EFFECT	Variation in sales caused by advertising exposure that occurs in the same period
CURRYOVER EFFECT	Part of the effect that occurs in periods of time following the launch of the advertisement
SHAPE EFFECT	Change in sales compared to the increase in investment in advertising in the same period
COMPETITIVE EFFECT	Reactivity of the brand in adv compared to the actions taken by competitors
DYNAMIC EFFECT	Variations of the effects of advertising over time
CONTENT EFFECT	Changes in sales in response to a change in ad content or creativity
MEDIA EFFECT	Changes in sales in response to promotional strategies of different types of discounts on different channels. Halo effect: customer approaching a product thanks to the positive experiences of purchasing products of the same brand; Cannibilization effect: negative impact on a product due to negative performance of other products of the same brand

Implementation of the econometric model: linear regression model

Linear regression models consider sales or the number of contracts signed as the dependent variable and all marketing activities and other external control factors among the independent variables.

A linear regression model is represented by the following formula:

$$Y_t = \beta_0 + \sum_{n=1}^{m} \beta_j \, x_{jt} + \varepsilon_t \, , \, \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$

Y_t = dependent variable (sales) at time t
x_{jt} = independent variables
m =number of marketing variables
β_0 = *MMM* baseline
β_j = coefficient for the marketing variable
ε_t = error

The main limitation concerns the only possibility to estimate the current effect.

It is an adequate model for a company with a stationary product, operating in a mature market, with moderate seasonality and made up of several competitors, in which it is assumed that each mic lever contributes to generating volumes independently of the others.

Implementation of the econometric model: multiplicative regression model

The multiplicative regression model considers a multiplicative relationship between the independent variables of the MMM. Describes the effect of m possible variables on sales.

A multiplicative regression model is represented by the following formula:

$$Y_t = \exp(\beta_0) + \prod_{n=1}^{m} \beta_j \, x_{jt} + \varepsilon_t$$

The logarithmic transformation of the target variable linearizes the model, making it possible to estimate as a simple additive model:

$$\text{Log}\,(Y_t\,) = \beta_0 + \sum_{n=1}^{m} \beta_j \log(\,x_{jt}) + \varepsilon_t \qquad , \varepsilon_t \sim \mathrm{N}(\mathrm{o}, \sigma^{\,2})$$

The three main benefits:

- The dependent variable undergoes the effect of the interaction of all the marketing variables considered;

- The composition of the equation manages to include a variety of forms of the response variable, making the model flexible and able to estimate relationships appropriate to them;

- The coefficient both estimates the effects of the independent variable on the dependent variable and represents a sales elasticity value with respect to advertising

The main limitation concerns the possibility to estimate current effect and shape effect.

Implementation of the econometric model: non-linear and hierarchical models

Non-linear and hierarchical models allow to introduce the estimation of two other effects: the saturation of the adv (S-shaped effect) and the carryover effect.

For the curryover effect we proceed by transforming the time series of the media variables through the adstock function:

$$adstock\ (x_{t-L+1,m}\ , \dots , x_{t,m}\ ;\ w_{m,L}\ ) = \frac{\sum_{l=0}^{L-1} w_m(l) x_{t-1,m}}{\sum_{l=0}^{L-1} w_m(l)}$$

w_m is a nonnegative function of weights
L is the maximum duration value of the carryover effect

Geometric decay function for the weight function: $w_m^g(l; \alpha_m) = \alpha_m^l$, l = 0, ..., L-1 , $0 < \alpha_m < 1$
α_m : adv effect retention rate

Adstock assumption assumes that the highest values occur in the same exposure period as the ad.
Some means to obtain an effect of the ADV require more time and therefore higher values may not coincide with the exposure period: therefore the delayed adstock function is introduced which also introduces control on the delay effect:

$$w_m^g(l; \alpha_m, \theta_m) = \alpha_m^{(l-\theta_m)^2} \quad , l = 0, \dots, L\text{-}1 \quad , \ 0 < \alpha_m < 1 \quad , \ 0 < \theta_m < L\text{-}1$$

θ_m major effect delay

Implementation of the econometric model: non-linear and hierarchical models

Non-linear and hierarchical models allow to introduce the estimation of two other effects: the saturation of the adv (S-shaped effect) and the carryover effect.

For the S-shaped effect we proceed by transforming the time series of the media variables through the Hill function:

$$Hill\ (x_{t,m}\ , K_m\ ;\ S_m\ ) = \frac{1}{1+(\frac{x_{t,m}}{K_m})^{-S_m}} \quad , x_{t,m} \geq 0$$

$S_m > 0$ shape parameter referred to the slope
K_m > 0 coincides with the saturation point

To allow for different effects for different media, a multiplicative coefficient is applied to the Hill function β:

$$\beta_m Hill_m(x_{t,m}) = \beta_m \frac{K_m^{S_m}\beta_m}{x_m^{S_m}+K_m^{S_m}}$$

The combination of the two separate effects introduces hierarchical models: the adstock function is applied first, then passing to the application of the Hill function or inverting the order of the transformations:

$$Y_t = \tau + \sum_{m=1}^{M} \beta_m Hill_m\ (x_{t,m}^*, K_m\ ;\ S_m\ ) + \sum_{c=1}^{C} \gamma_c z_{t,c} + \varepsilon_t$$

$x_{t,m}^*$ represents the adstock function
τ the sales baseline
γ_c the effects of the control variables z
ε_t error

Methods for estimating the parameters of the MMM and limitations

MAXIMUM LIKELIHOOD ESTIMATION

It consists in maximizing the logarithm function of the likelihood; the estimated parameters will represent the result of this transformation.
It is therefore a matter of solving a constrained nonlinear optimization problem.

BAYES THEOREM

The Bayesian method allows to introduce a priori knowledge for the estimation of the model.

With this approach you get:
- Samples of later distributions
- A credible range in which the unknown parameter can be included

The main issues that can question the reliability of the results of the Marketing Mix models are:

- Dataset size
- Limited range of data
- Selection bias:
  - Wrong choice of target audience
  - Seasonality
  - Funnel effect due to a distortion caused by the interference of the impact of an advertising channel with others inserted for the same marketing campaign

The main stages of development

The phases and sub-phases necessary for the definition of a Marketing Mix model are shown below:

ESPLORAZIONE DATI	SVILUPPO DEL MODELLO	BUSINESS INSIGHTS	INTEGRAZIONE NEI SISTEMI
• *Data collection*. • *Data understanding*. • *Data preparation* e *data cleaning*. • *Data exploration*. • Definizione del periodo temporale e delle potenziali variabili. • Costruzione della *customer table*.	• Selezione e trasformazione delle variabili di input. • Analisi di correlazione. • Test statistici. • Partizione dei dati tramite split *train-test* o *cross-validation*. • Sviluppo dei modelli di Marketing Mix. • Ottimizzazione e scelta del modello migliore.	• Interpretazione del modello e calcolo del ROI di campagna. • Presentazione dei risultati.	• Automatizzazione e integrazione del modello nei sistemi aziendali e nel processo di *customer engagement*.