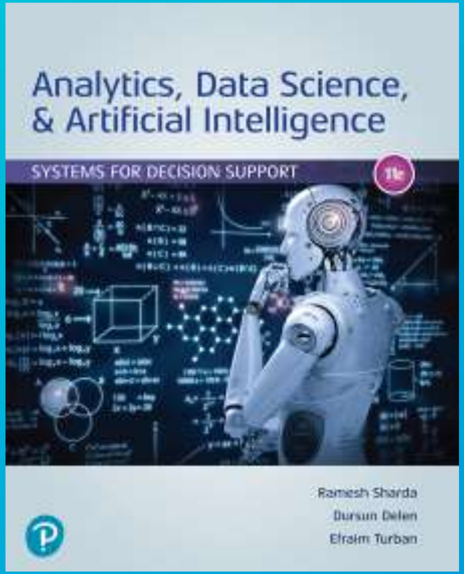
الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

College of Computing and Informatics

Week 12

Chapter 11:Group Decision Making, Collaborative Systems, and AI Support



Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Understand the basic concepts and processes of group work, communication, and collaboration
2. Describe how computer systems facilitate team communication and collaboration in an enterprise
3. Explain the concepts and importance of the time/place framework
4. Explain the underlying principles and capabilities of groupware, such as group support systems (GSS)
5. Understand how the Web enables collaborative computing and group support of virtual meetings
6. Describe collective intelligence and its role in decision making
7. Define crowdsourcing and explain how it supports decision making and problem solving
8. Describe the role of AI in supporting collaboration, group work, and decision making, and human–machine collaboration

Required Reading

- **Chapter 11**: “*Group Decision Making, Collaborative Systems, and AI Support* ” from “*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”.*

Recommended Video

- OnlineTutorial| Group Decision Support System (2020, April 23). [Video]. YouTube. https://www.youtube.com/watch?v=8qA15NF7rM8

11.2 Making Decisions in Groups: Characteristics, Processes, Benefits, and Dysfunctions

- Making Decisions in Groups
- Group Decision-Making Process
- Benefits and Limitations of Group Work

Making Decisions in Groups

- **Group work -** work done by two or more people together

- **Group decision making –** a situation in which people make decisions together.

Two types of Decisions Made by Groups

- Making a decision together.
- Supporting activities or tasks related to the decision-making process (e.g., select criteria for evaluating alternative solutions, prioritizing, and design strategy to implement them).

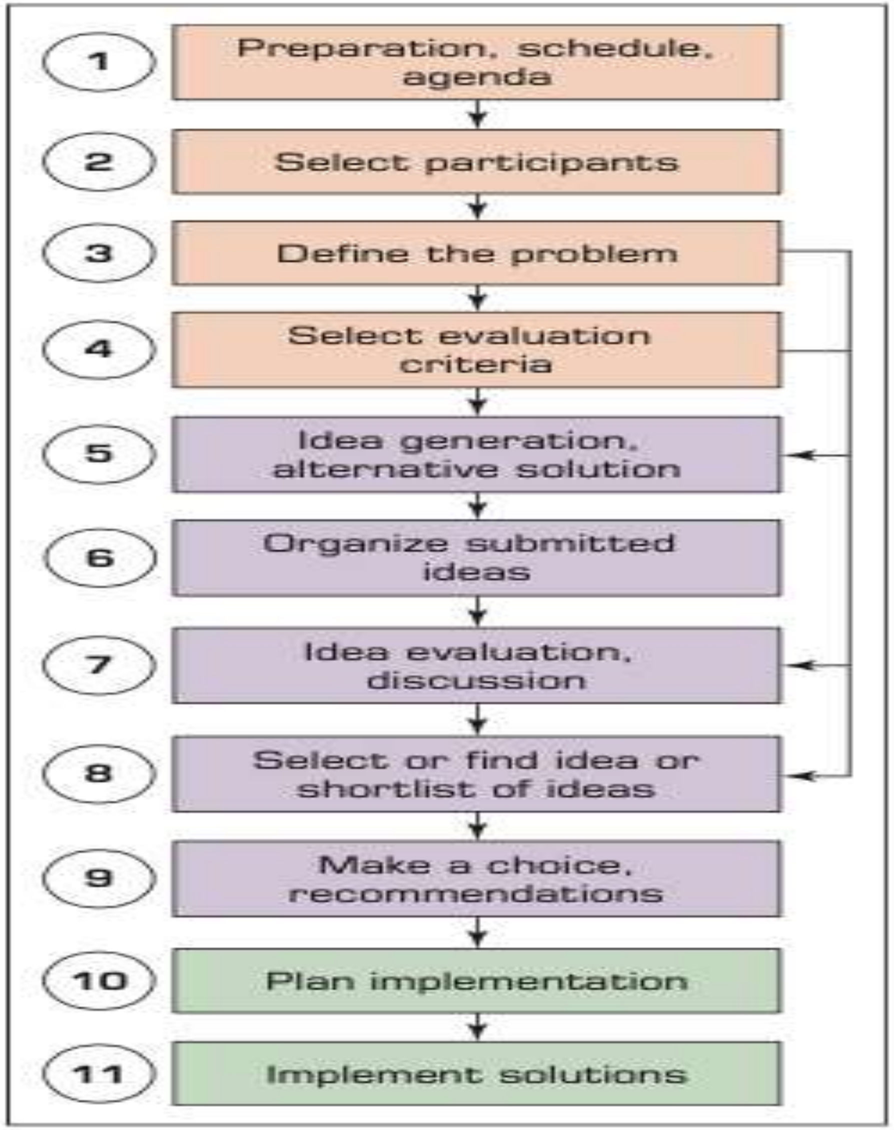
Making Decisions in Groups

Characteristics of Group Work

- Group members may be located in different places (virtual team)
- Group members may work at different times
- Group members may work for the same organization or different organizations
- A group can be permanent or temporary
- A group can be at one managerial level or span several levels
- A group can create synergy (leading to process and task gains) or result in conflict
- A group can generate productivity gains and/or losses
- It may be impossible or too expensive for all team members to meet in one place at the same time
- The expertise of a group's team members may be needed
- Concentrate on decision making or problem solving
- The decisions made by a group are easier to implement if supported by all (or at least most) members

Group Decision-Making Process

- The process of group decision making is similar to that of the general decision-making process but has more steps.



The Process of Group Decision Making

Group Decision-Making Process

Group Decision Facts

- The decisions made need to be implemented.
- Group members are typically of equal or nearly equal status.
- The outcome depends on the knowledge, opinions, and judgments of its participants and the support they give to the outcome.

Group Decision-Making Process

Difficulties associated with Group work

- Fear to speak up.
- Unable to reach consensus.
- Lack focus.
- Tendency to produce poor-quality compromises.
- Often non-productive time (e.g., socializing, preparing, waiting for latecomers).
- Tendency to repeat what has already been said (because of failure to remember or process).
- Meeting costs can be high (e.g., travel, participation time spent).
- Incomplete or inappropriate use of information.
- Too much information (i.e., information overload).
- Incomplete or incorrect task analysis.
- Inappropriate or incomplete representation in the group.
- Attention or concentration blockage.

Benefits and Limitations of Group Work

Benefits of Working in Groups (Process Gains)	Dysfunctions of Face-to-Face Group Process (Process Losses)
• Individuals at understanding problems. They can teach each other. • People readily take ownership of problems and their solutions. • Group members have their egos embedded in the final decision, so they are committed it. • Groups are better than individuals at catching errors. • A group has more information and knowledge than any one member does. • Members can combine their knowledge to create new knowledge (e.g., through brainstorming). • A group may produce synergy during problem solving, therefore the effectiveness and/or quality of group work can be greater than the sum of what individual members produce. • Working in a group may stimulate the creativity of the participants and the process. • Working together could allow a group to have better and more precise communication. • Risk propensity is balanced. Groups moderate high-risk takers and encourage conservatives.	• Social pressures of conformity may result in groupthink (i.e., people begin to think alike and not tolerate new ideas; they yield to conformance pressure). • It is a time-consuming, slow process. • Some relevant information could be missing. • A meeting can lack coordination, have a poor agenda, or be poorly planned. • A meeting may be dominated by time, topic, opinion of one or a few individuals, or fear of contributing because of the possibility of conflicts. • Some group members can tend to influence the agenda while some try to rely on others to do most of the work (free riding). The group may ignore good solutions, have poorly defined goals, or be composed of the wrong participants.

11.3 Supporting Group Work and Team Collaboration with Computerized Systems

- Overview of Group Support Systems (GSS)
- Time/Place Framework

Overview of Group Support Systems (GSS)

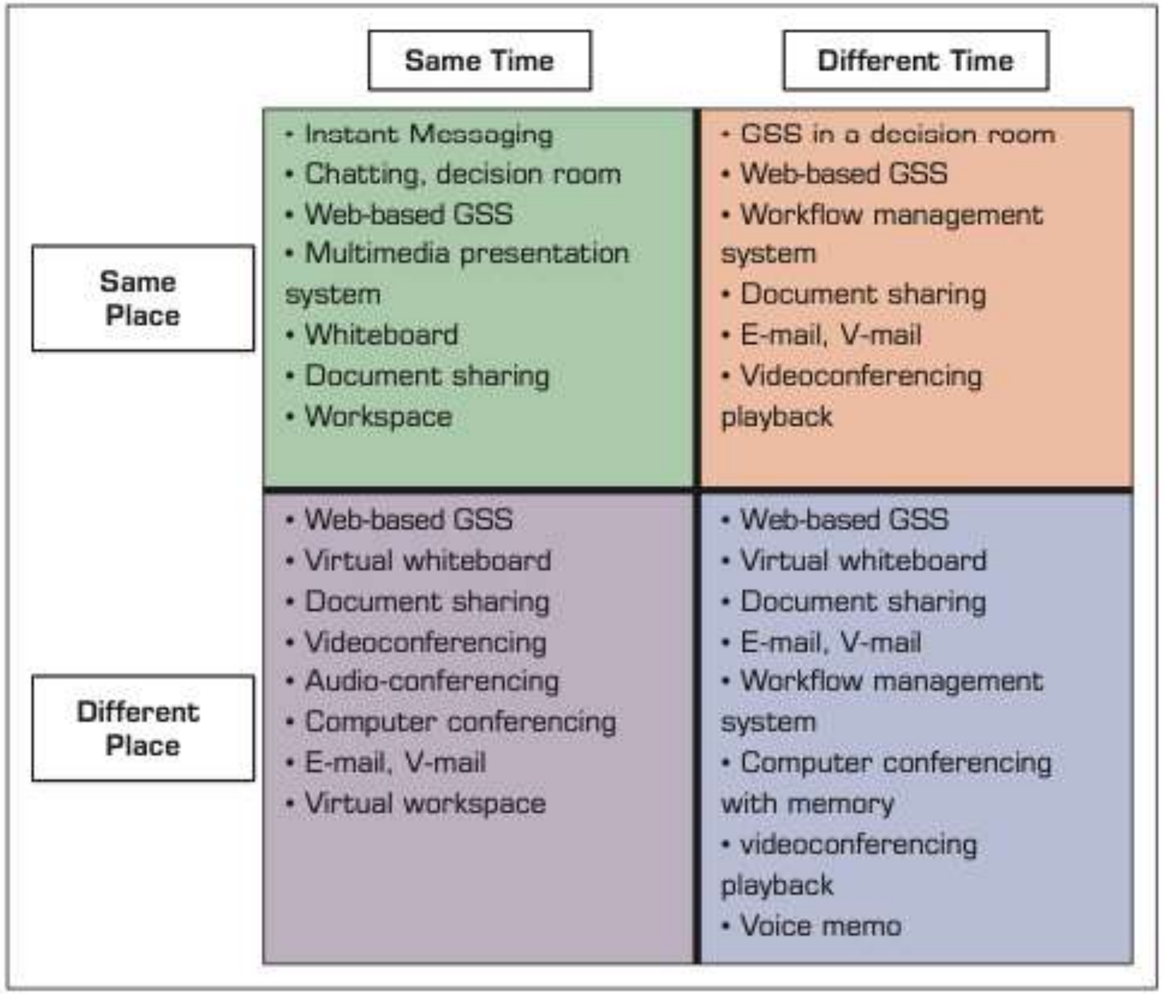
- Software products that provide collaborative support to groups
- Goal is to support group work
- Reduce the chance for process losses and increase benefit.
- Provide direct or indirect support by improving communication between team members
  - Direct support contains capabilities for brainstorming, conferencing, scheduling group meetings; planning; resolving conflicts; videoconferencing; sharing electronic documents; voting; formulating policy; and analysing enterprise data.
  - Indirect support: Internet, intranets, and IoT support virtual meetings and decision making through collaborative tools and access to data analysis, information, and knowledge.

Time/Place Framework

- The time/place framework - convenient way to describe the communication and collaboration patterns and support of group work.

- Communication can be synchronous (i.e., same time) or asynchronous (i.e., sent and received at different times).

- Different technologies can support different time/place settings.



11.4 Electronic Support to Group Communication and Collaboration

- Groupware for Group Collaboration

11.5 Direct Computerized Support for Group Decision Making

- Group decision support system (GDSS)
- Brainstorming for Idea Generation and Problem Solving
- Group Support System (GSS)

Groupware for Group Collaboration

Groupware - computerized tools for supporting group work.

- E.g., e-mail programs, chat rooms, IM, and teleconferences

- Their primary objective is to provide indirect support.

- Groupware products provides a mechanism for team members to share opinions, data, information, knowledge, and other resources.

- Synchronous versus Asynchronous Products
  - E.g.,
    - Web conferencing and IM (synchronous mode)
    - Google Drive, dropbox.com, Microsoft SharePoint,... (asynchronous modes)

Groupware for Group Collaboration

Features of groupware products that support commutation, collaboration, and coordination.

TABLE 11.1 Groupware Products and Features

General (Can Be Either Synchronous or Asynchronous)

- Built-in e-mail, messaging system
- Browser interface
- Joint Web page creation
- Active hyperlink sharing
- File sharing (graphics, video, audio, or other)
- Built-in search functions (by topic or keyword)
- Workflow tools
- Corporate portals for communication, collaboration, and search
- Shared screens
- Electronic decision rooms
- Peer-to-peer networks

Synchronous (Same Time)

- IM
- Videoconferences, multimedia conferences
- Audioconferences
- Shared whiteboard, smart whiteboard
- Instant videos
- Brainstorming
- Polling (voting) and other decision support (activities such as consensus building, scheduling)
  - Chats with people
  - Chats with bots

Asynchronous (Different Times)

- Virtual workspaces
- Tweets
- Ability to receive/send e-mail, SMS
- Ability to receive notification alerts via e-mail or SMS
- Ability to collapse/expand discussion threads
- Message sorting (by date, author, or read/unread)
- Auto responders
- Chat session logs
- Electronic bulletin boards, discussion groups
- Blogs and wikis
  - Collaborative planning and/or design tools

Groupware for Group Collaboration

- Virtual Meeting Systems
  - Support virtual team members in different locations
  - E.g., webex.com, gotomeeting.com, Skype.com, …

- Collaborative Workflow
  - software products that address project-oriented and collaborative processes
  - leading vendors FileNet and Action Technologies ,…

- Digital Collaborative Workspace
  - people can work together from any location at the same or at a different time
  - Slack workspace, Cisco’s Webex,…

- Collaborative Networks and Hubs
  - The purpose of a collaborative hub is to be a centre point for group collaboration.

Groupware for Group Collaboration

- Social Collaboration refers to collaboration conducted within and between socially oriented groups.
  - usually done on social media sites, and it is enabled by the Internet, IoT, and diversified social collaboration software.

Collaboration in social networks: Business-related collaboration is most evidenced on Facebook and LinkedIn.

- Instagram, Pinterest, and Twitter support collaboration as well.

There are software platforms specifically for forming teams and supporting their activities.

- E.g., Wrike, Ryver, Azendoo, Zimbra social platform, Samepage, Zoho, Asana, Jive, Chatter, and Social Tables…

Groupware for Group Collaboration

- Other tools that support collaboration
  - **Notejoy** makes collaborative notes for team
  - **Kahootz** brings stakeholders together to form communities of interest
  - **Nowbridge** offers team connectivity, ability to see participants
  - **Walkabout** Workplace (is a 3D virtual office for remote teams
  - **RealtimeBoard** is a enterprise visual collaboration
  - **Quora** is a popular place for posting questions to the crowd
  - **Pinterest** provides an e-commerce workspace that allows collection of text and images on selected topics
  - **IBM connection closed** offers a comprehensive communication and collaboration tool set
  - **Skedda** schedules space for coworking
  - **Zinc** is a social collaboration tool
  - **Scribblar** is an online collaboration room for virtual brainstorming
  - **Collokia** is a machine learning platform for workflow

Group decision support system (GDSS)

- Major organizational decisions are made by groups, such as executive committees and special task forces.

- Group decision support system (GDSS) is an interactive computer-based system that facilitates the solution of semi-structured or unstructured problems by a group of decision makers.

- Goal: improve the productivity of decision-making meetings by speeding up the decision-making process and/or to increase the quality of the resulting decisions.

- It provides direct support in face-to-face settings and in virtual meetings, attempting to increase process gains, and reducing process losses of group works.

Group decision support system (GDSS)

Major characteristics and capabilities of a GDSS:

- supports the *process* of group decision makers mainly by providing automation of subprocesses (e.g., brainstorming) and using information technology tools.
- It is a specially designed information system, not merely a configuration of already existing system components.
- It can be designed to address one type of problem or make a variety of group-level organizational decisions.
- It encourages generation of ideas, resolution of conflicts, and freedom of expression.
- It contains built-in mechanisms that discourage development of negative group behaviors, such as destructive conflict, miscommunication, and groupthink.

Group decision support system (GDSS)

There are two options for deploying GDSS technology:

- Decision room
  - GDSS is installed in expensive, customized, special- purpose facilities room.

- Internet-based groupware
  - The most common approach to GSS and GDSS delivery approach.
  - Allows group members to work from any location at any time (e.g., WebEx, GoToMeeting, Adobe Connect, IBM Connections, Microsoft Teams).

The major limitations of electronic software support are:

  - Increased cognitive load
  - Fear of using new technology
  - Need for technical assistance

Brainstorming for Idea Generation and Problem Solving

- A major activity in group decision making is idea generation.
- Groups not only generate more ideas but also better ones.
- Brainstorming is a process for generating creative ideas.
- Computer programs can support the various brainstorming activities.
  - Support is usually for online brainstorming, synchronously or asynchronously.
- Artificial intelligence can supports brainstorming
  - E.g., pattern recognition and identifying ideas that are similar to each other, idea generation and voting.

Brainstorming for Idea Generation and Problem Solving

The major features of brainstorming software packages:

- Creation of a large number of ideas.
- Large group participation.
- Real-time updates.
- Information color coding.
- Collaborative editing.
- Design of brainstorming sessions.
- Idea sharing.
- People participation.
- Idea mapping (e.g., create mind maps).
- Text, video, documents, etc. posting.
- Remote brainstorming.
- Creation of an electronic archive.
- Reduction of social loafing.

Group Support System (GSS)

GSS is any combination of hardware and software that enhances group work.

GSS improves group work

- Supporting parallel processing of information and idea generation (brainstorming).
- Enabling the participation of larger groups with more complete information, knowledge, and skills.
- Permitting the group to use structured or unstructured techniques and methods.
- Offering rapid, easy access to external information.
- Allowing parallel computer discussions.
- Helping participants frame the big picture.
- Providing anonymity, which allows shy people to contribute to the meeting (i.e., to get up and do what needs to be done).
- Providing measures that help prevent aggressive individuals from controlling a meeting.
- Providing multiple ways to participate in instant anonymous voting.

11.6 Collective Intelligence and Collaborative Intelligence

- Collective Intelligence Definitions and Benefits
- Collaborative Intelligence

Collective Intelligence Definitions and Benefits

- Collective intelligence (CI) - total intelligence of a group
- Based on the premise that the combined wisdom of several collaborating people is greater than that of individuals working separately.
- CI benefits are the ability to solve complex problems and/or design new products and services that result from innovations (e.g., MIT center for CI)

Collective Intelligence Definitions and Benefits

Types of collective intelligence - three major areas of applications-

- Cognition
- Cooperation
- Coordination

- Several configurations of collective intelligence can be supported differently by technology (e.g., internet, IoT).

- Collaboration platforms (e.g., Microsoft Teams and Slack) can facilitate collective intelligence.

Collaborative Intelligence

How to make people collaborate in groups? *(Coleman ,2011)*

1. willingness to share
2. knowing how to share
3. being willing to collaborate
4. knowing what to share
5. knowing how to build trust
6. understanding team dynamics
7. using correct hubs for networking
8. mentoring and coaching properly
9. being open to new ideas
10. using computerized tools and technology

How to Create Business Value from Collaboration?

The IBM Study concludes that “Collective intelligence is a powerful resource for creating value using the experience and insights of vast numbers of people around the world.”

11.7 Crowdsourcing as a Method for Decision Support

- The Essentials of Crowdsourcing
- The Process Of Crowdsourcing
- The Role Of Crowdsourcing In Decision Making

The Essentials of Crowdsourcing

- A process of outsourcing work to a crowd.

- Method of collective intelligence

Several reasons for outsourcing work:

  - Necessary skills may not be available internally
  - speed of execution is needed
  - problems are too complex to solve
  - Or special innovation is needed (innocentive.com)

Goal: improve decision making, assist in solving difficult problem, and wide exposure to expertise.

The Essentials of Crowdsourcing

Some crowdsourcing use case

- Airbnb is using user-submitted videos (15 seconds each) that describe travel sites.
- Dell’s Idea Storm (**ideastorm.com**) enables customers to vote on features of Idea Storm
- The LEGO company has a platform called LEGO Ideas through which users can
- submit ideas for new LEGO sets and vote on submitted ideas by the crowd.

Crowdsourcing vendors

- E.g., Logo design (Design Bill), Problem solving (InnoCentive, NineSigma, IdeaConnection), Business innovation (Chardix), Brand names (Name This), Product and manufacturing design (Pronto ERP),Data cleansing (Amazon Mechanical Turk), Software testing (uTest), Trend watching(TrendWatching), Images (Flickr Creative Commons)…

The Essentials of Crowdsourcing

Major Types Of Crowdsourcing

1. **Collective intelligence (or wisdom):** People in crowds are solving problems and providing new insights and ideas
2. **Crowd creation:** People are creating various types of content and sharing it with others (for pay or free).
3. **Crowd voting:** People are giving their opinions and ratings on ideas, products, or services (e.g., voting in *American Idol* competitions).
4. **Crowd support and funding.** People are contributing and supporting endeavors for social or business causes, such as offering donations, and micro-financing new ventures.

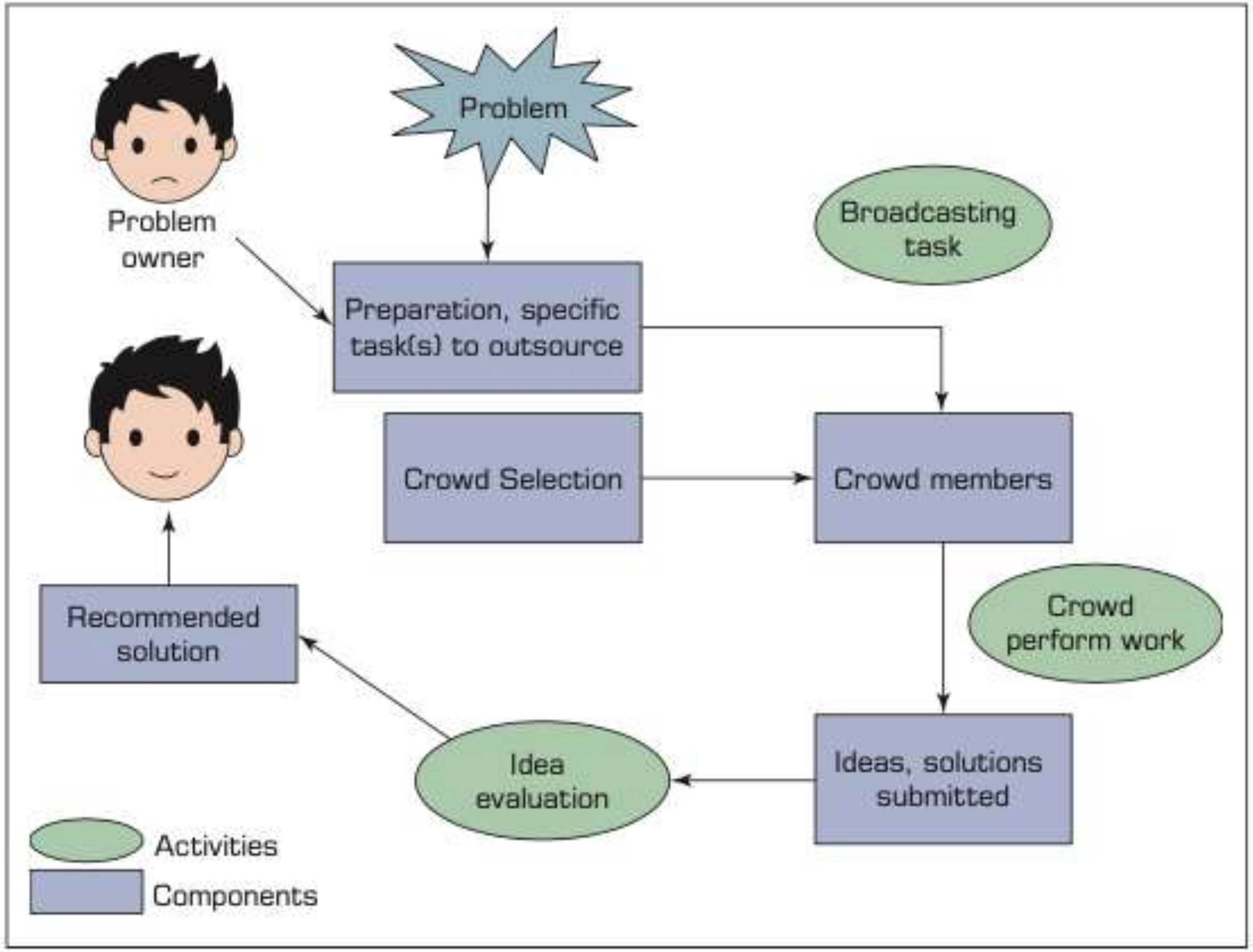
The Process Of Crowdsourcing

The Crowdsourcing Process

1. Identify the problem and the task(s) to be outsourced.
2. Select the target crowd (if not an open call).
3. Broadcast the task to the crowd (or to an unidentified crowd in an open call).
4. Engage the crowd in accomplishing the task (e.g., idea generation, problem solving).
5. Collect user-generated content.
6. Have the quality of submitted material evaluated by the management that initiated the request, by experts, or by a crowd.
7. Select the best solution (or a short list).
8. Compensate the crowd (e.g., the winning proposal).
9. Implement the solution.

The Process Of Crowdsourcing

- The process of crowdsourcing can differs from application to application



The Role Of Crowdsourcing In Decision Making

- Crowds can provide ideas in a collaborative or a competitive mode.

- Their role may differ at different stages of the decision-making process
  - E.g., use a crowd to decide how to respond to a competitor’s act or to help decide whether a proposed design is useful.

Potential roles of a crowd according to Simon’s model:

- Three major phases: intelligence, design, and choice.
- Crowdsourcing can provide different types of support to managerial decision-making process.
- Most of the applications are in the design phase (e.g., idea generation and co-creation) and in the choice phase (voting).
- Support can also be provided in all phases of the process.

11.8 Artificial Intelligence and Swarm AI Support of Team Collaboration and Group Decision Making

- The Essentials of Crowdsourcing
- The Process Of Crowdsourcing
- The Role Of Crowdsourcing In Decision Making

AI Support of Group Decision Making

AI can support many activities in group decision making

- **Meeting preparation**

  AI can assist in scheduling meetings so that all can participate.
- **Problem identification**

  AI can be used to identify potential or difficult to pinpoint problems.
- **Idea generation & organization**

  Team members can increase their creativity when they use AI for support.
- **Idea Group interaction and collaboration.**
- **Predictions**
- **Multinational groups**
- **Bots**

  Group members may consult Alexa and other bots. Chatbots can provide answers to queries in real time.
- **Other advisors**

  IBM Watson can provide useful advice during meetings

AI Support of Team Collaboration

How AI may impact collaboration? *(Cisco Systems survey)*

- Virtual assistants increase employees' productivity, creativity, and job satisfaction.
- Bots enable employees to focus on high-value tasks.
- Bots are accepted as part of workers' teams.
- Bots improve conference calls. They also can take meetings notes and schedule meetings.
- AI can use facial recognition to sign in eligible people to meetings.
- Personal characteristics are likely to influence how people feel about AI in the workplace.
- Employees in general like to have AI in their teams.
- Security is a major concern when AI, such as virtual assistants, is used in teams.
- The major AI tools that are most useful are NLP and voice response;
- AI can also summarize the key topics of meetings and understand participants' needs.
- AI can be aware of organizational goals and workers' skills and can make suggestions accordingly.

Swarm Intelligence and Swarm AI

- Swarm intelligence - collective behavior of decentralized, self-organized systems, natural or artificial.

- Swarm AI provides the algorithms for the interconnections among people to enable the knowledge, intuition, experience, and wisdom of individuals to merge into single improved swarm intelligence.

- Swarm AI has had significant success as method of communication in crowdsourcing

- Swarm AI can be used for making predictions in difficult-to-assess situations.
  - e.g., predicting winners in the regular NFL season, predicting the top recipients of the Oscars in 2018,…

11.9 Human–Machine Collaboration

- Robots as Coworkers: Opportunities and Challenges
- Teams of collaborating Robots

Human–Machine Collaboration

- Machines that once supported manufacturing work are used now also in support of cognitive, including managerial, and work.

- Human–machine/AI collaboration will shape the future of work

- For people and machines to work in teams, it is necessary to make special preparations.
  - humans controlling and monitoring production and robots doing physical work that requires speed, power, accuracy, or nonstop attention..

Human–Machine Collaboration

- AI enables the automation of nonmanual activities.

- Today human–machine collaboration in cognitive jobs become possible with the advancement of machine learning and deep learning.
  - E.g., IBM Watson, chatbots, medical diagnoses of complex situations, investments decisions…

- Robots complement human capabilities
  - E.g., Amazon’s distribution centers mobile robots do a variety of tasks, mostly in hauling materials and helping to fulfill customer orders

Robots as Coworkers: Opportunities and Challenges

- Robots will become cognitive coworkers and help people be more productive.

- Collaboration reduced idle time by 85 percent (Tobe, 2015)

- Collaborative robots (co-bots) are designed to work with people, assisting in executing various tasks.

- Blending humans and AI to best serve customers

Robots as Coworkers: Opportunities and Challenges

Some challenges must be considered

- Designing a human–machine team that capitalizes on the strength of each partner
- Exchanging information between humans and robots
- Preparing company employees in all departments for the collaboration (Marr,2017)
- Changing business processes to accommodate human–robot collaboration (Moran,2018)
- Ensuring the safety of robots and employees that work together

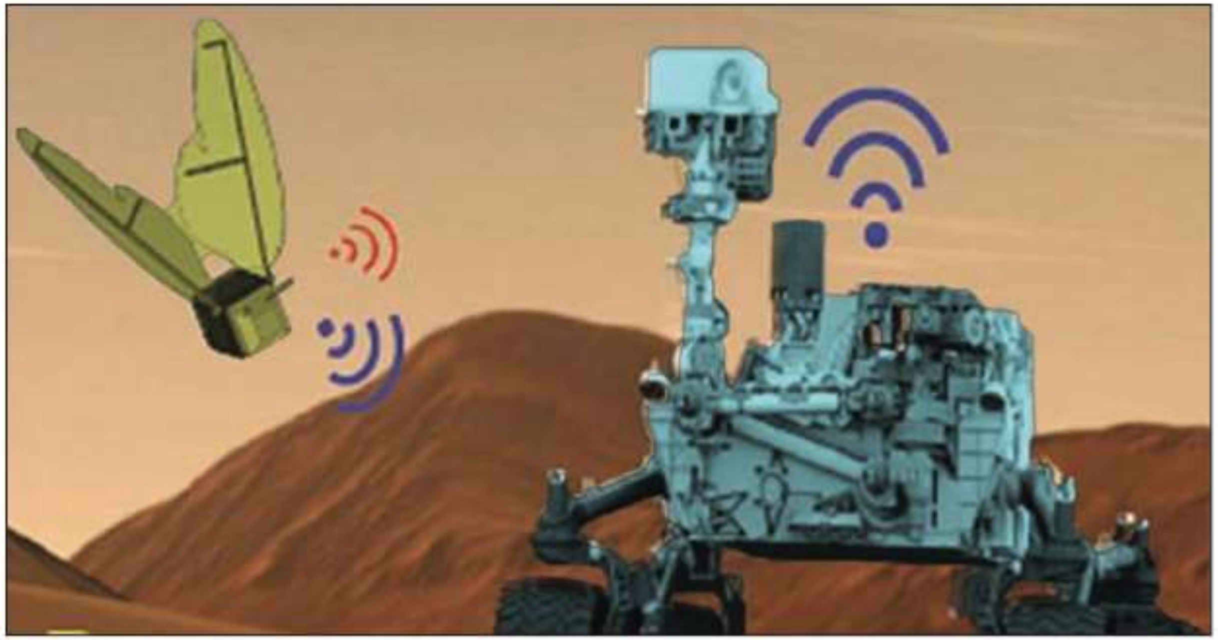
Robots as Coworkers: Opportunities and Challenges

Technologies That Support Robots As Coworkers *(Yurieff,2018b)*

- Virtual reality can be used as a powerful training tool (e.g., for safety).
- A robot is working with an ad agency in Japan to generate ideas.
- A robot can be your boss.
- Robots are coworkers in providing parts out of bins in assembly lines and can check quality together with humans.
- AI tools measure blood flow and volume of the cardiac muscles in seconds. This information facilitates the decisions made by radiologists.

Teams of collaborating Robots

- Future directions in robotics is creating teams of robots that are designed to do complex work



Team of Robots Prepares to Go to Mars

Main Reference

- **Chapter 11**: “*Group Decision Making, Collaborative Systems, and AI Support* ” from “*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”.*

Week self-review exercises

- **Application case 11.1 – 11.3** from *“Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”*

Thank You

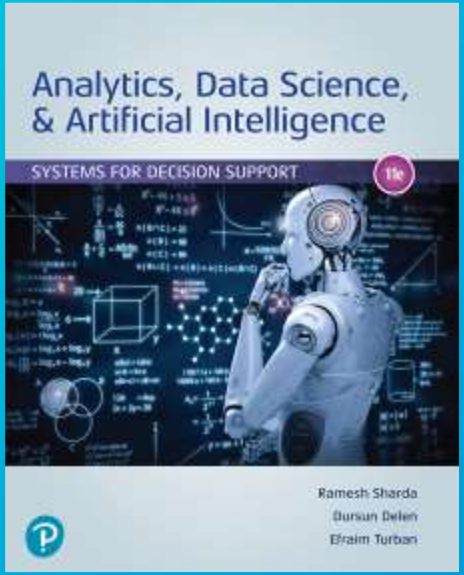
الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

College of Computing and Informatics

Week 10

Chapter 7: Text Mining, Sentiment Analysis, and Social Analytics



Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Describe text analytics and understand the need for text mining.
2. Differentiate among text analytics, text mining, and data mining
3. Understand the different application areas for text mining.
4. Know the process of carrying out a text mining project.
5. Appreciate the different methods to introduce structure- to text-based data.
6. Describe sentiment analysis and develop familiarity with popular applications of sentiment analysis and learn the common methods for sentiment analysis.
7. Become familiar with speech analytics as it relates to sentiment analysis.
8. Learn three facets of Web analytics—content, structure, and usage mining.
9. Know social analytics including social media and Social network analyses.

Required Reading

- **Chapter 7:** “Text Mining, Sentiment Analysis, and Social Analytics” from “*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*”.

Recommended Reading

- Sentiment Analysis: A Definitive Guide. MonkeyLearn. https://monkeylearn.com/sentiment-analysis/

Recommended Video

- What is Text Mining? (2015, Oct 8). [Video]. YouTube. https://www.youtube.com/watch?v=I3cjbB38Z4A

- Natural Language Processing In 5 Minutes | What Is NLP And How Does It Work? | Simplilearn (2021, Mar 17). [Video]. YouTube. https://www.youtube.com/watch?v=CMrHM8a3hqw

7.2 Text Analytics and Text Mining Overview

- Text Mining Concepts
- Text analytics versus Text Mining
- Data Mining versus Text Mining
- Text Mining Application Area
- Text Mining Terminology

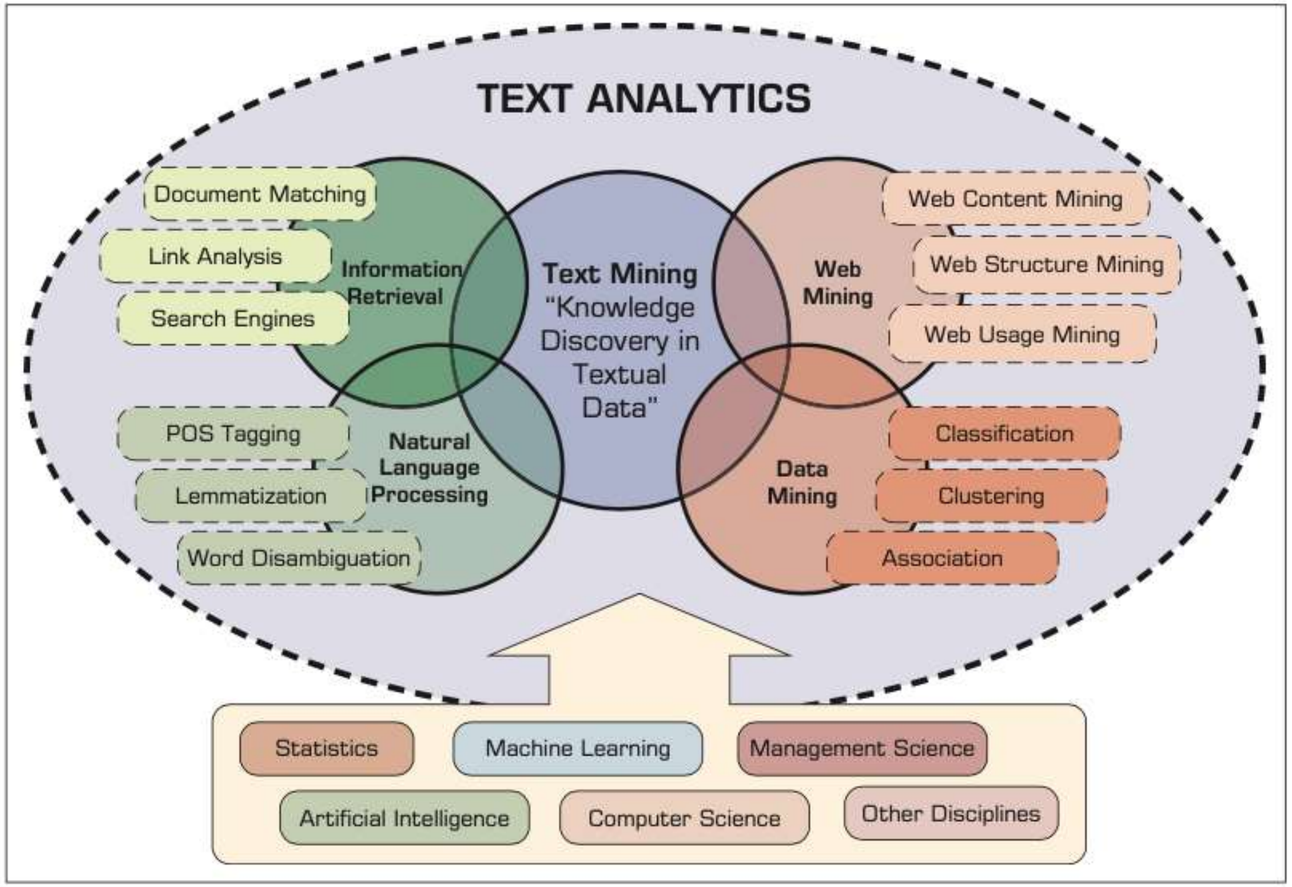
Text Mining Concepts

- Most of corporate data are captured and stored in some sort of unstructured form (e.g., text)
- knowledge is derived from data and information → power in today's business world.
- Businesses that effectively and efficiently tap into their text data sources will have the necessary knowledge to make better decision.
- To stay competitive text mining is needed!
  - E.g., Amadori Group Converts Consumer Sentiments into Near-Real-Time Sales using text mining.

What is Text Mining?

- Text mining is the discovery of knowledge from largely unstructured (text-based) data sources.
- One of the fastest-growing branches of business intelligence field.
- A semiautomated process of extracting patterns (useful information and knowledge) unstructured data sources.

Text Analytics and Text Mining



A relationships between text analytics and text mining along with other related application areas and enabling disciplines

Text analytics versus Text Mining

- Both seek to turn unstructured textual data into actionable information through the application of natural language processing (NLP) and analytic.
- Text analytics is a broader concept that includes: information retrieval (e.g., searching and identifying relevant documents for a given set of key terms), as well as information extraction, data mining, and Web mining.
- Text mining is primarily focused on discovering new and useful knowledge from the textual data sources.
- Even though the two can be defined somewhat differently at times, text analytics and text mining are usually used synonymously.

Data Mining versus Text Mining

- Both has the same purpose of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.

- Both uses the same processes (semi-automate).

- They Differ in input to the process → the nature of the data.

Data mining	Text mining
Input → Structured data found in databases	Input → Unstructured data such as as Word documents, PDF files, text excerpts, and XML files.

- Text mining uses NLP to first induce structure into the text collection and then uses data mining algorithms such as classification, clustering, association, and sequence discovery to extract knowledge from it.

Text Mining Concepts

- Benefits of text mining are obvious
  - E.g., the free-form text-based interactions with customers in the form of complaints (or compliments) can be used to objectively identify product and service characteristics that are deemed to be less than perfect and can be used as input to better product development and service allocations.
- Some text-rich data environments
  - E.g., law (court orders), academic research (research articles), finance (quarterly reports), medicine (discharge summaries), biology (molecular interactions), technology (patent files), and marketing (customer comments).
- Area where the automated processing of unstructured text has had much impact is Electronic communication records (e.g., Email)
  - filter junk e-mail
  - automatically prioritize and categorize e-mail
  - generate automatic responses

Text Mining Application Area

Most popular application areas of text mining:

Information extraction	Identifying key phrases and relationships within text.
Topic tracking	Predicting other documents of interest to the user, based on a user profile and documents that a user views.
Summarization	Summarizing a document to save the reader time.
Categorization	Identifying the main themes of a document and then placing the document into a predefined set of categories based on those themes.
Clustering	Grouping similar documents without having a predefined set of categories.
Concept linking	Connecting related documents by identifying their shared concepts
Question answering	Finding the best answer to a given question through knowledge-driven pattern matching.

Text Mining Terminology

- **Unstructured or semi-structured data**
- **Corpus:** a large and structured set of texts prepared for knowledge discovery.
- **Terms:** single word or multiword phrase extracted directly from the corpus
- **Concepts:** features generated from a collection of documents using manual, statistical, rule-based, or hybrid categorization methodology.
- **Stemming:** process of reducing inflected words to their stem.
- **Stop words:** noise words - filtered out prior to or after processing natural language data (e.g., (*a, an, the*)
- **Synonyms and polysemes**
- **Tokenizing:** a categorized block of text in a sentence.

Text Mining Terminology

- **Term dictionary:** a collection of terms specific to a narrow field that can be used to restrict the extracted terms within a corpus.

- **Word frequency:** number of times a word is found in a specific document.

- **Part-of-speech tagging:** process of marking the words to a particular part of speech (nouns, verbs, adjectives, adverbs, etc.)

- **Morphology:** studies the internal structure of words.

- **Term-by-document matrix:** common representation schema of the frequency-based relationship between the terms and documents (occurrence matrix).

- **Singular value decomposition:**
  - Latent semantic indexing - dimensionality reduction method.

7.3 Natural Language Processing (NLP)

Natural Language Processing (NLP)

- Introducing structure to a collection of text-based documents methods
  - bag-of-words an early approach.
  - natural language processing a new approach.

- The goal of NLP is to move beyond semantics-based text mining (bag-of-words) to a true understanding and processing of natural language that considers grammatical and semantic constraints as well as the context.

- **NLP is:**
  An important component of text mining
  a subfield of artificial intelligence and computational linguistics
  studies the problem of "understanding" the natural human language

Natural Language Processing (NLP)

- A major discussion topics in NLP is the definition and scope of computers “understanding”

  - Natural language is context driven and vague
  - True understanding of meaning requires extensive knowledge of a topic

Will computers ever be able to understand natural language the same way and with the same accuracy that humans do?

Longer way to go to really understand natural human language!

Natural Language Processing (NLP)

- Challenges commonly associated with the implementation of NLP:
  - **Part-of-speech tagging** difficulty in marking up terms in a text (context is all you need!)
  - **Text segmentation** single-word boundaries problem (e.g., Chinese language).
  - **Word sense disambiguation** words with more than one meaning.
  - **Syntax ambiguity** grammar for natural languages is ambiguous
  - **Imperfect or irregular input** (e.g., grammatical errors in texts).
  - **Speech acts** sentence structure might not contain enough information to define acts.

To have algorithms that are capable of automatically reading and obtaining knowledge from text -> Dream of AI community!

Natural Language Processing (NLP)

- To automatically identify the concepts and relationships between them → WordNet
  - A laboriously hand-coded database of English words, their definitions, sets of synonyms, and various semantic relations between synonym sets.
  - A major resource for NLP applications
  - very expensive to build and maintain manually → Need automation
- Sentiment Analysis
  - A technique used to detect favorable and unfavorable opinions toward specific products and services - a settled opinion reflective of one's feelings.
  - SentiWordNet - is an extension of WordNet to be used for sentiment identification.

NLP Task Categories

- NLP has successfully been applied to a variety of domains
  - Question answering
    - answering a question posed in natural language (e.g., chatbots)
  - Automatic summarization
    - creation of a shortened version of a textual document
  - Natural language generation & understanding
  - Machine translation (e.g., google translate)
  - Foreign language reading & writing
  - Speech recognition

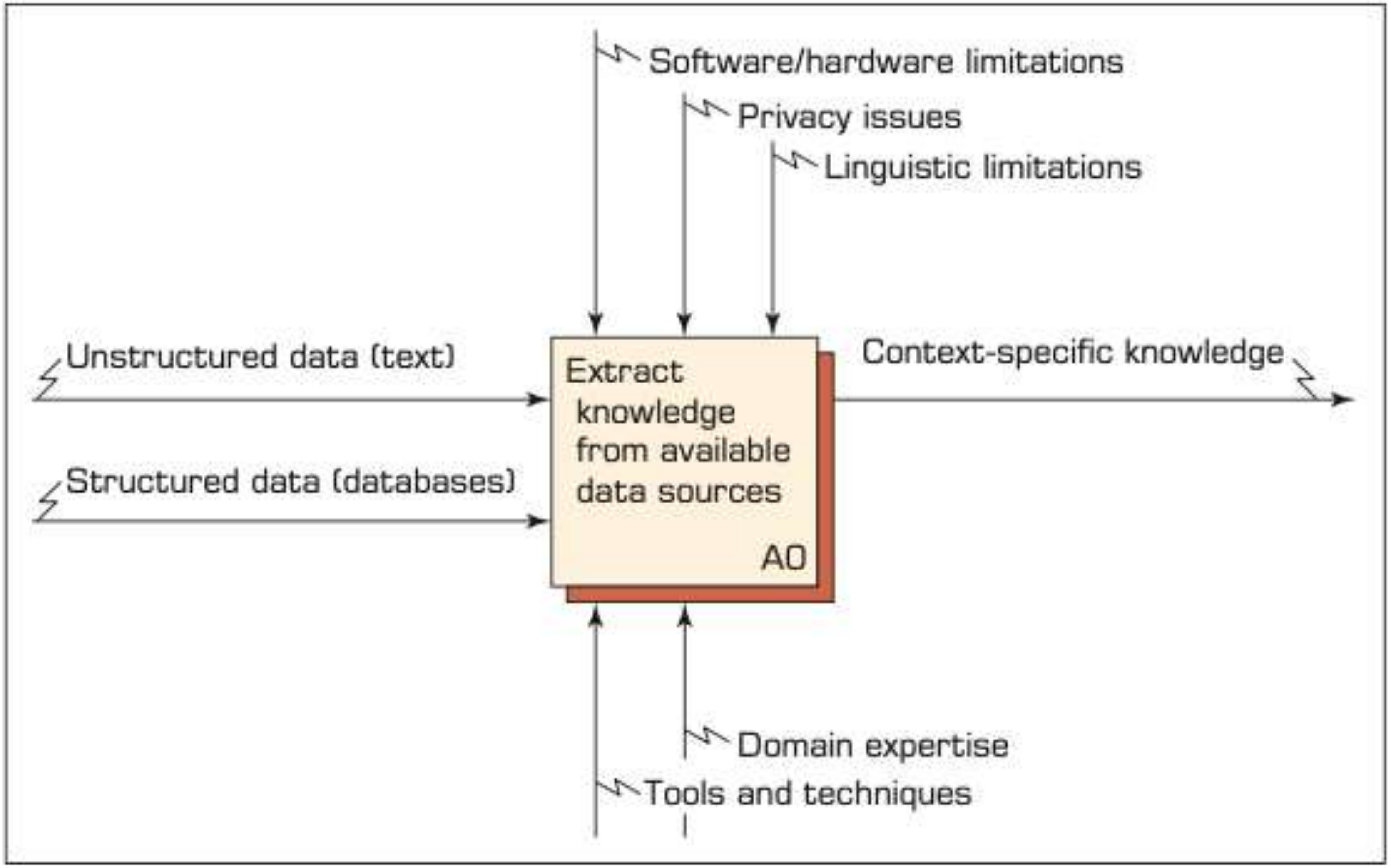
    Conversion of spoken words to machine-readable input
  - Text to speech (a.k.a speech synthesis)
  - Text proofing, optical character recognition

7.4 Text Mining Applications
7.5 Text Mining Process

Text Mining Applications

- **Marketing applications**
  - Increase cross-selling and up-selling by analysing the unstructured data generated by call centres.
  - Enables better CRM
- **Security applications**
  - Deception detection
- **Biomedical Applications**
  - Experimental techniques (e.g., DNA microarray analysis)
- **Academic applications**
  - Research stream analysis

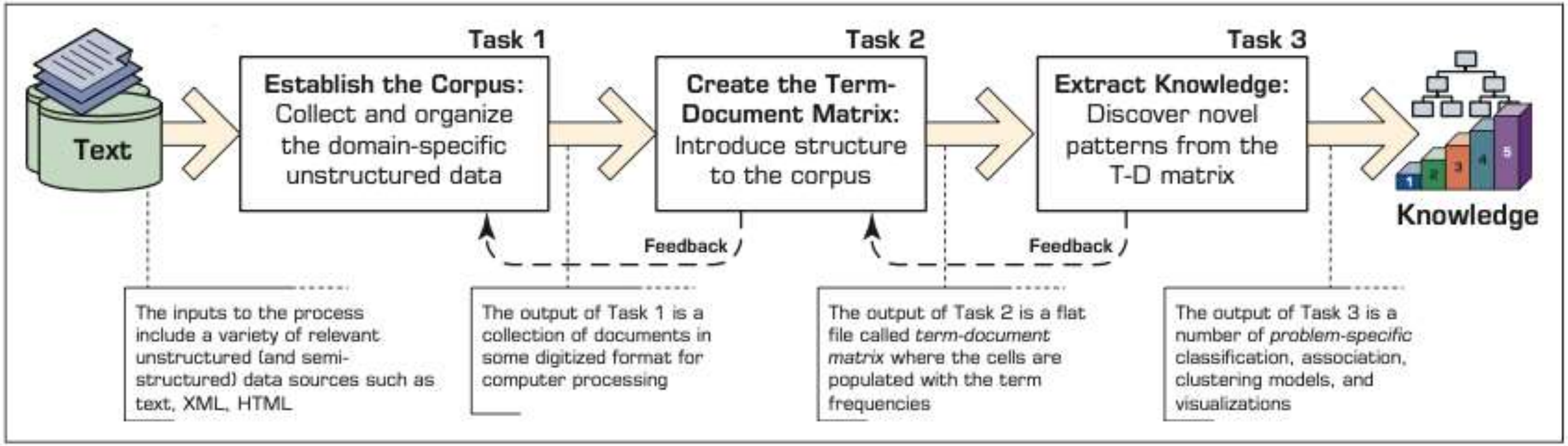
Text Mining Process



Context Diagram for the Text Mining Process

Text Mining Process

- The text mining process can be broken down into three consecutive tasks, each of which has specific inputs to generate certain outputs



The Three-Step/Task Text Mining Process

Text Mining Process

Task 1: Establish the corpus

- Collect all the documents (unstructured data) related to the context (domain of interest) being studied. (e.g., textual documents, XML files, emails, Web pages, and short notes, etc.)

- Transform and organize the collection in a manner such that they are all in the same representational form (e.g., ASCII text files)

- Store the collection in a common place (e.g., in a file folder or a list of links to a collection of Web pages)

Text Mining Process

Task 2: Create the Term-by-Document Matrix (TDM)

- Rows represent the documents and the columns represent the terms.

- The relationships between the terms and documents are characterized by indices (i.e., occurrences of the term)

Documents \ Terms	Investment Risk	Project Management	Software Engineering	Development	SAP	...	
Document 1	1			1			
Document 2		1					
Document 3			3		1		
Document 4		1					
Document 5			2	1			
Document 6	1			1			
...							

Text Mining Process

In TDM:

- Are all terms important when characterizing documents?
  - Stop words
  - Predetermined terms (dictionary)
  - Synonyms
  - Stemming
- What is the best representation of the indices (values in cells)?
  - Row counts; binary frequencies; log frequencies;
  - Inverse document frequency

Text Mining Process

Task 2: Create the Term-by-Document Matrix (TDM)

- TDM is a sparse matrix (most of the cells filled with zeros).

Approaches to reduce the dimensionality of the TDM

- Manual - a domain expert goes through it and eliminate terms that do not make much sense
- Eliminate terms with very few occurrences in very few documents
- Transform the matrix using singular value decomposition (SVD) - similar to principle component analysis

Text Mining Process

Task 3: Extract the knowledge/novel patterns

Main categories of knowledge extraction methods

- Classification (text categorization)
- Clustering (text natural groupings)
  - Improve search recall
  - Improve search precision
  - Scatter/gather
  - Query-specific clustering
- Association
- Trend Analysis

7.6 Sentiment Analysis

- Sentiment Analysis Overview
- Sentiment Analysis Applications
- Sentiment Analysis Process
- Methods for Polarity Identification

Sentiment Analysis Overview

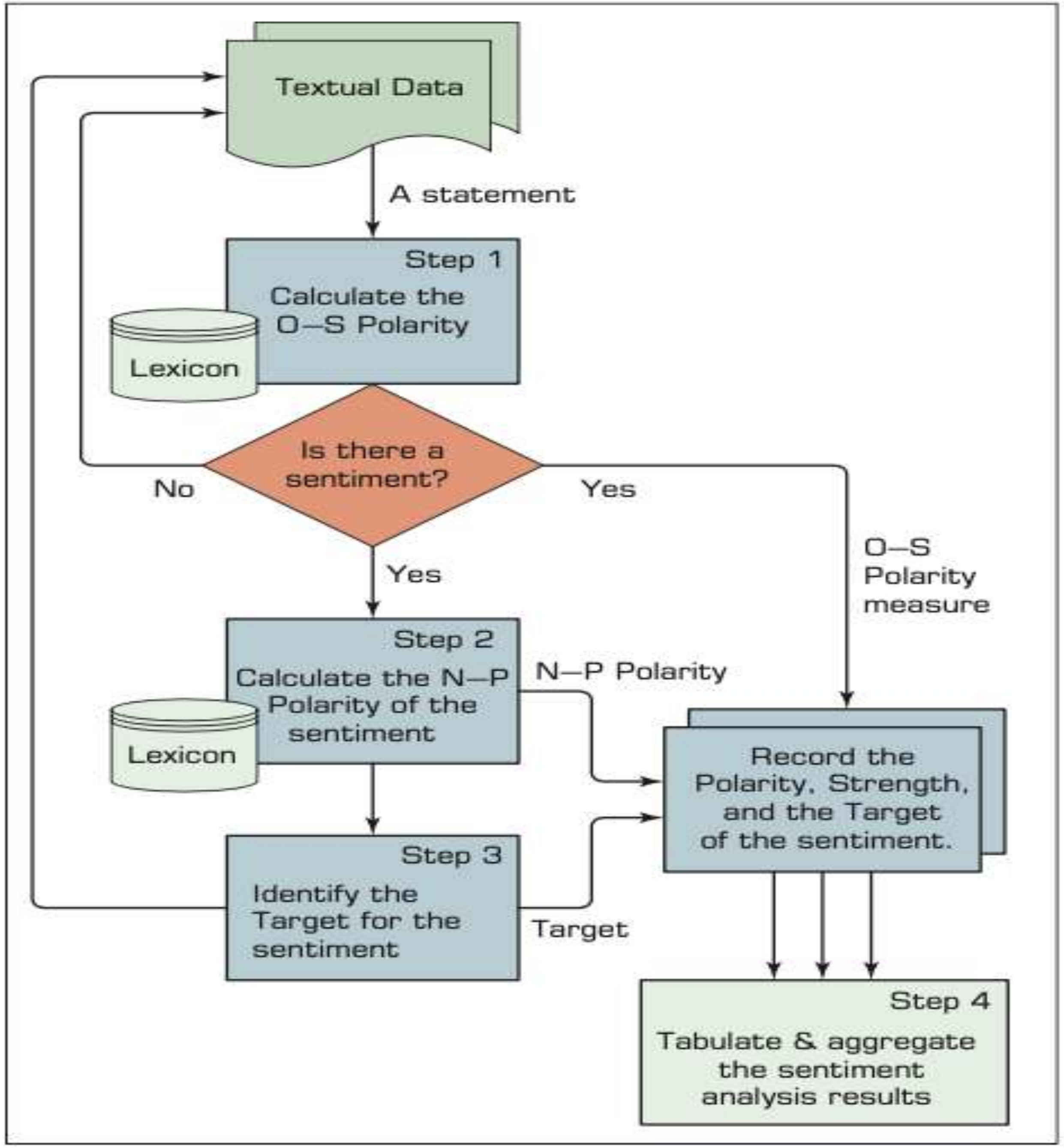
- Sentiment linked to belief, view, opinion, and conviction.
- Sentiment Analysis → opinion mining, subjectivity analysis, appraisal extraction with some connections to affective computing.
- Closely related to computational linguistics, NLP, and text mining.
- Seeks to find out “What do people feel about a certain topic?”
- Deals with differentiating between two classes, positive and negative, versus Neutral - a range of polarity.
- Two types of sentiment: explicit (e.g., “It’s a wonderful day”) or implicit in which text implies an opinion (e.g., “The handle breaks too easily”).

Sentiment Analysis Applications

- Some application examples that illustrate the power and the widespread coverage of sentiment analysis:
    - Voice of the customer (VOC) - an integral part of an analytic CRM
    - Voice of the Market (VOM) - understanding aggregate opinions and trends.
    - Voice of the Employee (VOE)
    - Brand Management
    - Financial Markets
    - Politics
    - Government Intelligence

Sentiment Analysis Process

Multistep Process to Sentiment Analysis



Sentiment Analysis Process

Step 1 – Sentiment Detection

- comes after the retrieval and preparation of the text documents (also called detection of objectivity).
- The goal is to differentiate between a fact and an opinion: objectivity versus subjectivity

Step 2 – N-P Polarity Classification

- Given an opinionated piece of text, the goal is to classify the opinion as falling under one of two opposing sentiment polarities: negative versus positive

Sentiment Analysis Process

Step 3 – Target Identification

- The goal is to accurately identify the target of the expressed sentiment (e.g., a person, a product, an event, etc.)
- Task level of difficulty depends on domain of the analysis.

Step 4 – Collection and Aggregation

- Once the sentiments of all text data points in the document are identified and calculated, they are aggregated and converted to a single sentiment measure (e.g., summing up the polarities).
- From word level to statement, paragraph or document level.

Methods for Polarity Identification

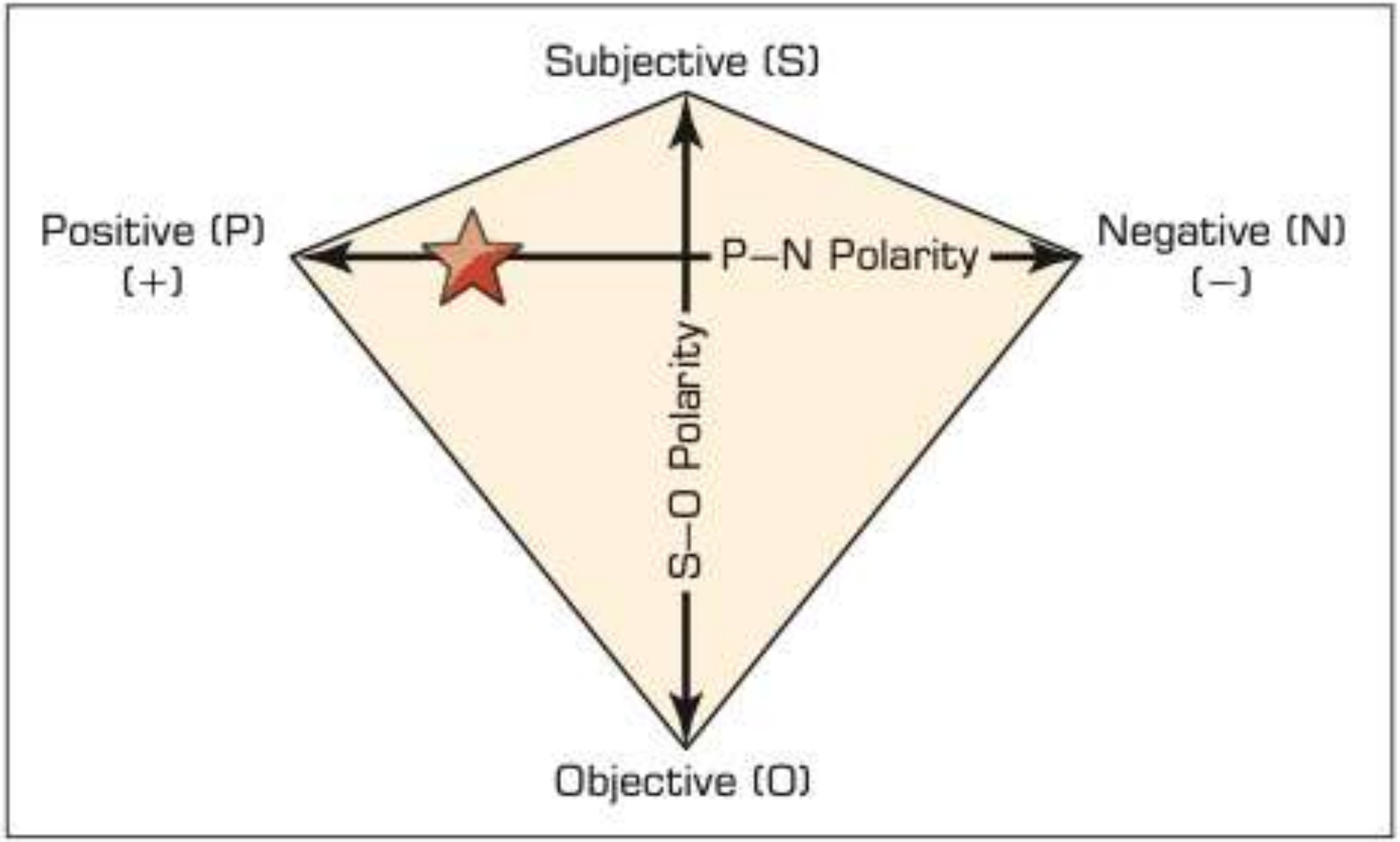
Polarity Identification – negative vs positive

- Can be made at the word, term, sentence, or document level.

- Two dominant techniques have been used for identification of polarity at word level:
  1. Using a lexicon as a reference library (e.g., WordNet, SentiWordNet)
  2. Using a collection of training documents
     - statistical analysis / machine learning
     - Large labeled textual data sets for prediction (e.g., Cornell Movie-Review, Stanford—Large Movie Review data sets).

- Identifying semantic orientation of sentences and phrases or documents
  - Extend semantic orientation of individual words to phrase, sentence or documents in which the word appears.
  - Aggregation typically accomplished by some type of sentiment polarity averaging.

P-N Polarity and S-O Polarity



P–N Polarity and S–O Polarity Relationship

7.7 Web Mining Overview

- Web Content and Web Structure Mining

7.8 Search Engines

- Anatomy of a Search Engine
- Search Engine Optimization
- Methods for Search Engine Optimization

7.8 Web Usage Mining (Web Analytics)

- Web Analytics Technologies
- Web Analytics Metrics

Web Mining Overview

- The growth of the Internet has made data creation, data collection, and data/information/opinion exchange easier.

- The amount of information on the Web is growing rapidly → creating the largest data and text repository.

- Web provides rich data for knowledge discovery (e.g., HTML ,XML, hyperlinks and usage information) → Web mining

- Due to the size and complexity, mining the Web is challenging
  - The Web is too big for effective data mining
  - The Web is too complex
  - The Web is too dynamic
  - The Web is not specific to a domain
  - The Web has everything

Web Mining Overview

- Web mining is the process of discovering intrinsic relationships (i.e., interesting and useful information) from Web data (textual, linkage, or usage information).

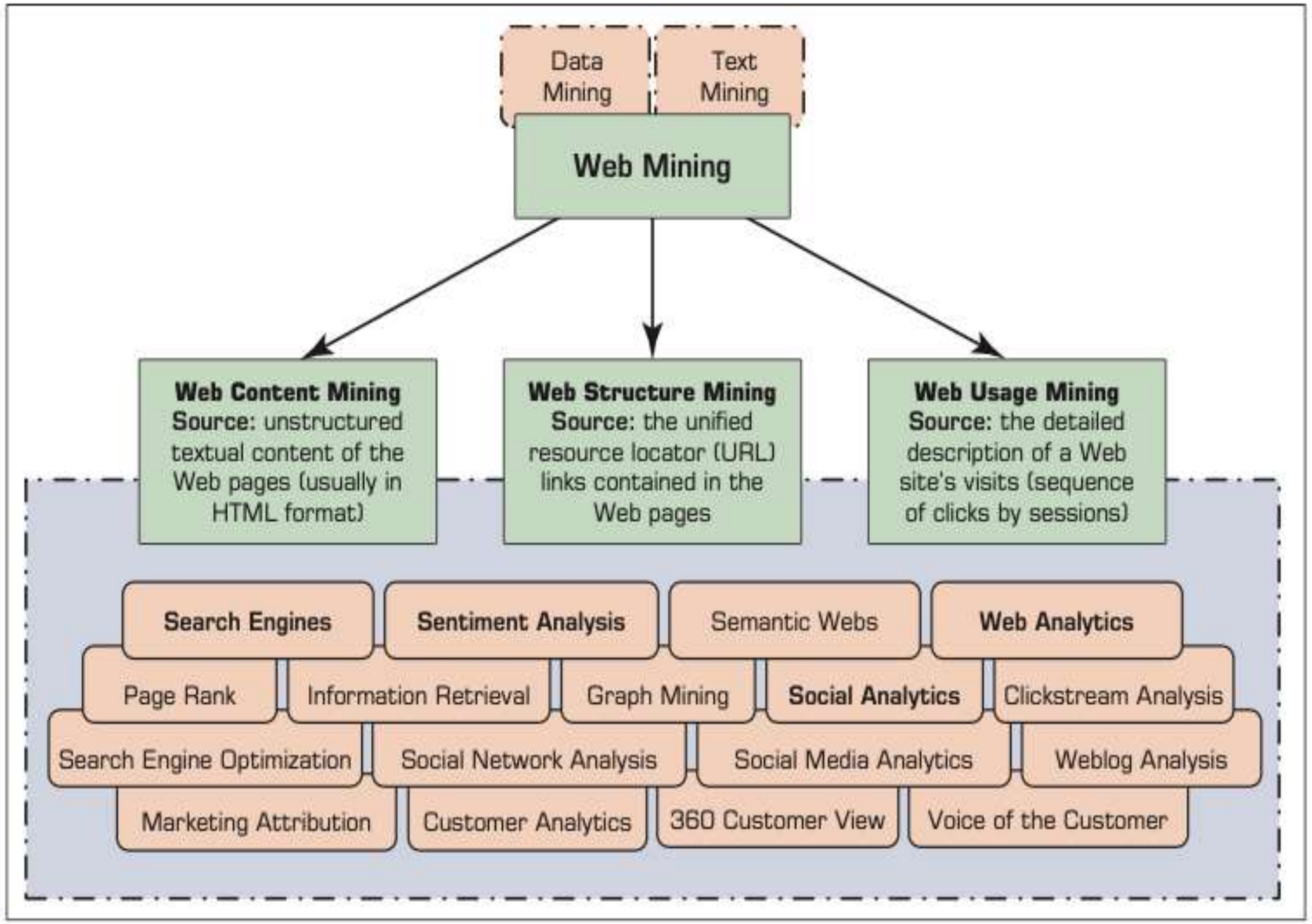
- The goal is to turn business transactions, customer interactions, and Web site usage data into actionable information (i.e., knowledge) to promote better decision making.

- It is the same as data mining but uses data generated over the Web.

Web Mining Overview

- Web mining consists of three areas:
  - Content mining
  - Structure mining
  - Usage mining
- Web Mining versus Web Analytics
  - Web analytics → Web site usage focused data
  - Web mining → all data generated via the Internet (i.e., transaction, social).

Taxonomy of Web Mining



Web Content and Web Structure Mining

- Automatic extraction of useful information from Web pages (textual content) → Web content mining

- Data are collected Web crawlers (a.k.a spiders).

- It can be used to enhance search results produced by search engines.

- Web pages include hyperlinks: Authoritative pages and Hubs
  - hyperlink-induced topic search (HITS): referenced algorithm used to calculate hubs and authorities

- Web structure mining generate interesting information from the links embedded in Web documents.

  - used to identify authoritative pages and hubs

Search Engines

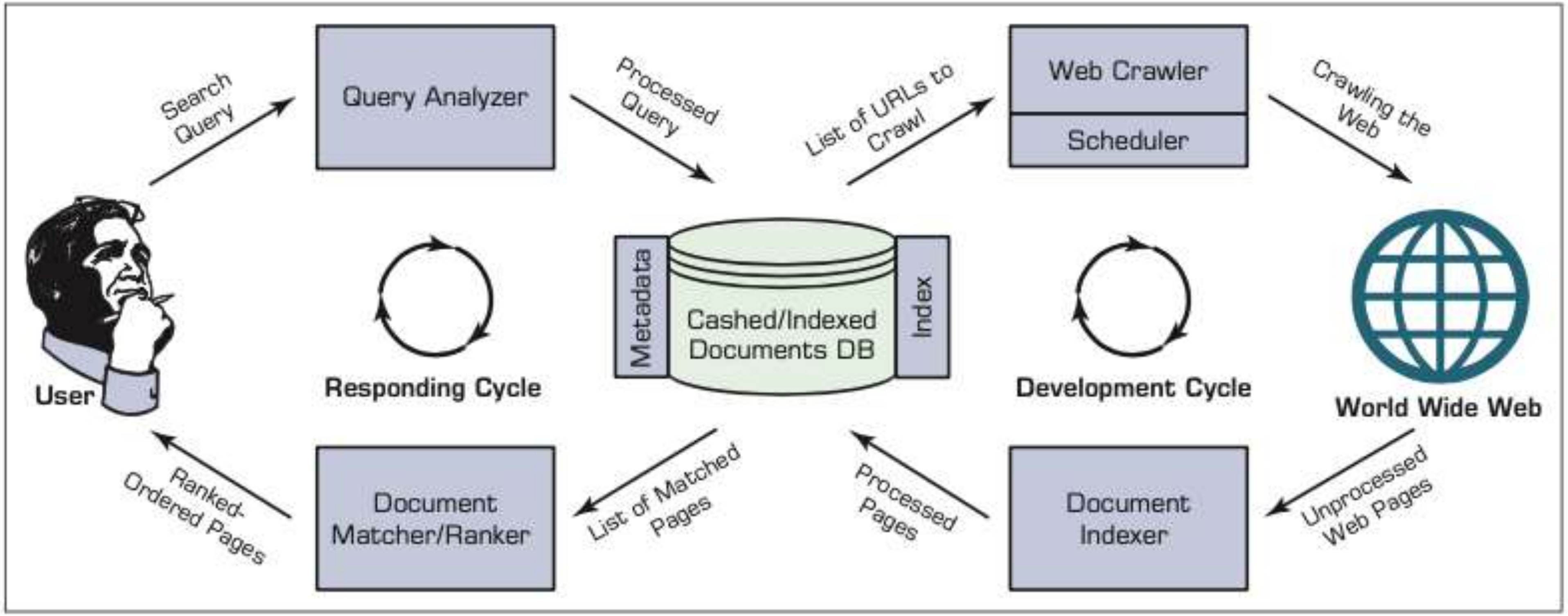
- Search engine→ Software program that searches for documents (Internet sites or files) based on the keywords (individual words, multi-word terms, or a complete sentence) users have provided that relate to the subject of their inquiry.
- E.g., Google, Bing, and Yahoo.

- Popular term for information retrieval system

- The two metrics to evaluate search engines:
  - Effectiveness/quality in finding the right documents/pages.
  - Efficiency/speed of returning a response quickly.

Search Engines

Structure of a Typical Internet Search Engine



Search Engine

Top 15 Most Popular Search Engines

Rank	Name	Estimated Unique Monthly Visitors
1	Google	1,600,000,000
2	Bing	400,000,000
3	Yahoo! Search	300,000,000
4	Ask	245,000,000
5	AOL Search	125,000,000
6	Wow	100,000,000
7	WebCrawler	65,000,000
8	MyWebSearch	60,000,000
9	Infospace	24,000,000
10	Info	13,500,000
11	DuckDuckGo	11,000,000
12	Contenko	10,500,000
13	Dogpile	7,500,000
14	Alhea	4,000,000
15	ixQuick	1,000,000

Anatomy of a Search Engine

- **Development Cycle** (two main components)
  - Web crawler: find and fetch web pages
  - Document indexer: process the web pages or files and place them into the document database.
  - Steps:
    1. Preprocessing The Documents
       - Collecting, formatting, and storing.
    2. Parsing The Documents
    3. Creating The Term-by-document Matrix
       - word/term values: numeric, binary, TF/IDF

Anatomy of a Search Engine

- **Response Cycle**
  - Query Analyzer
    - Responsible for receiving a search request from the and converting it into a standardized data structure.
  - Document Matcher/Ranker
    - The structured query data are matched against the document database to find the most relevant documents/pages and rank them in the order of relevance/importance.
    - Old approach - keyword match
    - Google approach - PageRank
      - Algorithmic way to rank order documents/pages based on their relevance and value/importance.

Search Engine Optimization

- SEO is the intentional activity of affecting the visibility of an e-commerce site or a Web site in a search engine’s natural (unpaid or organic) search results.

- Part of Internet marketing strategy:
  - considers how search engines work
  - what people search for
  - the actual search terms or keywords typed into search engines
  - which search engines are preferred by their targeted audience

- Optimizing a Web site can involve editing its content, HTML, and associated coding to both increase its relevance to specific keywords and to remove barriers to the indexing activities of search engines.

Methods for Search Engine Optimization

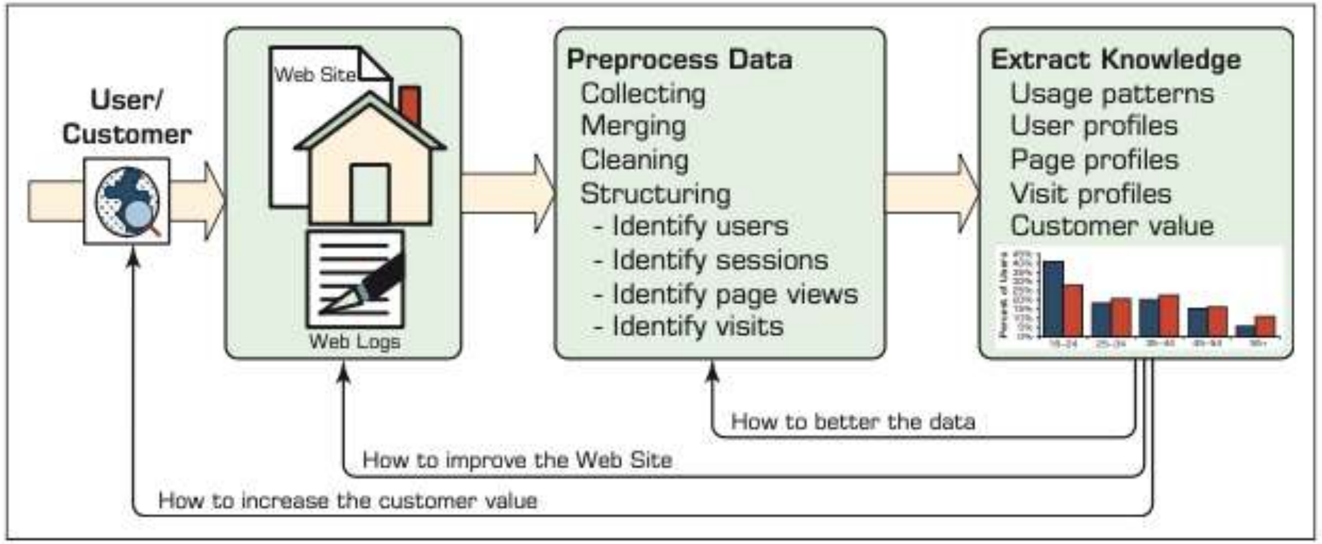
- **White-Hat SEO** search engines recommend technique
  - tend to produce results that based on good site design
  - conforms to the search engine’s guidelines and involves no deception
  - ensuring that the content a search engine indexes and ranks is the same content a user will see.

- **Black-hat SEO** search engines do not approve
  - attempts to improve rankings in ways that are not approved by the search engines or involve deception.
  - spamdexing (a.k.a search spam, search engine spam, or search engine poisoning)

Web Usage Mining (Web Analytics)

Web usage mining is the extraction of useful information from data generated through Web page visits and transactions (e.g., server logs, cookies, page metadata).

- Analysis of the information collected by Web servers can help us better understand user behaviour → clickstream analysis
  - E.g., knowing when visitors access a site, determining where to place online advertisements.



Extraction of Knowledge from Web Usage Data

Web Analytics Technologies

- Web analytics tools help to measure, collect, and analyze Internet data to
  - better understand and optimize web usage and traffic.
  - to assess and improve the effectiveness of e-commerce Web sites.
  - help companies measure the results of traditional print or broadcast advertising campaigns

Two main categories of Web analytics:

- **Off-site Web analytics:** refers to Web measurement and analysis about you and your products that take place outside the Web site
  - E.g., Web site's potential audience, share of voice (visibility or word of mouth), and buzz (comments or opinions) that is happening on the Internet.
- **On-site Web analytics:** measure visitors' behavior once they are on the Web site.
  - E.g., drivers and conversions.
  - Google Analytics is the most widely used on-site Web analytics service.
  - to method are used to produce Web traffic reports: server log file analysis and page tagging.

Web Analytics Metrics

- Provide nearly real-time valuable marketing data.
- Metrics categories:
  - Web site usability: How were they using my Web site?
  - Traffic sources: Where did they come from?
  - Visitor profiles: What do my visitors look like?
  - Conversion statistics: What does it all mean for the business?

Web Analytics Metrics

Web Site Usability

1. Page views
2. Time on site
3. Downloads
4. Click map
5. Click paths

Visitor Profiles

1. Keywords
2. Content groupings
3. Geography
4. Time of day
5. Landing page

Traffic Sources

1. Referral Web sites
2. Search engines
3. Direct
4. Offline campaigns
5. Online campaigns

Conversion Statistics

1. New visitors
2. Returning visitors
3. Leads
4. Sales/conversions
5. Abandonment rates

7.10 Social Analytics

- Social Network Analysis
- Social Network Analysis Metrics
- Social media & analytics

Social Network Analysis

- Social network - social structure composed of individuals (or groups/organizations) linked to one another with some type of connections/relationships
  - Help study relationships between individuals, groups, organizations

- Social analytics- analyses of social structure and dynamics.

- Interdisciplinary field that emerged from social psychology, sociology, statistics, and graph theory.

- Typical social network types
  - Communication networks, community networks, criminal networks, and innovation networks.

Social Network Analysis Metrics

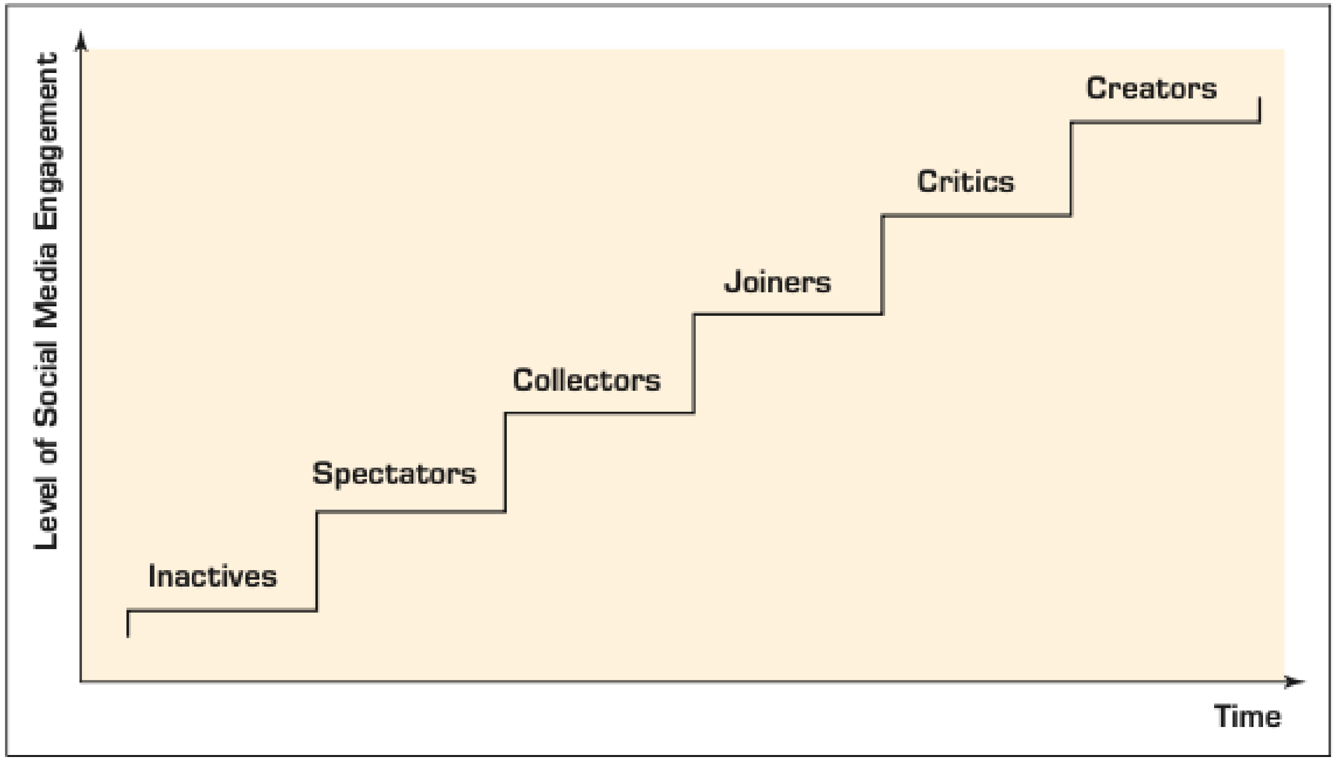
- Connections
  - Homophily
  - Multiplexity
  - Mutuality/reciprocity
  - Network closure
  - Propinquity
- Segmentation
  - Cliques and social circles
  - Clustering coefficient
  - Cohesion
- Distribution
  - Bridge
  - Centrality
  - Density
  - Structural holes
  - Tie strength

Social Media

- Enabling technologies of social interactions among people in virtual communities and networks (e.g., forums, social blogs, microblogging, wikis, social networks)
- Build on technological foundations of Web 2.0
- Social media are different from traditional/industrial media, such as newspapers, television, and film
  - Quality
  - Reach
  - Frequency
  - Accessibility
  - Usability
  - Immediacy
  - Updatability

Social Media

How Do People Use Social Media?



Six different engagement levels

Social media analytics

- Systematic and scientific ways to consume the vast amount of content created by Web-based social media outlets tools, and techniques for the betterment of an organization’s competitiveness.

- Rapidly growing movement in analytics.

- Measuring the social media impact:
    - Descriptive analytics
    - Social network analysis
    - Advanced analytics

Social media analytics

Best Practices in Social Media Analytics

- Think of measurement as a guidance system, not a rating system
- Track the elusive sentiment
- Continuously improve the accuracy of text analysis
- Look at the ripple effect
- Look beyond the brand
- Identify your most powerful influencers
- Look closely at the accuracy of your analytic tool
- Incorporate social media intelligence into planning

Main Reference

- **Chapter 7:** “Text Mining, Sentiment Analysis, and Social Analytics” from *“Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”.*

Week self-review exercises

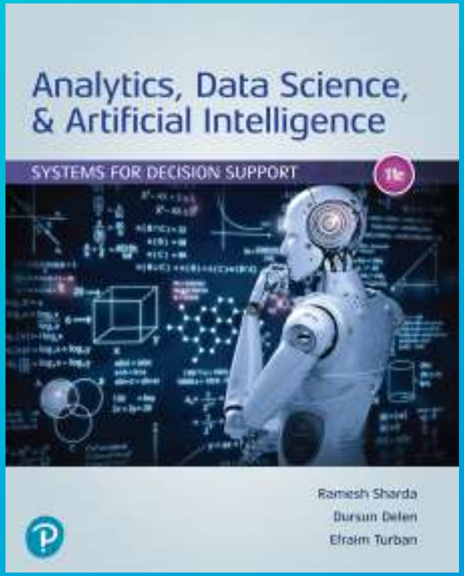
- **Application case 7.1 – 7.8** from from *“Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”*

Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

College of Computing and Informatics



Week 9

Chapter 6:Deep Learning and Cognitive Computing

Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Learn what deep learning is and how it is changing the world of computing.
2. Know the placement of deep learning within the broad family of artificial intelligence (AI) learning methods.
3. Understand how traditional “shallow” artificial neural networks (ANN) work.
4. Become familiar with the development and learning processes of ANN.
5. Develop an understanding of the methods to shed light into the ANN black box.
6. Know the underlying concept and methods for deep neural networks.
7. Become familiar with different types of deep learning methods.
8. Understand how convolutional neural networks (CNN) work.
9. Learn how recurrent neural networks (RNN) and long short-memory networks (LSTM) work.
10. Become familiar with the computer frameworks for implementing deep learning.
11. Know the foundational details about cognitive and learn how IBM Watson works and what types of application it can be used for.

Required Reading

- **Chapter 6**: "*Deep Learning and Cognitive Computing*" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".

Recommended Reading

- Black, D. (2018, January 23). AI Definitions: Machine Learning vs. Deep Learning vs. Cognitive Computing vs. Robotics vs. Strong AI. Datanami. https://www.datanami.com/2018/01/23/ai-definitions-machine-learning-vs-deep-learning-vs-cognitive-computing-vs-robotics-vs-strong-ai/

Recommended Video

- What is Cognitive AI? Cognitive Computing vs Artificial Intelligence (2020, Jan 15). [Video]. YouTube. https://www.youtube.com/watch?v=Zsl7ttA9Kcg

6.2 Introduction to Deep Learning

- Introduction to Deep Learning
- Classic Machine-Learning vs Deep Learning

Introduction to Deep Learning

- Deep learning is among the latest trends in AI that come with great expectations.

- The initial idea of deep learning goes back to the late 1980s.

- Goal: mimic the thought process of humans—using mathematical algorithms to learn from data pretty much the same way that humans learn (similar to those of the other machine-leaning methods).

- It has added the ability to automatically acquire the features required to accomplish highly complex and unstructured tasks (e.g. image recognition) to the classic machine-learning methods that contribute to the superior system performance.

- The recent emergence and popularity of deep learning can largely be attributed to very large data sets and rapidly advancing commuting infrastructures.

- Many deep learning applications have promised to make our life easier.
  - E.g., Google Home, Amazon’s Alexa, Google Translate, …)

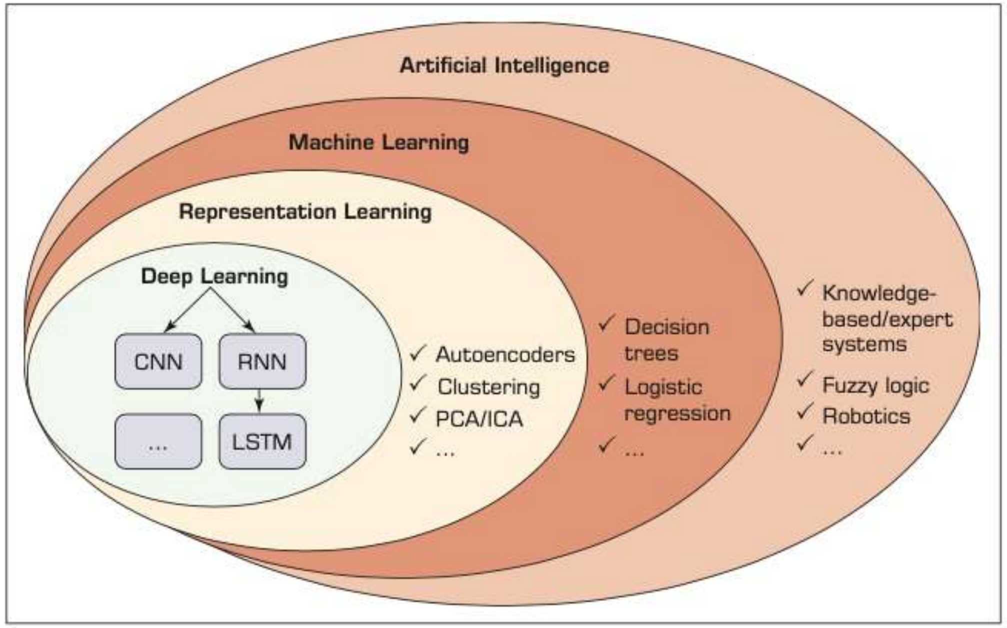
Introduction to Deep Learning

- Deep learning is an extension of neural networks with the idea that deep learning is able to deal with more complicated tasks with a higher level of sophistication.
- Neural networks are extended by employing many layers of connected neurons along with much larger data sets to automatically characterized variables and solve the problems.
- The initial idea of deep learning had to wait more than two decades until some advanced computational and technological infrastructure emerged, because of:
  1. Very high computational requirement.
  2. The need for very large data sets.

Introduction to Deep Learning

Placement of Deep Learning within the Overarching AI-Based Learning Methods

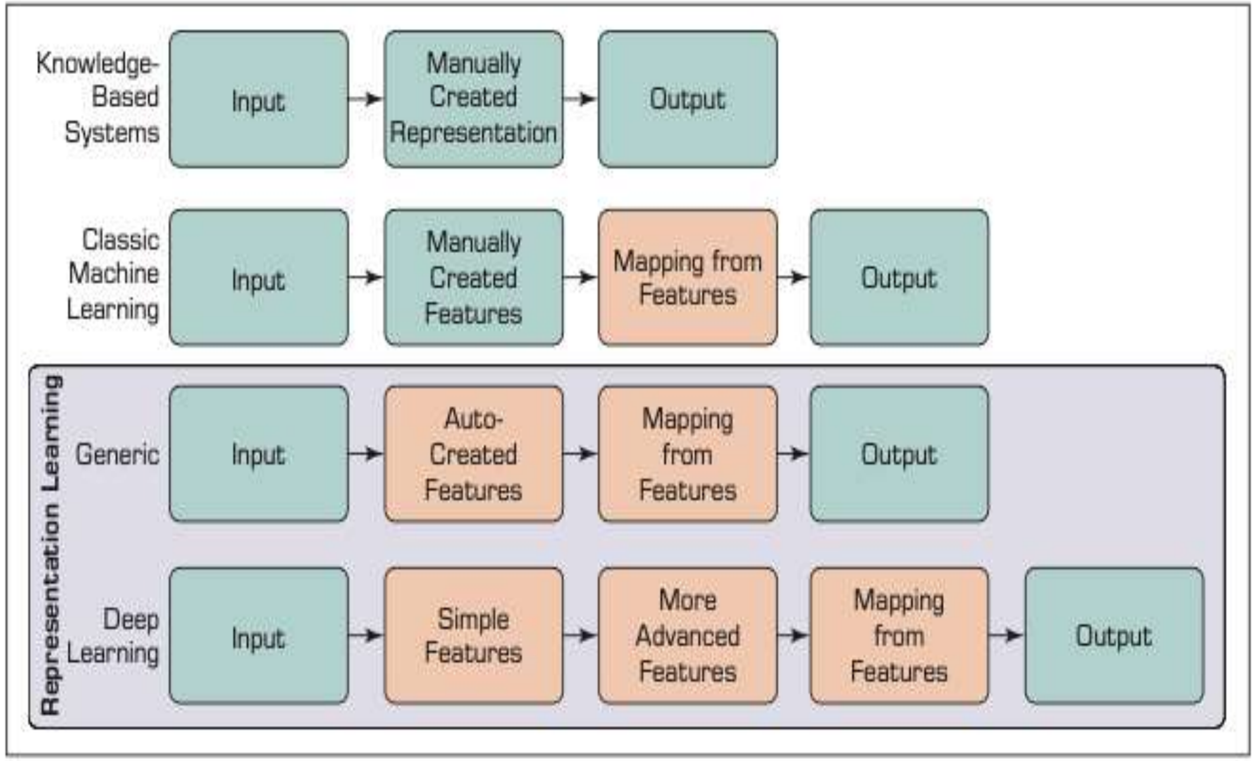
- Deep learning is categorized as part of the representation learning within the AI learning family of methods
- Representation learning focus on learning and discovering features by the system in addition to discovering the mapping from those features to the output/target.



Classic Machine-Learning vs Deep Learning

- In Knowledge-based systems and classic machine-learning methods, features (i.e., the representation) are created manually by data scientists to achieve the desired output.

- Deep learning enables the computer to derive some complex features from simple concepts that would be very effort intensive to be discovered by humans manually.

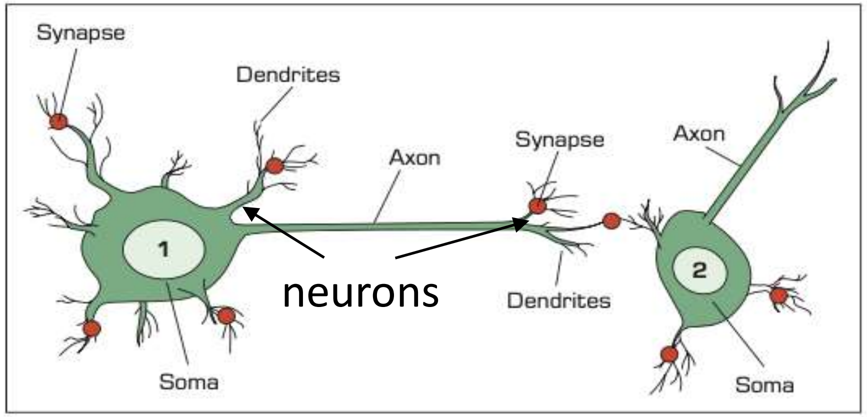


6.3 Basics of “Shallow” Neural Networks

- Artificial Neural Networks (ANN)
- Elements of an Artificial Neural Network
- Common Transfer Functions in Neural Networks

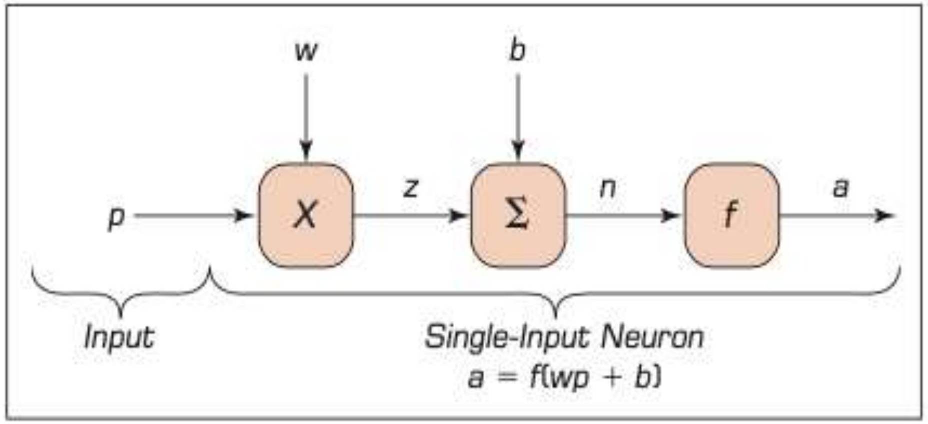
Artificial Neural Networks (ANN)

- The human brain has a set of billions of interconnected neurons that facilitate our thinking, learning, and understanding of the world around us.
- Artificial neural networks emulate the way the human brain works.
- The basic processing unit is a neuron. Multiple neurons are grouped into layers and linked together.



A Biological Neural Network: Two Interconnected Cells/Neurons.

VS



ANN with single neuron, single inputs and outputs

Processing Information in ANN

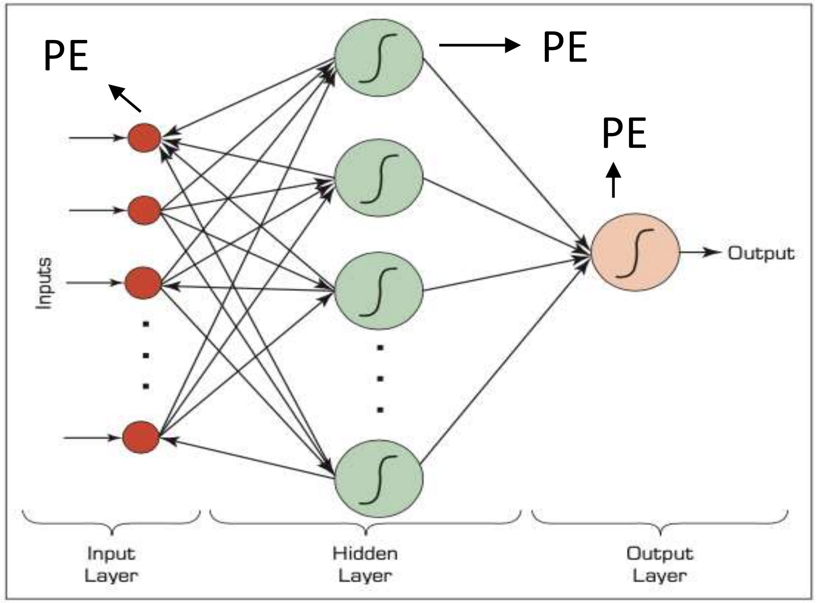
- The basic processing unit is a neuron (processing element – PE).

- PE: perform a set of predefined mathematical operations on the numerical values coming from the input or from the other neuron outputs to create and push out its own outputs.

- A neuron can have more than a single input *p*, each of the individual input values would have its own adjustable weight *w.*

- In a neural network, knowledge is stored in the weight associated with the connections between neurons.

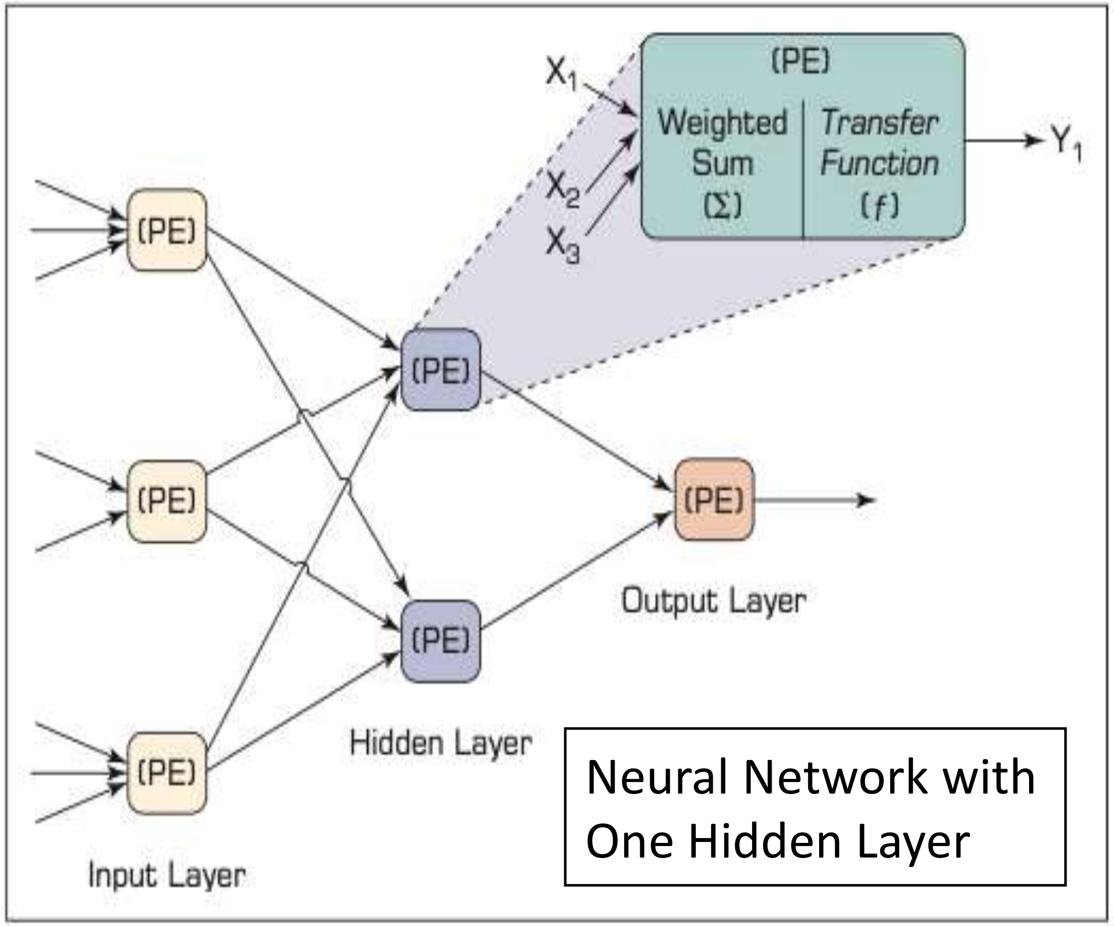
- Multiple neurons are grouped into layers and linked together.



Typical Neural Network with Three Layers and Eight Neurons.

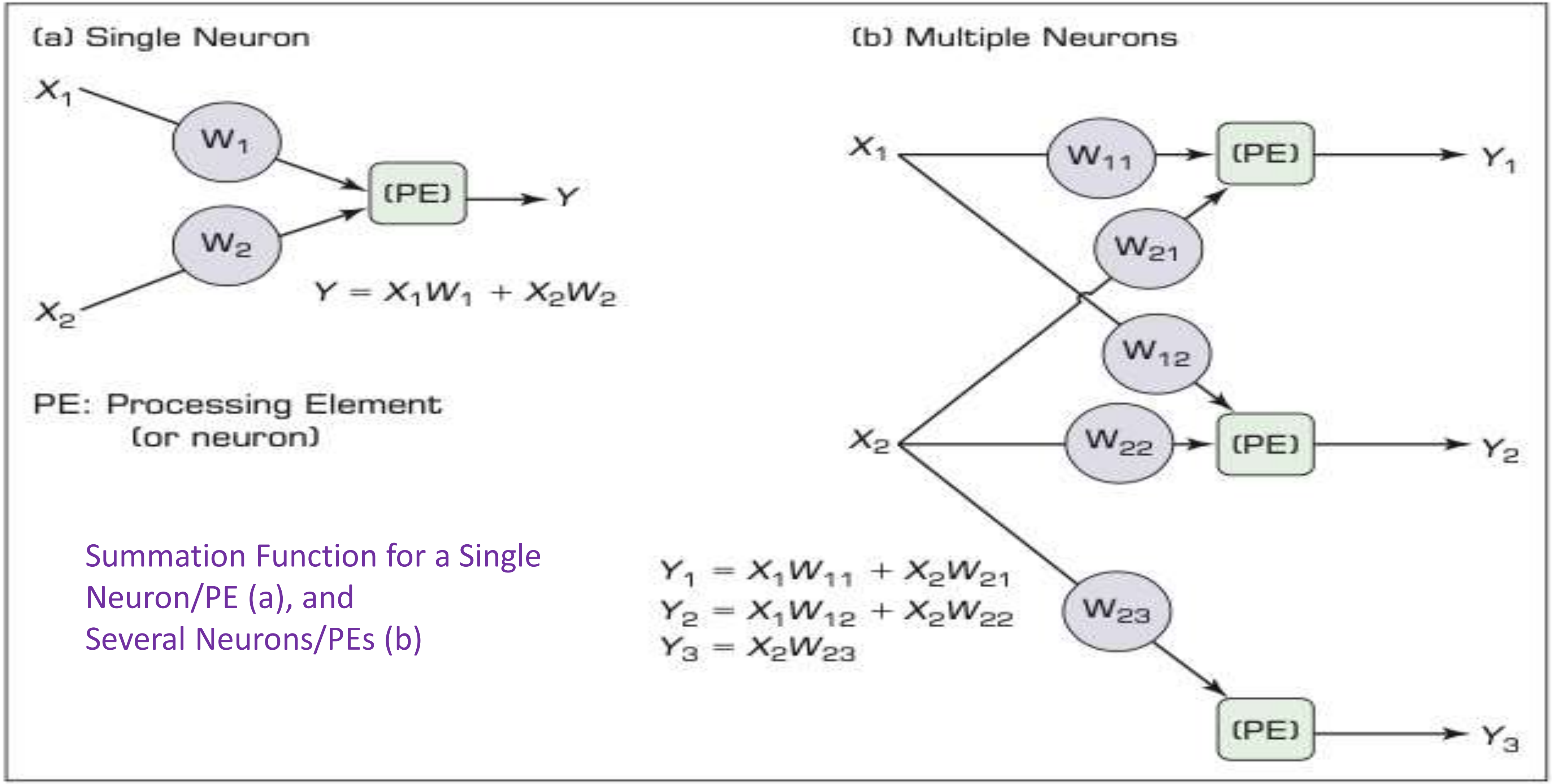
Elements of an Artificial Neural Network

- Processing element (PE)
- Network architecture
  - Hidden layers
  - Parallel processing
- Network information processing
  - Inputs
  - Outputs
  - Connection weights
  - Summation function
  - Transfer Function



Neural Network with One Hidden Layer

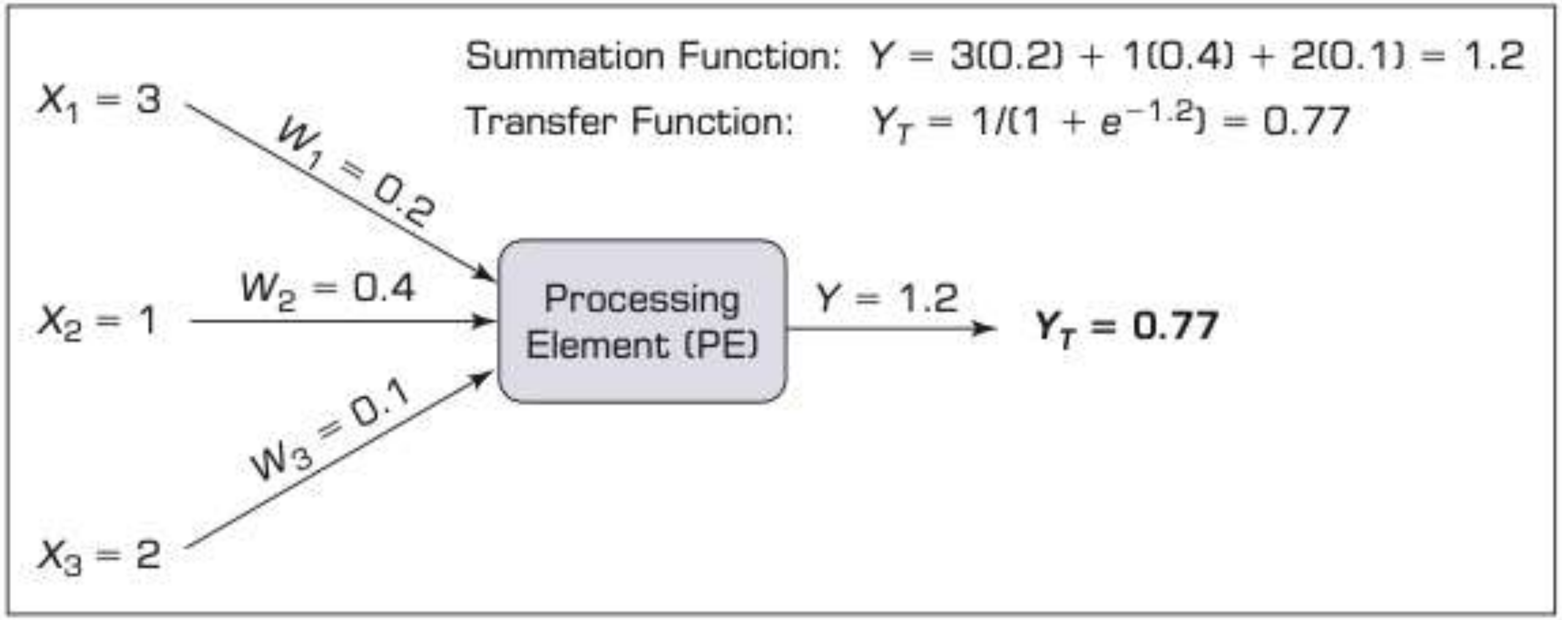
Elements of an Artificial Neural Network



Summation Function for a Single Neuron/PE (a), and
Several Neurons/PEs (b)

Elements of an Artificial Neural Network

- Various types of transfer functions are commonly used in the design of neural networks.
- Common Transfer Function types (Linear function, Sigmoid (log) function [0 1] and Tangent Hyperbolic function [-1 1]).
- Example of ANN Transfer Function (sigmoid-type activation function)

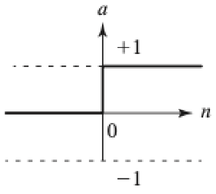
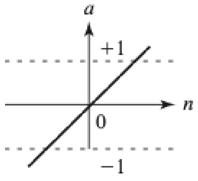
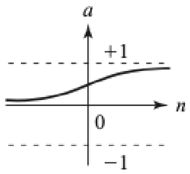
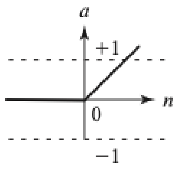


Common Transfer Functions in Neural Networks

- The selection of proper transfer functions for a network requires a broad knowledge of neural networks (e.g. characteristics of the data as well as the specific purpose for which the network is created).

- There are some guidelines for choosing the appropriate transfer function especially for the neurons located at the output layer of the network.

- E.g., if the nature of the output for a model is binary, it is advised to use Sigmoid transfer functions at the output layer so that it produces an output between 0 and 1.

Transfer Function	Form	Operation
Hard limit	 $a = hardlim\ (n)$	$a = +1$ if $n > 0$ $a = 0$ if $n < 0$
Linear	 $a = purelin\ (n)$	$a = n$
Log-Sigmoid	 $a = logsig\ (n)$	$a = \frac{1}{1 + e^{-n}}$
Positive linear (a.k.a. rectified linear or ReLU)	 $a = poslin(n)$	$a = n$ if $n > 0$ $a = 0$ if $n < 0$

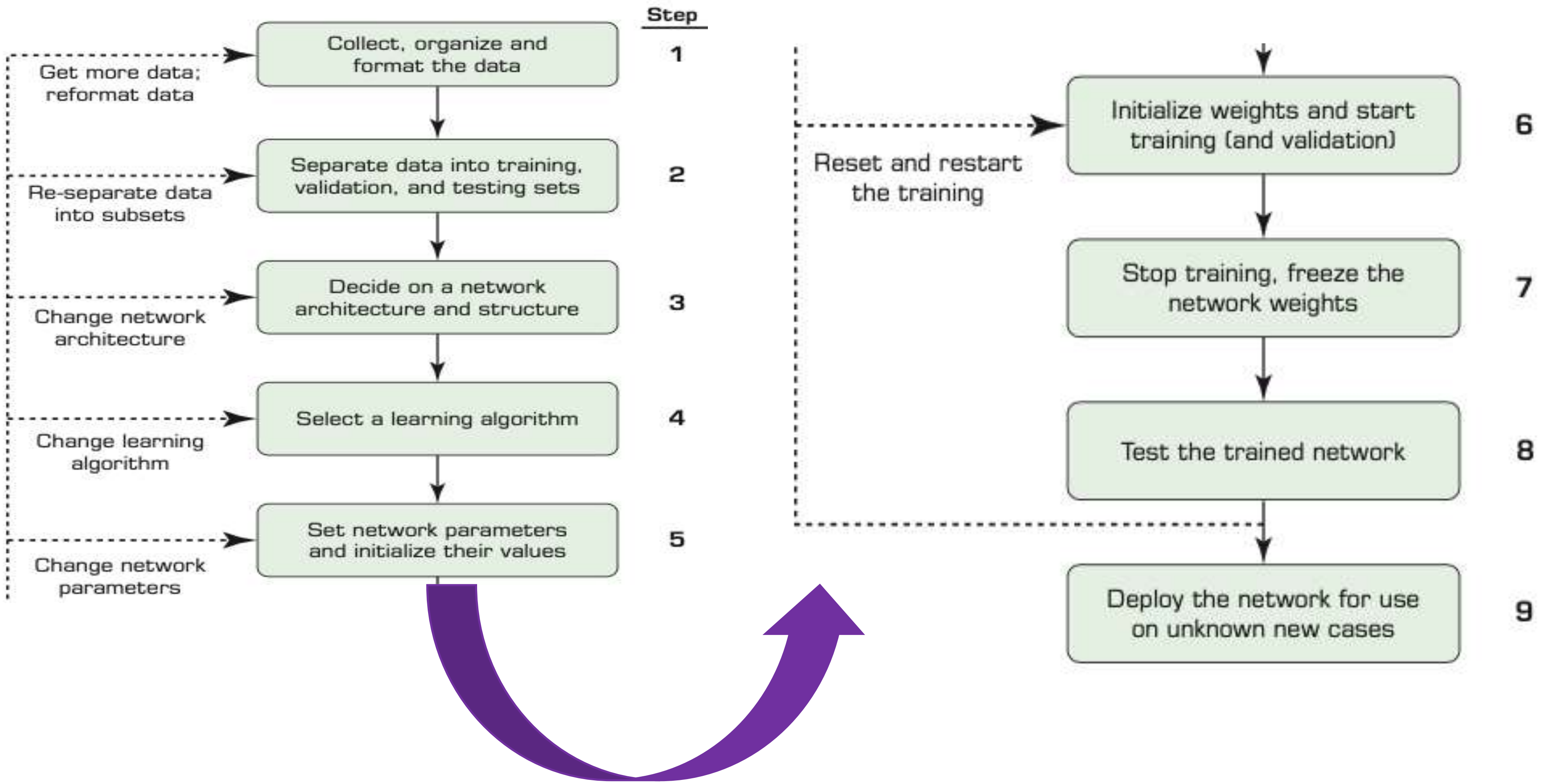
some of the most common transfer functions and their corresponding operations

6.4 Process of Developing Neural Network–Based Systems

- Development Process of an ANN Model
- Learning Process in ANN
- Backpropagation Learning for ANN
- Overfitting in ANN

Development Process of an ANN Model

- Developing neural network–based systems requires a step-by-step process.

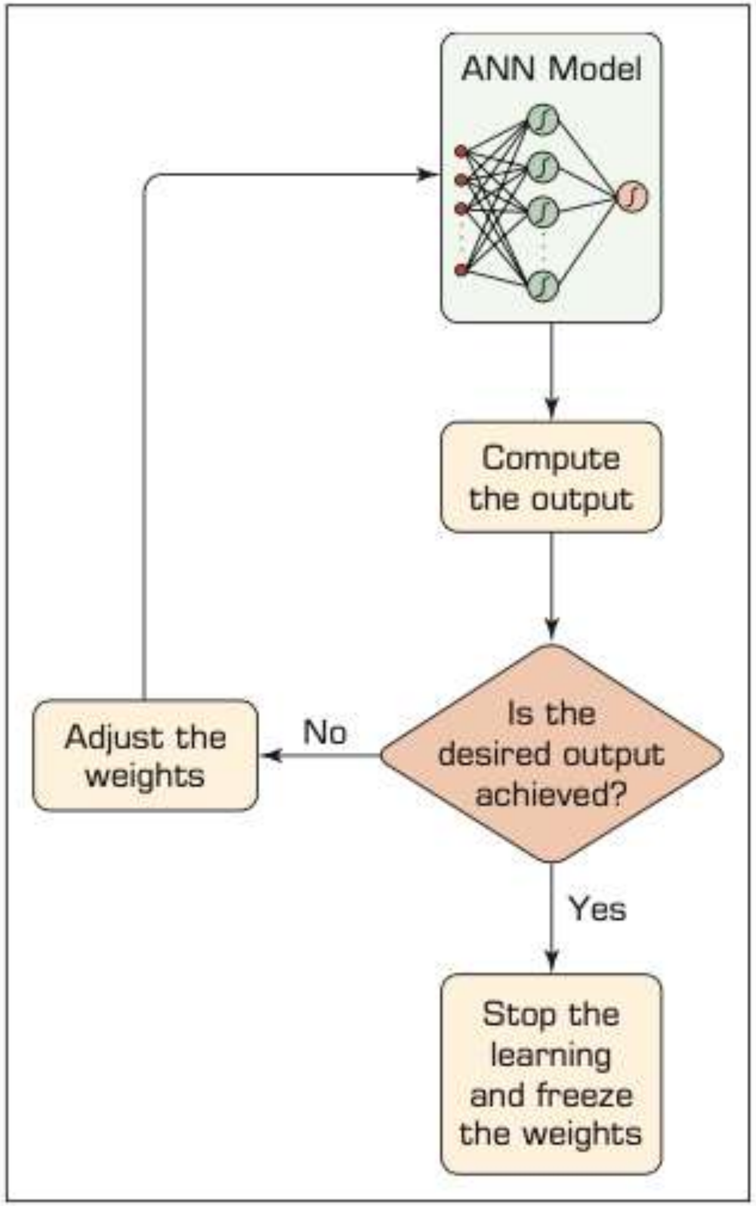


Learning Process in ANN

- A supervised learning process.

- The learning process is inductive; that is, connection weights are derived from existing cases.

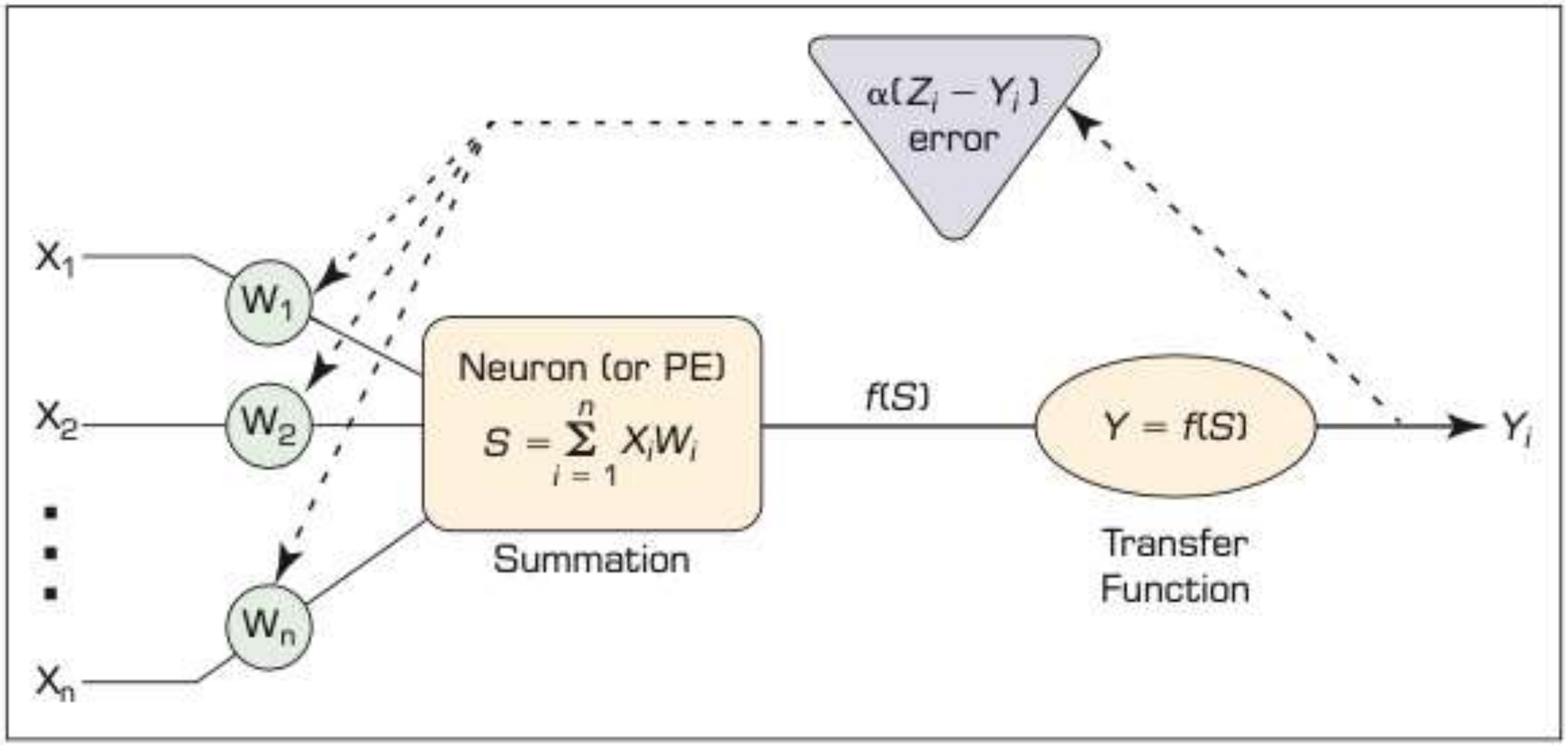
- **The usual process of learning involves three tasks:**
  - Compute temporary outputs.
  - Compare outputs with desired targets.
  - Adjust the weights and repeat the process.



Supervised Learning Process of an ANN.

Backpropagation Learning for ANN

- Backpropagation is the most popular supervised learning paradigm for ANN.



Backpropagation of Error for a Single Neuron

Backpropagation Learning for ANN

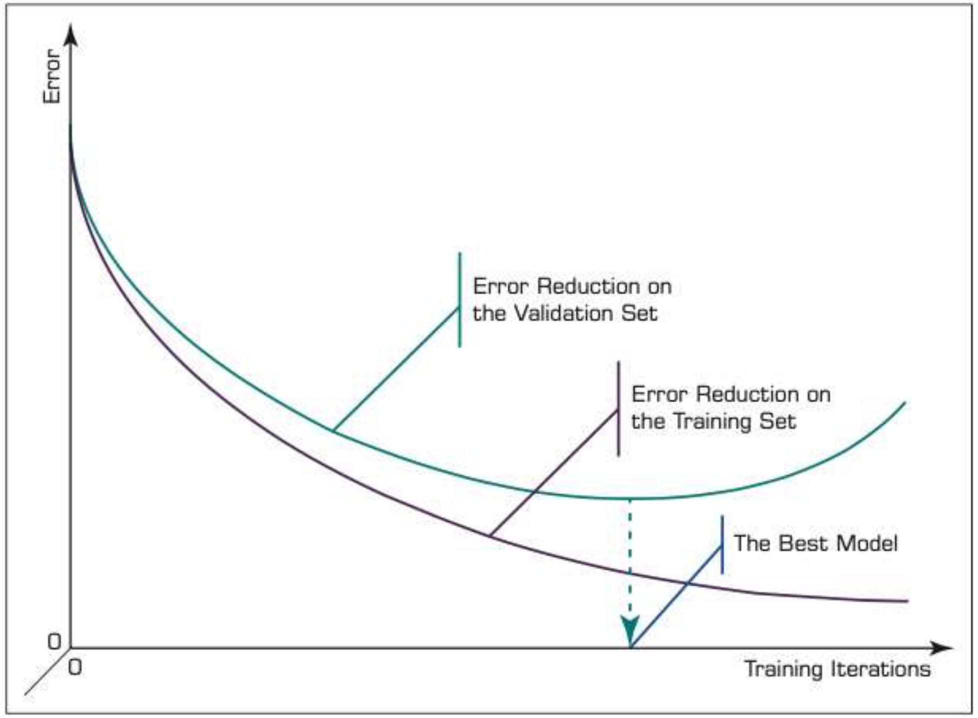
The learning algorithm procedure:

1. Initialize weights with random values and set other parameters.
2. Read in the input vector and the desired output.
3. Compute the actual output via the calculations, working forward through the layers.
4. Compute the error.
5. Change the weights by working backward from the output layer through the hidden layers.

Overfitting in ANN

- Occurs when neural networks are trained for a large number of iterations with relatively small data sets.

- To prevent overfitting, the training process is controlled by an assessment process using a separate validation data set.



Overfitting in ANN—Gradually Changing Error Rates in the Training and Validation Data Sets As the Number of Iterations Increases.

6.5 ILLUMINATING THE BLACK BOX OF ANN

- Sensitivity Analysis on ANN Models

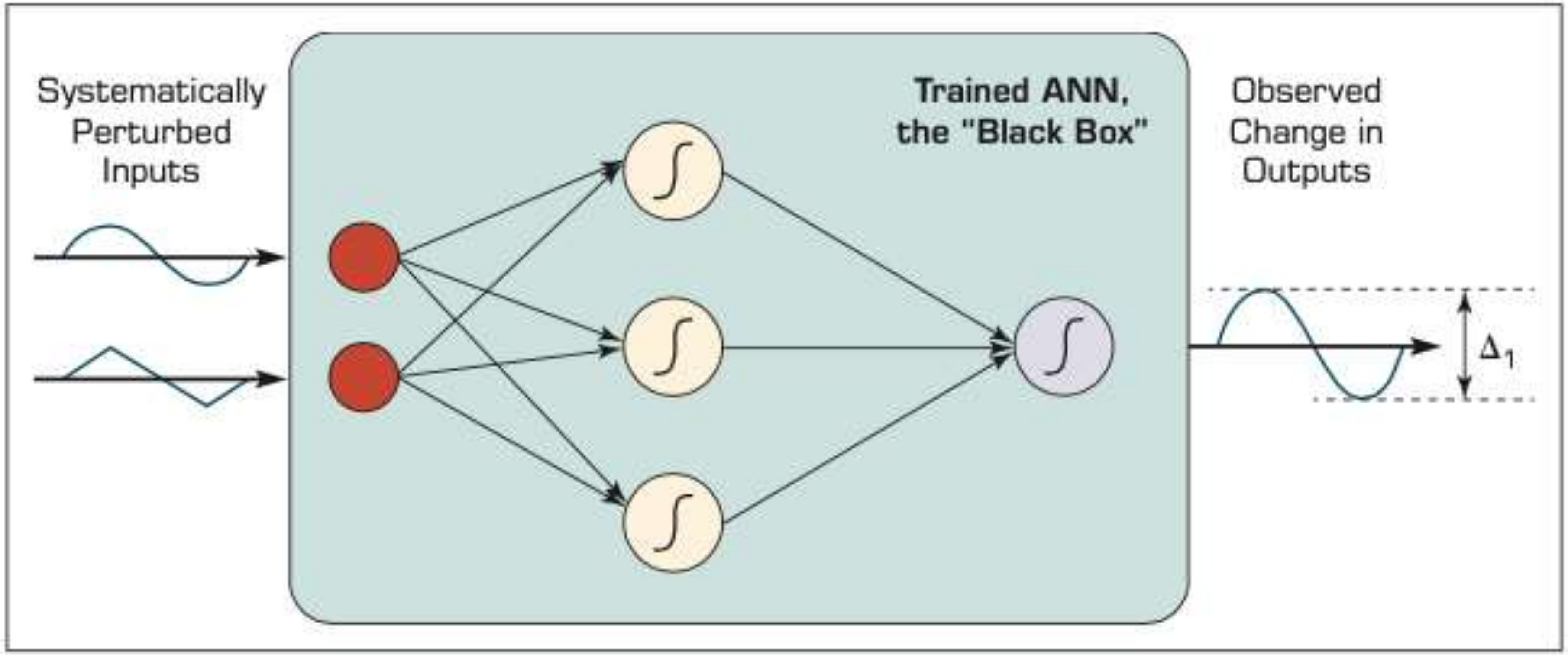
Sensitivity Analysis on ANN Models

- ANNs are known as black-box models.
  - But, “how the model does what it does?”
- ANNs lack of explanation/transparency -> black-box syndrome!
- To shed light into the black-box syndrome sensitivity analysis is applied.

- Sensitivity analysis:

1. Preformed on a trained ANN
2. Perturbed the inputs to the network systematically within the allowable value ranges.
3. The corresponding change in the output is recorded for each and every input variable.
4. The relative importance of input variables are illustrated in the result.

Sensitivity Analysis on ANN Models



- Sensitivity analysis extract the cause-and-effect relationships among the inputs and the outputs of a trained neural network model.

6.6 Deep Neural Networks

- Deep Neural Networks

- Feedforward Multilayer Perceptron (MLP)

6.7 Convolutional Neural Networks (CNNs)

- Image Processing Using Convolutional Networks

- Text Processing Using Convolutional Networks

6.8 Recurrent Networks And Long Short-term Memory Networks (RNNs) and Long Short-Term Memory Networks (LSTM)

Deep Neural Networks

- Most neural network applications involved network architectures with only a few hidden layers and a limited number of neurons in each layer.

- Deep neural networks broke the generally accepted notion of "no more than two hidden layers are needed to formulate complex prediction problems."

- They promote increasing the hidden layer to arbitrarily large numbers to better represent the complexity in the data set.

- Different types of deep networks involve various modifications to the architecture of standard neural networks.
  - Typically equipped with distinct capabilities of dealing with particular data types for advanced purposes (e.g. image or text processing).

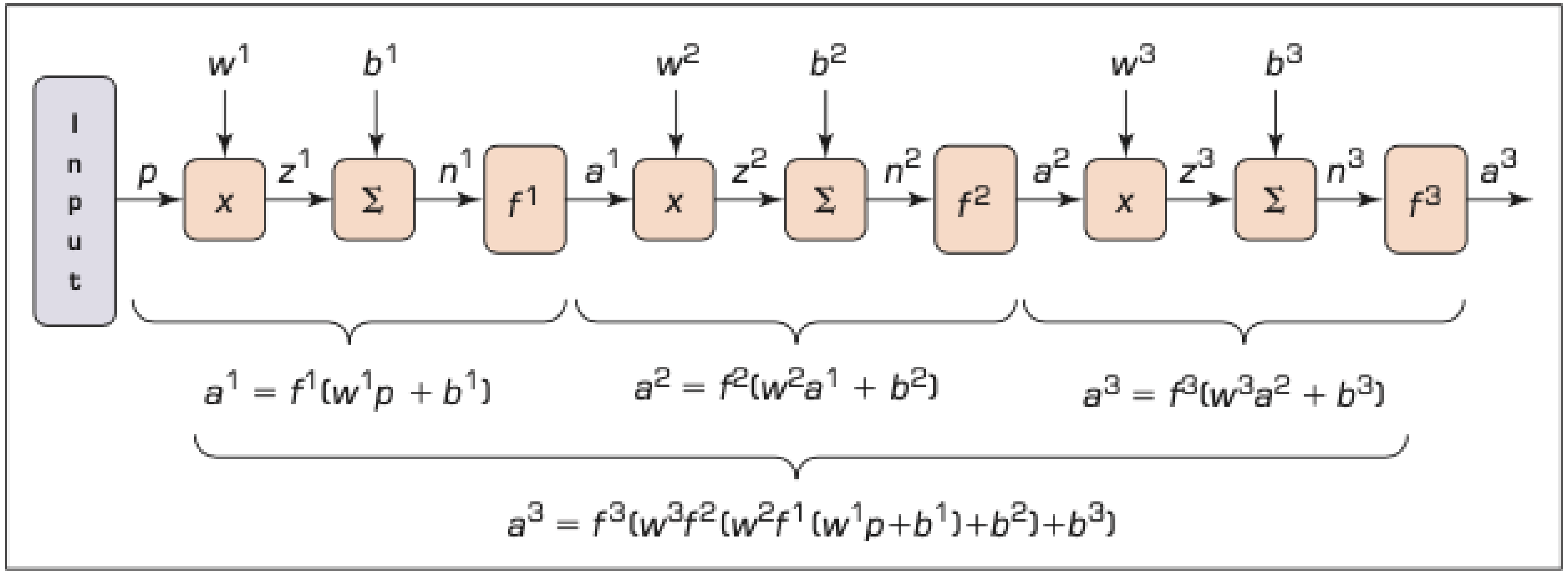
Feedforward Multilayer Perceptron (MLP)

- MLP deep networks (a.k.a deep feedforward networks) are the most general type of deep networks.
- MLP Consists of an input layer, an output layer, and a number of hidden layers.
- The nodes in one layer are connected to the nodes in the next layer.
- Each node at the input layer typically represents a single attribute that may affect the prediction.
- The flow of information is always forwarding and no feedback connections, hence it is called “called feedforward network”.

More Hidden Layers versus More Neurons?

- it is still an open research question, practically using more layers in a network seems to be more and computationally more efficient than using many neurons in a few layers.

Feedforward Multilayer Perceptron (MLP)



The First Three Layers in a Typical MLP Network.

Convolutional Neural Networks (CNNs)

- CNNs (by LeCun et al., 1989) are arguably the most popular and most successful deep learning methods.

- A variations of the deep MLP architecture.

- CNNs initially designed for computer vision applications (e.g., image processing, video processing, text recognition)

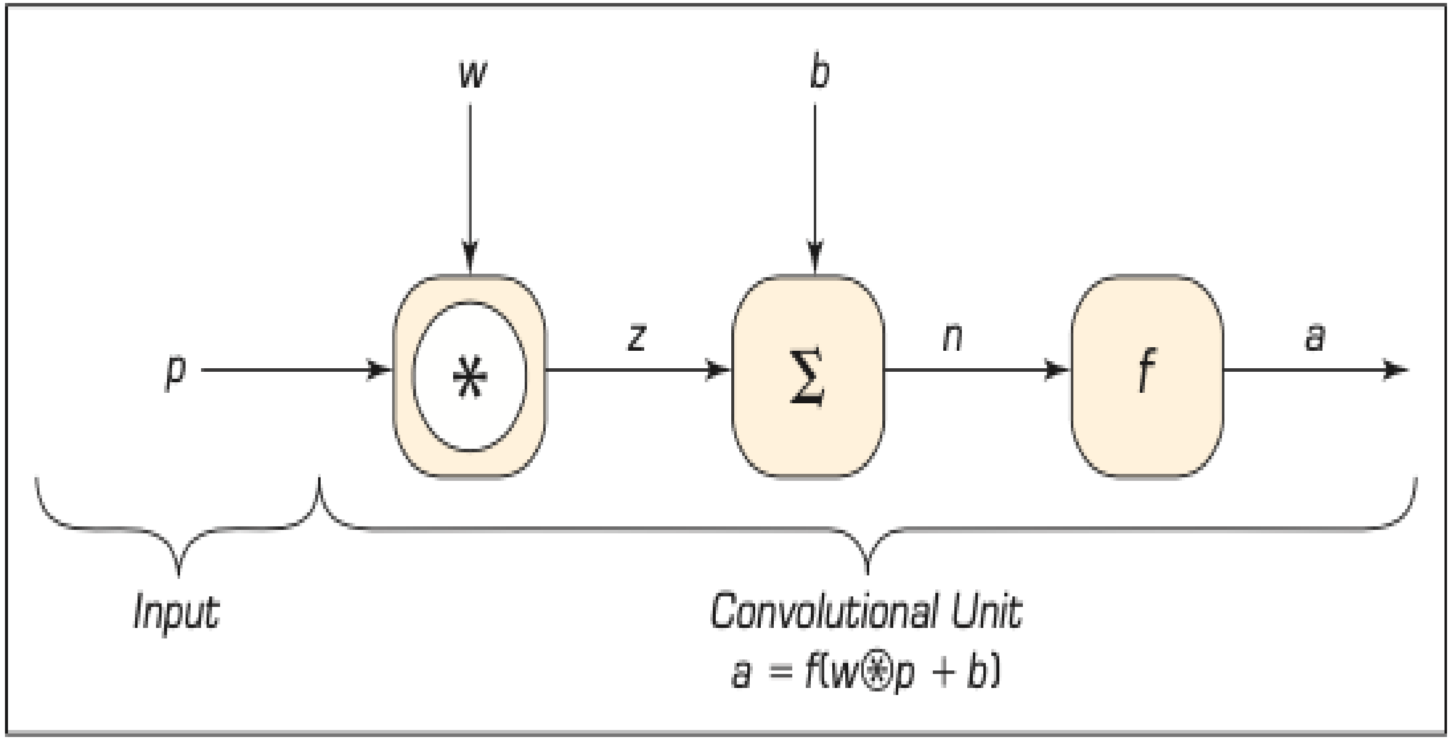
- CNNs are also applicable to nonimage or non-text data sets.

- The main characteristic of the convolutional networks is having at least one layer involving a convolution weight function instead of general matrix multiplication.

- A layer containing a convolution function in a CNN is called a convolution layer.

A typical convolutional unit

- Convolution layer shown by the symbol * is a linear operation that essentially aims at extracting simple patterns from sophisticated data patterns.



Convolution Function

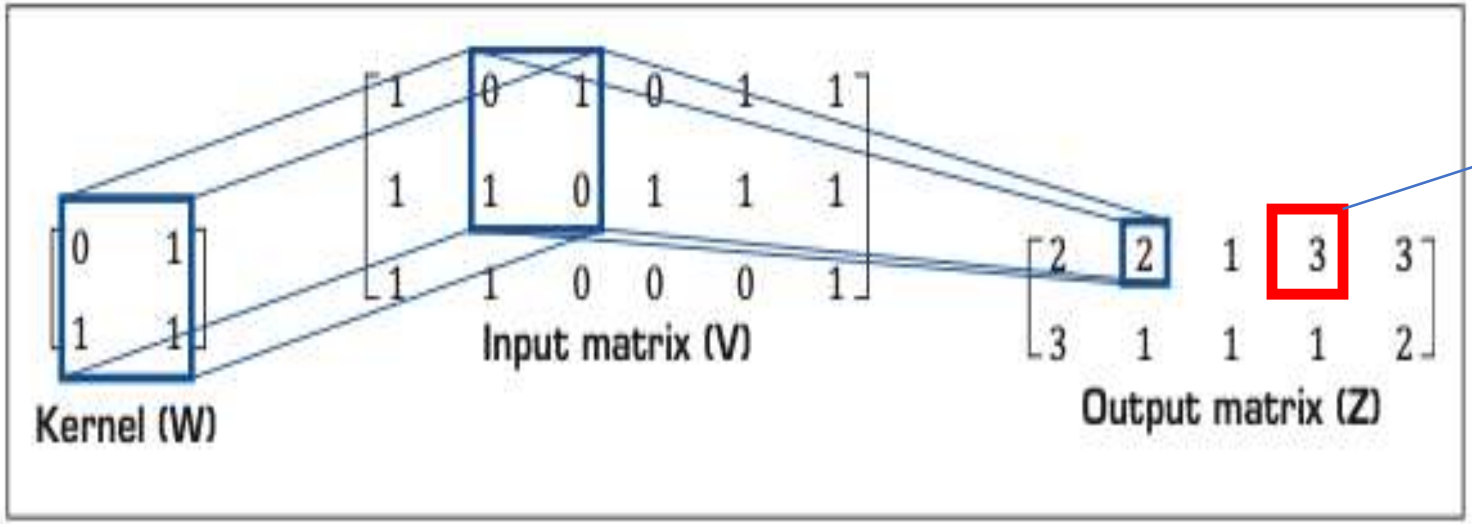
- In a typical neural network, having only a single input layer for an image-processing task with images of size 150 * 150 pixels requires thousands of weight parameters to be defined and trained, which dramatically increase the required time and processing power to train a network.

- Solution: convolution function.
  - In CNNs, the convolution function is a method to address the issue of having too many network weight parameters by introducing the notion of parameter sharing that offer computational efficiency.
  - Convolution functions can extract simple patterns like the existence of horizontal or vertical lines or edges in different parts of the picture.

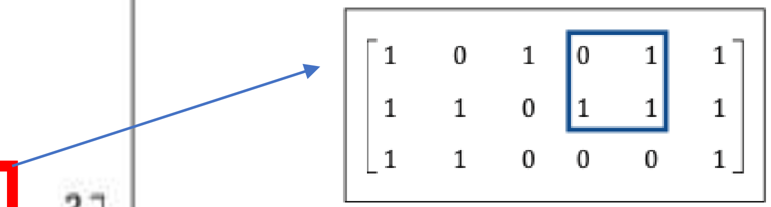
Cont..

- Instead of having a weight for each input in a convolution layer, there is a set of weights referred to as the convolution kernel or filter.

- Filter is shared between inputs and moves around the input matrix to produce the outputs.



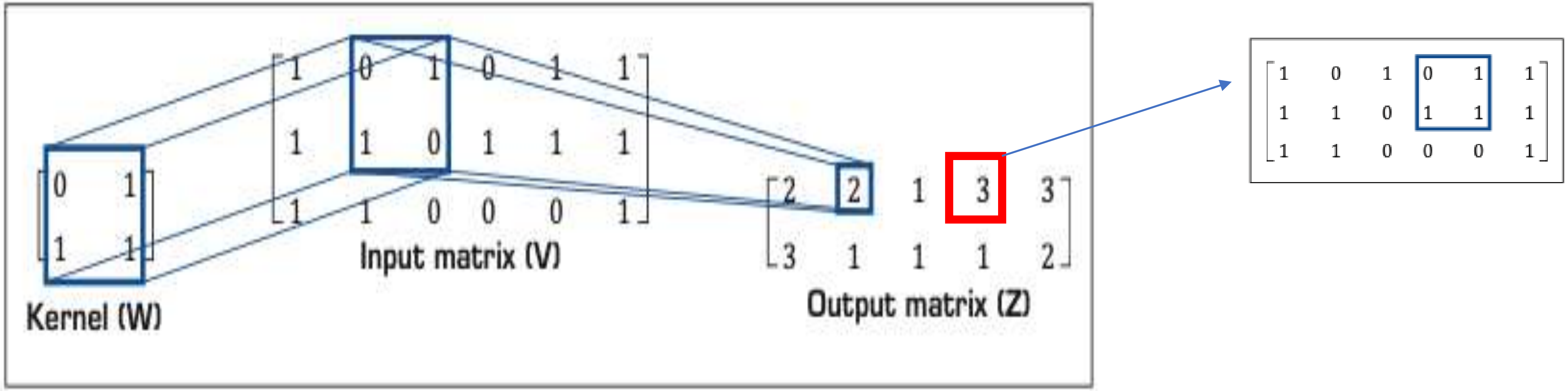
Convolution of a 2 : 2 Kernel by a 3 : 6 Input Matrix.



The Output of Convolution Operation Is Maximized When the Kernel Exactly Matches the Part of Input Matrix That Is Being Convoluted by.

Cont...

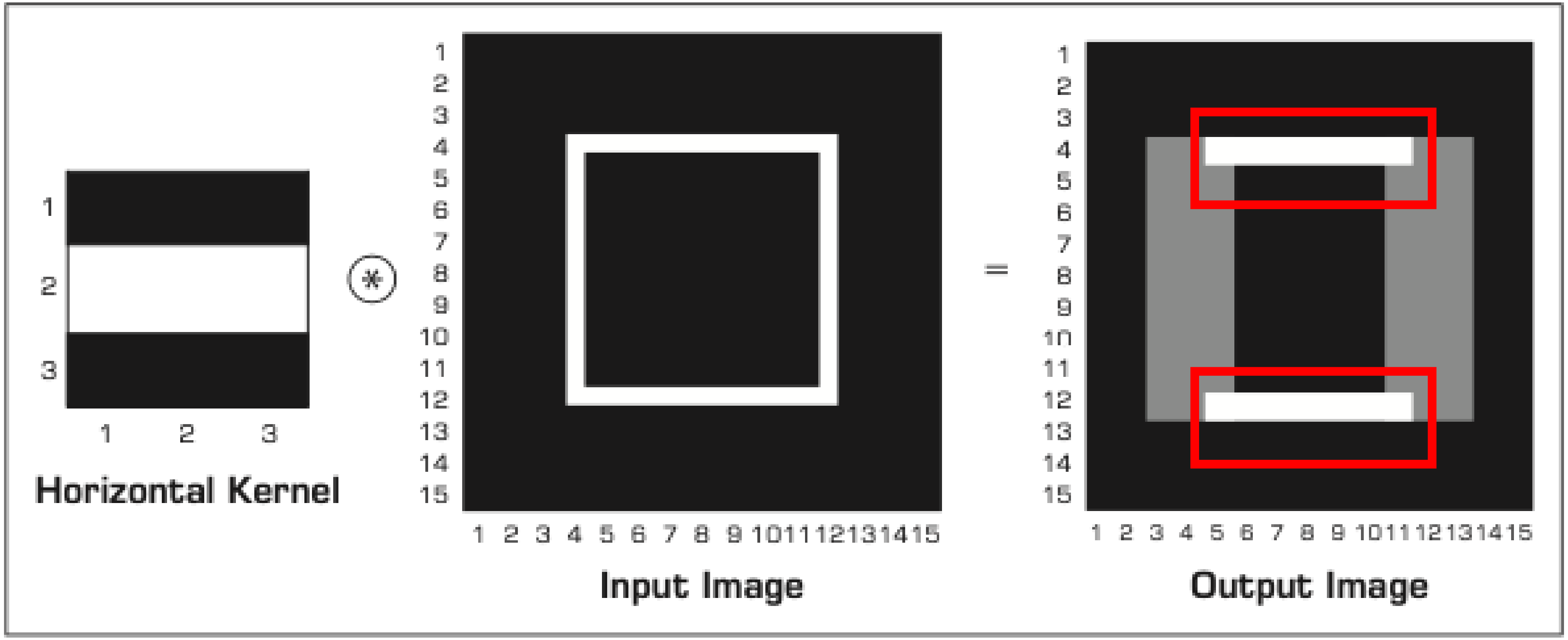
- The output of convolution operation is maximized when the kernel exactly matches the part of input matrix that is being convoluted by.



Convolution of a 2 : 2 Kernel by a 3 : 6 Input Matrix.

Cont…

- Example of Using Convolution for Extracting Features (Horizontal Lines in This Example) from Images.



The horizontal kernel produces an output in which the location of horizontal lines (as a feature) in the original input image is identified.

Cont…

- A convolution layer in a network will have a property called equivariance for translation purposes

- Equivariance: any changes in the input will lead to a change in the output in the same way.

- E.g., moving an object in the input image by 10 pixels in a particular direction will lead to moving its representation in the output image by 10 pixels in the same direction.

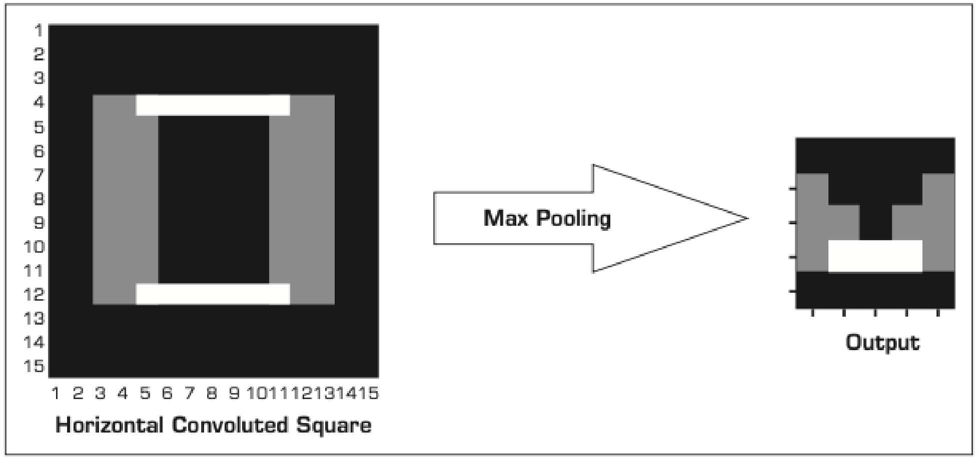
- This feature is especially useful for analyzing time-series data using convolutional networks where convolution can produce a kind of timeline that shows when each feature appears in the input.

Pooling

- A convolution layer is often followed by another layer known as the pooling (a.k.a. sub-sampling) layer.

- The purpose of a pooling layer is to consolidate elements in the input matrix in order to produce a smaller output matrix while maintaining the important features.

- There are different types of pooling operation (e.g., average pooling, max pooling ,...).

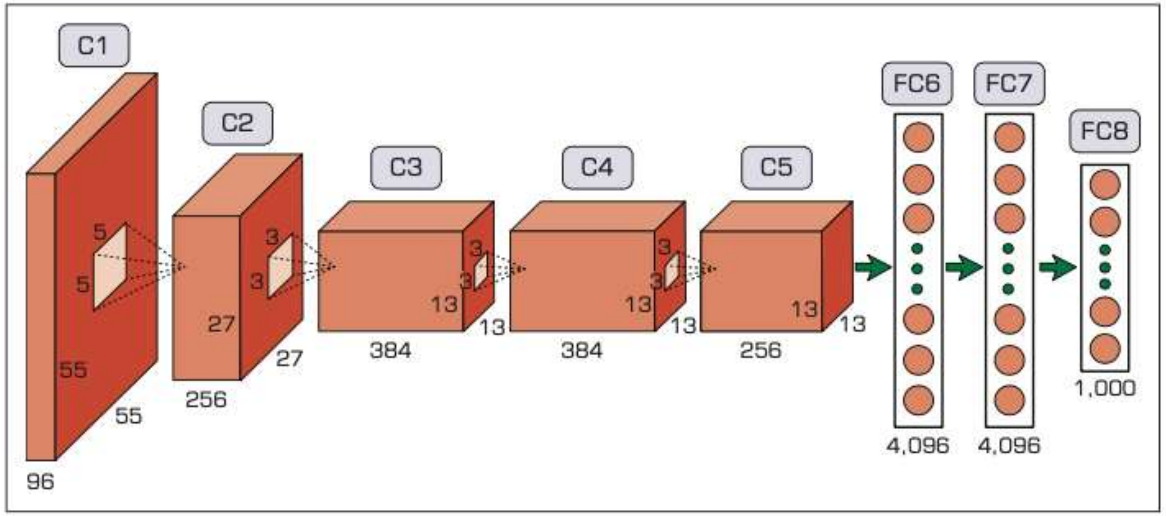


An Example of Applying Max Pooling on an Output Image to Reduce Its Size

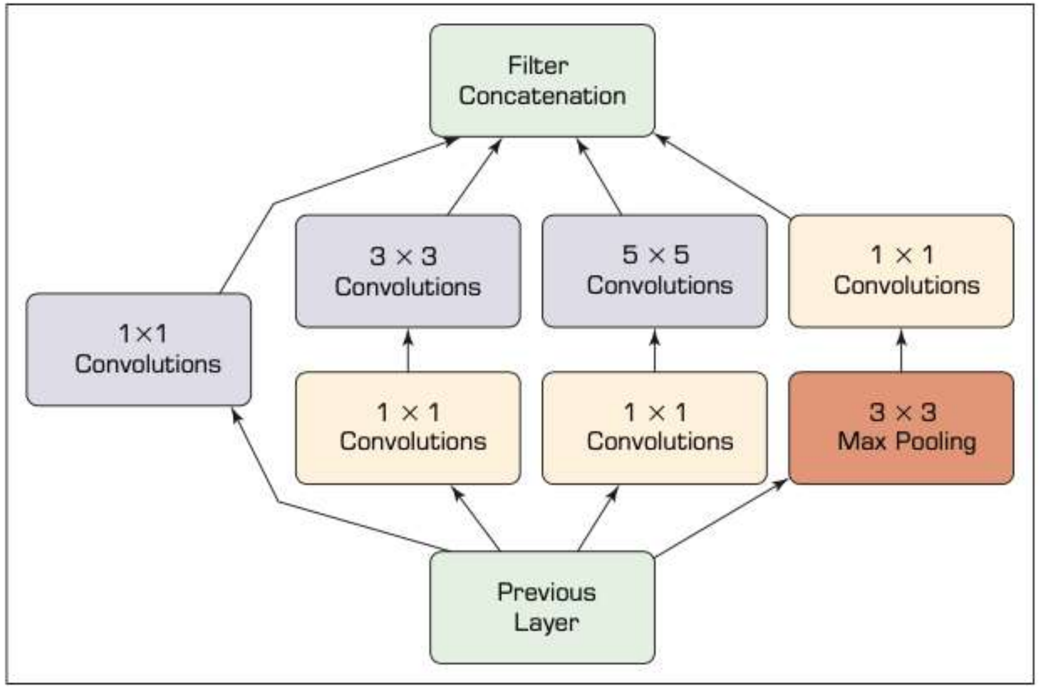
Image Processing Using Convolutional Networks

- Image classification networks traditionally involve two pipelines: visual feature extraction and image classification.
- ImageNet is an ongoing research project that provides researchers with a large database of images, each linked to a set of synonym words (known as synset) from WordNet (a word hierarchy database) - most widely used for benchmarking.
- AlexNet is one of the first convolutional networks designed for image classification using the ImageNet data set. Its success rapidly popularized the use and reputation of CNNs.
- GoogLeNet (a.k.a. Inception), a deep convolutional network architecture designed by Google researchers, was the winning architecture at ILSVRC 2014.
- Google Lens is an app that uses deep learning artificial neural network algorithms to deliver information about the images captured by users from their nearby objects.

Image Processing Using Convolutional Networks



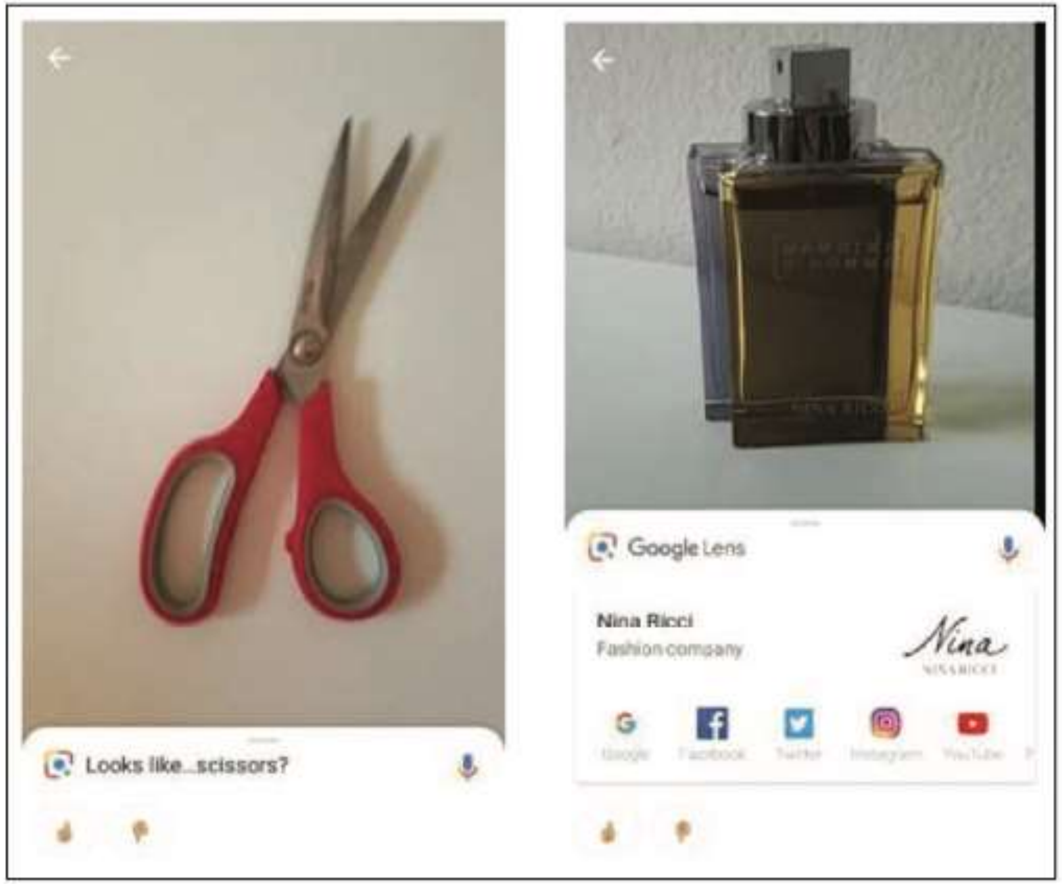
Architecture of AlexNet



Conceptual Representation of the Inception Feature in GoogLeNet.

Cont…

- Two Examples of Using the Google Lens, a Service Based on Convolutional Deep Networks for Image Recognition.



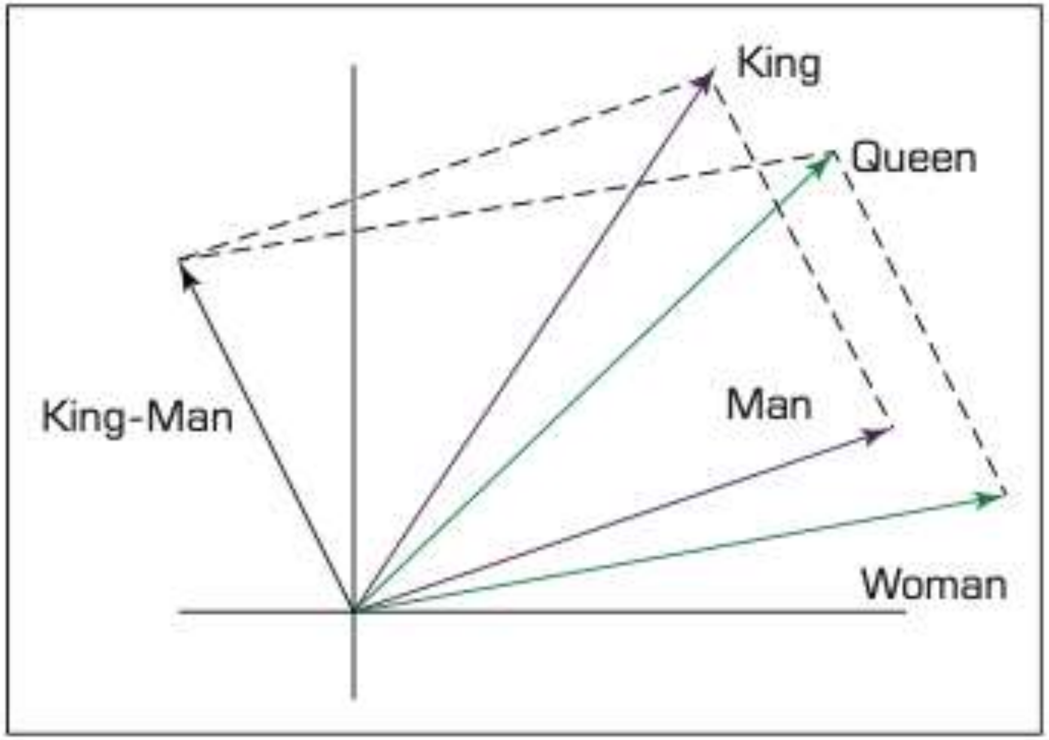
Text Processing Using Convolutional Networks

- CNNs have been shown to be useful in some large-scale text mining tasks.
- Google’s word2vec project remarkably increased the use of CNN-type deep learning for text mining applications.
- Word2vec is a two-layer neural network that gets a large text corpus as the input and converts each word in the corpus to a numeric vector of any given size (typically ranging from 100 to 1,000) with very interesting features.
- Word2vec itself is not a deep learning algorithm, its outputs (word vectors, known as word embeddings) that have been widely used in many deep learning research and commercial projects as inputs.

Word2vec

- One of the most interesting properties of word vectors created by the word2vec algorithm is maintaining the words' relative associations.

- E.g.: vector operations *vector*('*King*') - *vector*('*Man*') + *vector*('*Woman*') will result in a vector very close to *vector* ('*Queen*')



Typical Vector Representation of Word Embeddings in a Two-Dimensional Space

Word2vec

- Word2vec takes into account the contexts in which a word has been used and the frequency of using it in each context in guessing the meaning of the word.

- It is able to handle and correctly represent words including typos, abbreviations, and informal conversations.
  - E.g., the words *Frnce, Franse*, and *Frans* would all get roughly the same word embeddings as their original counterpart *France*.

- Word embeddings are also able to determine other interesting types of associations such as distinction of entities or geopolitical association.

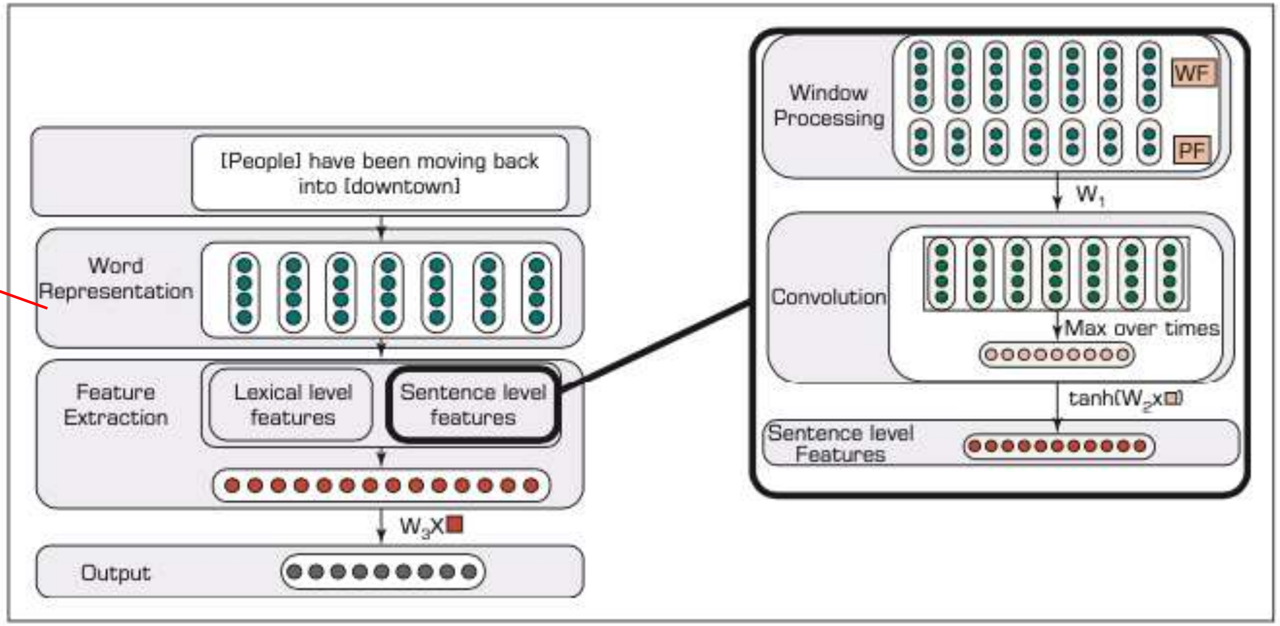
TABLE 6.2 Example of the word2vec Project Indicating the Closest Word Vectors to the Word "Sweden"

Word	Cosine Distance
Norway	0.760124
Denmark	0.715460
Finland	0.620022
Switzerland	0.588132
Belgium	0.585635
Netherlands	0.574631
Iceland	0.562368
Estonia	0.547621
Slovenia	0.531408

Word2vec

- Various types of deep networks have been applied to the word embeddings created by word2vec to accomplish different objectives (e.g., relation extraction from textual data sets).

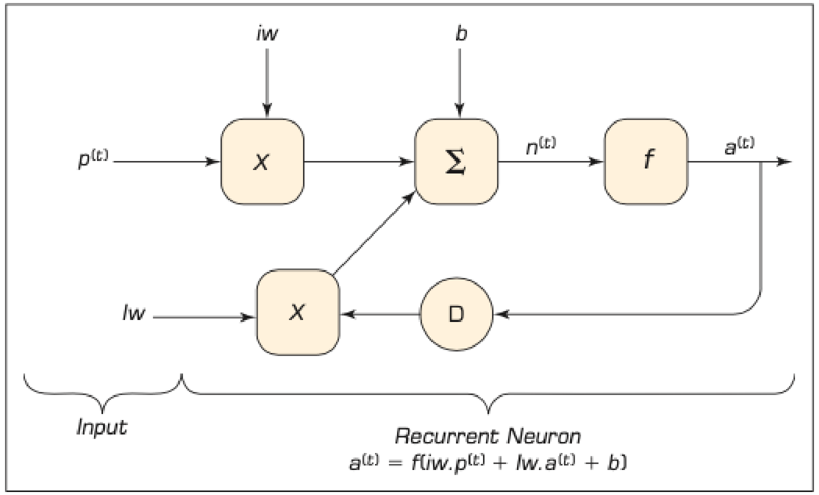
word embedding (i.e., vector) associated with one of the words involved in the sentence.



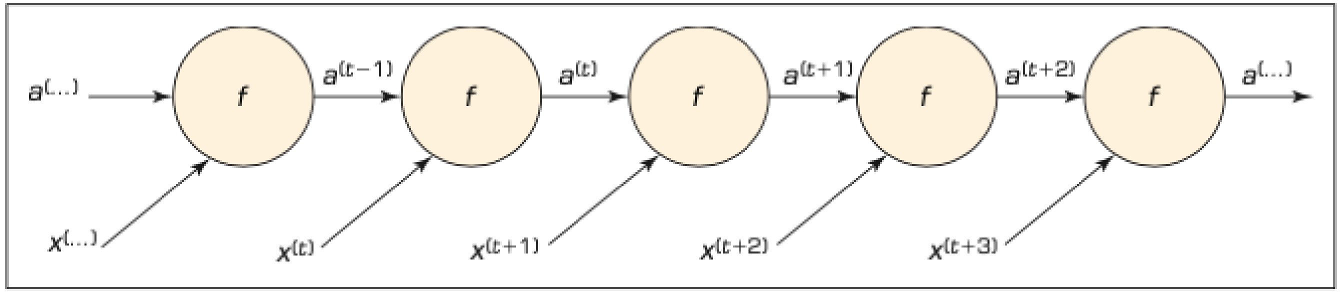
CNN Architecture for Relation Extraction Task in Text Mining

Recurrent Networks And Long Short-term Memory Networks (RNNs)

- The neural networks in which feedback connections are allowed are called recurrent neural networks (RNN).
- RNN is another deep learning architecture designed to process sequential inputs.
- RNNs have memory to remember previous information in determining context-specific, time-dependent outcomes.



Typical Recurrent Unit



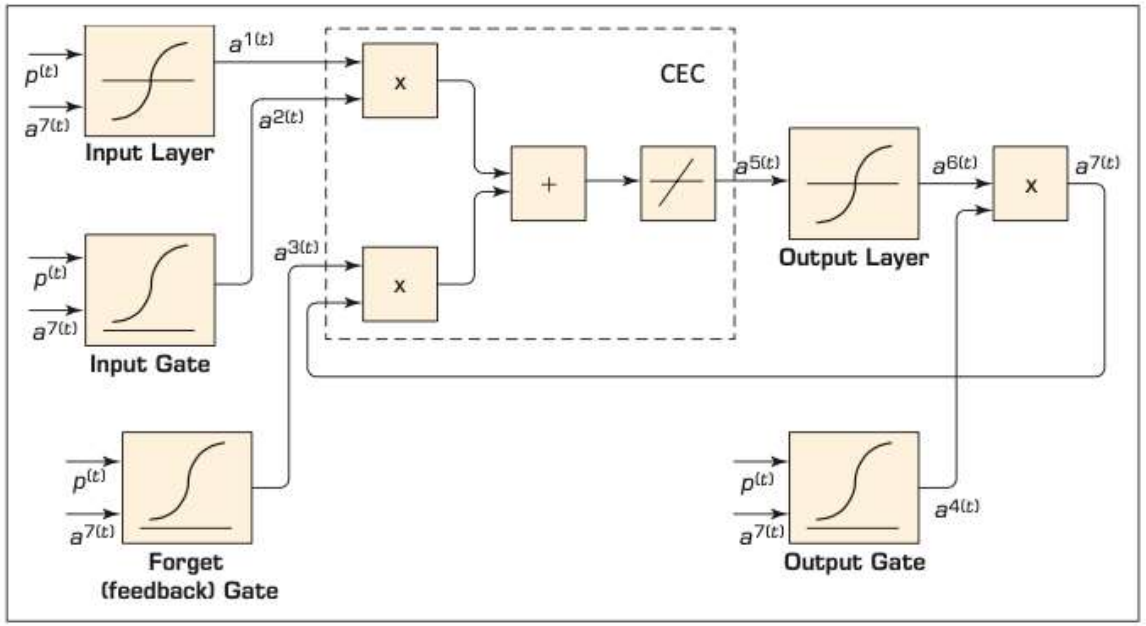
Unfolded View of a Typical Recurrent Network

Long Short-Term Memory Networks (LSTM)

- LSTM is a variation of RNN.
- LSTM network is known as the most effective sequence modelling technique and is the base of many practical applications.
- LSTM refers to a network in which we are trying to remember what happened in the past (i.e., feedbacks; previous outputs of the layers) – A memory concept.
- A long short-term memory means keeping the effect of previous outputs for a longer time.
- Many application areas
  - Image captioning, handwriting recognition and generation, parsing, machine translation (Two emerging applications: Google Neural Machine Translator and Microsoft Skype Translator).

Long Short-Term Memory Networks (LSTM)

- The memory concept is incorporated in LSTM networks by incorporating four additional layers into the typical recurrent network architecture:
  - Three gate layers (input gate, forget (a.k.a. feedback) gate, and output gate).
  - Constant Error Carousel (CEC) layer, also known as the state unit.



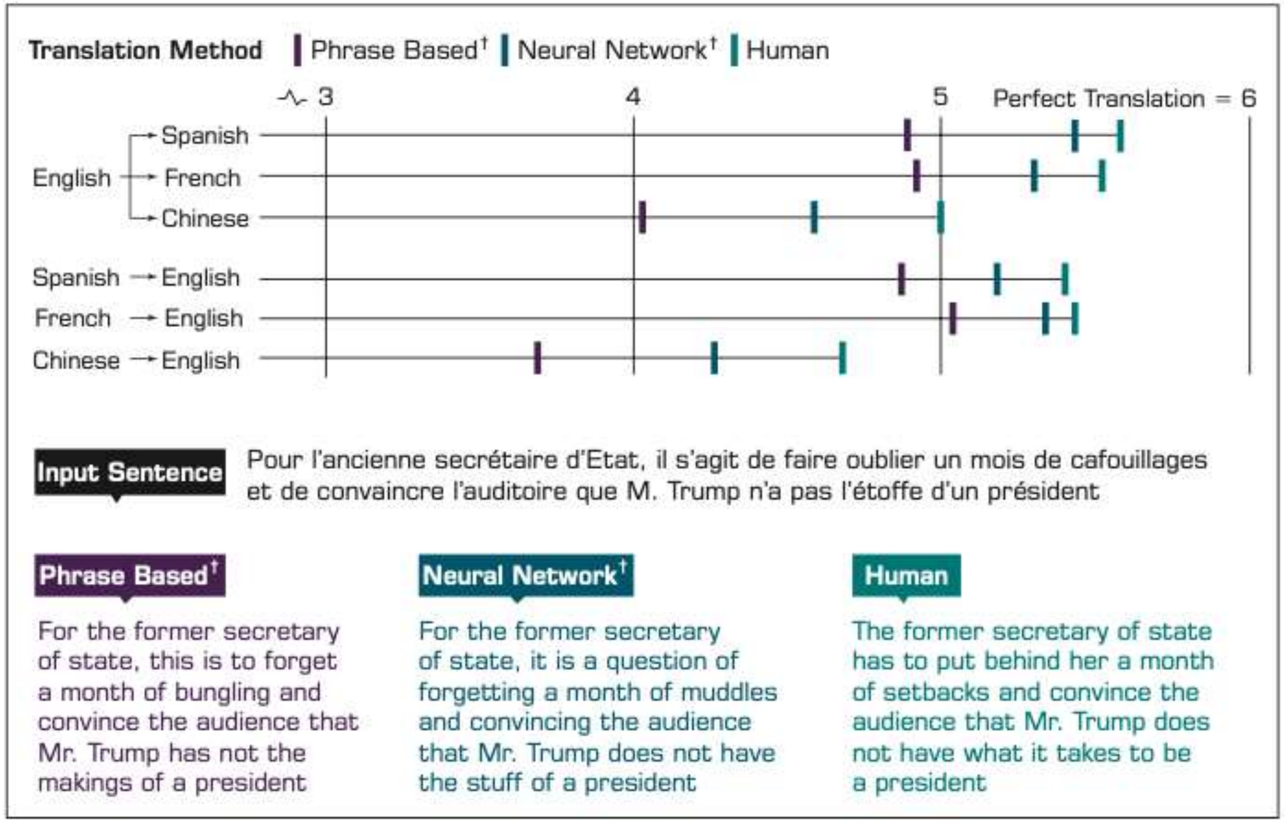
Typical Long Short-Term Memory (LSTM) Network Architecture

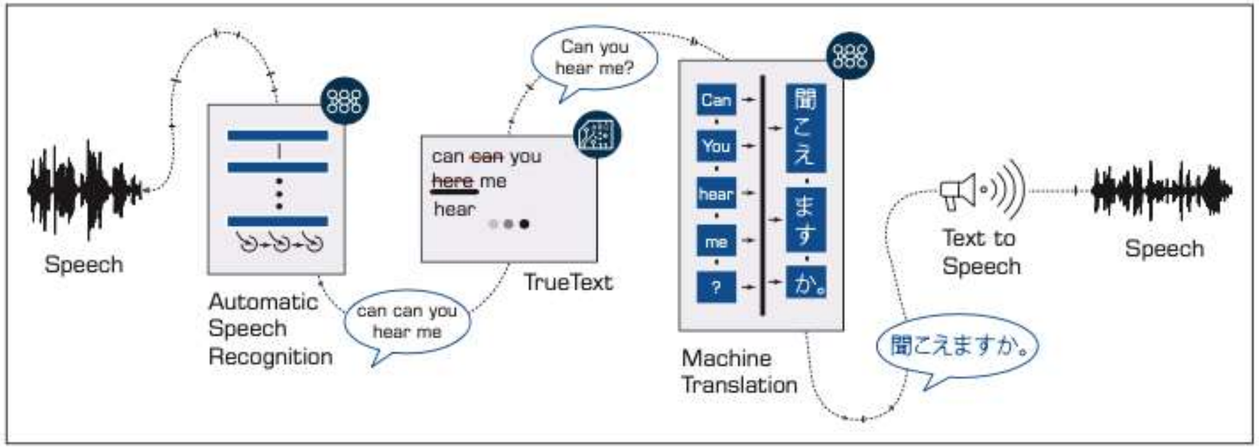
- The gates in the LSTM are in charge of controlling the flow of information through the network and dynamically change the time scale of integration based on the input sequence.

LSTM Networks Applications

LSTM networks have been widely used in many sequence modelling applications:

- Image captioning.
- Handwriting recognition and generation.
- Parsing.
- Speech recognition.
- Machine translation.





6.9 Deep COMPUTER FRAMEWORKS FOR IMPLEMENTATION OF DEEP LEARNING

- Frameworks

Frameworks

- Deep learning implementation frameworks (open-source) include:
  - **Torch**: is a scientific computing framework for implementing machine-learning algorithms using GPUs.
  - **Caffe**: The deep learning libraries are written in the C++ programming language, everything is done using text files instead of code.
  - **TensorFlow**: a popular deep learning framework, It was originally developed by the Google Brain Group.
  - **Theano**: one of the first deep learning frameworks
  - **Keras**: functions as a high-level application programming interface (API) and is able to run on top of various deep learning frameworks including Theano and TensorFlow.

6.6 Cognitive Computing

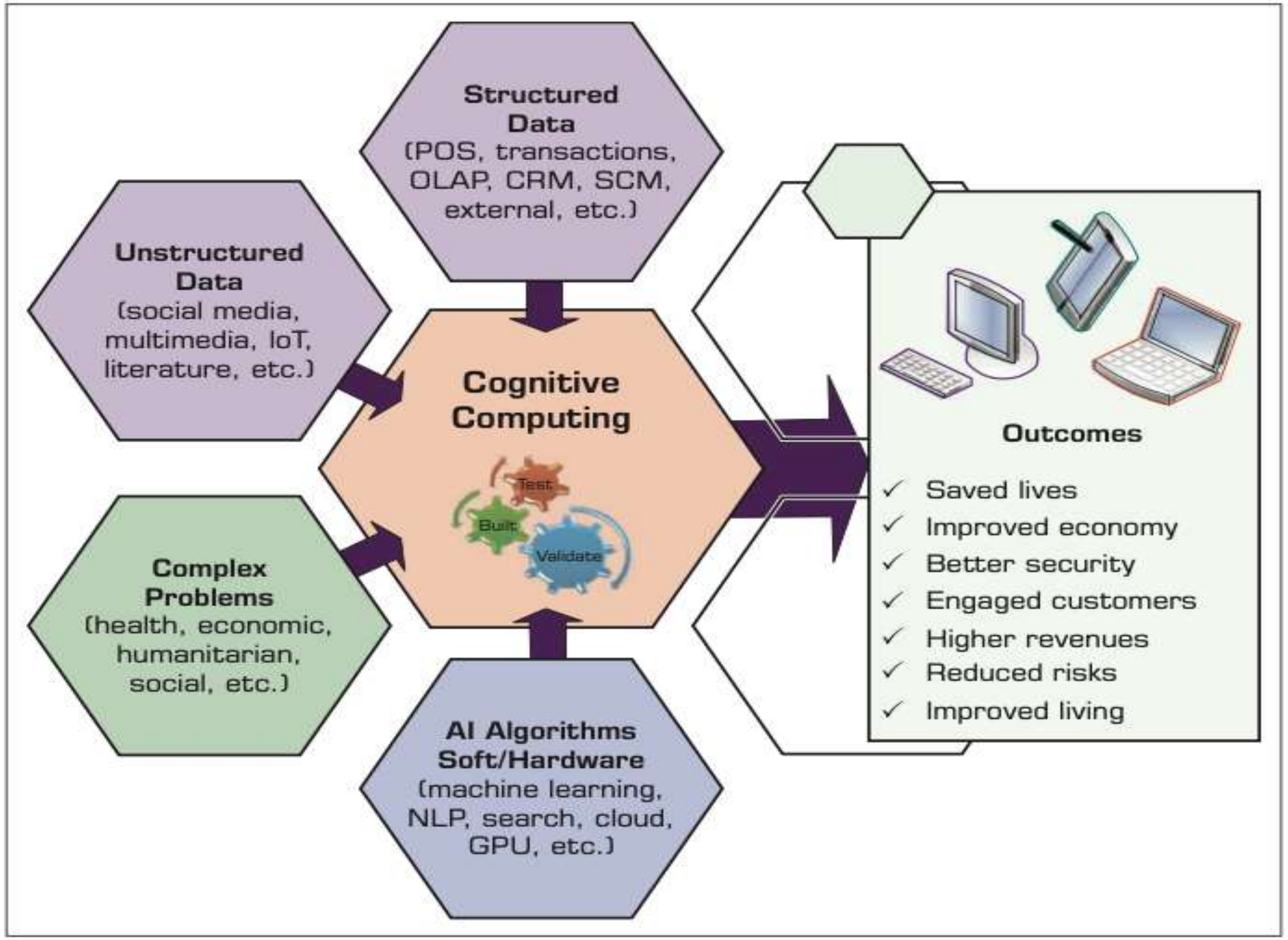
- Conceptual Framework for Cognitive Computing
- How Does Cognitive Computing Work?
- Cognitive Computing and AI
- Typical use cases for cognitive computing
- Cognitive analytics and Search
- IBM Watson

Cognitive Computing

- Cognitive computing makes a new class of problems computable.
- It address highly complex situations that are characterized by ambiguity and uncertainty.
- Handles the kinds of problems that are thought to be solvable by human ingenuity and creativity.
- Computing system offers a synthesis not just of information sources but also of influences, contexts, and insights that help users understand their problems.

Conceptual Framework for Cognitive Computing

- To provide the best possible answers to a given question or problem, cognitive computing:
  - finds and synthesizes data from various information sources,
  - And weighs the context and conflicting evidence inherent in the data.
  - And suggest an answer that is "best" rather than "right."



a general framework for cognitive computing where data and AI technologies are used to solve complex real-world problems

How Does Cognitive Computing Work?

- Cognitive computing works much like a human thought process, reasoning mechanism, and cognitive system.

- It includes self-learning technologies that use data mining, pattern recognition, deep learning, and NLP to mimic the way the human brain works.

- Cognitive systems may draw on multiple sources of vast amounts of information, including structured and unstructured data and visual, auditory, or sensor data solve the types of problems that humans are typically tasked.

- Over time, cognitive systems are able to refine the way in which they learn and recognize patterns and the way they process data to become capable of anticipating new problems and modelling and proposing possible solutions.

How Does Cognitive Computing Work?

The key attributes of cognitive computing capabilities:

- **Adaptability:** be flexible enough to learn as information changes and goals evolve.
- **Interactivity:** Users must be able to interact with cognitive machines and define their needs as those needs change.
- **Iterative and stateful:** ability to maintaining information about similar situations that have previously occurred.
- **Contextual:** must understand, identify, and mine contextual data, such as syntax, time, location, domain, requirements, and a specific user’s profile, tasks, or goals.

Cognitive Computing and AI

TABLE 6.3 Cognitive Computing versus Artificial Intelligence (AI)

Characteristic	Cognitive Computing	Artificial Intelligence (AI)
Technologies used	• Machine learning • Natural language processing • Neural networks • Deep learning • Text mining • Sentiment analysis	• Machine learning • Natural language processing • Neural networks • Deep learning
Capabilities offered	Simulate human thought processes to assist humans in finding solutions to complex problems	Find hidden patterns in a variety of data sources to identify problems and provide potential solutions
Purpose	Augment human capability	Automate complex processes by acting like a human in certain situations
Industries	Customer service, marketing, healthcare, entertainment, service sector	Manufacturing, finance, healthcare, banking, securities, retail, government

Typical use cases for cognitive computing

- Development of smart and adaptive search engines.
- Effective use of natural language processing.
- Speech recognition.
- Language translation.
- Context-based sentiment analysis.
- Face recognition and facial emotion detection.
- Risk assessment and mitigation.
- Fraud detection and mitigation.
- Behavioral assessment and recommendations.

Cognitive analytics

- Cognitive analytics is a term that refers to cognitive computing–branded technology platforms.
  - E.g., IBM Watson specialize in the processing and analysis of large unstructured data sets.
- The benefit of utilizing cognitive analytics over traditional Big Data analytics tools is that for cognitive analytics such data sets do not need to be pretagged.
- Cognitive analytics systems can use machine learning to adapt to different contexts with minimal human supervision.
  - These systems can be equipped with a chatbot or search assistant that understands queries, explains data insights, and interacts with humans in human languages.

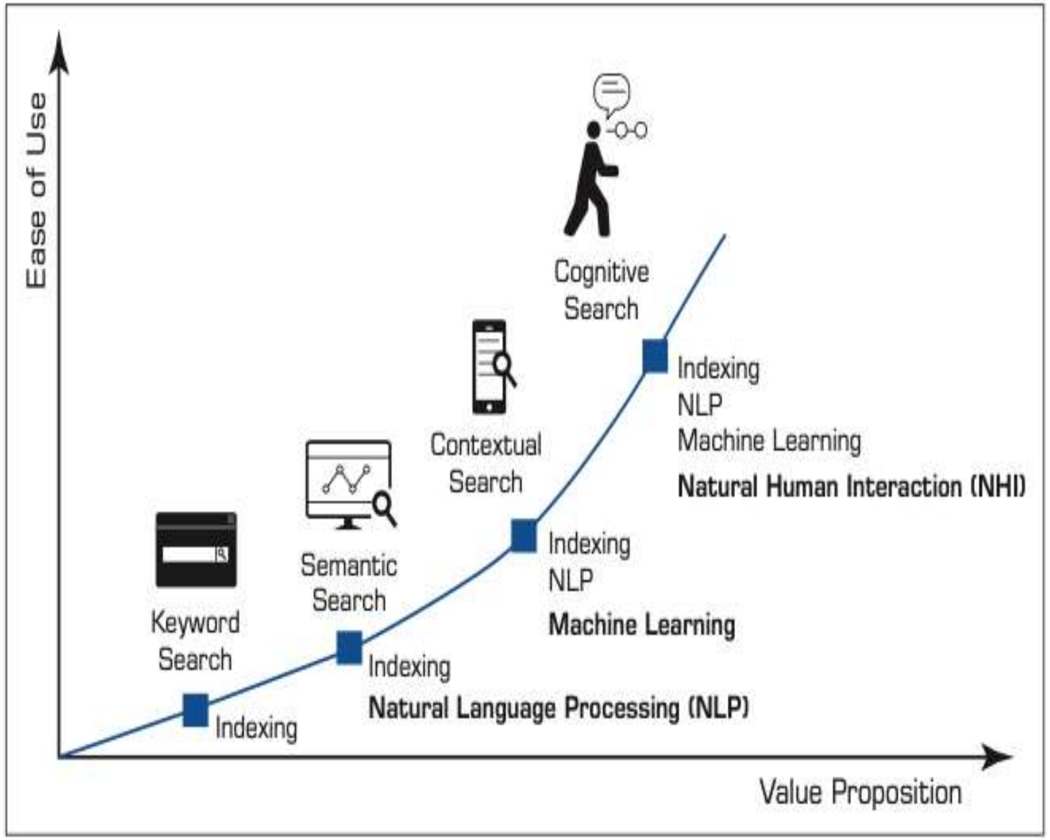
Cognitive Search

- Searching for information is a tedious task.
- Cognitive search is the new generation of search method that uses AI (e.g., advanced indexing, NLP, and machine learning) to return results that are much more relevant to the user than traditional search methods.
- It creates searchable information out of non-searchable content by leveraging cognitive computing algorithms to create an indexing platform.
- Cognitive search proposes the next generation of search tailored for use in enterprises.

Cognitive Search

Cognitive search is different from traditional search because, according to Gualtieri (2017), it:

- Can handle a variety of data types.
- Can contextualize the search space.
- Employ advanced AI technologies.
- Enable developers to build enterprise-specific search applications.



The progressive evolution of search methods.

IBM Watson

- IBM Watson is perhaps the smartest computer system built to date. It has coined and popularized the term cognitive computing.

- It is an extraordinary computer system—a novel combination of advanced hardware and software—designed to answer questions posed in natural human language.

- IBM Watson beat the best of men (the two most winning competitors) at the quiz game Jeopardy!, showcasing the ability of commuters to do tasks that are designed for human intelligence.

- Watson and systems like it are now in use in many application areas including:
  - Healthcare, finance, security, retail, education, government and research.

How Does Watson Do It?

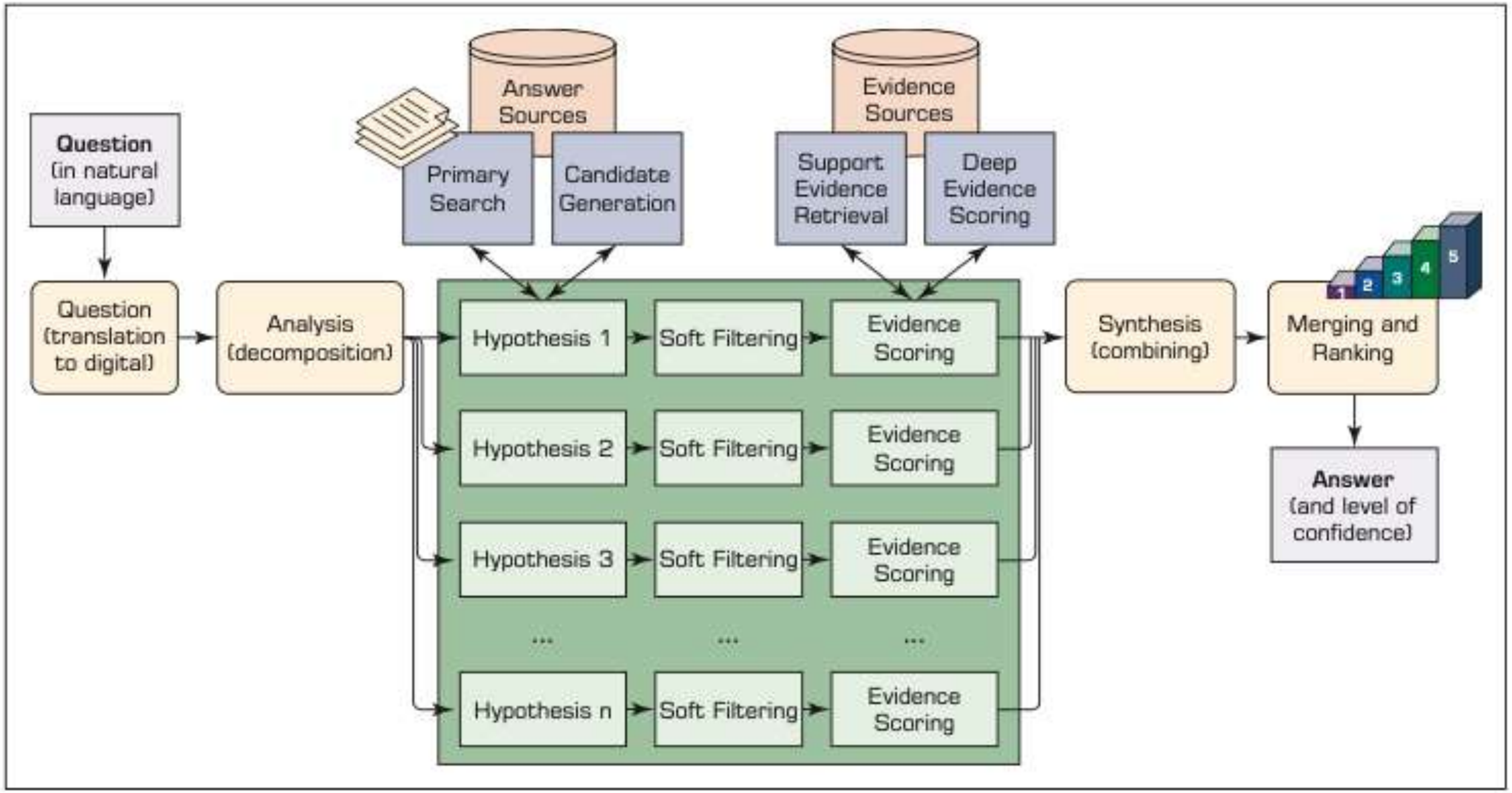
- DeepQA is the system behind Watson, which is a massively parallel, text mining–focused, probabilistic evidence–based computational architecture.

- Goal: to bring their strengths to bear and contribute to improvements in accuracy, confidence, and speed.

Principles in DeepQA

- Massive parallelism.
- Many experts.
- Pervasive confidence estimation.
- Integration of shallow and deep knowledge.

How Does Watson Do It?



A High-Level Depiction of DeepQA Architecture

Main Reference

- **Chapter 6**: "*Deep Learning and Cognitive Computing*" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".

Week self-review exercises

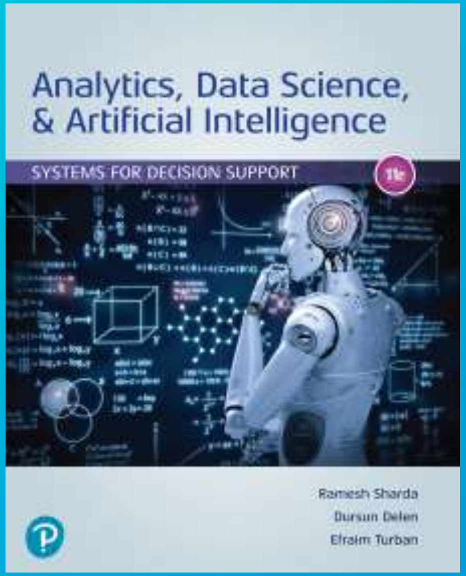
- **Application case 6.1 to 6.8** from *"Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*"

Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

College of Computing and Informatics



Week 14

Chapter 13:The Internet of Things as a Platform for Intelligent Applications

Chapter 14:The Internet of Things as a Platform for Intelligent Applications

Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Describe the IoT and its characteristics
2. Discuss the benefits and drivers of IoT and understand how IoT works
3. Describe sensors and explain their role in IoT applications and describe typical IoT applications in a diversity of fields
4. Describe smart appliances and homes and understand the concept of smart cities, their content, and their benefits
5. Describe the landscape of autonomous vehicles
6. Discuss the major issues of IoT implementation
7. Describe the major implementation issues of intelligent technologies
8. Discuss legal, privacy and ethical issues
9. Describe the major impacts on organizations and society
10. Discuss and debate the impacts on jobs and work
11. Describe the major influencing technology trends
12. Describe the highlights of the future of intelligent

Required Reading

- **Chapter 13**: "*The Internet of Things as a Platform for Intelligent Applications*" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".

- **Chapter 14**: "*Implementation Issues: From Ethics and Privacy to Organizational and Societal Impacts*" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".

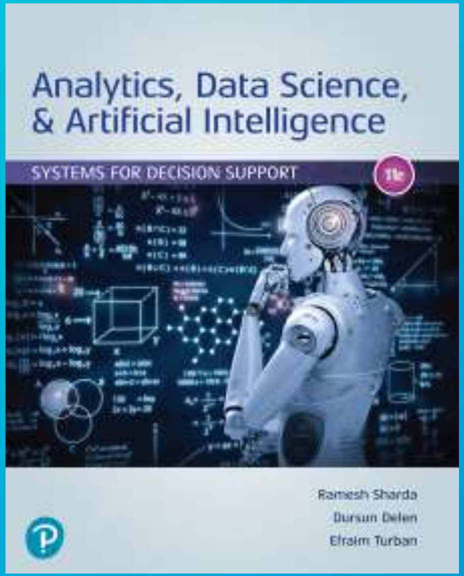
Recommended Readings

- Radu Stefan & George Carutasu, 2020. "How to Approach Ethics in Intelligent Decision Support Systems," Springer Proceedings in Business and Economics, in: Gabriela Prostean & Juan José Lavios Villahoz & Laura Brancu & Gyula Bakacsi (ed.), Innovation in Sustainable Management and Entrepreneurship, chapter 0, pages 25-40, Springer. https://ideas.repec.org/h/spr/prbchp/978-3-030-44711-3_3.html

- Rhem, A.J. AI ethics and its impact on knowledge management. *AI Ethics* **1,** 33–37 (2021). https://link.springer.com/article/10.1007/s43681-020-00015-2

Recommended Video

- Does AI make better decisions than humans? Thinking Ethics of AI (2020, Oct 19). [Video]. YouTube. https://www.youtube.com/watch?v=2E7l1hdjHsg



Week 14 Part 1

Chapter 13:The Internet of Things as a Platform for Intelligent Applications

13.2 Essentials of IoT

- Definitions and Characteristics
- The IoT Ecosystem
- IoT and Decision Support

Definitions and Characteristics

- The IoT is a revolutionary technology that can change everything.

Internet of Things (IoT) - is a network of connected computing devices including different types of objects (e.g., digital machines). Each object in the network has a unique identifier (UID), and it is capable of collecting and transferring data automatically across the network.

- The collected data has no value until it is analyzed.

- IoT allows people and things to interact and communicate at any time, any place, regarding any business topic or service.

- The opportunities of this disruptive technology are cutting costs, creating new business models and improving quality.

Definitions and Characteristics

- IoT is a connected network in which:
  - Large numbers of objects (things) can be connected.
  - Each thing has a unique definition (IP address).
  - Each thing has the ability to receive, send, and store data automatically.
  - Each thing is delivered mostly over the wireless Internet.
  - Each thing is built upon machine-to-machine (M2M) communication.

- IoT is changing everything which are facilitated by AI systems, to enhance analytics and automate or support decision making.

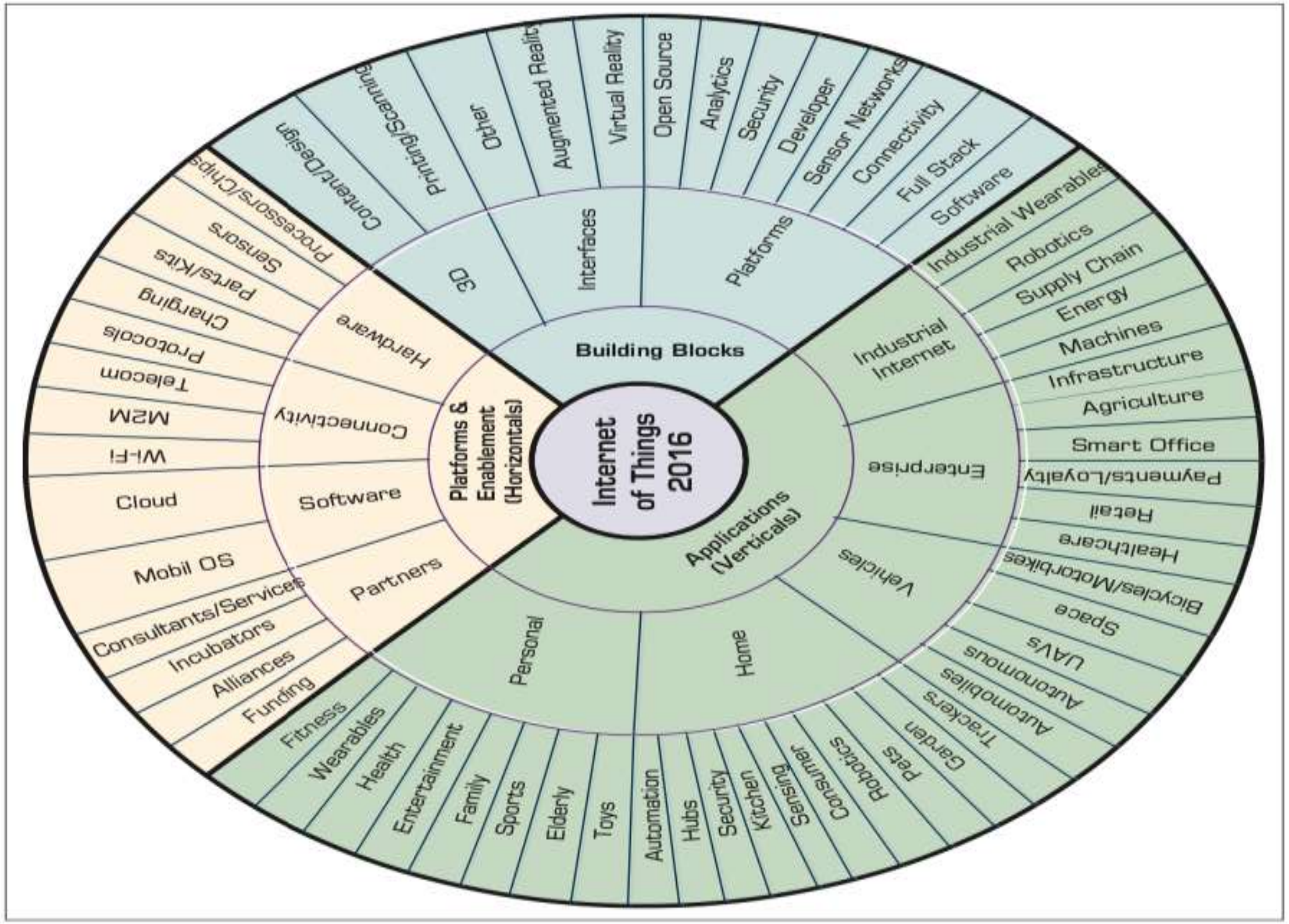
IoT examples

- Autonomous vehicle - a vehicle needs to have enough sensors that automatically monitor the situation around the car and take appropriate actions.

- A refrigerator could automatically order food (e.g., milk) when it detects that the food has run out!.

- Fleet tracking systems allow logistics and transport firms to optimize routing, track vehicle speeds and locations, and analyze driver and route efficiencies.

The IoT Ecosystem



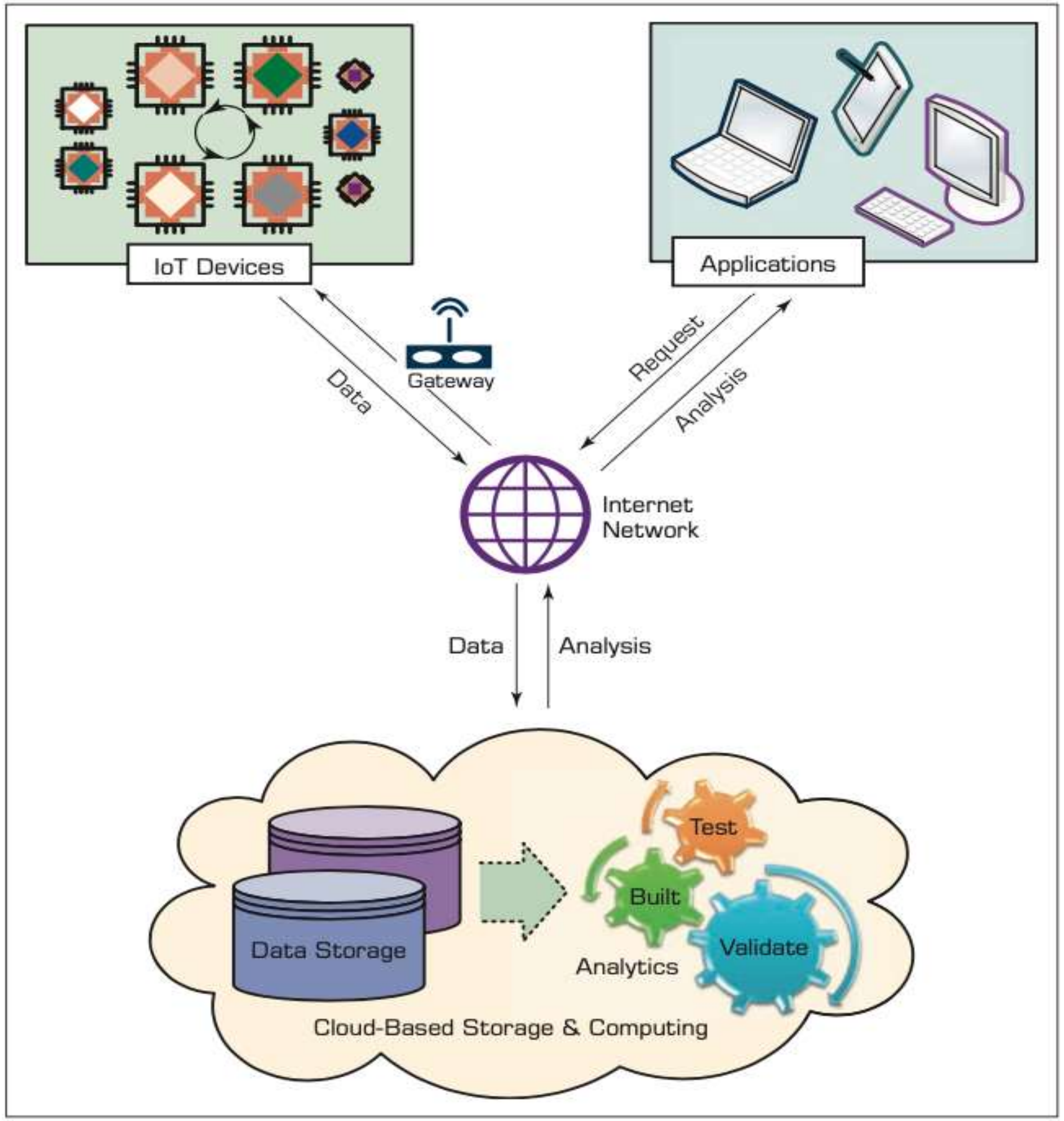
The IoT Ecosystem

IoT technology infrastructure

- Hardware
- Connectivity
- Software backend
- Applications

IoT platforms

- Amazon AWS IoT, Microsoft Azure IoT suite, Predix IoT, and IBM Watson IoT platform….



The Building Blocks of IoT

IoT and Decision Support

- The IoT may generate a huge amount of data (Big Data) that needs to be analyzed by various business intelligence methods, including deep learning, or advanced AI methods.
- Generated data can be submitted as support to decision makers or is inputted to automated decision support entities.
- Large-scale IoT usually needs to filter the collected data and "clean" them before they can be used for decision support, particularly if they are used as a base for automated decision making.

13.3 Major Benefits and Drivers of IoT

Major Benefits of IoT

- Reduces cost by automating processes.
- Improves workers' productivity.
- Creates new revenue streams.
- Optimizes asset utilization (e.g., see the opening vignette).
- Improves sustainability.
- Changes and improves everything.
- May anticipate our needs (predictions).
- Enables insights into broad environments (sensors collect data).
- Enables smarter decisions/purchases.
- Provides increased accuracy of predictions.
- Identifies problems quickly (even before they occur).
- Provides instant information generation and dissemination.
- Offers quick and inexpensive tracking of activities.
- Makes business processes more efficient.
- Enables communication between consumers and financial institutions.
- Facilitates growth strategy.
- Fundamentally improves the use of analytics (see the opening vignette).
- Enables better decision making based on real-time information.
- Expedites problem resolution and malfunction recovery.
- Supports facility integration.
- Provides better knowledge about customers for personalized services and marketing.

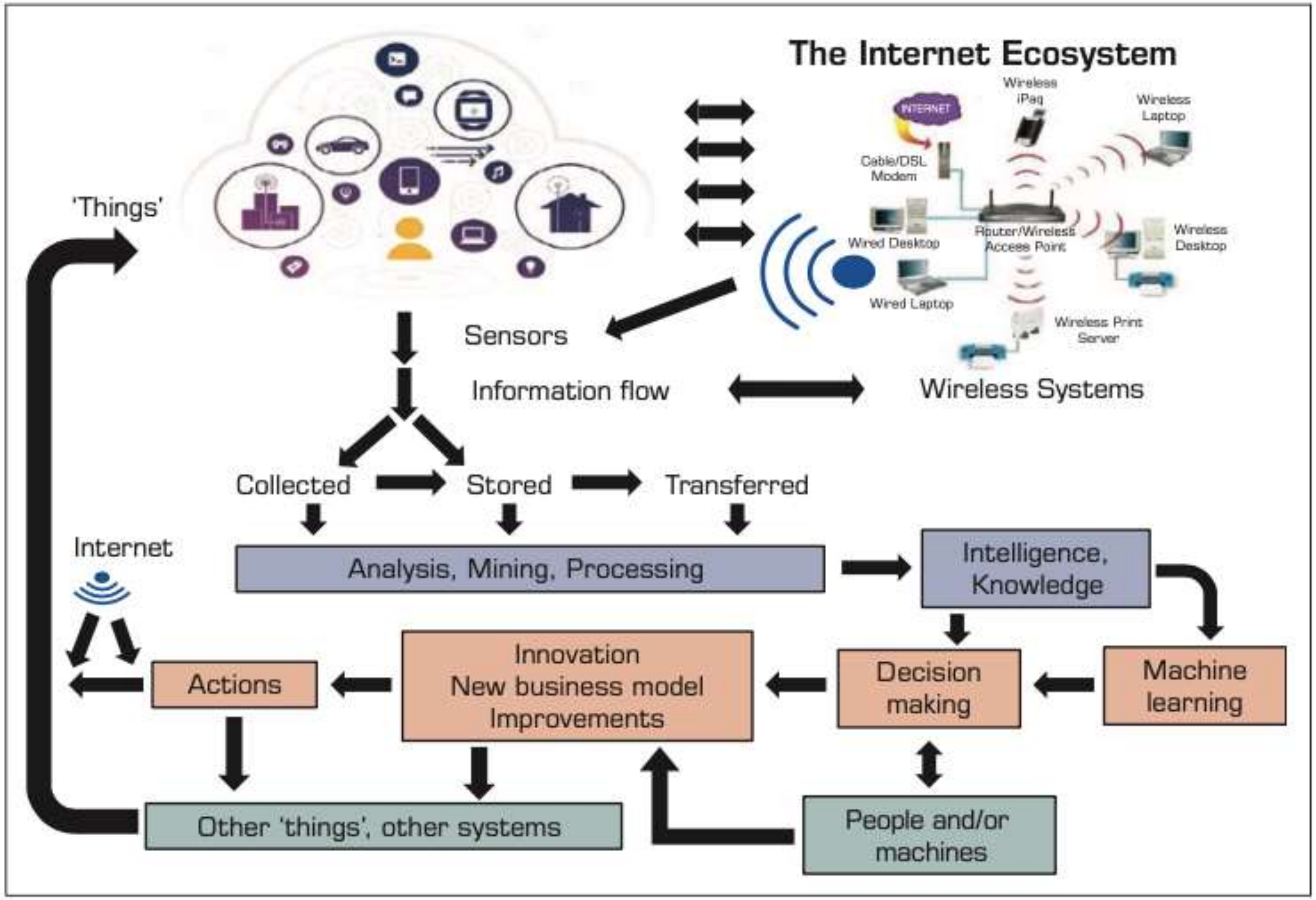
Major Drivers of IoT

- The number of “things”—20 to 50 billion—may be connected to the Internet by 2020–2025.
- Connected autonomous “things”/systems (e.g., robots, cars) create new IoT applications.
- Broadband Internet is more widely available, increasing with time.
- The cost of devices and sensors is continuously declining.
- The cost of connecting the devices is decreasing.
- Additional devices are created (via innovations) and are interconnected easily (e.g., see Fenwick, 2016).
- More sensors are built into devices.
- Smartphones’ penetration is skyrocketing.
- The availability of wearable devices is increasing.
- The speed of moving data is increasing to 60 THz.
- Protocols are developing for IoT (e.g., WiGig).
- Customer expectations are rising; innovative customer services are becoming a necessity.
- The availability of IoT tools and platforms is increasing.
- The availability of powerful analytics that are used with IoT is increasing.

13.4 How IoT Works

How IoT Works

- IoT encompasses infrastructure, platform, or framework that is used to support applications.
- Sensor to insight: applications create knowledge or deliver new information.
- Sensor to action cycle



13.5 Sensors and Their Role in IoT

Introduction to Sensor Technology

Sensor - is an electronic device that automatically collects data about events or changes in its environment. The collected data are sent to other electronic devices for processing.

- Sensors often generate signals that are converted to human-readable displays.

- They are essential components in robotics and autonomous vehicles in addition to IoT.

- Each sensor usually has a limit on the maximum distance that it can detect.
  - very short range sensors (e.g., proximity sensors) are more reliable than those that operate in larger ranges.

Introduction to Sensor Technology

How sensors work with IoT

- In large-scale applications, sensors collect data that are transferred to processing in the "cloud."
  - Several platforms are used for this process (e.g., Rockwell Automation).

There are several types of sensors

- A well-known type of sensor that plays an important role in IoT is radio-frequency identification (RFID)
  - part of a broader ecosystem of data capture technologies.
  - use of radio-frequency waves to identify objects.
  - Several forms of RFID play a major role in IoT applications.
- Smart sensors
  - Senses the environment and processes the input it collects by using its built-in computing capabilities (e.g., a micro- processing).
  - crucial and an integral element in the IoT
  - Provide different levels of capabilities (e.g., special software for data conversion, digital processing, and communication capability to external devices.
- Other sensors: temperature and humidity sensors …..

13.6 Selected IoT Applications

13.7 Smart Homes and Appliances

13.8 Smart Cities and Factories

13.9 Autonomous (Self-driving) Vehicles

IoT Applications Examples

- **Hilton Hotel:** Guests can check in directly to their rooms with their smartphones (no check-in lobby is needed, no keys are used).
- **Ford:** Users can connect to apps by voice. Auto-paying for gas and pre-ordering drinks at Starbucks directly from Ford's cars are in development.
- **Tesla:** Tesla's software autonomously schedules a valet to pick up a car and drive it to Tesla's facility when a car needs repair or schedule service. Tesla trucks, man- aged by IoT, will be driverless one day.
- **Apple:** Apple enables users of iPhones, Apple Watches, and Home kits to streamline shopping with Apple Pay.
- **Starbucks Clover Net in the Cloud:** This system connects coffee brewers to customers' preferences. It also monitors employee's performance, improves recipes, tracks consumption patterns, and so on.

How IoT Is Driving Marketing

- **Disruptive data collection**
  - IoT collects more data about customers from more data sources than other technologies do.
  - E.g., data from wearables, smart homes, and data about changes in consumer preferences and behavior….
- **Real-time personalization**
  - IoT can provide more accurate information about specific customers buying decisions
  - (e.g., IoT can identify customer expectations and direct customers to specific brands).
- **Environmental attribution**
  - IoT can monitor environments regarding ad delivery for specific places, customers, methods, and campaigns.
  - IoT can facilitate research of business environment; factors such competition, pricing, weather conditions, and new government regulations are observed.
- **Complete conversation path**
  - IoT initiatives expand and enrich the digital channel of conversations between customers and vendors, especially those using wireless digital engagement.
  - IoT also provides insight on consumer purchasing paths.
  - Marketers will receive improved customized market research data (e.g., by following the manner of customers’ engagement and how customers react to promotions).

Smart Homes and Appliances

Smart home - is a home with automated components that are interconnected (frequently wirelessly), such as appliances, security, lights, and entertainment, and are centrally controlled and able to communicate with each other.

- Goal – is to provide comfort, security, low energy cost, and convenience.
- Communicate via smartphones or the Internet.
  - XIO, UPB, Z-Wave, and EnOcean are connections protocols that offers scalability.
- Smart appliance includes features that can remotely control the appliance operations, based on the user preferences.
  - It may utilize a Home Network or the Internet to communicate with other devices in the smart home.

Smart Homes and Appliances

The major components of smart homes

Smart Homes and Appliances

- Google’s nest a leading manufacturer of IoT smart home.
  - producer of programmable self-learning, sensor- driven, Wi-Fi–enabled products (e.g., Learning thermostat).
- Amazon Echo and Google Home (Popular smart-home starter kits).
  - The virtual personal assistant enables people to converse by voice with chatbots such as Alexa/Echo and Google Assistant to manage appliances in smart homes.

Smart home trends:

- TVs that can be used as a smart Hub for home appliances is coming from Samsung.
- Dolby Atmos products include speakers, receivers, and other entertainment items.
- DIY home smart security cameras make sure there is an intruder, not just the cat,
- before alerting the police.
- Water controls for faucets, sprinklers, and flood detectors are available.

Smart Homes and Appliances

Barriers to Smart Home Adoption:

- **Compatibility:**
  - Too many products and vendors.
  - Industry standards are needed.
- **Communication**
  - Capabilities and benefits of a smart home need to be clearly communicated to users.
- **Concentration**
  - Brands need to focus on population segments that are most interested in smart homes.

- Other issues related to smart home adaption:
  - cost justification
  - invasion of privacy
  - security
  - ease of use

Smart Cities and Factories

Smart cities - digital technologies (mostly mobile based) facilitate better public services for citizens, better utilization of resources, and less negative environmental impact.

- Goal: provide a better life for residents by automating as many as possible public services such as transportation, utilities, social services, security, medical care, education, and economy.

- Smart cities can have several smart entities (e.g., universities and factories).

- A large number of vendors, research institutions, and governments are providing technology support for smart cities (e.g., Bosch).

- Examples:
  - "In Zaragoza, Spain, a 'citizen card' can get you on the free city-wide Wi-Fi network, unlock a bike share, check a book out of the library, and pay for your bus ride home."
  - Hong Kong has a project called a smart mobility for the improvement of road safety. It can manages speed and lane violations and traffic congestion.

Smart Buildings: From Automated to Cognitive Buildings

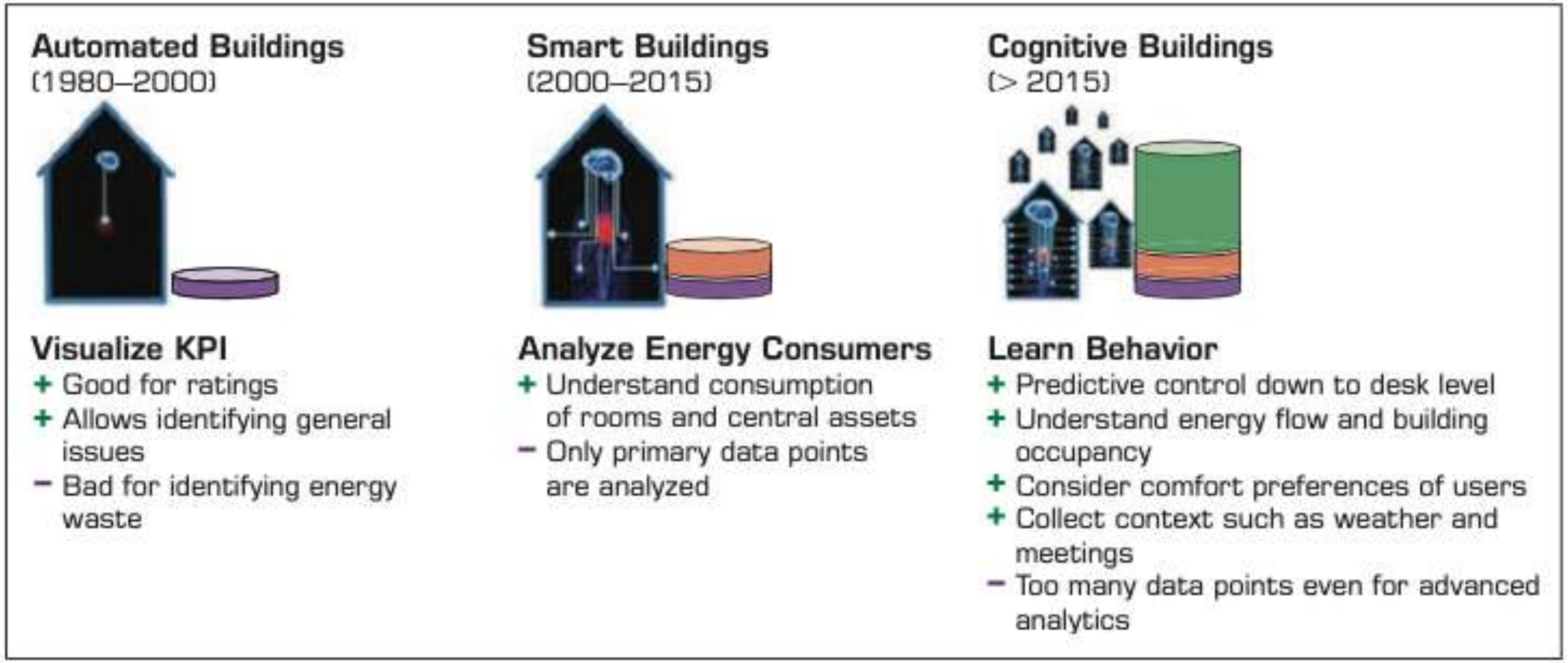
IMB's cognitive buildings - able to learn the behavior of a building's system in order to optimize it by autonomously integrating the IoT devices with the IoT operation to increases the productivity of existing systems.

Highlights of a cognitive building:

- By applying advance analytics, buildings can provide insights in near real time.
- It learns and reasons from data and interacts with humans.
- The system can detect and diagnose abnormal situations and propose remedies.
- It has the ability to change building temperature subject to humans' preferences.
- It is aware of its status and that of its users.
- It is aware of its energy status and adjusts it to be comfortable to dwellers.
- Its users can interact with the building via text messages and voice chatting.
- Robots and drones are starting to operate inside and outside the building without human intervention.

IBM's Cognitive Building Maturity Framework

Automated buildings → Smart building → Cognitive building



Smart (Digital) Factories

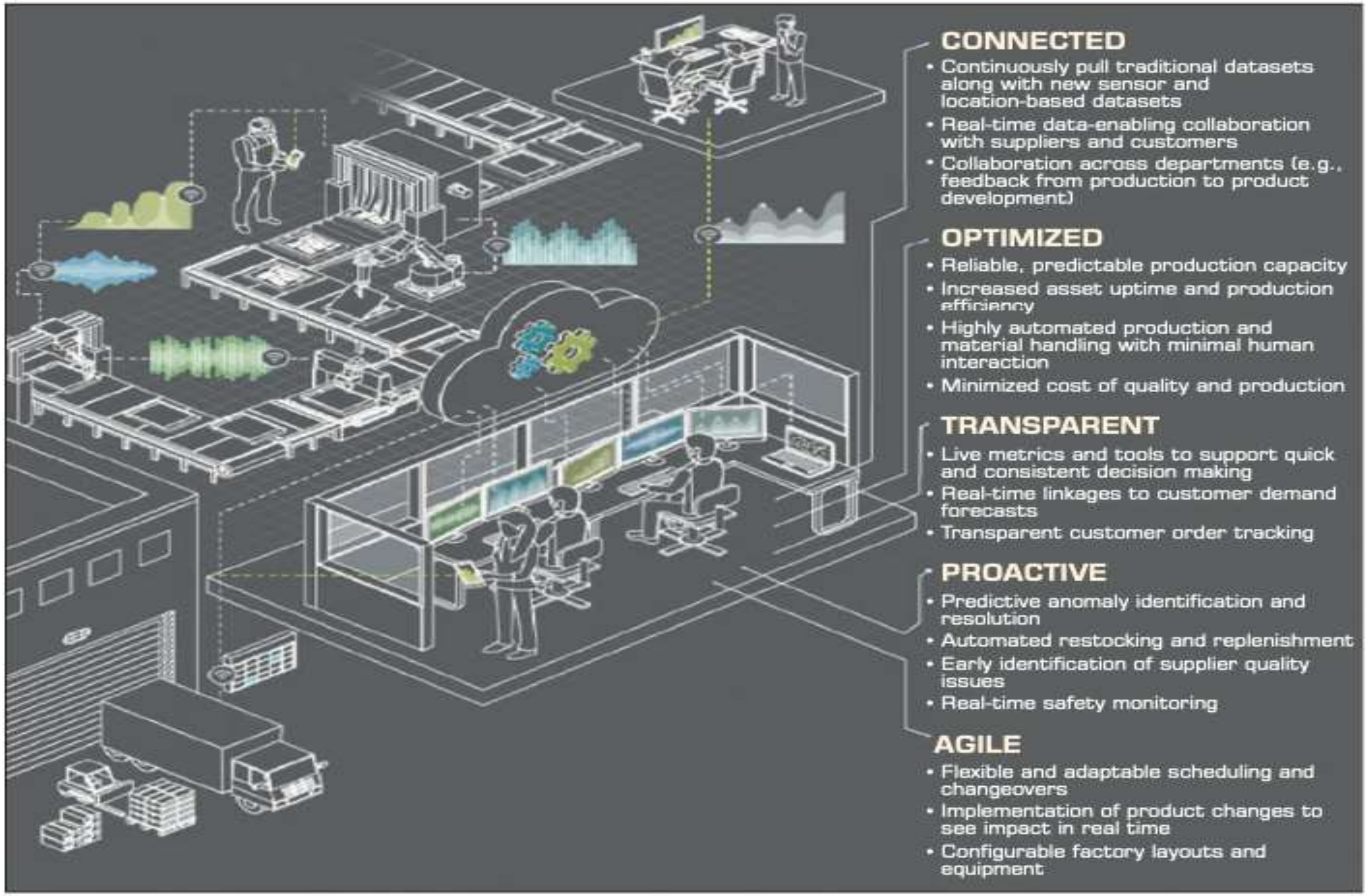
- Factories are getting smarter with AI technologies and IoT applications.

Future industrial production will be

  - fully digitized and connected, fast, and flexible (Tomás, 2016).
  - The entire production chain, from raw material suppliers, logistics, and manufacturing to sales, will be connected to IoT systems for planning, coordination, and control.
  - Planning will be based on analytic predictions of demand.
  - Production processes will be automated as much as possible and wirelessly controlled.
  - Logistics will be provided on demand quickly, and quality control will be automated.
  - IoT combined with sensors will be used for both predictive and preventive maintenance.
  - Example: Smart Bike Production in a Smart Factory

- Smart factory - is "a flexible system that can self-optimize performance across a broader network, self-adapt to and learn from new conditions in real or near real time, and autonomously run entire production processes. (Deloitte University Press)

The major characteristics of a smart factory (Deloitte)



Examples of Smart City Initiatives

Improving Transportation in the Smart City

- A major area of improvement in a smart city is transportation.
  - Sensors are taking data from cars on the roads to help generate data that can analyzed and results are transmitted to drivers.
  - air quality sensors have been placed on bicycles and cars.
  - E.g., Valerann start-up - developed smart road studs to replace the reflective studs of today's technology. They can transmit information of what they sense about what is occurring on the roads

Combining Analytics and IoT in Smart City Initiatives

- The SAS Analytics Model for Smart Cities: Three major phases
  - Sense: Using sensors, sense anything that matters.
  - Understand the signals in the data: Using data mining algorithms, the entire relevant ecosystem is analyzed for pattern recognition.
  - Act: Decisions can be made quickly as all relevant data are in place.

Autonomous (Self-driving) Vehicles

- Autonomous vehicles – are electric cars, and they can create a revolution by reducing emissions, accidents, fatalities.

- The initial efforts to commercialize a self-driving car were started by Google in the 1990s.
  - Google Chauffeur is becoming a reality.
  - Several car manufacturers are ready to sell or operate such cars (e.g., BMW, Mercedes, Ford, GM, Tesla,…).

- The Self Drive Act in USA aims to regulate the safety of the passengers in autonomous vehicles.

- **Flying Cars**
  - drones that can carry people already exist.
  - the navigation of a large number of flying cars may be a problem.
  - E.g., Airbus, Uber developed flying taxi, and Toyota is working on making a flying car.

Implementation Issues in Autonomous Vehicles

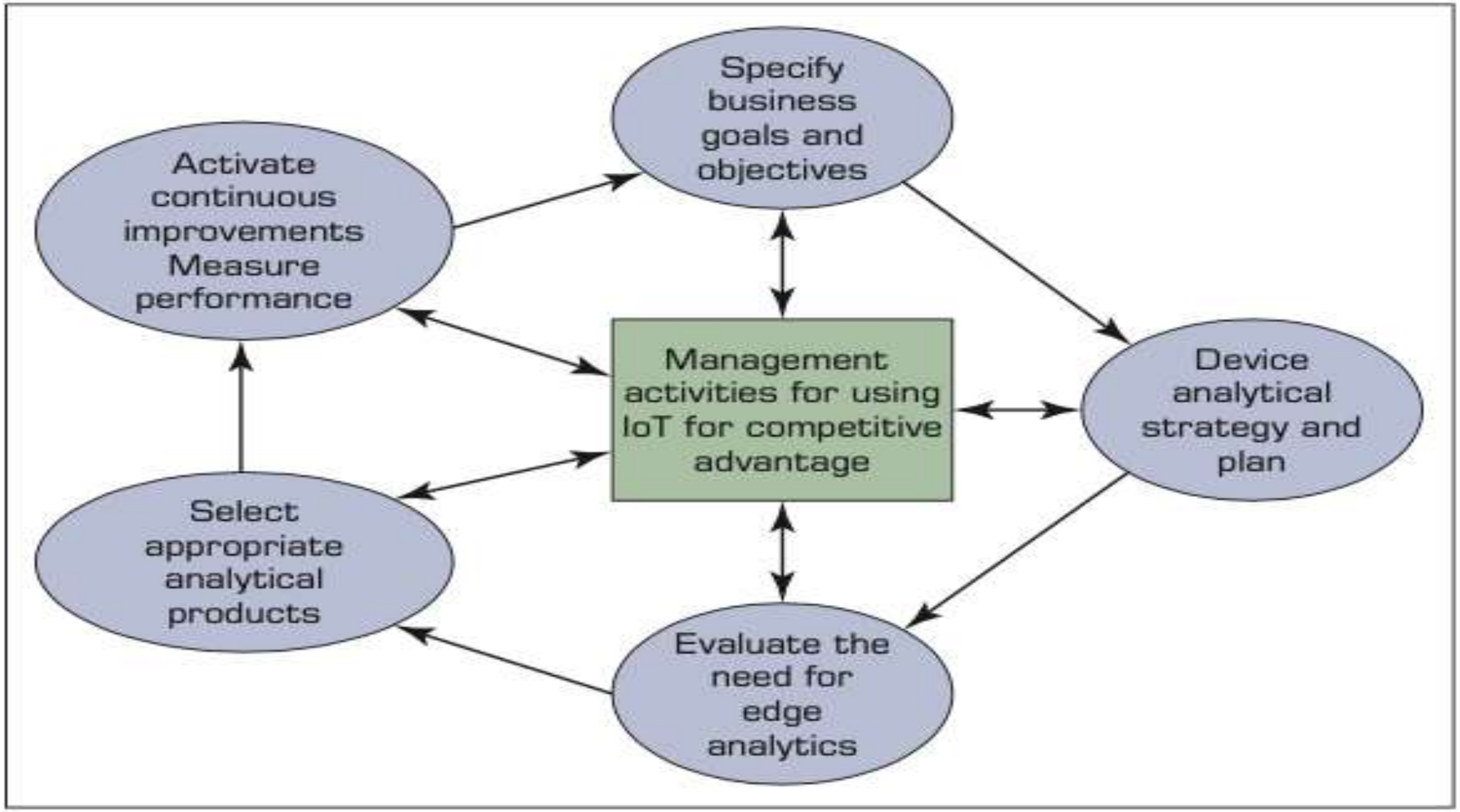
- The cost of real-time 3D map technologies needs to be reduced and their quality needs to be increased.
- AI software must be nimble and its capabilities increased (E.g., needs to deal with many unexpected conditions, including that of the behavior of drivers of other cars).
- Customers seem to acknowledge that such cars are coming. But they resist boarding one. However, some daring people expect these cars to do a better job than humans in driving.
- The technology needs more research, which is very expensive. One reason is that the many sensors in the cars and on the road need to be improved and their cost need to be reduced.
- The IoT is connecting many objects for autonomous vehicles, including those in clouds. The IoT systems themselves need to be improved. For example, data transmission delays must be eliminated.

13.10 Implementing IoT and Managerial Considerations

Major Implementation Issues

1. Organizational alignment
2. Interoperability challenges
   - E.g., technological issues regarding connectivity, Big Data processing Issues
3. Security
   - Data security
4. Privacy
   - Privacy protection system and policy are required
5. Connection of the silos of data
   - The need for a “fabric” and connectivity
6. Preparation of existing IT architectures and operating models for IoT can be a complex issue in many organizations
7. Management
8. Connected customers

Strategy for Turning Industrial IoT into Competitive Advantage



The IoT Strategy Cycle

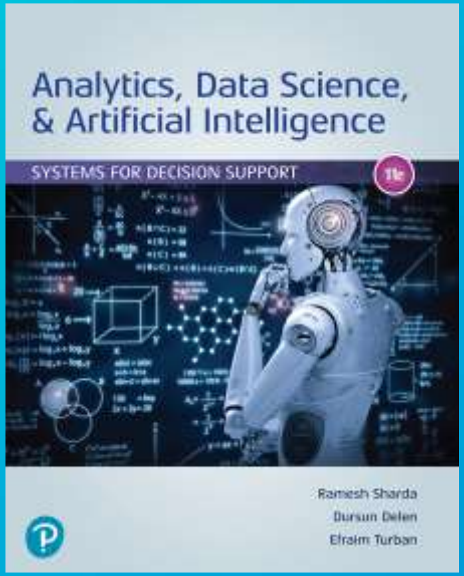
Strategy for Turning Industrial IoT into Competitive Advantage

Weldon (2015) suggests the following steps for successful IoT implementation:

1. Develop a business case to justify the IoT project including a cost-benefit analysis and a comparison with other projects.
2. Develop a working prototype. Experiment with it. Learn and improve it.
3. Install the IoT in one organizational unit; experiment with it. Learn lessons.
4. Plan an organization-wide deployment if the pilot is a success. Give special attention to data processing and dissemination.

AI enhancement of IoT

- Using machine learning that to provide insights about data.
- Help in creating devices ("things") that can self-diagnose problems and even repair them.
- Embodied cognition that injects AI capabilities into objects to enable the objects to understand their environments and then self-learn and improve their operation.
- Help the integration of IoT with other IT systems.



Week 14 Part 2

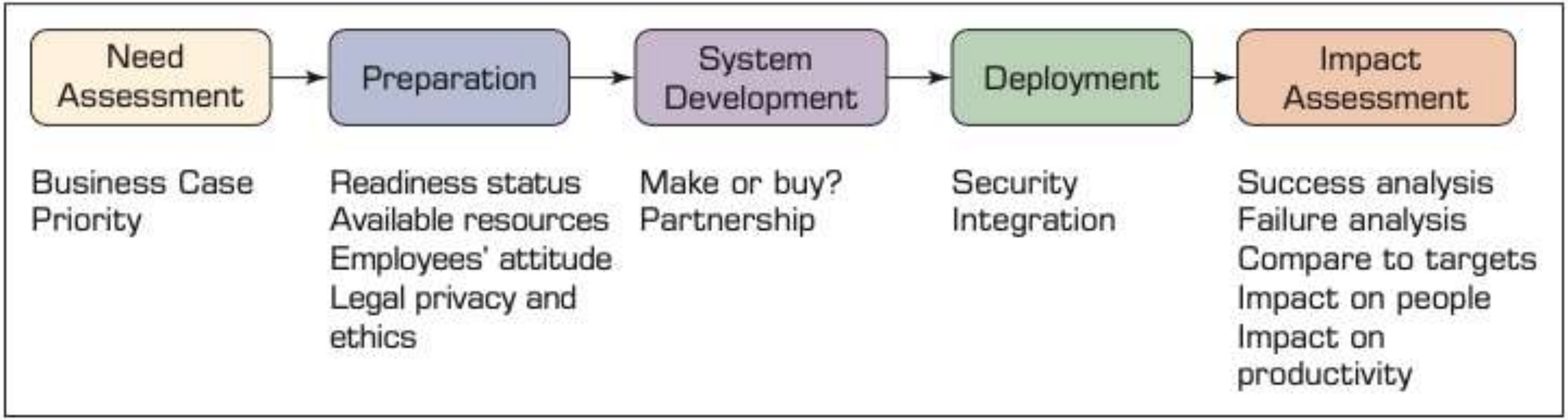
Chapter 14:The Internet of Things as a Platform for Intelligent Applications

14.2 Introduction Implementing Intelligent Systems: An Overview

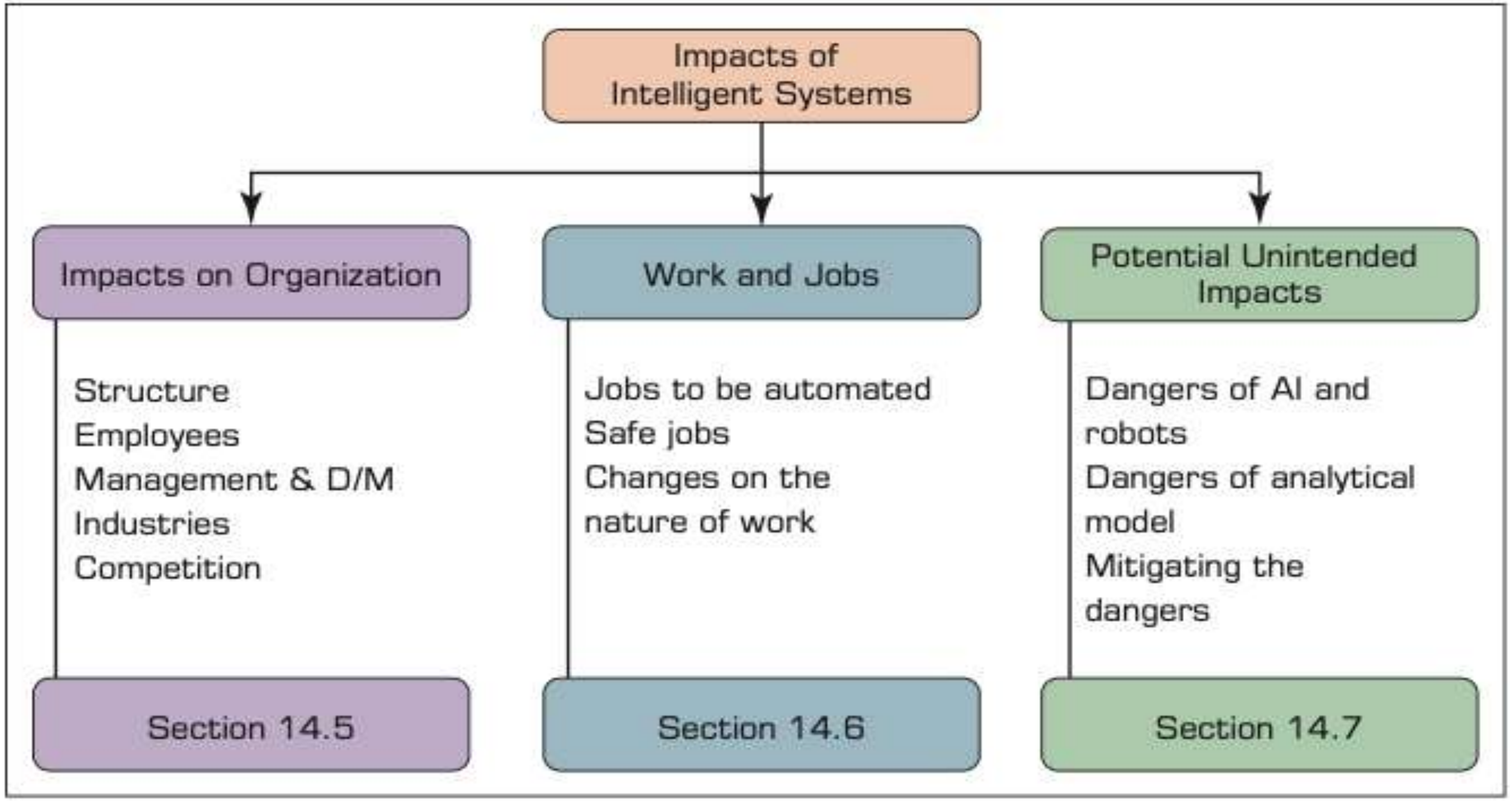
- The Intelligent Systems Implementation Process
- The Impacts of Intelligent Systems

The Intelligent Systems Implementation Process

The major steps of implementation of intelligent systems



The Impacts of Intelligent Systems



Impact Landscape

14.3 Legal, Privacy, and Ethical Issues

Legal Issues

- Serious legal issues may develop with the introduction of intelligent systems
  - liability and privacy are the dominant problem areas.

Representative issues to consider

- Who is liable if an enterprise finds itself bankrupt as a result of using the advice of an AI-based application?
- Will the enterprise itself be held responsible for not testing the system adequately before entrusting it with sensitive or volatile issues?
- Will auditing and accounting firms share the liability for failing to apply adequate auditing tests?
- Will the software developers of intelligent systems be jointly liable?
- As self-driving cars become more common, who is liable for any damage or injury when a car’s sensors, network, or AI system fail to function as planned?

Legal Issues

AI potential legal issues to consider

- What is the value of an expert opinion in court when the expertise is encoded in a computer?
- Who is liable for wrong advice (or information) provided by an intelligent application? What happens if a manager enters an incorrect judgment value into an intelligent application and the result is damage or a disaster?
- Who owns the knowledge in a knowledge base (e.g., the knowledge of a chatbot)?
- Can management force experts to contribute their expertise to an intelligent system? How will they be compensated?
- Is it okay for self-driving cars with in-vehicle back-up drivers to drive on public roads? (Yes, in a few states, notably in California.)
- Who should regulate driverless car: cities, states, or the federal government?
- And more in the book….

Legal Issues

Legal issues of intelligent technologies

- E.g., robots’ legal rights. Do we need these rights?
- there are very few laws regarding intelligent technologies.
- Most of the laws relate to safety.

AI and law - AI applications to the legal profession

- Analyzing legal-related data (e.g., regulatory conflicts) to detect pattern
- Providing legal advice to consumers (e.g., see **DoNotPay.com**).
- Document review
- Analyzing contracts
- Supporting legal research
- Predicting results (e.g., likelihood to win)
- AI impact on the legal profession

Privacy Issues

- Privacy - is the right to be left alone and the right to be free from unreasonable personal intrusions.

- Privacy has long been related to legal, ethical, and social issues in many countries.

- Privacy are getting more important as the amount of data generated on the Internet is increasing exponentially, and in many cases it is lightly secured.

- Who Owns Our Private Data?
  - A major battle is brewing about who owns the user data that are being generated from the use of smartphones, cars, and so on.
  - Intelligent systems professionals and users must be aware of the legal and ethical issues involved in collecting information that may be privileged or protected.

Privacy Issues

Privacy issues as it relates to AI

- Collecting information about individuals
  - Intelligent technologies aim to pro- vide targeted services and marketing to consumers.
- Virtual personal assistants
  - Amazon’s Echo/Alexa and similar devices listen to what is going on.
- Mobile user privacy
- Privacy in IoT networks
- Recent technology issues in privacy and analytics
  - Many companies have started to employ intelligent technologies to develop profiles of users on the basis of their device usage, surfing, and contacts.

Ethics Issues

Ethical Issues of Intelligent Systems

- What are their impact on jobs?
- How do machines (i.e., robots) affect our behavior and interactions?
- How can wealth created by intelligent machines be distributed (e.g., Kaplan, 2016)?
- How can intelligent applications mistakes be guarded against? For example, how long should training programs in machine learning be?
- Can intelligent systems be fair and unbiased? How can bias in creation and operation of AI systems be eliminated?
- How can intelligent applications be keep safe from adversaries?
- More in the book …

Ethics Issues

Topics in Intelligent Systems Ethics

- Machine ethics is a part of the ethics of AI that is concerned with the moral behavior of artificially intelligent beings.
- Robotics is concerned with the moral behavior of designers, builders, and users of robots.
- Microsoft's Tay chatbot was closed due to its inability to understand many irrelevant and offending comments.
- Some are afraid that algorithm-based technologies, including AI, may become racists.
- self-driving cars may one day face a decision of whom to save and whom to kill (Spangler, 2017).
- Voice technologies enable the identification of callers to AI machines. This may be great on one hand, but it creates privacy concerns on the other.

Computer Ethics

- Focuses on the behavior of people toward information systems and computers in general.

The ten commandments of computer ethics

1. Thou shalt not use a computer to harm other people.
2. Thou shalt not interfere with other people’s computer work.
3. Thou shalt not snoop around in other people’s files.
4. Thou shalt not use a computer to steal.
5. Thou shalt not use a computer to bear false witness.
6. Thou shalt not use or copy software for which you have not paid.
7. Thou shalt not use other people’s computer resources without authorization.
8. Thou shalt not appropriate other people’s intellectual output.
9. Thou shalt not think about the social consequences of the program you write.
10. Thou shalt not use a computer in ways that show consideration and respect.

14.5 Impacts of Intelligent Systems on Organizations

Impacts of Intelligent Systems

- Separating the impact of intelligent systems from that of other computerized systems is a difficult task.
  - trend toward integrating, or even embedding, intelligent systems with other computer-based information systems

- Explosive growth in analytics, AI, and cognitive computing is going to have a major impact on the future of organizations.

- The impact can be divided into three general categories:
  - organizational
  - Individual
    - The impact of analytics on individuals varies it can be positive, neutral, or negative.
  - Societal

Impacts of Intelligent Systems

- New organizational units and their management
  - Creating an analytics department, a BI department, a data science department, and/or an AI department.
- Transforming businesses and increasing competitive advantage
  - The impact of intelligent systems ranges from full automation of many tasks, including managerial ones, to an increase in human-machine collaboration
  - Using intelligent systems to gain competitive advantage
    - E.g., using robots, Amazon enabled the company to reduce cost and control online commerce.
- Redesign of an organization through the use of analytics
  - studying organizational dynamics, personnel behavior, and redesigning the organization to better achieve its goals .
  - People Analytics
- Industrial Restructuring

Impacts of Intelligent Systems

- Impact on managers' activities, performance and job satisfaction
  - Some jobs may be substantially enriched by intelligent technologies, other jobs may become more routine and less satisfying.

  - Intelligent technologies can change the manner in which many decisions are made and can consequently change managers' job responsibilities.

  - Potential impacts of intelligent system on managers' jobs:
    - Less expertise (experience) is required for making many decisions.
    - Faster decision making is possible because of the availability of information and the automation of some phases in the decision-making process.
    - Less reliance on experts and analysts is required to provide support to top managers and executives.
    - Power is being redistributed among managers.
    - More in the book…

Impact on Decision Making

- Intelligent systems are used to support real-time decision making and improve or automate decision making.
- E.g., SAS Real-Time Decision Manager (RTDM) - an analytics-based integrated product
  - It combines SAS analytics with business logic and contact strategies to deliver enhanced real-time recommendations and decisions
  - **Key benefits**
    - Makes the right decisions every time, all the time.
    - Realizes customer needs with the right offer, at the right time, in the right channel.
    - Better allocates valuable IT resources.
  - **key features:** Real-time analytics, Rapid decision process construction, Enterprise data throughout, Campaign testing. Automated self-learning analytical process and Connectivity.

14.6 Impacts on Jobs and Work

Overview

Impacts of intelligent systems is on jobs and work:

- Automation segment, which is related to cognitive computing and AI, will accelerate polarization of the labor market in the future.
- Intelligent systems will create many new jobs as automation always has.
  - AI will create 2.3 million jobs in 2020 while eliminating 1.8 million (de Vos,2018).
- There will be a need to retrain many people.
- The nature of work will be changed.
  - Jobs may be redesigned either to be low skilled in order to be automated, or to be high skilled so that they will be executed exclusively by humans.
  - There is a chance for massive unemployment.
  - Jobs that require “middle skills” such as specialized knowledge that is applied over and over with some adaptation, are at the greatest risk of disappearing.
    - AI puts many jobs at risk (e.g., taxi driver, Robo advisors may cause some people to lose their jobs)
  - Changes in jobs and business processes will impact training, innovation, wages, and the nature of work itself.

Which Jobs Are Most in Danger? Which Ones Are Safe?

- Intelligent systems may actually add jobs
  - One needs to consider the great benefits of AI and the fact that human and machine intelligence will complement each other in many jobs.

TABLE 14.1 Ten Top Safe and at Risk Occupations

Probability of Job Loss	
Low-Risk Jobs	
0.0036	First-Line supervisors of firefighting and prevention workers
0.0036	Oral and maxillofacial surgeons
0.0035	Healthcare social workers
0.0035	Orthotists and prosthetists
0.0033	Audiologists
0.0031	Mental health and substance abuse social workers
0.0030	Emergency management directors
0.0030	First-Line supervisors of mechanics, installers, and repairers
0.0028	Recreational therapists
High-Risk Jobs	
0.99	Telemarketers
0.99	Title examiners, abstractors, and searchers
0.99	Sewers, hand
0.99	Mathematical technicians
0.99	Insurance underwriters
0.99	Watch repairer
0.99	Cargo and freight agents
0.99	Tax preparers
0.99	Photographic process workers and processing machine operators
0.99	New account clerks

Success Tips for Implementing AI Provided by the Executives

- Digital capabilities need to come before AI.
- Machine learning is powerful, but it is not the solution to all problems.
- Do not put technology teams solely in charge of intelligent technologies.
- Adding a business partner may help with AI-based projects.
- Prioritize a portfolio approach to AI initiatives.
- The biggest challenges will be people and business processes.
- Not every business is using intelligent systems, but almost all those that use them increase income and profit.
- Top leadership support is necessary for a transformation to AI.

Dealing With The Changes In Jobs and The Nature Of Work

- Use learning and education to facilitate the change.
- Involve the private sector in enhancing training and retraining.
- Have governments provide incentives to the private sector so employees can invest in improved human capital.
- Encourage private and public sectors to create appropriate digital infrastructure.
- Innovative income and wage schemes need to be developed.
- Carefully plan the transition to the new work.
- Deal properly with displaced employees.
- Properly handle new technology-enabled technologies.
- Focus on new job creation, particularly digital jobs.
- Properly capture the productivity increase opportunities.

14.7 Potential Dangers Of Robots, AI, and Analytical Modeling

Potential Dangers Of Robots, AI, and Analytical Modeling

- There is a heated debate regarding the future of AI and particularly robots.
  - Optimistic approach (Utopia) and the pessimistic approach (Dystopia) (Dickson, 2017).

Position of AI Dystopia

- Elon Musk predicts that World War III will start because of AI. “Robots will kill us all, one day”.
- Bill Gates: suggested taxing the manufacturers and users of robots and other AI machines.
- Stephen Hawking: “The development of full artificial intelligence could spell the end of the human race.”

The AI Utopia’s Position

- Contribution of AI to the quality of life.
- AI will support humans and enable innovations. They will become more productive and will have time to do more innovative tasks.

Potential Dangers Of Robots, AI, And Analytical Modeling

- Some issues related to the utopia
  - AI will be so great that people will have a problem of what to do with their free time.
  - The road to AI Utopia could be rocky. It will take time to stabilize and adjust work and life of living with robots, chatbots, and other AI applications.
  - One day we will not drive anymore and everything will be different. The changes may be rapid and turbulent and we may even face disasters, as projected by the Dystopia camp.
- The open AI project and the Friendly AI
  - The major objective of open AI is to enact the path to safe artificial general intelligence (AGI).
  - Idea of friendly AI is that AI machines should be designed so that they will benefit humans rather than harm them.

Potential Dangers Of Robots, AI, And Analytical Modeling

The O’Neil Claim of Potential Analytics’ Dangers

- Managers and data science professionals should be aware of the social and long-term effects of mathematical models and algorithms.

Models must satisfy three conditions (O’Neil,2016)

1. They must be transparent.
2. The model must have clear quantifiable objectives.
3. The models must have a self-correcting mechanism (a process in place to audit the models regularly and new inputs and outputs are constantly being considered).

14.8 Relevant Technology Trends

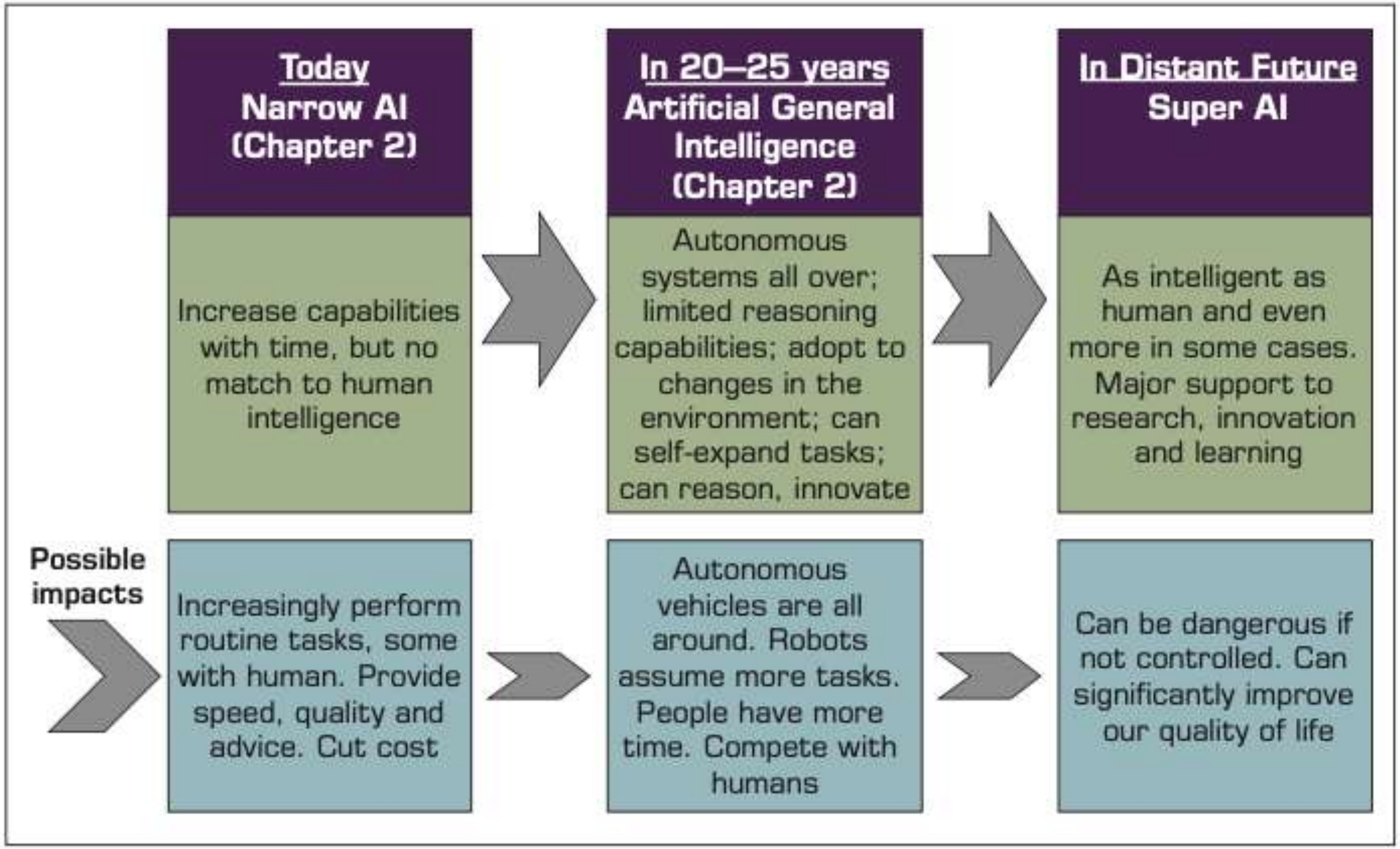
Gartner’s Top Strategic Technology Trends for 2018 and 2019

- AI Foundation and Development.
- Intelligent Apps and Analytics.
- Intelligent and Autonomous Things.
- Digital Twin.
- Empowered Cloud (Cloud to the Edge).
- Conversational Human-Machine Platforms.
- Immersive Experience.
- Blockchain.
- Augmented Analytics.

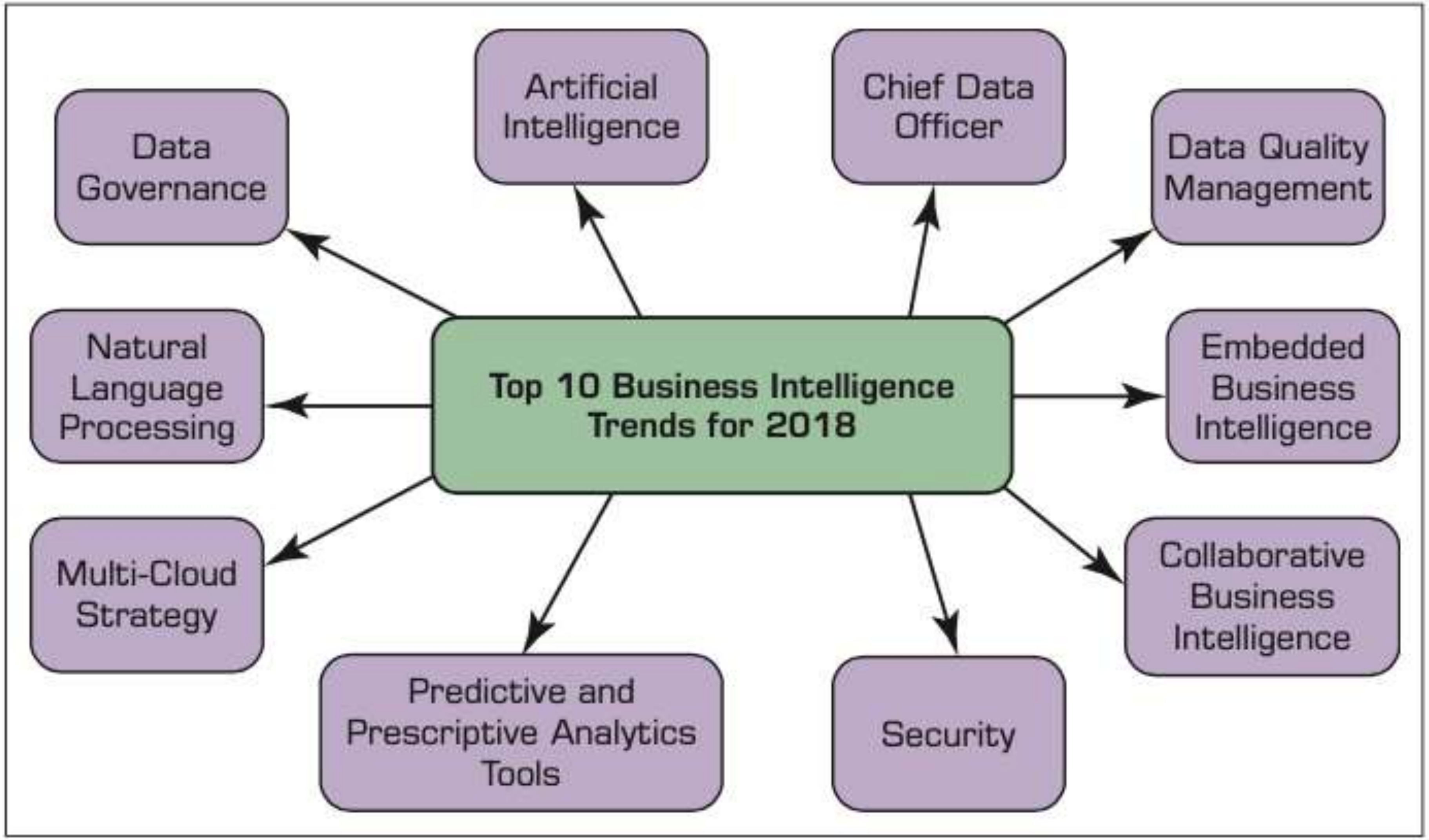
List Of Technology Trends In Intelligent Systems *Sommer (2017)*

- Data literacy will spread both in organizations and in society.
- Information points will be connected via hybrid multi-cloud systems.
- The mystery of rural networks will be exposed by deep learning theory.
- Self-service systems will use data catalogs as their frontier.
- Need to focus on Application Programming Interfaces (APIs).
- Analytics become conversational (e.g., via chatbots).
- Analytics will include immersive capabilities.
- Using augmented intelligence users will be turned to participants.

List Of Impact on AI and Analytics



The future of BI and analytics



Ambient Computing (Intelligence)

- Ambient computing - (a futuristic paradigm computing) it refers to electronic environments (e.g., network devices such as sensors) that are sensitive and responsive to people and their environments.

- Ambient devices can support people in whatever task they are doing.

Potential benefits of ambient intelligence

- Recognize individuals and other “things” and their context at any given time and place.
- Integrate into the environment and existing systems.
- Anticipate people’s desires and needs without asking (e.g., context awareness).
- Deliver targeted services based on people’s needs.
- Be flexible (i.e., can change their actions in response to people’s needs or activities).
- Be invisible.

Main Reference

- **Chapter 13**: "*The Internet of Things as a Platform for Intelligent Applications*" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".

- **Chapter 14**: "*Implementation Issues: From Ethics and Privacy to Organizational and Societal Impacts*" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".

Week self-review exercises

- **Application case 13.1 – 13.5** from *"Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support"*.

- **Application case 14.1 – 14.3** from *"Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support"* .

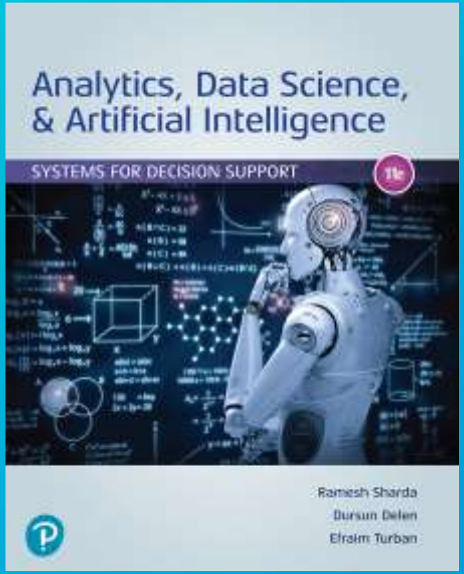
Thank You

IT445
Decision Support Systems

College of Computing and Informatics

Week 3

Chapter 2: Artificial Intelligence Concepts, Drivers, Major Technologies, and Business Applications



Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Understand the concepts of artificial intelligence (AI)
2. Become familiar with the drivers, capabilities, and benefits of AI
3. Describe human and machine intelligence
4. Describe the major AI technologies and some derivatives
5. Discuss the manner in which AI supports decision making
6. Describe AI applications in accounting, human resource management, marketing, financial Services and in Production-Operation Management (POM)

Required Reading

- **Chapter 2**: "Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support" from "Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".

Recommended Reading

- **AI-powered decision support systems, what are they?**
  https://blog.pwc.lu/ai-powered-decision-support-systems-what-are-they/

Recommended Videos

- **Artificial intelligence and decision-making (by Thorbjørn Knudsen)**
  https://www.youtube.com/watch?v=Sujww4njwE4

2.2 Introduction To Artificial Intelligence

- Definition of AI
- Major Characteristics of AI Machines
- Major Elements of AI
- AI Applications
- Major Goals of AI
- Drivers of AI
- Benefits of AI
- Some Limitations of AI Machines
- Three Flavors of AI Decisions
- Artificial Brain

Definition of AI

- Artificial intelligence has several definitions that is concerned with two basic ideas:
  - The study of human thought processes (to understand what intelligence is)
  - The representation and duplication of those thought processes in machines (e.g., computers, robots)

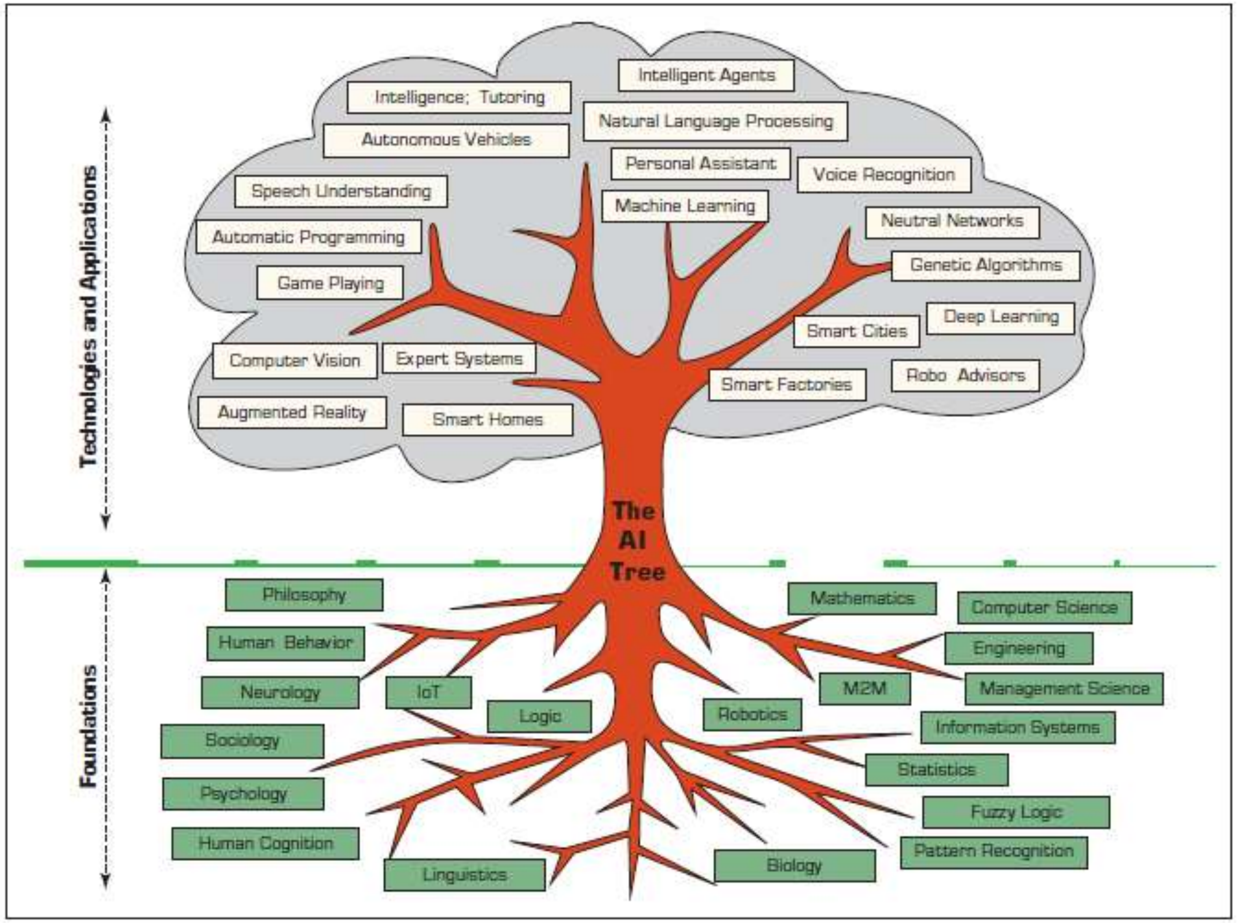
- Another definition of AI is "the capabilities of a machine to imitate intelligent human behavior"

Major Characteristics of AI Machines

- There is an increasing trend to make computers “smarter”.
  - Web 3.0 enables computerized systems that exhibit more intelligence than Web 2.0.
- Several applications are already based on multiple AI techniques.
  - Machine translation of languages is helping people who speak different languages to collaborate in real time as well as to buy online products that are advertised in different languages.

Major Elements of AI

- AI components can be divided into **two** groups: Foundations, and Technologies & Applications.



AI Applications

Smart or intelligent applications include:

- Machines to answer customers’ questions asked in natural languages
- Knowledge-based systems which can provide advice, assist people to make decisions, and even make decisions on their own
- Automatic generating of online purchasing orders and arranging fulfillment of orders placed online.
- Shipping prices are determined automatically based on the dimensions, weight, and packaging.

Major Goals of AI

- The overall **goal of AI** is to create intelligent machines that are capable of executing a variety of tasks currently done by people.
- **AI machines** should be able to reason, think abstractly, plan, solve problems, and learn.
- Some specific goals are to:
  - Perceive and properly react to changes in the environment that influence specific business processes and operations
  - Introduce creativity in business processes and decision making

Drivers of AI

- **The use of AI has been driven by the following:**

  - People's interest in smart machines and artificial brains
  - The low cost of AI applications versus the high cost of manual labor (doing the same work)
  - The desire of large tech companies to capture competitive advantage and market share of the AI market and their willingness to invest billions of dollars in AI
  - The pressure on management to increase productivity and speed
  - The availability of quality data contributing to the progress of AI
  - The increasing functionalities and reduced cost of computers in general
  - The development of new technologies, particularly cloud computing

Benefits of AI

- AI has the ability to complete certain tasks faster than humans.
- The consistency of AI work. AI machines do not make mistakes, stop, or sleep.
- AI systems allow for continuous improvement projects.
- AI can be used for predictive analysis via its capability of pattern recognition.
- AI can manage delays and blockages in business processes.
- AI machines can work autonomously or be assistants to humans.
- AI machines can learn, improve its performance, and work in hazardous environments.
- AI machines can facilitate innovations by human (i.e., support research and development)
- AI excels in fraud detection and in security facilitations.
- AI can free employees to work on more complex and productive jobs.
- AI can solve difficult problems that previously were unsolved.

Some Limitations of AI Machines

- The following are the **major limitations** of AI machines:
  - Lack human touch and feel
  - Lack attention to non-task surroundings
  - Can lead people to rely on AI machines (e.g., people may stop to think on their own)
  - Can be programmed to create destruction
  - Can cause many people to lose their jobs
  - Can start to think by themselves, diminishing with time. However, risks exist. Therefore, it is necessary to properly causing significant damage
- Some of the limitations are diminishing with time. However, risks exist. Therefore, it is necessary to improve AI development and minimize the risks.

The Three Flavors of AI Decisions

- The capabilities of AI systems can be divided into three levels

○ **Assisted Intelligence:** This is equivalent mostly to the weak AI, which works only in narrow domains. It requires clearly defined inputs and outputs. Examples are some monitoring systems and low-level virtual personal assistants.

○ **Autonomous AI:** These systems are in the realm of the strong AI but in a narrow domain. Eventually, the computer will take over as very narrow expert and have absolute decision-making power.

○ **Augmented Intelligence:** Most of the existing AI applications are between assisted and autonomous, which is referred to as augmented intelligence. The technology focuses on augmenting computer abilities to extend human cognitive abilities, resulting in high performance

Artificial Brain

- The **artificial brain** is a machine that is desired to be as intelligent, creative, and self-aware as humans. To date, no one has created such a machine.
- The following are some differences between traditional and augmented AI:
  - Augmented machines extend rather than replace human decision making
  - Augmentation excels in solving complex human and industry problems in specific domains in contrast with strong, general AI.
  - In contrast with a “black box” model of some AI and analytics, augmented intelligence provides insights and recommendations, including explanations.

2.3 Human and Computer Intelligence

- A. What Is Intelligence?
- B. How Intelligent Is AI?
- C. Measuring AI

What Is Intelligence?

- **Intelligence** is a broad term measured by an IQ test.
- To understand what artificial intelligence is, it is useful to first examine those abilities that are considered signs of human intelligence:
  - Learning or understanding from experience
  - Making sense out of ambiguous, incomplete, or even contradictory messages and information
  - Responding quickly and successfully to a new situation
  - Understanding/inferring in a rational way, and solving problems
  - Applying knowledge to manipulate environments and situations
  - Recognizing & judging the relative importance of elements in a situation

How Intelligent Is AI?

- AI machines have demonstrated superiority over humans in playing complex games such as chess, Jeopardy!, and Go by defeating the world's best players.
- Despite this many AI applications still show significantly less intelligence than humans.

TABLE 2.1 Artificial Intelligence versus Human Intelligence

Area	AI	Human
Execution	Very fast	Can be slow
Emotions	Not yet	Can be positive or negative
Computation speed	Very fast	Slow, may have trouble
Imagination	Only what is programmed for	Can expand existing knowledge
Answers to questions	What is in the program	Can be innovative
Flexibility	Rigid	Large, flexible
Foundation	A binary code	Five senses
Consistency	High	Variable, can be poor
Process	As modeled	Cognitive
Form	Numbers	Signals
Memory	Built in, or accessed in the cloud	Use of content and scheme memory

Measuring AI

- The Turing Test is a well-known attempt to measure the intelligence level of AI machines.
- It **aims** to determine whether a computer exhibits intelligent behavior. A computer can be considered smart only when a human interviewer asking the same questions to both an unseen human and an unseen computer cannot determine which is which.
- To pass the **Turing Test**, a computer needs to be able to understand a human language (NLP), to possess human intelligence (e.g., have a knowledge base), to reason using its stored knowledge, and to be able to learn from its experiences (machine learning).

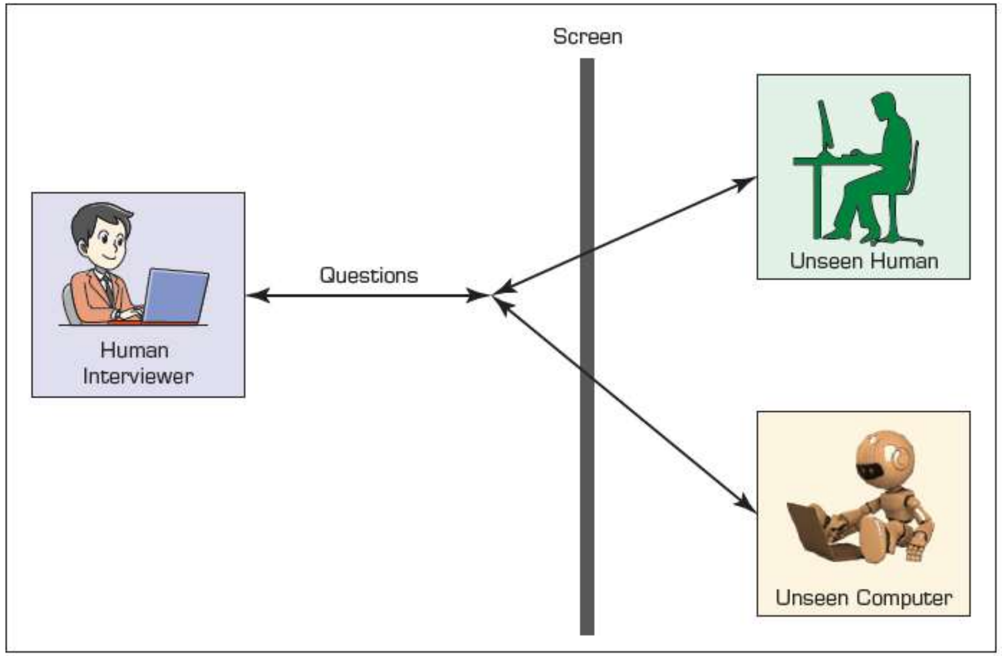


FIGURE 2.3 A Pictorial Representation of the Turing Test

2.4 Major AI Technologies And Some Derivatives

- Intelligent Agents
- Machine Learning
- Deep Learning
- Machine and Computer Vision
- Robotic Systems
- Natural Language Processing
- Knowledge and Expert Systems and Recommenders
- Chatbots
- Emerging AI Technologies

Intelligent Agents

- An **intelligent agent (IA)** is a small computer software program that observes and acts upon changes in its environment by running specific tasks autonomously.

- An IA directs an agent's activities to achieve specific goals related to the changes in the surrounding environment.

- IAs have the ability to learn by using their knowledge

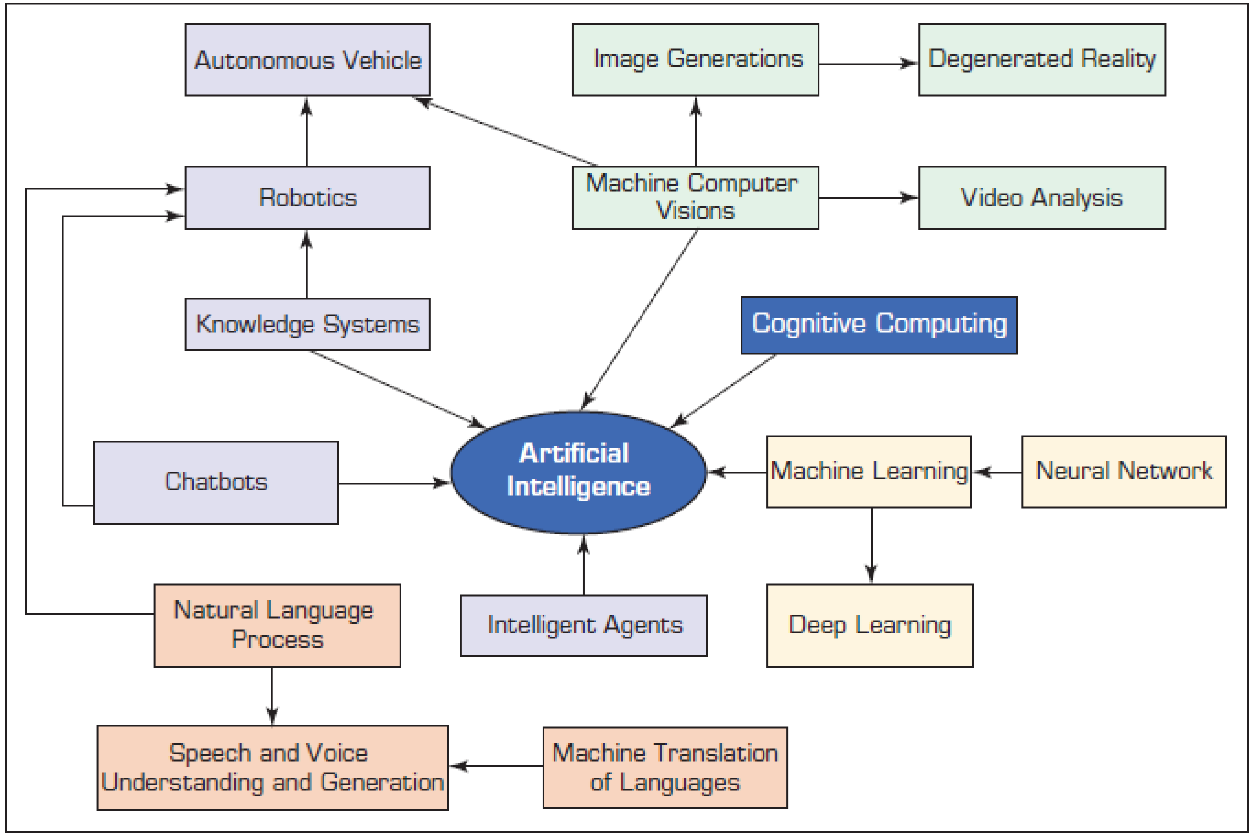


FIGURE 2.4 The Major AI Technologies

Intelligent Agents (cont.)

- IAs are tools for **overcoming Internet** information overload and making computers viable decision support tools.

- **Example 1:** An example of an intelligent software agent is a virus detection program. It resides in a computer, scans incoming data, and removes viruses automatically while learning to detect new virus types and detection methods.

- **Example 2:** Allstate Business Insurance uses an intelligent agent to reduce call center traffic and provide human insurance agents during the rate-quoting process with business customers.

Machine Learning

- **Machine Learning** (ML) is a discipline concerned with design & development of algorithms that allow computers to learn based on incoming data.
- ML allows computer systems to monitor and sense their environment, so that the machines can adjust their behavior to deal with the changes
- ML scientists teach computers to identify patterns and make connections by showing the machines a large volume of examples and related data.
- **ML used** for predicting, recognizing patterns, & supporting decision makers.
- **ML example** situations to a machine-learning program it can find hidden elements. An example is computers detecting credit card fraud.
- **ML applications** are expanding due to the availability of Big Data sources, especially those provided by the IoT.

Deep Learning

- One subset of machine learning is called deep learning; a technology tries to mimic how the human brain works.
- **Deep learning** (DL) uses artificial neural technology and deals with complex applications that regular machine learning and AI technologies can not handle.
- DL delivers systems that think and learn, enabling self-direction. This allows DL to tackle previously unsolvable problems using powerful learning algorithms.
- For example, DL is a key technology in autonomous vehicles by helping to interpret road signs and road obstacles.
- DL is mostly useful in real-time interactive applications in the areas of machine vision, scene recognition, robotics, and speech and voice processing.

Machine and Computer Vision

- **Machine vision** includes "technology and methods used to provide imaging-based automated inspection and analysis for applications such as robot guidance, process control, autonomous vehicles, and inspection."
- **Computer vision** "is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do."
- Both computer vision and machine vision automate many human tasks (e.g., inspection). However, Machine vision is treated more as an engineering subfield, while computer vision belongs to the computer science area.

Machine and Computer Vision

- The two technologies are combined with image processing that facilitates complex applications, such as visual quality control.
- Applied area of machine vision is scene recognition, done by computer vision.
- Video analytics is a derivative application of computer vision, where techniques are applied to videos to enable pattern recognition and identify events.
- Example of Applications:
    - The machine vision wood identification project developed a prototype machine vision system for wood identification to help identify illegal logging.
    - AI computer vision mixed with deep learning identifies illegal animal poachers.
    - Facial recognition that employ smart glasses to identify potential suspects.

Robotic Systems

- A robot is a device guided by a program to perform manual/mental tasks.
- An "intelligent" robot has a sensory apparatus (e.g., camera) to collects information about the surroundings and can respond to the changes in the environment.
- Autonomous robots (programmed to do tasks completely on their own, even repair themselves), are equipped with AI intelligent agents.
- Example: Walmart Is Using Robots to Properly Stock Shelves

  ○ In Walmart, 2-foot-tall robots use cameras/sensors to scan the shelves for misplaced, missing, or mispriced items. The results are transmitted to humans who take corrective actions. The robots carry out their tasks faster and frequently more accurately than humans.

Natural Language Processing

- **Natural language processing** (NLP) allows users to communicate with a computer in their native language. NLP includes two subfields:
  - NLP that investigates methods of enabling computers to comprehend instructions or queries provided in English or other human languages.
  - NLP generation that strives to have computers produce ordinary spoken language so that people can understand the computers more easily.
- **Speech/Voice Understanding**: recognition & comprehension of spoken languages by a computer. This has been adopted in automated call centers.
- **Machine Translation Of Languages:** uses computer programs to translate words and sentences from one language to another.

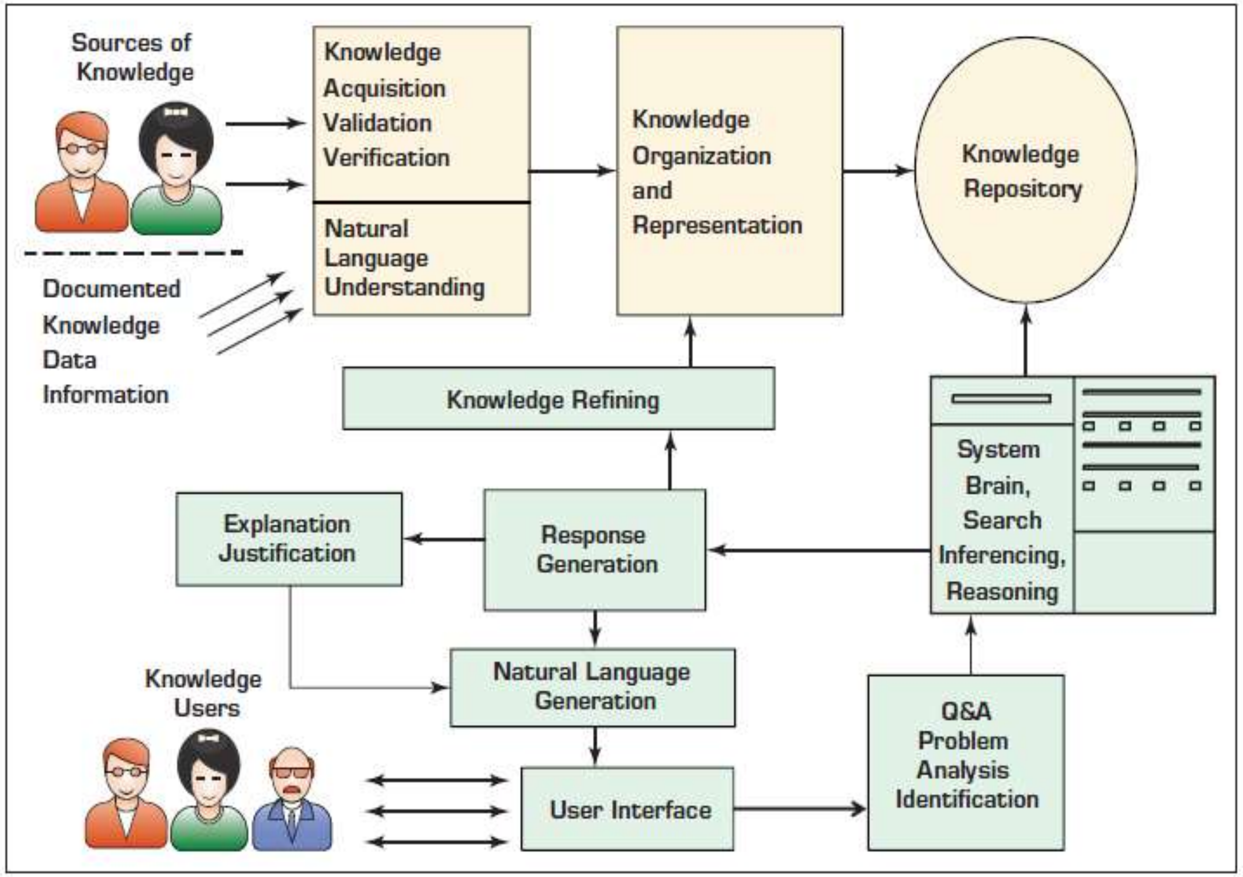
Knowledge and Expert Systems and Recommenders

- These systems are computer programs that store knowledge, which their applications use to generate expert advice and/or perform problem solving.
- **Knowledge-based expert** systems help people to verify information and make certain types of automated routine decisions.
- **Recommendation systems** are knowledge-based systems that make recommendations to people. Another knowledge system is chatbots.
- **Knowledge Sources And Acquisition For Intelligent:** Intelligent systems must gain knowledge through knowledge acquisition.
- Knowledge acquisition includes extracting and structuring knowledge from data, and then experts may be used to verify it.

Knowledge and Expert Systems and Recommenders

- **Knowledge Representation:** the organization and storage of acquired knowledge. The method of representation depends on the use of knowledge, reasoning from knowledge, user interaction with the knowledge, and more.

- **Reasoning From Knowledge:** processing a users' requests and gives answers. The difference among intelligent technologies is their type of reasoning.



Chatbots

- Robots come in several shapes and types, of which is a chatbot. A **chatbot** is a conversional robot that is used for chatting with people.
- Depending on the purpose of the chat, which can be done in writing or by voice, bots can be in the form of intelligent agents that retrieve information or personal assistants that provide advice.
- In either case, chatbots are usually equipped with NLP that enables conversations in natural human languages rather than in a programmed computer language.

Emerging AI Technologies

- Several new AI technologies are emerging. Here are a few examples:
  - **Effective computing.** Technologies that detect emotional conditions of people and suggest how to deal with discovered problems
  - **Biometric analysis**. Technologies that verify an identity based on unique biological traits that are compared to stored ones (e.g., facial recognition).
- **Cognitive Computing:** The application of knowledge derived from cognitive science so that computers can exhibit and/or support decision-making and problem-solving capabilities.
- **Augmented Reality:** Augmented reality (AR) refers to the real time integration of digital information with a users environment (mostly vision and sound). The technology provides a real-world interactive experience with the environment.

2.5 AI Support For Decision Making

- Issues and Factors in Using AI in Decision Making
- AI Support of the Decision-Making Process
- Automated Decision Making

Issues and Factors in Using AI in Decision Making

- These factors determine the justification of AI usage and its chance of success:
  - The nature of the decision (E.g., routine decisions = likely automated)
  - The method of support, what technologies are used
- **Cost-benefit & Risk Analyses:** necessary for large-scale decisions, but hard to compute with AI models due to difficulties in measuring costs, risks, & benefits.
- **Using Business Rule:** AI systems can be based on business rules, whose quality determines that of the automated.
- **AI Algorithms:** are the basis for automated decisions & decision support.
- **Speed:** Decision automation is dependent on the speed decisions needs to be made. Some decisions takes too much time to get all the relevant input data.

AI Support of the Decision-Making Process

➢ AI support can be applied to the various **steps** of the **decision-making process**:

1. **Problem Identification:** collecting data through technology that can be used by AI algorithms. Performance levels of machines are compared to standards, and trend analysis can point to opportunities.

2. **Generating/Finding Alternative Solutions:** matching problem characteristics with best practice, or proven solutions stored in databases. Tools such as case-based reasoning and neural computing are for this purpose.

3. **Selecting a Solution:** evaluate proposed solutions, predict future impacts, asses chance of success, or predict a reply to actions taken by a competitor.

4. **Implementing the Solutions:** demonstrates the superiority of proposals and to assess resistance to changes.

Automated Decision Making

- The process of automated decision making starts with knowledge acquisition and creation of a knowledge repository.
- The system generates and submits responses to user's questions. Solutions are evaluated to improve the knowledge repository, and complex situations are forwarded to humans.
- Companies use automated decision making for both their external operations (e.g., sales) and internal operations (e.g., resource allocation)

    Example: Supporting Nurses' Diagnosis Decisions

- Researchers used AI tools to conduct data mining to predict the probable success of automated nursing diagnoses based on patient characteristics.

2.6 AI Applications in Accounting

- AI in Accounting: An Overview
- Accounting Applications in Small Firms
- Job of Accountants

AI in Accounting: An Overview

- The CEO of an Accounting Software for businesses noticed trends among professional accountants: an increase in their use of AI, including bots in professional routines.

- It was observed that the major drivers for this are perceived savings in time and money and increased accuracy and productivity. The adoption has been rapid and it has been followed by significant improvements.

- An example is the execution of compliance procedures. For instance, machine learning is used for detecting anomalous data (e.g., fraudulent invoices).

Accounting Applications in Small Firms

- Crowe Horwath accounting firm uses AI to solve **healthcare billing problems**.
- Other applications like analyzing real estate contracts & risk analysis are used.

- A comprehensive study of AI in accounting report the advantage of AI as:
  - Provides cheaper, better data to support decision making
  - Generating insight from data analysis
  - Freeing time of accountants to concentrate on problem solving

- The report points to the use of the following:
  - Machine learning for detecting fraud and predicting fraudulent activities
  - Knowledge-based systems for verifying of accounting tasks
  - Deep learning to analyze unstructured data, such as contracts and e-mails

Job of Accountants

- AI and analytics will automate many routine tasks done today by accountants, many of whom may lose their jobs
- On the other hand, accountants will need to manage the used AI-based accounting systems
- Accountants will need to drive AI innovation in order to succeed or possibly even survive in the business

2.7 AI Applications in Financial Services

- AI Activities in Financial Services
- AI in Banking: An Overview
- Illustrative AI Applications in Banking
- Insurance Services

AI Activities in Financial Services

- These activities may be found across various types of financial services:
  - Extreme personalization (e.g., chatbots, assistants, robo investment advisors)
  - Shifting customer behavior both online and in brick-and-mortar branches
  - Facilitating trust in digital identity
  - Revolutionizing payments
  - Sharing economic activities (e.g., person-to-person loans)
  - Offering financial services 24/7 and globally (connecting the world)

AI in Banking: An Overview

- AI is transforming the **banking industry** in areas of AI IT, finance and accounting, marketing & sales, human resource management, customer service, and operations.

- The following are some points on the relation between AI and banking:

  - AI technologies in banking include many analytical tools.
  - These technologies help banks improve both their front-office and back-office operations.
  - Major activities are the use of chatbots to improve customer service and communicating with customers, and robo advising is used by some financial institutions.
  - Facial recognition is used for safer online banking.
  - Advanced analytics helps customers with investment decisions.
  - AI algorithms help banks identify and block fraudulent activities (E.g., money laundering).
  - AI algorithms can help in assessing the creditworthiness of loan applicants.

Illustrative AI Applications in Banking

- The following are applications of AI in banking:
  - **AI machines** that step up employee surveillance, and prevent illegal activities
  - **Applications** that ensures individuals pay what they owe, and tries to lower tax bills.
  - **Bots** that assist staff members to find the appropriate answers to queries in real time.
  - **Robots** that analyze customer facial expression and determine their nationality. Then determines and selects the matching language of the customers.

- Example: Banks Manage Compliance and Supports Decision Making

  - Banks must examine huge amounts of data to ensure compliance with government regulations. Thus, a set of cognitive tools dealing with regulatory compliance was created.
  - One of the tools deals with financial crimes, flagging suspicious transactions and fraud. The second tool monitors compliance, and the third deals with large volumes of data.

Insurance Services

- Advancements in AI are improving issuing policies and handling claims. Incoming claims are analyzed by AI, and are sent to appropriate available adjusters.
- The AI software can help in data collection and analysis and in data mining old claims.
- The machines provide speed, accuracy, and efficiency in performing this process. AI improves analysis results and enhance customer experience.

- **Example: Metromile Uses AI in Claim Processing**

- Metromile is a vehicle insurance company, using the pay-per-mile model. It uses an AI-based program to automate accident data, process claims, and pay customer claims.
- The AI bot simulates the accidents' major points and makes a verification based on decision rules; authorization for payments provides for successful verification.
- Only complex cases are sent to investigation by human processors.

2.8 AI in Human Resource Management (HRM)

- AI in HRM: An Overview
- AI in Onboarding
- Introducing AI to HRM Operations

AI in HRM: An Overview

- The following are ways in which AI can transform HRM in recruiting:
  1. Reducing human bias & Improving relationships with current employees
  2. Increasing efficiency, productivity, and insight in evaluating candidates

- The impact of AI into the following areas:

- **Recruitment (Talent Acquisition):** AI improves recruiting by helping recruiters and job seekers by using AI algorithms to suggest matches. This removes biases and prejudice.
- **AI Facilities Training:** AI methods can be used to facilitate learning. For example, chatbots can be used as a source of knowledge to answer learners' queries.
- **AI Supports Performance Analysis (Evaluation):** AI enables HR management to conduct performance analysis and measure the performance of employees. The performance is compared to objectives, which are provided to employees and teams.
- **AI Use In Retention And Attrition Detection:** Machine learning can be used to detect reasons why employees leave companies by identifying influencing patterns.

AI in Onboarding

- The HR department needs help introducing new employees to the organizational culture and operating processes.
- Some new employees require much attention and AI can help prepare customized onboarding paths that are best for the newcomers.
- Results showed that those employees supported by AI-based plans tend to stay longer in organizations
- The use of chatbots in HRM is increasing rapidly. Their ability to provide current information to employees anytime is a major reason.

Introducing AI to HRM Operations

- Introducing AI to HRM operations is like introducing AI to functional areas.
- The following are suggested activities of AI in this area:
  - Experiment with a variety of chatbots
  - Develop a team approach involving other functional areas
  - Properly plan a technology roadmap for both the short and long term, including shared vision with other functional areas
  - Identify new roles & modifications to existing roles in transformed environments
  - Train and educate the HRM team to understand AI and gain expertise in it

2.9 AI in Marketing, Advertising, and CRM

- Overview of Major Applications
- Customer Experiences and CRM
- Other Uses of AI in Marketing

Overview of Major Applications

➢ The following are examples of AI in marketing:

1. Product & personal recommendations
2. Smart search engines
3. Fraud and data breaches detection
4. Social semantics
5. Website design
6. Producer pricing
7. Predictive customer service
8. Ad targeting
9. Speech/Image recognition
10. Language translation
11. Customer Segmentation
12. Sales forecasting
13. Content generation
14. Using bots, robo advisors, etc.

Customer Experiences and CRM

- A major impact of AI technologies is changing customer experiences through:

1. Use NLP for creating user documentation. This improves customer–machine dialogue.
2. Use visual categorization to organize images (E.g., IBM's Visual Recognition & Clarifai)
3. Provide personalized and segmented services by analyzing customer data. This includes improving shopping experience and CRM.

- Example: Salesforce's AI Einstein

- Salesforce Einstein is an AI set of technologies used for enhancing customer interactions and supporting sales. The system delivers dynamic sales dashboards to sales reps, tracks performance and manages teamwork by using sales analytics. The AI product also can provide predictions and recommendations.
- Einstein's automatically prioritized sales leads make sales reps more productive when dealing with sales leads and potential opportunities. The sales reps also get insights about customers' sentiments, competitors' involvement, and other information.

Other Uses of AI in Marketing

- The following show the diversity of AI technologies used in marketing:

- **Bots** can mimic the expertise of in-store sales people, and make shopping easier.
- It provides lead generation. AI can help generate sales leads by analyzing customers' data. The program can generate predictions. Insights can be generated by intelligent analytics.
- It can increase customer loyalty using personalization.
- It can improve the sales pipeline. Robots can convert unknown visitors into customers using the **following three stages**:
  1. Prepare a list of target customers in the database
  2. Send information, ads, videos, and so on to prospects on the list created earlier
  3. Provide the company sales department with a list of leads that successfully convert potential customers to buyers.

2.10 AI Applications in Production-Operation

- AI in Manufacturing
- Implementation Model
- Intelligent Factories
- Logistics and Transportation

AI in Manufacturing

- To handle increasing labor costs, customers' requirement changes, increased competition, etc. companies use high levels of automation/digitization.
- Companies need to be more agile, and react quicker and more effectively.
- They also need to be more efficient and improve customers' (organizations' and individuals') experiences.
- To achieve these goals, they need to automate processes and make use of AI and other cutting-edge technologies.

Implementation Model

- **The five-component model to use intelligent technologies includes:**
  1. Streamlining processes, including minimizing waste, redesigning processes, and using business process management (BPM)
  2. Outsourcing certain business processes, including going offshore
  3. Using intelligence in decision making by deploying AI and analytics
  4. Replacing human tasks with intelligent automation
  5. Digitizing customers’ experiences

- Companies use robots for complex tasks, where mental tasks are automated. AI and sensors allow supporting or automating production decisions in real time.

Intelligent Factories

- Companies employ smart or intelligent factories, who use complex software and sensors.
  - An example of a lead supplier is General Electric, which provides software such as OEE Performance Analyzer and Production Execution Supervisor. .
- Many small vendors are specializing in different aspects of AI for manufacturing.
  - For example, BellHawk Systems Corporation, provides services to small companies, specializes in real-time operations tracking.

Logistics and Transportation

- AI and intelligent robots are used extensively in corporate logistics and internal and external transportation, as well as in supply chain management.
- Example: DHL Supply Chain

○ DHL is a global delivery company developing innovative logistics and transportation business models, mostly with AI, IoT, and machine learning.

○ Several of the IoT projects are linked to machine learning, in the areas of sensors, communication, device management, security, and analysis.

○ Tagging and tracking items using Radio Frequency Identification (RFID) and Quick Response (QR) code allow for item tracking along the supply chain.

Main Reference

- **Chapter 2**: "Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support" from "Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".

Week self-review exercises

- **Application Case 2.1 - 2.7** from "Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support"

Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

College of Computing and Informatics

Week 7

Chapter 8 – Part 2 (sections 8.5 & 8.6: Modeling in LP: An Example)

Analytics, Data Science, & Artificial Intelligence Systems For Decision Support

Chapter 2: Working with Data-Analysis Tools

Excel Data Analysis For Dummies

Chapters 21 & 25: “Customer Value, Monte Carlo Simulation, and Marketing Decision Making” & “ Using Classification Trees for Segmentation”

Marketing Analytics: Data-Driven Techniques with Microsoft Excel

This Presentation is mainly dependent on this textbooks

Contents

Weekly Learning Outcomes

1. Describe how spreadsheets can be used for analytical modelling and solutions

2. Perform calculations using Excel to help construct and interpret decision trees

3. Utilize Monte Carlo Simulation and Linear Programming Optimization in Excel

Required Reading

- **Chapter 8:** "Prescriptive Analytics: Optimization and Simulation" (sections 8.5 & 8.6: Modeling in LP: An Example) from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".
- **Chapter 2:** Working with Data-Analysis Tools from "*Excel Data Analysis For Dummies*"
- **Chapters 21 & 25**: "Customer Value, Monte Carlo Simulation, and Marketing Decision Making" & " Using Classification Trees for Segmentation" from "*Marketing Analytics: Data-Driven Techniques with Microsoft Excel*"

Recommended Reading

- Decision Table: Clearer Logic and Better Decision
  https://warren2lynch.medium.com/decision-table-clearer-logic-and-better-decision-603c3cff9f24

Recommended Video

- Monte Carlo Simulations in Excel
  https://www.youtube.com/watch?v=UeGncSFijUM
- Linear Programming (LP) Optimization with Excel Solver
  https://www.youtube.com/watch?v=6xa1x_Iqjzg

Practical Portion

Chapter 2: Working with Data-Analysis Tools

Excel Data Analysis For Dummies

Chapter 2: Working with Data-Analysis Tools

- Working with Data Tables
- Creating a basic data table
- Creating a two-input data table
- Skipping data tables when calculating workbooks
- Analyzing Data with Goal Seek
- Analyzing Data with Scenarios
- Create a Scenario
- Apply a Scenario
- Edit a Scenario
- Delete a Scenario
- Optimizing Data with Solver
- When should you use Solver?
- Loading the Solver Add-in
- Optimizing a result with Solver
- Adding Constraints to a Solver
- Save a Solver solution as a scenario

Working with Data Tables

- If you want to study the effect that different input values have on a formula, one solution is to set up the worksheet model and then manually change the formula's input cells.
- The problem with modifying the values of a formula input is that you see only a single result at one time.
- A better solution is to set up a data table, which is a range that consists of the formula you're using and multiple input values for that formula. Excel automatically creates a solution to the formula for each different input value.

Creating a basic data table

The most basic type of data table is one that varies only one of the formula's input cells. Here are the steps to follow to create a one-input data table:

1. Type the input values.
   - To enter the values in a column, start the column one cell down and one cell to the left of the cell containing the formula, as shown in Figure 2-1.
   - To enter the values in a row, start the row one cell up and one cell to the right of the cell containing the formula.
2. Select the range that includes the input values and the formula.

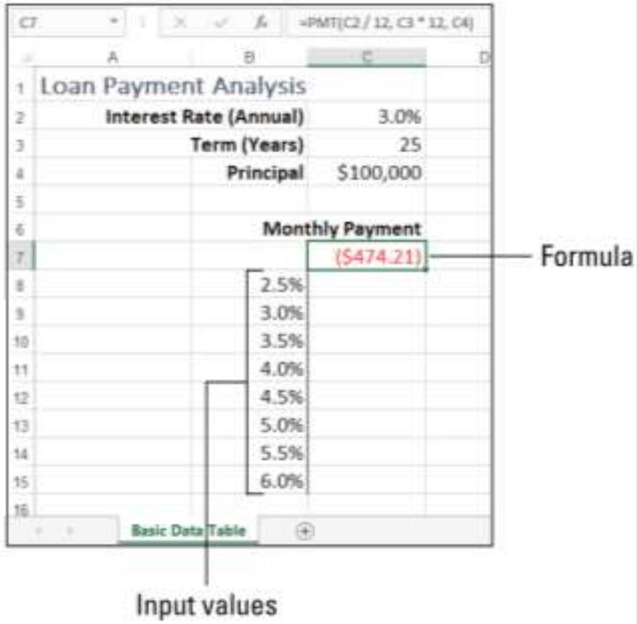


FIGURE 2-1: This data table has the input values in a column.

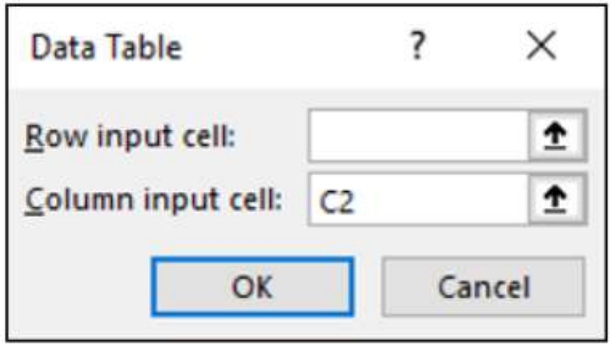


FIGURE 2-2: Enter the address of the input cell.

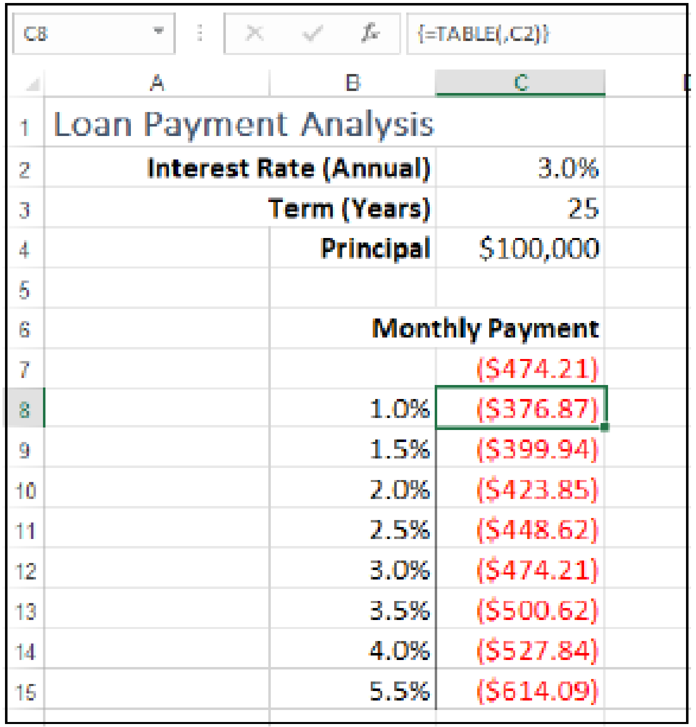
Creating a basic data table

3. Choose Data ➪ What-If Analysis ➪ Data Table to open the Data Table dialog box.

4. Enter the address of the input cell; the cell referenced by the formula you want the data table to vary.

5. Click OK.

- When you see the data table results, you might find that all the calculated values are identical.
- The problem most likely is Excel's current calculation mode.
- Choose Formulas ➪ Calculation Options ➪ Automatic, and the data table results should recalculate to the correct values.



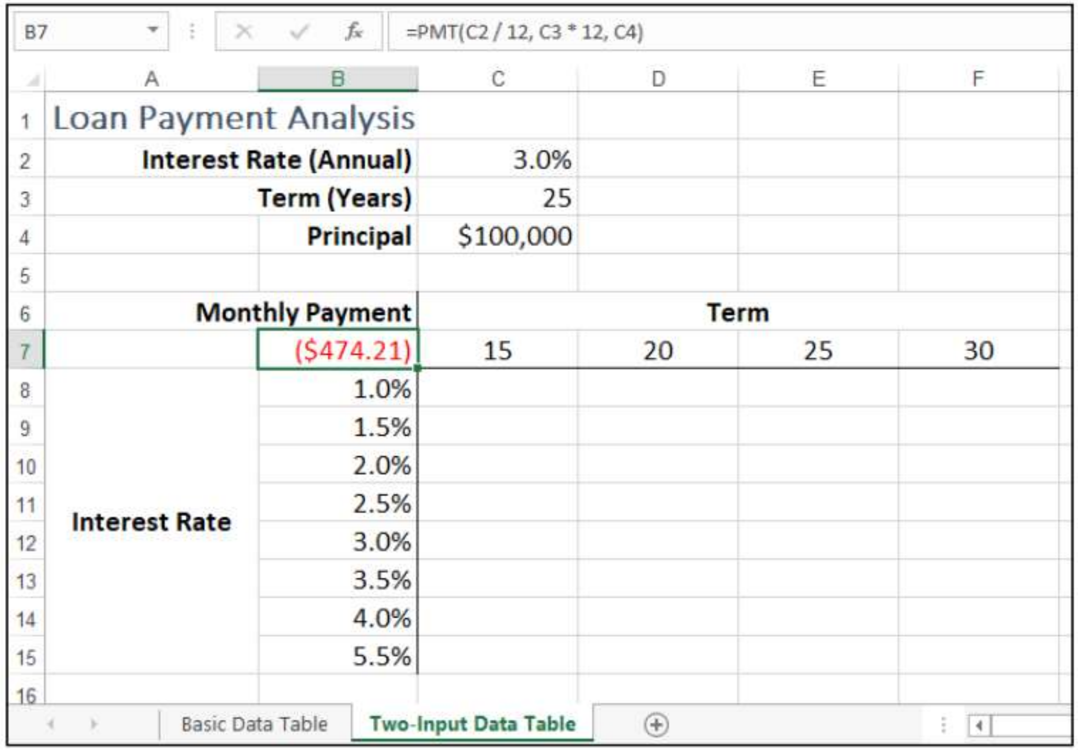
	A	B	C
1	Loan Payment Analysis		
2		Interest Rate (Annual)	3.0%
3		Term (Years)	25
4		Principal	$100,000
5			
6			Monthly Payment
7			($474.21)
8		1.0%	($376.87)
9		1.5%	($399.94)
10		2.0%	($423.85)
11		2.5%	($448.62)
12		3.0%	($474.21)
13		3.5%	($500.62)
14		4.0%	($527.84)
15		5.5%	($614.09)

FIGURE 2-3: The data table results.

Creating a two-input data table

- A two-input data table is one that varies two formula inputs at the same time.
- To set up a two-input data table, you must set up two ranges of input cells.
- Here are the steps to follow:

1. Type the input values:
2. Choose Data ➪ What-If Analysis ➪ Data Table to open the Data Table dialog box.
3. In the Row Input Cell text box, enter the cell address of the input cell that corresponds to the row values you entered.



B7 =PMT(C2 / 12, C3 * 12, C4)

	A	B	C	D	E	F
1	Loan Payment Analysis					
2		Interest Rate (Annual)	3.0%			
3		Term (Years)	25			
4		Principal	$100,000			
5						
6		Monthly Payment		Term		
7		($474.21)	15	20	25	30
8		1.0%				
9		1.5%				
10		2.0%				
11	Interest Rate	2.5%				
12		3.0%				
13		3.5%				
14		4.0%				
15		5.5%				
16						

FIGURE 2-4: For a two-input data table, enter one set of values in a column and the other in a row.

Creating a two-input data table

5. In the Row Input Cell text box, enter the cell address of the input cell that corresponds to the row values you entered.

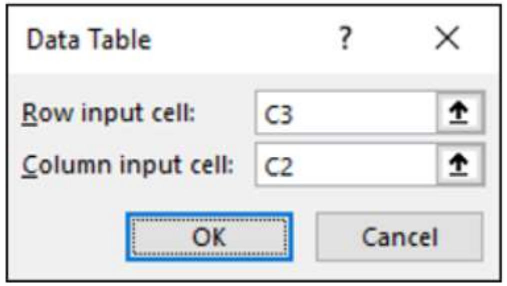
6. In the Column Input Cell text box, enter the cell address of the input cell you want to use for the column values.

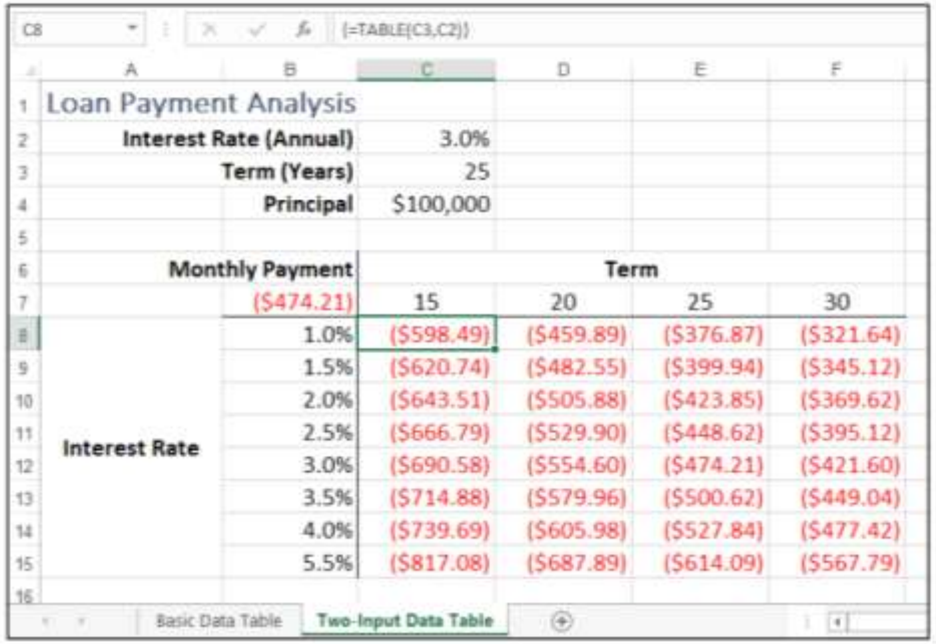
6. Click OK.

- When you run the Data Table command, Excel enters an array formula in the interior of the data table.
- The formula is a TABLE function with the following syntax:

{=TABLE(row_input_ref, column_input_ref)}

FIGURE 2-5: Enter the addresses of the input cells.





	A	B	C	D	E	F
1	Loan Payment Analysis					
2		Interest Rate (Annual)	3.0%			
3		Term (Years)	25			
4		Principal	$100,000			
5						
6		Monthly Payment		Term		
7		($474.21)	15	20	25	30
8		1.0%	($598.49)	($459.89)	($376.87)	($321.64)
9		1.5%	($620.74)	($482.55)	($399.94)	($345.12)
10		2.0%	($643.51)	($505.88)	($423.85)	($369.62)
11		2.5%	($666.79)	($529.90)	($448.62)	($395.12)
12	Interest Rate	3.0%	($690.58)	($554.60)	($474.21)	($421.60)
13		3.5%	($714.88)	($579.96)	($500.62)	($449.04)
14		4.0%	($739.69)	($605.98)	($527.84)	($477.42)
15		5.5%	($817.08)	($687.89)	($614.09)	($567.79)
16						

FIGURE 2-6: The two-input data table results.

Skipping data tables when calculating workbooks

- If you're working with a large data table, you can reduce the time it takes for Excel to recalculate the workbook if you configure Excel to bypass data tables when it's running the recalculation. Here are the two methods you can use:
  - Choose Formulas ➪ Calculation Options ➪ Automatic Except for Data Tables.
  - Choose File ➪ Options to open the Excel Options dialog box, choose Formulas, select the Automatic Except for Data Tables option, and then click OK.
- When you want to recalculate a data table, you can repeat either of the preceding procedures and then choose the Automatic option.
- If you prefer to leave the Automatic Except for Data Tables option selected, you can still recalculate the data table by selecting any cell inside the data table and either choosing Formulas ➪ Calculate Now or pressing F9.

Analyzing Data with Goal Seek

- Goal Seek works by trying dozens of possibilities that enable it to get closer to a solution. When Goal Seek finds a solution, it stops and shows you the result.
- You must do three things to set up your worksheet for Goal Seek:

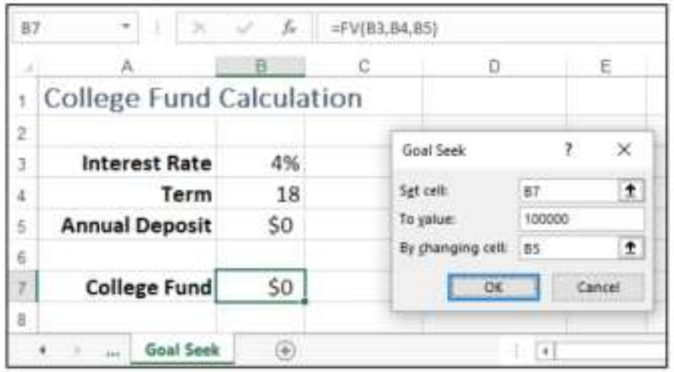
1. Set up one cell as the changing cell, which is the formula input cell value that Goal Seek will manipulate to reach the goal. In the college fund example, the formula cell that holds the annual deposit is the changing cell.
2. Set up the other input values for the formula and give them proper initial values. In the college fund example, you enter four percent for the interest rate and 18 years for the term.
3. Create a formula for Goal Seek to use to reach the goal. In the college fund example, you use the FV() function, which calculates the future value of an investment given an interest rate, term, and regular deposit.

Skipping data tables when calculating workbooks

Here are the steps to follow to get Goal Seek on the job:

1. Select Data ➪ What-If Analysis ➪ Goal Seek.
2. In the Set Cell box, enter the address of the cell that contains the formula you want Goal Seek to work with.
3. In the To Value text box, enter the value that you want Goal Seek to find.
4. In the By Changing Cell box, enter the address of the cell that you want Goal Seek to modify.
5. Click OK.
6. Then, Click OK to accept the solution.

FIGURE 2-7: Using Goal Seek to calculate the annual deposit required to end up with $100,000 in a college fund.



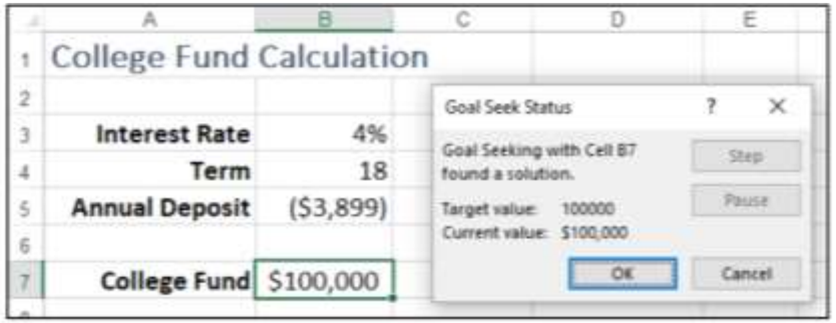


FIGURE 2-8: Goal Seek took all of a second or two to find a solution.

Analyzing Data with Scenarios

- In Excel, each of coherent set of input values is called a scenario. By creating multiple scenarios, you can quickly apply different value sets to analyze how the result of a formula changes under different conditions.

1. Excel enables you to enter up to 32 changing cells in a single scenario.
2. no matter how many changing cells you have in a scenario, Excel enables you to show the scenario's result with just a few taps or clicks.
3. The number of scenarios you can define is limited only by the available memory on your computer, you can use as many scenarios as you need to analyze your data model.

- When building a worksheet model, you can use a couple of techniques to make the model more suited to scenarios:
    - Group all your changing cells in one place and label them.
    - Make sure that each changing cell is a constant value.

Create a scenario

Here are the steps to follow to create a scenario for a worksheet model:

1. Choose Data ➪ What-If Analysis ➪ Scenario Manager.
2. Click Add.
3. In the Scenario Name box, type a name for the scenario.
4. In the Changing Cells box, enter the cells you want to change in the scenario.

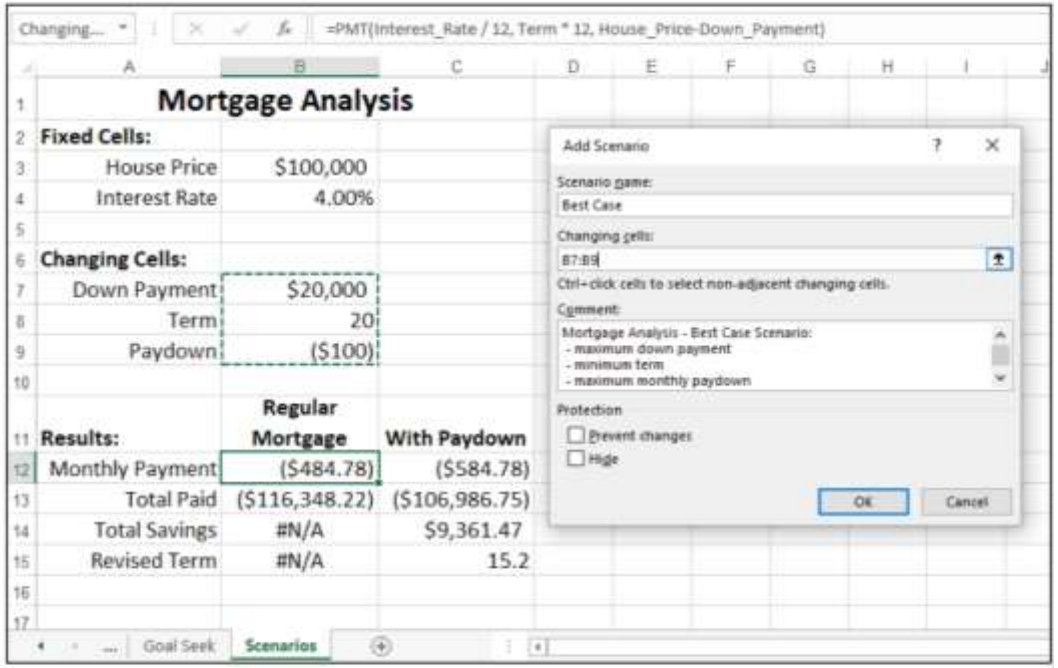


FIGURE 2-9: Creating a scenario for a mortgage analysis.

Create a scenario (cont.)

5. In the Comment box, enter a description for the scenario.
6. Click OK.
7. In the text boxes, enter a value for each changing cell.
8. To add more scenarios, click Add and then repeat Steps 3 through 7.
9. Click OK.

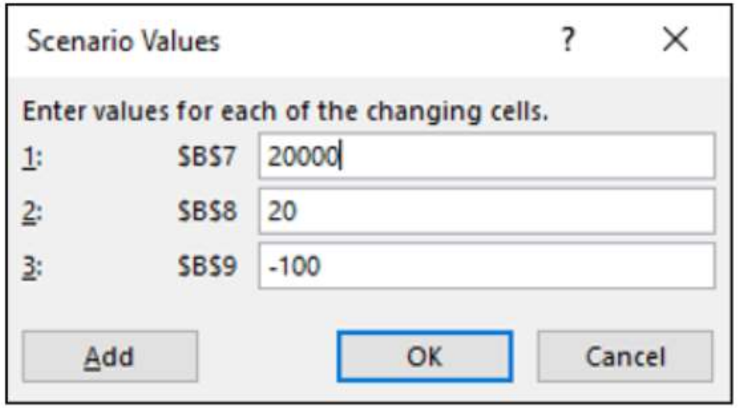


FIGURE 2-10: Example values for a scenario's changing cells.

Apply a scenario

The real value of a scenario is that no matter how many changing cells you've defined or how complicated the formula is, you can apply any scenario by:

1. Choose Data ➪ What-If Analysis ➪ Scenario Manager.
2. Select the scenario you want to display.
3. Click Show.
4. Feel free to repeat Steps 2 and 3 to display other scenarios.
5. When you've completed your analysis, click Close

Edit a scenario

If you need to make changes to a scenario, you can edit the name, the changing cells by:

1. Choose Data ➪ What-If Analysis ➪ Scenario Manager.
2. Select the scenario you want to modify.
3. Click Edit.
4. Modify the scenario name, changing cells, and comment, as needed. Click OK.
6. Modify the scenario values, as needed. Click OK.
7. Click Close

Delete a scenario

If you have a scenario that has worn out its welcome, you should delete it to reduce clutter in the Scenario Manager by:

1. Choose Data ➪ What-If Analysis ➪ Scenario Manager.
2. Select the cenario you want to remove.
3. Click Delete.
4. Click Close.

Optimizing Data with Solver

- Solver is a sophisticated optimization program that enables you to find the solutions to problems that would otherwise require high-level mathematical analysis.
- Solver, like Goal Seek, uses an iterative method to perform its calculations.
- Solver examines how results change with each iteration and, through some mathematical processes can tell in what direction it should head for the solution.
- Solver brings a number of advantages to the table:
  - Solver enables you to specify multiple adjustable cells.
  - Solver enables you to set up constraints on the adjustable cells.
  - Solver seeks not only a desired result but also the optimal one. For complex problems, Solver can generate multiple solutions.

When should you use Solver?

Many problems require the Solver approach. These problems cover many different fields and situations, but they all have the following characteristics in common:

- They have a single objective cell (also called the target cell) that contains a formula you want to maximize, minimize, or set to a specific value.
- The objective cell formula contains references to one or more variable cells (also called unknowns or changing cells).
- Optionally, there are one or more constraint cells that must satisfy certain criteria.

For example, Figure 2-11 shows a worksheet data model that's all set up for Solver.

	A	B	C	D
1				
2		Inflatable Dartboard	Dog Polisher	
3	Price	$24.95	$19.95	
4	Units	1	1	
5	Revenue	$25	$20	
6				
7	Unit Cost	$12.50	$9.50	
8	Variable Costs	$12	$9	
9	Fixed Costs	$100,000	$75,000	
10	Total Costs	$100,012	$75,009	
11				
12	Product Profit	-$99,987	-$74,989	
13				
14	Total Profit	-$174,975		
15				

FIGURE 2-11: The goal for this data model is to find the break-even point (where total profit is $0).

Loading the Solver add-in

- An **add-in is software** that adds one or more features to Excel. Installing add-ins gives you additional Excel features that aren’t available in the Ribbon by default.
- Several add-ins come standard with Excel, including Solver, which enables you to solve optimization problems.
- You install the bundled add-ins by using the Excel Options dialog box; you can find them in the Add-Ins section. After installation, add-ins appear on a tab related to their function.

Here are the steps to follow to load the Solver add-in:

1. Choose File ⇨ Options.
2. Choose Add-Ins.
3. In the Manage list, select Excel Add-Ins and then select Go.
4. Select the Solver Add-In check box
5. Click OK.

Optimizing a result with Solver

- When Solver finds a solution, you can choose either Keep Solver Solution or Restore Original Values. If you choose Keep Solver Solution, Excel permanently changes the worksheet. You cannot undo the changes.
- With your Solver-ready worksheet model ready to go, here are the steps to follow to find an optimal result for your model using Solver:

1. Choose Data ➪ Solver.
2. In the Set Objective box, enter the address of your model's objective cell.
3. In the To group, select an option:
    - **Max:** Returns the maximum possible value.
    - **Min:** Returns the minimum possible value.
    - **Value Of:** Enter a number to set the objective cell to that number.

Optimizing a result with Solver

4. In the By Changing Variable Cells box, enter the addresses of the cells you want Solver to change while it looks for a solution.

5. Click Solve.

6. In any Show Trial Solution dialog box that appears, click Continue to move things along.

7. Select the Keep Solver Solution option.

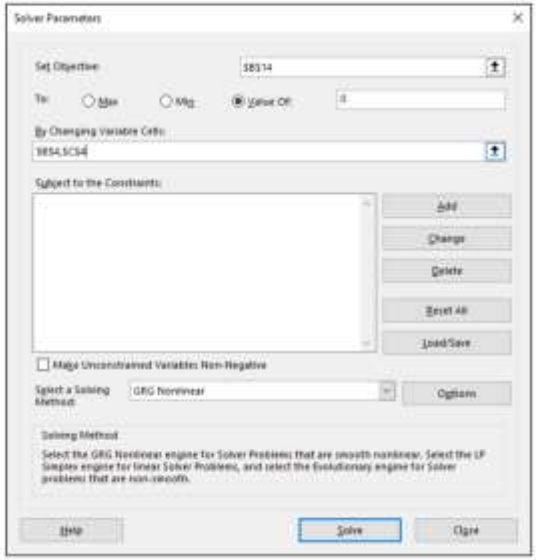


FIGURE 2-12: The completed Solver Parameters dialog box.

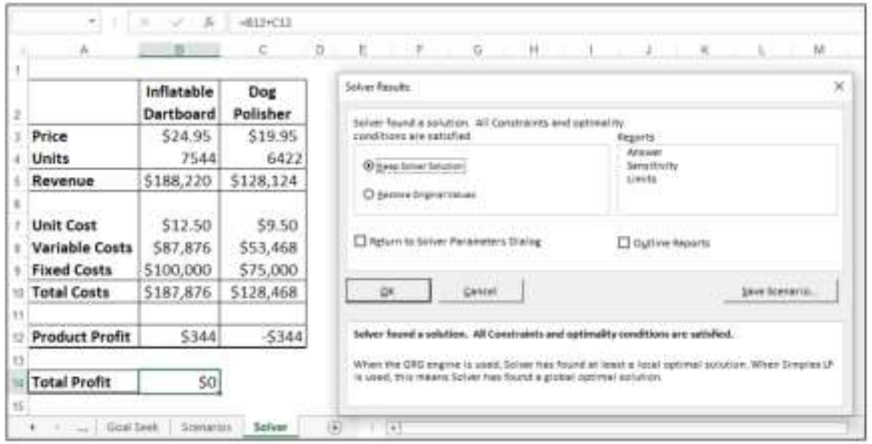


FIGURE 2-13: The Solver Results dialog box and the solution to the break-even problem.

Optimizing a result with Solver

- You can ask Solver to display one or more reports. In the Solver Results dialog box, use the Reports list to select each report you want to view:
  - **Answer:** Displays information about the model's objective cell, variable cells, and constraints.
  - **Sensitivity:** Attempts to show how sensitive a solution is to changes in the model's formulas. The Sensitivity report layout depends on the type of model you're using.
  - **Limits:** Displays the objective cell and its value, as well as the variable cells and their addresses, names, and values.
- Solver can use one of several solving methods. In the Solver Parameters dialog box, use the Select a Solving Method list to select one of the following:
  - **Simplex LP:** Use if your worksheet model is linear.
  - **GRG Nonlinear:** Use if your worksheet model is nonlinear and smooth.
  - **Evolutionary:** Use if your worksheet model is nonlinear and nonsmooth.

Adding constraints to Solver

- Restrictions and conditions are examples of what Solver calls constraints. Adding constraints tells Solver to find a solution so that these conditions are not violated.

Here's how to run Solver with constraints added to the optimization:

1. Choose Data ➪ Solver.
2. Use the Set Objective box, the To group, and the By Changing Variable Cells box to set up Solver as I describe in the previous section, "Optimizing a result with Solver."
3. Click Add.
4. In the Cell Reference box, enter the address of the cell you want to constrain.
5. In the drop-down list, select the operator you want to use.
6. If you chose a comparison operator in Step 5, in the Constraint box, enter the value by which you want to restrict the cell.

Adding constraints to Solver

- Restrictions and conditions are examples of what Solver calls constraints. Adding constraints tells Solver to find a solution so that these conditions are not violated.

Here’s how to run Solver with constraints added to the optimization:

1. Choose Data ➪ Solver.
2. Use the Set Objective box, the To group, and the By Changing Variable Cells box to set up Solver as I describe in the previous section, “Optimizing a result with Solver.”
3. Click Add.
4. In the Cell Reference box, enter the address of the cell you want to constrain.
5. In the drop-down list, select the operator you want to use.
6. If you chose a comparison operator in Step 5, in the Constraint box, enter the value by which you want to restrict the cell.

Adding constraints to Solver (cont.)

7. To specify more constraints, click Add and repeat Steps 4 through 6, as needed.
8. Click OK.
9. Click Solve.
10. In any Show Trial Solution dialog box that appears, click Continue to move things along.
11. Select the Keep Solver Solution option.
12. Click OK

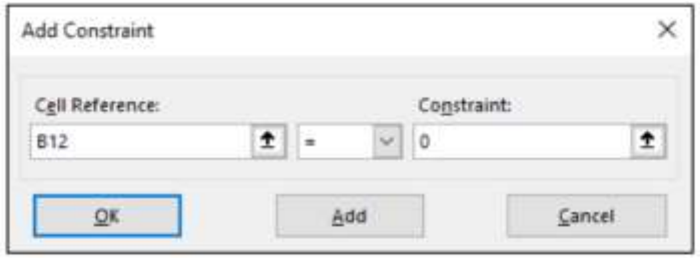


FIGURE 2-14: The completed Add Constraint dialog box.

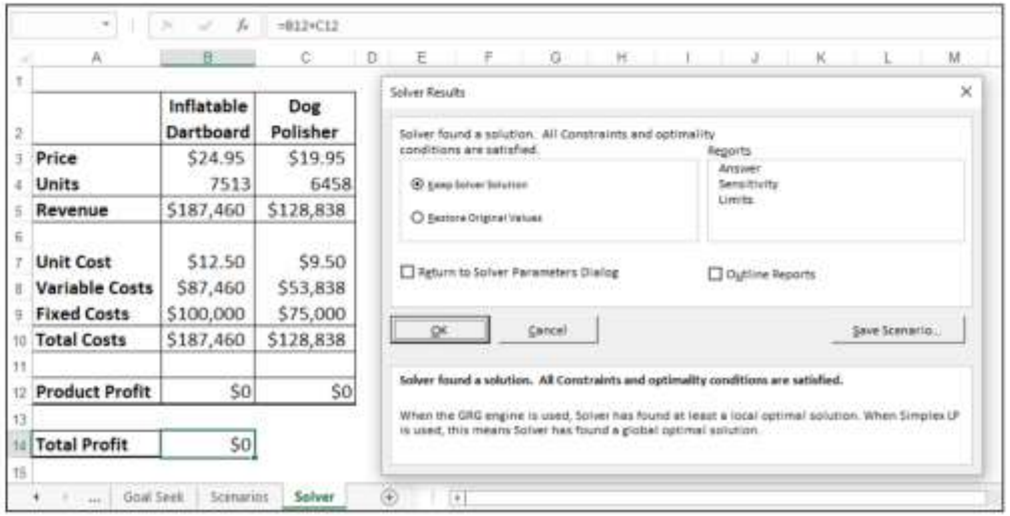


FIGURE 2-15: The Solver Results dialog box and the final solution to the break-even problem.

Save a Solver solution as a scenario

- Follow these steps to save a Solver solution as a scenario:

1. Choose Data ➪ Solver.
2. Use the Set Objective box, the To group, the By Changing Variable Cells box, and the Subject to the Constraints list to set up Solver.
3. Click Solve.
4. Anytime the Show Trial Solution dialog box appears, choose Continue.
5. Click Save Scenario.
6. In the Scenario Name dialog box, type a name for the scenario and then click OK.
7. Select the Keep Solver Solution option.
8. Click OK

Chapter 8 – Part 2 (sections 8.5 & 8.6: Modeling in LP: An Example)

Analytics, Data Science, & Artificial Intelligence Systems For Decision Support

8.5 Decision Modeling with Spreadsheets

- Application Case 8.4: Pennsylvania Adoption Exchange Uses Spreadsheet Model to Better Match Children with Families
- Application Case 8.5: Metro Meals on Wheels Treasure Valley Uses Excel to Find Optimal Delivery Routes
- Modelling with Spreadsheets

Decision Modeling with Spreadsheets

- Models can be developed and implemented in a variety of programming languages and systems. We focus primarily on spreadsheets, modeling languages, and transparent data analysis tools.
- Spreadsheet packages were recognized as easy-to-use software due to their strength and flexibility.
- Spreadsheet packages were used in the development of a wide range of applications in business, mathematics, and science.
- As spreadsheet packages evolved, add-ins were developed for structuring and solving specific model classes.
- Among the add-in packages, many were developed for DSS development.

Application Case 8.4: Pennsylvania Adoption Exchange Uses Spreadsheet Model to Better Match Children with Families

- The Pennsylvania Adoption Exchange (PAE) was established to help county and nonprofit agencies find prospective families for orphan children who had not been adopted due to age or special needs.
- The PAE started collecting information about the orphans and families through online surveys that include a new set of questions. The PAE and consultants created a spreadsheet matching tool. In this model, caseworkers can specify the weight of the attributes for selecting a family for a child.
- Using this tool, the matching committee can compare a child and family on each attribute, thus making a more accurate match decision between a family and a child.
- **Results/Benefits:** Since use of the new spreadsheet, the percentage of children getting a permanent home has increased.

Application Case 8.5: Metro Meals on Wheels Treasure Valley Uses Excel to Find Optimal Delivery Routes

- Meals on Wheels Association of America is a not-for-profit organization that delivers approximately one million meals to homes of older people in need.
- Metro Meals on Wheels wanted a routing tool that could generate routing solutions for both one-way and round-trip directions for delivering meals.
- To solve the routing problem, a spreadsheet-based tool was developed.
- Excel's Visual Basic was used to access a developer's networking map application programming interface (API) called MapQuest.
- This API was used to create a travel matrix that calculated time and distance needed for delivery of the meal.
- **Results/Benefits**: Metro Meals saved $5,800 in 2015. The tool reduced the time spent on route planning for meal deliveries. Also, It increased volunteer satisfaction.

Modeling with Spreadsheets

- **Goal seeking** is an important spreadsheet feature. It is performed by indicating a target cell, its desired value, and a changing cell.
- Static/dynamic models can be built in a spreadsheet. For example, the monthly loan calculation model shown in DS498_week7-Ch8c.xlsx.

  - A change in the interest rate is immediately reflected in the monthly payment.
  - If we require a specific monthly payment, we can use goal seeking to determine an appropriate interest rate or loan amount.
  - The results can be observed and analyzed immediately.
  - The **static model** indicates a single month's performance, which is replicated.

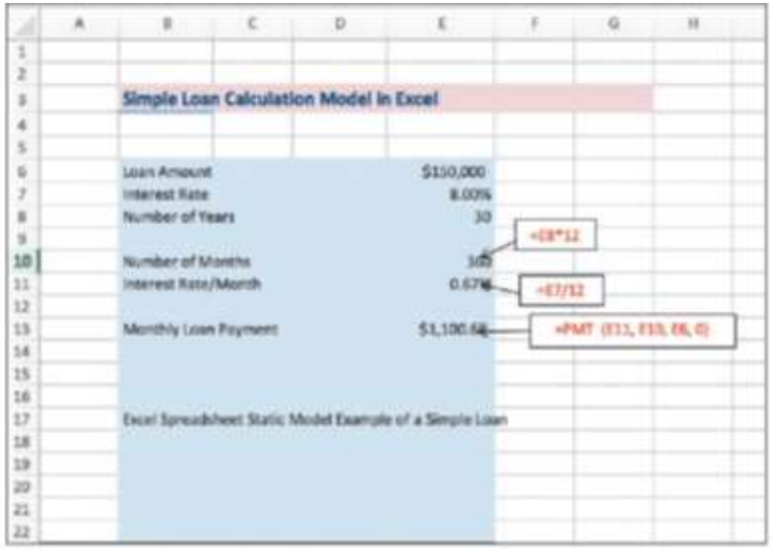


FIGURE 8.3 Excel Spreadsheet Static Model Example of a Simple Loan Calculation of Monthly Payments.

Modeling with Spreadsheets (cont.)

- A **dynamic model**, in contrast, represents behavior over time.

- The loan calculations in the spreadsheet DS498_week7-Ch8c.xlsx indicate the effect of prepayment on the principal over time.

- Risk analysis can be incorporated into spreadsheets by using built-in random-number generators to develop simulation models.

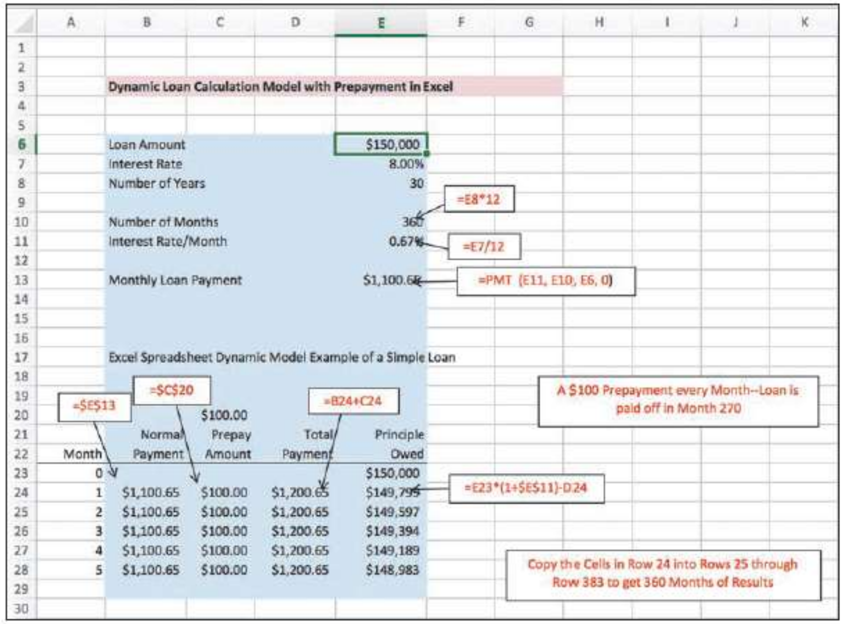


FIGURE 8.4 Excel Spreadsheet Dynamic Model Example of a Simple Loan Calculation of Monthly Payments and the Effects of Prepayment.

8.6 Mathematical Programming Optimization

- Modeling in LP: An Example

Modeling in LP: An Example

- **MBI Corporation**, which manufactures special-purpose computers, needs to **make a decision**: How many computers should it produce next month at the Boston plant?
- MBI is considering two types of computers:
  - CC-7, which requires 300 days of labor and $10,000 in materials, &
  - CC-8, which requires 500 days of labor and $15,000 in materials.
- The profit contribution of each CC-7 is $8,000, and of each CC-8 is $12,000.
- The plant has a capacity of 200,000 working days per month, and the material budget is $8 million per month.
- Marketing requires that at least 100 units of the CC-7 and at least 200 units of the CC-8 be produced each month.
- **Problem:** Maximize the company's profits by determining how many units of the CC-7 and how many units of the CC-8 should be produced each month.

Modeling in LP: An Example (cont.)

- The **problem** is to find the values of the decision variables $X1$, $X2$, such that the value of the result variable Z is maximized, subject to a set of linear constraints that express the technology, market conditions, and other uncontrollable variables.
- The mathematical relationships are all linear equations and inequalities.

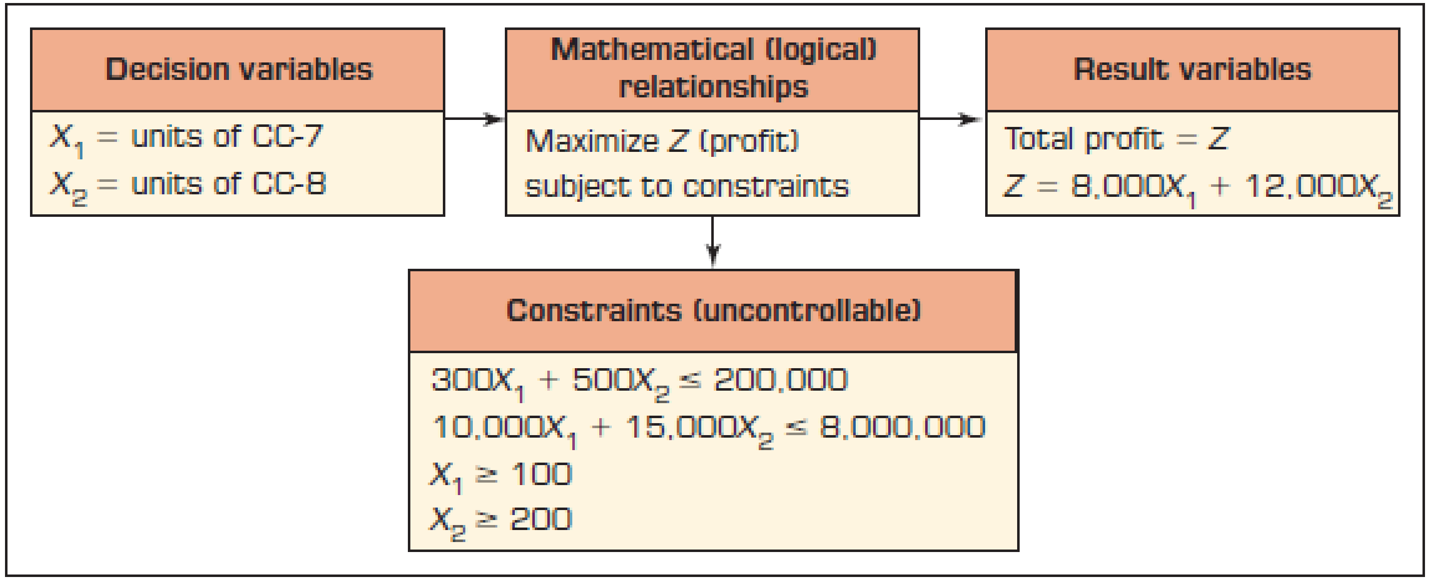


FIGURE 8.5 Mathematical Model of a Product-Mix Example.

Modeling in LP: An Example (cont.)

- Theoretically, any allocation problem of this type has an infinite number of possible solutions.
- Using special mathematical procedures, the LP approach applies a unique computerized search procedure that finds the best (optimal) solution/s (ex. maximizes total profit) in a matter of seconds.
- Excel 'add-in Solver' is used to obtain an optimal (best) solution to this problem. Open spreadsheet DS498_week7-Ch8d.xlsx

  1. Activate 'Solver' Under the Data tab and on the Analysis ribbon.

     If it is not there, you should be able to enable it by going to File -> Excel's Options Menu and selecting Add-ins.

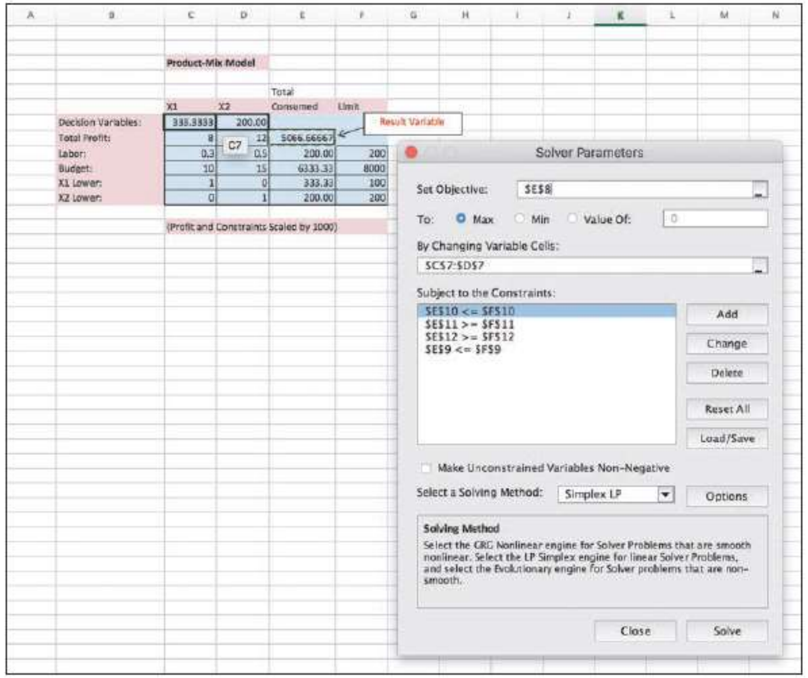


FIGURE 8.6 Excel Solver Solution to the Product-Mix Example.

Modeling in LP: An Example (cont.)

2. Enter these data directly into an Excel spreadsheet.
3. Identify the goal (by setting Target Cell equal to **Max**).
4. Identify decision variables (by setting By Changing Cells).
5. Identify constraints on labor capacity, budget, and the desired minimum production of the two products $X1$ and $X2$.
6. Clicking on the Solver Add-in opens a dialog box,
   - Specify the cells or ranges that define the objective function cell, decision/changing variables (cells), and the constraints.
   - Select the solution method (usually Simplex LP)
   - solve the problem.
7. Next, we select all three reports—Answer, Sensitivity, and Limits (optimal solution of $X1$ = 333.33, $X2$ = 200, Profit = $5,066,667)

Practical Portion

Chapter 21: Customer Value, Monte Carlo Simulation, and Marketing Decision Making

Marketing Analytics: Data-Driven Techniques with Microsoft Excel

Chapter 21: Customer Value, Monte Carlo Simulation, and Marketing Decision Making

- A Markov Chain Model of Customer Value
- Using Monte Carlo Simulation to Predict Success of a Marketing Initiative
- Using a One-Way Data Table to Simulate the Groupon Deal
- Using a Histogram to Summarize the Simulation Results
- Summary

A Markov Chain Model of Customer Value

- **Monte Carlo simulation** is a method to model uncertainty and estimate a range of outcomes under uncertainty by replaying a situation out many times.
- Imagine that a small mail-order firm mails out catalogs every three months. Based on analyses performed on historical purchase data, the marketing analyst found that each time a customer places an order, the profit earned follows a normal random variable with mean $60 and standard deviation $10.
- The probability that a customer orders from a catalog depends on
  - Recency: number of catalogs since last order, and
  - Frequency: total number of orders placed
- as shown in Figure 21-1 (next slide); see file DS498_week7_Markov.xls.

A Markov Chain Model of Customer Value

- For example, based on the preceding analysis, a customer (call her Miley) who has ordered twice and whose last order was three catalogs ago has a 6.9 percent Customer Value, Monte Carlo Simulation, and Marketing Decision Making chance of ordering.
- For example, in analyzing Miley’s future purchases you need only know she has ordered twice and her last order was three catalogs ago. You do not need to know, for example, the timing of other orders that were placed before the last order. This model is an example of a Markov Chain.

	C	D	E	F	G	H
1		Frequency				
2						
3	Recency	1	2	3	4	>=5
4	1	0.103	0.121	0.143	0.151	0.163
5	2	0.076	0.09	0.106	0.112	0.121
6	3	0.059	0.069	0.081	0.086	0.093
7	4	0.045	0.053	0.062	0.066	0.071
8	5	0.038	0.045	0.053	0.056	0.061
9	6	0.035	0.041	0.049	0.051	0.056
10	7	0.03	0.035	0.041	0.043	0.047
11	8	0.027	0.032	0.038	0.04	0.043
12	9	0.025	0.029	0.035	0.037	0.04
13	10	0.021	0.025	0.03	0.031	0.034
14	11	0.021	0.024	0.028	0.03	0.033
15	12	0.02	0.024	0.028	0.03	0.032
16	13	0.017	0.02	0.024	0.025	0.027
17	14	0.017	0.02	0.024	0.025	0.027
18	15	0.016	0.019	0.022	0.024	0.026
19	16	0.015	0.018	0.021	0.022	0.024
20	17	0.014	0.017	0.02	0.021	0.022
21	18	0.013	0.016	0.018	0.019	0.021
22	19	0.013	0.015	0.018	0.019	0.02
23	20	0.012	0.014	0.017	0.018	0.019
24	21	0.012	0.014	0.016	0.017	0.018
25	22	0.011	0.013	0.015	0.016	0.017
26	23	0.011	0.012	0.015	0.015	0.017
27	24	0.01	0.012	0.014	0.015	0.016

Figure 21-1: Chance of purchasing from Land’s End catalog

A Markov Chain Model of Customer Value (cont.)

- In a **Markov Chain** a process evolves from one state to another state, with the probability of going to the next state depending only on the current state.
- In this example the customer's current state is the number of periods in the past when last order occurred and the number of previous orders. Given this information you can determine how the value of a customer who has ordered once depends on recency.
- Assume you stop mailing to a customer after she fails to order 24 consecutive times. Also assume the annual discount rate (often called the weighted average cost of capital [WACC]) is 3 percent per period (or 1.034) per year.

A Markov Chain Model of Customer Value (cont.)

- The following steps describe how to determine the customer value for a customer who has bought one time (frequency = 1) and bought from the last received catalog (recency = 1).
- In the DS498_week7_Markov.xls file:

1. In cell C30 enter the recency level (in this case 1) with the formula

    =Initial_recency

2. In cell D30 enter the initial frequency with the formula

    =Original_Frequency

3. In cell E30 determine the probability that the customer orders in with the formula

    =IF(B30="no",0,INDEX(probs,C30,D30))

4. If you have ended the relationship with the customer, the order probability is 0.

A Markov Chain Model of Customer Value (cont.)

- The key to performing a **Monte Carlo simulation** is the **RAND() function**.
- When you enter the RAND() function in a cell, Excel enters a number that is equally likely to assume any value greater than 0 and less than 1.
- Values in different cells containing a RAND() function are independent; that is, the value of a RAND() in one cell does not have influence on the value of a RAND() in any other cell.
- In a Monte Carlo simulation, you use RAND() functions to model sources of uncertainty. Then you recalculate the spreadsheet many times (say 10,000) to determine the range of outcomes that can occur.
- The RAND() plays the role of "electronic dice."

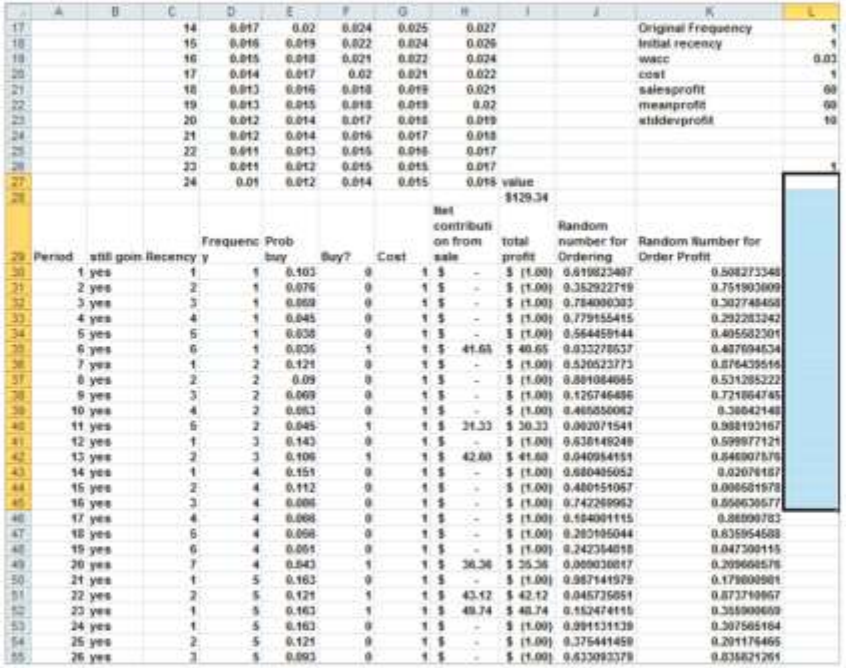


Figure 21-2: Land's End Customer Value model

A Markov Chain Model of Customer Value (cont.)

- During each three-month period, the Monte Carlo simulation uses two RAND() functions. One RAND() determines if the customer places an order. If the customer orders, the second RAND() function determines the profit (excluding mailing cost) generated by the order. The following steps continue the example, now walking you through using the RAND() function to perform the Monte Carlo simulation:

1. In cell F30 you can determine if the customer orders during period 1 by using the formula

   =IF(B30="no",0,IF(J30<E30,1,0))

- If you mailed a catalog and the random number in Column J is less than or equal to the chance of the customer placing an order, then an order is placed. Because the RAND() value is equally likely to be any number between 0 and 1, this gives a probability of E30 that an order is placed.

A Markov Chain Model of Customer Value (cont.)

2. Book the cost of mailing the catalog (if the customer is still with you) by using the formula

   =IF(B30="yes",$L$20,0) in cell G30

3. Book the profit from an order by using the formula in cell H30.
   =IF(AND(B30="yes",F30=1),NORMINV(J30,meanprofit,stddevprofit),0)

4. If an order is received, then the profit of the order is generated with the **NORMINV(J30,meanprofit,stddevprofit,0)** portion of the formula. If the random number in Column J equals x, then this formula returns the xth percentile of a normal random variable with the given mean and standard deviation. For example, if J30 contains a 0.5, you generate a profit equal to the mean, and if J30 contains a 0.841, you generate a profit equal to one standard deviation above the mean.

5. Calculate the total profit for the period in cell I30 with the formula =H30-G30.

A Markov Chain Model of Customer Value (cont.)

6. In cells C31 and D31 update recency and frequency based on what happened with the last catalog mailing. This is the key step in your model because it deter-mines how the customer’s state changes. In C31 the formula **=IF(F30=1,1,C30+1)** increases recency by 1 if a customer did not buy last period. Otherwise, recency returns to 1. In cell D31 the formula **=IF(D30=5,5,D30+F30)** increases frequency by 1 if and only if the customer placed an order last period. If the customer has ordered five times frequency remains at > = 5.
7. In cell B31 end the relationship with the customer if she has not ordered for 24 months with the formula **=IF(C30>=24,"no","yes").**When there is a “No” in Column B, all future cash flows are 0.
8. Copy the formulas from E30:K30 to E31:K109 and copy the formulas from B31:D31 to B32:D109 to arbitrarily cut off profits after 80 quarters (20 years).
9. In cell I28 compute the present value of all profits (assuming end-of-period profits) with the formula **=NPV(wacc,I30:I109)**

A Markov Chain Model of Customer Value (cont.)

You can "perform" the Monte Carlo simulation by:

1. Enter the possible recency values (1–24) in the range Q5:AN5.
2. Enter the integers 1 through 10,000 (corresponding to the 10,000 "iterations" of recalculating the spreadsheet) in the range P6:P10005.

   To accomplish this enter a 1 in P6, and from the Home tab, select Fill and then Series. Then complete the dialog box, as shown in Figure 21-3

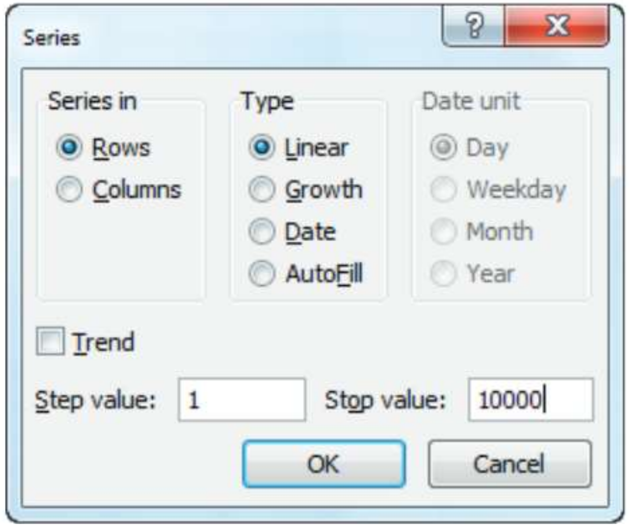


Figure 21-3: Filling in iteration numbers 1–10,000

3. Enter the output formula =I28 in the upper-left corner (cell P5) of the table range.
4. From the Data tab select What-If Analysis and then choose Data Table.
5. Fill in the Row input cell as L18 (Initial Recency Level).

A Markov Chain Model of Customer Value (cont.)

6. For the Column input cell choose any blank cell (such as AD2). In each column Excel sequentially places 1, 2, ...10,000 in the blank cell and recalculates the RAND() values for the column's recency level. After a few minutes you have "played out" 10,000 customers for each recency level. (Recall you have fixed Frequency = 1.)

7. Copy the formula **=AVERAGE(Q6:Q10005)** from Q4 to R4:AN4 to compute an estimate of the average profit for each recency level. The results are shown in Figure 21-4.

	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1												
2	frequency =1											
3												
4	mean profit	$14.48	$8.76	$4.57	$2.42	$0.79	$0.21	-$0.29	-$0.85	-$1.79	-$2.80	-$2.44
5	($16.94)	1	2	3	4	5	6	7	8	9	10	11
6	1	$99.31	$4.64	$20.04	-$15.42	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30
7	2	$94.40	-$16.44	$23.48	-$15.42	$1.25	-$14.32	-$13.75	$100.87	-$12.56	-$11.94	-$11.30
8	3	-$16.94	$44.92	-$15.94	-$15.42	$186.04	$21.98	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30
9	4	$7.93	-$16.44	-$15.94	-$15.42	$17.19	-$14.32	$18.94	-$13.17	-$12.56	-$11.94	-$11.30
10	5	-$16.94	-$16.44	-$15.94	-$15.42	-$14.88	$20.43	-$13.75	-$13.17	$79.70	-$11.94	-$11.30
11	6	$18.90	-$16.44	$29.76	$6.27	-$14.88	-$14.32	$36.04	-$13.17	-$12.56	-$11.94	$10.66
12	7	$1.29	-$16.44	-$15.94	-$15.42	-$14.88	$18.18	$4.27	-$13.17	-$12.56	$6.38	-$11.30
13	8	-$16.94	-$16.44	-$15.94	-$15.42	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	$96.57
14	9	$122.50	-$16.44	-$15.94	-$0.83	-$14.88	-$14.32	-$13.75	-$13.17	$17.88	-$11.94	$24.29
15	10	$125.81	-$16.44	-$15.94	-$15.42	-$14.88	-$14.32	-$13.75	$52.19	-$12.56	-$11.94	-$11.30
16	11	$117.76	$52.02	-$15.94	$46.97	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30
17	12	$54.03	$8.29	-$15.94	-$15.42	$66.78	$7.03	$33.38	-$13.17	-$12.56	$75.49	-$11.30
18	13	$68.79	-$16.44	-$15.94	$12.29	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	$16.28	-$11.30
19	14	$9.72	-$16.44	-$15.94	-$15.42	$2.91	$10.70	-$13.75	-$1.52	-$12.56	-$11.94	-$11.30
20	15	$40.58	$30.75	-$15.94	$115.56	-$14.88	$9.68	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30
21	16	$51.17	-$16.44	-$15.94	-$15.42	-$14.88	$123.73	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30
22	17	$4.02	$19.84	$51.08	$36.85	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30
23	18	-$16.94	$43.52	$44.90	-$15.42	-$14.88	-$14.32	-$13.75	$21.54	$8.26	-$11.94	-$11.30
24	19	-$16.94	-$16.44	-$15.94	-$15.42	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30
25	20	$18.15	-$16.44	$52.36	-$15.42	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	$29.71
26	21	-$16.94	-$3.41	$76.47	-$15.42	$28.66	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$0.64
27	22	-$16.94	-$16.44	-$15.94	-$15.42	$80.05	-$14.32	-$13.75	-$0.01	-$12.56	-$11.94	-$11.30
28	23	-$16.94	-$16.44	$25.66	$34.71	-$14.88	-$14.32	-$13.75	$14.06	-$12.56	-$11.94	-$11.30

Figure 21-4: Land's End data table

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative

- If a company repeatedly chooses marketing decisions that have much more than a 50 percent probability of making the company better off, then in the long run the company will succeed as result of marketing decisions.
- You can use Monte Carlo simulation to evaluate the probability that a marketing decision will improve a company's bottom line.
- In this section you will learn how to do this by analyzing whether a pizza parlor will benefit from a **Groupon offer**.
- Carrie has just been fired from the CIA and has purchased a suburban Virginia pizza parlor. She is trying to determine whether she should offer the local residents a Groupon offer. The restaurant hopes to recoup the loss on the pizzas sold with a Groupon offer via the customer value of new customers who become return customers.

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

To be specific, the terms of the Groupon offer follow:

- Customers are offered two pizzas (usually sold for $26) for $10.
- Carrie keeps half of the revenue ($5).
- Carrie's profit margin is 50 percent.
- In deciding whether to use Groupon, Carrie faces many sources of uncertainty:
- Fraction of customers who take the offer who are new customers
- Fraction of people who spend more than the deal size ($26)
- For customers who spend more than $26, the amount spent in excess of $26
- Fraction of new customers who return
- Annual profit generated by new customers who return
- Retention rate for new customers

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

- If a company repeatedly chooses marketing decisions that have much more than a 50 percent probability of making the company better off, then in the long run the company will succeed as result of marketing decisions.
- You can use Monte Carlo simulation to evaluate the probability that a marketing decision will improve a company’s bottom line. In this section you will learn how to do this by analyzing whether a pizza parlor will benefit from a Groupon offer.
- Carrie has just been fired from the CIA and has purchased a suburban Virginia pizza parlor. She is trying to determine whether she should offer the local residents a Groupon offer. The restaurant hopes to recoup the loss on the pizzas sold with a Groupon offer via the customer value of new customers who become return customers.

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

- To help Carrie determine the range of outcomes that will result from the Groupon offer, you can use the Monte Carlo simulation to model these sources of uncertainty and in doing so, estimate the chances that the Groupon offer will increase profitability.
- The key to the analysis is using the customer value concepts and the Monte Carlo simulation to estimate the probability that the benefit the pizza parlor gains from new customers will outweigh the lost profit on pizzas sold to customers who redeem the Groupon offer.
- To simplify modeling, assume each of the six uncertain quantities is equally likely to assume any value between a low value and a high value. The low and high values for these quantities are shown in Figure 21-5; see file DS498_week7_ Groupon.xlsx.

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

- When trying to determine these values, the marketing analyst can use historical information to establish upper and lower bounds. We surveyed 324 businesses that used Groupon and obtained estimates for some of the previously listed quantities:
    - 75 percent of people taking the deal are new customers.
    - 36 percent of all deal takers spent more than the deal size.
    - 20 percent of new customers returned later

	B	C	D	E	F	G
1		Groupon				
2						
3		margin	0.5			
4		2pizzas	$26.00			
5		weget	$5.00			
6		cost	$13.00	Low	High	
7		probnewcustomer	0.76	0.65	0.85	
8		newspendmorethandeal	0.4	0.3	0.42	
9		newpeoplewhoreturn	0.17	0.1	0.3	
10						
11		Look at 100 people who take deal				
12	new	76				
13	returnees	24				
14	howmanyspendmore	40	Low	High		
15	howmuchmore	$7.00	$3.00	$17.00		
16	valuenewcustomer	$59.07	Low	High		
17	anncustomerprofit	$21.00	$20.00	$40.00		
18	retentionrate	0.77	0.55	0.85		
19	newcomeback	12.92				
20						
21						
22						
23						
24	Loss					
25	fromnew	$608.00				
26	fromreturning	$504.00	assume they would have come anyway			
27	total loss	$1,112.00				
28	Benefits					
29	extraprofittoday	$140.00				
30	valuenewcustomers	$763.15				
31	total	$903.15				
32						
33	Net Gain	-$208.85				

Figure 21-5: Analysis of Groupon for Carrie's Pizza

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

Based on this information and on past Groupon offers, Carrie believes the following:

- Between 65-85 percent of customers taking the deal will be new customers.
- Between 30-42 percent of customers will spend an amount in excess of $26.
- Those who spend in excess of $26 will spend average between $3-$17 beyond $26.
- Between 10 percent and 30 percent of new customers will return.
- Average annual profit generated by a new customer is between $20 and $40.
- Average annual retention rate for new customers generated by Groupon is between 55-85 percent.

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

- You can use the Excel RANDBETWEEN function to ensure that an uncertain quantity (often called a random variable) is equally likely to lie between a lower limit L and an upper limit U.
- For integers, entering the formula RANDBETWEEN(L,U) in a cell makes it equally likely that any integer between L and U inclusive will be entered in the cell. For example, entering the formula **=RANDBETWEEN(65,85)** in a cell ensures that it is equally likely that any of 65, 66, ..., 84, or 85 is entered in the cell.

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

- Complete the following steps (see the file DS498_week7_Groupon.xlsx) to model the range of outcomes that will ensue if Carrie introduces a Groupon offer:

1. In D3:D5 enter Carrie's profit margin, the price of two pizzas without Groupon, and what Carrie receives from Groupon for two pizzas.
2. In D6 compute the cost of producing two pizzas with the formula

   =(1-margin)*_2pizzas.
3. In cell D7 enter the fraction of Groupon offer takers who are new customers with the formula **=RANDBETWEEN(100*E7,100*F7)/100**. This is equally likely to enter a .65, .66, ..., .84, .85 in cell D7.
4. In cell D8 generate the fraction of new customers who spend more than $26 with the formula **=RANDBETWEEN(100*E8,100*F8)/100.**

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

5. Copy this formula from cell D8 to D9 to generate the fraction of new Groupon customers who return.
6. Assume without loss of generality that 100 customers take the Groupon offer and you generate the random net gain (or loss) in profit from these 100 offer takers. Include gains or losses today and gains from added new customers.
7. In cell C12 compute the number of your 100 offer takers who will be new customers with the formula =**100*probnewcustomer**.
8. In cell C13 compute the number of offer takers who are returnees with the formula =**100*(1-probnewcustomer)**.
9. In cell C14 compute the number of offer takers who spend more than the deal with the formula =**100*newspendmorethandeal**.

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

10. In cell C15 determine the average amount spent in excess of $26 by those spending more than $26 with the formula **=RANDBETWEEN(D15,E15)**. Copy this formula to C17 to determine the average level of annual customer profit for new customers created by the Groupon offer.

11. In the worksheet basic model, attach your Customer Value template. Then in C16 use the formula **='basic model'!E5*C17** to compute the lifetime value of a customer based on mid-year cash flows.

12. In cell C18 compute the average retention rate for new customers with the formula **=RANDBETWEEN(100*D18,100*E18)/100**.

13. In cell C19 compute the number of the 100 offer takers who are returning new customers with the formula **=C12*newpeoplewhoreturn**

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

In the range C25:C33 you can compute your gain or loss from the 100 offer takers:

1. To begin in cell C25, use the formula **=(cost-weget)*C12** to compute your loss today from the new customers among the offer takers as $8 * number of new customers.
2. To simplify your work assume all offer takers who were previous customers would have shown up anyway. Because each of these returning customers would have paid $26, you lose $26 – $5 = $21 on each of these customers.
3. Then in cell C26 use the formula **=C13*(_2pizzas-weget)** to compute your loss on these customers.
4. In cell C27 the formula **=SUM(C25:C26)** computes your total loss today on the 100 people who took the Groupon offer.

Using Monte Carlo Simulation to Predict Success of a Marketing Initiative (cont.)

5. In C29 with the formula **=margin*C15*C14**, compute the extra profit earned today by multiplying your 50 percent profit margin by the amount in excess of $26 spent today by offer takers.
6. In C30 use the formula **=C19*C16** to compute the value of the new customers by multiplying the number of returning new customers times the average value for each new customer.
7. In cell C31 use the formula **=C29+C30** to compute the total benefits created by the Groupon offer.
8. In cell C33 use the formula **=C31-C27** to compute the total benefits less today's losses.

Using a One-Way Data Table to Simulate the Groupon Deal

1. Use FILL SERIES (from the Home tab) to enter the iteration numbers (1, 2, ..., 10,000) in the range I9:I10008.
2. Use a one-way data table to “trick” Excel into replaying your spreadsheet 10,000 times. Recalculate your total gain on the 100 offer takers, so enter the total gain in cell J8 with the formula =C33.
3. Select the data table range (I8:J10008), and from the Data tab, choose What-If Analysis and select Data Table. In a one-way data table, there is no row input cell, so all you need to do is choose any blank cell (such as N7) as the column input cell. Then Excel places 1, 2, ..., 10,000 in N7 and each time recalculates Carrie’s net gain. During each recalculation each RANDBETWEEN function recalculates, so you play out the modeled uncertainty 10,000 times.
4. The resulting simulated profits are shown in the range J9:J10008 of Figure 21-6.

Using a One-Way Data Table to Simulate the Groupon Deal (cont.)

	I	J	K	L	M	N
1						
2						
3		average gain	$144.11			
4		probability profitable	0.5384			
5						
6						
7						
8		-$208.85				
9	1	-253.5384849			-$1,000.00	0
10	2	27.29517303			-$500.00	802
11	3	-143.2669584			$0.00	3814
12	4	35.30855332			$500.00	3156
13	5	-251.2044892			$1,000.00	1413
14	6	171.9389368			$1,500.00	570
15	7	647.8345245			$2,000.00	184
16	8	-334.9786087			$2,500.00	52
17	9	1035.426236			$3,000.00	8
18	10	157.4370552			>3000	1
19	11	240.1467352				
20	12	129.2907669				
21	13	-388.6731532				
22	14	61.43800788				
23	15	292.5474792				
24	16	588.1872725				
25	17	-0.674150668				
26	18	405.2020103				
27	19	16.15306931				
28	20	43.19811399				
29	21	271.3136001				
30	22	-591.9734436				

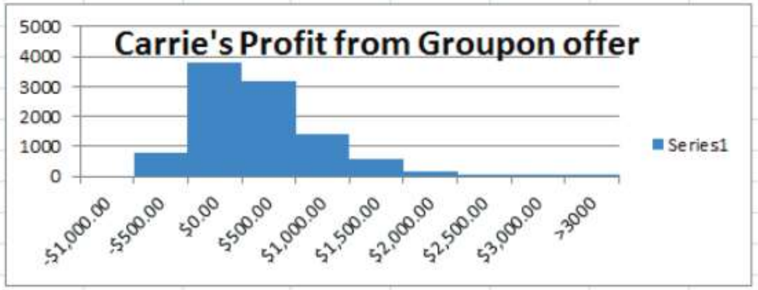


Figure 21-6: Simulation results for Carrie's Pizza

Using a One-Way Data Table to Simulate the Groupon Deal (cont.)

5. In cell K3 compute the average profit over your 10,000 iterations earned from the 100 deal takers with the formula **=AVERAGE(J9:J10008).** You find an average gain of $144.11, which indicates that on average the Groupon deal can improve Carrie's bottom line.
6. In cell K4 compute the probability that the deal increases profits with the formula **=COUNTIF(J9:J10008,">0")/10000**. There is a 53.8 percent chance the deal yields a favorable result.
7. This finding indicates that the deal is of marginal value to the pizza parlor. From cell K3, you find that the average profit per customer equals $144.11. This again indicates that on average, the Groupon deal does just a little better than breaking even

Using a Histogram to Summarize the Simulation Results

To create a bar graph of your simulation results, proceed as follows:

1. Enter the boundaries of bin ranges (–\$1000, –\$500, ..., \$3000) in M9:M17. Append a label >3000 for all iterations in which profit is more than \$3000.

2. Select the range N9:N18 and array enter by selecting Ctrl+Shift+Enter, the formula **=FREQUENCY(J9:J10008,M9:M17)**.

3. In N9 this computes the number of iterations in which profit is <=\$1000; in N10 this array formula computes the number of iterations in which profit is >–\$1000 and <=–\$500 (802); in N18 this array formula computes the number of iterations (1) in which profit is >\$3000.

4. A column graph summarizing these results is shown in Figure 21-6.

Summary

- You can use the Excel data table feature combined with Excel's **RAND()** and **RANDBETWEEN** functions to simulate uncertainty in situations in which customer value involves uncertain quantities (random variables).
- If an uncertain event has a probability x of occurring, then the event occurs if a value of RAND() is less than or equal to x.
- If an uncertain quantity (such as annual retention rate) is normally distributed with a given mean and sigma, then the function generates a normal random variable with the given mean and sigma.

  =NORMINV(RAND(),mean,standard dev)

- If an uncertain quantity (such as annual profit generated by a customer) is equally likely to be between two integers L and U, then it can be modeled with the function RANDBETWEEN(L,U).

Practical Portion

Chapter 25: Using Classification Trees for Segmentation

Marketing Analytics: Data-Driven Techniques with Microsoft Excel

Chapter 25: Using Classification Trees for Segmentation

- Introducing Decision Trees
- Constructing a Decision Tree
- Pruning Trees and CART
- Summary

Introducing Decision Trees

- Decision trees are used to predict a categorical dependent variable such as:
  - Will a family purchase a riding mower during the next year?
  - Will a person suffer a heart attack in the next year?
  - Will a voter vote Republican or Democratic in the next presidential election?

- You begin the tree with a root node that includes all combinations of attribute values, and then use an independent variable to “split” the root node to create the most improvement in class separation.

- You can see this concept more clearly using a simple example to illustrate the construction of a decision tree.

Constructing a Decision Tree

- Suppose that you want to come up with a simple rule to determine whether a person will buy **Greek yogurt**. Figure 25-1 contains data on a sample of 10 adults (see the DS498_week7_Greekyogurt.xlsx).
- For example, Person 1 is a single, high-income woman who did not buy Greek yogurt.
- In this example the dependent variable for each person is whether the person purchased Greek yogurt.

	B	C	D	E	F
1					
2	Person	Gender	Marital Status	Income Level	Bought?
3	1	Female	Single	High	No
4	2	Male	Married	Average	No
5	3	Male	Single	Low	No
6	4	Female	Married	High	No
7	5	Male	Divorced	Average	Yes
8	6	Male	Married	Low	No
9	7	Female	Divorced	High	No
10	8	Male	Single	Average	Yes
11	9	Male	Married	Low	No
12	10	Male	Single	Average	Yes

Figure 25-1: Data on Greek yogurt purchasers

Constructing a Decision Tree

- A node of a decision tree is considered pure if all data points associated with the node have the same value of the dependent variable. You should branch only on impure nodes.
- In this example, because the root node contains three purchasers of yogurt and seven nonpurchasers, branch on the root node. The goal in branching is to create a pair of child nodes with the least impurity.
- There are several metrics used to measure the impurity of a node. Then the impurity of a split is computed as a weighted average of the impurities for the nodes involved in the split, with the weight for a child node being proportional to the number of observations in the child node.
- In this section you use the concept of entropy to measure the impurity of a node.

Constructing a Decision Tree

- To define the entropy of a node, suppose there are c possible values (0, 1, 2, c-1) for the dependent variable.
- Assume the child node is defined by independent variable X being equal to a.
- Then the entropy of the child node is computed as the following equation:

$$(1)\ \text{Entropy} = \sum_{i=0}^{i=c-1} P(i|X = a) Log_2 (P(i|X = a).$$

- In Equation 1 the following is true:
  - P(i|X = a) is the fraction of observations in class i given that X = a.
  - Log2 (0) is defined to equal 0.

Constructing a Decision Tree (cont.)

- Entropy always yields a number between 0 and 1 and is a concept that has its roots in Information Theory.
- With two classes, a pure node has an entropy of 0 (−0*Log2 0 + 1*Log2 1) = 0.
- A split of a node can yield a maximum entropy value of 1 when one-half the observations associated with a node have c = 0 and c = 1.
- This shows that intuitively picking a split based on entropy can yield pure nodes. It can also be shown that with two nodes the entropy decreases as the fraction of nodes having c = 1 or c = 0 moves away from 0.5.
- This means that choosing splits with lower entropy will have the desired effect of decreasing impurity.

Constructing a Decision Tree (cont.)

- Suppose there are S possible values (s = 1, 2, ..., S) for the attribute generating the split and there are in observations having the independent variable = i. Also assume the parent node has N total observations. Then the impurity associated with the split is defined by the following equation:

$$(2)\ \text{Impurity} = \Sigma_{i=1}^{i=S} \frac{Entropy(i)n_i}{N}$$

- For this example suppose that you split the root node based on gender. From Equation 1 you can find that:
  - Entropy (Female) = −[(3 / 3) * (Log2 (3 / 3) + (0/3) * Log2 (0 / 3)] = 0
  - Entropy (Male) = −[(4 / 7) * Log2 (4 / 7) + (3 / 7) * Log2 (3 / 7)] = 0.985

- Because the data set contains three women and seven men Equation 2 shows you that the impurity of the split is (3 / 10) * 0 + (7 / 10) * (0.985) = 0.69.

Constructing a Decision Tree (cont.)

- You should split on the independent variable whose child nodes yield the lowest level of impurity.
- Using the Excel COUNTIFS function, it is easy to compute the level of impurity resulting from a split on each independent variable (see Figure 25-2).
- Then you split the root node using the independent variable that results in the lowest impurity level.

	B	C	D	E	F	G	H	I	J	K
1										
2	Person	Gender	Marital Status	Income Level	Bought?					
3	1	Female	Single	High	No					
4	2	Male	Married	Average	No					
5	3	Male	Single	Low	No					
6	4	Female	Married	High	No					
7	5	Male	Divorced	Average	Yes					
8	6	Male	Married	Low	No					
9	7	Female	Divorced	High	No					
10	8	Male	Single	Average	Yes					
11	9	Male	Married	Low	No					
12	10	Male	Single	Average	Yes					
13			Buy							
14			Yes	No	Total	Fraction			Entropy	Impurity
15	Gender	Female	0	3	3	0.3	0	0	0	0.68966
16		Male	3	4	7	0.7	-0.524	-0.461	-0.98523	
17			Buy							
18			Yes	No						
19	Income Level	High	0	3	3	0.3	0	0	0	0.324511
20		Average	3	1	4	0.4	-0.311	-0.5	-0.81128	
21		Low	0	3	3	0.3	0	0	0	
22			Buy							
23			Yes	No						
24	Marital Status	Single	2	2	4	0.4	-0.5	-0.5	-1	0.6
25		Married	0	4	4	0.4	0	0	0	
26		Divorced	1	1	2	0.2	-0.5	-0.5	-1	

Figure 25-2: Impurity calculations from each split

Constructing a Decision Tree (cont.)

1. Copy the formula =**COUNTIFS($C$3:$C$12,$C15,$F$3:$F$12,D$14)** from D15 to D15:E16 to compute the number of females and males who buy and do not buy Greek yogurt. For example, you find four males do not buy Greek yogurt.

2. Copy the formula **=COUNTIFS($E$3:$E$12,$C19,$F$3:$F$12,D$14**) from D19 to D19:E21 to count the number of people for each income level that buy and do not buy Greek yogurt. For example, three average income people bought Greek yogurt.

3. Copy the formula **=COUNTIFS($D$3:$D$12,$C24,$F$3:$F$12,D$14)** from D24 to the range D24:E26 to count how many people for each marital status buy or do not buy Greek yogurt. For example, two single people buy Greek yogurt.

4. Copy the formula **=SUM(D15:E15)** from F15 to the range F15:F26 to compute the number of people for the given attribute value. For example, cell F25 tells you there are four married people in the population.

Constructing a Decision Tree (cont.)

5. Copy the **formula =F15/SUM($F$15:$F$16)** from G15 to G15:G26, to compute the fraction of observations having each possible attribute value. For example, from G16 you can find that 70 percent of the observations involve males.

6. Copy the formula **=IFERROR((D15/$F15)*LOG(D15/$F15,2),0)**
   - from H15 to H15:I26 to compute for each attribute value category level combination the term P(i|X=a)*Log2(P(i|X=a).
   - You need IFERROR to ensure that when P(i|X=a)=0 Log2(0) the undefined value is replaced by 0.
   - In general, entering **IFERROR(formula, anything)** will enter the value computed by the formula as long as the formula does not return an error. If the formula does return an error, IFERROR returns whatever is entered after the comma (in this case 0.)

Constructing a Decision Tree (cont.)

7. Copy the formula =**SUM(H15:I15**) from J15 to J15:J26 to compute via Equation 1 the entropy for each possible node split.

8. Copy the formula =-**SUMPRODUCT(G15:G16,J15:J16**) from K15 to K16:K24 to compute via Equation 2 the impurity for each split.

9. The impurity for income of 0.325 is smaller than the impurities for gender (0.69) and marital status (0.60), so begin the tree by splitting the parent node on income.

10. This yields the three nodes shown in Figure 25-3.

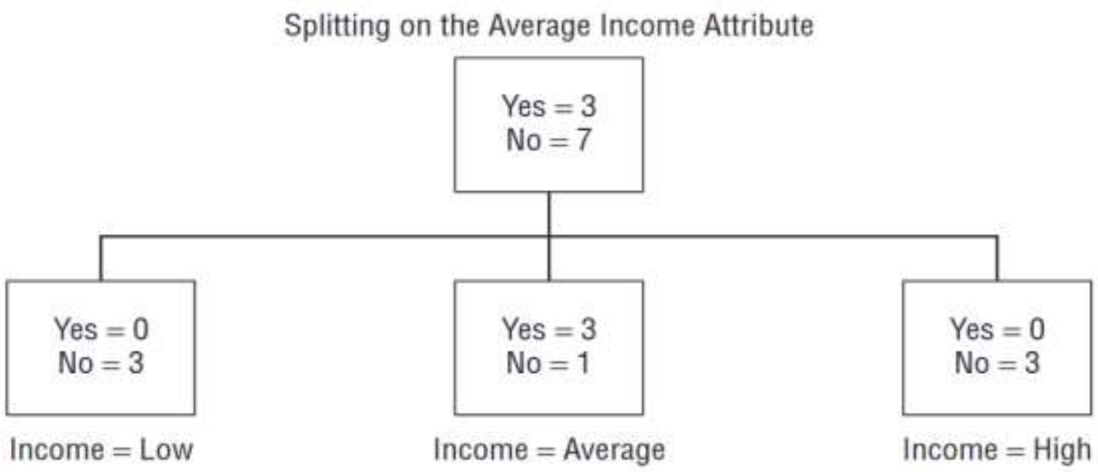


Figure 25-3: Splitting root node with Income variable

Note that this figure was drawn and was not created by Excel.

Constructing a Decision Tree (cont.)

- The Income = Low and Income = High nodes are pure, so no further splitting is necessary.
- The Income = Average is not pure, so you need to consider splitting this node on either gender or marital status.

- Splitting on gender yields an impurity of 0.811, whereas splitting on marital status yields an impurity of 0.
- Therefore, split the Income = Average node on marital status. Because all terminal nodes are pure (that is, each respondent for a terminal node is in the same class), no further splitting is needed, and you obtain the decision tree, as shown in Figure 25.4.

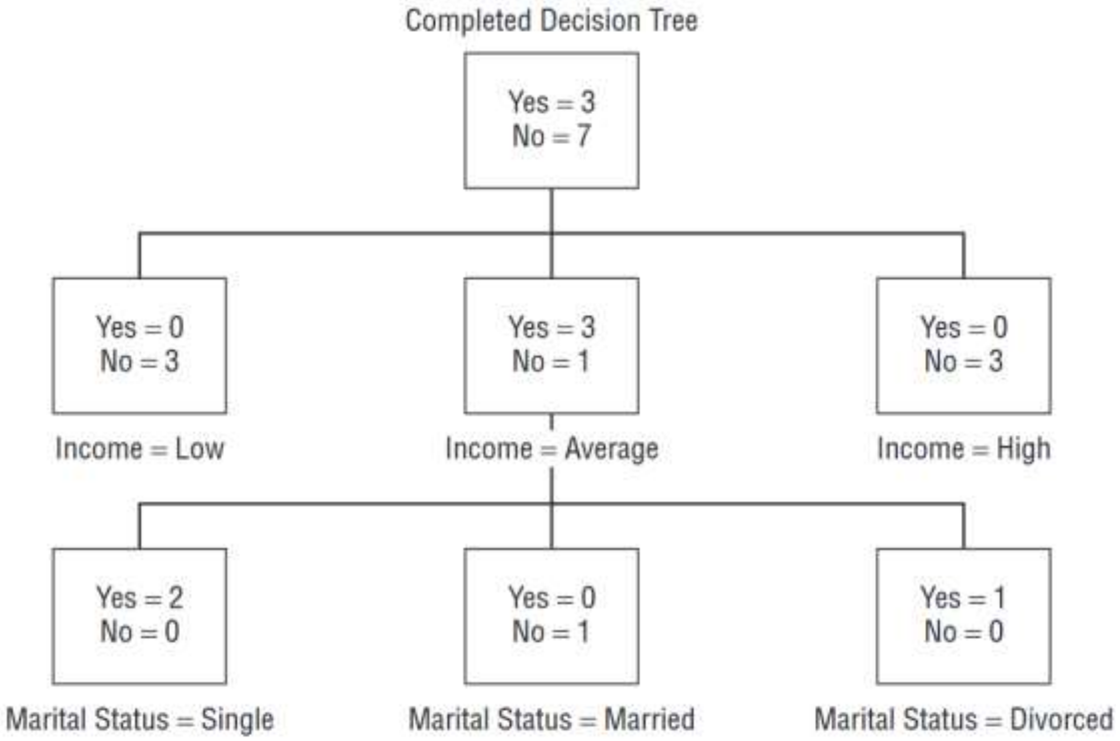


Figure 25-4: Tree after marital status and income splits

Pruning Trees and CART

- Extensive calculation is required to create decision trees.
- Fortunately, widely used statistical packages such as SAS, STATISTICA, R, SPSS, and XLMINER can quickly churn out decision trees.
- A key issue in creating a decision tree is the size of the tree. By adding enough nodes you can always create a tree for which the terminal nodes are all pure.
- Unfortunately, this usually results in overfitting, which means the tree would do poorly in classifying out of sample observations.
- **CART** “prunes” a tree by trading off a cost for each node against the benefit generated by the tree. The benefit derived from a tree is usually measured by the misclassification rate.

Pruning Trees and CART (cont.)

- To compute the misclassification rate for a tree, assume at each terminal node all observations are assigned to the class that occurs most frequently.

- All other observations associated with a node are assumed to be misclassified.

- Then the misclassification rate is the fraction of all misclassified observations.

- To illustrate the calculation of the misclassification rate, compute the misclassification rate for the tree, as shown in Figure 25-5.

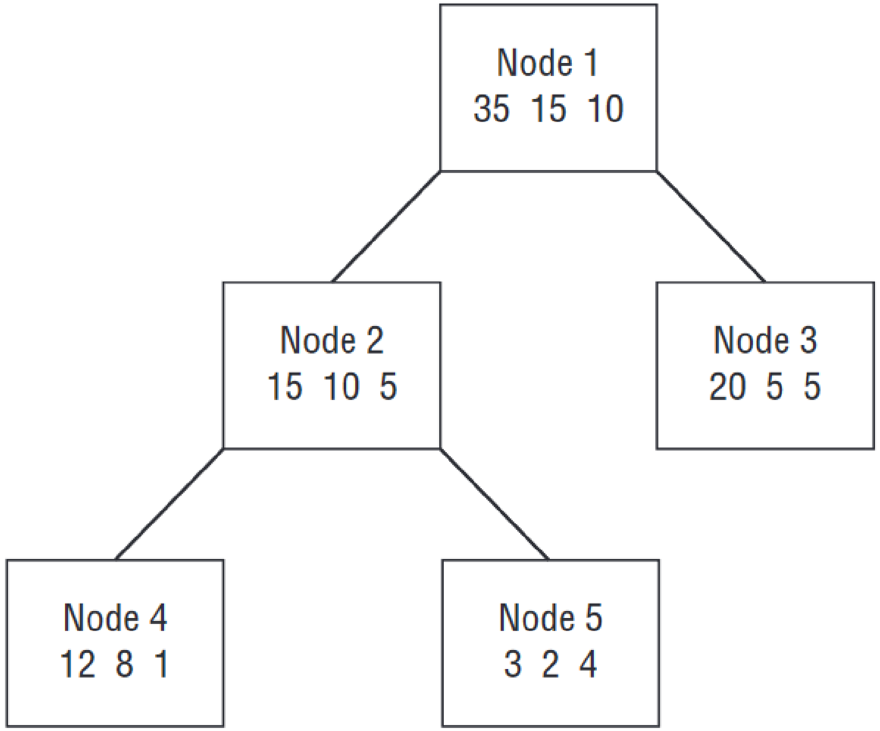


Figure 25-5: Tree for illustrating misclassification rate

Pruning Trees and CART (cont.)

For each terminal node the number of misclassified observations is computed as follows:

- For Node 3, 20 observations are classified in Class 1, 5 in Class 2, and 5 in Class 3. For misclassification purposes you should assume that all Node 3 observations should be classified as Class 1. Therefore 10 observations (those classified as Class 2 or 3) are assumed to be misclassified.
- Node 4 observations are classified as Class 1, so the 9 observations classified as Class 2 or 3 are misclassified.
- Node 5 observations are classified as Class 3, so the 5 observations classified as Class 1 or 2 are misclassified.
- In total 24/60 = 40 percent of all observations are misclassified.

Summary

In this chapter you learned the following:

- You can use decision trees to determine simple classification rules that can be understood by people with little statistical training.
- To create a decision tree, split a parent node using a division of attribute values that creates the greatest reduction in overall impurity value.
- The widely used CART algorithm uses pruning of a decision tree to avoid overfitting and creates effective parsimonious classification rules.

Main Reference

- **Chapter 8 (sections 8.5 & 8.6: Modeling in LP: An Example):** "Prescriptive Analytics: Optimization and Simulation" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".
- **Chapter 2:** Working with Data-Analysis Tools from "*Excel Data Analysis For Dummies*"
- **Chapters 21 & 25**: "Customer Value, Monte Carlo Simulation, and Marketing Decision Making" & " Using Classification Trees for Segmentation" from "*Marketing Analytics: Data-Driven Techniques with Microsoft Excel*"

Week self-review exercises

- **Section 8.6:** Apply the example of "swing states" for the 2016 election, created by Professor Rick Wilson of Oklahoma State University. from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*"

Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

College of Computing and Informatics

Week 11

Data analytics: Getting started with Orange

Contents

- Orange Workflow Overview.
- Visualizing data from input data file.
- Analyzing data with regression models and decision trees.
- Analyzing data with deep learning models.
- Text processing and classification.

Weekly Learning Outcomes

1. Understand Data workflow in orange.
2. Inspect data with orange visualisation.
3. Develop predictive models in orange, including regression, decision trees and deep learning models.
4. Perform Textual data analysis with orange.
5. Assess quality of various predication methods.

Required Reading

- Orange Visual Programming Documentation (Release 3). Orange (2021). https://buildmedia.readthedocs.org/media/pdf/orange-visual-programming/latest/orange-visual-programming.pdf
  - Chapter 1 – section 1.2, 1.3 (subsection: 1-2), 1.4, 1.5, Chapter 2 – section 2.1 (subsection: 1-10, 34) 2.2 (subsection: 1,3,4,5,16) 2.3 (subsection: 5,9,10,13) and section 2.4, (subsection: 2,4,6)
  - AJDA. (2017, August 4). *Text Analysis: New Features*. Orangedatamining.Com. https://orangedatamining.com/blog/2017/08/04/text-analysis-new-features/

Recommended Readings

- Zupan, D. (2018, May). *Introduction to Data Mining: Working notes for the hands-on course with Orange Data Mining*. University of Ljubljana. https://file.biolab.si/notes/2018-05-intro-to-datamining-notes.pdf
  - Lesson: 1,2,3,4,5,6,7,8,10,14,17,31, 32.
- Foong, N. W. (2019, August 7). *Data Science Made Easy: Interactive Data Visualization using Orange*. Medium, Towardsdatascience.com. https://towardsdatascience.com/data-science-made-easy-interactive-data-visualization-using-orange-de8d5f6b7f2b
- Analytics Vidhya, A. (2017, September 7). *Building Machine Learning Model is fun using Orange*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2017/09/building-machine-learning-model-fun-using-orange/
- Foong, N. W. (2019b, August 29). *Data Science Made Easy: Image Analytics using Orange*. Medium, Towardsdatascience.com. https://towardsdatascience.com/data-science-made-easy-image-analytics-using-orange-ad4af375ca7a

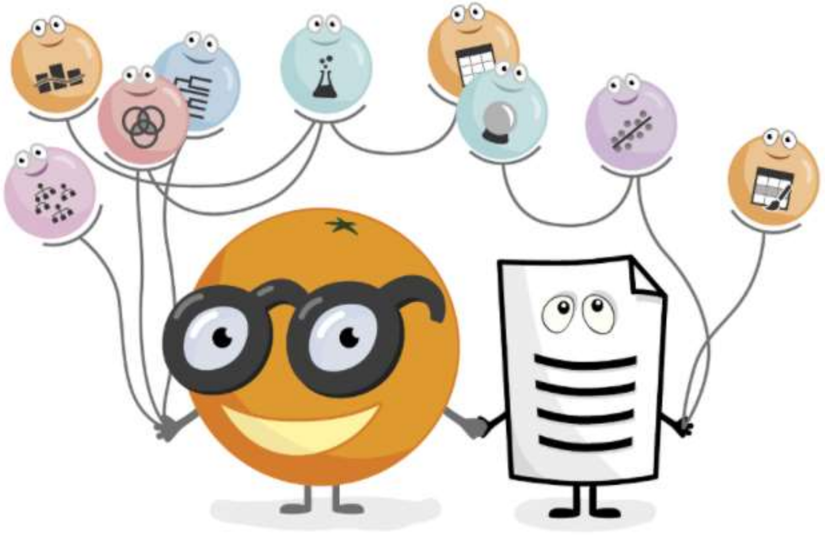
This Presentation is mainly dependent on the above recourses.

Recommended Videos

- Getting Started with Orange 01: Welcome to Orange (2015, December 21). [Video]. YouTube. https://www.youtube.com/channel/UClKKWBe2SCAEyv7ZNGhIe4g
- Orange Data Mining tool. (2016, May 5). [Video]. YouTube. https://www.youtube.com/watch?v=rrsRBSCHDXw
- Getting Started with Orange 16: Text Preprocessing (2017, Jun 20). [Video]. YouTube. youtube.com/watch?v=V70UwJZWkZ8
- Text Mining: Twitter Data Analysis (2020, August 4). [Video]. YouTube. https://www.youtube.com/watch?v=HDkI6G4slzQ
- Getting Started with Orange 18: Text Classification (2017, Jun 28). [Video]. YouTube. https://www.youtube.com/watch?v=zO_zwKZCULo

Orange workflow Overview

Orange workflow



“Orange is a component-based data mining software. It includes a range of data visualization, exploration, preprocessing and modeling techniques. It can be used through a nice and intuitive user interface or, for more advanced users, as a module for the Python programming language.” *(Orange official GitHub page).*

Orange workflow

- The core principle of Orange is visual programming.

- The basic processing unit of any data manipulation in Orange are called widgets.
  - Each analytical step/action is contained within a widget.
  - Widgets communicate by sending information along with a communication channel and the output from one widget is used as input to another.

- A workflow is the sequence of steps/actions that is performed to accomplish a particular task.
  - Widgets are are placed on the canvas and connected into an analytical workflow.
  - Orange analytical workflow is executed from left to right and never passes data backwards.

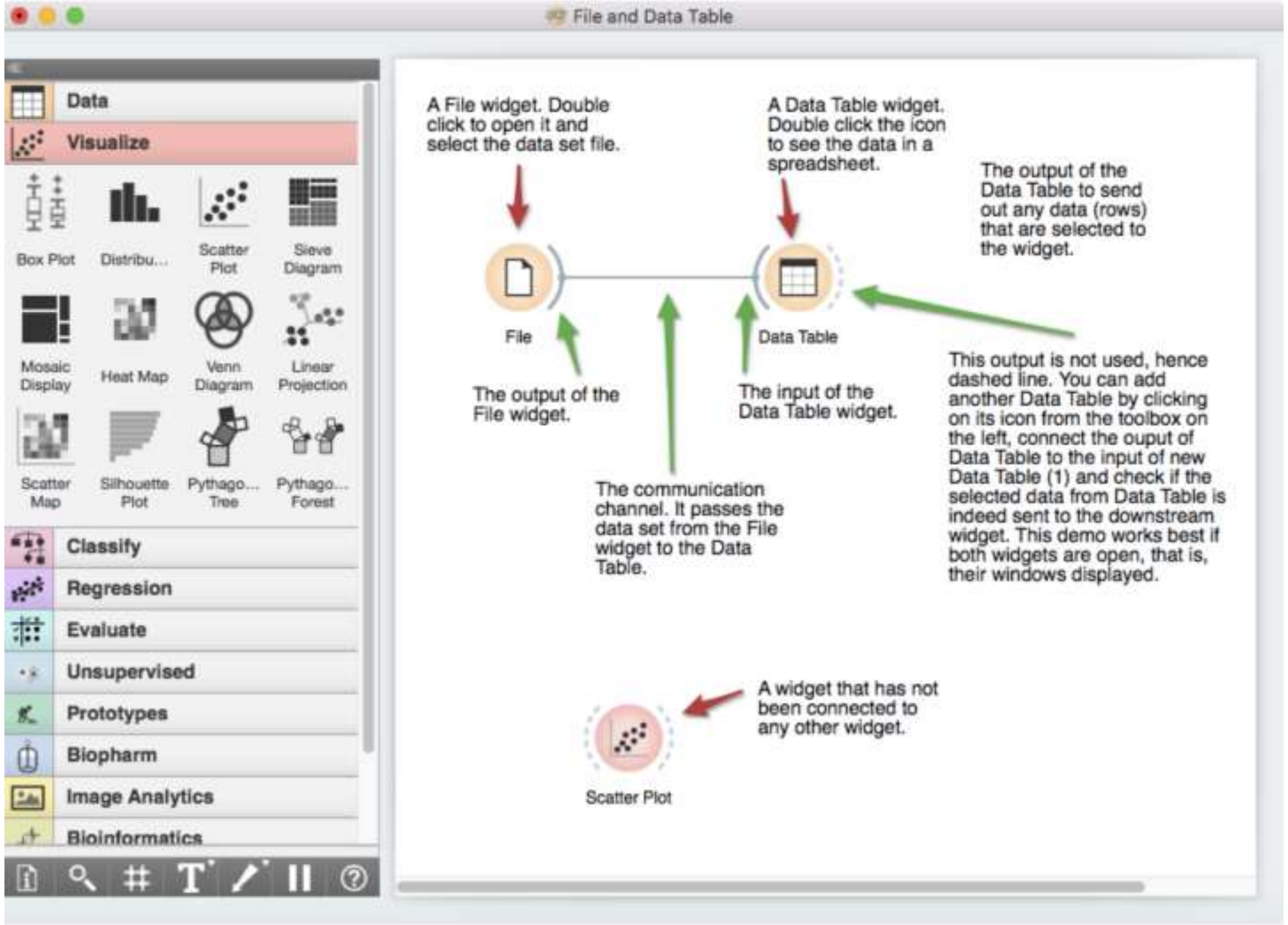
- Orange workflows consist of components that
  - Read
  - Process
  - Visualize data

Simple Orange workflow - Files and Data Tables

- File widget: reads the data.

- Data Table widget: a viewer and shows the data in a spreadsheet. It passes onwards only the selection.

- The data is always available in the File widget.



Workflow with two connected widgets

Simple Orange workflow - Files and Data Tables

- Most Orange workflows would probably start with the File widget.
- Orange can import any comma, .xlsx or tab-delimited data file or URL.

Example:

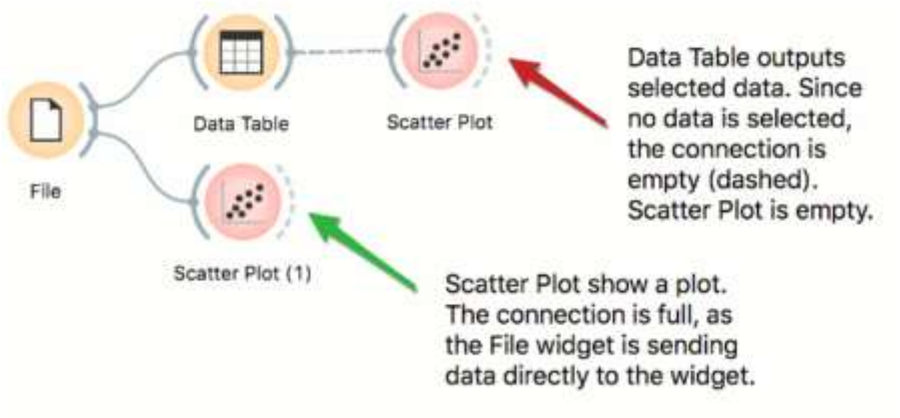
File widget is used to read the data that is sent to both the Data Table and the Box Plot widgets.

Workflows with subsets

- Visualizations in Orange are interactive, which means the user can select data instances from the plot and pass them downstream.

Example:

- **Selecting subsets**
  - **Step 1:** Place File widget on the canvas.
  - **Step 2**: Connect Scatter Plot to it.



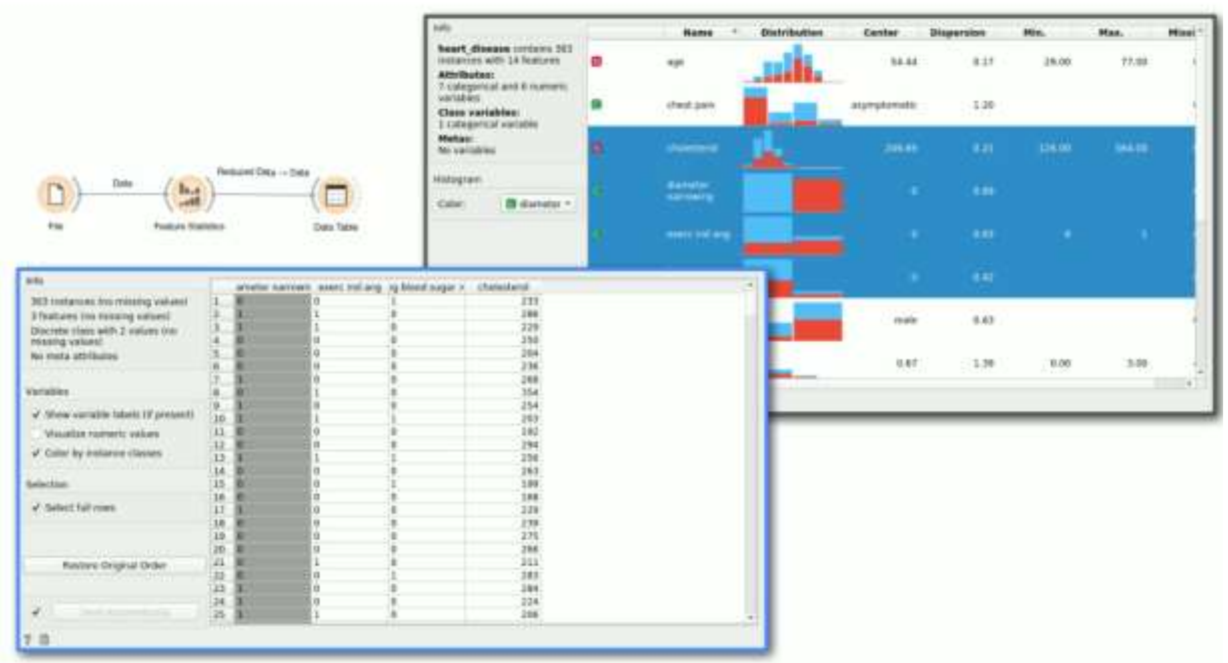
  - **Step 3:** Click and drag a rectangle around a subset of points.
  - **Step 4:** Connect Data Table to Scatter Plot → Data Table will show selected points.

- **Highlighting workflows**
  - Connect Data Table to Scatter Plot.
  - Select a subset of points from the Data Table → Scatter Plot will highlight selected points.

Workflows - data exploration

- Feature Statistics widget provides a quick way to inspect and find interesting features in a given data set.

- Example: Heart-disease data exploration

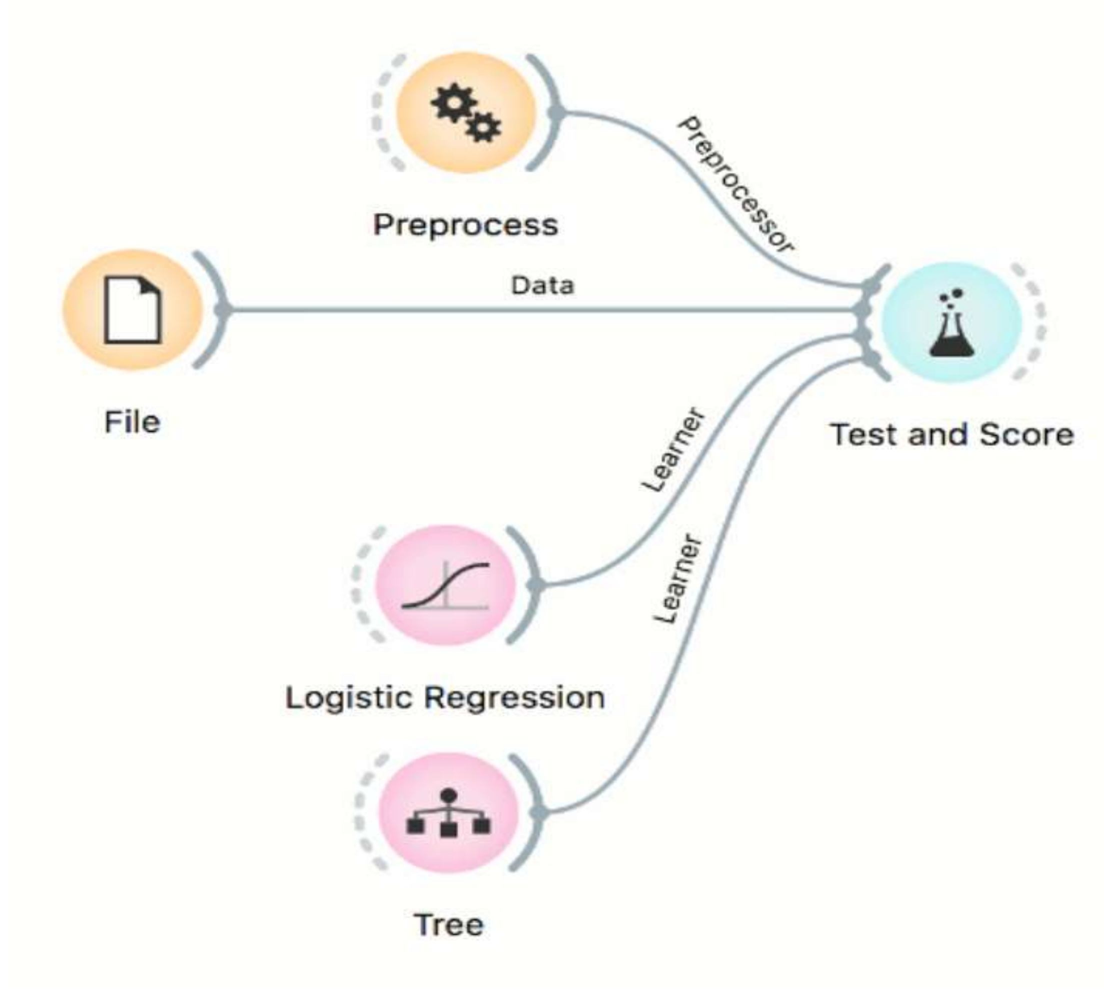


- Select a subset of potentially interesting features, or simply select the features we want to keep.
- The widget will outputs a new data set with only these selected feature.

Workflows with Models

- Predictive models are evaluated in Test and Score widget.

- Test and Score accepts several inputs:
    1. Data (data set for evaluating models).
    2. Learners (algorithms to use for training the model).
    3. Optional preprocessor (for normalization or feature selection).



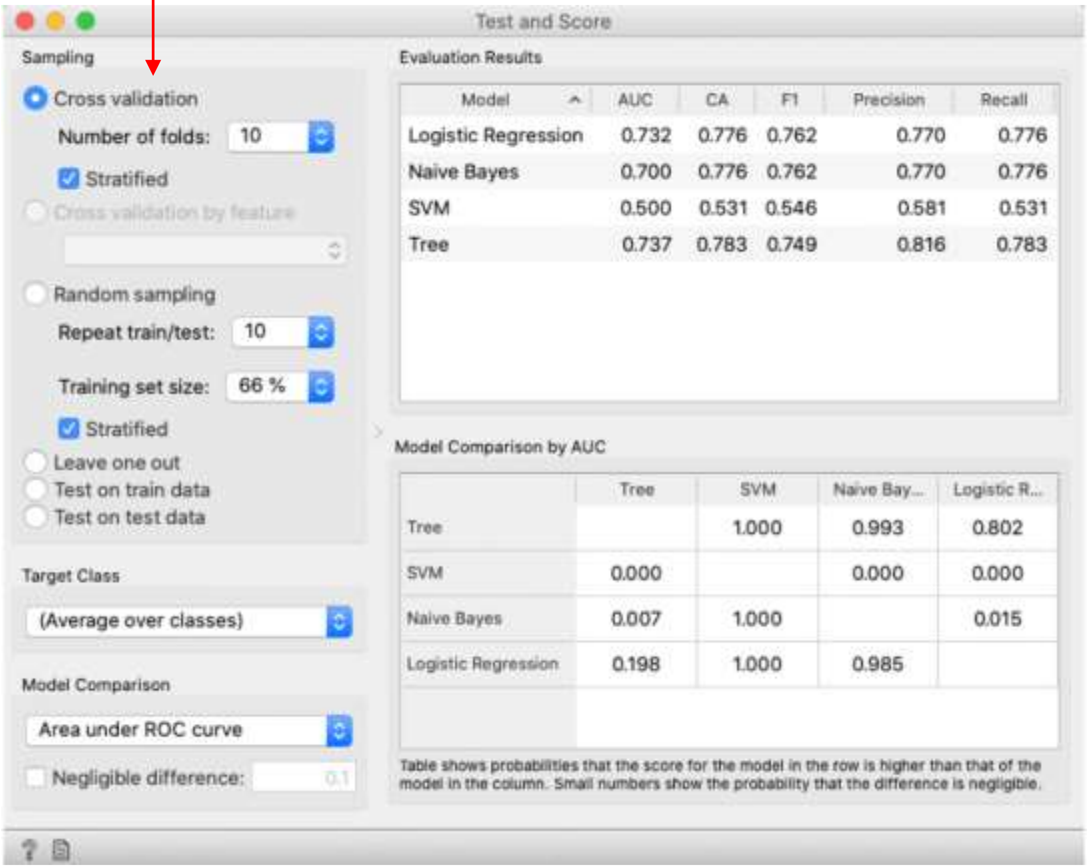
Workflows with Models

Sampling setting (e.g., performs cross-validation or some other train-and-test procedures).

The widget does two things:

1. It shows evaluation results (results of testing different classification/regression algorithms).

2. It outputs evaluation results, which can be used by other widgets for analysing the performance of classifiers, such as confusion matrix.



Workflows with Models - Evaluation

- Confusion matrix widget: show proportions between the predicted and actual class.

Inputs:

- Evaluation results: results of testing classification algorithms.

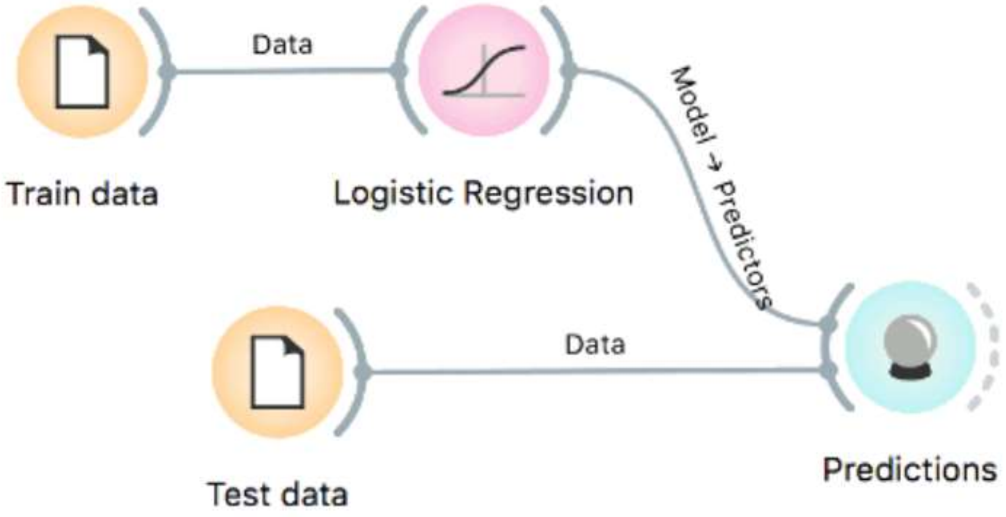
Outputs:

- Selected Data: data subset selected from confusion matrix.
- Data: data with the additional information on whether a data instance was selected.

The test results are fed into the Confusion Matrix, where we can observe how many instances were misclassified and in which way.

Workflows with Models

- Predictions on new data are done in Predictions:
  - The training data is first passed to the model.
  - Once the model is trained, it is passed to Predictions.
  - The Predictions widget also needs data to predict on (Test data), which are passed as a second input.

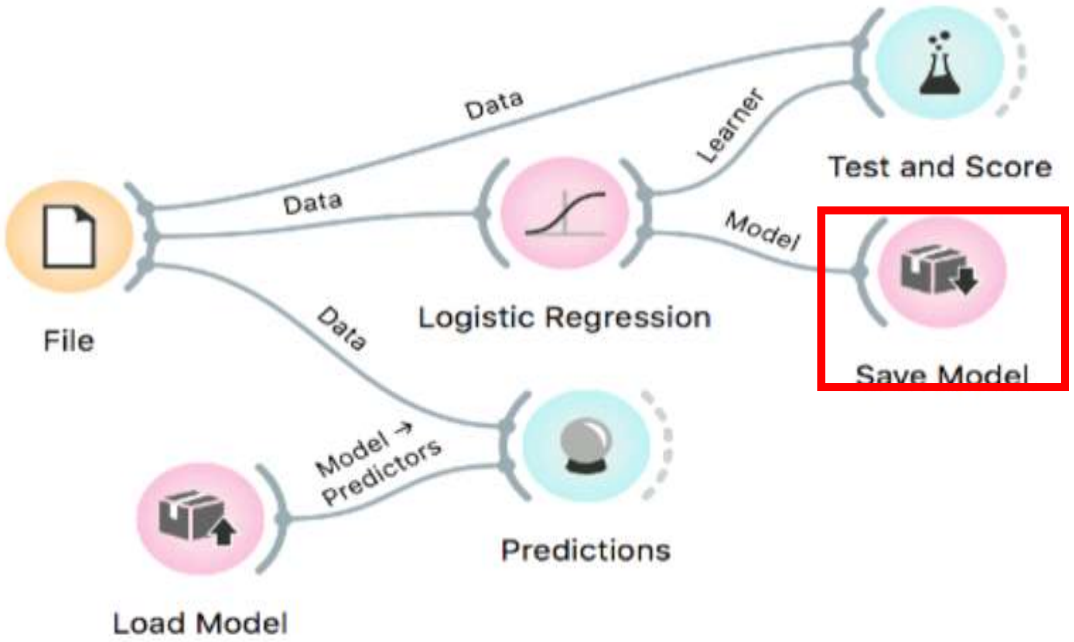


Workflows with Models

- Predictive models can be saved and reused in different Orange. Workflows.

- To save a model:
  1. Models first require data for training.
  2. They output a trained model, which can be saved with Save Model widget in the pickle format.

- Trained model can be loaded and used in Predictions and/or elsewhere.

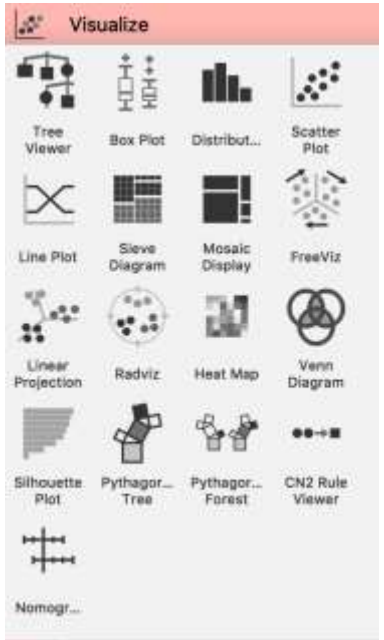


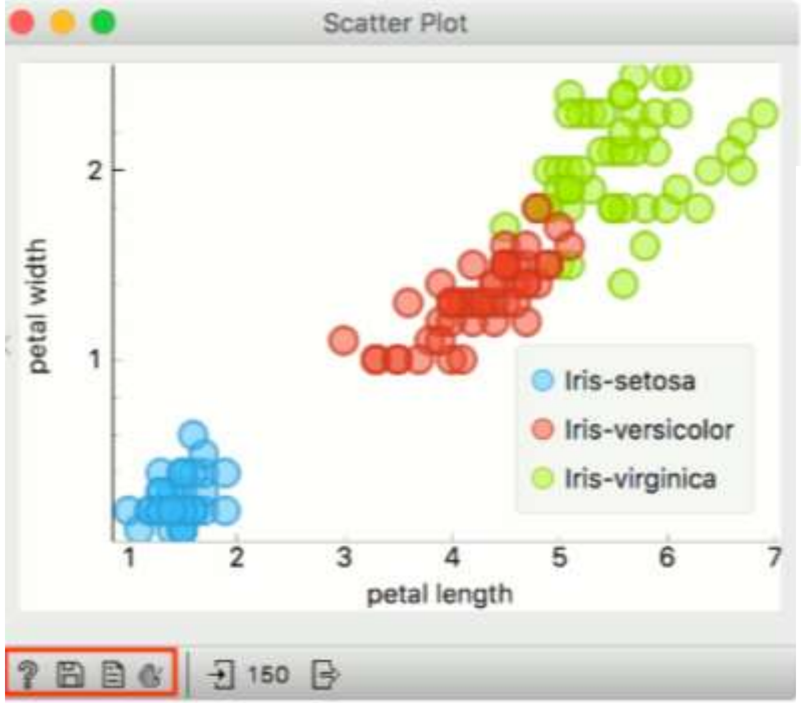
Visualizing data from input data file

Visualizations in Orange

- Visualizations are an essential part of data science.

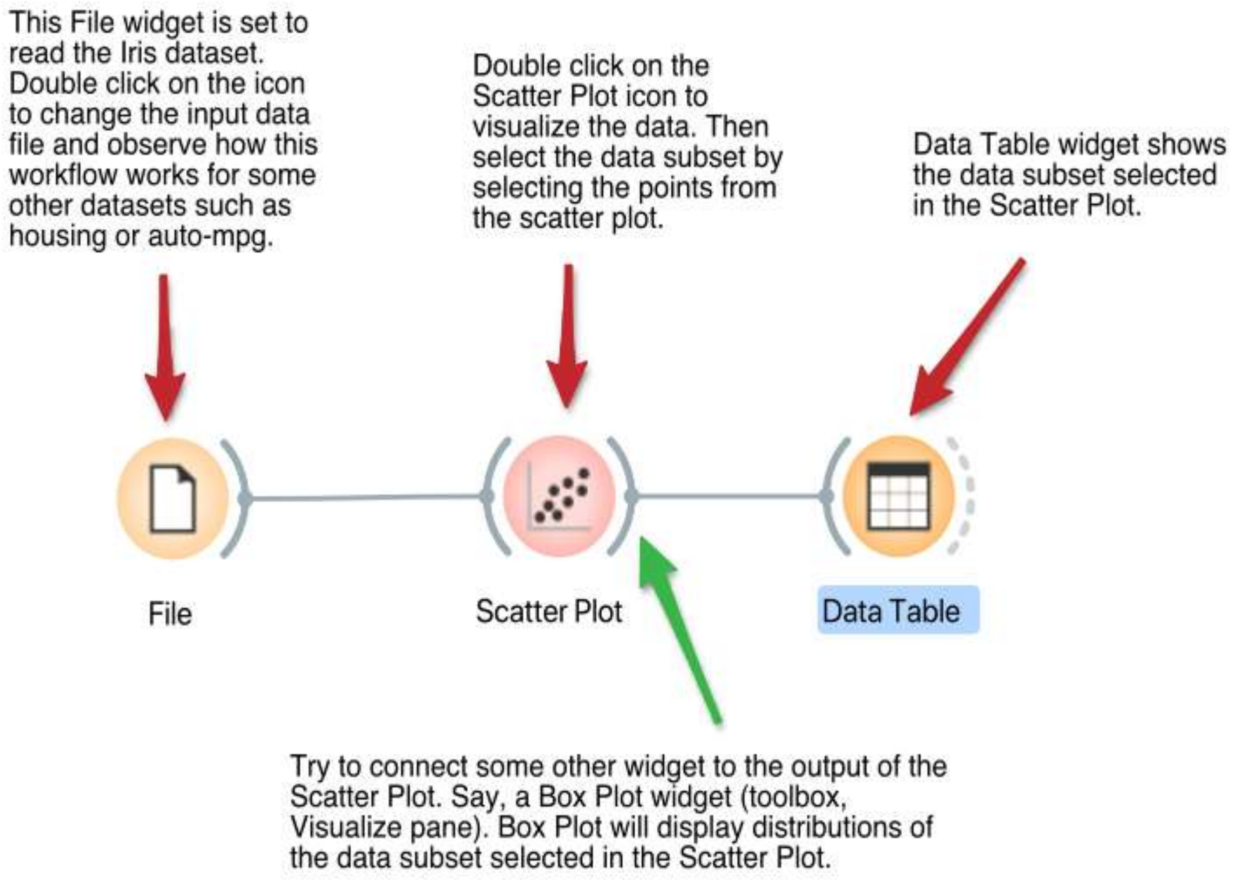
- Visualizations in Orange are interactive.





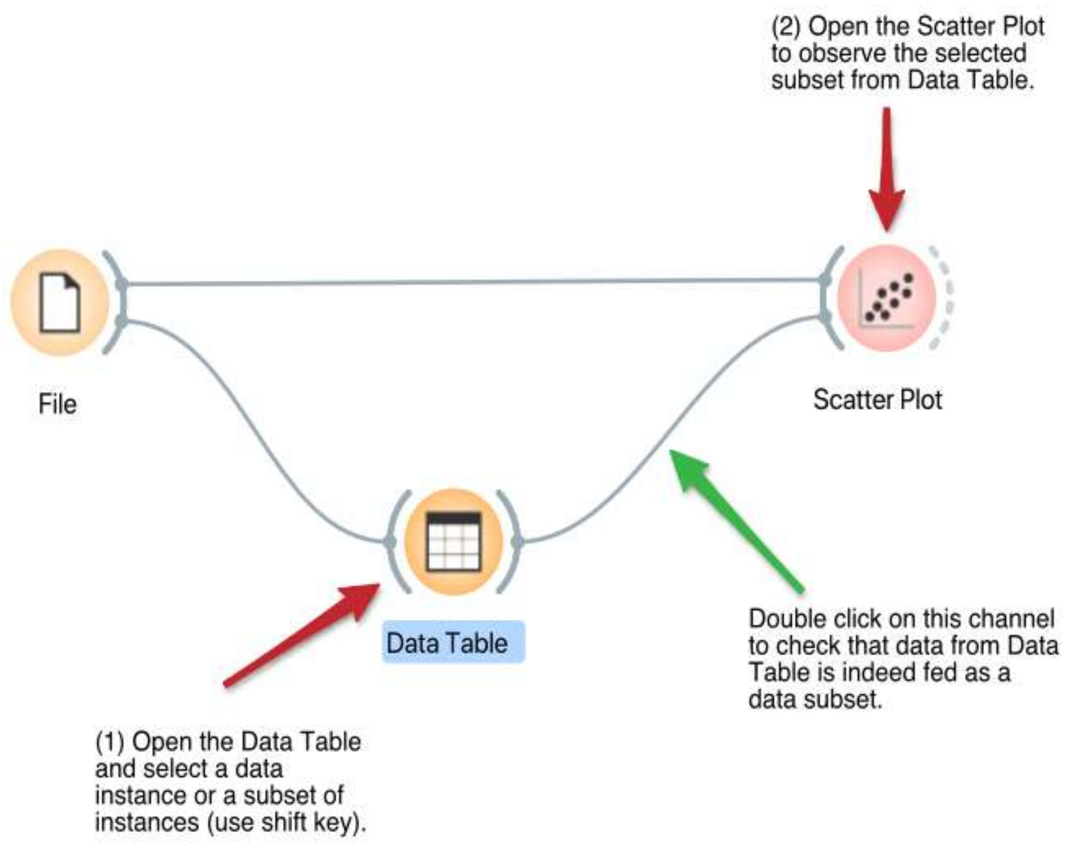
Visualizations in Orange

Interactive visualization workflow example



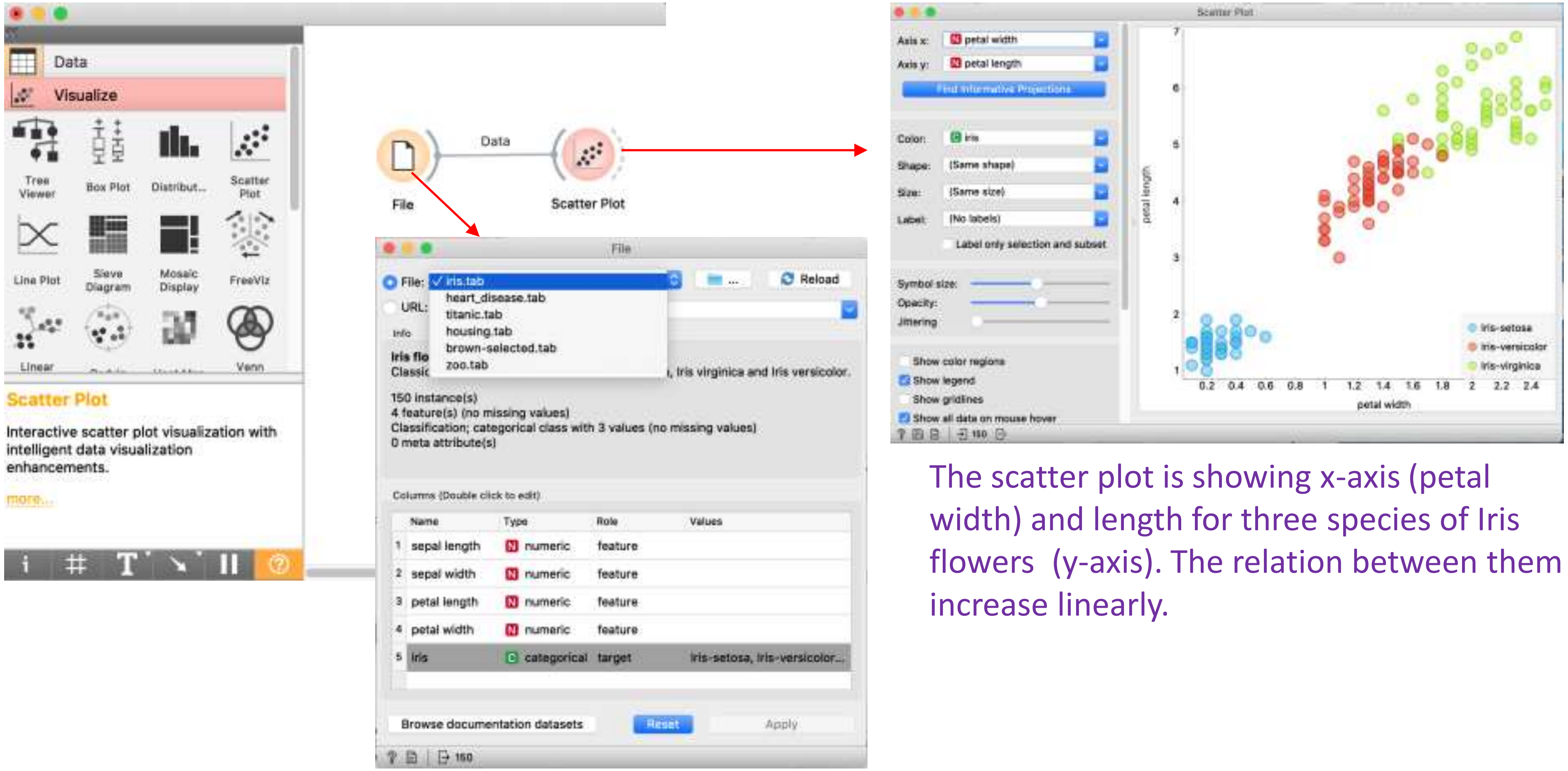
Visualizations in Orange

Visualization workflow of data subsets example



Visualizations in Orange

- Exercise: Build a simple workflow with File and Scatter Plot for Iris dataset.



The scatter plot is showing x-axis (petal width) and length for three species of Iris flowers (y-axis). The relation between them increase linearly.

Visualizations in Orange

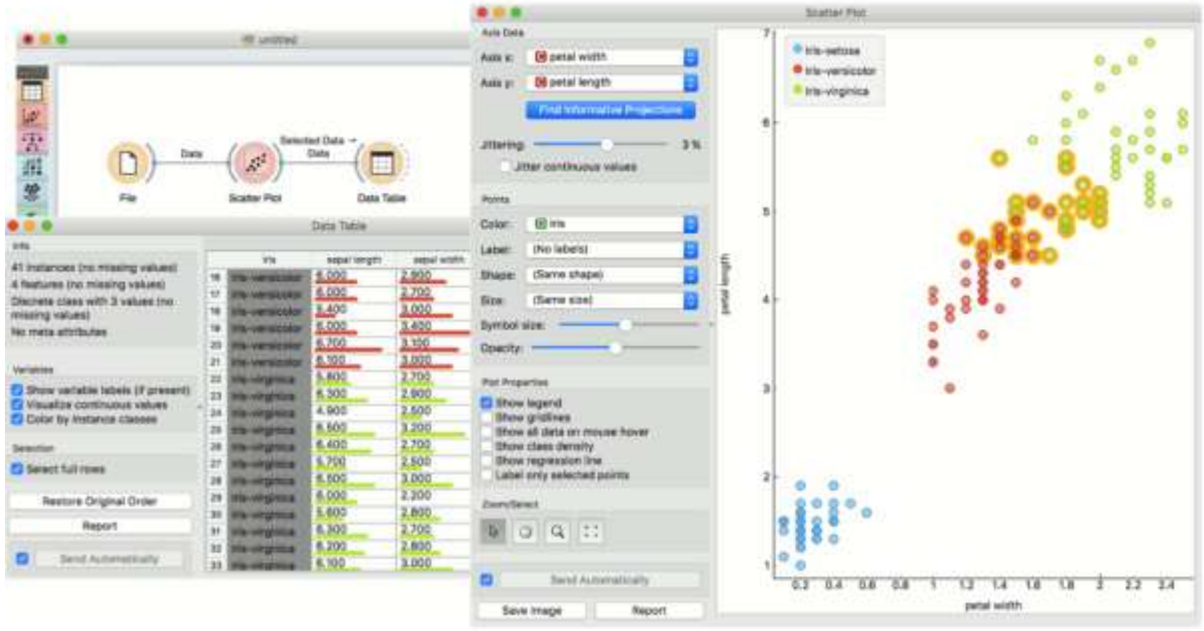
- Scatter Plot supports zooming-in and out of part of the plot and a manual selection of data instances.

- Example: Explorative data analysis using Iris dataset.

1. Selected data instances from a rectangular region on Scatter plot.

2. Sent them to the Data Table widget.

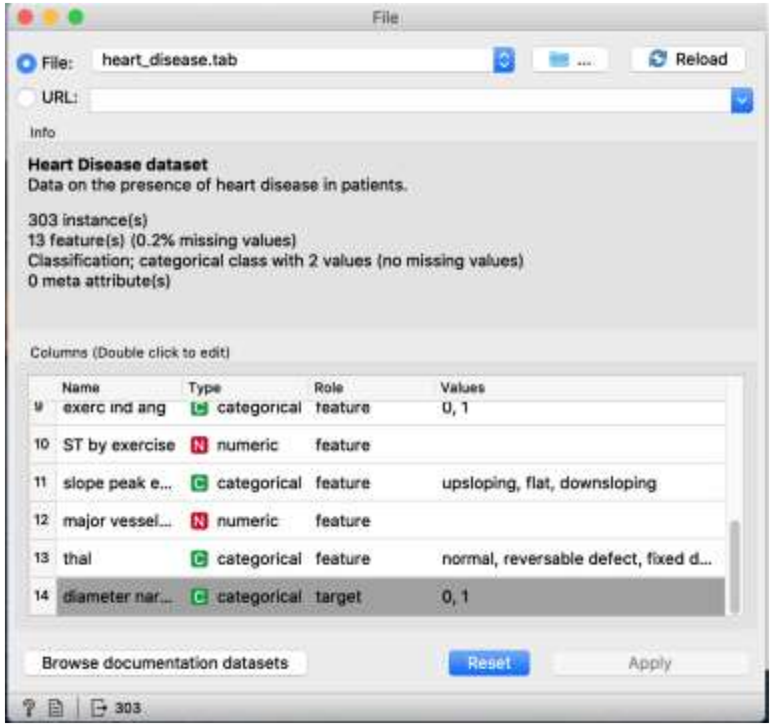
3. Explore the relationship between any two variables.



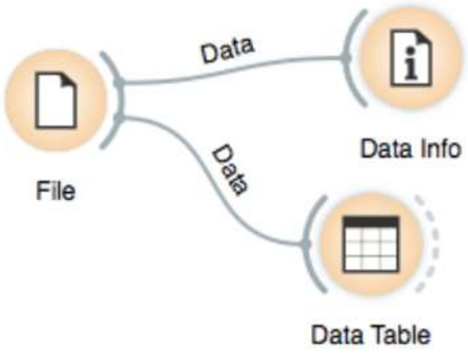
Visualizations in Orange

- Basic Data Exploration from input data file.

Example: heart_disease



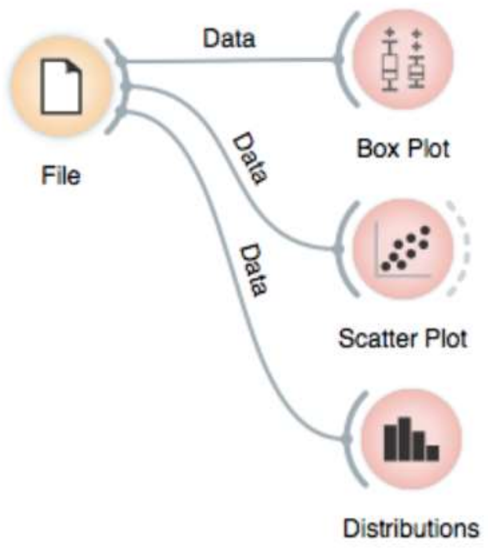
How does the data look?



Cont…

Example: heart_disease

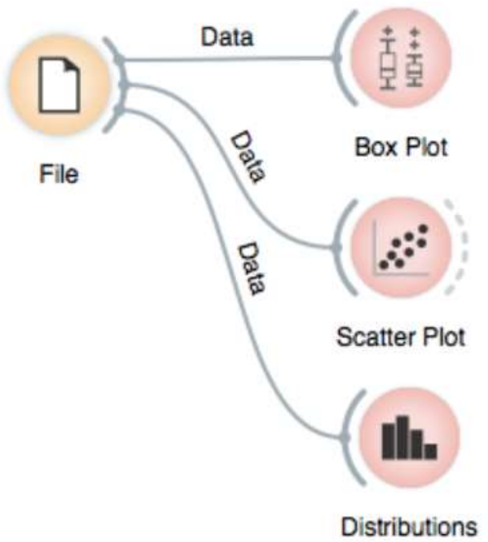
Explore the data with standard visualizations tell us anything interesting!

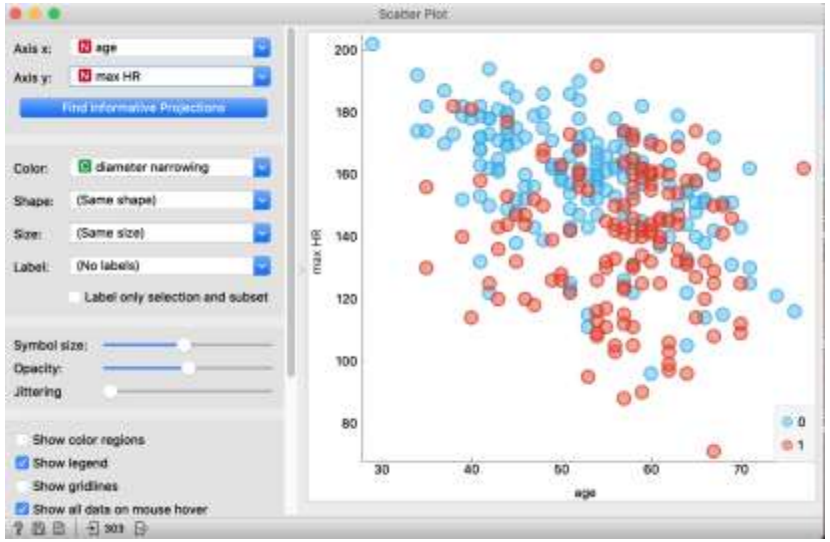


- The Box Plot widget is most commonly used immediately after the File widget to observe the statistical properties of a dataset and discover any anomalies, such as duplicated values, outliers, …)
- The Scatter Plot widget provides a 2-dimensional scatter plot visualization for continuous attributes.
- The Distributions widget displays the value distribution of discrete or continuous attributes.

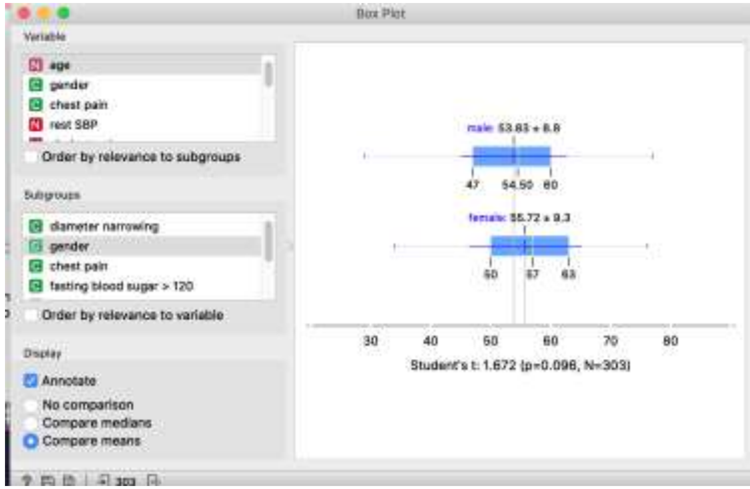
Cont…

Example: heart_disease

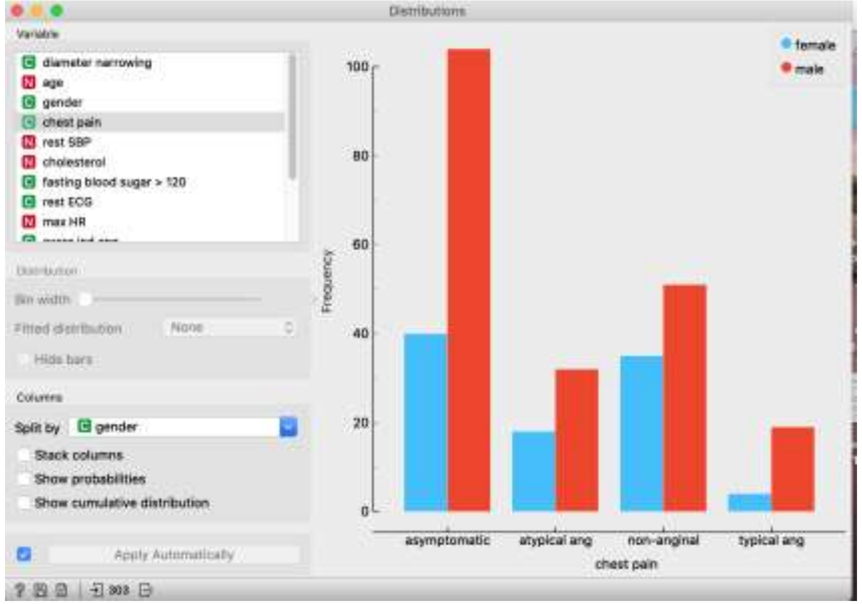




Max - HR decreases with age.



Box plot for attribute 'age' grouped by 'gender'

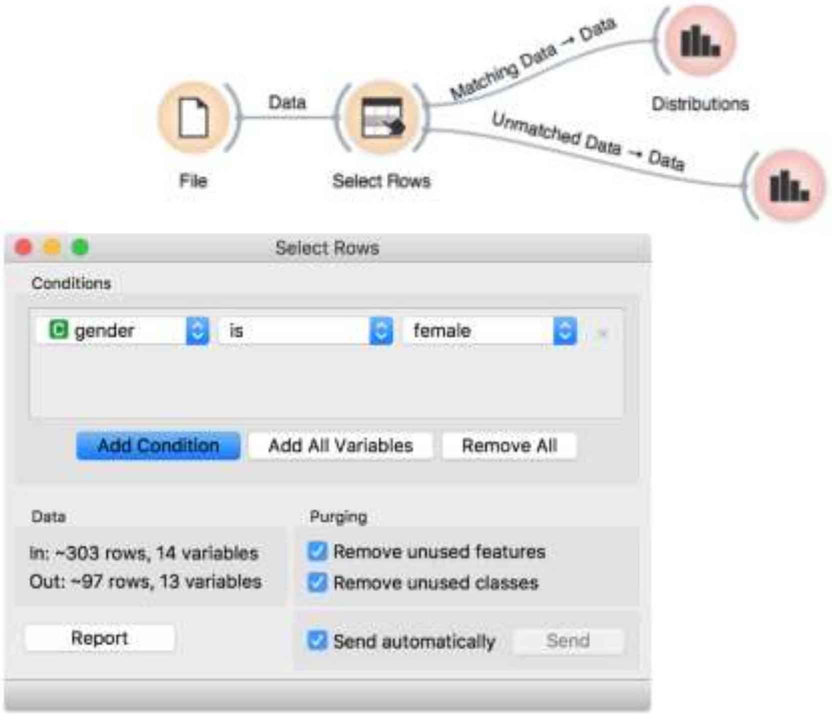


Distribution of 'chest pain' with columns split by 'gender'

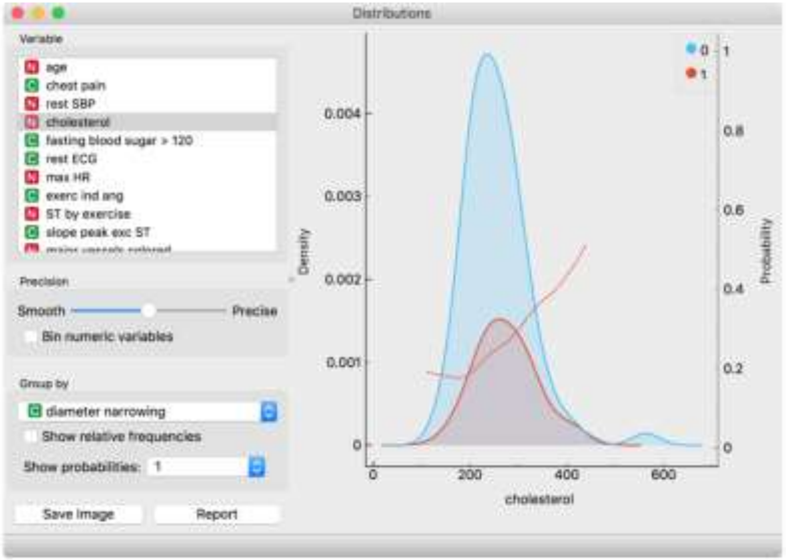
Cont…

Example: heart_disease

- Data can also be split by the value of features and analyse it separately.
- Split data by gender – use select Rows widget



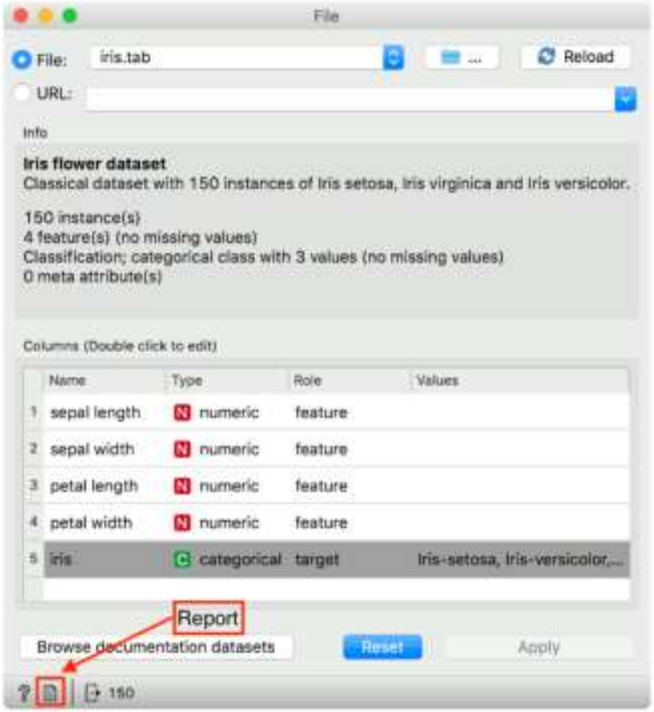
Choose the female patients in Select Rows widget



Selection of data instances works well with visualisation of data distribution and explore the data.

Visualizations in Orange

- Reports
  - Reports allow to trace back analytical steps as it saves the workflow at which each report segment was created.
  - Reports can be saved in .html, .pdf or .report format.

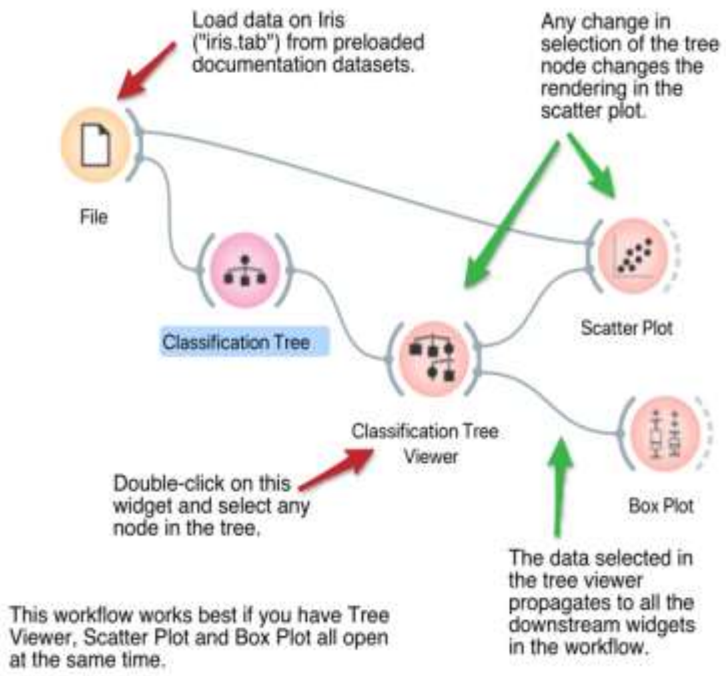


Analyzing data with regression models and decision trees

Analyzing data with decision trees

- Decision tree is is one of the oldest, but still popular, machine learning methods.

- Decision trees workflow example

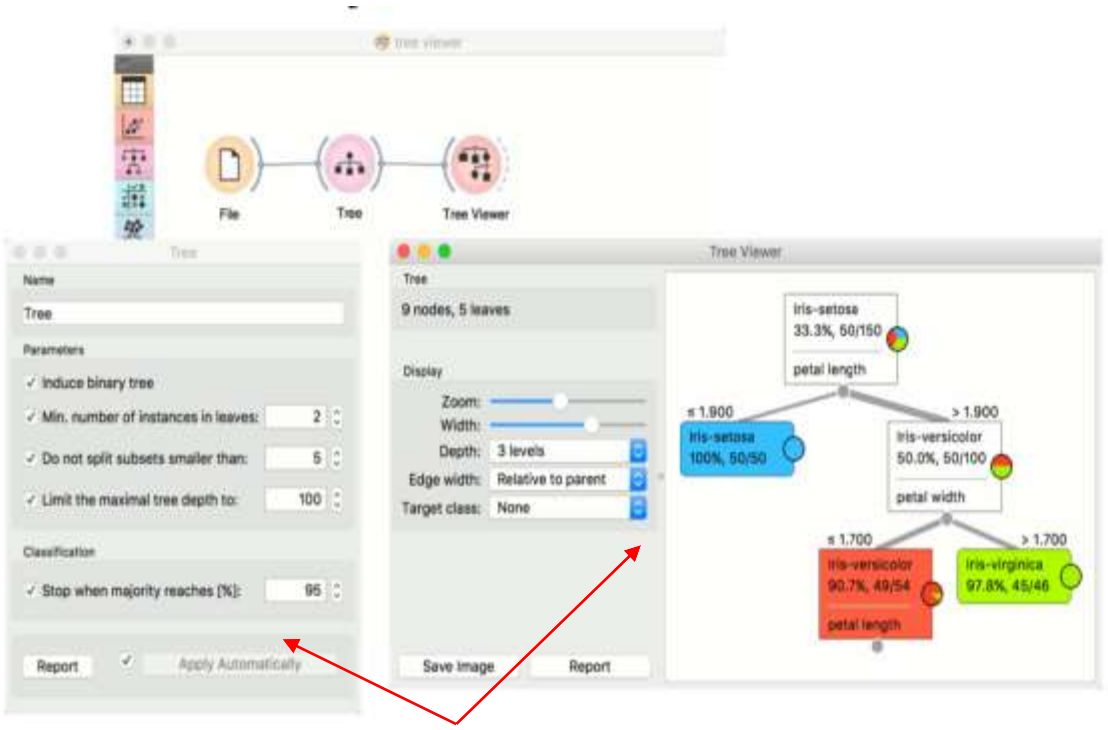


- Decision trees in Orange does not use any data pre-processing.

Analyzing data with decision trees

Tree viewer

- This widget cab be used for or visualizing decision trees.
- To enable explorative data analysis, Select a node, which instruct the widget to output the data associated with that node.
- If both the viewer and Tree are open, any re-run of the tree induction algorithm will immediately affect the visualization.

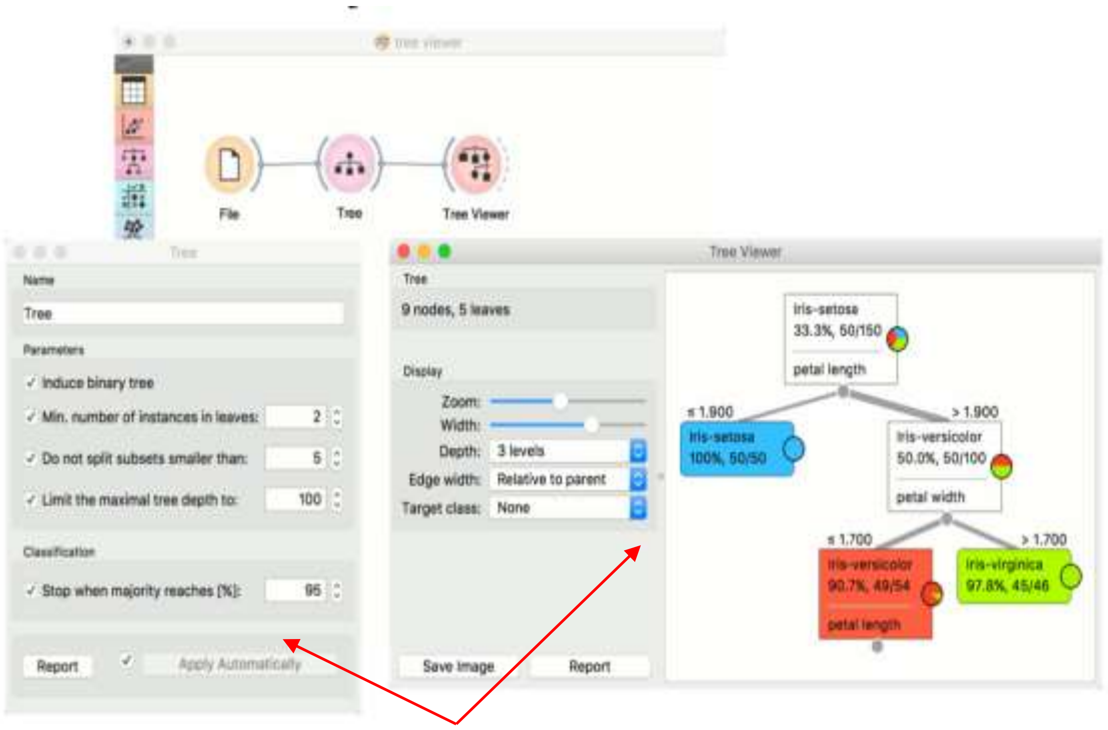


Explore how the parameters of the decision tree algorithm influence the structure of the resulting tree.

Analyzing data with decision trees

Tree parameters:

- **Induce binary tree:** build a binary tree (split into two child nodes).
- **Min. number of instances in leaves:** if checked, the algorithm will never construct a split which would put less than the specified number of training examples into any of the branches.
- **Do not split subsets smaller than:** forbids the algorithm to split the nodes with less than the given number of instances.
- **Limit the maximal tree depth:** limits the depth of the classification tree to the specified number of node levels.
- **Stop when majority reaches [%]:** stop splitting the nodes after a specified majority threshold is reached.

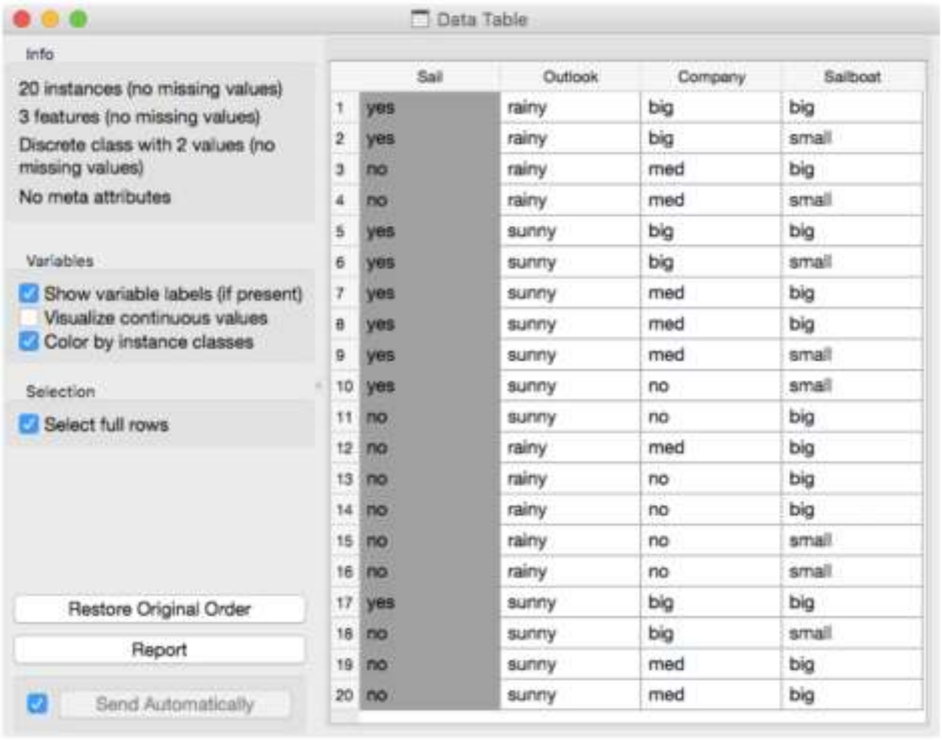


Explore how the parameters of the decision tree algorithm influence the structure of the resulting tree.

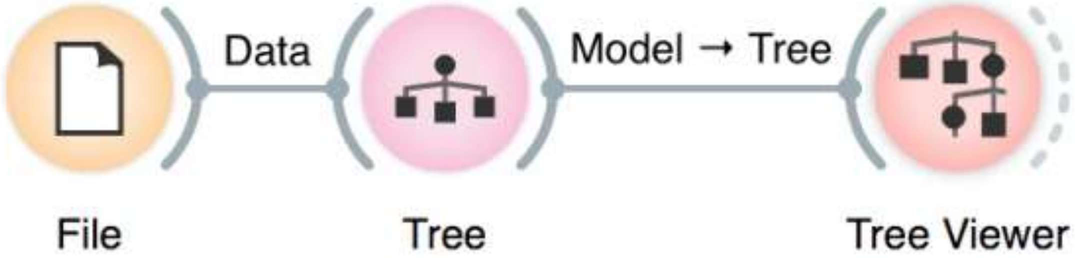
Analyzing data with decision trees

Example: Using sailing data, predict the conditions under which a friend skipper went sailing

Load the data → Build a tree → visualize it in the Tree Viewer



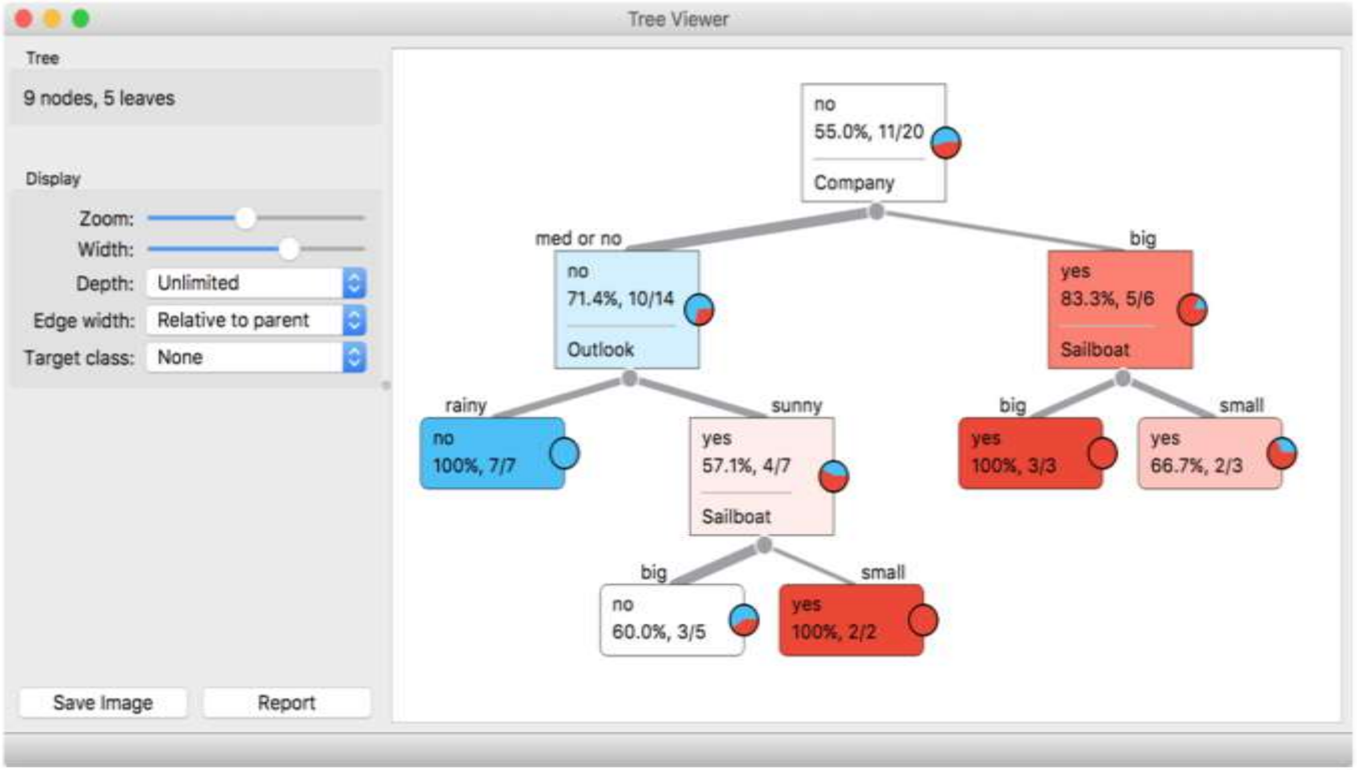
	Sail	Outlook	Company	Sailboat
1	yes	rainy	big	big
2	yes	rainy	big	small
3	no	rainy	med	big
4	no	rainy	med	small
5	yes	sunny	big	big
6	yes	sunny	big	small
7	yes	sunny	med	big
8	yes	sunny	med	big
9	yes	sunny	med	small
10	yes	sunny	no	small
11	no	sunny	no	big
12	no	rainy	med	big
13	no	rainy	no	big
14	no	rainy	no	big
15	no	rainy	no	small
16	no	rainy	no	small
17	yes	sunny	big	big
18	no	sunny	big	small
19	no	sunny	med	big
20	no	sunny	med	big



Analyzing data with decision trees

Example: Using sailing data, predict the conditions under which a friend skipper went sailing

- Trees place the most useful feature at the root.
- The most useful feature is the feature that splits the data into two purest possible subsets.
- These are then split further, again by the most informative features.
- This process of breaking up the data subsets to smaller ones repeats until we reach subsets where all data belongs to the same class.
- These subsets are represented by leaf nodes in strong blue or red.
- The process of data splitting can also terminate when it runs out of data instances or out of useful features (the two leaf nodes in white).

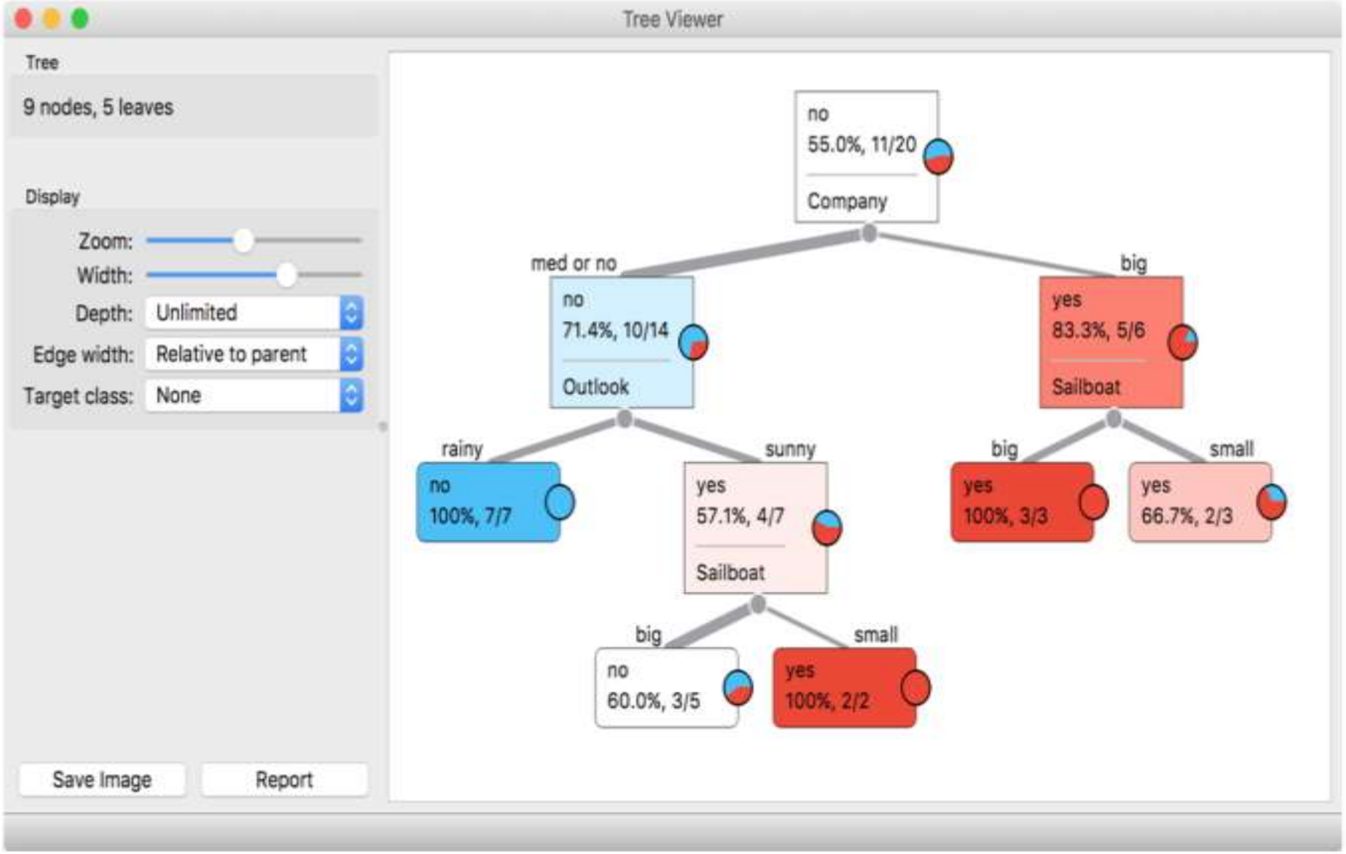


Analyzing data with decision trees

Example: Using sailing data, predict the conditions under which a friend skipper went sailing

According to the decision tree results:

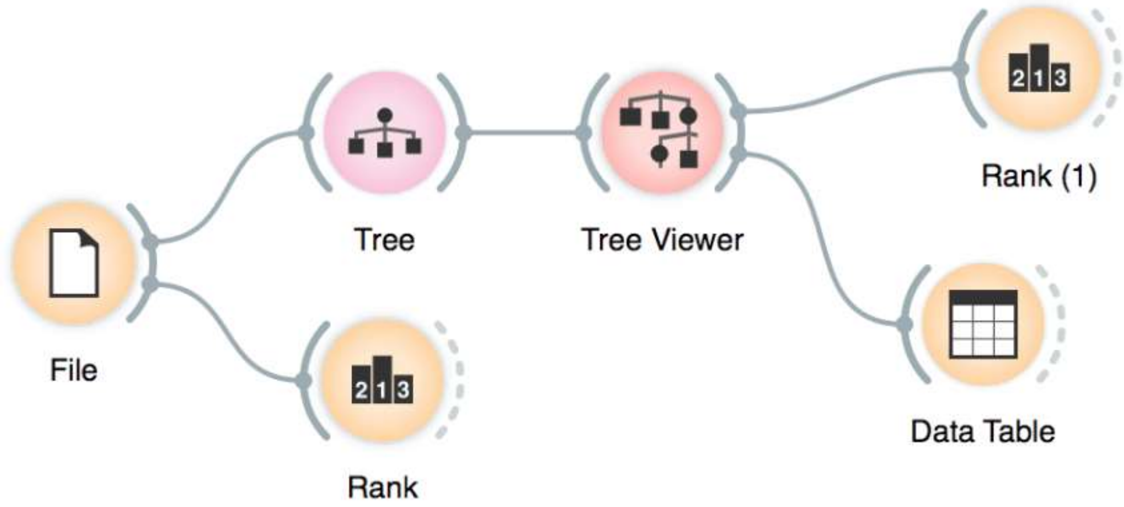
- It looks like this skipper is a social person; as soon as there's company, the probability of her sailing increases.
- When joined by a smaller group of individuals, there is no sailing if there is rain (Thunderstorms? Too dangerous?)
- When she has a smaller company, but the boat at her disposal is big, there is no sailing either.

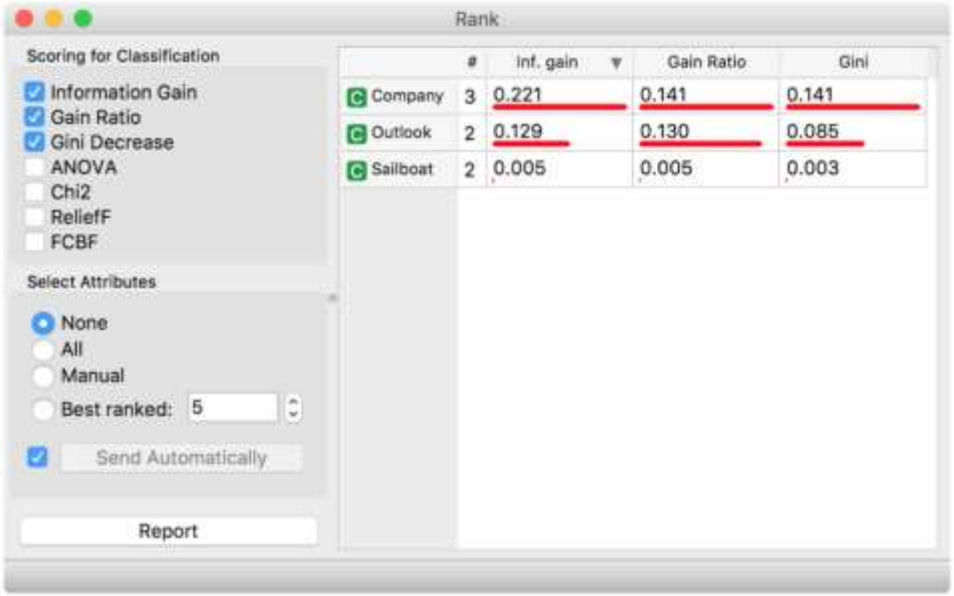


Analyzing data with decision trees

Example: Using sailing data, predict the conditions under which a friend skipper went sailing

- What are the most the most useful” feature?
- **Rank widget -** estimates the quality of data features and ranks them according to how much information they carry.

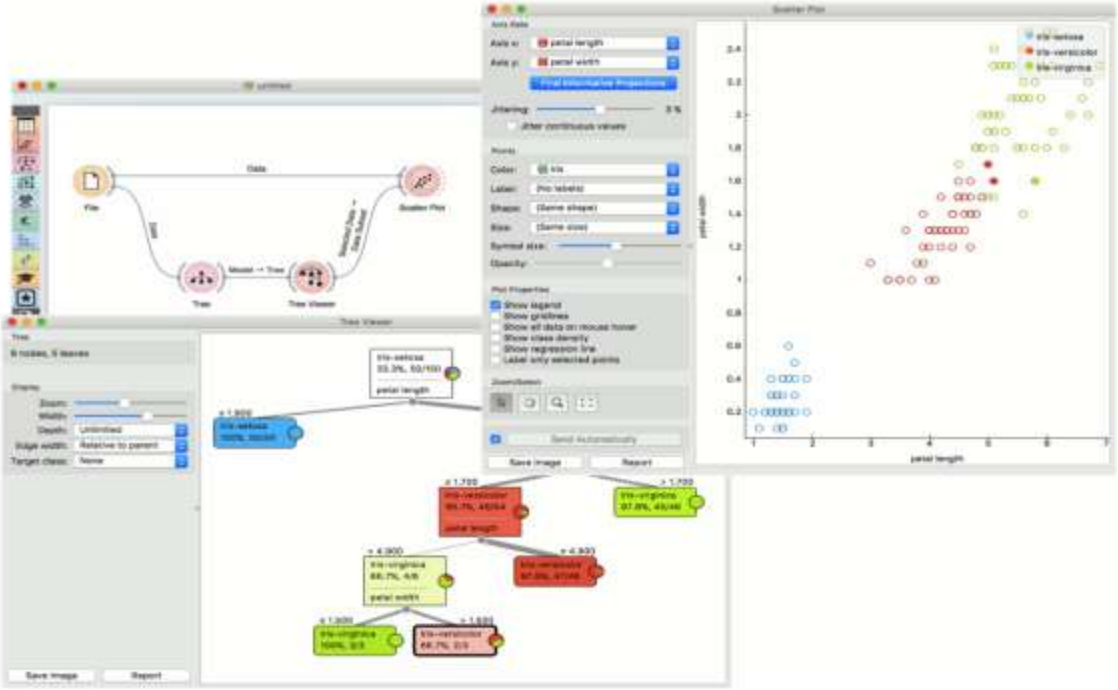




Analyzing data with decision trees

Model inspection example

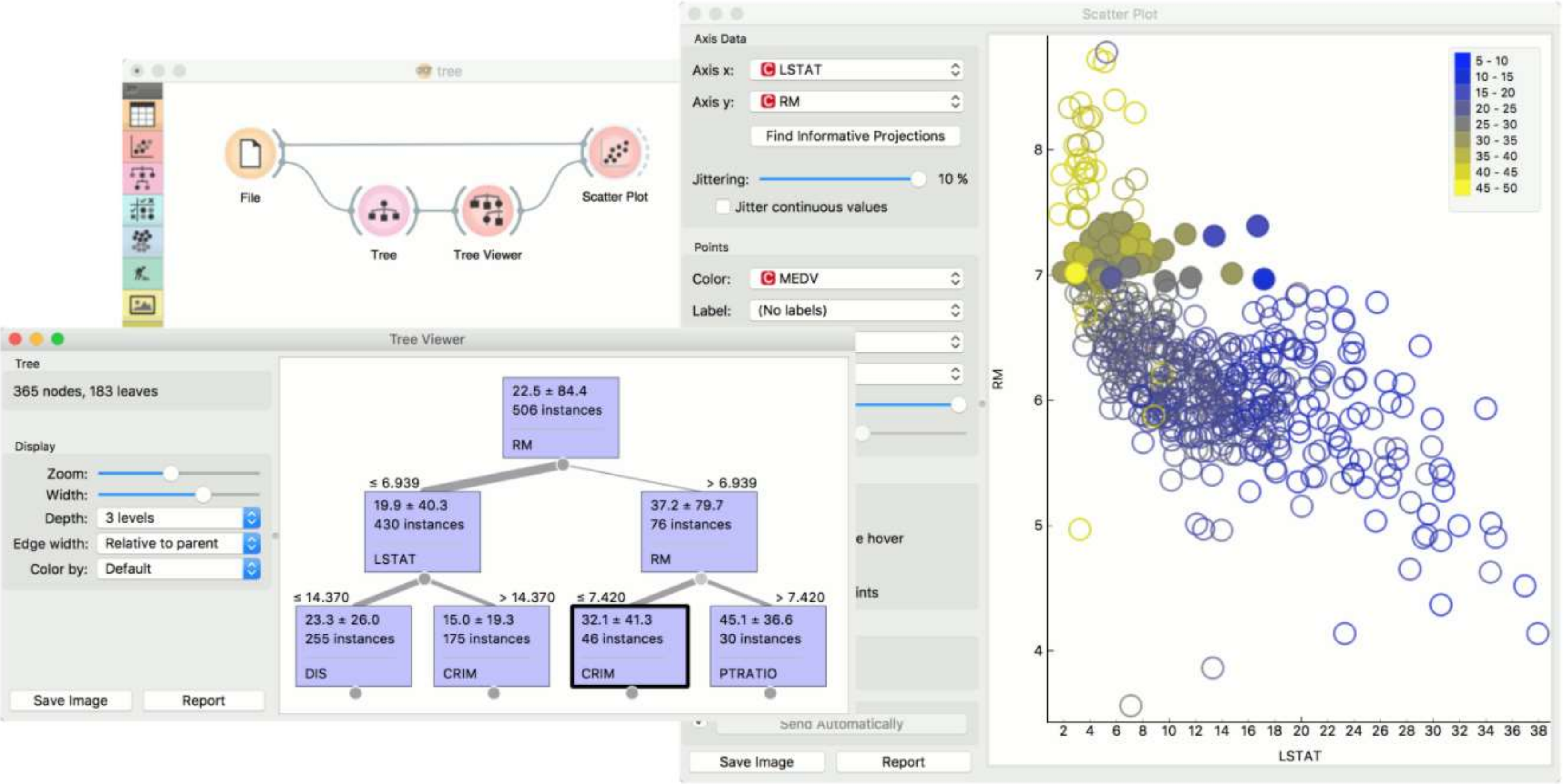
- To inspect a model, combine Tree and Scatter Plot widgets to display instances taken from a chosen decision tree node.



Iris dataset: Model inspection example

Analyzing data with decision trees

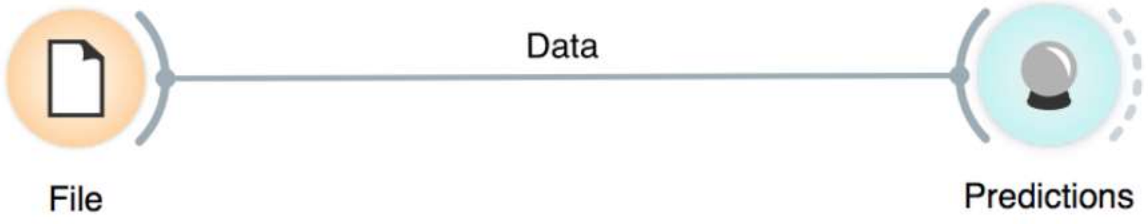
- Decision trees works for regression tasks.



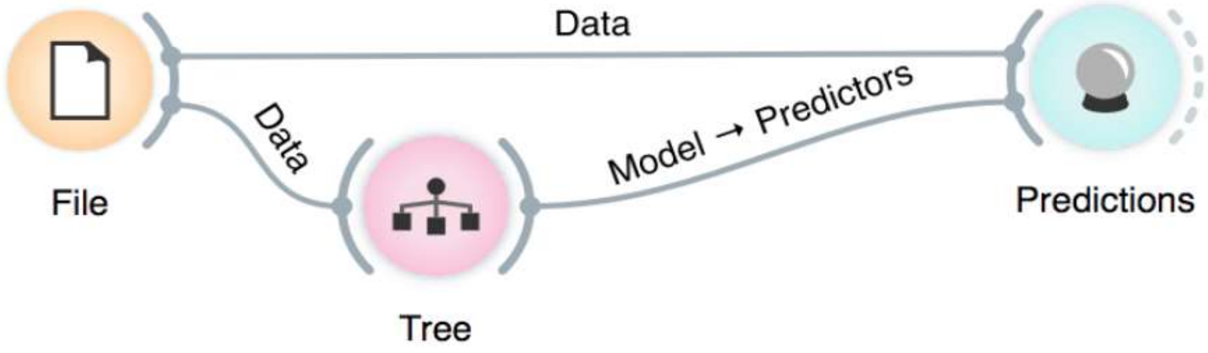
Decision tree for housing dataset example

Prediction models

- Predictions widget will shows the data, but makes no predictions.



- To analyse data with prediction model, a predictive model is needed.
  - The Predictions widget uses the model to make predictions about the data and shows them in the table
- E.g.:



Model widget is channel that carries a model

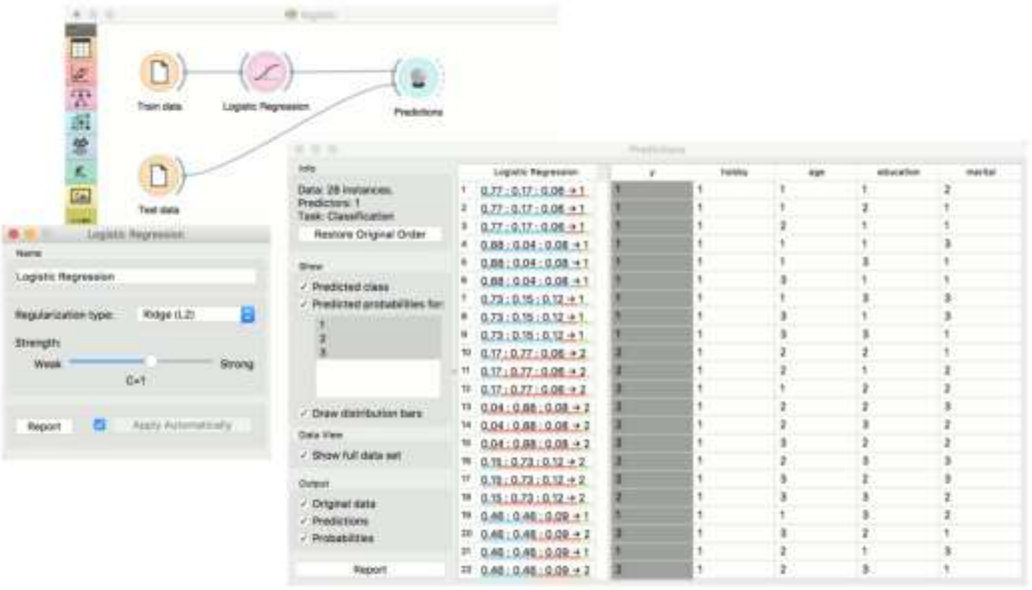
1. The data is fed into the model widget to infer a predictive model.
2. The Predictions widget gets the data from the File widget and also a predictive model from the model widget.

Analyzing data with regression models

Two regression models are available:

- Liner Regression works with continuous data.
- Logistic Regression only works for classification tasks (It learns a Logistic Regression model from the data).

Example: Demonstrate prediction results with logistic regression on hayes-roth dataset.



Traning:

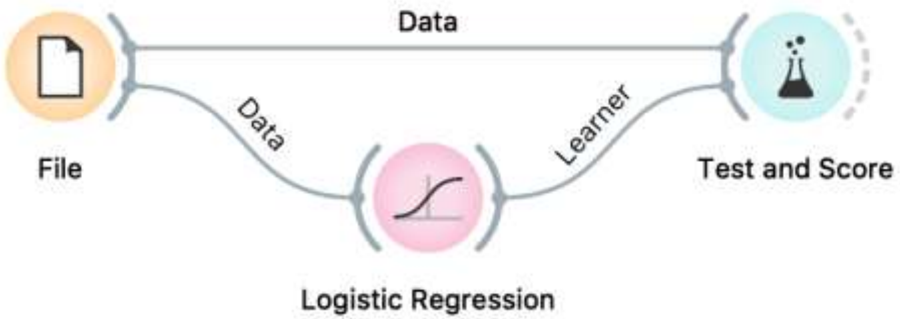
1. First load hayes-roth dataset to File widget.
2. Pass the data to Logistic Regression model for training.
3. Pass the trained model to Prediction widget.

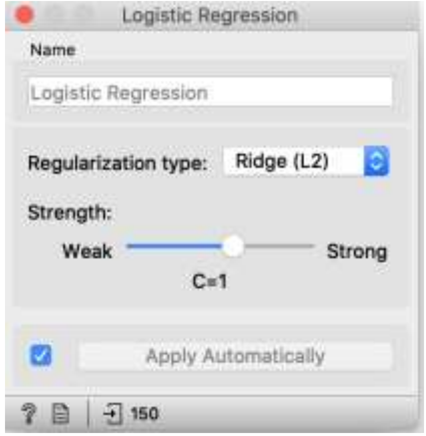
Testing: predict class value on a new dataset

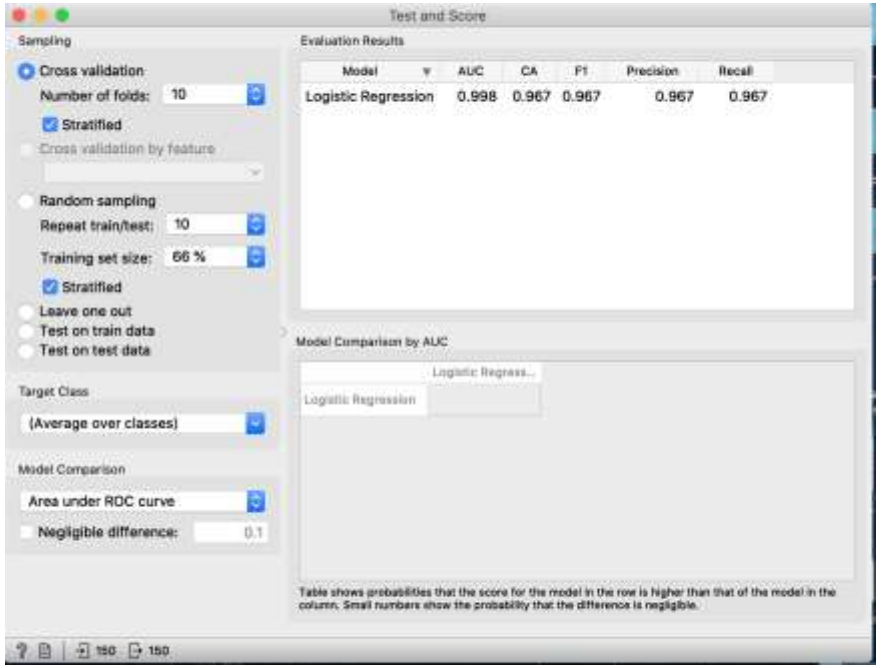
1. Load hayes-roth_test in the second File widget
2. Connect it to Predictions.
3. Observe class values predicted with Logistic Regression directly in Predictions.

Analyzing data with regression models

Example: Predict Iris flower type using Logistic Regression.

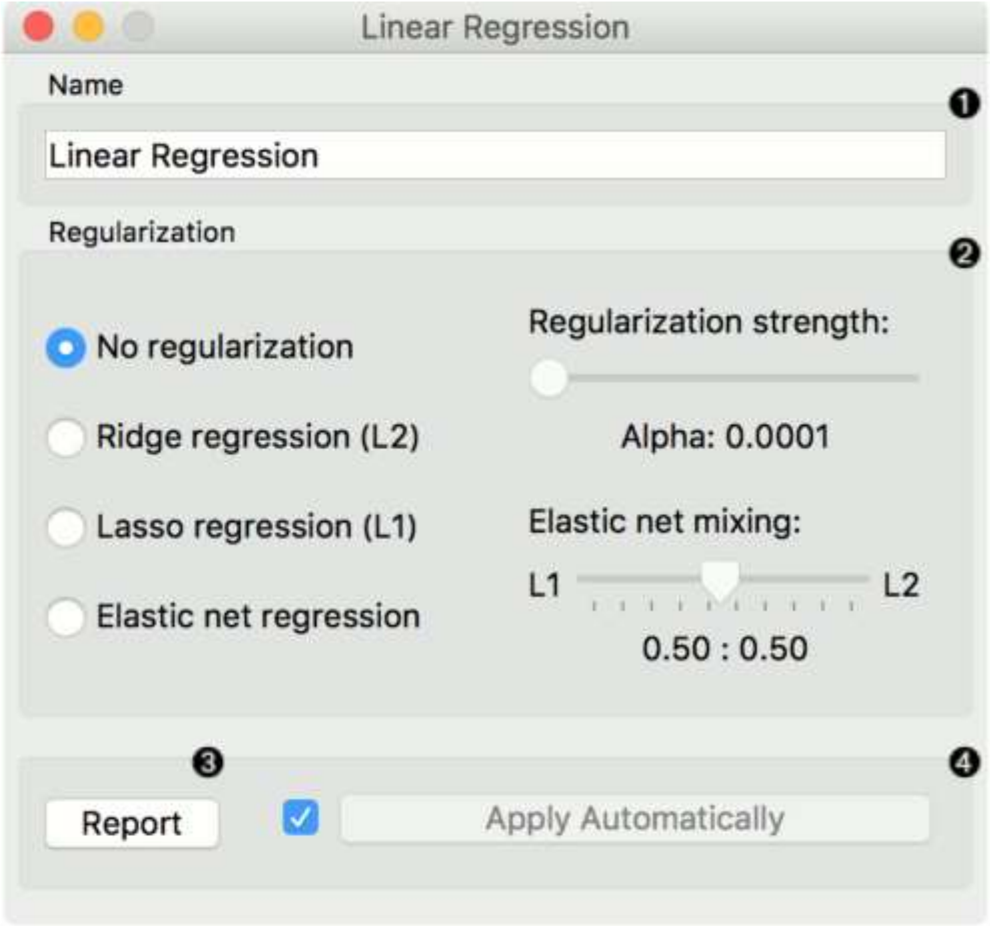






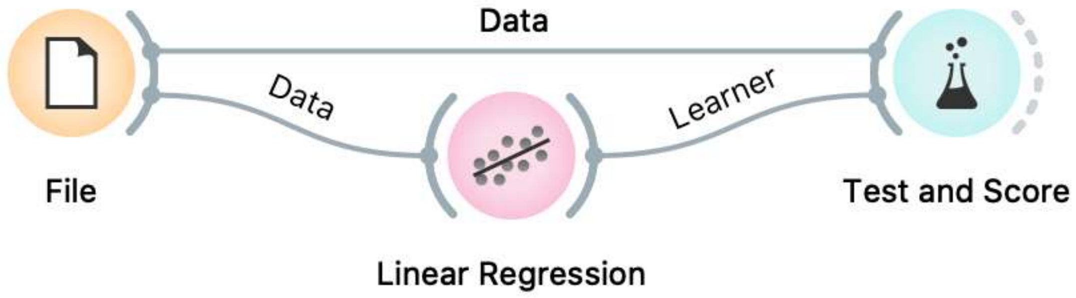
Analyzing data with regression models

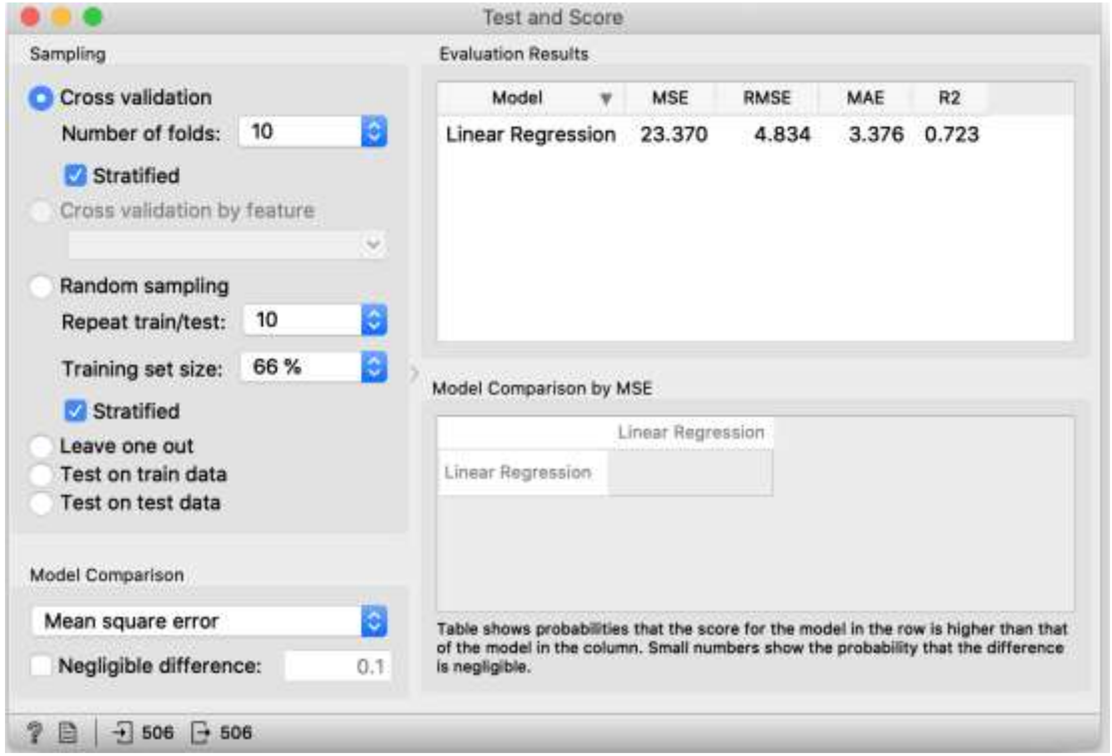
- Linear Regression widget constructs a learner/predictor that learns a linear function from its input data.
- The model can identify the relationship between a predictor *x* and the response variable *y*.
- Linear regression works only on regression tasks.



Analyzing data with regression models

- Example: Train a Linear Regression on housing dataset and evaluated its performance in Test & Score.





Analyzing data with deep learning models

Analyzing data with deep learning models

- Neural Network widget is a multi-layer perceptron (MLP) algorithm with backpropagation.

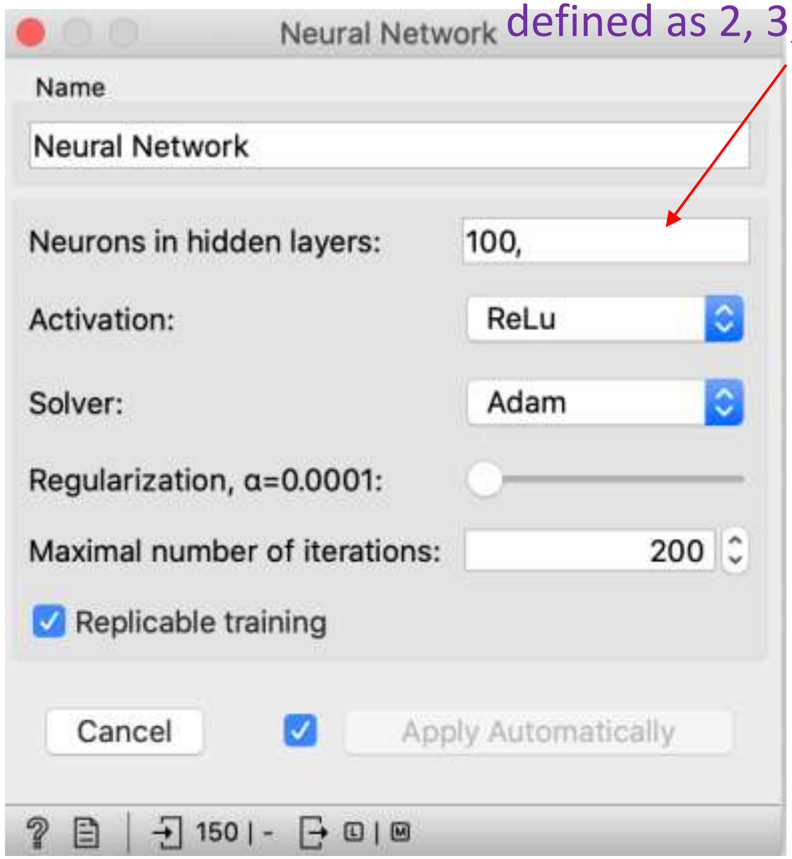
Inputs

- Data: input dataset
- Preprocessor: preprocessing method(s)

Outputs

- Learner: multi-layer perceptron learning algorithm
- Model: trained model

neural network with 3 layers can be defined as 2, 3, 2



Analyzing data with deep learning models

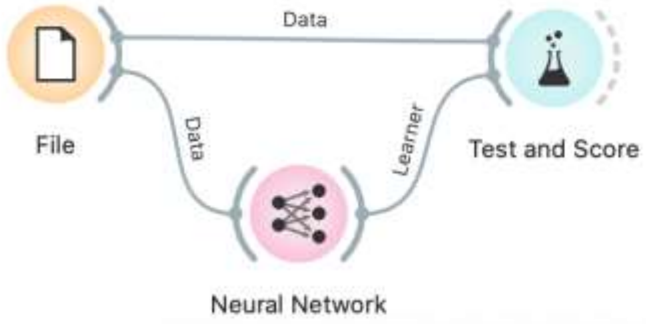
- Neural Network uses default preprocessing when no other preprocessors are given. It executes them in the following order:

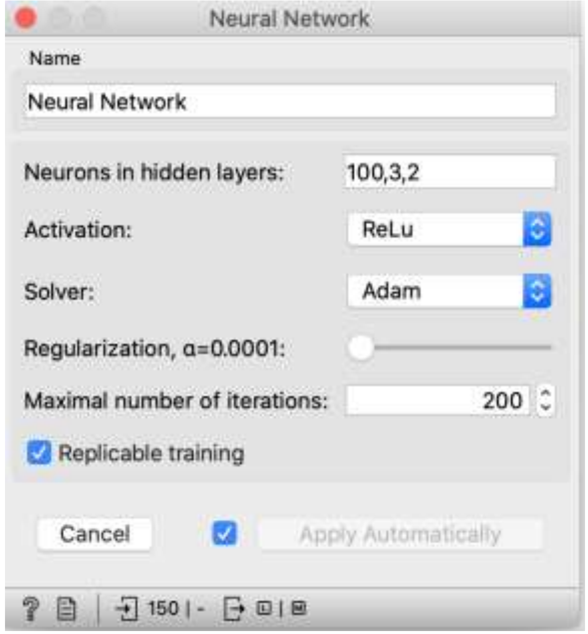
    1. Removes instances with unknown target values
    2. Continuizes categorical variables (with one-hot-encoding)
    3. Removes empty columns
    4. Imputes missing values with mean values
    5. Normalizes the data by centering to mean and scaling to standard deviation of 1

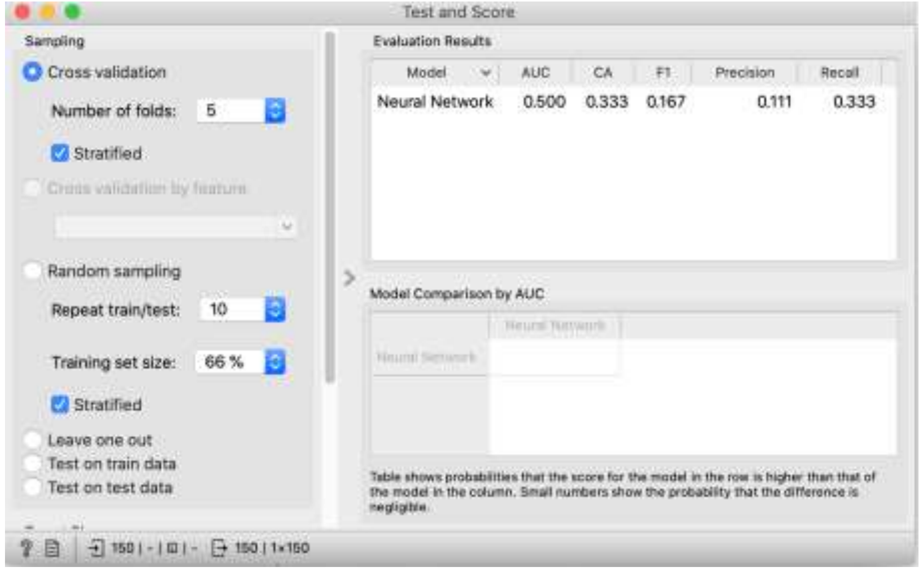
- To remove default preprocessing, connect an empty Preprocess widget to the learner.

Analyzing data with deep learning models

Example: Neural Network Workflow for classification task on the iris data.



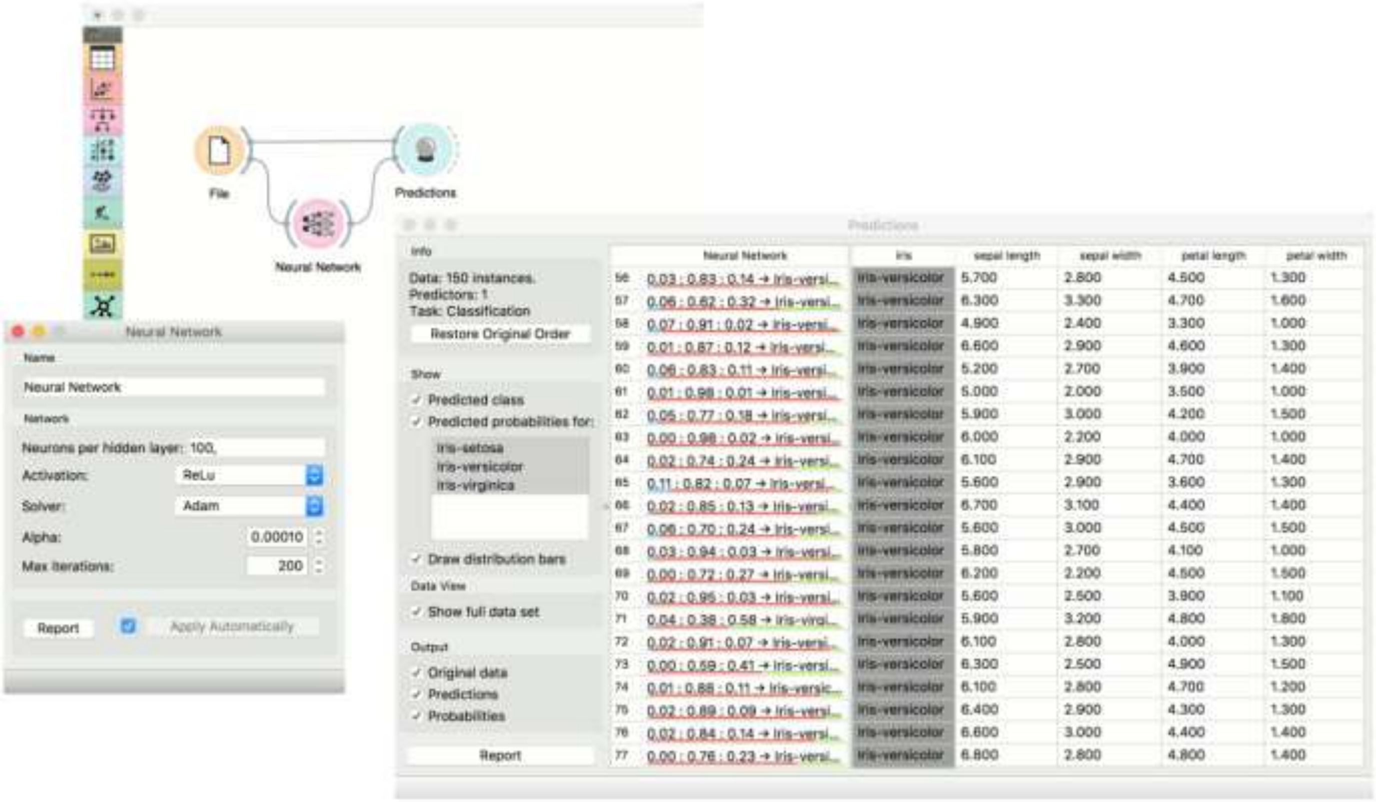




Analyzing data with deep learning models

Example: Neural Network Workflow for a prediction task on the iris data.

1. Input the Neural Network prediction model into Predication.
2. Observe the predicted values.



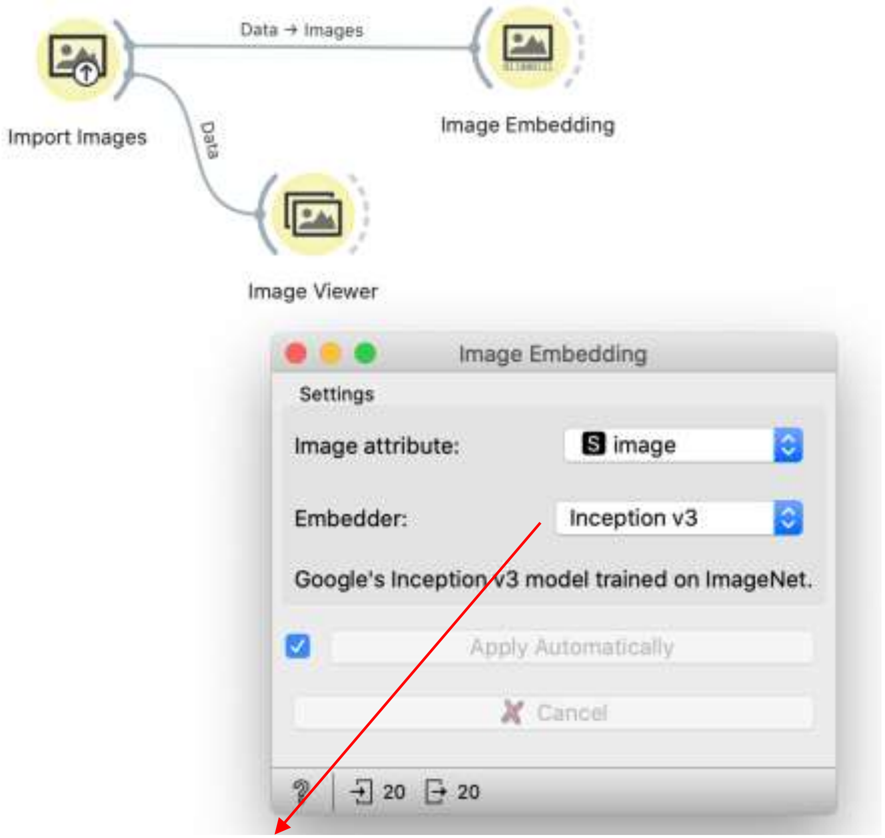
Analyzing data deep learning models

Example: image analytics workflow on domestic animal image dataset using Image Analytics add-on.

1. Import the image data via the Import Images widget.

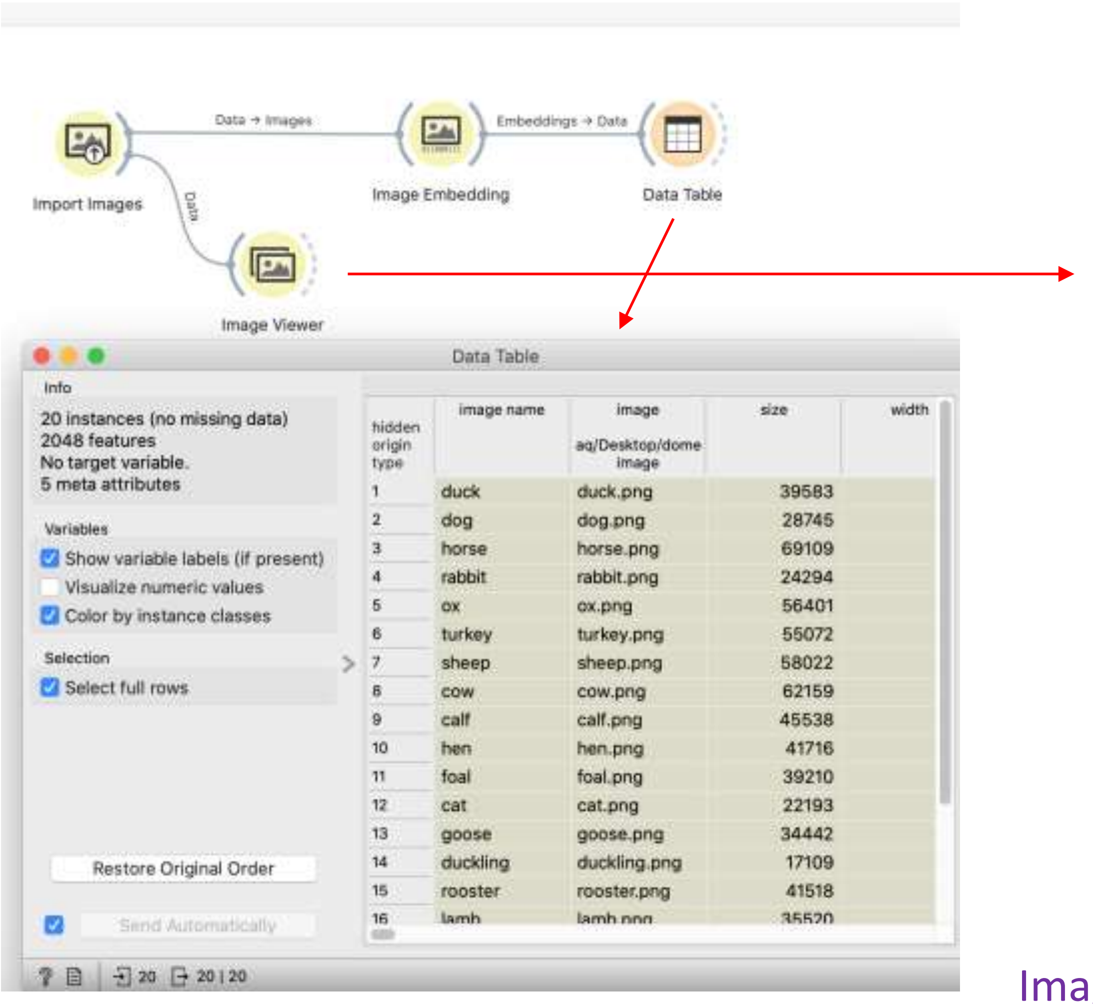
2. display all of the loaded images using Image Viewer widget.

3. For image data analysis Image embeddings widget must be used as classification and regressions tasks accept data in the form of numbers.



The most important parameters for the Image Embedding interface is the Embedder. Supported Deep network embeddings: SqueezeNet, Inception v3, VGG-16, VGG-19, Painters, Deeploc.

Analyzing data deep learning models



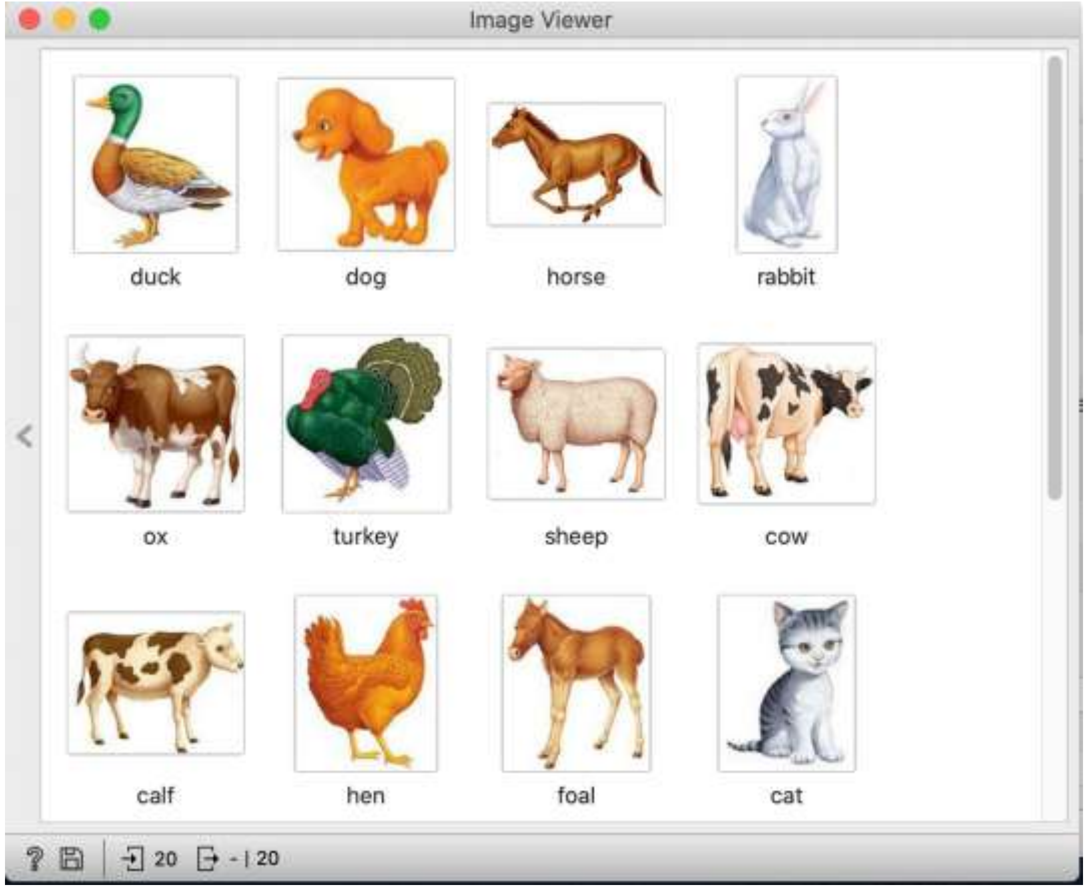
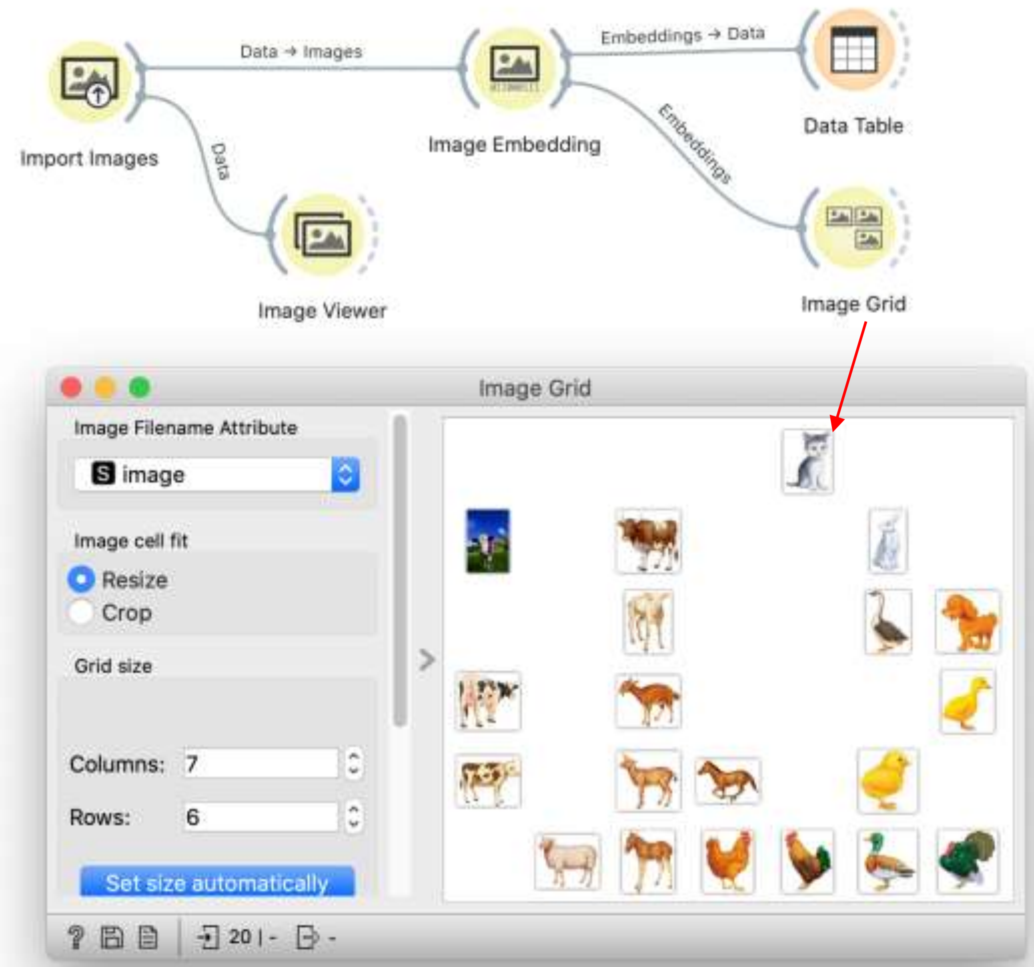


Image Embedding widget convert images to a vectors of numbers.

Analyzing data deep learning models

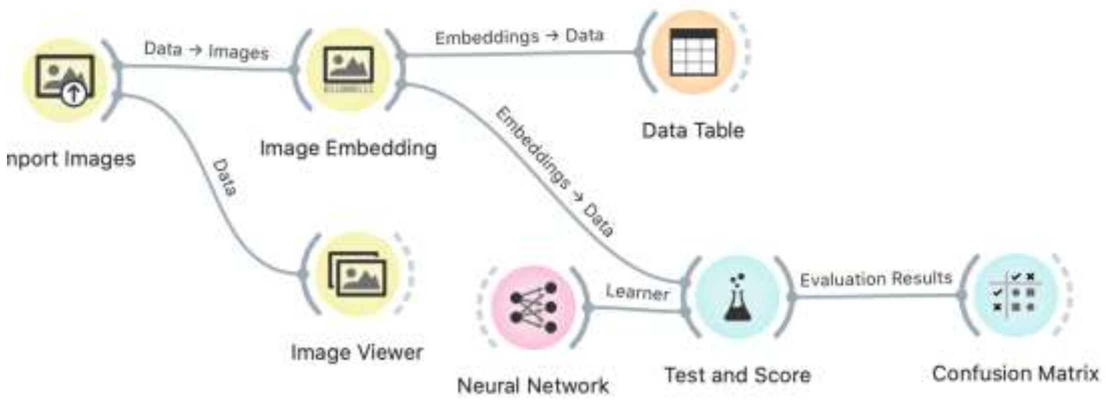
- Image Grid widget display images from a dataset in a similarity grid such that images with similar content are placed closer to each other.

- Image Grid widget can be used for image comparison, while looking for similarities or discrepancies between selected data instances.

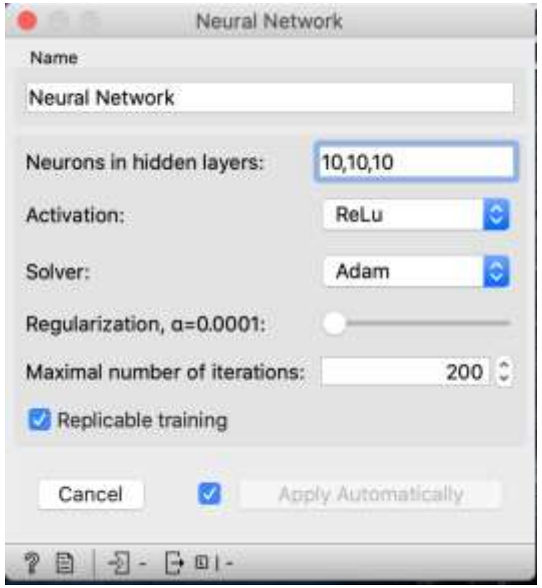


Analyzing data deep learning models

Example: Workflow for image analytics for Classification task on on animal image dataset.

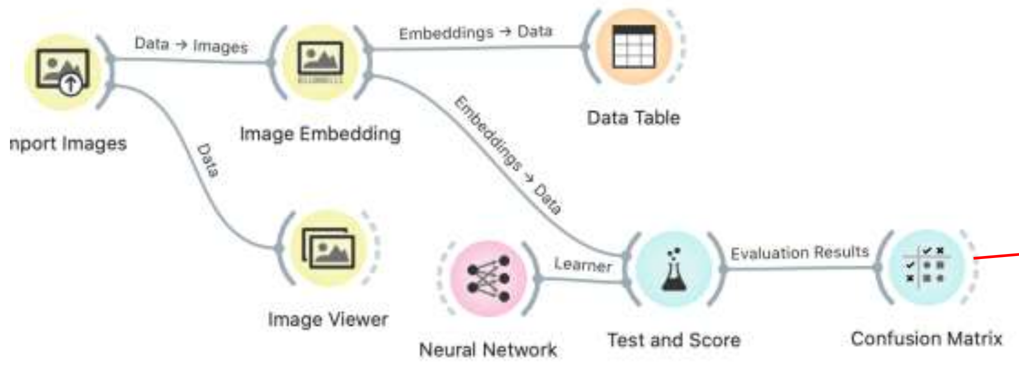


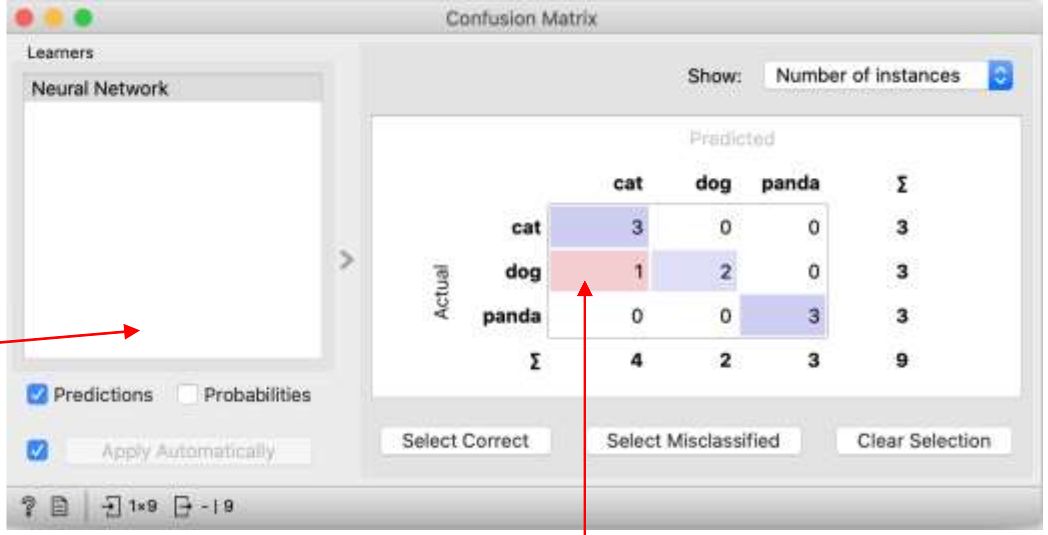
1. Pass Image Embedding to Test and Score.
2. Use Neural Network learner with 3 layers with 10 neurons each.
3. Input the learner into Test and Score.
4. Observe the predicted values.

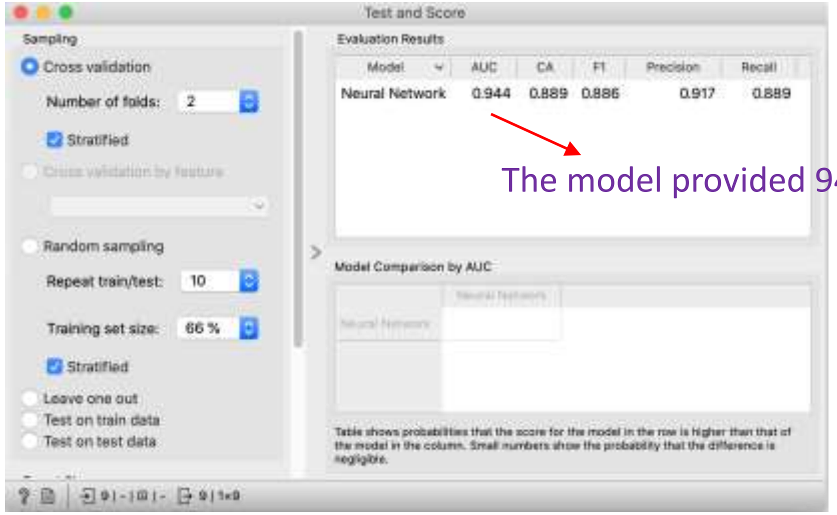


Analyzing data deep learning models

Example: Workflow for image analytics for Classification task on on animal image dataset.







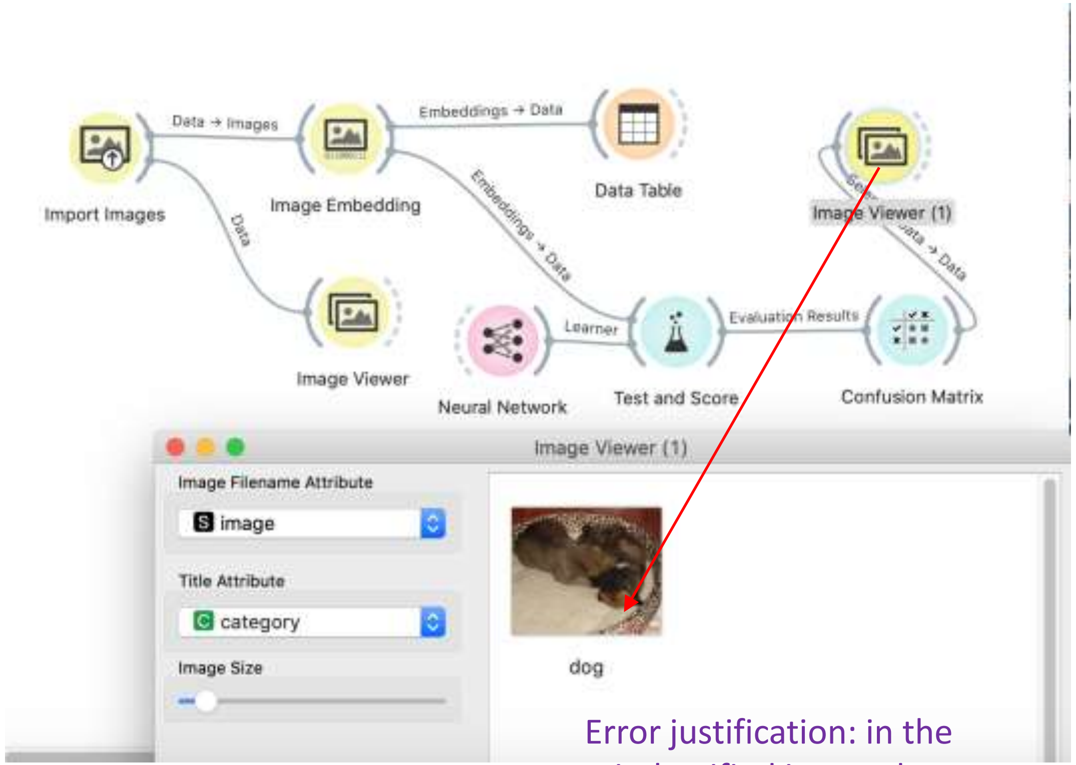
The model provided 94% accuracy

Misclassification: the model predicted the image as cat instead of dog.

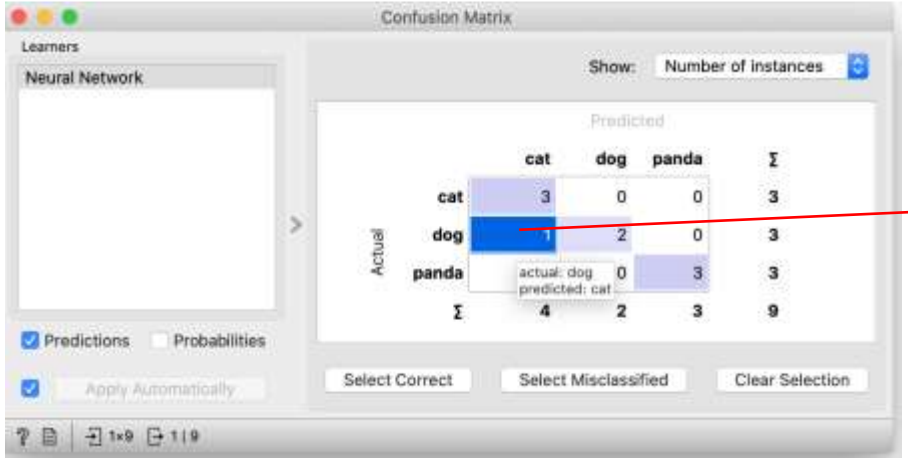
Analyzing data deep learning models

Example: Workflow for image analytics for Classification task on on animal image dataset.

Use image viewer to investigate the Misclassification example



Error justification: in the misclassified image dog looks like a cat.



Select the misclassified example

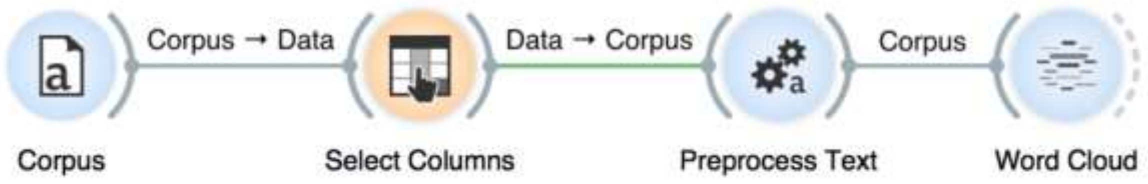
Text processing and classification

Textual data analysis

- Orange support textual data analysis through Text add-on.

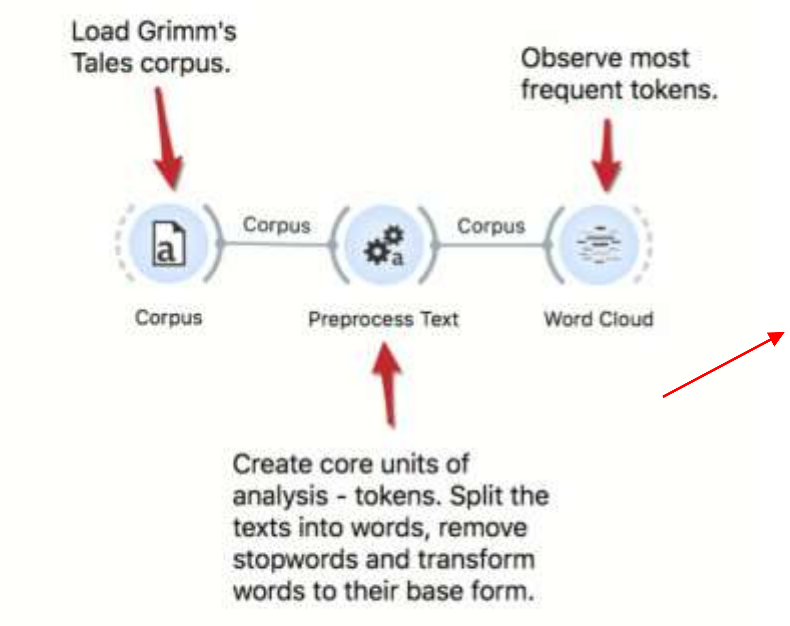
- Common text widgets:
  - Text preprocessing: preprocessing text (e.g., removing stopworks, lowercase, …).
  - Corpus viewer: to view corpus content.
  - Sentiment Analysis: enables basic sentiment analysis of corpora.
  - Topic Modelling: uncover the hidden thematic structure in a corpus.
  - Word Cloud: display word frequency.

- Typical textual data analysis workflows:



Textual data analysis

Typical text pre-process workflow example:



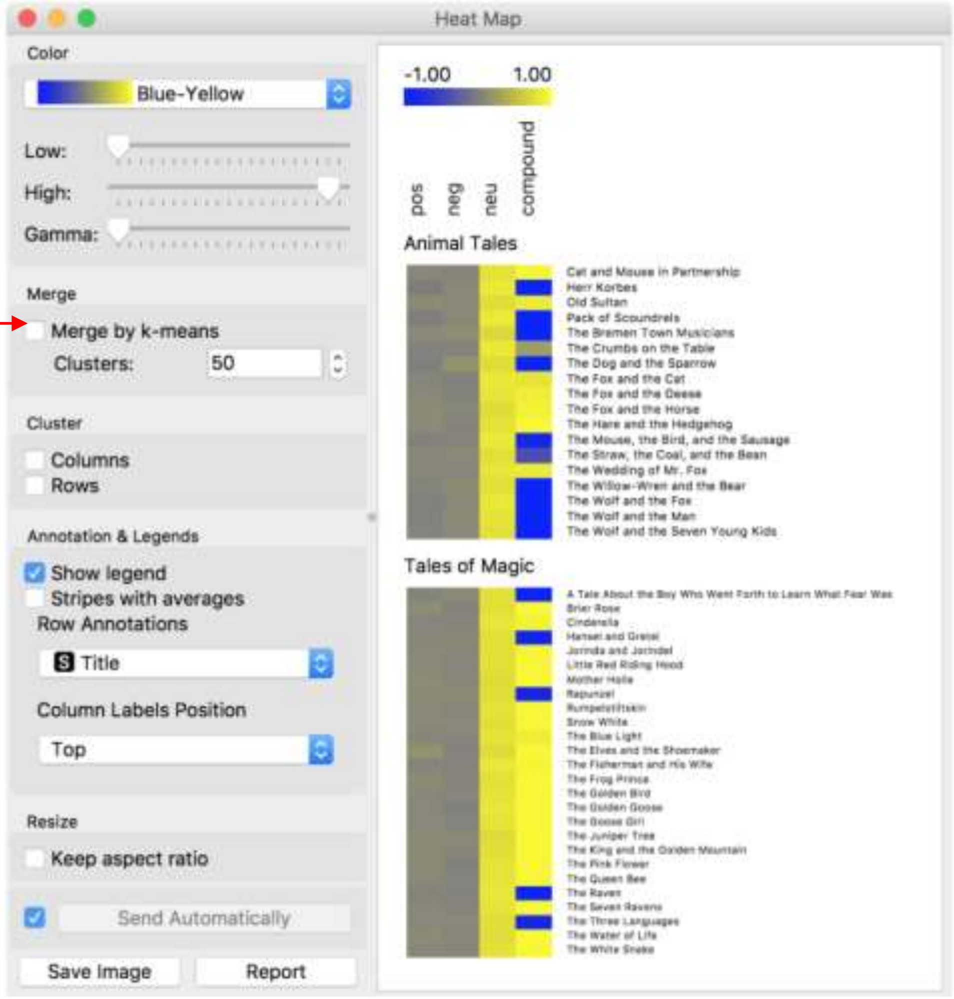
Results of preprocessing can be observe in a Word Cloud

This workflow uses simple reprocessing for creating tokens from documents:

1. it applies lowercase
2. splits text into words
3. it removes frequent stopwords.

Textual data analysis

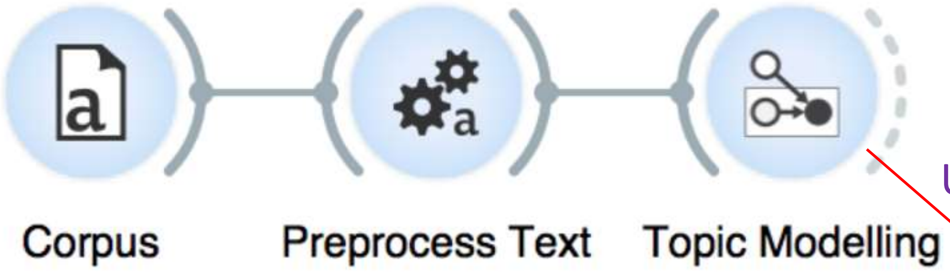
Typical Sentiment workflow example:



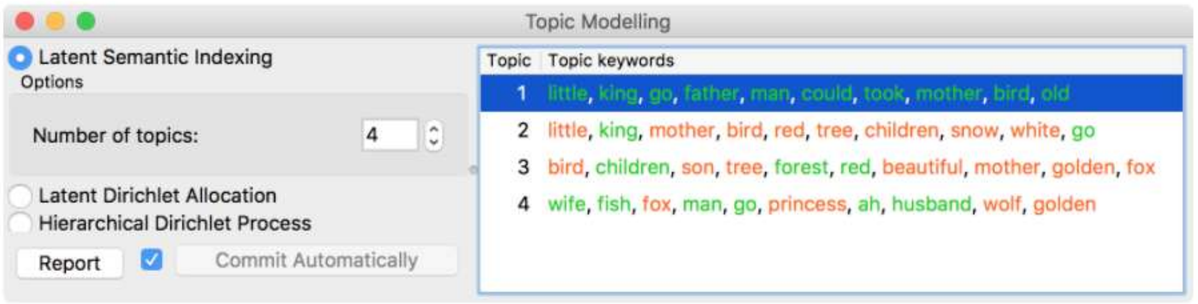
Yellow represent a high, positive score, while blue represent a low, negative score.

Textual data analysis

Typical Topic modelling workflow example:



Uncover latent topics in the data

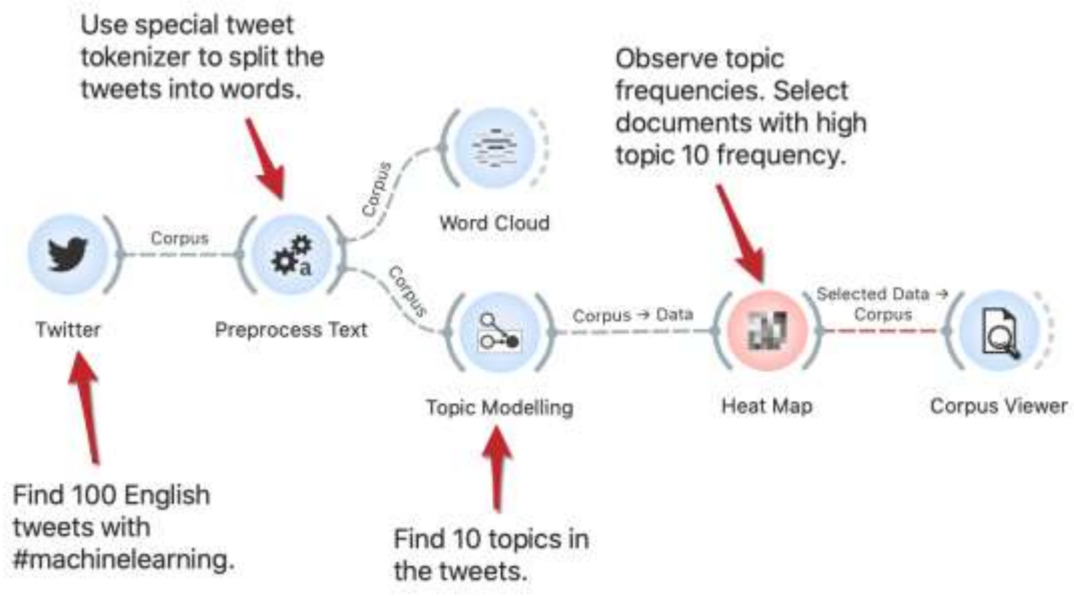


Topic Modelling, for example, colors words by their weights - positive weights are colored green and negative red.

Textual data analysis

- Tweets are a valuable source of information, for social scientists, marketing managers, linguists, economists, and so on.

Twitter Data Analysis workflow example:

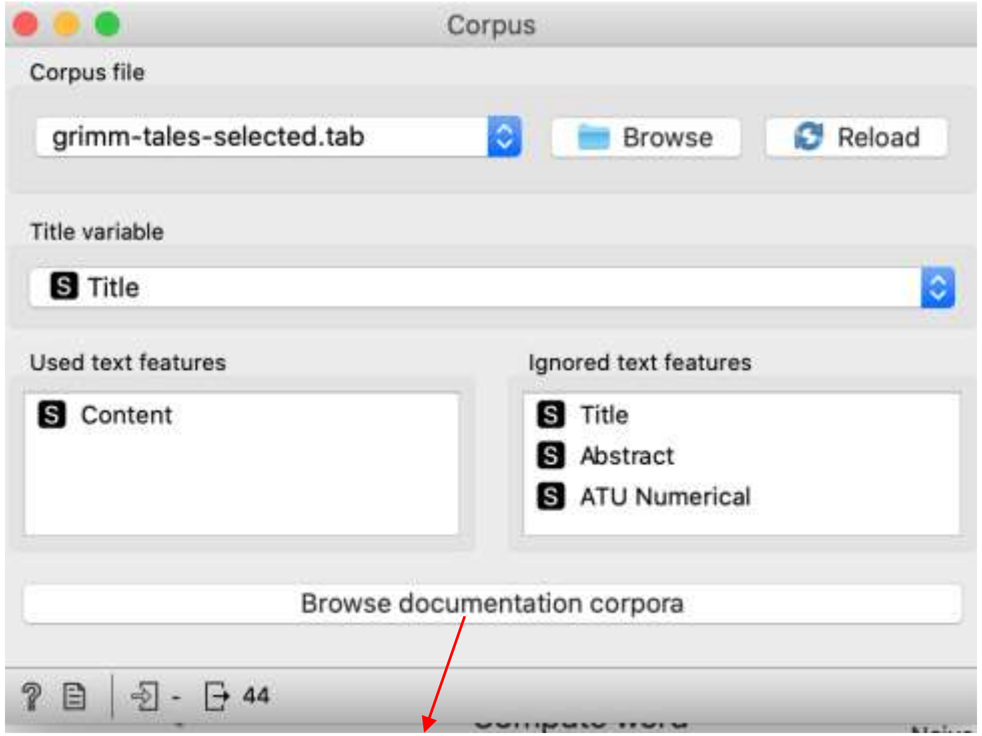


Text classification

- Predictive models can be used to classify documents by authorship, their type, sentiment and so on.

Text classification workflow example:

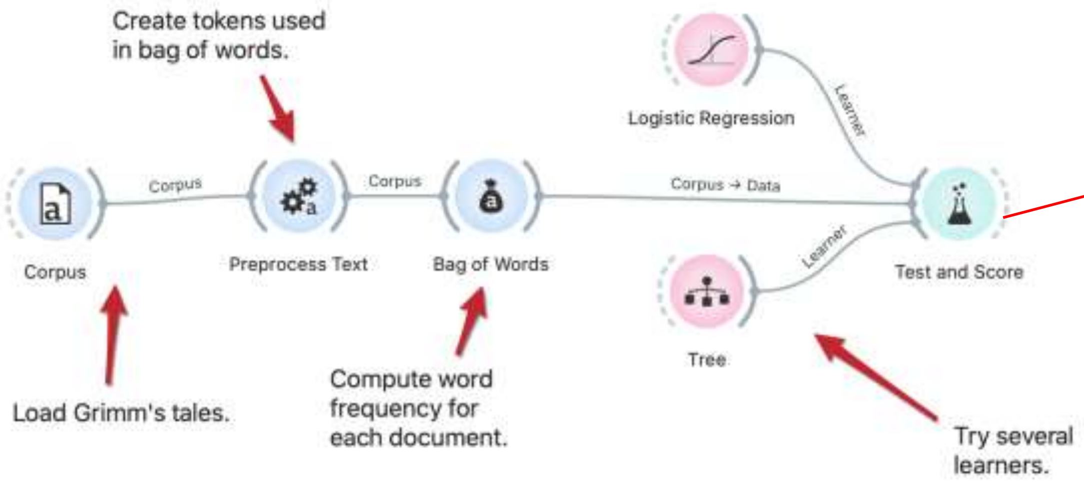
- Data: Grimm tales data
- Task: classify documents by their topic of the tale.
- Predication models: Logistic Regression and Decision Tree.

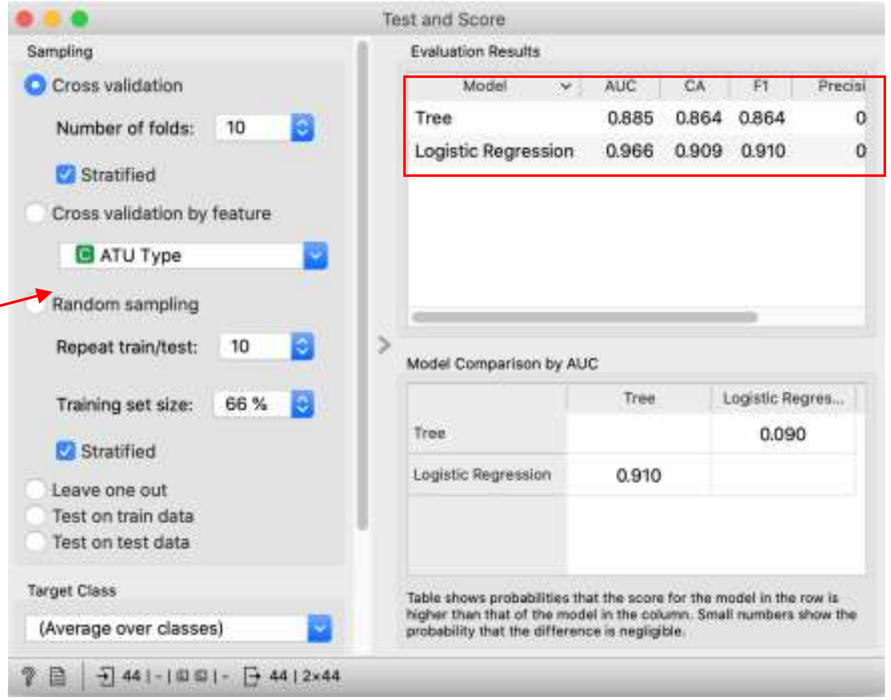


Load Grimm tales data

Text classification

Text classification workflow example:



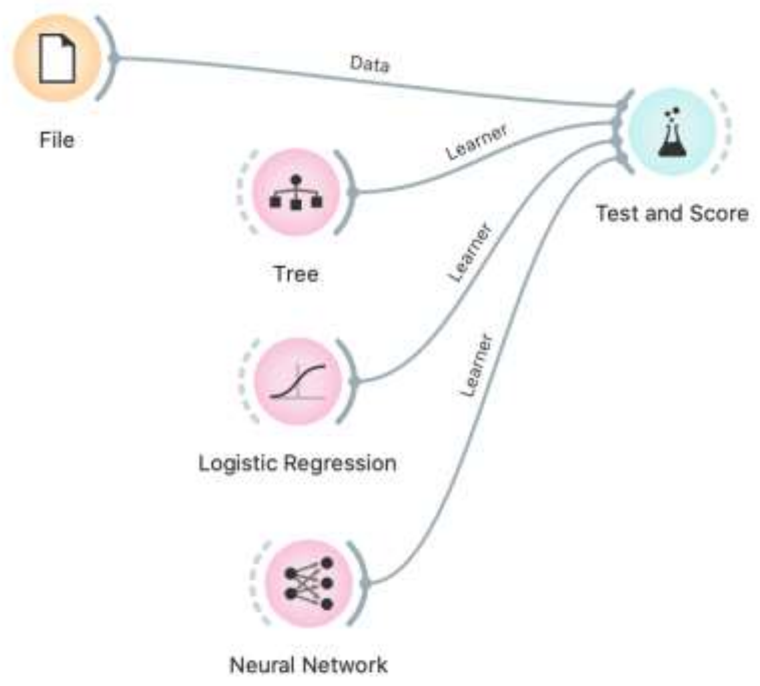


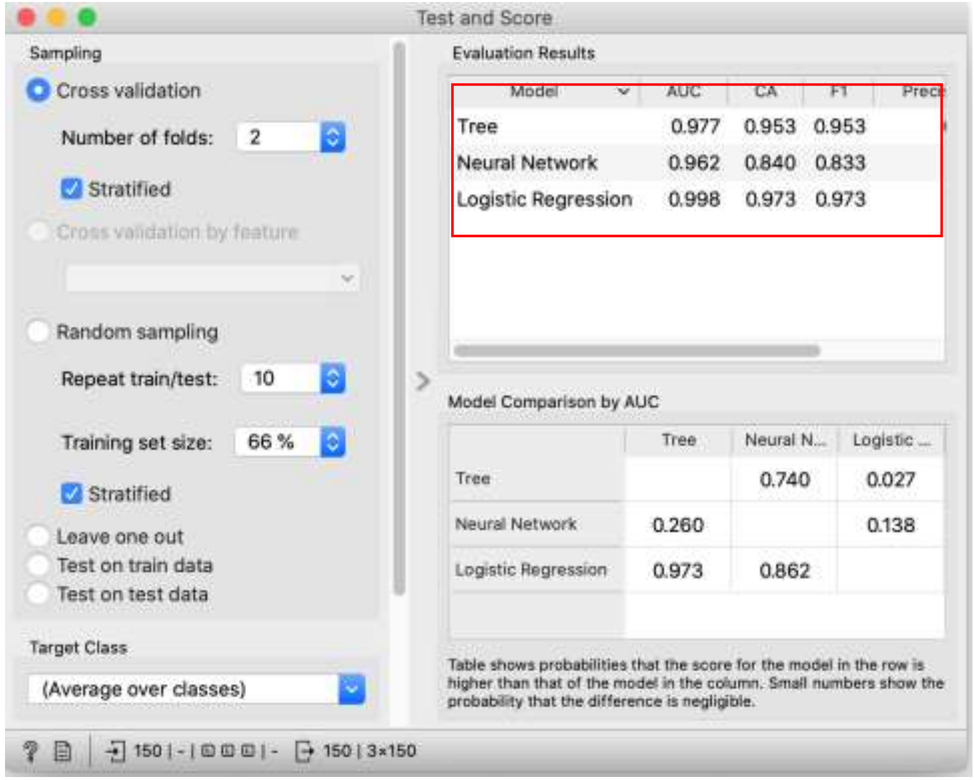
Given to tales of different class Logistic regression can correctly distinguish between them in over 90% of the cases. Better than Decision Tree!

Predicative model quality

Predicative model quality

Example: Explore the performance of different predictive models on iris dataset.





Logistic Regression outperformer other classifiers

Main Reference

- Orange Visual Programming Documentation (Release 3). Orange (2021). https://buildmedia.readthedocs.org/media/pdf/orange-visual-programming/latest/orange-visual-programming.pdf
- AJDA. (2017, August 4). *Text Analysis: New Features*. Orangedatamining.Com. https://orangedatamining.com/blog/2017/08/04/text-analysis-new-features/
- Zupan, D. (2018, May). *Introduction to Data Mining: Working notes for the hands-on course with Orange Data Mining*. University of Ljubljana. https://file.biolab.si/notes/2018-05-intro-to-datamining-notes.pdf
- Foong, N. W. (2019, August 7). *Data Science Made Easy: Interactive Data Visualization using Orange* [Post]. Medium, Towardsdatascience.com. https://towardsdatascience.com/data-science-made-easy-interactive-data-visualization-using-orange-de8d5f6b7f2b
- Analytics Vidhya, A. (2017, September 7). *Building Machine Learning Model is fun using Orange*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2017/09/building-machine-learning-model-fun-using-orange/
- Orange Data Mining. Workflows, accessed 10 August 2021 https://orangedatamining.com/workflows/
- Foong, N. W. (2019b, August 29). Data Science Made Easy: Image Analytics using Orange. Medium, Towardsdatascience.com. https://towardsdatascience.com/data-science-made-easy-image-analytics-using-orange-ad4af375ca7a

This Presentation is mainly dependent on the above recourses.

Week self-review exercises

- Download orange and understand its workflow:
  - https://orangedatamining.com/getting-started/
  - https://orangedatamining.com/workflows/

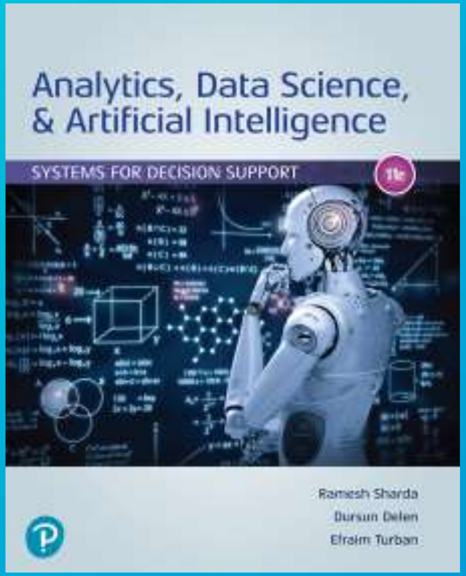
- Hands-on practice on using orange for data analysis, visualisation, developing predictive models and textual data analysis: https://orangedatamining.com/blog/

Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

College of Computing and Informatics



Week 2

Chapter 1: Overview of Business Intelligence, Analytics, Data Science, and Artificial Intelligence: Systems for Decision Support

Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Understand the need for computerized support of managerial decision making
2. Understand the development of systems for providing decision-making support
3. Recognize the evolution of such computerized support to the current state of analytics/data science and artificial intelligence
4. Describe the business intelligence (BI) methodology and concepts
5. Understand the different types of analytics and review selected applications
6. Understand the basic concepts of artificial intelligence (AI) and see selected applications
7. Understand the analytics ecosystem to identify various key players and career opportunities

Required Reading

- **Chapter 1**: “An overview of Decision Support system, Business Intelligence, Analysis, and AI” from “*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*”.

Recommended Video

- Understanding Decision Support Systems
  https://www.youtube.com/watch?v=k31CSoO7wC4
- What is Business Intelligence (BI)?
  https://www.youtube.com/watch?v=hDJdkcdG1iA

1.2 Changing Business Environments and Evolving Needs for Decision Support and Analytics

- The Decision Making Process
- The Influence of the External and Internal Environments on the Process
- Data and Its Analysis in Decision Making
- Technologies for Data Analysis and Decision Support

The Decision Making Process

- The business world is full of uncertainties and rapid changes.
- Thus, the decision making process can determine the success or failure of an organization, and how well it performs.
- In the past decision making was based on creativity, instinct, and experience.
- Now, decision making is more grounded in scientific approach, and utilizes systematic quantitative methods.

The Decision Making Process (cont.)

Managers usually undergo the decision making process using the following steps:

1. Understand the decision you have to make
2. Collect all the information *(Define the problem)*
3. Identify the alternatives *(Construct a model & Identify possible solutions)*
4. Evaluate the pros and cons
5. Select the best alternative *(Compare, choose and recommend the best solution)*
6. Make the decision
7. Evaluate the impact of your decision

The Influence of External & Internal Environments

- Predicting the consequences and the future of any given decisions is complex. This is due to the uncertainty that can arise from multiple factors including:
    1. Political Factors (E.g. government policies, political instability)
    2. Economic Factors (E.g. competition, changing demand)
    3. Sociological and psychological factors
    4. Environment Factors

- Because of these constant changes, the trial-and-error approach to management is unreliable and unsustainable. Managers must begin to use the new tools and techniques of their fields.

Data and Its Analysis in Decision Making

- The amount of data doubles every two years.

- Decision making process requires an organization to collect and analyze vast stores of data.

- Computer applications have moved from

  transaction-processing and monitoring activities => to
  problem analysis and solution applications.

- These activities are done with cloud-based technologies, mostly accessed through mobile devices.

Data and Its Analysis in Decision Making (cont.)

- The foundation of modern management lies in

    1. analytics and BI tools such as data warehousing, data mining, online analytical processing (OLAP), dashboards, and

    2. the use of cloud-based systems for decision support.

Technologies for Data Analysis and Decision Support

- The following developments have contributed to the growth of decision support and analytics technologies:

  1. Group communication and collaboration
  2. Improved data management
  3. Managing giant data warehouses & Big Data
  4. Analytical Support
  5. Overcoming cognitive limits in processing and storing information
  6. Knowledge Management
  7. Anywhere, anytime support
  8. Innovation and artificial intelligence

1.3 Decision-Making Processes and Computer Decision Support Framework

- Simon’s Process: Intelligence, Design, Choice, and Implementation Phases
- The Classical Decision Support System Framework
- DSS Application
- Characteristics of DSS
- Components of a Decision Support System

Simon's Process: Intelligence, Design, and Choice

- Decision-making process phases:

1. **Intelligence:** The decision maker examines reality, identifies and defines the problem.
2. **Design:** A model representing the system is constructed by and then validated.
    1. making assumptions to simplify reality and
    2. identifying relationships between variables.
3. **Choice:** Selection of a proposed solution to the model . This solution is tested to determine its viability.
4. **Implementation:** Successful implementation results in solving the real problem. Failure leads to a return to an earlier phase of the process.

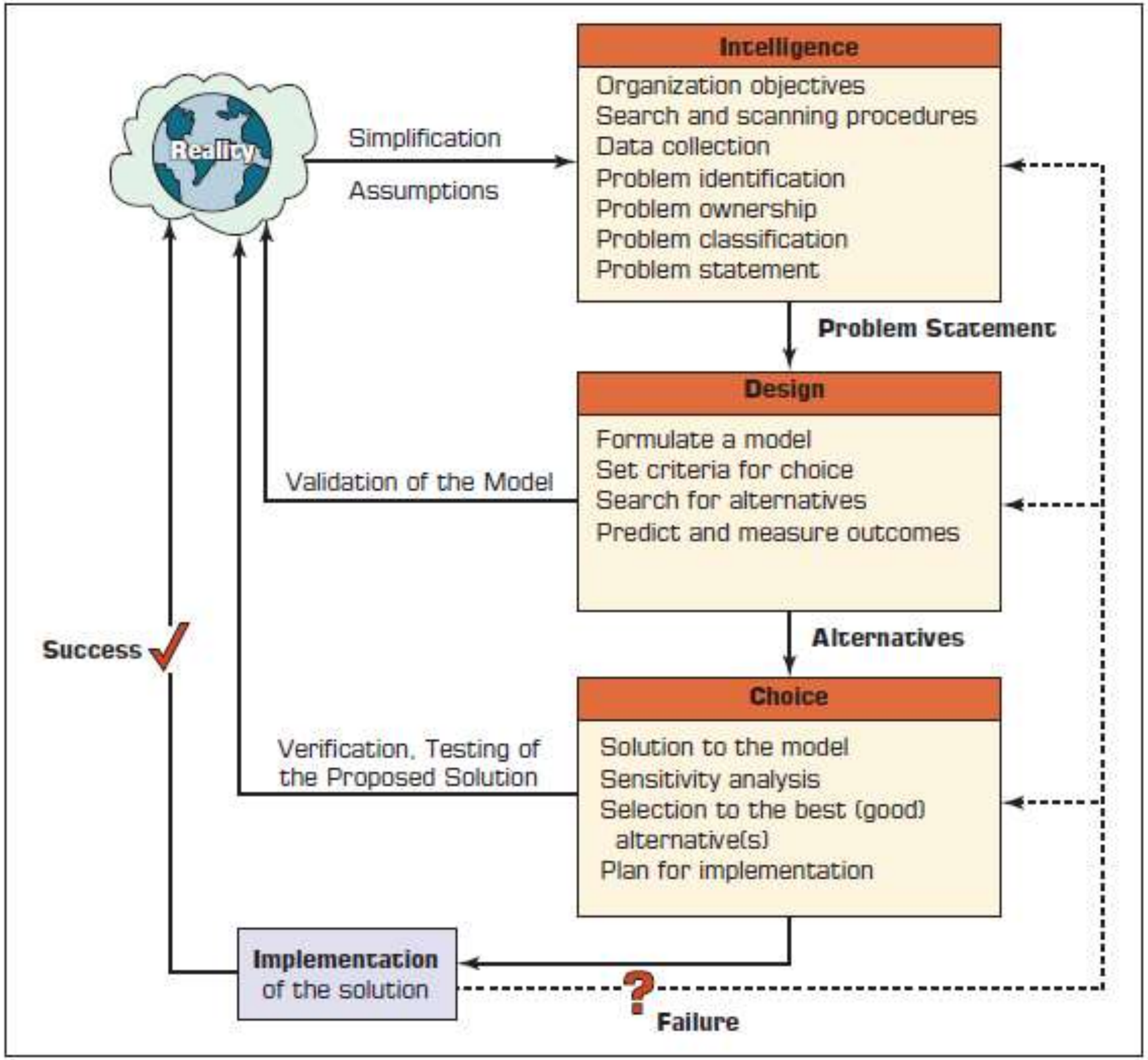


FIGURE 1.1 The Decision-Making/Modeling Process.

The Classical Decision Support System Framework

- The **framework** for computerized decision support can be represented in a 3x3 **matrix** with two dimensions:

1. **Degrees of structure**:
   - **Structured:** Routine/repetitive problems for which standard solution methods exist
   - **Semi structured:** Fall between structured and unstructured problems
   - **Unstructured:** Complex problems where there are no clear, standard solution methods
2. **Types of controls:**
   - **Strategic planning:** Defining long-range goals/policies for resource allocation
   - **Management control:** The efficient use of resources in the accomplishment of organizational goals
   - **Operational control:** the efficient and effective execution of specific tasks

Type of Decision	Type of Control: Operational Control	Managerial Control	Strategic Planning
Structured	1 Monitoring accounts receivable Monitoring accounts payable Placing order entries	2 Analyzing budget Forecasting short-term Reporting on personnel Making or buying	3 Managing finances Monitoring investment portfolio Locating warehouse Monitoring distribution systems
Semistructured	4 Scheduling production Controlling inventory	5 Evaluating credit Preparing budget Laying out plant Scheduling project Designing reward system Categorizing inventory	6 Building a new plant Planning mergers and acquisitions Planning new products Planning compensation Providing quality assurance Establishing human resources policies Planning inventory
Unstructured	7 Buying software Approving loans Operating a help desk Selecting a cover for a magazine	8 Negotiating Recruiting an executive Buying hardware Lobbying	9 Planning research and development Developing new technologies Planning social responsibility

FIGURE 1.2 Decision Support Frameworks.

DSS Application

- **Key difference between DSS and BI applications:**

1. **Business intelligence (BI) systems** monitor situations and identify problems and/or opportunities using analytic methods.

2. **DSS** is a methodology for supporting decision making:-

   - It uses an interactive, adaptable **computer-based information system** (CBIS) developed for supporting the solution to a specific unstructured management problem.

   - It uses **data**, provides a user-friendly **interface**, and incorporates the **decision maker's insights**.

   - It includes **models** that developed through an interactive and iterative process.

Characteristics of DSS

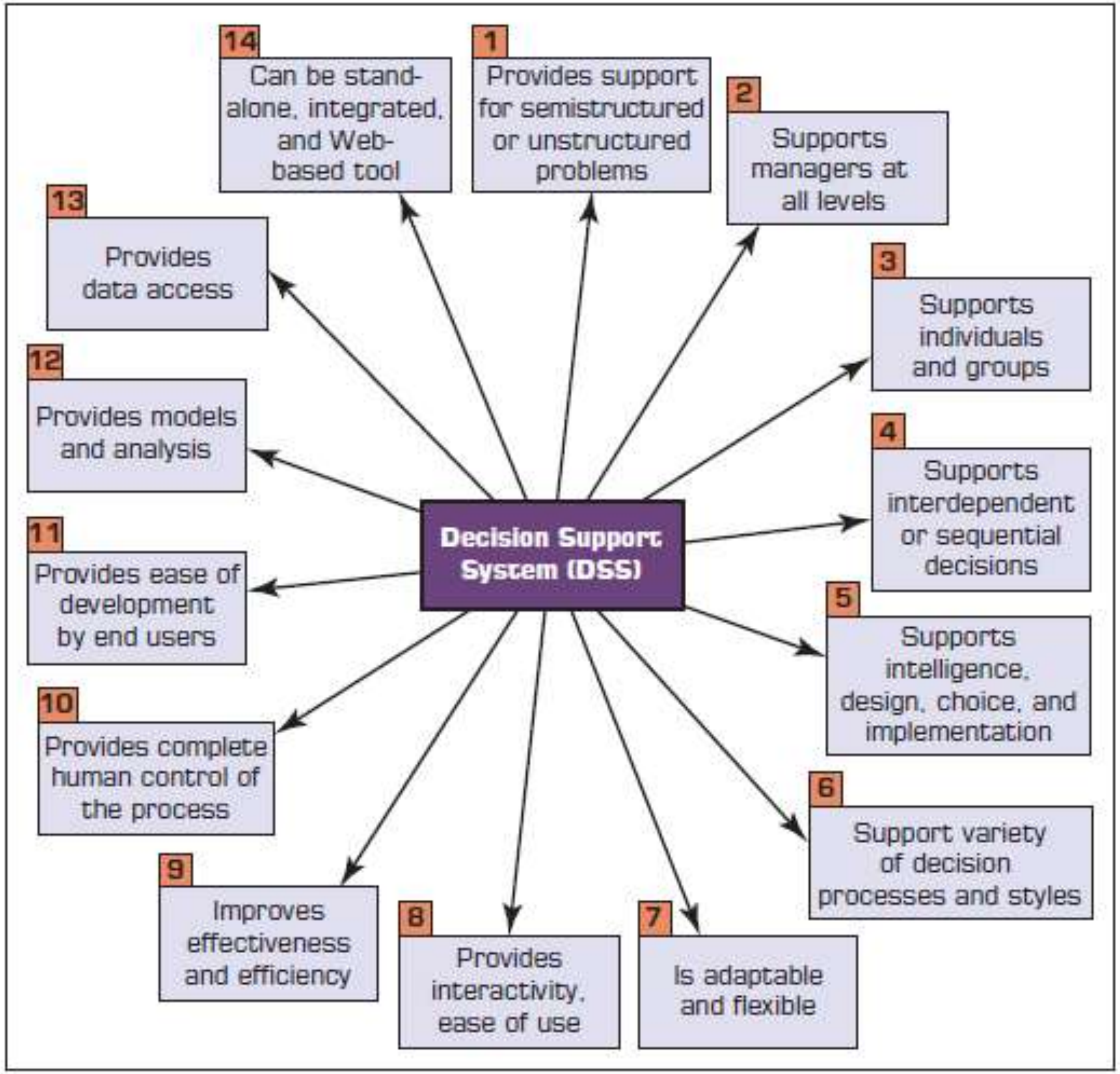


FIGURE 1.3 Key Characteristics and Capabilities of DSS.

Components of a Decision Support System

- A DSS can be composed of:

1. Data management subsystem
2. Model management subsystem
3. User interface subsystem
4. Knowledge-based management subsystem

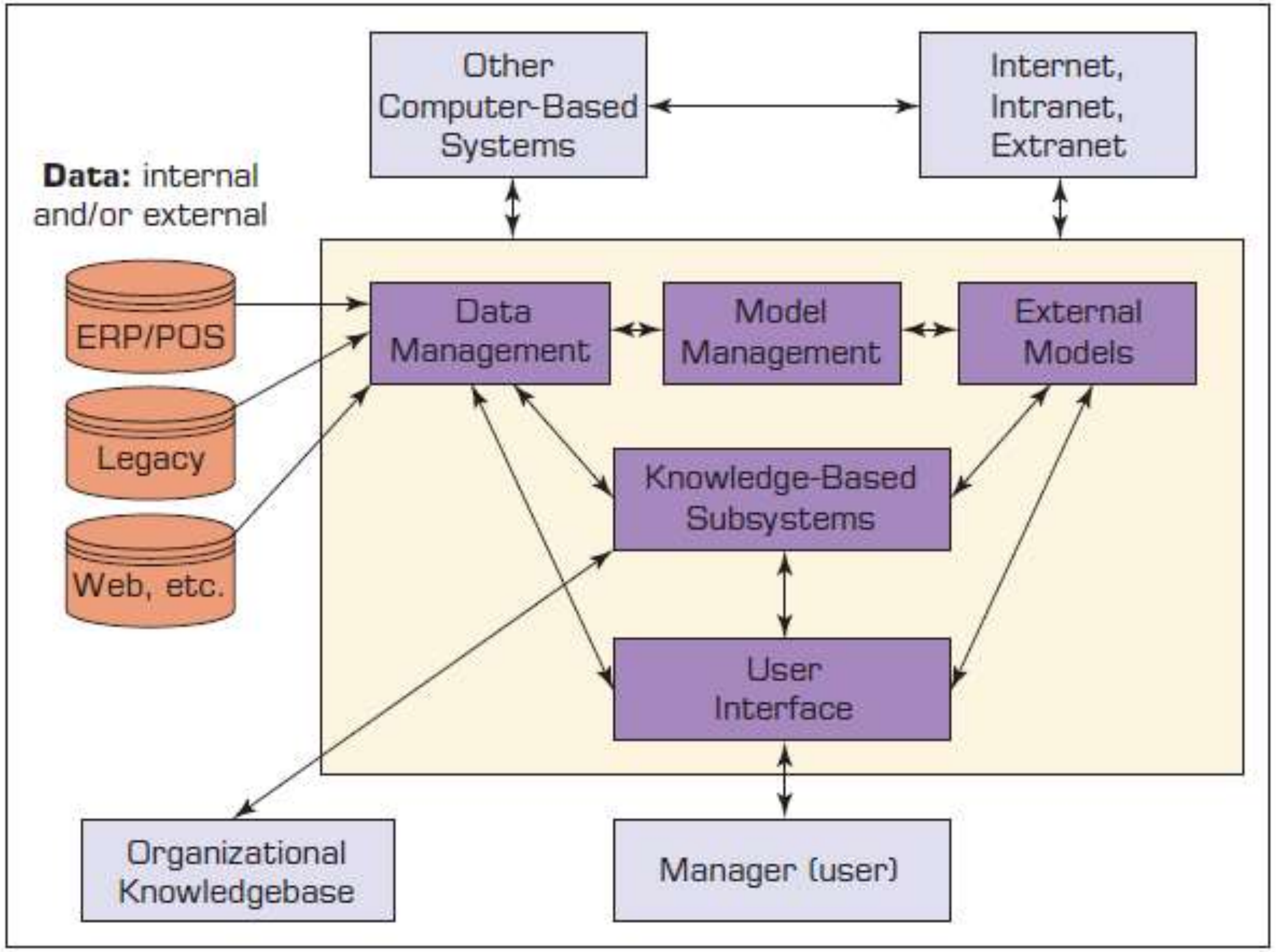


FIGURE 1.4 Schematic View of DSS.

1. Data Management Subsystem

- The data management subsystem is composed of the following elements:
    1. DSS database
    2. Database management system
    3. Data directory
    4. Query facility

- It can be interconnected with the corporate data warehouse (a repository for corporate relevant decision-making data)

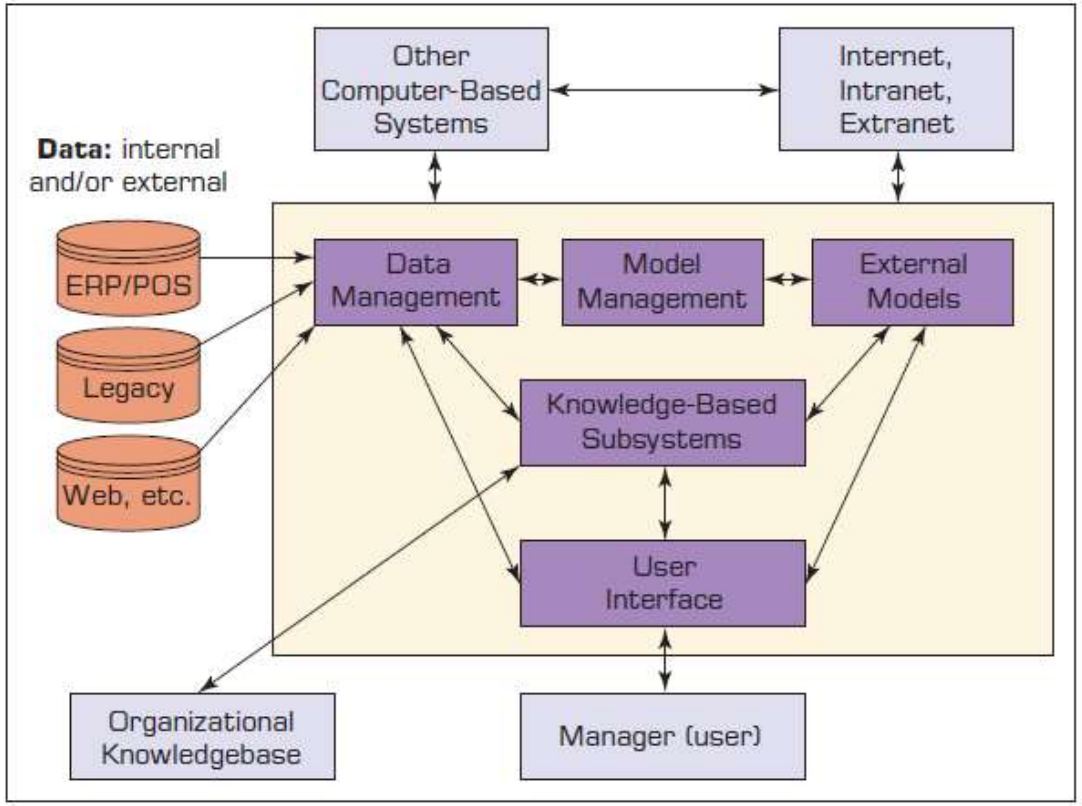


FIGURE 1.4 Schematic View of DSS.

2. Model Management Subsystem

- The model management subsystem is a component that includes quantitative models that provide the
  - system's analytical capabilities and
  - appropriate software management.

- The model management subsystem of a DSS is composed of the following elements:
  1. Model base
  2. MBMS (Model Base Management System)
  3. Modeling language
  4. Model directory
  5. Model execution, integration, and command processor

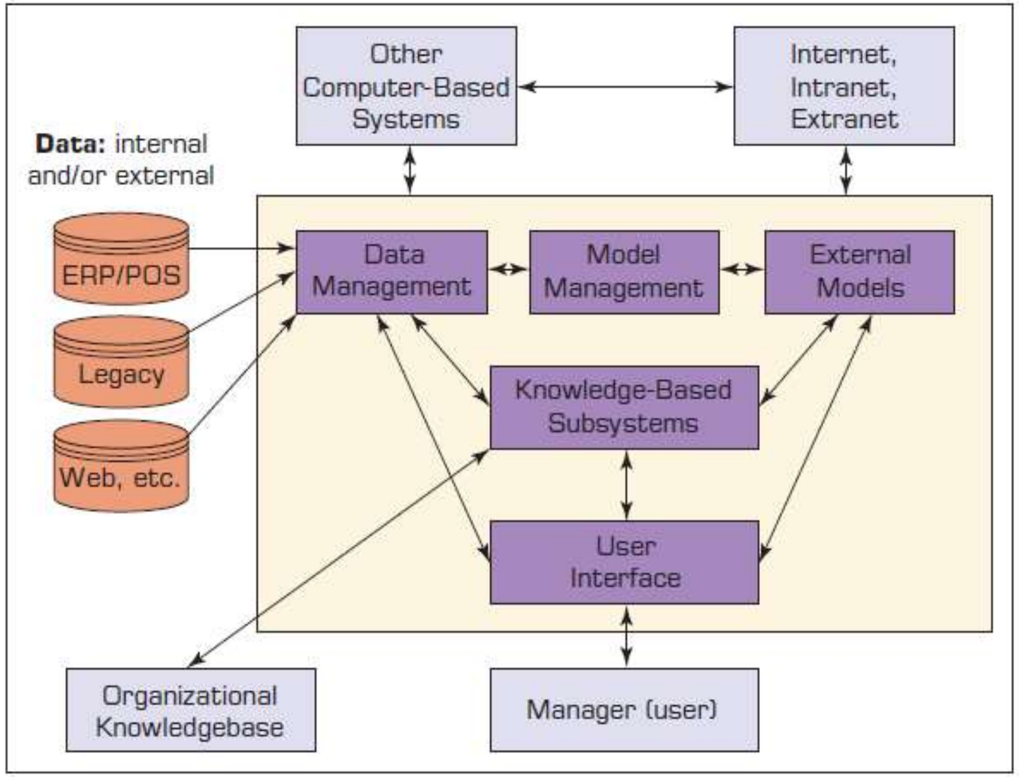


FIGURE 1.4 Schematic View of DSS.

3. The User Interface Subsystem

- The user is considered part of the system.

- The user communicates with and commands the DSS through the user interface subsystem.

- The Web browser has been recognized as an effective DSS GUI because it is
  - flexible,
  - user-friendly, and a
  - gateway to almost all sources of necessary information and data.

- Web browsers have led to the development of portals and dashboards (front end of many DSS).

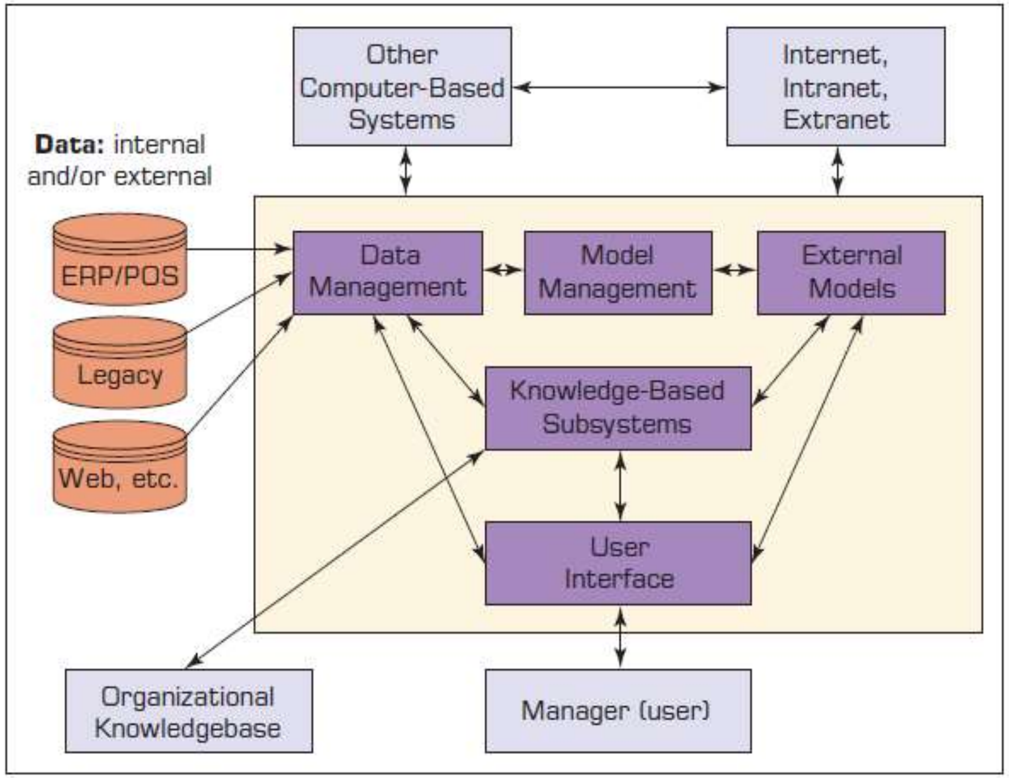


FIGURE 1.4 Schematic View of DSS.

4. The Knowledge-Based Management Subsystem

- The knowledge-based management subsystem provides intelligence
  - to augment the decision maker's own intelligence
  - to help understand a user's query and provide a consistent answer.

- Interconnected with the organization's knowledge repository or connected to thousands of external knowledge sources.

- User interface developments are closely tied to the major new advances in their knowledge-based systems.

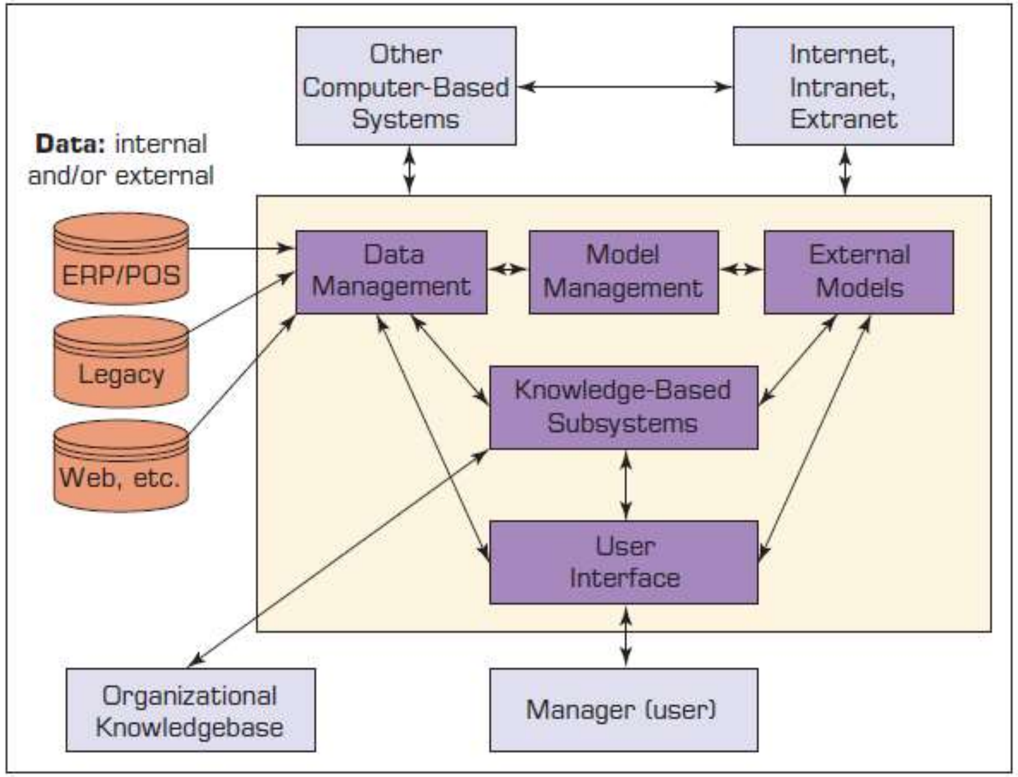


FIGURE 1.4 Schematic View of DSS.

1.4 Evolution of Computerized Decision Support to Business Intelligence/Analytics/Data Science

- Evolution Of Computerized Decision Support
- Framework for Business Intelligence
- Architecture of BI
- The Origins and Drivers of BI
- Data Warehouse as a Foundation for Business Intelligence
- Transaction Processing versus Analytic Processing
- A Multimedia Exercise in Business Intelligence

Evolution Of Computerized Decision Support

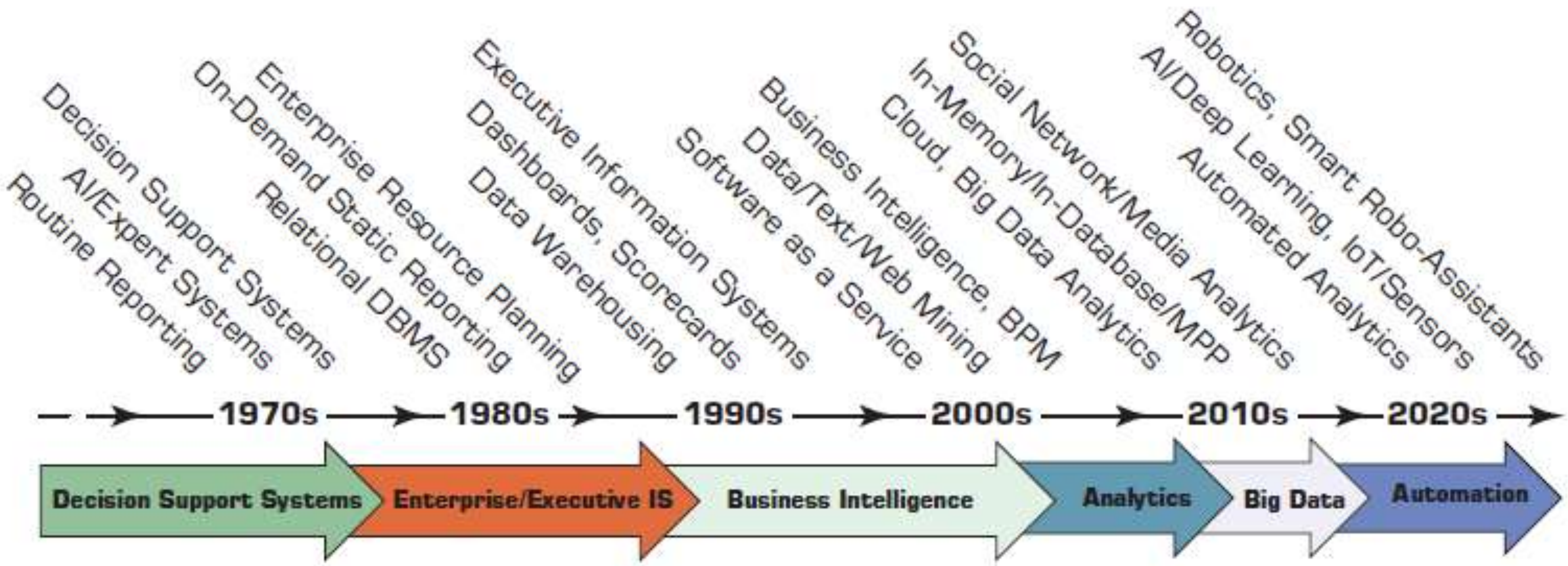


FIGURE 1.5 Evolution of Decision Support, Business Intelligence, Analytics, and AI.

Evolution Of Computerized Decision Support (cont.)

1. **Management Information Systems (MIS):** a variety of reports used to understand and address the changing business needs and challenges.

2. **Decision Support Systems (DSS):** combination of individuals intellectual resources with the capabilities of the computer to improve the quality of decisions.
   - DSS designed and developed specifically for executives and their decision-making needs

3. **Executive Information Systems (EIS):** the need for more versatile reporting leads to the development of EISs;
   - These systems were designed as graphical dashboards that allow decision makers to keep track of the key performance indicators.

A Framework for Business Intelligence (BI)

- **BI** is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies.
- **BI process** is based on the transformation of data -> information -> decisions -> actions
- BI's major **objective** is to enable
  1. interactive access to data (in real time),
  2. data manipulation,
  3. appropriate analyses.
- By analyzing **historical** and **current** data, **situations**, and **performances**, decision makers able to make more informed and better decisions.

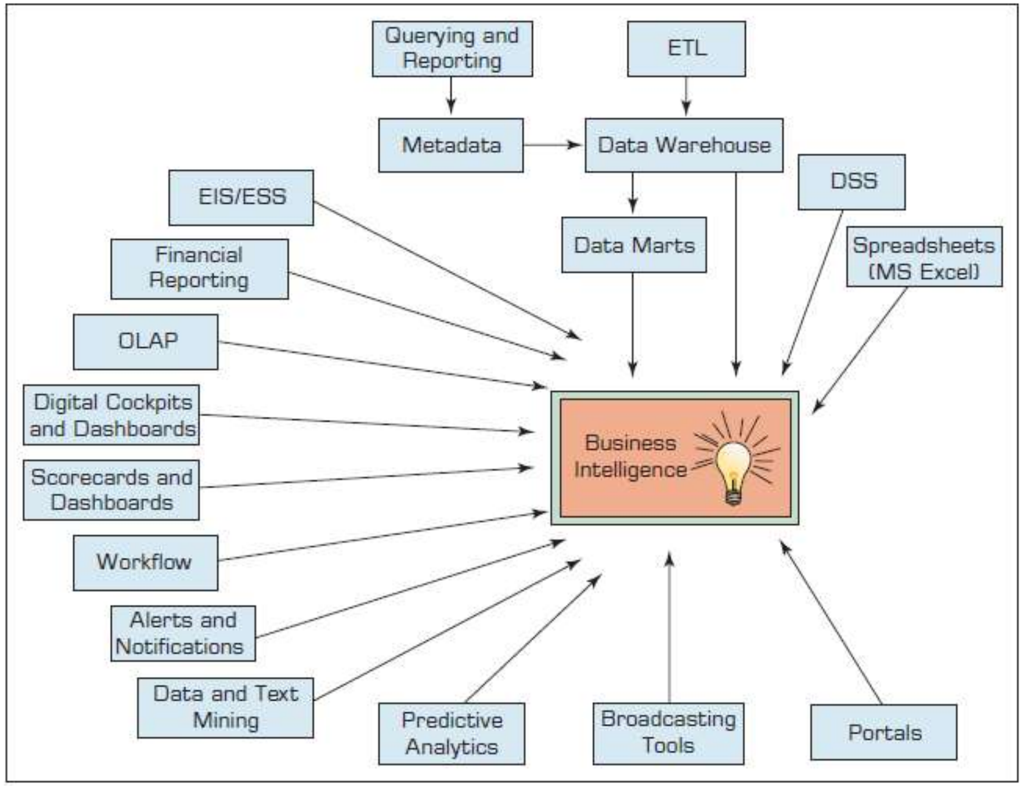
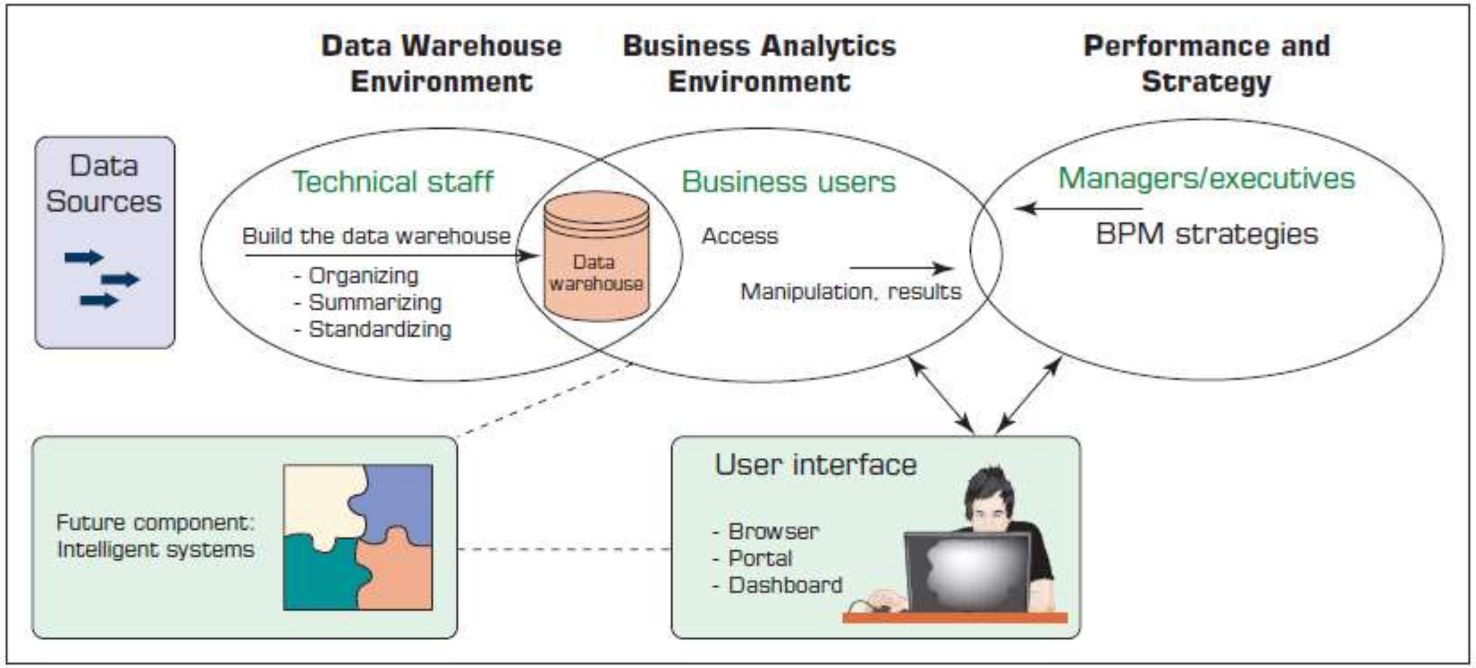


FIGURE 1.6 Evolution of Business Intelligence (BI).

Architecture of BI

- BI has four major components:

1. **DW** (with its source data)
2. **Business analytics** (a collection of tools for manipulating, mining, and analyzing the data in the DW)
3. **BPM** (for monitoring & analyzing performance)
4. **User interface**



The Origins and Drivers of BI

- Business cycle times are extremely compressed; faster, and more informed.
- Managers need the right information at the right time and in the right place (core of the modern approaches to BI).
- Organizations are being driven to capture, understand, and harness their data to support decision making and improve business operations.
- Legislation/regulations require leaders to document their business processes and sign off on the validity of the information they report to stakeholders.

Data Warehouse (DW) as a Foundation for BI

- BI systems rely on DWs as the information source for creating insight and supporting managerial decisions.
- **DW** is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process.
- The three main types of data warehouses are:
  - **Data Marts (DM):** a subset of a DW, typically consisting of a single subject area
  - **Operational Data Stores (ODS):** provides a form of customer information file whose contents are updated throughout the course of business operations.
  - **Enterprise Data Warehouses (EDW):** a large-scale data warehouse used across the enterprise for decision support.

Data Warehouse (DW) as a Foundation for BI (cont.)

- Data from many different sources can be extracted, transformed, and loaded into a DW for further access and analytics for decision support.

- Data is structured to be available in a form ready for analytical processing activities:
  - OLAP,
  - data mining,
  - querying,
  - reporting,
  - decision support applications

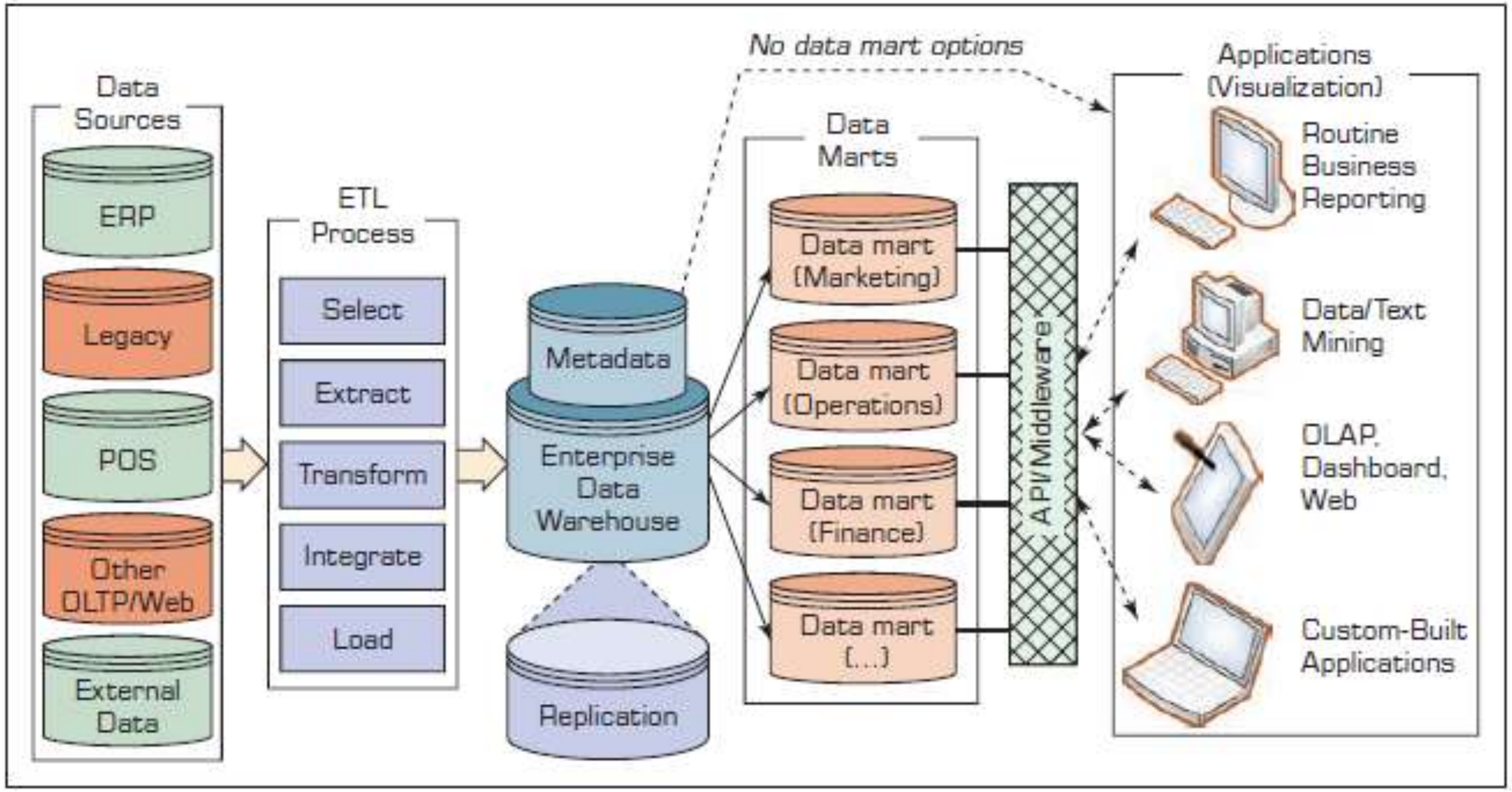


FIGURE 1.8 Data Warehouse Framework and Views.

Transaction Processing versus Analytic Processing

- **Transaction processing systems** are constantly involved in handling updates to what may be known as operational databases.

- **OLTP** (OnLine Transaction Processing) systems handle a company's routine ongoing business, where the computer responds immediately to user requests.

- The OLTP system is efficient for transaction processing, but inefficient for end-user ad hoc reports, queries, and analysis.

- In contrast, a **DW** is typically a distinct system that provides storage for data that will be used for analysis by **OLAP** (OnLine Analytical Processing) systems.

Multimedia Exercise in Business Intelligence

- The **fundamental reasons** for investing in BI must be aligned with the company's business strategy.
- **BI** improve company business processes and transforming it to be more data driven.
- **BI tools** are sometimes integrated among themselves, resulting in six key trends:
    1. Big Data
    2. Focus on customer experience as opposed to just operational efficiency.
    3. Mobile and even newer user interfaces—visual, voice, mobile.
    4. Predictive and prescriptive analytics, machine learning, artificial intelligence.
    5. Migration to cloud.
    6. Much greater focus on security and privacy protection.

1.5 Analytics Overview

- Overview
- Descriptive, Predictive, Prescriptive Analytics
- Big Data

Analytics Overview

- **Analytics** is the process of developing actionable/recommendations for actions based on insights generated from historical data.
- To solve real problems analytics represents the combination of
    - computer technology +
    - management science techniques +
    - statistics.

Descriptive, Predictive, Prescriptive Analytics

- Three types of analytics:
  1. **Descriptive**
  2. **Predictive**
  3. **Prescriptive**

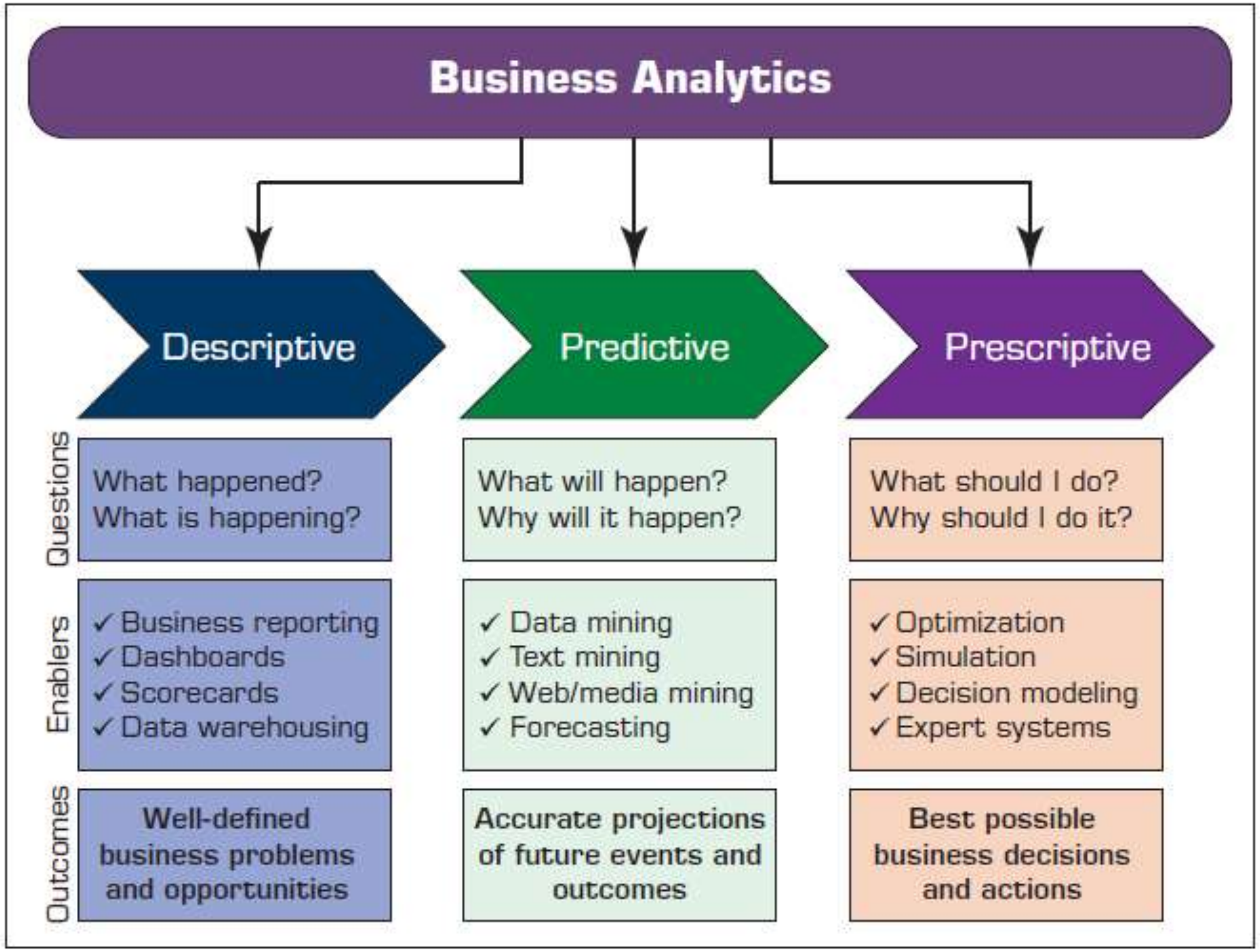


FIGURE 1.9 Three Types of Analytics.

Big Data

- **Big Data** refers to data that cannot be stored in a single storage unit.

- Different **forms**:
  - structured
  - unstructured

- **Characteristics** of big data:
  - volume
  - velocity
  - variety

- These are evolving quickly to encompass stream analytics, IoT, cloud computing, and deep learning– enabled AI.

Big Data (cont.)

- There are two aspects to managing data on this scale:

1. **Storing:**
   - an extremely expensive storage solution **vs.** store data in chunks on different machines connected by a network.
   - Hadoop Distributed File System (HDFS): store a copy or two of this chunk in different locations on the network, both logically and physically.

2. **Processing:**
   - To process vast amounts of data **computation done by one powerful computer vs.** processing data sets with a parallel distributed algorithm on a cluster
   - MapReduce programming paradigm

1.6 Analytics Examples in Selected Domains

- Sports Analytics
- Sports Analytics Example 1: The Business Office
- Sports Analytics Example 2: The Coach
- Healthcare Analytics Example 1: Humana
- Retail Analytics
- Image Analytics

Sports Analytics

- **Sport analytics** focuses on gathering data about athletes and teams to create insights that improve sports decisions, such as
  - deciding which players to recruit,
  - how much to pay them,
  - who to play, etc.

- For teams, it involves business decisions such as
  - ticket pricing
  - roster /schedule decisions,
  - analysis of each competitor's strengths and weaknesses

- Sport is a big business.

Sports Analytics Example 1: The Business Office

- Dave works as a business analyst for a major pro baseball team, focusing on revenue.
- His responsibility include analyzing ticket sales; why season ticket holders renew (or do not renew) their tickets as well as what factors drive last-minute individual seat ticket purchases, and ticket pricing.
- Some of the analytical techniques Dave uses include simple statistics on behavior such as overall attendance and answers to survey questions about likelihood to purchase again.
- Dave uses dynamic pricing. This is a rich research area for many sports teams and has huge upside potential for revenue enhancement. For example, his pricing takes into account the team's record, who they are playing, game dates and times, etc.
- Dave builds regression models to pick out key factors driving these historic behaviors and create PMs (Predictive Models) to identify how to spend marketing resources to drive revenues. He builds churn models for season ticket holders to create segments of customers who will renew, will not renew, or are fence-sitters, which then drives more refined marketing campaigns.

Sports Analytics Example 2: The Coach

- Bob Breedlove is the football coach for a major college team. His areas of focus include recruiting the best high school players, developing them to fit his offense and defense systems, and getting maximum effort from them on game days. Sample questions in his area of responsibility include: Whom do we recruit? What drills help develop their skills?
- Dar Beranek specializes in helping the coaches make tactical decisions. She is working with a team of student interns who are creating opponent analytics. They used the coach's annotated game film to build a cascaded decision tree model to predict whether the next play will be a running play or passing play.

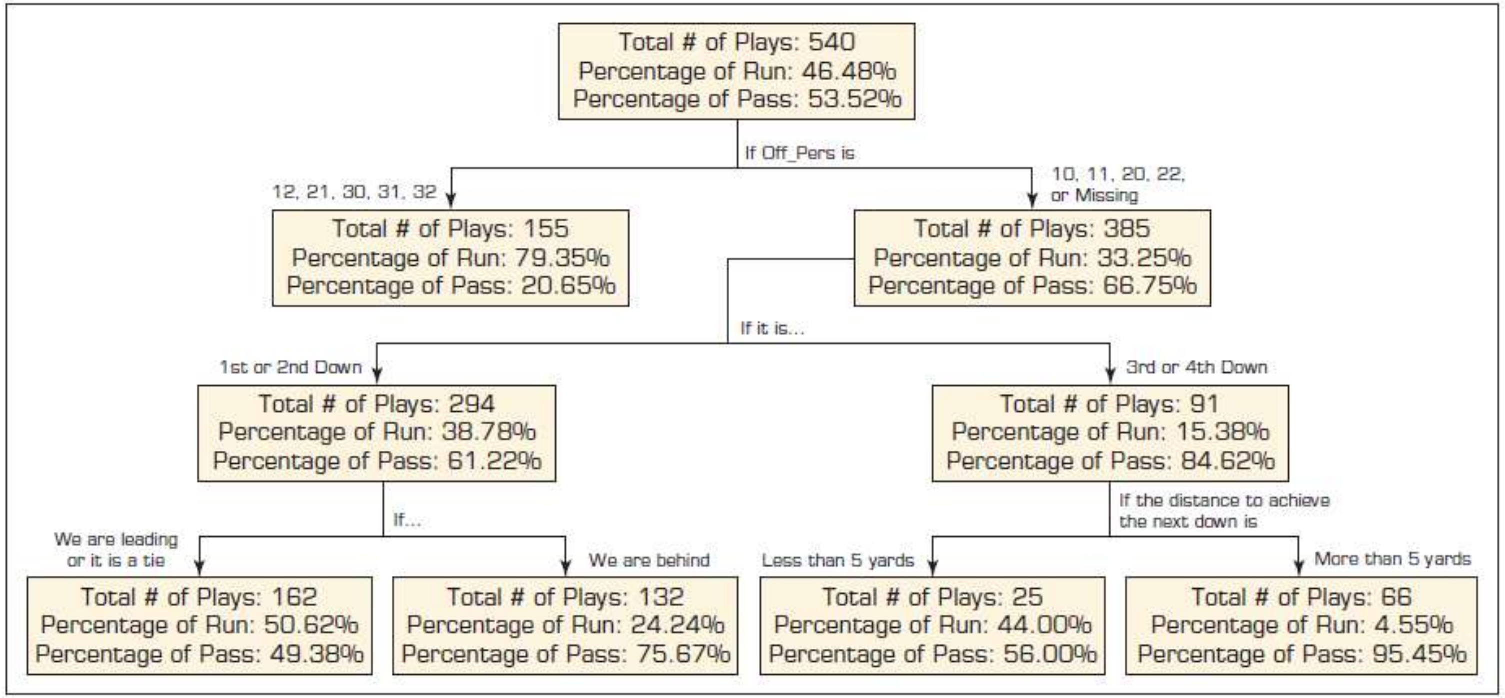


FIGURE 1.12 Cascaded Decision Tree for Run or Pass Plays.

Healthcare Analytics Example 1: Preventing Falls in Seniors

- Accidental falls are a major health risk for adults age 65+ years. The costs of falls pose a significant strain on the healthcare system; estimated to be $34 billion in 2013 alone.
- Humana is a major health insurance company, with 3.2 million members, most of whom are seniors. Keeping its senior members well and helping them live safely is a key business objective. Fall prevention is an important component.
- As such, there is a need for an accurate method to identify individuals at greatest risk of falling so that they can be proactively managed for fall prevention. Administrative medical and pharmacy claims, clinical data, temporal clinical patterns, and other data is needed to identify individuals at high risk of falling over a time horizon.
- Today, the Falls PM (Predictive Models) is central to Humana's ability to identify seniors who could benefit from fall mitigation interventions. Using the PM, Humana was able to identify 20,000 consumers at a high risk of falling who benefited from this program. Identified consumers wear a device that detects falls and alerts a service for immediate assistance.

Retail Analytics

- Goal of the analytics industry is to use information about business suppliers, customers, employees, and stakeholders to make better decisions.

- Example:
    - Amazon's enormous investments in analytics to power their value chain.
    - Walmart & Target have invested millions of dollars in analytics for their supply chains.

- Analytics—descriptive, predictive, or prescriptive—can play a role in making better data-driven decisions.

Retail Analytics (cont.)

- Examples of key questions that can be answered through analytics, and the potential business value derived from fielding such analytics.

TABLE 1.1 Examples of Analytics Applications in the Retail Value Chain

Analytic Application	Business Question	Business Value
Inventory Optimization	1. Which products have high demand? 2. Which products are slow moving or becoming obsolete?	1. Forecast the consumption of fast-moving products and order them with sufficient inventory to avoid a stock out scenario. 2. Perform fast inventory turnover of slow-moving products by combining them with one in high demand.
Price Elasticity	1. How much net margin do I have on the product? 2. How much discount can I give on this product?	1. Markdown prices for each product can be optimized to reduce the margin dollar loss. 2. Optimized price for the bundle of products is identified to save the margin dollar.
Market-Basket Analysis	1. What products should I combine to create a bundle offer? 2. Should I combine products based on slow-moving and fast-moving characteristics? 3. Should I create a bundle from the same category or a different category line?	1. The affinity analysis identifies the hidden correlations between the products, which can help in following values: a. Strategize the product bundle offering based on focus on inventory or margin. b. Increase cross-selling or up-selling by creating bundle from different categories or the same categories, respectively.
Shopper Insight	1. Which customer is buying what product at what location?	1. By customer segmentation, the business owner can create personalized offers resulting in better customer experience and retention of the customer.
Customer Churn Analysis	1. Who are the customers who will not return? 2. How much business will I lose? 3. How can I retain the customers? 4. What demography of customer is my loyal customer?	1. Businesses can identify the customer and product relationships that are not working and show high churn. Thus, they can have better focus on product quality and the reason for that churn. 2. Based on the customer lifetime value (LTV), the business can do targeted marketing resulting in retention of the customer.

Retail Analytics (cont.)

TABLE 1.1 Examples of Analytics Applications in the Retail Value Chain

Analytic Application	Business Question	Business Value
Channel Analysis	1. Which channel has lower customer acquisition cost? 2. Which channel has better customer retention? 3. Which channel is more profitable?	1. Marketing budget can be optimized based on insight for better return on investment.
New Store Analysis	1. What location should I open? 2. What and how much opening inventory should I keep?	1. Best practices of other locations and channels can be used to get a jump-start. 2. Comparison with competitor data can help to create a differentiator to attract the new customers.
Store Layout	1. How should I do store layout for better topline? 2. How can I increase my in-store customer experience?	1. Understand the association of products to decide store layout and better alignment with customer needs. 2. Workforce deployment can be planned for better customer interactivity and thus satisfying customer experience.
Video Analytics	1. What demography is entering the store during the peak period of sales? 2. How can I identify a customer with high LTV at the store entrance so that a better personalized experience can be provided to this customer?	1. In-store promotions and events can be planned based on the demography of incoming traffic. 2. Targeted customer engagement and instant discount enhances the customer experience resulting in higher retention.

Image Analytics

- The benefits of satellite data are significant to scientists who need to regularly monitor global change, land usage, and weather.

- Examples of image analytics include:
    - **World Bank** researchers used satellite data to propose strategic recommendations for urban planners and officials from developing nations.
    - **EarthCast** provides accurate weather updates for a large commercial U.S. airline. It map out conditions along a flight path.
    - **Imazon** started using satellite data to develop a picture of close real-time information on Amazon deforestation.

1.7 Artificial Intelligence Overview

- What Is Artificial Intelligence?
- The Major Benefits of AI
- The Landscape of AI
- The Three Flavors of AI Decisions
- Technology Insights 1.1

What Is Artificial Intelligence (AI)?

- The **major goal** of AI is to create intelligent machines that can do tasks currently done by people.

- **AI tasks** include
  - reasoning,
  - thinking,
  - learning, and
  - problem solving.

- **AI can also be defined as:**
  - Technology that can learn to do things better over time.
  - Technology that can understand human language.
  - Technology that can answer questions.

The Major Benefits of AI

- Significant reduction in the cost of performing work. This reduction continues over time while the cost of doing the same work manually increases with time.

- Work can be performed much faster.

- Work is consistent in general, more consistent than human work.

- Increased productivity and profitability as well as a competitive advantage are the major drivers of AI.

The Landscape of AI

- We defined the landscape/ecosystem of AI into 5 categories:

1. **Major Technologies:** machine learning, deep learning, intelligent agents
2. **Knowledge-Based Technologies:** expert systems, recommendation engines, chatbots.
3. **Biometric-Related Technologies:** natural language processing, image recognition
4. **Support Theories, Tools, & Platforms:** computer science, cognitive science, mathematics, statistics, sensors, augmented reality, neural networks, APIs, knowledge management.
5. **AI Applications:** smart cities, smart homes, automatic decisions, translation, robotics, fraud detection.

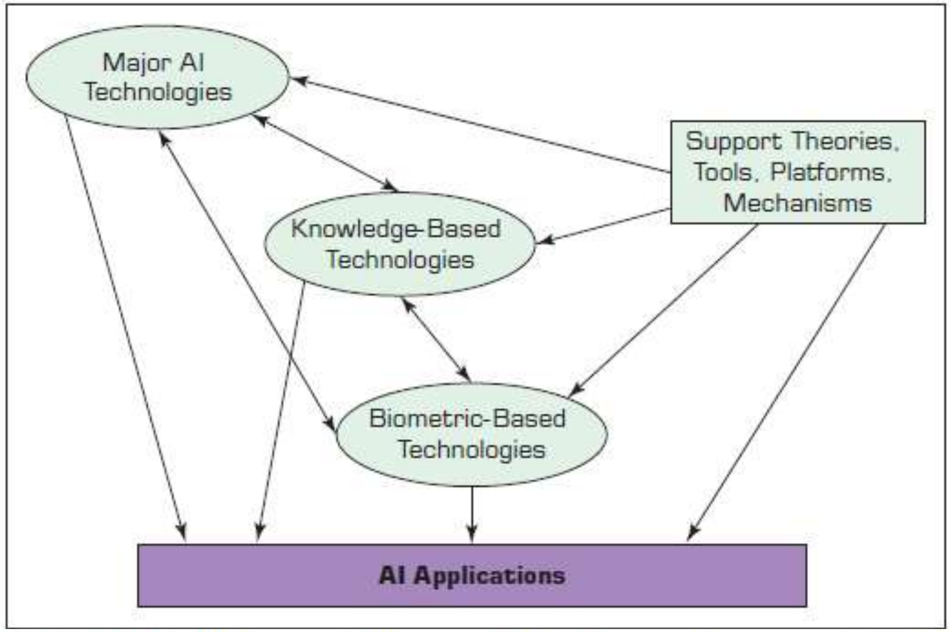


FIGURE 1.16 The Landscape (Ecosystem) of AI. *Source:* Drawn by E. Turban.

The Landscape of AI (cont.)

- AI **Applications** are in area like
  - business,
  - medicine & healthcare
  - transportation
  - education

- AI field divided into two major categories of applications:

  1. **Narrow (Weak):** Focuses on one narrow field (Domain). Examples include SIRI and Alexa that operate in limited, predefined areas.

  2. **General (Strong):** To exhibit real intelligence, machines need to perform the full range of human cognitive capabilities (e.g. reasoning and problem solving).

The Three Flavors of AI Decisions

- The capabilities of AI systems can be divided into three levels:

1. **Assisted Intelligence:** equivalent mostly to the weak AI, which works only in narrow domains. It requires clearly defined inputs and outputs such as monitoring systems.

2. **Autonomous AI:** in the realm of the strong AI but in a narrow domain. Eventually, the computer will take over as very narrow expert and have absolute decision-making power.

3. **Augmented Intelligence:** between assisted and autonomous. The technology focuses on augmenting computer abilities to extend human cognitive abilities, resulting in high performance. Examples: Cybercrime fighting, E-Commerce decisions, High-frequency stock market trading.

Technology Insights 1.1

- **Differences** between traditional and augmented AI:

  1. Augmented machines extend rather than replace human decision making

  2. Augmentation excels in solving complex human and industry problems in specific domains in contrast with strong, general AI.

  3. In contrast with a "black box" model of some AI and analytics, augmented intelligence provides insights and recommendations, including explanations.

1.7 Convergence of Analytics and AI

- Major Differences between Analytics and AI
- Why Combine Intelligent Systems?
- How Convergence Can Help?
- Big Data Is Empowering AI Technologies
- The Convergence of AI and the IoT
- The Convergence with Blockchain and Other Technologies

Major Differences between Analytics and AI

- **Analytics** processes historical data using statistical, management science and other computational tools to describe situations, predict results, and propose recommendations for solutions to problems.

- **AI**'s major objective is to mimic the manner in which people think, learn, reason, make decisions, and solve problems.

- The emphasis in **AI** is on knowledge and intelligence as major tools for solving problems rather than relying on computation, which we do in analysis.

Why Combine Intelligent Systems?

- Analytics, AI and their different technologies have **limitations**, resulting in only a small chance that they can be used to reach organizational excellence.
- There are several reasons for this situation including:
  1. Predictive models have unintended effects
  2. The results of analytics may be good for some applications but not for others
  3. Models are as good as their input data and assumptions
  4. Data could be incomplete/inaccurate.
  5. Data collected from different sources can vary in format and quality.
- A major reason for the high failure rate of AI is that some of its technologies need a large amount of data (Big Data). Without this continuous flow of data, there would not be good learning in **AI**.

How Convergence Can Help?

- **BI** and its **analytics** answer most of the why and what questions regarding the sufficiency of problem solving.

- The next generation of **business intelligence platforms** will **use AI** to automatically locate, visualize, and narrate important things. This can also be used to create automatic alerts and notifications.

- Machine learning and deep learning can support **analytics** by conducting pattern recognition and more accurate predictions. AI will help to compare actual performance with the predicted one.

Big Data Is Empowering AI Technologies

- Technologies and methods that enable capturing, cleaning, and analyzing **Big Data**, can also enable companies to make real-time decisions.
- The availability of new **Big Data analytics** enables new capabilities in **AI technologies** that were not possible until recently.
- **Big Data** can empower **AI** due to:
  - The new capabilities of processing Big Data at a much reduced cost.
  - The availability of large data sets online.
  - The scale up of algorithms, including deep learning, is enabling powerful AI capabilities.

The Convergence of AI and the IoT

- The **IoT** collects a large amount of data from sensors and other "things." These data need to be processed for decision support.
- Combining **AI** and **IoT** can
  - lead to the "next-level solutions and experiences." The emphasis in such combination is on learning more about customers and their needs.
  - facilitate competitive analysis and business operation.
- Three examples of combining **AI** and **IoT**:
  - The smart thermostat of Nest Labs
  - Automated vacuum cleaners
  - Self-driving vehicles

The Convergence with Blockchain & Other Technologies

- Several experts raise the possibility of the convergence of **AI**, **analytics**, and **blockchain**. This convergence also can include the **IoT**.
- The **blockchain** technology can add security to data shared by all parties in a distributed network, where transaction data can be recorded.
- This combination can be very useful in complex applications such as autonomous vehicles.

1.8 Overview of the Analytics Ecosystem

- Analytics Ecosystem

Analytics Ecosystem

- **Purpose** of analytics ecosystem is to be aware of organizations, new offerings, and opportunities in sectors allied with analytics.
- The **components** of the ecosystem are represented by the petals of an analytics flower.
- Grouped into **categories**:
    - inner petals
    - outer petals
    - seed

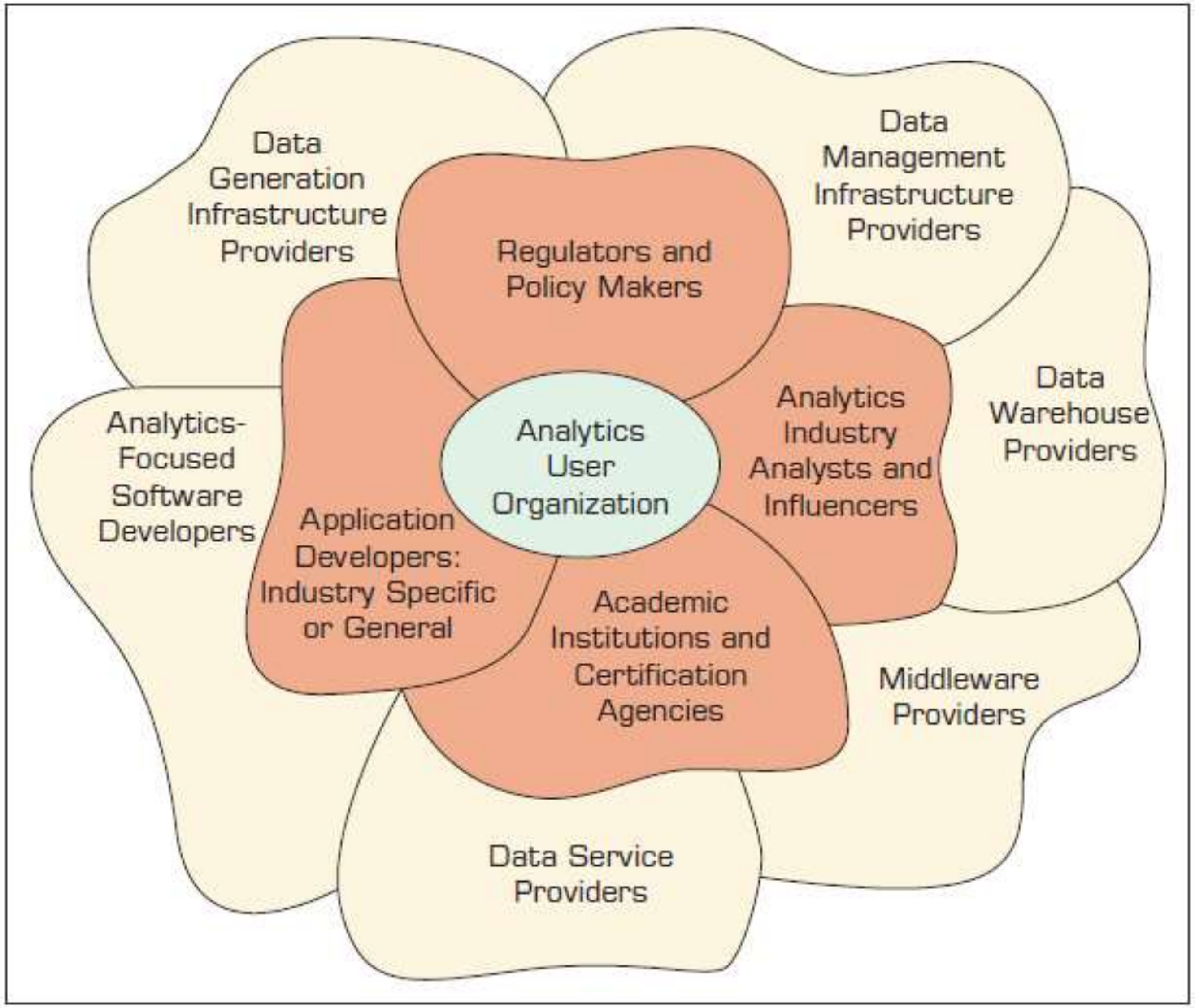


FIGURE 1.17 Analytics Ecosystem.

Main Reference

- **Chapter 1:** “An overview of Decision Support system, Business Intelligence, Analysis, and AI” from “Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”.

Week self-review exercises

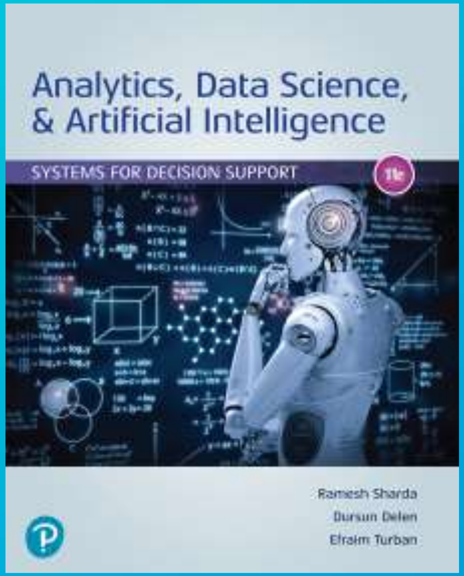
- **Application Case 1.1 to Application Case 1.10** from “Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”

Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

College of Computing and Informatics



Week 6

Chapter 8- part1: Prescriptive Analytics: Optimization and Simulation

Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Understand the applications of prescriptive analytics techniques in combination with reporting and predictive analytics
2. Understand the basic concepts of analytical decision modelling
3. Understand the concepts of analytical models for selected decision problems, including linear programming and simulation models for decision support
4. Explain the basic concepts of optimization and when to use them
5. Describe how to structure a linear programming
6. Explain what is meant by sensitivity analysis, what-if analysis, and goal seeking
7. Understand the concepts and applications of different types of simulation
8. Understand potential applications of discrete event simulation

Required Reading

- **Chapter 8 (sections 8.2 to 8.4 and 8.6 – 8.10):** “Prescriptive Analytics: Optimization and Simulation” from “*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*”.

Recommended Reading

- ***Artificial Intelligence: Systems for Decision Support”: Monte Carlo Simulation***

https://www.palisade.com/risk/monte_carlo_simulation.asp

- ***It’s Only Logical: Decision Tables and Decision Trees***

https://medium.com/analysts-corner/its-only-logical-decision-tables-and-decision-trees-12a8b52243ea

Recommended Video

- **Decision Trees & Decision Tables**

https://www.youtube.com/watch?v=A5-w3mof-3I

8.2 Model-Based Decision Making

- Prescriptive Analytics Model Examples
- Identification of the Problem and Environmental Analysis
- Model Categories
- Current Trends in Modeling

Prescriptive Analytics Model Examples

- **Modeling** is a key element for prescriptive analytics.
- Depending on the problem we are addressing, there are many classes of models, and there are often many specialized techniques for solving each one.
- Prescriptive analytics involves the application of mathematical models, sometimes the term data science is more commonly associated with the application of such mathematical models.

Identification of the Problem and Environmental Analysis

- It is important to analyze the scope of the domain and the forces and dynamics of the environment when making a decision.
- A decision maker needs to identify the organizational culture and the corporate decision-making processes.
- **Environmental scanning** and analysis is the monitoring, scanning, and interpretation of collected information.
- Variable Identification is critical, as are the relationships among the variables.
- **Influence diagrams** can facilitate the identification process.
- A **cognitive map**, can help a decision maker develop a better understanding of a problem, and variable interactions.

Identification of the Problem and Environmental Analysis

- **Predictive analytics** (forecasting) is essential for constructing and manipulating models because the results of an implemented decision occur in the future.
- There is no point in running a what-if (sensitivity) analysis on the past because decisions made then have no impact on the future.
- Online commerce and communication has created an immense need for forecasting and an abundance of available information for performing it.
- These activities occur quickly, yet information about such purchases is gathered and should be analyzed to produce forecasts.
- Forecasting models use product life-cycle needs and information about the marketplace to analyze the situation.

Model Categories

- The following table classifies decision models into seven groups and lists several representative techniques for each category.
- Each technique can be applied to a static or a dynamic model, which can be constructed under assumed environments of certainty, uncertainty, or risk.

TABLE 8.1 Categories of Models

Category	Process and Objective	Representative Techniques
Optimization of problems with few alternatives	Find the best solution from a small number of alternatives	Decision tables, decision trees, analytic hierarchy process
Optimization via algorithm	Find the best solution from a large number of alternatives, using a step-by-step improvement process	Linear and other mathematical programming models, network models
Optimization via an analytic formula	Find the best solution in one step, using a formula	Some inventory models
Simulation	Find a good enough solution or the best among the alternatives checked, using experimentation	Several types of simulation
Heuristics	Find a good enough solution, using rules	Heuristic programming, expert systems
Predictive models	Predict the future for a given scenario	Forecasting models, Markov analysis
Other models	Solve a what-if case, using a formula	Financial modeling, waiting lines

Model Categories (cont.)

- To expedite model construction, we can use special decision analysis systems that have modeling languages and capabilities embedded in them.
- These include spreadsheets, data mining systems, online analytic processing (OLAP) systems, and modeling languages.
- **Model Management:** Models must be managed to maintain integrity, and applicability. This is done with the aid of model-based management systems, which are analogous to database management systems (DBMS).
- **Knowledge-based Modeling DSS**: use quantitative models, whereas expert systems use qualitative, knowledge-based models in their applications.

Current Trends in Modeling

- One trend in modeling involves the development of model libraries and solution technique libraries.
- There is a clear trend toward developing and using cloud-based tools and software to run software to perform modeling, optimization, simulation, etc.
- With management models, the amount of data and model sizes is large, necessitating data warehouses and parallel computing for solutions.
- There is a trend toward making analytics models transparent to decision makers, and using influence diagrams (a model of a model to help in analysis).
- Many decision makers accustomed to slicing and dicing data cubes are now using OLAP systems that access data warehouses.

8.3 Structure of Mathematical Models for Decision Support

- Components of Decision Support Mathematical Models
- The Structure of Mathematical Models

Components of Decision Support Mathematical Models

- Quantitative models are made up of **four** basic components: result variables, decision variables, uncontrollable variables, and intermediate result variables.
- Mathematical relationships link these components together.
- In non-quantitative models, the relationships are symbolic or qualitative. The results of decisions are determined based on decisions made, uncontrollable variable, and relationships among variables.
- The modeling process involves identifying the variables and relationships. Solving a model determines the values of these and the result variable(s).

Components of Decision Support Mathematical Models

- **Result/Outcome/Output Variables:** reflect level of effectiveness of a system.
- **Decision Variables:** Decision variables describe alternative courses of action. The decision maker controls the decision variables.
- **Uncontrollable Variables/Parameters:** fixed/varying factors that affect the result variables but are not under the decision maker control.
- **Intermediate Result Variables:** reflect intermediate outcomes in models.

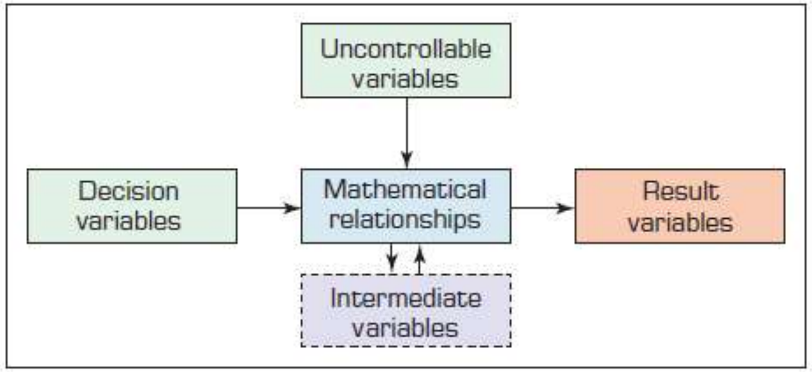


FIGURE 8.1 The General Structure of a Quantitative Model.

The Structure of Mathematical Models

- Components of a quantitative model are linked by mathematical expressions.
- A very simple financial model is: $P = R - C$
- Another financial model is the simple present-value cash flow model.
- It is possible to determine the present value of a payment of $100,000 to be made 5 years from today, at a 10% interest rate, as follows:

$$P = 100{,}000/(1 + 0.1)^5 = 62{,}092$$

8.4 Certainty, Uncertainty, and Risk

- Decision Making under Certainty
- Decision Making under Uncertainty
- Decision Making under Risk (Risk Analysis)

Decision Making under Certainty

- In decision making under certainty, complete knowledge is available so decision maker know the outcome of each course of action.
- This is done with structured problems and short time horizons (up to 1 year).
- Outcomes are not 100% known, but this assumption simplifies the model.
- The decision maker is viewed as a perfect predictor of the future because it is assumed that there is only one outcome for each alternative.
- Certainty models are easy to develop and solve, and yield optimal solutions.
- Financial models are constructed under assumed certainty.

Decision Making under Uncertainty

- Decision maker considers situation where several outcomes are possible for each course of action.
- In contrast to the risk situation, the decision maker does not know, or cannot estimate, the probability of occurrence of the possible outcomes.
- Modeling such situations involves assessment of the decision maker's attitude toward risk.
- Instead of dealing with uncertainty, manager's sometimes attempt to obtain more information so that the problem can be treated under certainty.
- If more information is not available, the problem must be treated under a condition of uncertainty, which is less definitive than the other categories.

Decision Making under Risk (Risk Analysis)

- **Risk analysis** is a decision-making method that analyzes risk associated with alternatives, each with a given probability of occurrence.
- The probabilities that the given outcomes will occur are assumed to be known or can be estimated. Under these assumptions, the decision maker can assess the degree of risk associated with each alternative (calculated risk).
- Risk analysis can be performed by calculating the expected value of each alternative and selecting the one with the best expected value.

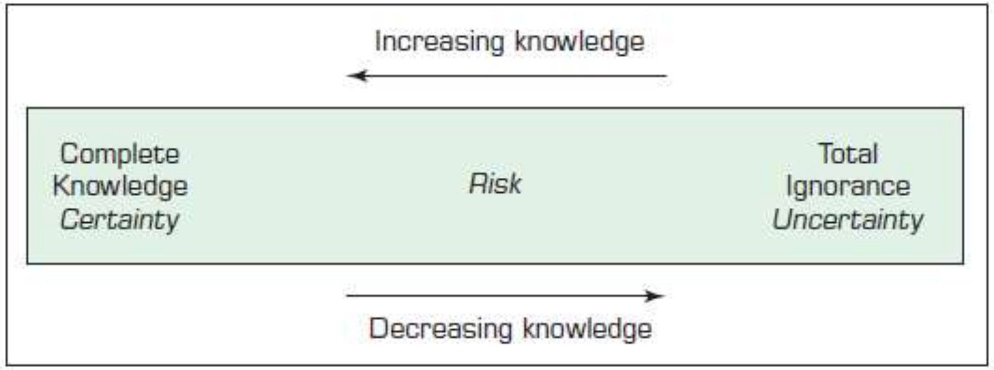


 FIGURE 8.2 **The Zones of Decision Making.**

8.6 Mathematical Programming Optimization

- Overview
- Linear Programming Model
- Implementation

Overview

- Mathematical programming are tools that helps decision makers allocate scarce resources among competing activities to optimize a measurable goal.
- In linear programming (LP), one of the tools, all variable relationships are linear. Applications include supply chain management, product decisions, etc.

LP allocation problems usually display the following characteristics:

- Limited quantity of resources, most are used in product/service production
- Two or more ways resources can be used. Each is called a solution/program.
- Each activity where resources are used, yields a return in terms of stated goal

Linear Programming Model

- **The LP allocation model is based on the following economic assumptions:**
  - Returns from allocations are independent and measured by a common unit
  - The total return is the sum of the returns yielded by the different activities.
  - All data are known with certainty, and resources are used economically

- Allocation problems have a large number of possible solutions. Depending on the assumptions, the number of solutions can be either infinite or finite.

- Different solutions yield different rewards. The solution with highest degree of goal attainment is called optimal solution, and found by a special algorithm

Linear Programming Model (cont.)

- **Every LP model is composed of:**

- Decision variables: Unknown values that are being searched for
- Objective Function: A linear mathematical function that relates the decision variables to the goal, measures goal attainment, and is to be optimized
- Coefficients: indicate contribution to objective of one unit of a decision variable
- Constraint: Linear (in)equalities that limit resources
- Capacities: Describe upper and lower limits on the constraints and variables
- Input/output & Coefficients: Indicate resource utilization for a decision variable

Implementation

- Implement the model in "standard form", where constraints are written with decision variables on the left and a number on the right.
- Alternatively, use spreadsheet to calculate the model in a less rigid manner.
- LP models can be specified directly in a number of user-friendly modeling systems. Models are specified in the same way they are defined algebraically.
- **Optimization models can be solved by mathematical programming methods:**
  - Assignment & Network models for planning and scheduling
  - Dynamic, Goal, Linear, Nonlinear and Integer programming
  - Investment & Replacement
  - Simple inventory models & Transportation

8.7 Multiple Goals, Sensitivity Analysis, What-If Analysis, and Goal Seeking

- Multiple Goals
- Sensitivity Analysis
- Types of Sensitivity Analysis
- What-If Analysis
- What-If Analysis Example
- Goal Seeking
- Goal Seeking Example

Multiple Goals

- Today's management systems are complex, and managers want to attain simultaneous goals, some of which may conflict.
- Transform a multiple-goal problem into a single-measure-of effectiveness problem before comparing the effects of the solutions.
- **Certain difficulties may arise when analyzing multiple goals:**
  - It is difficult to obtain the organization's goals explicitly
  - Goals are viewed differently at various levels of the organization
  - Goals and their importance change in response to the organization
  - Complex problems are solved by decision makers with different agendas

Sensitivity Analysis

- Sensitivity analysis assesses impact of input data changes on proposed solution.

- **Sensitivity analysis allows for:**
  - Adaptation to conditions of different decision-making situations
  - Provides a better understanding of the model
  - Permits the input of data to increase model confidence.

Sensitivity Analysis (cont.)

Sensitivity analysis tests relationships such as the following:

- Impact of parameter change, and decision variables on outcome variable(s)
- The effect of uncertainty in estimating external variables
- The effect of different dependent interactions among variables
- The robustness of decisions under changing conditions

Sensitivity analyses are used for:

- Revising models to eliminate too-large sensitivities
- Detailing variables and obtaining estimates of sensitive external variables
- Altering a real-world system to reduce actual sensitivities

Types of Sensitivity Analysis

- The two types of sensitivity analyses are automatic and trial and error.

Automatic Sensitivity Analysis:

- This is performed in standard quantitative model implementations such as LP.
- It is usually limited to one change at a time, and only for certain variables.
- It is powerful because of its ability to establish ranges and limits very fast

Trial-and-error Sensitivity Analysis:

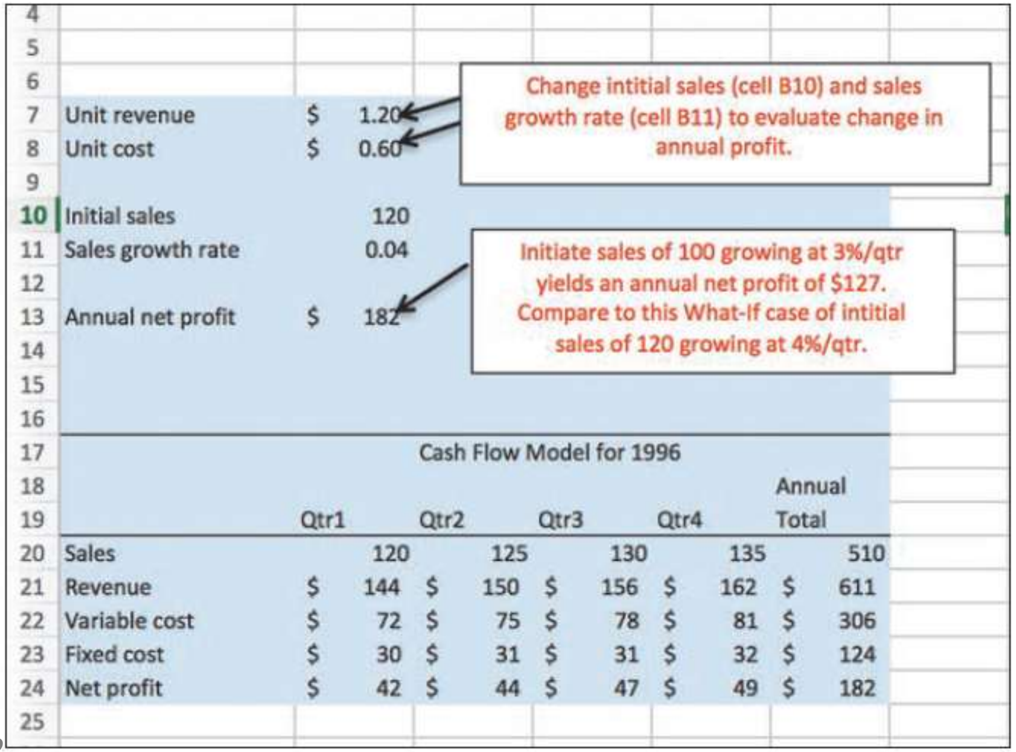
- Impact of changes in variable(s) is determined by trial-and-error approach
- When changes are repeated, better and better solutions may be discovered.
- Such experimentation has **two** approaches: what-if analysis and goal seeking.

What-If Analysis

- **What-if analysis** is structured as: *What will happen to the solution if an input variable, assumption, or parameter value is changed?*
  - Total inventory cost if the carrying inventories cost increases by 10%?
  - Market share if the advertising budget increases by 5%?
- With the appropriate user interface, managers can ask a computer model these types of questions and get immediate answers.
- Performs multiple cases and change the percentage, or other data as needed.
- What-if analysis is common in many decision systems. Users are given the opportunity to change their answers to the system's questions, and a revised recommendation is found.

What-If Analysis Example

A what-if query for a cash flow problem: the user changes the cells containing the initial sales (from 100 to 120) and the sales growth rate (from 3% to 4% per quarter), the program immediately re-computes the value of the annual net profit cell (from $127 to $182) .



4						
5						
6						
7	Unit revenue	$ 1.20				
8	Unit cost	$ 0.60				
9						
10	Initial sales	120				
11	Sales growth rate	0.04				
12						
13	Annual net profit	$ 182				
14						
15						
16						
17			Cash Flow Model for 1996			
18						Annual
19		Qtr1	Qtr2	Qtr3	Qtr4	Total
20	Sales	120	125	130	135	510
21	Revenue	$ 144	$ 150	$ 156	$ 162	$ 611
22	Variable cost	$ 72	$ 75	$ 78	$ 81	$ 306
23	Fixed cost	$ 30	$ 31	$ 31	$ 32	$ 124
24	Net profit	$ 42	$ 44	$ 47	$ 49	$ 182
25						

Goal Seeking

- **Goal seeking** calculates the values of the inputs necessary to achieve a desired level of an output (goal). The following are some examples:

  - Annual R&D budget is needed for an annual growth rate of 15% by 2018?
  - How many nurses needed to reduce the average waiting time of a patient in the emergency room to less than 10 minutes?

- **Computing A Break-even Point By Using Goal Seeking:**

  - Some modeling software packages can directly compute break-even points, which is an important application of goal seeking.
  - Determining the value of the decision variables that generate zero profit.
  - Sensitivity analysis as the prewritten routines present a limited opportunity for asking what-if questions.

Goal Seeking

- In a financial planning model, the internal rate of return (IRR) is the interest rate that produces a net present value (NPV) of zero.
- Given a stream of annual returns in Column E, we can compute the NPV of planned investment through goal-seeking.
- An NPV equal to zero determines the IRR of this cash flow, including the investment. We set the NPV cell to 0 by changing the interest rate cell.
- **The answer is 38.77059%.**

5					
6					
7	Investment Problem		Initial Investment:		$ 1,000.00
8	Example of GoalSeeking		Interest Rate:		10%
9					
10	Find the Interest Rate		Annual		NPV
11	(the Internal Rate of	Year	Returns		Calculations
12	Return-IRR)	1	$ 120.00		$109.09
13	that yields an NPV	2	$ 130.00		$118.18
14	of $0	3	$ 140.00		$127.27
15		4	$ 150.00		$136.36
16		5	$ 160.00		$145.45
17		6	$ 152.00		$138.18
18		7	$ 144.40		$131.27
19		8	$ 137.18		$124.71
20		9	$ 130.32		$118.47
21		10	$ 123.80		$112.55
22					
23			The NPV	Solutions:	$261.55
24					

8.8 Decision Analysis with Decision Tables & Decision Trees

- Decision Tables
- Decision Tables Example
- Decision Trees

Decision Tables

- Decision tables organize information in systematic, tabular form for analysis.
- **Treating Uncertainty**: Several methods are available for handling uncertainty.

- Optimistic approach: assumes and selects best outcomes for alternatives
- Pessimistic approach: assumes worst outcome for alternatives; selects the best
- Another approach simply assumes that all states of nature are equally possible.

- When possible, analysts should attempt to gather information to treat the problem under assumed certainty.

- **Treating Risk:** The most common method for solving this risk analysis problem is to select the alternative with the greatest expected value.

Decision Tables Example

- An investor estimates: solid growth (50%), stagnation (30%), and inflation (20%)
- Expected value is computed by multiplying result probabilities and adding them

TABLE 8.3 Investment Problem Decision Table Model

	State of Nature (Uncontrollable Variables)		
Alternative	**Solid Growth (%)**	**Stagnation (%)**	**Inflation (%)**
Bonds	12.0	6.0	3.0

Bond investment yields an expected return of 12(0.5) + 6(0.3) + 3(0.2) = 8.4%

- Suppose a financial advisor presents a $1,000 investment with 0.9999 chance to double your money, and 0.0001 chance you'll lose $500,000.

0.9999 ($2,000 - $1,000) + .0001(-$500,000 - $1,000) = $999.90 - $50.10 = $949.80

- The potential loss could be catastrophic for any investor

Decision Trees

- **Decision trees** are alternative representations of a decision table; it shows a problem's relationship graphically and handles complex situations compactly.
- TreeAge Pro & PrecisionTree are systems that show decision trees in practice.
- You can apply mathematical programming to decision-making situations under risk. These include simulation, certainty factors, and fuzzy logic.
- A simplified investment case of multiple goals is shown in the table. The three goals are yield, safety, and liquidity. This situation is under assumed certainty.

TABLE 8.4 Multiple Goals

Alternative	Yield (%)	Safety	Liquidity
Bonds	8.4	High	High
Stocks	8.0	Low	High
CDs	6.5	Very high	High

8.9 Introduction to Simulation

- Major Characteristics of Simulation
- Advantages of Simulation
- Disadvantages of Simulation
- The Methodology of Simulation
- Simulation Types
- Monte Carlo Simulation
- Discrete Event Simulation

Major Characteristics of Simulation

- **Simulation** involves building a model of reality to the extent practical.
- Simulation models may suffer from fewer assumptions about the decision situation as compared to other prescriptive analytic models.
- Simulation is a technique for conducting experiments. Therefore, it involves testing specific values of the decision or uncontrollable variables in the model and observing the impact on the output variables.
- Simulation is used only when a problem is too complex to be treated using numerical optimization techniques.
- Complexity in this situation means either that the problem cannot be formulated for optimization, that the formulation is too large, that there are too many interactions among the variables.

Advantages of Simulation

Simulation is used in decision support modeling for the following reasons:

- The theory is straightforward, and model is built from manager’s perspective.
- Time compression is attained quickly to give idea of policies’ long-term effects.
- Descriptive rather than normative, allowing managers to ask what-if questions, and use a trial-and-error approach with less expense and risk.
- Requires intimate knowledge; model builder constantly interact with manager.
- Can handle a variety of problem types and higher-level managerial functions.
- Produces performance measures, and includes real complexities of problems.
- Can readily handle relatively unstructured problems.

Disadvantages of Simulation

- An optimal solution cannot be guaranteed, but relatively good ones are generally found.
- Simulation model construction can be a slow and costly process, although newer modeling systems are easier to use than ever.
- Solutions and inferences from a simulation study are usually not transferable to other problems because the model incorporates unique problem factors.
- Simulation is sometimes so easy to explain to managers that analytic methods are often overlooked.
- Simulation software sometimes requires special skills because of the complexity of the formal solution method.

The Methodology of Simulation

Simulation involves setting up a model of a real system through the steps:

- **Define the problem:** Examine problem, and specify need for simulation
- **Construct model:** Determine variables, relationships, and gather data.
- **Test and validate model:** Ensure model properly represents studied system.
- **Design experiment:** There are two conflicting objectives: accuracy and cost.
- **Conduct experiment:** can involve issues like number generation.
- **Evaluate results:** Statistical tools/sensitivity analyses used to interpret results.
- **Implement results:** Managerial involvement leads to implementation success.

The Methodology of Simulation

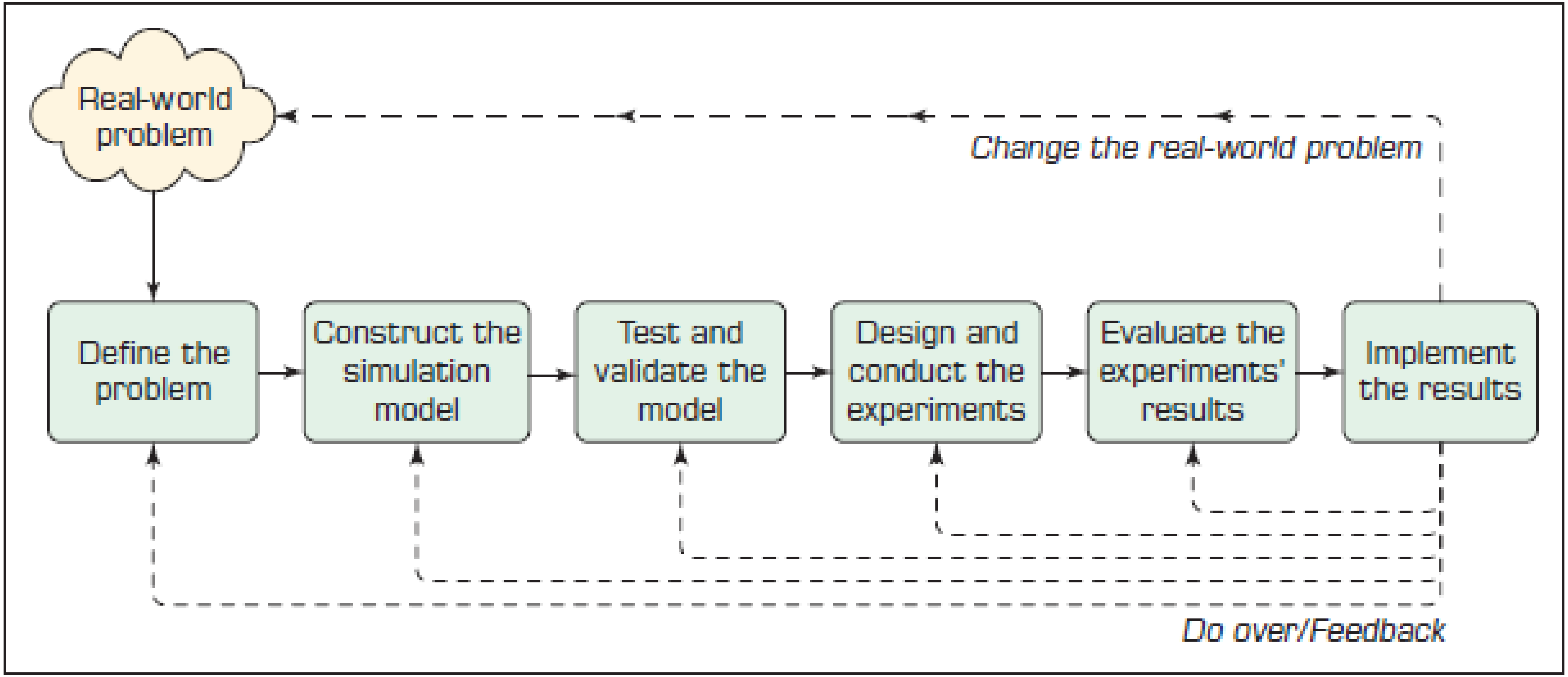


FIGURE 8.12 **The Process of Simulation.**

Simulation Types

- **Simulation model** consists of relationships that present the real-world operations. Simulation results depend on the set of parameters given as inputs.
- There are various simulation paradigms such as Monte Carlo simulation, discrete event, agent based, or system dynamics.
- The level of abstraction in a problem can determine simulation technique.
- Discrete events and agent based models are used for low levels of abstraction.
- They consider individual elements such as people in the simulation models, whereas systems dynamics is more appropriate for aggregate analysis.
- Here we introduce the major types of simulation: probabilistic simulation, time-dependent and time-independent simulation, and visual simulation.

Simulation Types

- **Probabilistic Simulation**: One or more independent variables are probabilistic. They follow probability distributions, which can be discrete or continuous:

  - Discrete involves situations with limited event numbers and finite values.
  - Continuous distributions are situations with unlimited numbers of possible events that follow density functions, such as the normal distribution.

- **Time-dependent Versus Time-independent Simulation**:

  - Time-independent refers to situations where time of event occurrence is unimportant. For example, we may know that the demand for a product is 3 units/day, but do not care when during the day the item is demanded.
  - However, in waiting-line problems applicable to e-commerce, it is important to know the precise time of arrival. This is a time-dependent situation.

Monte Carlo Simulation

- In business decision problems, we employ probabilistic simulations. The **Monte Carlo simulation** is commonly used.
- This method begins with building a model of the decision problem without having to consider the uncertainty of any variables.
- Then we recognize that certain variables are uncertain or follow an estimated probability distribution. This estimation is based on analysis of past data.
- Then we begin running sampling experiments. This consists of generating random values of uncertain parameters and then computing values of the variables that are impacted by such parameters or variables.
- We then analyze the behavior of these performance variables by examining their statistical distributions.

Discrete Event Simulation

- **Discrete event simulation** refers to building a model of a system where the interaction between different entities is studied.
- An example of this is modeling the customers arriving at various rates and the server serving at various rates, we can estimate the average system performance, waiting time, number of waiting customers, etc. Such systems are viewed as collections of customers, queues, and servers.
- There are thousands of documented applications of discrete event simulation models in engineering, business, and so on.
- Tools for building discrete event simulation models have been around for a long time, but these have evolved to take advantage of developments in graphical capabilities for building and understanding the results of such simulation models.

8.10 Visual Interactive Simulation

- Conventional Simulation Inadequacies
- Visual Interactive Simulation
- Visual Interactive Models and DSS

Conventional Simulation Inadequacies

- **Simulation** is a well-established, useful, descriptive, mathematics-based method for gaining insight into complex decision-making situations.

- Simulation does not usually allow decision makers to see how a solution to a complex problem evolves over time, nor interact with the simulation.

- Simulation reports statistical results at the end of a set of experiments. Decision makers are thus not an integral part of simulation development and experimentation, and their experience and judgment cannot be used directly.

- If the simulation results do not match the intuition or judgment of the decision maker, a confidence gap in the results can occur.

Visual Interactive Simulation

- Visual interactive simulation (VIS), visual interactive modeling (VIM) and visual interactive problem solving, is a simulation method that lets decision makers see what a model is doing, how it interacts with made decisions.
- Users employ knowledge to try different decision strategies while interacting with the model. Decision makers can contribute to model validation.
- VIS uses animated computer graphic displays to present the impact of different managerial decisions. It differs from regular graphics in that the user can adjust the decision-making process and see results of the intervention.
- VIS can represent static or dynamic systems. Static models display a visual image of the result of one decision alternative at a time. Dynamic models display evolving systems over time. The evolution is represented by animation.

Visual Interactive Models (VIM) and DSS

- **VIM in DSS** has been used in several operations management decisions.
- The method consists of priming a visual interactive model of a company with its current status.
- Waiting-line management is a good example of VIM. Such a DSS usually computes measures of performance for the various decision alternatives.
- Complex waiting-line problems require simulation. VIM can display the size of the waiting line as it changes during the simulation runs and can graphically present the answers to what-if questions regarding changes in input variables.
- The VIM approach can be used with AI. Integration of the two techniques adds several capabilities that range from the ability to build systems graphically to learning about the dynamics of the system.

Main Reference

- **Chapter 8 (sections 8.2 to 8.4 and 8.6 to 8.10: Modeling in LP: An Example):** “Prescriptive Analytics: Optimization and Simulation” from “*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*”.

Week self-review exercises

- **Application Case 8.5 to Application Case 8.9** from ““Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”.

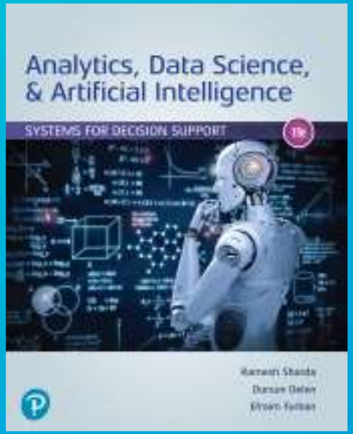
Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

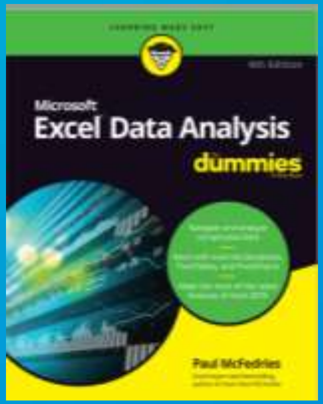
College of Computing and Informatics

Week 4



Chapter 3 – part1 (sections 3.2 to 3.6): Nature of Data, Statistical Modeling, and Visualization

Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support



Chapter 3, 5, 11 & 13: Getting Started with Data Analysis

Excel Data Analysis For Dummies

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Understand the nature of data as they relate to business intelligence (BI) and analytics
2. Learn the methods used to make real-world data analytics ready
3. Describe statistical modelling and its relationship to business analytics
4. Learn about descriptive and inferential statistics
5. Demonstrate the ability in utilizing Excel to perform data preprocessing and data analysis

Required Reading

- **Chapter 3 (sections 3.2 to 3.6)**: “Nature of Data, Statistical Modeling, and Visualization” from “*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*”.
- ***Chapter 3, 5, 11 & 13:*** *“Getting Started with Data Analysis”* from *“Excel Data Analysis For Dummies”*

Recommended Video

- **How to Clean Up Raw Data in Excel**
  https://www.youtube.com/watch?v=3es54FafNC0
- **10 Super Neat Ways to Clean Data in Excel**
  https://www.youtube.com/watch?v=e0TfIbZXPeA
- **Business Analytics with Excel | Data Science Tutorial | Simplilearn**
  https://www.youtube.com/watch?v=W3vrMSah3rc
- **Predictive Modelling in Excel – How to Create a Linear Regression Model from Scratch**
  https://www.analyticsvidhya.com/blog/2020/06/predictive-modeling-excel-linear-regression/

Chapter 3 – part1 (sections 3.2 to 3.6): Nature of Data, Statistical Modeling, and Visualization

Analytics, Data Science, & Artificial Intelligence Systems For Decision Support

3.2 Nature of Data

- Overview of the Nature of Data
- The Characteristics of Analytics–Ready Data

Overview of the Nature of Data

- Data is the main element for BI, data science, and business analytics. It can be small or large; structured or unstructured.

- Modern-day data collection mechanisms use radio frequency identification (RFID)–based computerized networks. These automated data collection systems collects volumes of data and enhances its quality and integrity.

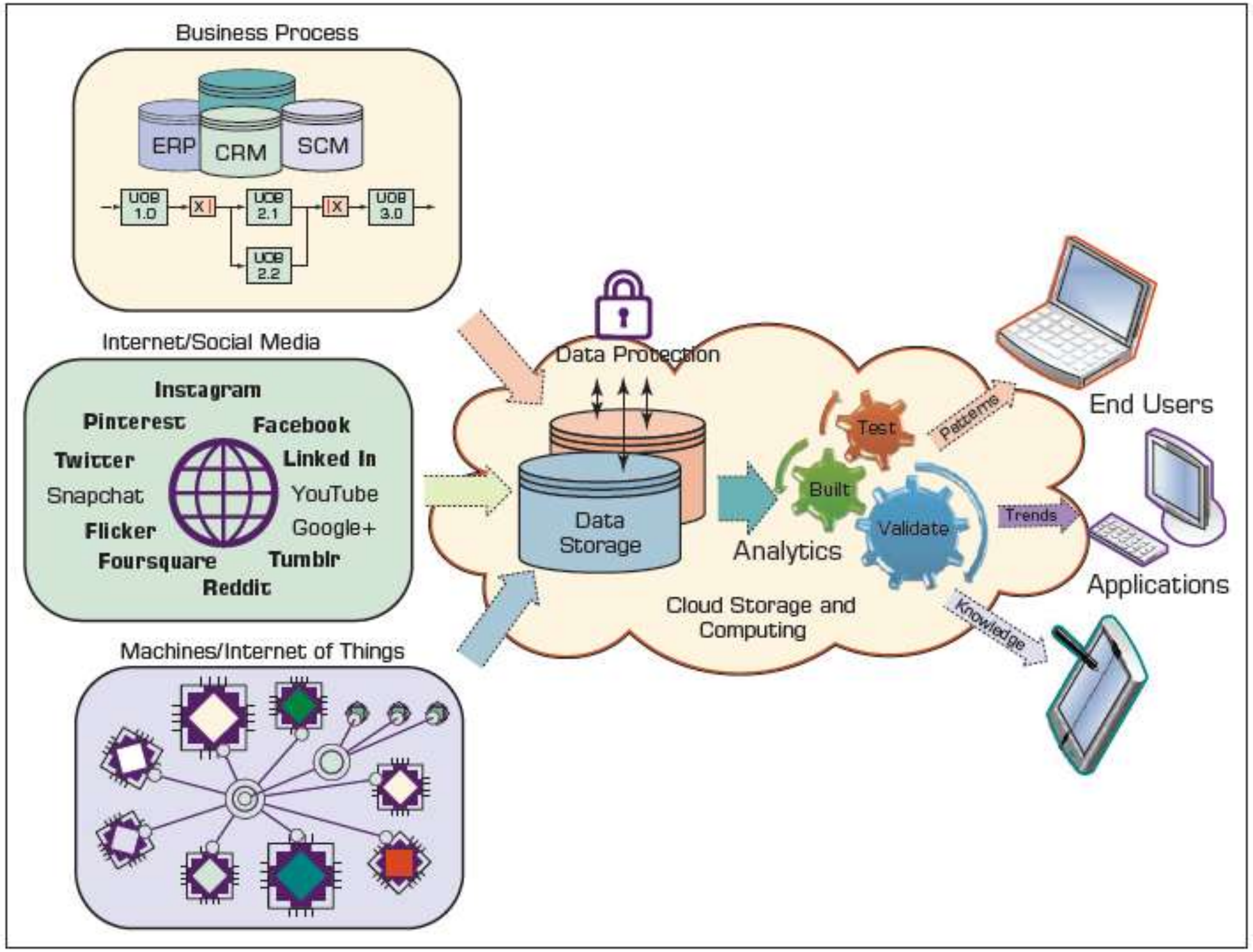


FIGURE 3.1 A Data to Knowledge Continuum.

Overview of the Nature of Data (cont.)

- Data must be analytics-ready, relevant, comply with quality/quantity metrics, and structured in place with key variables and normalized values.
- Predictive algorithms require a flat file and target variable; thus data analytics for prediction must be in that form.
- It is imperative to match the data to the needs and wants of a specific predictive algorithm and/or a software tool.

B. The Characteristics of Analytics – Ready Data

- **Source Reliability:** The originality/appropriateness of data storage medium
- **Content Accuracy:** Data is a good match for the analytics problem
- **Accessibility:** Data is easily and readily obtainable
- **Security & Privacy:** Data is secured, accessible to those with authorization
- **Richness:** All required data elements are included in the data set
- **Consistency:** Data is accurately collected and combined/merged
- **Currency/Timeliness:** Data is up-to-date for a given analytics model
- **Granularity:** variables are defined at the lowest level of detail for their use
- **Validity:** match/mismatch between actual & expected values of a variable
- **Relevancy:** variables in the data set are relevant to the conducted study

3.3 Simple Taxonomy of Data

- Taxonomy of Data Overview

Taxonomy of Data Overview

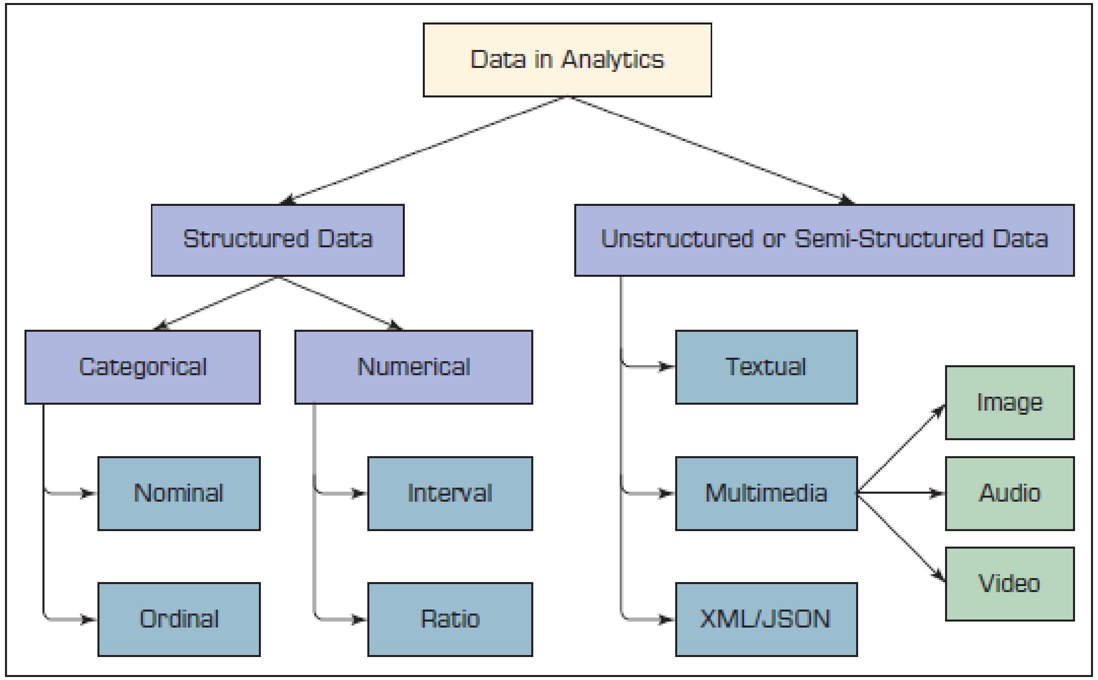
- The term **data** refers to a collection of facts usually obtained as the result of experiments, observations, transactions, or experiences.
- **Data** can be **classified** as structured, unstructured or semistructured. It can also be classified as static or dynamic (i.e., temporal or time series).
- **Unstructured/semistructured** data is composed of any combination of textual, imagery, voice, and Web content.
- **Data types** include textual, spatial, video/voice, and need to be converted into categorical/numeric representation before processing by analytics methods.
- **Predictive analytics/ data mining** methods and **machine-learning algorithms** are very selective about the type of data that they can handle.
- **Incompatible data** types lead to incorrect models or halt model development.

Taxonomy of Data Overview (cont.)

- **Categorical/Discrete Data:** labels of multiple classes used to divide a variable into specific groups, and have finite number of values with no continuum.

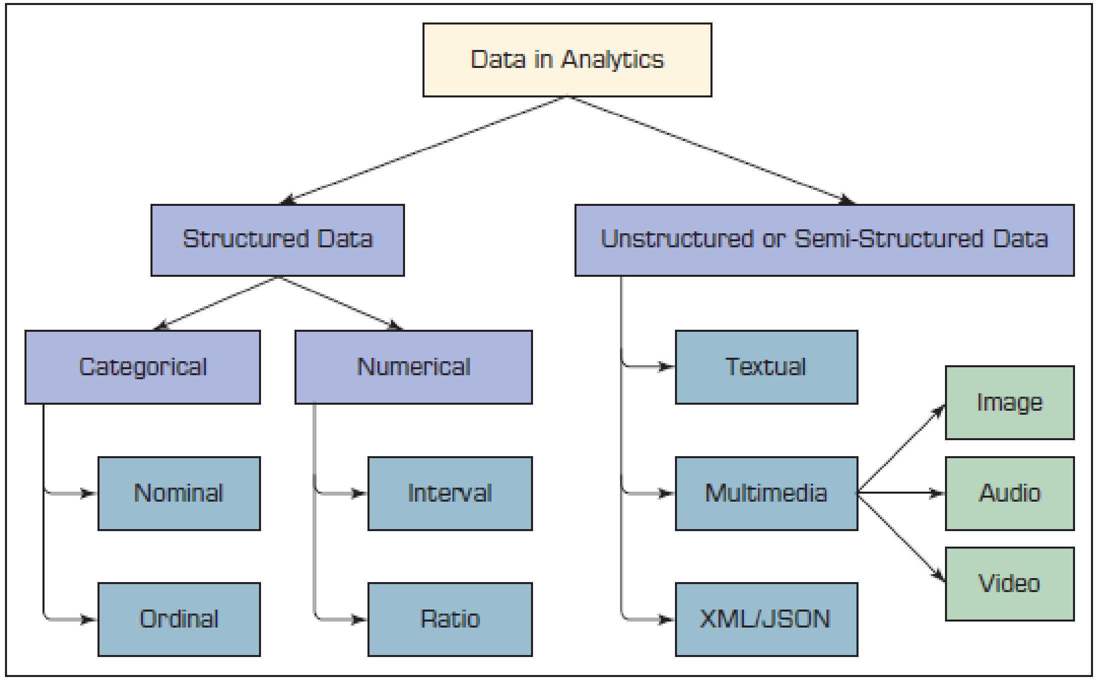
  - **Nominal Data:** measurements of simple codes assigned to objects as labels, which are not measurements. For example, the variable marital status.

  - **Ordinal Data:** codes assigned to objects or events as labels that also represent the rank order among them. For example, the pain scale.



Taxonomy of Data Overview (cont.)

- **Numeric Data**: measures on a specific scale that allow for interim values.
  - **Interval Data:** variables that can be measured on interval scales. An example is temperature Celsius scale as there is not an absolute zero value.
  - **Ratio Data**: the ratio between a magnitude of a continuous quantity and a unit magnitude (E.g. mass, length, time). The distinguishing feature is the possession of a nonarbitrary zero value.



3.4 Art & Science of Data Preprocessing

- Steps for Data Preprocessing

A. Steps for Data Preprocessing

- Data's original form can be inaccurate. Data preprocessing can convert the raw data to a form for analytics algorithms.

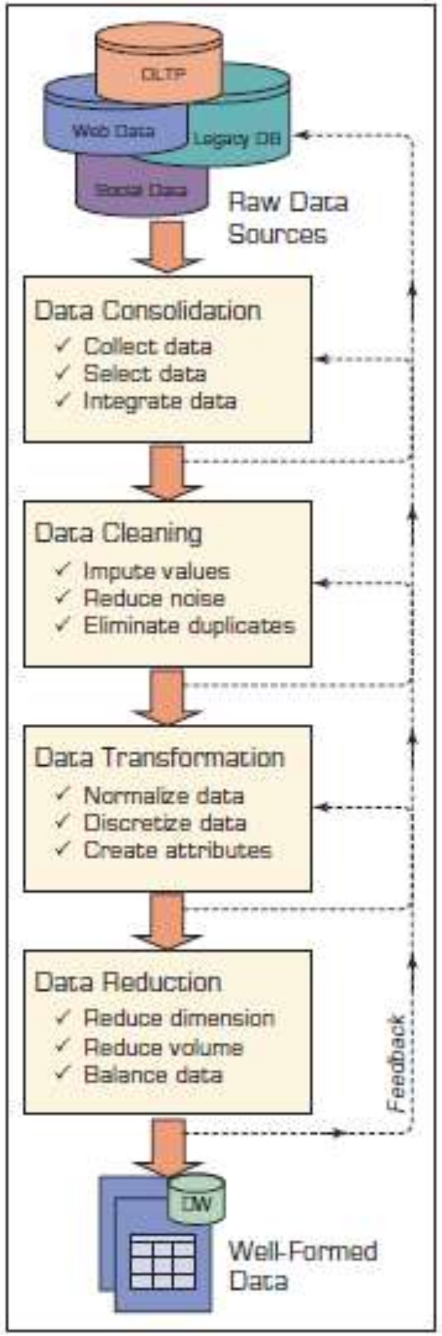
- **Steps of Data Preprocessing:**

1. Relevant data is collected, records/variables are selected, and the multiple data sources are integrated/merged.

2. Data is cleaned. Values are identified, while missing ones are imputed or ignored. Noisy values in the data are identified and smoothed out.

3. Data is transformed. It is normalized between a minimum and maximum for all variables to mitigate potential bias. Or it undergoes discretization.

4. Data reduction. One can use literature; consult experts; or a combination of techniques to reduce data dimensions to more manageable subsets.



3.5 Statistical Modelling for Business Analytics

- Statistics Overview
- Descriptive Statistics for Descriptive Analytics
- Measures of Centrality Tendency
- Arithmetic Mean
- Median
- Mode
- Measures of Dispersion
- Range
- Variance
- Standard Deviation
- Mean Absolute Deviation
- Quartiles and Interquartile Range
- Box-and-Whiskers Plot
- Shape of a Distribution

Statistics Overview

- **Statistics** —mathematical techniques to help interpret data—and statistical methods (E.g., regression, k-means clustering) is part of descriptive analytics.
- Descriptive analytics includes **statistics** & **online analytics processing** (OLAP)
- **OLAP:** Analyzing, characterizing, and summarizing structured data stored in organizational databases using cubes. The OLAP branch of descriptive analytics has also been called business intelligence.

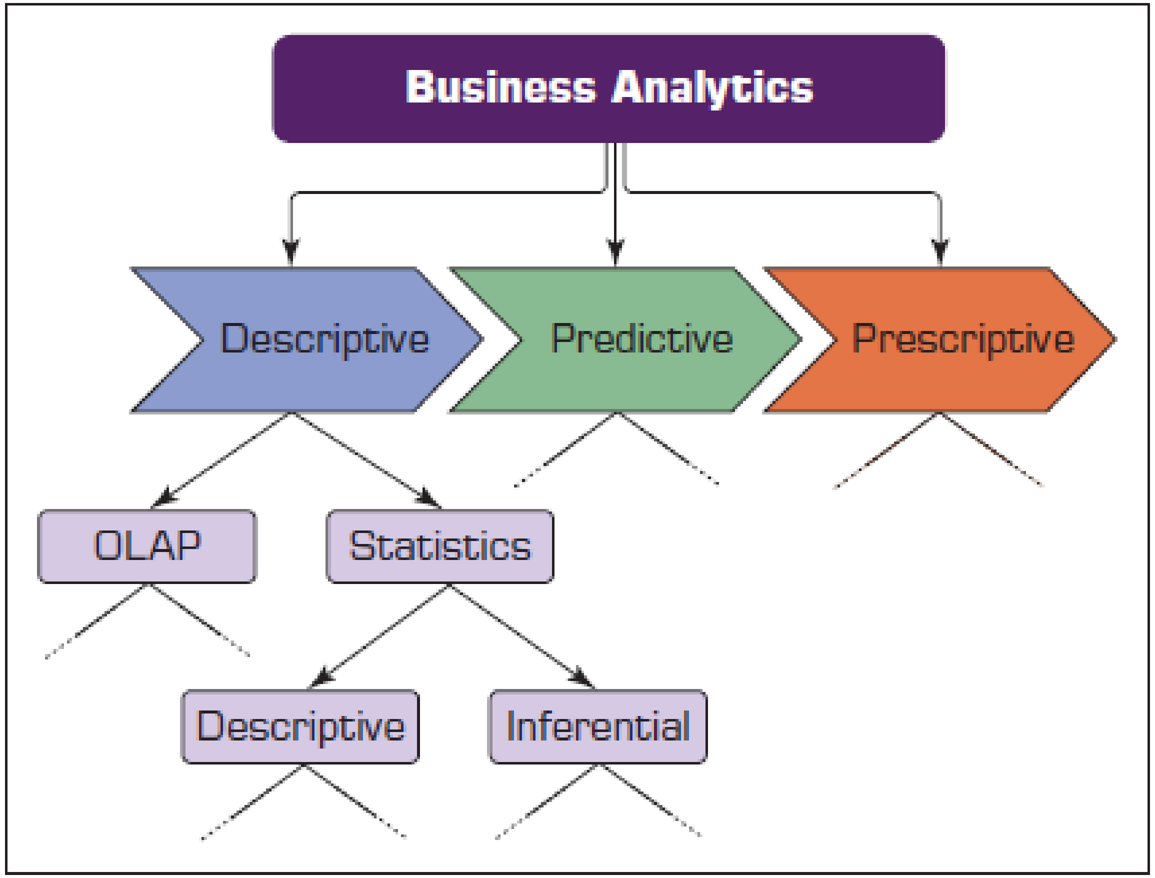


FIGURE 3.7 Relationship between Statistics and Descriptive Analytics.

Statistics Overview (cont.)

- **Statistics:** Helps to characterize data, either one variable at a time or multivariable, all together using either descriptive or inferential methods.
    - Descriptive statistics is all about describing the sample data.
    - Inferential statistics is about drawing inferences or conclusions about the characteristics of the population.

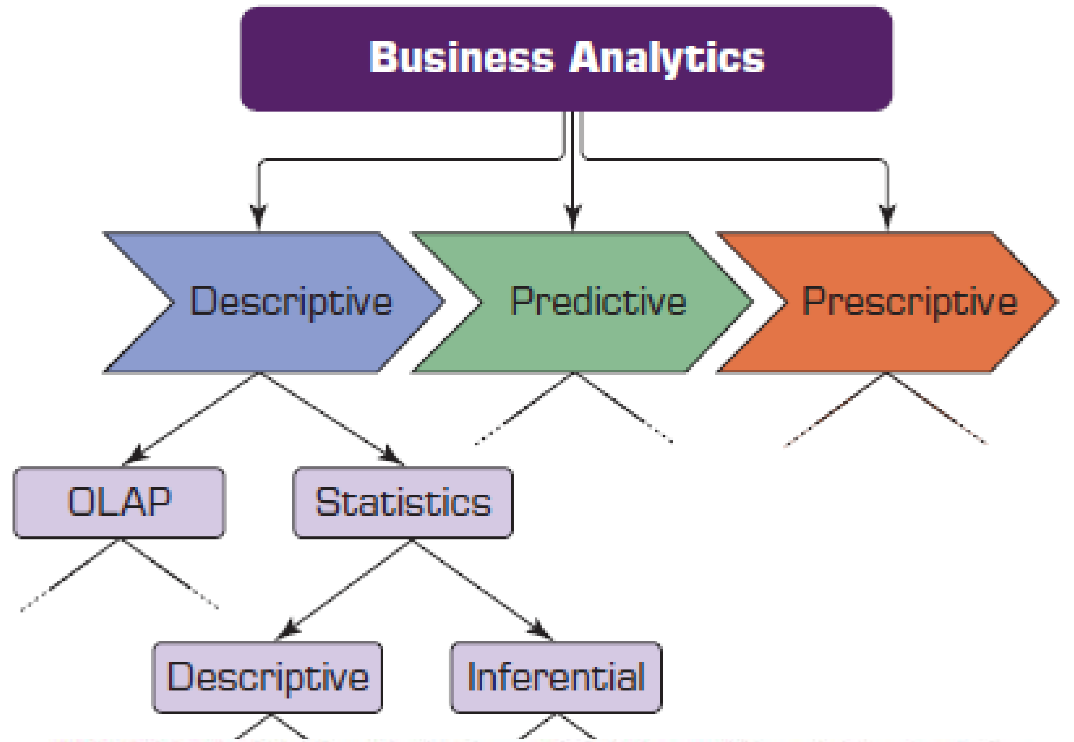


FIGURE 3.7 Relationship between Statistics and Descriptive Analytics.

Descriptive Statistics for Descriptive Analytics

- **Descriptive statistics** describes the basic characteristics of the data at hand.
- Descriptive statistics summarizes data to draw understandable patterns, and doesn’t allow making conclusions beyond the sample of the analyzed data.
- In business analytics, it helps understand and explain/present our data in a meaningful manner using aggregated numbers, data tables, or charts/graphs.
- This helps in the decision-making processes, and data scientists to characterize and validate the data for more sophisticated analytics tasks.
- Descriptive statistics allows analysts to identify data concertation, outliers, and unexpectedly distributed data values for numeric variables.
- Methods in descriptive statistics are classified as **measures of central tendency** or **measures of dispersion**.

Measures of Centrality Tendency

- Measures of centrality are the mathematical methods by which we estimate or describe central positioning of a given variable of interest.
- A measure of central tendency is a single numerical value that describes a set of data by simply identifying or estimating the central position within the data.
- The mean is the most commonly used measure of central tendency. Median and mode are also used to describe the centrality of a given variable.
- Although, the mean, median, and mode are all valid measures of central tendency, under different circumstances, one of these measures of centrality becomes more appropriate than the others.

Arithmetic Mean

- The **arithmetic mean** is the sum of all the values divided by the number of observations in the data set. It is used with continuous or discrete data.
- For a given variable x, if we happen to have n values, we can write the arithmetic mean of the data sample as follows:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- The mean is calculable for both interval- and ratio-type numeric data.
- One major downside is that the mean can be affected by outliers, which pull the mean in their direction and bias the centrality representation.

	Division
	1
	2
	2
	3
	3
	3
	4
	4
Sum	22
Avg/Mean	2.75
median	3
mode	3

Median

- The **median** is the number in the middle of a given set of data that has been arranged/sorted in order of magnitude.
- If the number of observations is an odd number just sort the observations based on their values and pick the value right in the middle.
- If the number of observations is an even number, identify the two middle values, and then take the simple average of these two values.
- The median is meaningful and calculable for ratio, interval, and ordinal data.
- Contrary to the mean, the median is not affected by outliers or skewed data.

	Division
	1
	2
	2
	3
	3
	3
	4
	4
Sum	22
Avg/Mean	2.75
median	3
mode	3

Mode

- The **mode** is the observation or value that occurs most frequently in a data set.
- The mode is most useful for data small data sets of unique values.
- Although useful for nominal data, mode is not a good representation of centrality, and should not be used as the only measure of central tendency.

	Division
	1
	2
	2
	3
	3
	3
	4
	4
Sum	22
Avg/Mean	2.75
median	3
mode	3

Which central tendency measure is the best?

- Use the mean when data doesn't have outliers or significant skewness
- Use the median when the data have outliers and/or it is ordinal in nature
- Use the mode when the data are nominal.
- Best practice is to use all three together.

	Division
	1
	2
	2
	3
	3
	3
	4
	4
Sum	22
Avg/Mean	2.75
median	3
mode	3

Measures of Dispersion

- **Measures of dispersion** are mathematical methods used to estimate the degree of variation in a given variable.
- To describe this dispersion, a number of statistical measures are developed; the most notable are range, variance, and standard deviation.
- The measures of data dispersion are important as they give us an indication of how well the centrality measures represent the sample data.
- If the dispersion of values is large, the mean is not a good representation of the data. Large dispersion indicates large differences between individual scores.
- It is a positive sign to see a small variation within each data sample, as it may indicate homogeneity, similarity, and robustness within the collected data.

Range

- The **range** is the simplest measure of dispersion. It is the difference between the largest and the smallest values in a given data set (i.e., variables).
- We calculate range by simply identifying the smallest value in the data set (minimum), identifying the largest value in the data set (maximum), and calculating the difference between them
- **Range = maximum - minimum**

	Division
	1
	2
	2
	3
	3
	3
	4
	4
Range	3

Variance

- **Variance** is a method used to calculate the deviation of all data points in a given data set from the mean.
- The larger variances indicate more dispersion, whereas smaller variances indicate compression.
- The formula for a data sample can be written as:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

- Because differences are squared, larger deviations from the mean contribute more to the value of variance.
- Therefore, instead of variance, we use a more meaningful dispersion measure, called standard deviation.

	Division
	1
	2
	2
	3
	3
	3
	4
	4
Range	3
Avg/Mean	2.75
Variance	1.07142857

Standard Deviation

- The **standard deviation** is a measure of spread of values within a set of data.
- It can be calculated by simply taking the square root of the variations.
- The following formula shows the standard deviation from a given sample of data points:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

	Division
	1
	2
	2
	3
	3
	3
	4
	4
Range	3
Avg/Mean	2.75
Variance	1.07142857
Standard Deviation	1.03509834

Mean Absolute Deviation

- **Mean absolute deviation** can also be used to measure dispersion in a data set. It is a simpler way to calculate the overall deviation from the mean.

- The mean absolute deviation is calculated by measuring the absolute values of the differences between each data point and the mean, then summing them.

- This process provides a measure of spread without being specific about the data point being lower or higher than the mean.

- The following formula shows the calculation of the mean absolute deviation:

$$MAD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

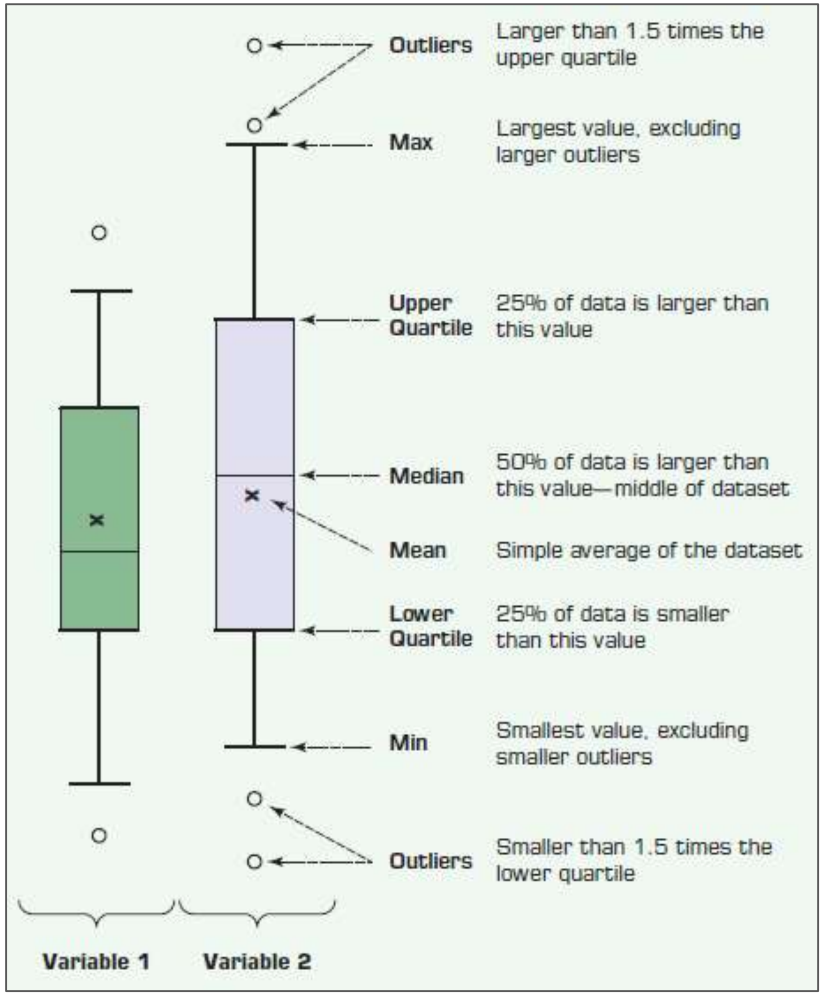
	Division
	1
	2
	2
	3
	3
	3
	4
	4
Range	3
Avg/Mean	2.75
Variance	1.07142857
Standard Deviation	1.03509834
Mean Absolute Deviation	0.8125

Quartiles and Interquartile Range

- Quartiles help identify spread within a subset of the data.
- A **quartile** is a quarter of the number of data points given in a data set; determined by sorting the data and splitting into four disjoint smaller sets.
- Quartiles are much less affected by outliers or skewness than the equivalent measures in the whole data set.
- A common way of expressing quartiles is as **an interquartile range**. This describes the difference between the third quartile (Q3) and the first quartile (Q1), telling us the range of the middle half of the scores in the distribution.
- The quartile-driven descriptive measures (both centrality and dispersion) are best explained with a popular plot called a **box-and-whiskers plot**.

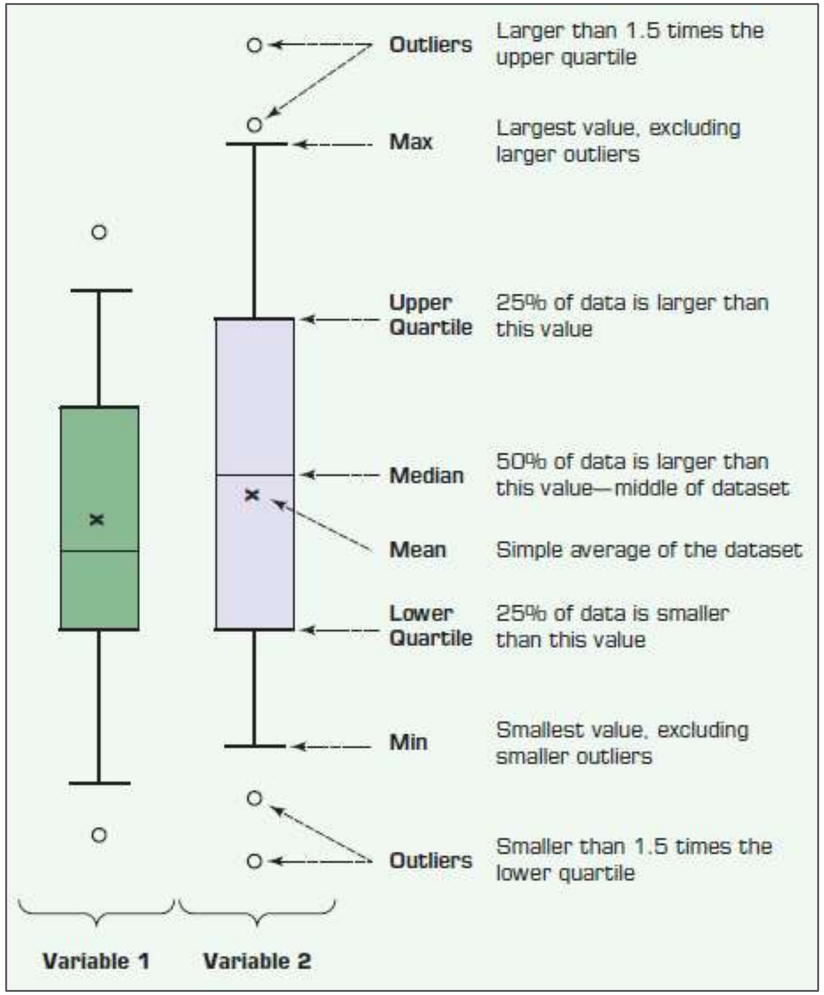
Box-and-Whiskers Plot

- The **box-and-whiskers plot** is a graphical illustration of several descriptive statistics about a given data set.
    - The y-axis is the measure of magnitude (the numerical value of the variable), and the x-axis shows different classes such as categories.
    - The plot shows the centrality (mean, median) as well as the dispersion (the density of the data within the middle)



Box-and-Whiskers Plot (cont.)

- The minimum and maximum ranges (shown as extended lines from the box; calculated as 1.5 times the upper or lower end of the quartile box)
- The outliers that are larger than the limits of the whiskers.
- The length of whiskers on the side of the box indicate the skewness.



Shape of a Distribution

- **Distribution** is the frequency of data points plotted over labels/numerical ranges. A normal distribution; it is symmetric on both sides of the mean.

- As dispersion increases, so does standard deviation, and distribution looks wider.

- There are two commonly used measures to calculate the shape characteristics of a distribution: **skewness** and **kurtosis**.

Shape of a Distribution (Skewness)

- **Skewness** is a measure of asymmetry in a unimodal distribution of the data.

$$Skewness = S = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{(n-1)s^3}$$

- If the distribution sways left, then it produces a positive skewness measure; if the distribution sways right, then it produces a negative skewness measure.

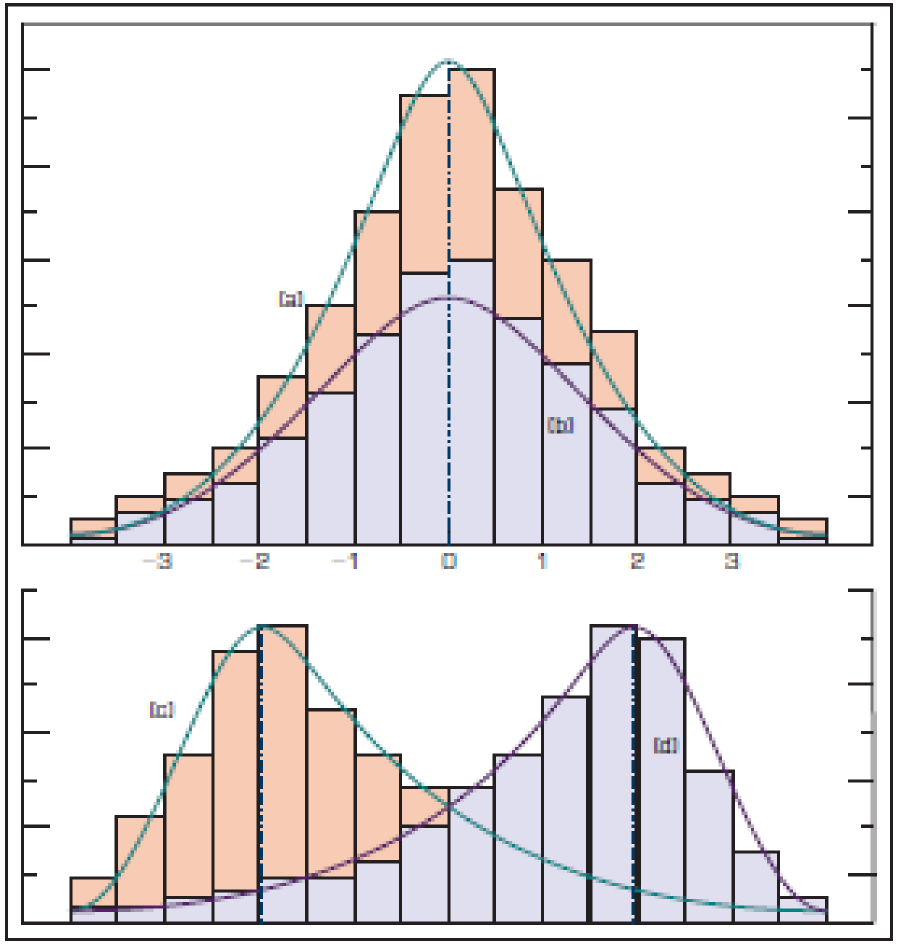


FIGURE 3.9 Relationship between Dispersion and Distribution Shape Properties.

Shape of a Distribution (Kurtosis)

- **Kurtosis** measures the degree to which a distribution is more or less peaked than a normal distribution.

- A positive kurtosis indicates a relatively peaked/tall distribution, a negative kurtosis indicates a relatively flat/short distribution. As a reference point, a normal distribution has a kurtosis of 3. The formula for kurtosis:

$$Kurtosis = K = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{ns^4} - 3$$

3.6 Regression Modelling for Inferential Statistics

- Overview
- How Do We Develop the Linear Regression Model?
- How Do We Know If the Model Is Good Enough?
- The Most Important Assumptions in Linear Regression
- Logistic Regression
- Time-Series Forecasting

Overview

- **Regression** is a statistical technique to model the dependence of a variable (response or output variable) on one (or more) explanatory (input) variables.
- Relationships between variables can be represented as a linear function.
- Regression captures the functional relationship between characteristics of the real world and describe this relationship with a mathematical model.
- It can be used for:
  - **Hypothesis Testing:** Investigating the potential relationships between different variables
  - **Prediction/Forecasting**: Estimating response variable values based on explanatory variables

Overview

- **Correlation vs. Regression:**
  - **Correlation** makes no a prior assumption of variable dependency and gives an estimate on the degree of association between the variables.
  - **Regression** describes the dependence of a response variable on explanatory variables where it assumes a causal effect from the explanatory variable(s) to the response variable.
- **Simple vs. Multiple Regression:**
  - **Simple regression** is when the equation is built between one response variable and one explanatory variable.
  - **Multiple regression** is the extension of simple regression when the explanatory variables are more than one.

How Do We Develop the Linear Regression Model?

- To understand the relationship between two variables, draw a scatter plot where the y-axis is the response variable values and the x-axis is the explanatory variable values.

- **Simple regression** tries to find the signature of a straight line passing through the plotted dots in a way that minimizes distance between dots and line.

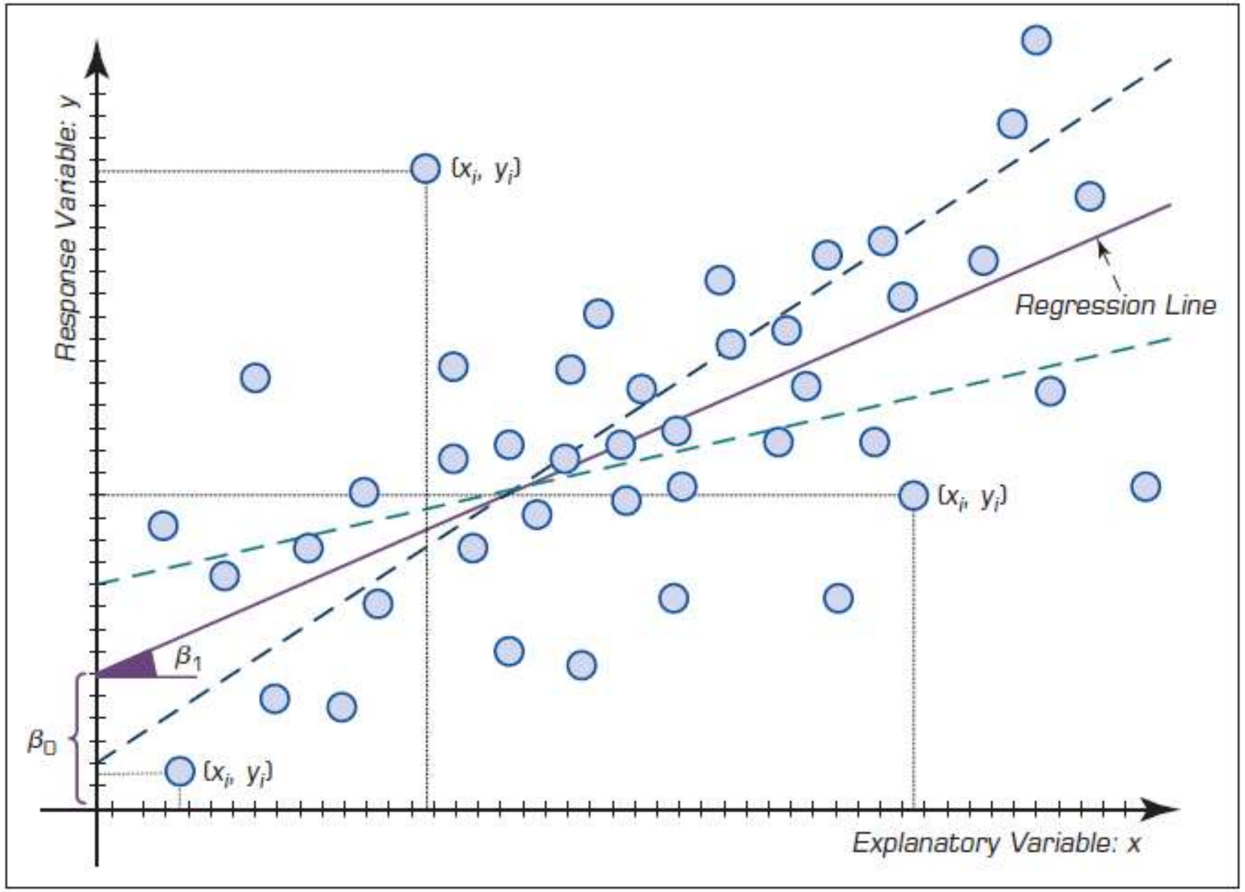


FIGURE 3.13 A Scatter Plot and a Linear Regression Line.

How Do We Develop the Linear Regression Model?

- The most commonly used method to identify regression is the **ordinary least squares** (OLS).

- OLS aims to minimize the sum of squared residuals and leads to a mathematical expression for the estimated value of the regression line (b parameters).

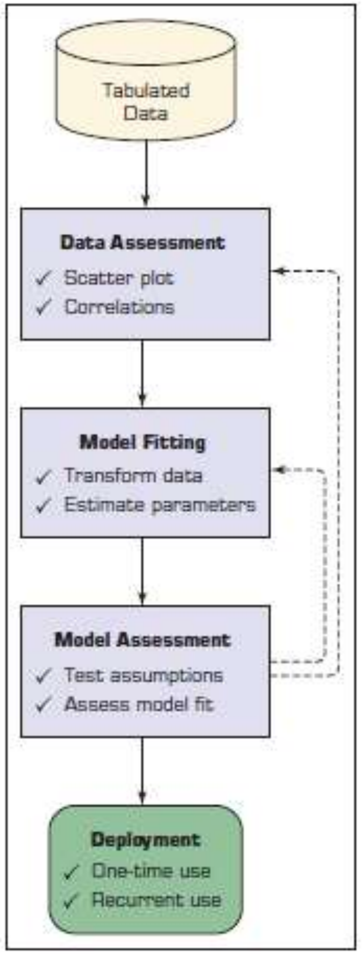


FIGURE 3.14 A Process Flow for Developing Regression Models.

How Do We Develop the Linear Regression Model?

- **Simple linear regression**: $y = \beta_0 + \beta_1 x$

- **Ordinary Least Squares methods** (OLS) determines values of two coefficients, the simple equation is use β_1 o forecast the values of y for given values of x.

- The sign and the value of reveal the direction and the strengths of relationship between the two variables.

β_i

- If the model is of a multiple linear regression type, then there would be one coefficient multiplied with each additional explanatory variable. Then, summed together to establish:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdot + \beta_n x_n$$

How Do We Know If the Model Is Good Enough?

- Regression model must be assessed for the degree it represents the response variable. A fitting regression model predicts values close to observed values.
- For numerical assessment, R^2 , overall F-test, and root mean square error (RMSE) are used in evaluating the fit of a regression model
- R^2 has an intuitive scale, and ranges from 0 to 1; 0 indicates the model's relationship/prediction is not good, and 1 indicates a perfect fit.
- An R^2 value of 0.3 in social sciences can be considered good enough fit, but a 0.7 in engineering might be considered as not a good enough fit.
- The improvement in the regression model can be achieved by adding more explanatory variables or using different data transformation techniques, which would result in comparative increases in an R^2 value.

The Most Important Assumptions in Linear Regression

- Linear regression models have highly restrictive assumptions. The validity of the linear model depends on ability to comply with the assumptions:

1. **Linearity:** relationship between response and explanatory variables is linear.
2. **Error Independence:** Response variable errors are uncorrelated.
3. **Error Normality:** The errors of the response variable are normally distributed.
4. **Constant Variance/Homoscedasticity:** The response variables have the same variance in their error regardless of explanatory variables' values.
5. **Multicollinearity:** The explanatory variables are not correlated. This can be triggered by having two or more perfectly correlated explanatory variables presented to the model.

Logistic Regression

- **Logistic regression** is a probability-based classification algorithm using supervised learning capable of predicting **multiclass** output variables.
- In predictive analytics, logistic regression models develop probabilistic models between one or more explanatory variables and a response variable.
- Logistic regression takes the natural logarithm of the odds of the response variable to create a continuous criterion.
- The logistic regression coefficients are estimated using the maximum likelihood method. It is not possible to find a closed-form expression for the coefficient values that maximizes the likelihood function.
- Thus, a starting solution is proposed, then parameters are continuously revised for improvement, until no improvement can be achieved.

Logistic Regression

- The **logistic function**, f(y) is the core of logistic regression,
- Which can take values only between 0 and 1.
- The following equation is a simple mathematical representation of this function:

$$f(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

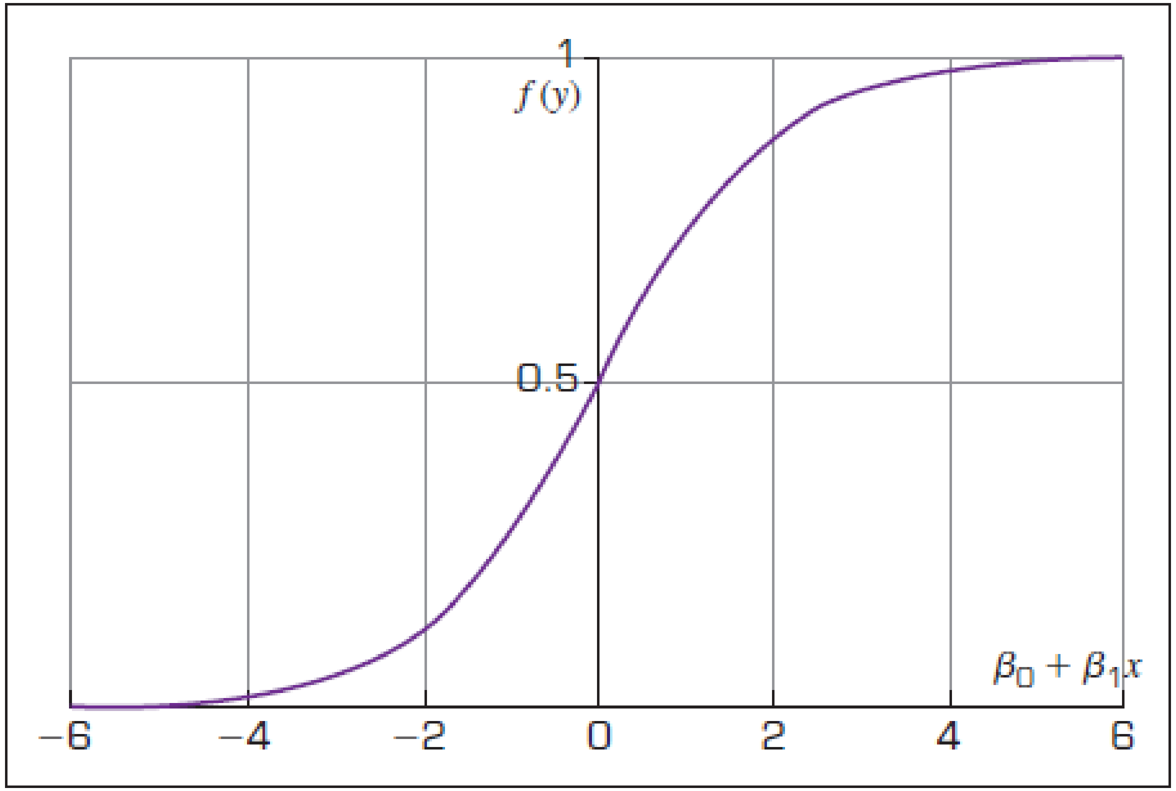


FIGURE 3.15 The Logistic Function.

Time-Series Forecasting

- A **time series** is a sequence of data points of the variable of interest, measured and represented at successive points in time spaced at uniform time intervals.
- A time series is developed when the variable we are interested in might not have distinctly identifiable explanatory variables.

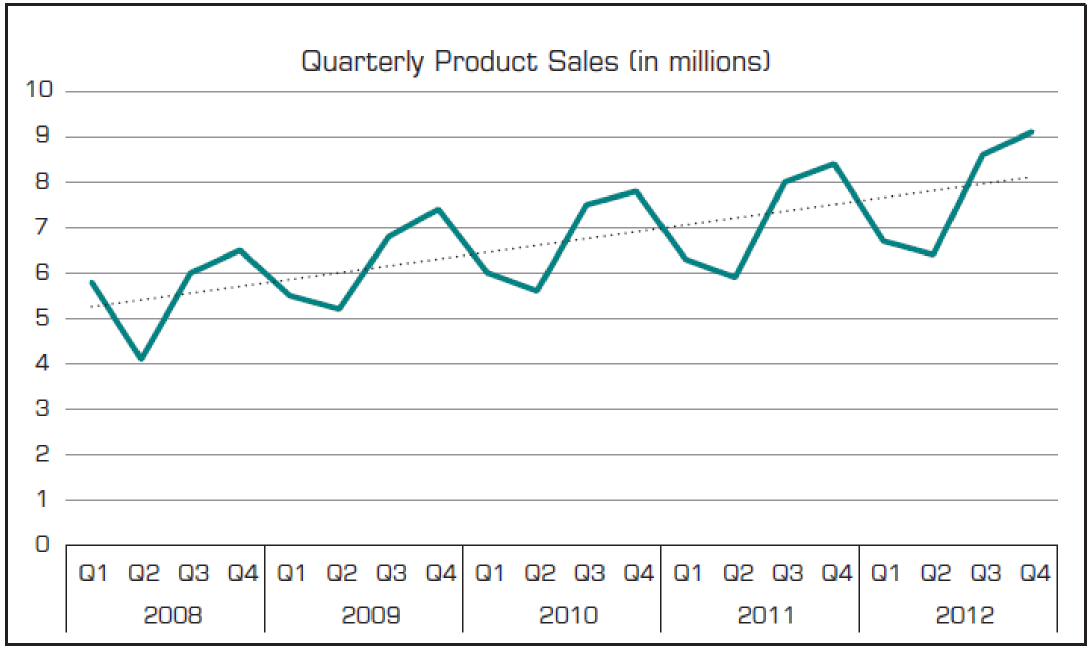


FIGURE 3.17 A Sample Time Series of Data on Quarterly Sales Volumes.

Time-Series Forecasting

- **Time-series forecasting** is the use of mathematical modeling to predict future values of the variable of interest based on previously observed values.

- Time-series forecasting techniques are the averaging methods: simple average, moving average, weighted moving average, and exponential smoothing.

- The accuracy of a method is usually assessed by computing its error via mean absolute error, mean squared error, or mean absolute percent error.

Practical Portion

Chapter 3, 5, 11 & 13: Getting Started with Data Analysis

Excel Data Analysis For Dummies

Practical Portion

Chapter 3: Introducing Excel Tables

Excel Data Analysis For Dummies

Chapter 3: Introducing Excel Tables

- Concepts
- Terms
- Building a Table
- Quick Statistical Measures Available on the Status Bar
- Analyzing Table Information
- Adding a column subtotal
- Sorting table records
- Filtering table records
- Clearing a Filter
- Applying a predefined AutoFilter

Concepts

Basic Concepts:

- **A table is a kind of database:** Microsoft Access is a powerful database management tool that lets you work with large, complex databases.
- **A table has its advantages:** Because a table is a collection of rows and columns on a worksheet, it looks very much like a regular Excel range
- **A table makes data analysis easier:** Tables are also useful tools for analyzing your data.

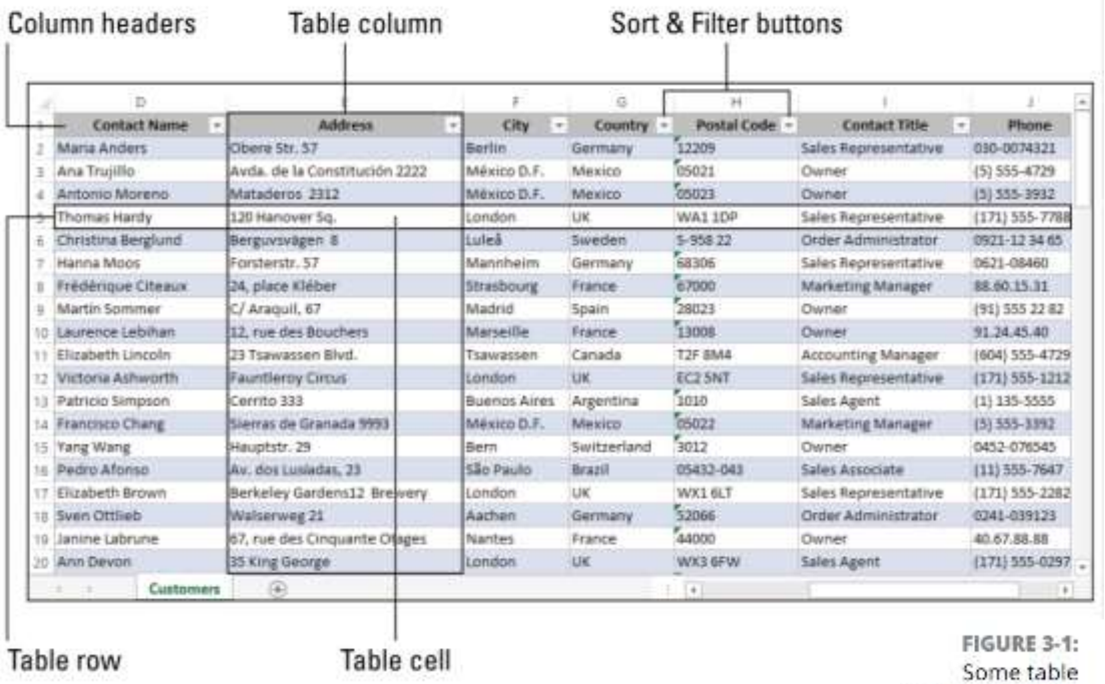


FIGURE 3-1: Some table terminology you should know.

Terms

Table Terms:

- **Table column:** A single type of information, such as names, addresses, or phone numbers. In an Excel table, each column is the equivalent of a data-base field.
- **Table row:** A set of associated table cells, such as the data for a single contact. In an Excel table, each row is the equivalent of a database record.
- **Table cell:** An item in a table column that represents a single instance of that column's data, such as a name, address, or phone number. In an Excel table, each cell is equivalent to a database field value.
- **Headers:** The unique names you assign to every table column that serve to label the data in each column. These names are always found in the first row of the table.
- **Sort & Filter buttons:** An Excel feature that gives you access to a set of commands that perform various actions on a column, such as sorting or filtering the column data.

Building a Table

- **Getting the data from an external source:** The usual way to create an Excel table is to import the information from an external source, such as another workbook, an - Access database, a text file, or even a webpage.

Converting a range to a table:

- Decide whether you want your table to have column headers.
- Column headers must be unique and must be text or text formulas.
- Excel can often automatically identify the size and shape of the range that contains your data.

Note: to avoid confusing Excel, make sure that you have no blank rows or columns in your range.

Building a Table (cont.)

These images are for illustration only. Open Excel file ‘DS498_week4-Ch3’ and apply the steps:

1. Select a cell within the range that you want to convert to a table.

2. Choose Insert ➪ Tables ➪ Table or press Ctrl+T.

3. If Excel got it wrong, drag over the correct range.

4. If your range has labels that you want as column headers, make sure the ‘My Table Has Headers’ checkbox is selected.

5. Click OK

Note: The Design tab is displayed whenever you select a cell anywhere inside the table. It is full of useful table tools.

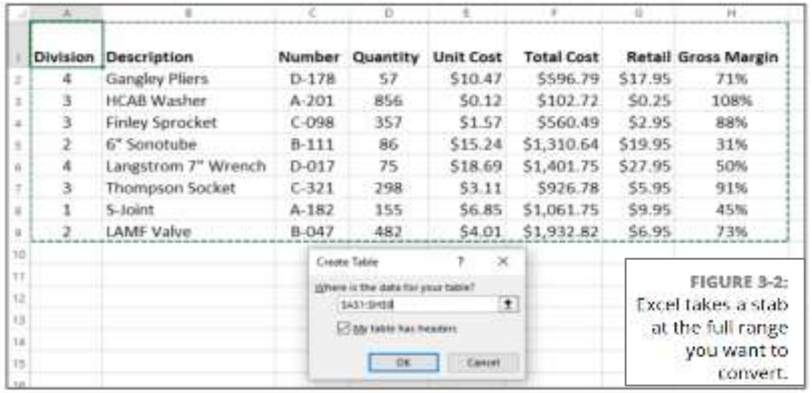


FIGURE 3-2: Excel takes a stab at the full range you want to convert.

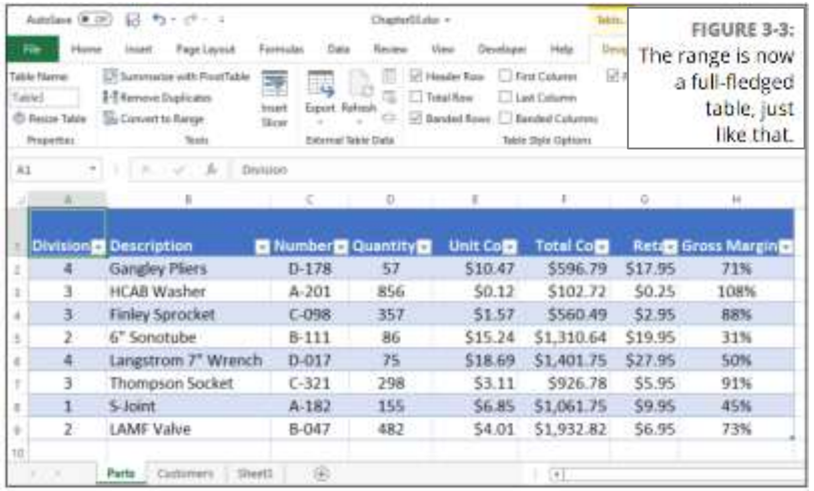


FIGURE 3-3: The range is now a full-fledged table, just like that.

Quick Statistical Measures Available on the Status Bar

Measure	What It Displays
Average	The average of the cells in the selected range.
Count	The number of cells that hold labels, values, or formulas. In other words, use this statistical measure when you want to count the number of cells that are not empty.
Numerical Count	The number of cells in a selected range that hold values or formulas.
Minimum	The smallest value in the selected range.
Maximum	The largest value in the selected range.
Sum	The total of the values in the selected range.

Analyzing Table Information

- **Displaying simple statistics:**

  Average: 295.75 Count: 8 Sum: 2366

- Excel displays the status bar statistics when you select any range, so you can get those stats without bothering to convert a range to a table

Division	Description	Number	Quantity	Unit Co	Total Co	Reta	Gross Margin
4	Gangley Pliers	D-178	57	$10.47	$596.79	$17.95	71%
3	HCAB Washer	A-201	856	$0.12	$102.72	$0.25	108%
3	Finley Sprocket	C-098	357	$1.57	$560.49	$2.95	88%
2	6" Sonotube	B-111	86	$15.24	$1,310.64	$19.95	31%
4	Langstrom 7" Wrench	D-017	75	$18.69	$1,401.75	$27.95	50%
3	Thompson Socket	C-321	298	$3.11	$926.78	$5.95	91%
1	S-Joint	A-182	155	$6.85	$1,061.75	$9.95	45%
2	LAMF Valve	B-047	482	$4.01	$1,932.82	$6.95	73%

FIGURE 3-4: Select a column's cells, and Excel displays a few stats in the status bar.

Adding a column subtotal

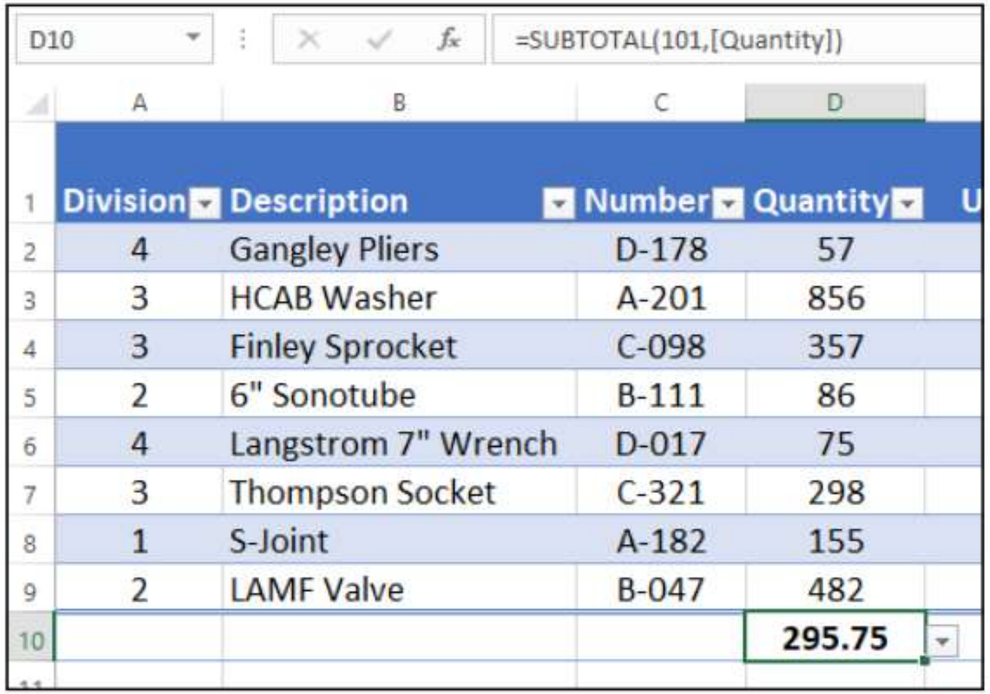
The image are for illustration only. Open Excel file ‘DS498_week4-Ch3’ and apply the steps:

1. Select the data in the column you want to total.

2. Click the Quick Analysis smart tag or press Ctrl + Q. The Quick Analysis options appear.

3. Click the Totals tab.

4. Select the type of calculation you want to use.



D10 =SUBTOTAL(101,[Quantity])

	A	B	C	D
1	Division	Description	Number	Quantity
2	4	Gangley Pliers	D-178	57
3	3	HCAB Washer	A-201	856
4	3	Finley Sprocket	C-098	357
5	2	6" Sonotube	B-111	86
6	4	Langstrom 7" Wrench	D-017	75
7	3	Thompson Socket	C-321	298
8	1	S-Joint	A-182	155
9	2	LAMF Valve	B-047	482
10				295.75

FIGURE 3-5: Excel adds a total row to the bottom of the table and displays the result of the calculation.

Technology Insights: How to Calculate Descriptive Statistics in Microsoft Excel

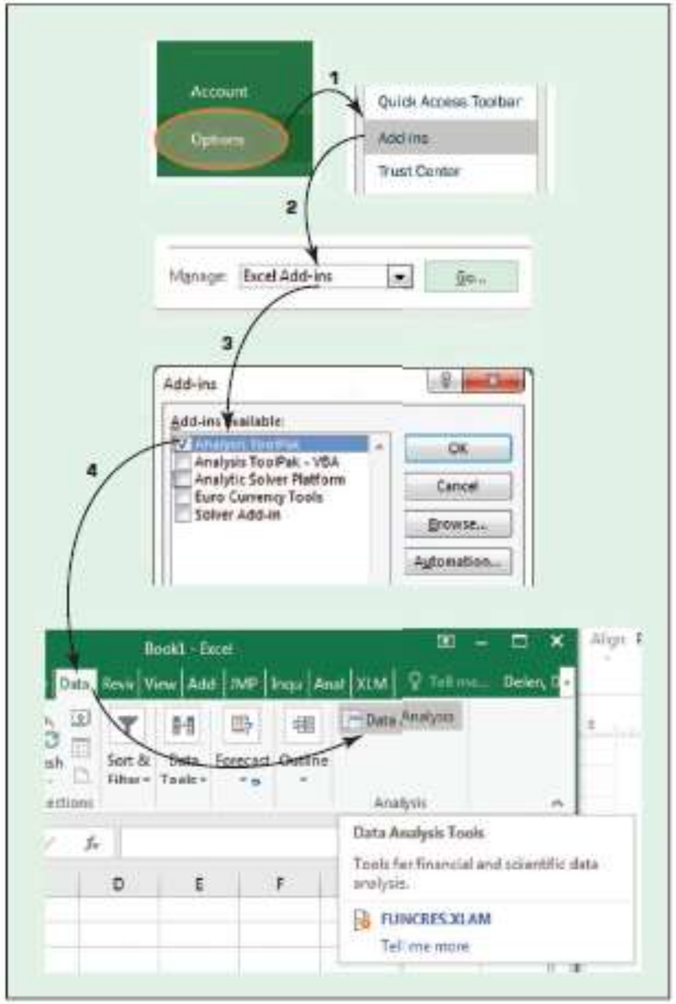


FIGURE 3.10 Activating Statistics Function in Excel 2016.

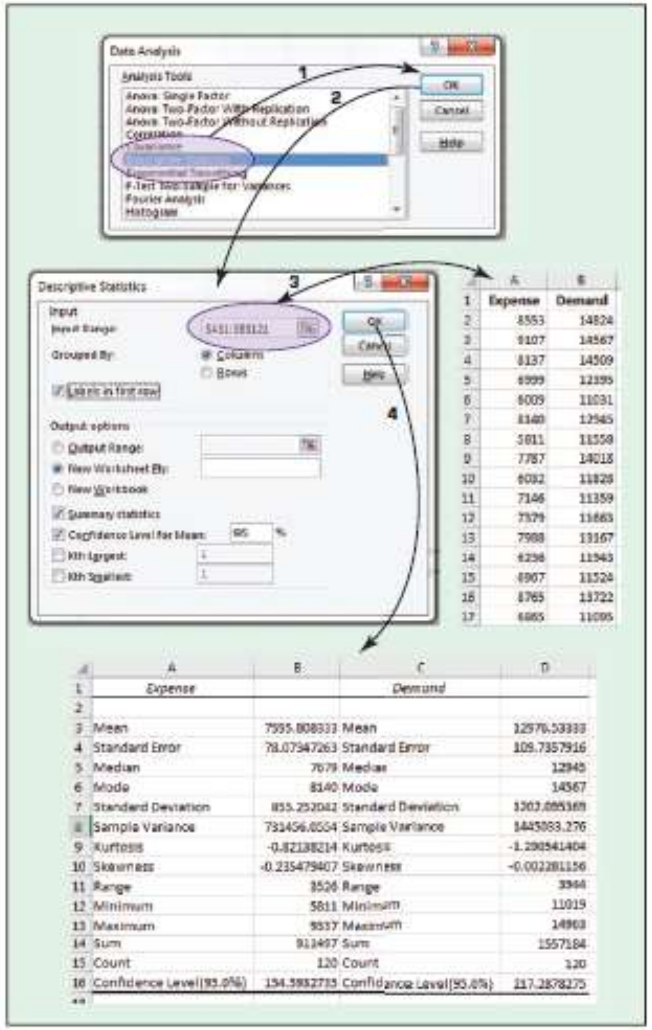


FIGURE 3.11 Obtaining Descriptive Statistics in Excel.

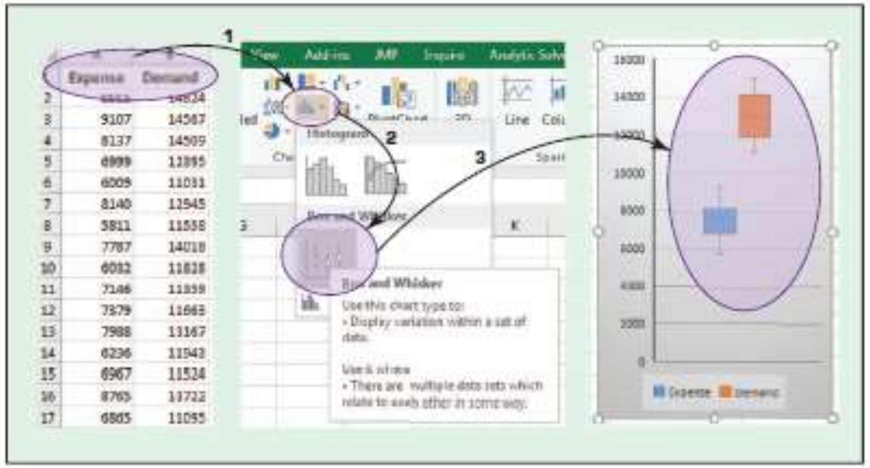


FIGURE 3.12 Creating a Box-and-Whiskers Plot in Excel 2016.

Sorting table records

- Sorting enables you to get a feel for how your data is distributed overall. For example, you might notice that most of the values cluster around the low end of the range of values.

- Sorting enables you to identify certain types of trends in the data. For example, you might notice that records (that is, rows) with low values in the sorted column all come from the same geographic area, or that high values in another table all come from the same division of the company.

- Sorting enables you to identify outliers, which are data points that are significantly outside the norm for your data. For example, if your sort shows that most of your column values lie between 1 and 100, but one row contains the value 250, you'll want to investigate why that value is so much greater than the others

Sorting table records

To sort a table based on the values in a column, follow these steps:

1. Click the Sort & Filter button for the column you want to sort.
   - Excel displays the Sort & Filter menu.

2. Select the sort option you want:
   - Sort Smallest to Largest: Sorts the column values in ascending numeric order.
   - Sort Largest to Smallest: Sorts the column values in descending numeric order

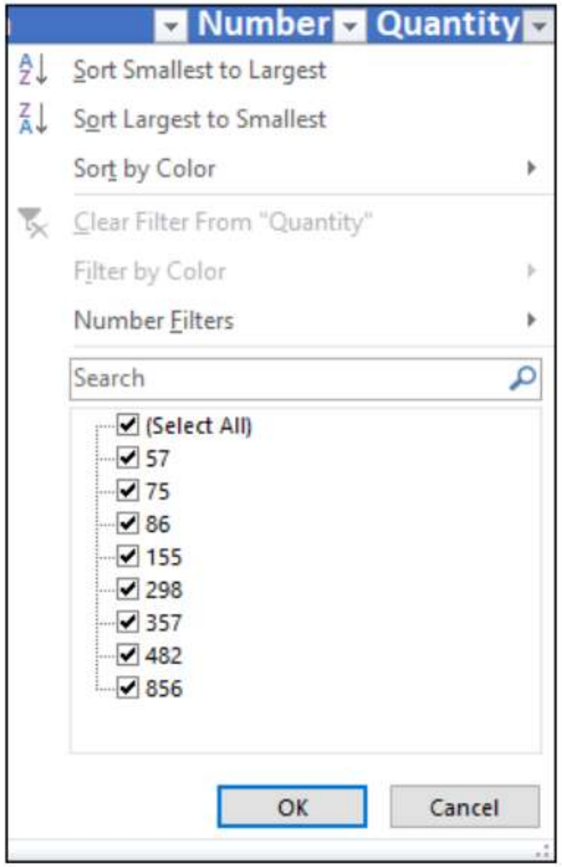


FIGURE 3-6: Selecting the Quantity column's Sort & Filter button displays this menu.

Sorting table records (cont.)

To use the Custom Sort command, follow - these steps:

1. Click any Sort & Filter button in the table.

    Excel displays the Sort & Filter menu.

2. Choose Sort by Color ➪ Custom Sort.

    Excel displays the Sort dialog box.

3. Use the Sort By drop-down list to select the field that you want to use for sorting.
4. Use the Sort On list to select Cell Values.
5. Use the Order list to select a sort order.

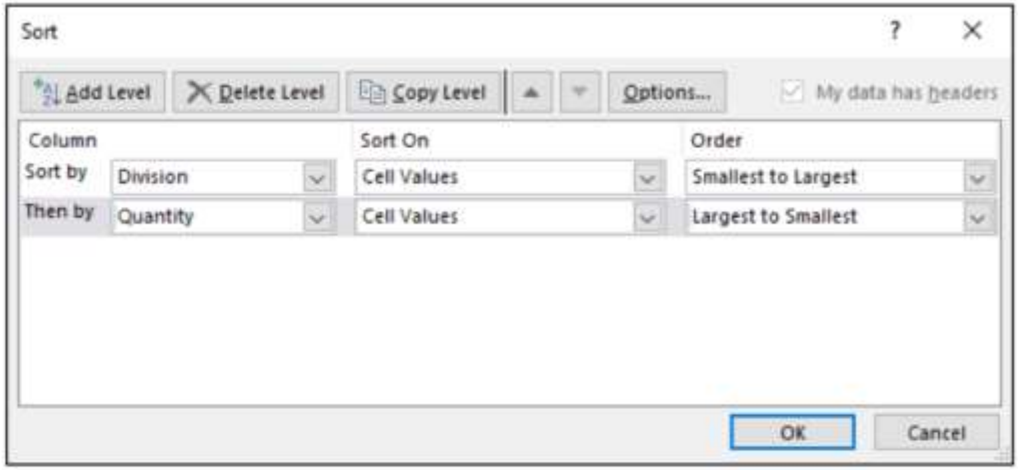


FIGURE 3-7: The Sort dialog box set up for a two-level sort.

Note: If you're using conditional formatting, you can also choose to sort on **Cell Color, Font Color**, or Conditional Formatting Icon.

Filtering table records

To apply an Auto Filter to a table, follow these steps:

1. Click the Sort & Filter button for the column you want to filter.

   Excel displays the column's Sort & Filter menu. Above the OK and Cancel buttons, you see a list of check boxes, where the name of each check box is a unique value from the column.

2. Deselect the Select All check box to deselect all the check boxes.

3. Select the check box for each column value you want to see in the filtered table.

   Sort & Filter menu for the Division column, with only the 3 check box selected.

4. Click OK.

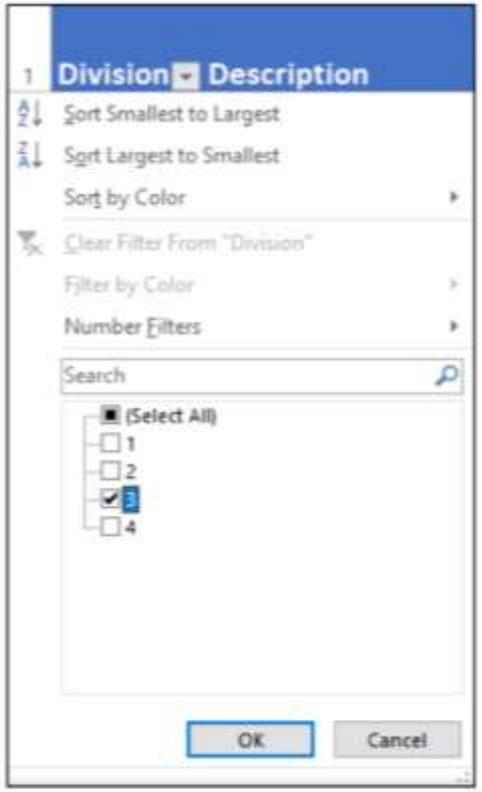


FIGURE 3-8: Select the check box beside each column value that you want to include in your filtered table.

Clearing a filter

Filtered icon

FIGURE 3-9: The Parts table filtered to show only the records from Division 3.

	A	B	C	D	E	F	G	H
1	Division	Description	Number	Quantity	Unit Co	Total Co	Reta	Gross Margin
3	3	HCAB Washer	A-201	856	$0.12	$102.72	$0.25	108%
4	3	Finley Sprocket	C-098	357	$1.57	$560.49	$2.95	88%
7	3	Thompson Socket	C-321	298	$3.11	$926.78	$5.95	91%

1. To clear a filter from a single column, select the column's Sort & Filter button
2. Then select the Clear Filter From "Column" command from the menu (where Column is the name of the filtered column).

- If you have multiple filters applied to the table, you can clear all the filters in one fell swoop by choosing Data ➪ Clear (look for it in the Sort & Filter group)

Applying a predefined AutoFilter

1. Click the Sort & Filter button for the column you want to filter.

   Excel displays the column's Sort & Filter menu.

2. Choose the X Filters command, where X refers to the type of data in the column.

   You have three possible commands:

   Number Filters: Appears when the column contains numeric data.

   Date Filters: Appears when the column contains date values, time values, or both.

   Text Filters: Appears when the column contains text data.

Applying a predefined AutoFilter (cont.)

3. Select the filter operator you want to apply.

   Excel displays a menu of predefined filter operators. The contents of the menu depend on the data type of your column. For example, if your column contains numeric data, you see the following operators: Equals, Greater Than, Less Than

   Custom Filter: Displays the Custom AutoFilter dialog box, which enables you to create your own filter condition.

4. Complete the AutoFilter condition.

5. Click OK.

   Excel filters your table according to your custom AutoFilter.

FIGURE 3-10: The Custom AutoFilter dialog box.

FIGURE 3-11: The filter shown in Figure 3-10 applied to the Parts table.

Division	Description	Number	Quantity	Unit Co	Total Co	Reta	Gross Margin
3	HCAB Washer	A-201	856	$0.12	$102.72	$0.25	108%
3	Finley Sprocket	C-098	357	$1.57	$560.49	$2.95	88%
3	Thompson Socket	C-321	298	$3.11	$926.78	$5.95	91%
1	S-Joint	A-182	155	$6.85	$1,061.75	$9.95	45%
2	LAMF Valve	B-047	482	$4.01	$1,932.82	$6.95	73%

Practical Portion

Chapter 5: Scrub-a-Dub-Dub: Cleaning Data

Excel Data Analysis For Dummies

Chapter 5: Scrub-a-Dub-Dub: Cleaning Data

- Editing Your Imported Workbook
- Deleting columns/rows & Erasing Content
- Formatting numeric values
- Replacing data in fields
- The CLEAN & CONCAT function
- The EXACT & LEN function
- The FIND & LOWER function
- The MID function
- The NUMBERVALUE function
- The PROPER & REPLACE function
- The SEARCH & UPPER function
- The SUBSTITUTE & VALUE function
- The TEXT & TRIM function
- The TEXTJOIN function
- Converting text function formulas to text
- Using Validation to Keep Data Clean

Editing Your Imported Workbook

If you take a look at the workbook DS498_week4-Ch5 shown in Figure 5-1, you see that the data, although some-what neatly formatted, suffers from quite a few problems:

- The data is a regular range, not an Excel table.
- The workbook has several blank rows and a blank column.
- The “numbers” you see in columns B, C, and F are text values. You see an error indicator in each cell, and when you click the error icon, Excel tells you the cell contains a number formatted as text.
- The column are poorly sized for the data they contain. For example, columns A, B, C, and F are too wide, whereas column E is too narrow (which is why those pound signs (#) appear in some cells)

	A	B	C	D	E	F
1	Nuclear Power By Country (2016)					
2	Country	Reactors	Megawatts		Gigaw	Energy Share
3						
4	Argentina	3	1633		5716	4.5%
5	Armenia	1	375		2411	32.5%
6	Belgium	7	5918		####	49.9%
7	Brazil	2	1884		####	2.7%
8	Bulgaria	2	1926		####	34.3%
9	Canada	19	13554		####	14.6%
10	China	39	34514		####	3.9%
11	Czech Republic	6	3930		####	33.1%
12						
13	Finland	4	2769		####	33.2%
14	France	58	63130		####	71.6%
15	Germany	8	10799		####	11.6%
16	Hungary	4	1889		####	50.0%
17	India	22	6255		####	3.2%
18	Iran	1	915		6366	2.2%
19	Japan	42	39752		####	3.6%
20						
21	Korea, Republic of	25	23070		####	27.1%
22	Mexico	2	1552		####	6.0%

FIGURE 5-1: This worksheet needs a good scrubbing.

Deleting columns/rows & Erasing Content

- To **delete unnecessary columns**, either
  - click the column header to select the entire column and then choose Home ➪ Delete, or
  - right-click the column header and then click Delete.
- To **delete unnecessary rows**, either
  - click the row header to select the entire row and then choose Home ➪ Delete, or
  - right-click the row header and then click Delete.
- To **erase the contents of a cell or range** that contains data you don't need, select the worksheet cell or range and then choose Home ➪ Clear ➪ Clear All.

  Excel erases both the contents of the cells in the selected range and any formatting assigned to those cells

Formatting numeric values

- Select the range you want to format, click the Number Format drop-down list on the Home tab, and then select the format you want to apply.
- **OR**

1. Select the range you want to format,
2. click the Number Format drop-down list on the Home tab,
3. then select More Number Formats.

   Excel displays the Format Cells dialog box with the Number tab displayed.
4. Select a category and then select among the options to specify the format you want to use. While you're here

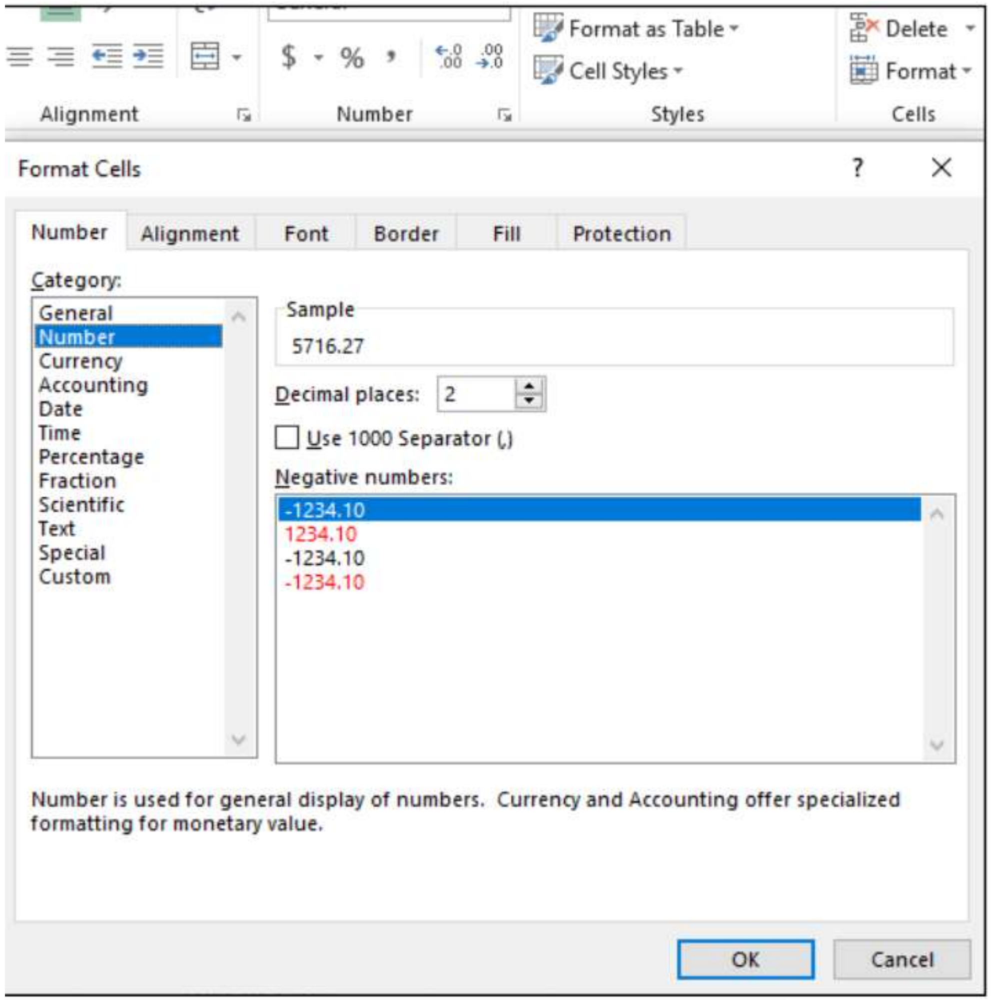


FIGURE 5-2: Format your numeric values here.

Replacing data in fields

Some common reasons to replace text include:

- repeated characters,
- Misspelled words,
- Abbreviations or acronyms

1. To use Replace, choose Home ➪ Find & Select ➪ Replace to open the Find and Replace dialog box with the Replace tab displayed.

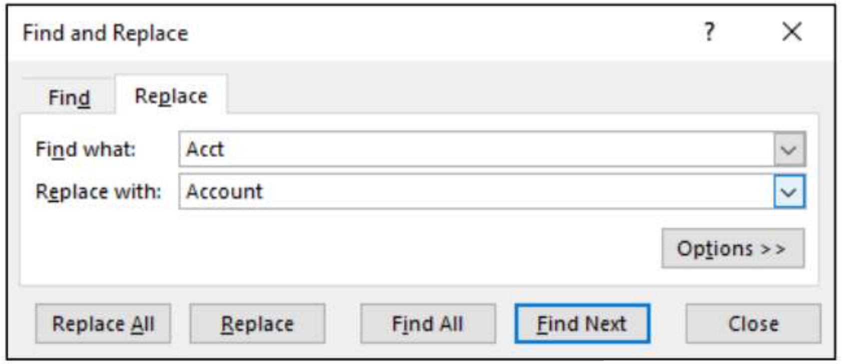


FIGURE 5-3: Fix funky data with Excel's Replace feature.

2. Enter incorrect text you want to find and then enter the correct text. You can proceed by:

   - Click Find Next to find the next instance. If you want to change, click Replace; otherwise, click Find Next, **OR**
   - If you're absolutely certain that you want to replace every last instance, click the Replace All button to make all the changes at one time

The CLEAN & CONCAT function

CLEAN

- You can use the CLEAN function to remove nonprintable characters from text or expunge the line feeds and carriage returns, making the data display on a single line.
- The CLEAN function uses the following syntax: **CLEAN(text)**
  - For example, to clean the text stored in cell A2: **CLEAN(A2)**

CONCAT

- The CONCAT function joins two or more chunks of text into a single string using the following syntax: **CONCAT(text1, text2, text3,...)**
- For example, **CONCAT(A1, " ", B1)** tells Excel to take the text from cell A1, tack on a space, and then add the text from cell B1.
- If A1 contains “Paul” and B1 contains “McFedries,” the CONCAT function returns the following string: **Paul McFedries**

The EXACT & LEN function

EXACT

- The EXACT function compares two strings. If the two strings are the same, it returns the logical value TRUE. If the two strings differ, the EXACT function returns FALSE.
- The EXACT function is case sensitive, and uses the following syntax: **EXACT(text1,text2)**
- For example, to check whether the two strings "Redmond" and "redmond" are the same, use the following: **EXACT("Redmond","redmond").** This should return FALSE.

LEN

- The LEN function counts the number of characters in a string, and uses the syntax: **LEN(text)**
- For example, to measure the length of the string in cell I81 reading Semper fidelis, use the following formula: **LEN(I81)**
- The function returns the value 14. Spaces are counted as characters, too.

The FIND & LOWER function

FIND

- The FIND function finds the starting character position of one string within another string, and uses the syntax: **FIND(find_text,within_text,[start_num])**
- For example, to find at what point the two-letter state abbreviation WA begins in the string *Redmond WA 98052*, use the following: **FIND("WA","Redmond WA 98052",1)**
- This returns the value 9 because WA begins at the ninth position (spaces are counted).

LOWER

- If an all-uppercase column should really be all-lowercase, use the LOWER function.
- The function uses the following syntax: **LOWER(text)**
- For example, to convert the string PROFESSIONAL to professional, use the following formula: **LOWER("PROFESSIONAL").** The function returns “professional”.

The MID function

- The MID function returns a chunk of text from inside of a string. The function uses the following syntax: **MID(text, start_num, num_char)**
- For example, to grab the text fragment tac from the string tic tac toe, use the following formula: **=MID("tic tac toe",5,3)**
- For example, here are some part numbers that are imported from a database:

  LDW-2125-X52

  MP-9790-C78 ...

- Suppose you want to extract the four digits that appear between the hyphens (-), use the FIND function to locate the first hyphen, and add 1 to get the starting point of the four-digit numeric value. Here's expression to use if the text is in cell A2: **FIND("-", A2) + 1**
- You can then plug this expression into your MID function: **MID(A2, FIND("-", A2) + 1, 4)** For the first example string, this expression return 2125.

The NUMBERVALUE function

- The NUMBERVALUE function converts digits formatted as a string to a true numeric value; the syntax: **NUMBERVALUE(text, decimal_separator, group_separator)**
- For example, the formula in the previous section to extract four digits from within a string returns those digits as a string instead of a number. Assuming that one such string resides in cell B2, you use the following expression to convert it to a number: **NUMBERVALUE(B2)**
- As another example, suppose your imported data includes the following values as strings:

  71.970,53 …

- To convert these strings to numbers, you use the following version of NUMBER-VALUE (assuming that the first string is in B2): **NUMBERVALUE(B2, ",", ".")**
- This tells Excel to treat the comma (,) as the decimal separator and the period (.) as the group separator . For the first string, this function returns the numeric value 71970.53.

The PROPER & REPLACE function

PROPER

- The PROPER function capitalizes the first letter of every word in a string. The function uses the syntax: **PROPER(text)**
- For example, to capitalize the initial letters in the string SALES REPRESENTATIVE , use the following formula: **PROPER("SALES REPRESENTATIVE")**
- The function returns the string Sales Representative.

REPLACE

- The REPLACE function replaces a portion of a string with new text. The function uses the syntax: **REPLACE(old_text, start_num, num_chars, new_text)**
- For example, to replace the string man with the string Representative in the string Salesman, use the following expression: **REPLACE("Salesman",6,3," Representative")**
- The function returns the string Sales Representative.

The SEARCH & UPPER function

SEARCH

- The SEARCH function performs a case-insensitive search to return the starting position of a specified string within a larger string. The function uses the syntax: **SEARCH(find_text,within_text,start_num)**
- For example, to identify the position at which the text fragment Churchill starts in the string Mr. Churchill, use the following formula: **SEARCH("Churchill","Mr. Churchill",1)**
- The function returns the value 5.

UPPER

- The UPPER function returns an all-uppercase version of a string. The function uses the following syntax: **UPPER(text)**
- For example, to convert the string professional to PROFESSIONAL, you can use the following formula: **UPPER("professional").** The function returns “PROFESSIONAL”.

The SUBSTITUTE & VALUE function

SUBSTITUTE

- The SUBSTITUTE function replaces occurrences of text in a string. The function uses the syntax: **SUBSTITUTE(text,old_text,new_text,instance_num)**
- For example your data comes with a phone number column in the following format:

  123 555 6789

- Assuming that one such string is in cell A2, here's a SUBSTITUTE expression that replaces all the spaces with dashes: **SUBSTITUTE(A2, " ", "-")**

VALUE

- The VALUE function converts a string that looks like a numeric value to an actual number. The function uses the following syntax: **VALUE(text)**
- For example, to convert the string $123,456.78 — assume that this isn't a numeric value but a string — you use the following expression: **VALUE("$123,456.78")**
- The function returns the numeric value 123456.78.

The TEXT & TRIM function

TEXT

- The TEXT function formats a value and then returns the value as text. The function uses the syntax: **TEXT(value,format_text)**
- For example, the formula: **=TEXT(1234.5678,"$##,###.00")** returns the text $1,234.57.

TRIM

- Some of the biggest causes of untidiness in imported data are extra characters.
- The aptly named TRIM function removes extra spaces, line feeds, carriage returns, and tabs from a string. The function uses the syntax: **TRIM(text)**
- For example, to trim the text stored in cell A2, use the following syntax: **TRIM(A2)**

The TEXTJOIN function

- The TEXTJOIN function combines two or more strings into a single string with a specified character — called the delimiter — between each of the original strings. The function uses the syntax: **TEXTJOIN(delimiter,ignore_empty, text1, text2,...))**
- For example, suppose you imported data that includes separate fields for three types of code: Manufacturer, Category, and Location. Also you can generate the values of a Part Number field by combining these three codes into a single string, separated by hyphens. If the three codes are in cells A2, B2, and C2, the TEXTJOIN expression is: **=TEXTJOIN("-", TRUE, A2, B2, C2)**

D2 =TEXTJOIN("-", TRUE, A2, B2, C2)

	A	B	C	D	E	F
1	Manufacturer	Category	Location	Part Number		
2	LDW	2125	X52	LDW-2125-X52		
3	MP	9790	C78	MP-9790-C78		
4	PNH	7793	W40	PNH-7793-W40		
5	SA	8703	I16	SA-8703-I16		
6	RB	3024	Z87	RB-3024-Z87		
7	N	4191	W23	N-4191-W23		

FIGURE 5-4: The TEXTJOIN function doing its thing.

Converting text function formulas to text

- When you’re using text functions to clean up imported data, the most common technique is to start a new column beside the data and create a formula that uses a text function to convert the first cell of the imported data into the format you want. You then fill that formula down the column to get cleaned-up versions of all the cells.
- The aforementioned technique has one inconvenient fact: Your cleaned-up versions appear as formula results instead of plain text or numbers. You can convert those formulas to the strings and numbers that they return by:

1. **Select the worksheet range that holds the formulas.**
2. **Choose Home ➪ Copy.**
3. **Choose Home ➪ Paste ➪ Paste Values.**
    - Because you left the formula range selected, the paste replaces the formulas with the values they returned.

Using Validation to Keep Data Clean

- The Data Validation command enables you to supply messages that give data input information and error messages that attempt to help someone correct data-entry errors.
- To use Data Validation, follow these steps:

1. Select the worksheet range where the to-be-validated data will go.
2. Choose Data ➪ Data Validation. Excel displays the Data Validation dialog box, as shown in Figure 5-5.

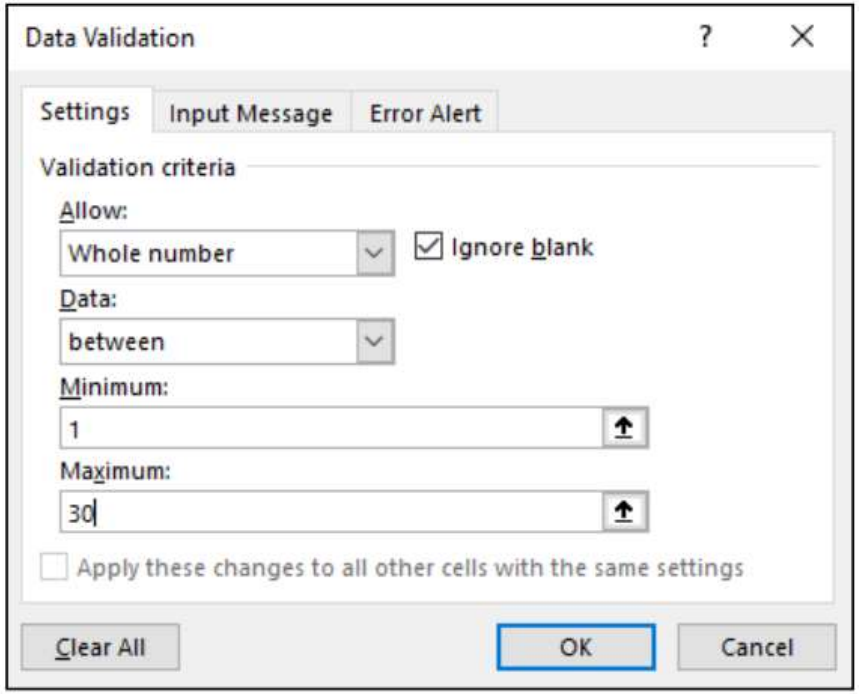


FIGURE 5-5: Keep data clean with the Data Validation dialog box.

Using Validation to Keep Data Clean

3. On the Settings tab of the Data Validation dialog box, use the controls in the Validation Criteria area to describe what is valid data.
4. Fine-tune the validation.
5. (Optional) Consider expanding the scope of the data validation.

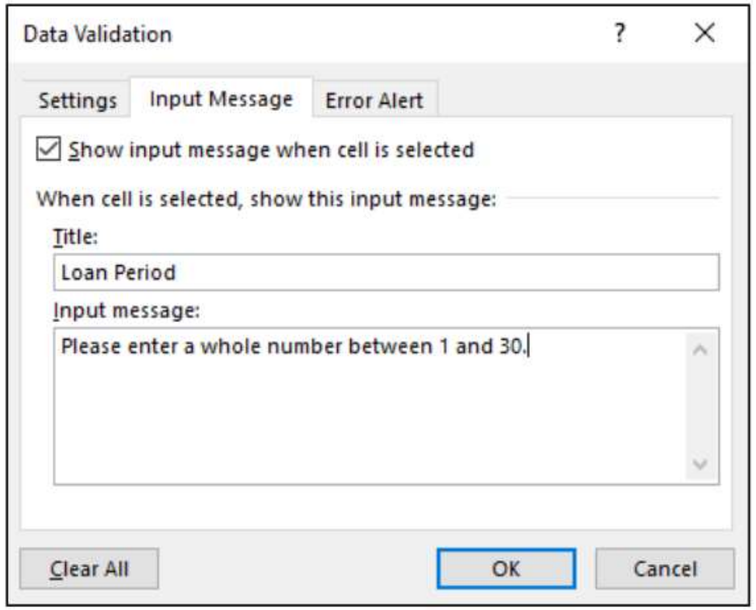


FIGURE 5-6: Create a data entry instruction message.

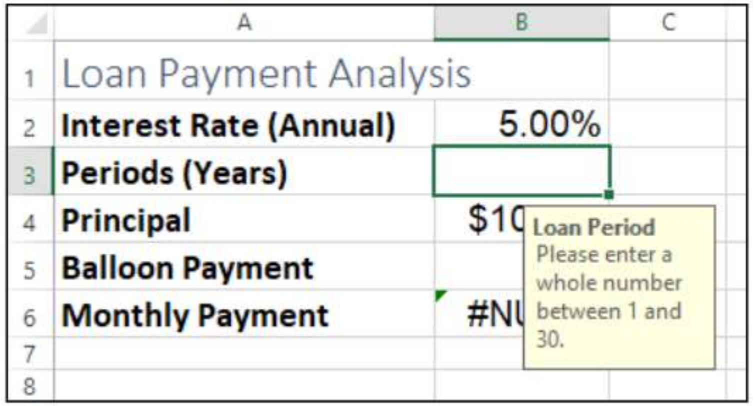


FIGURE 5-7: When the user selects the cell, Excel displays the message.

Using Validation to Keep Data Clean

6. Provide an input message from the Input Message tab of the Data Validation dialog box.
7. Provide an error message from the Error Alert tab of the Data Validation dialog box (Figure 5-8).

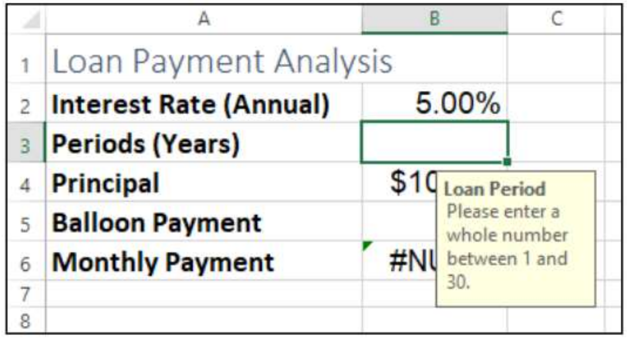


FIGURE 5-7: When the user selects the cell, Excel displays the message.

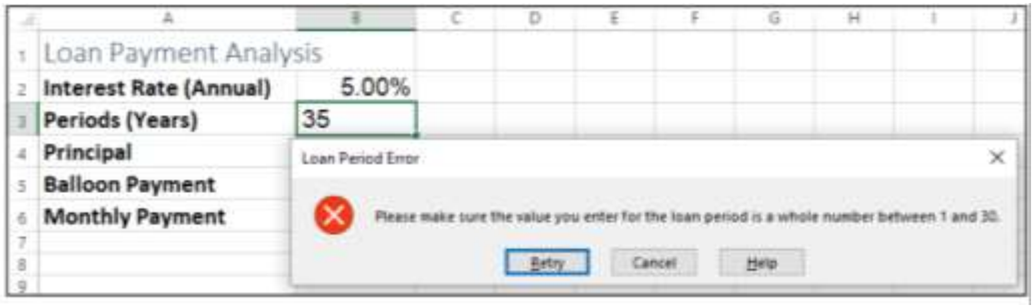


FIGURE 5-9: If the user enters invalid data, the error message rears its head.

FIGURE 5-8: Create an annoying data-entry error message.

Practical Portion

Chapter 11: Analyzing Data with Statistics

Excel Data Analysis For Dummies

Chapter 11: Analyzing Data with Statistics

- Counting Numbers
- Counting empty & nonempty cells
- Counting cells that match criteria
- Calculating average & conditional average
- Calculating median & mode
- Finding Rank
- Determining Nth Largest & Smallest Value
- Calculating the Standard Deviation
- Finding the Correlation

Counting numbers

- Take a look at the workbook DS498_week4-Ch11 for the following analysis of data:
- The COUNT function tallies the number of cells within a specified range that hold numeric values.

  =COUNT(value1[, value2, ...])

- Here, value1, value2, and so on are cell or range references.
- For example, to use the COUNT function to return how many numeric values are in the range B3:B12 in the worksheet shown in Figure 11-1, use the following formula: **=COUNT(B3:B12)**
- As shown in cell E3, COUNT returns 6 (five numbers plus one date).

E3 =COUNT(B3:B12)

	A	B	C	D	E
1	Data From Server				
2	ID	Data			
3	1	735		COUNT	6
4	2	TRUE		COUNTA	9
5	3	636		COUNTBLANK	1
6	4	Five		COUNTIF	3
7	5			COUNTIFS	2
8	6	#N/A			
9	7	995			
10	8	747			
11	9	8/23/2019			
12	10	894			
13					

FIGURE 11-1: COUNT returns the tally of the numeric values in a range.

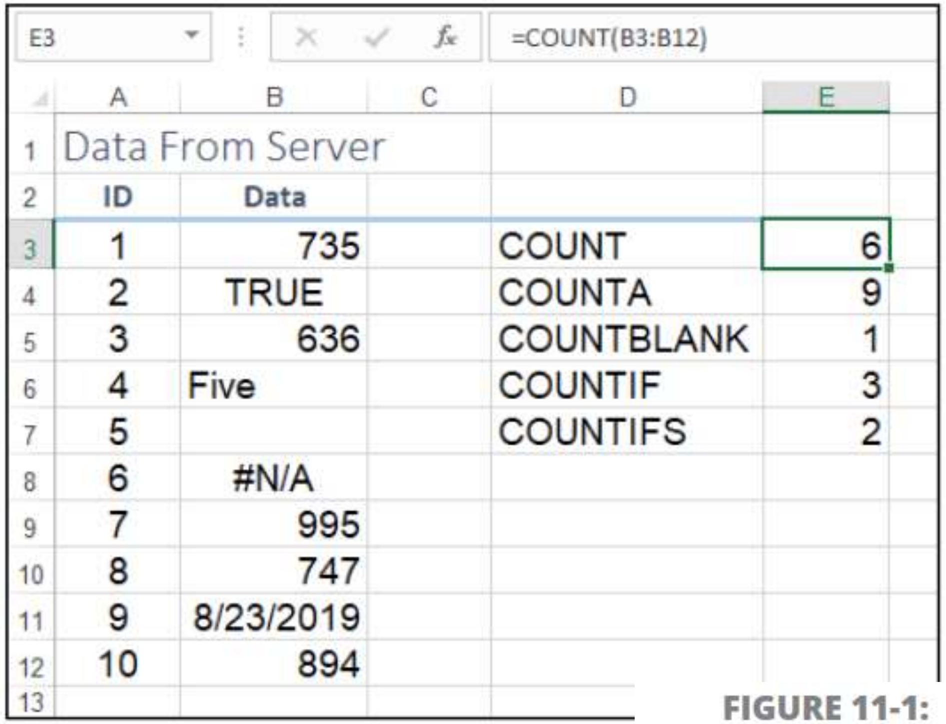
Counting empty & nonempty cells

Nonempty Cells:

- The COUNTA function counts the number of cells within a specified range that are nonempty.
- For example, to use COUNTA in the range B3:B12: **=COUNTA(B3:B12).** COUNTA returns the value 9.

Empty Cells:

- The COUNTBLANK function counts the number of cells within a specified range that are empty.
- For example, to use COUNTBLANK in the range B3:B12: **=COUNTBLANK(B3:B12).** COUNTBLANK returns the value 1.



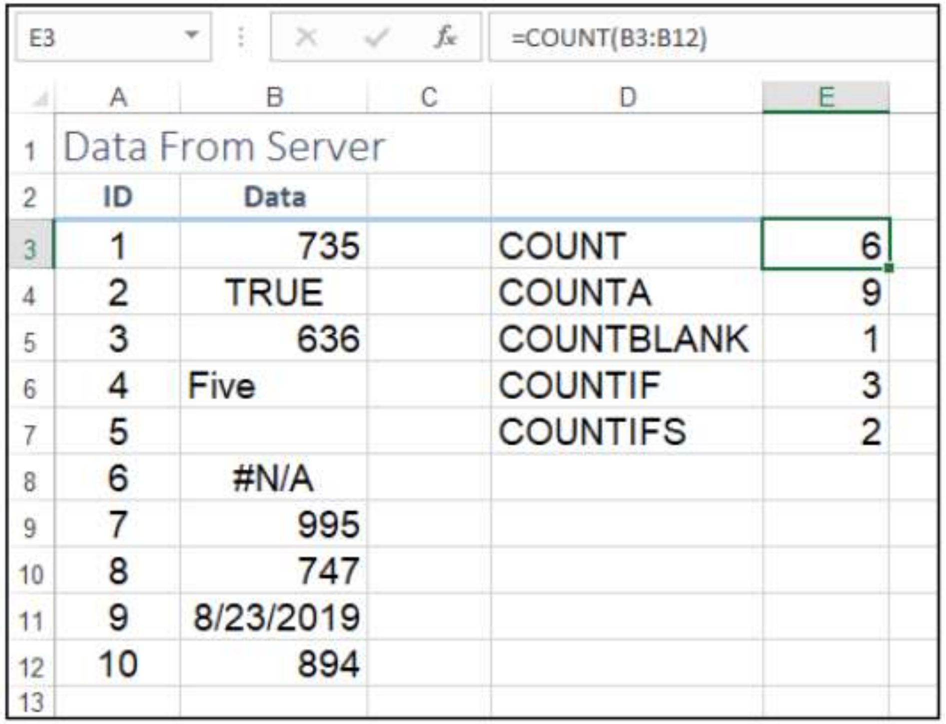
E3 =COUNT(B3:B12)

	A	B	C	D	E
1	Data From Server				
2	ID	Data			
3	1	735		COUNT	6
4	2	TRUE		COUNTA	9
5	3	636		COUNTBLANK	1
6	4	Five		COUNTIF	3
7	5			COUNTIFS	2
8	6	#N/A			
9	7	995			
10	8	747			
11	9	8/23/2019			
12	10	894			
13					

FIGURE 11-1: COUNT returns the tally of the numeric values in a range.

Counting cells that match criteria

- If you have your own criteria for what should or shouldn't get counted, you can use the COUNTIF function to apply that condition.
- The syntax is: **=COUNTIF(range, criteria)**
- For example, looking back at the range B3:B12, suppose you want to know how many cells contain a value greater than 800: **=COUNTIF(B3:B12,">800")** This returns the value 3.
- You can use any of the standard logical operators when building your criteria expression.



	A	B	C	D	E
1	Data From Server				
2	ID	Data			
3	1	735		COUNT	6
4	2	TRUE		COUNTA	9
5	3	636		COUNTBLANK	1
6	4	Five		COUNTIF	3
7	5			COUNTIFS	2
8	6	#N/A			
9	7	995			
10	8	747			
11	9	8/23/2019			
12	10	894			
13					

FIGURE 11-1: COUNT returns the tally of the numeric values in a range.

Calculating average & conditional average

Average

- The AVERAGE function uses the following syntax: **AVERAGE(number1[, number2, ...])**
- For example, to determine the average of the values in the range D3:D19, you use the following formula: **=AVERAGE(D3:D19)**

Conditional Average

- The AVERAGEIF function averages cells in a range that meet the condition you specify.
- AVERAGEIF takes up to three arguments: **=AVERAGEIF(range, criteria[, average_range])**
- For example, if you want to get the average of the values in the Gross Margin column, you use: **=AVERAGE(H3:H10)**

F13 =AVERAGEIF(E3:E10, "<10", H3:H10)

	A	B	C	D	E	F	G	H
1	Parts Database							
2	Division	Description	Number	Quantity	Cost	Total Cost	Retail	Gross Margin
3	4	Gangley Pliers	D-178	57	$10.47	$ 596.79	$17.95	71.4%
4	3	HCAB Washer	A-201	856	$ 0.12	$ 102.72	$ 0.25	108.3%
5	3	Finley Sprocket	C-098	357	$ 1.57	$ 560.49	$ 2.95	87.9%
6	2	6" Sonotube	B-111	86	$15.24	$1,310.64	$19.95	30.9%
7	4	Langstrom 7" Wrench	D-017	75	$18.69	$1,401.75	$27.95	49.5%
8	3	Thompson Socket	C-321	298	$ 3.11	$ 926.78	$ 5.95	91.3%
9	1	S-Joint	A-182	155	$ 6.85	$1,061.75	$ 9.95	45.3%
10	2	LAMF Valve	B-047	482	$ 4.01	$1,932.82	$ 6.95	73.3%
11								
12			Total cost of Division 3 parts:			$1,589.99		
13		Average gross margin for parts under $10:				81.2%		
14	Average gross margin for parts over $15 and less than 100 units:					40.2%		

FIGURE 11-2:
A parts database.

Calculating median & mode

Median

- To calculate the median, you use Excel's MEDIAN function: **MEDIAN(number1[, number2, ...])**
- In the Product Defects, the median value of the Defects column (D3:D19) is given by: **=MEDIAN(D3:D19)**

Mode

- In the Product Defects, the mode value of the defects column (D3:D19) is given by the formula in cell I3: **=MODE(D3:D19)**

I2 =MEDIAN(D3:D19)

Product Defects

Workgroup	Group Leader	Defects	Units	% Defective
A	Hammond	8	969	0.8%
B	Hammond	4	815	0.5%
C	Hammond	14	1,625	0.9%
D	Hammond	3	1,453	0.2%
E	Hammond	9	767	1.2%
F	Hammond	11	1,023	1.1%
G	Hammond	15	1,256	1.2%
H	Hammond	8	781	1.0%
L	Bolter	7	1,109	0.6%
M	Bolter	11	1,021	1.1%
N	Bolter	6	812	0.7%
O	Bolter	11	977	1.1%
P	Bolter	5	1,182	0.4%
Q	Bolter	7	961	0.7%
R	Bolter	12	689	1.7%
T	Bolter	19	1,308	1.5%

Median Defects:	9
Mode Defects:	11

FIGURE 11-3: MEDIAN returns the median value of a set of numeric values.

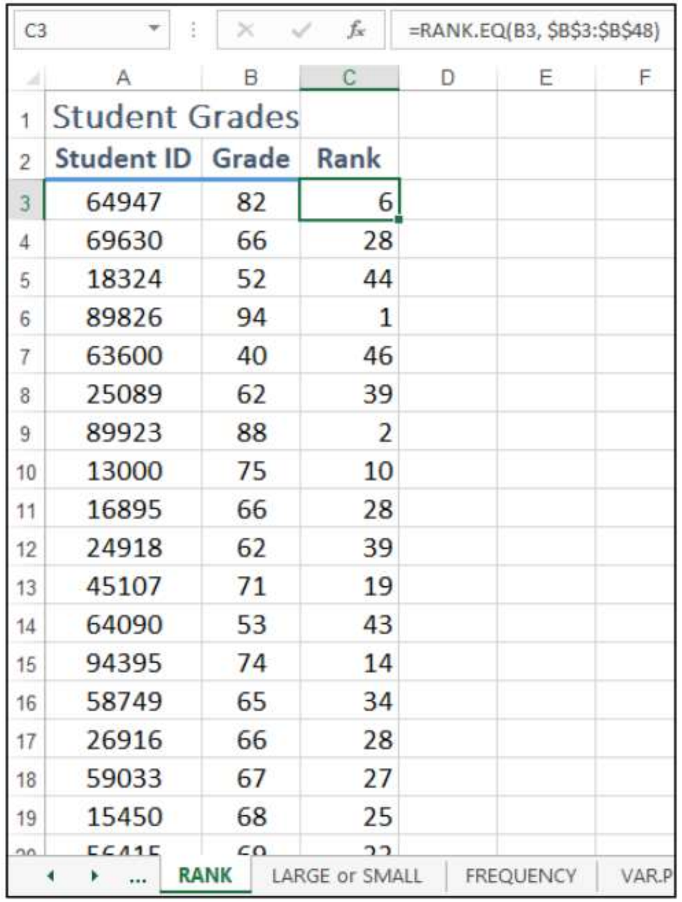
Finding Rank

Rank

- RANK.EQ determines an item's rank relative to other items in a list: **RANK.EQ(number, ref[, order])**
- For example, in Figure 11-4 (range B3:B48) to find the rank of the grade in cell B3: **=RANK.EQ(B3, $B$3:$B$48).** The returned value is 6.

Rank Average

- With RANK.AVG, if two or more numbers have the same rank, Excel averages the rank: **RANK.AVG(number, ref[, order])**
- For example, in the list 100, 95, 90, 85, 85, 80, 70, the RANK.AVG function ranks the number 85 as 4.5, which is the average of 4 and 5.


C3 =RANK.EQ(B3, $B$3:$B$48)

	A	B	C	D	E	F
1	Student Grades					
2	Student ID	Grade	Rank			
3	64947	82	6			
4	69630	66	28			
5	18324	52	44			
6	89826	94	1			
7	63600	40	46			
8	25089	62	39			
9	89923	88	2			
10	13000	75	10			
11	16895	66	28			
12	24918	62	39			
13	45107	71	19			
14	64090	53	43			
15	94395	74	14			
16	58749	65	34			
17	26916	66	28			
18	59033	67	27			
19	15450	68	25			
20	[illegible]	[illegible]	[illegible]			

RANK | LARGE or SMALL | FREQUENCY | VAR.P

FIGURE 11-4: RANK returns the ranking of a value in a set.

Determining Nth Largest & Smallest Value

LARGE Function

- LARGE returns the nth highest value in a list: **LARGE(array, n)**
- For example, given B3:B48, what's the minimum mark required to crack the top 10 grades?
- The LARGE function (D4) returns the value 75: **=LARGE(B3:B48, 10)**

SMALL Function

- SMALL returns the nth smallest value in an array or range: **SMALL(array, n)**
- For example, given B3:B48, what's the lowest grade?
- The formula (D7) returns the value 40: **=SMALL(B3:B48, 1)**

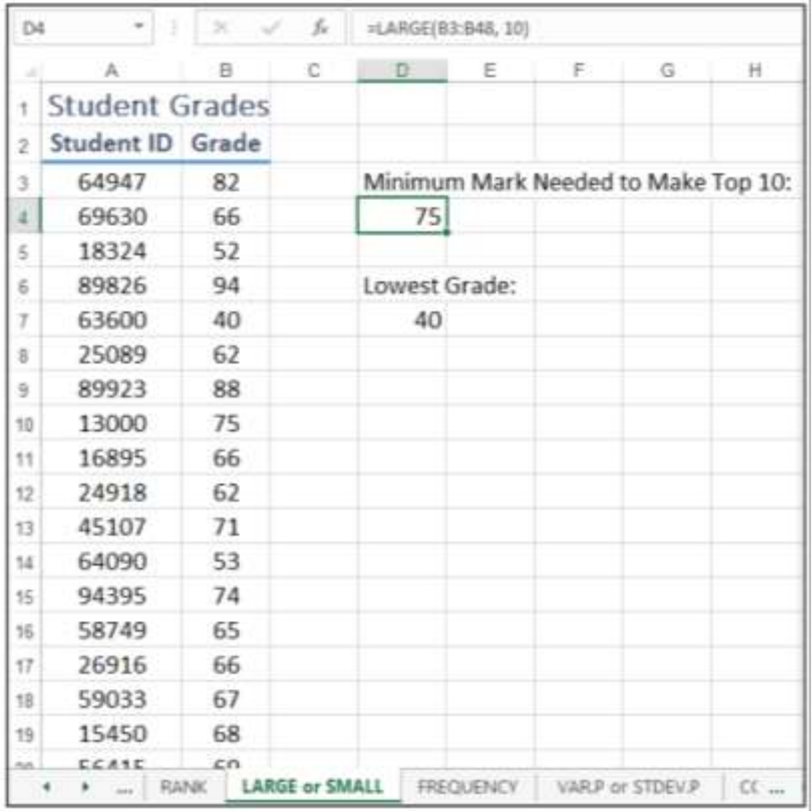


FIGURE 11-5: LARGE returns the *n*th largest value in a range or array.

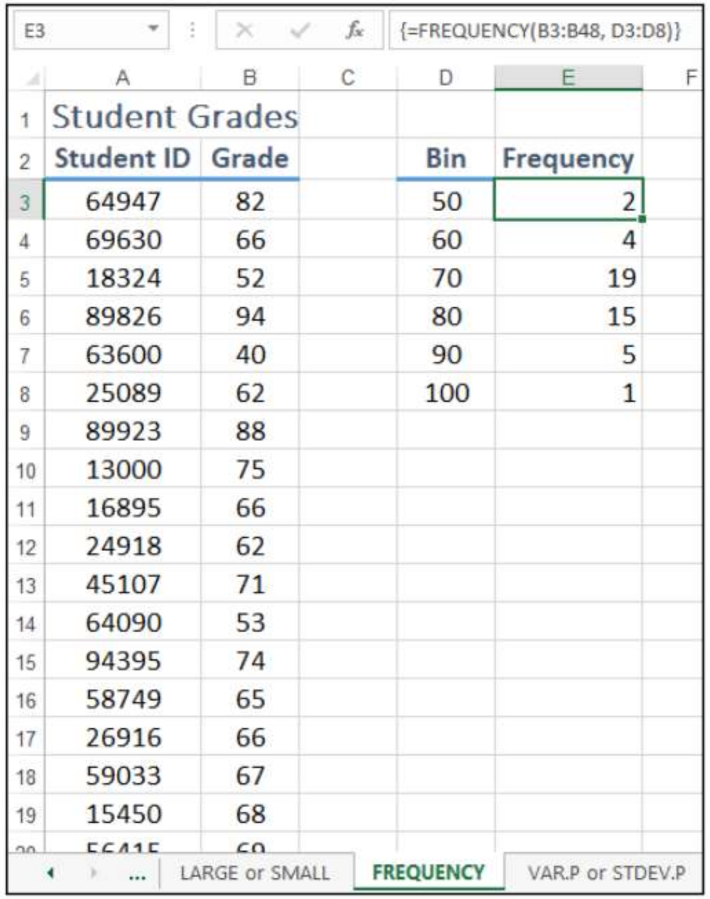
Determining Nth Largest & Smallest Value

- You can use the FREQUENCY function to return the number of occurrences in each group:

  FREQUENCY(data_array, bins_array)

- Here are the steps to follow:

1. Select the cells where you want the grouped frequency distribution to appear.
2. Type =frequency(.
3. Enter or select the items you want to group.
4. Type a comma and then enter or select the list of groupings.
5. Type)
6. Hold down Ctrl+Shift and then click the Enter button or press Ctrl+Shift+Enter.



E3 {=FREQUENCY(B3:B48, D3:D8)}

	A	B	C	D	E
1	Student Grades				
2	Student ID	Grade		Bin	Frequency
3	64947	82		50	2
4	69630	66		60	4
5	18324	52		70	19
6	89826	94		80	15
7	63600	40		90	5
8	25089	62		100	1
9	89923	88			
10	13000	75			
11	16895	66			
12	24918	62			
13	45107	71			
14	64090	53			
15	94395	74			
16	58749	65			
17	26916	66			
18	59033	67			
19	15450	68			

LARGE or SMALL | FREQUENCY | VAR.P or STDEV.P

FIGURE 11-6: FREQUENCY tells you how many items in a range appear in each bin.

Determining Nth Largest & Smallest Value

- In Excel, the variance is calculated using VAR.S or VAR.P:

  VAR.S(number1[, number2, ...])

  VAR.P(number1[, number2, ...])

- Use VAR.S if data represents a sample; use VAR.P if data represents the entire population.
- For example, in the Product Defects worksheet; calculate the variance of the Defects column (D3:D19) with the following formula (see cell H3): **=VAR.P(D3:D19)**

H3 =VAR.P(D3:D19)

Product Defects

Workgroup	Group Leader	Defects	Units	% Defective
A	Hammond	8	969	0.8%
B	Hammond	4	815	0.5%
C	Hammond	14	1,625	0.9%
D	Hammond	3	1,453	0.2%
E	Hammond	9	767	1.2%
F	Hammond	11	1,023	1.1%
G	Hammond	15	1,256	1.2%
H	Hammond	8	781	1.0%
L	Bolter	7	1,109	0.6%
M	Bolter	11	1,021	1.1%
N	Bolter	6	812	0.7%
O	Bolter	11	977	1.1%
P	Bolter	5	1,182	0.4%
Q	Bolter	7	961	0.7%
R	Bolter	12	689	1.7%
T	Bolter	19	1,308	1.5%

Defects Variance: 17.2

Defects Standard Deviation: 4.2

FIGURE 11-7: VAR.P returns the variance of data that represents an entire population.

Calculating the Standard Deviation

- To calculate the standard deviation, you can use the STDEV.S or STDEV.P function.

  STDEV.S(number1[, number2, ...])

  STDEV.P(number1[, number2, ...])

- Use STDEV.S for sample; STDEV.P for entire population.
- For example, in the Product Defects worksheet shown earlier in Figure 11-7, I calculated the standard deviation of the Defects column (D3:D19) with the following formula (see cell H3): **=STDEV.P(D3:D19)**

H3 =VAR.P(D3:D19)

Product Defects

Workgroup	Group Leader	Defects	Units	% Defective
A	Hammond	8	969	0.8%
B	Hammond	4	815	0.5%
C	Hammond	14	1,625	0.9%
D	Hammond	3	1,453	0.2%
E	Hammond	9	767	1.2%
F	Hammond	11	1,023	1.1%
G	Hammond	15	1,256	1.2%
H	Hammond	8	781	1.0%
L	Bolter	7	1,109	0.6%
M	Bolter	11	1,021	1.1%
N	Bolter	6	812	0.7%
O	Bolter	11	977	1.1%
P	Bolter	5	1,182	0.4%
Q	Bolter	7	961	0.7%
R	Bolter	12	689	1.7%
T	Bolter	19	1,308	1.5%

Defects Variance: 17.2

Defects Standard Deviation: 4.2

FIGURE 11-7: VAR.P returns the variance of data that represents an entire population.

Finding the Correlation

- Correlation is a measure of the relationship between two sets of data:

  CORREL(array1, array2)

- Figure 11-8 shows a worksheet that has advertising costs in the range C3:C14 and sales in the range D3:D14. Cell F3 calculates the correlation: **=CORREL(C3:C14, D3:D14)**
- CORREL returns the correlation coefficient, between –1 and 1.
- The sign suggests whether the relationship is positive (+) or negative (–).

F3 =CORREL(C3:C14, D3:D14)

	A	B	C	D	E	F
1	Correlation Between Advertising and Sales					
2			**Advertising**	**Sales**		
3	**Fiscal**	1st Quarter	512,450	8,123,965	**Correlation:**	0.74
4	**2017**	2nd Quarter	447,840	7,750,500		
5		3rd Quarter	500,125	7,860,405		
6		4th Quarter	515,600	8,005,800		
7	**Fiscal**	1st Quarter	482,754	8,136,444		
8	**2018**	2nd Quarter	485,750	7,950,426		
9		3rd Quarter	460,890	7,875,500		
10		4th Quarter	490,400	7,952,600		
11	**Fiscal**	1st Quarter	510,230	8,100,145		
12	**2019**	2nd Quarter	515,471	8,034,125		
13		3rd Quarter	525,850	8,350,450		
14		4th Quarter	520,365	8,100,520		
15						

FIGURE 11-8: CORREL calculates the correlation between two sets of values.

Correlation Coefficient	**Interpretation**
1	The data sets are perfectly and positively correlated. For example, a 10-percent increase in advertising produces a 10-percent increase in sales.
Between 0 and 1	The data sets are positively correlated. The higher the number is, the higher the correlation is between the data.
0	No correlation exists between the data.
Between 0 and –1	The data sets are negatively correlated. The lower the number is, the more negatively correlated the data is.
–1	The data sets have a perfect negative correlation. For example, a 10-percent increase in advertising leads to a 10-percent decrease in sales.

Practical Portion

Chapter 13: Analyzing Data with Inferential Statistics

Excel Data Analysis For Dummies

Chapter 13: Analyzing Data with Inferential Statistics

- Determining the Regression
- Calculating the Correlation
- Sampling Data (recommended)
- Using the t-Test Tools (recommend)
- Calculating the Covariance (recommended)

Determining the Regression

1. Choose Data ➪ Data Analysis.
2. Use the Analysis Tools list to select the Regression tool and then click OK.
3. Identify your Y and X values.
4. (Optional) Set the constant to zero.
5. (Optional) Calculate a confidence level in your regression analysis.
6. Select a location for the regression analysis results.

Determining the Regression

7. Identify what data you want returned. Figure 13-9 shows the dialog box
8. Click OK. Figure 13-10 shows a portion of the regression analysis results.

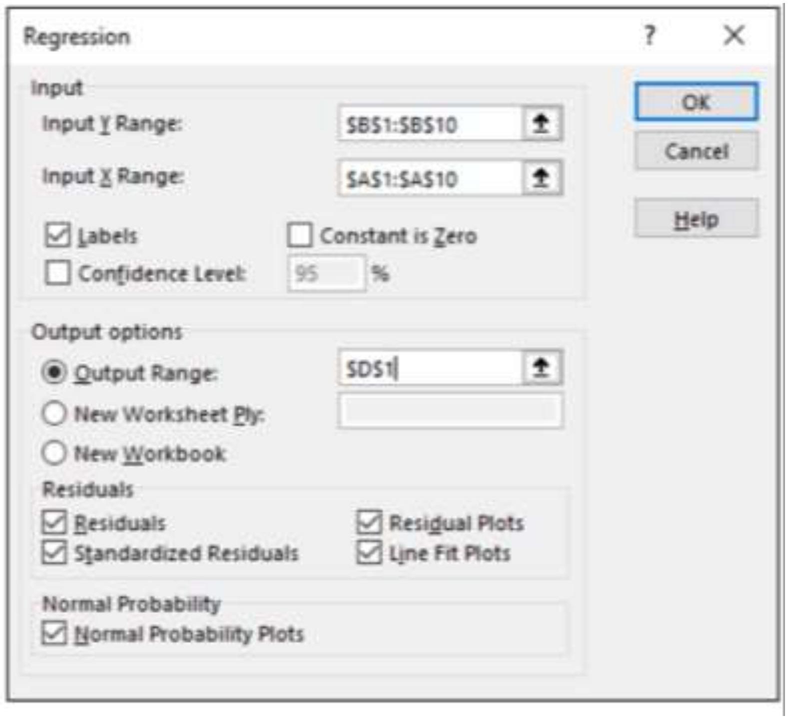


FIGURE 13-9: The completed Regression dialog box.

FIGURE 13-10: The regression analysis results.

Calculating the Correlation

To use the Correlation analysis tool, follow the steps:

1. Choose Data ➪ Data Analysis.
2. Use the Analysis Tools list to select the Correlation tool and then click OK.
3. Identify the range of X and Y values that you want to analyze.
4. Select an output location.
   - Figure 13-11 shows a completed version of the dialog box.
5. Click OK.
   - Figure 13-12 shows the correlation results for list price versus units sold.

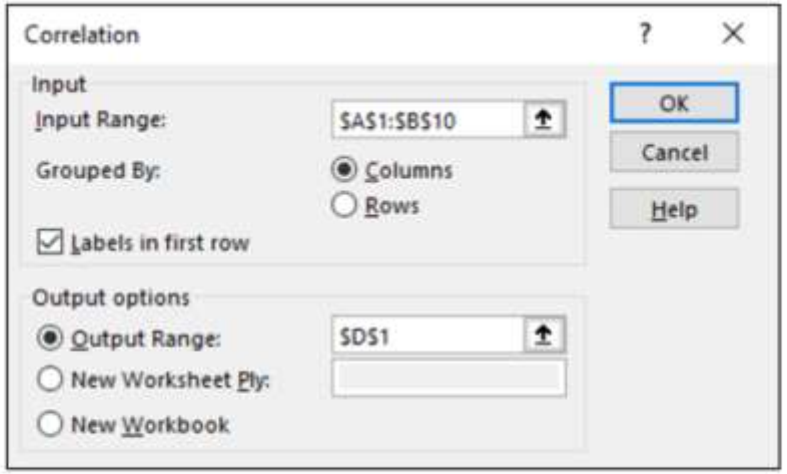


FIGURE 13-11: The completed Correlation dialog box.

	A	B	C	D	E	F
1	**List Price**	**Units Sold**			*List Price*	*Units Sold*
2	$6.99	9,985		List Price	1	
3	$7.99	8,110		Units Sold	-0.96666	1
4	$8.99	6,039				
5	$9.99	5,721				
6	$10.99	4,555				
7	$11.99	4,075				
8	$12.99	3,290				
9	$13.99	2,592				
10	$14.99	2,289				

FIGURE 13-12: The worksheet showing the correlation results for the list price and units sold data.

Recommended Part

Chapter 13: Analyzing Data with Inferential Statistics

Excel Data Analysis For Dummies

Sampling Data

- When the population is huge, and analyses would take up too much time, you extract a sample and use inferential statistics to draw conclusions about the population.
- To sample items from a worksheet, take the following steps:

1. Choose Data ➪ Data Analysis.
2. In the Analysis Tools list, select Sampling and then click OK.
3. Use the Input Range box to specify the range of cells from which you want to extract your sample.

Sampling Data (cont.)

4. Choose a sampling method (Periodic/Random). Figure 13-2 shows the sampling dialog box.
5. Click OK. Figure 13-3 shows an example; sample data is the Defects column (D3:D22).

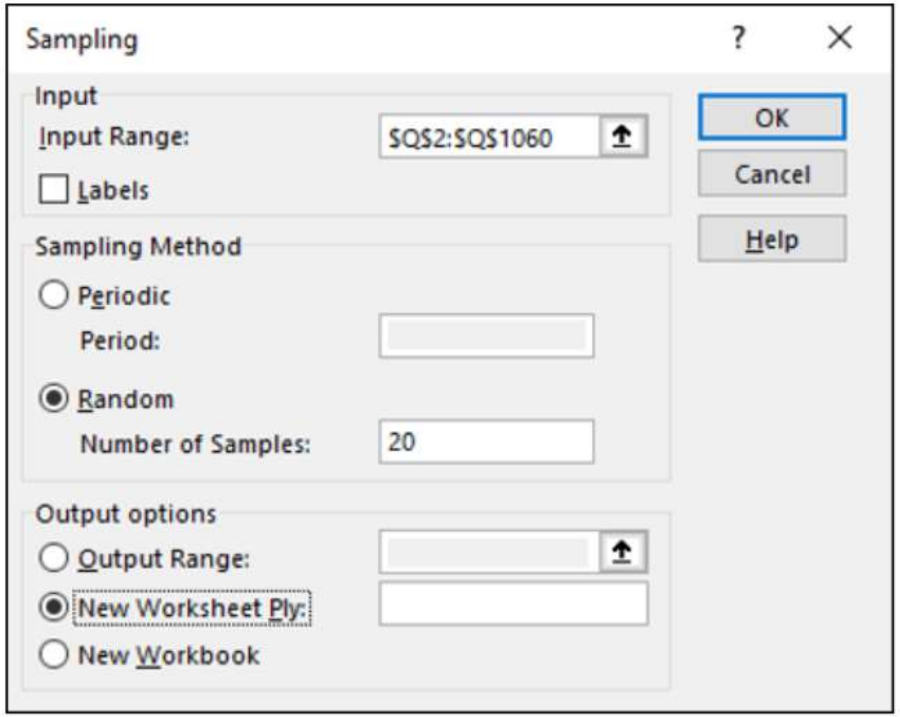


FIGURE 13-2: The Sampling dialog box, ready to sample.

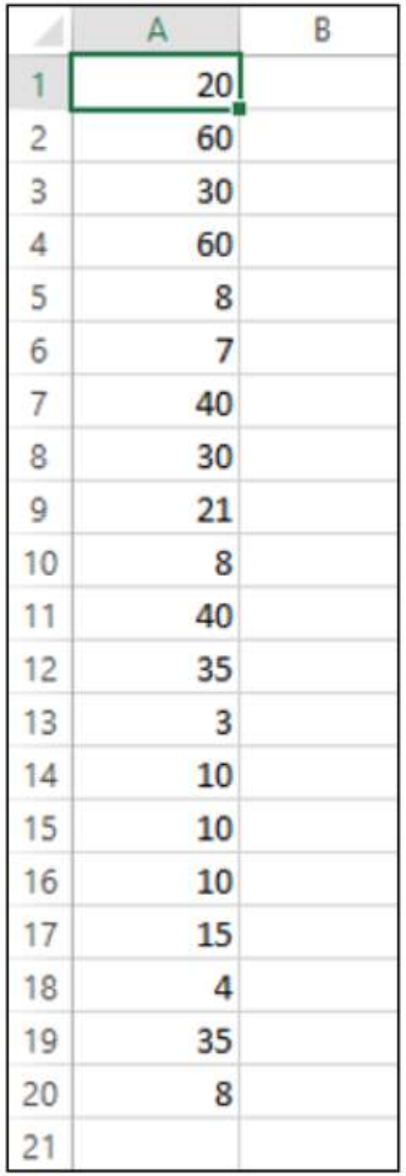


FIGURE 13-3: The sample extracted by the Sampling tool.

Using the t-Test Tools

- The Excel Analysis ToolPak add-in provides three tools for working with t-values and t-tests, which can be useful for inferences about very small data sets:
  - **Paired Two-Sample for Means:** for performing a paired two-sample t-test.
  - **Two-Sample Assuming Equal Variances:** for a two-sample test where you assume that the variances of both samples equal each other.
  - **Two-Sample Assuming Unequal Variances:** for a two-sample test where you assume that the two-sample variances are unequal.
- Take a look at the workbook DS498_week4-Ch13 for performing the t-test.

Using the t-Test Tools

To perform a t-test calculation, follow these steps:

1. Choose Data ➪ Data Analysis.
2. Use the Analysis Tools list to select the t-test tool you want to use and then click OK.
3. In the Variable 1 Range and Variable 2 Range input text boxes, identify the sample values by telling Excel in what worksheet ranges you've stored the two samples.

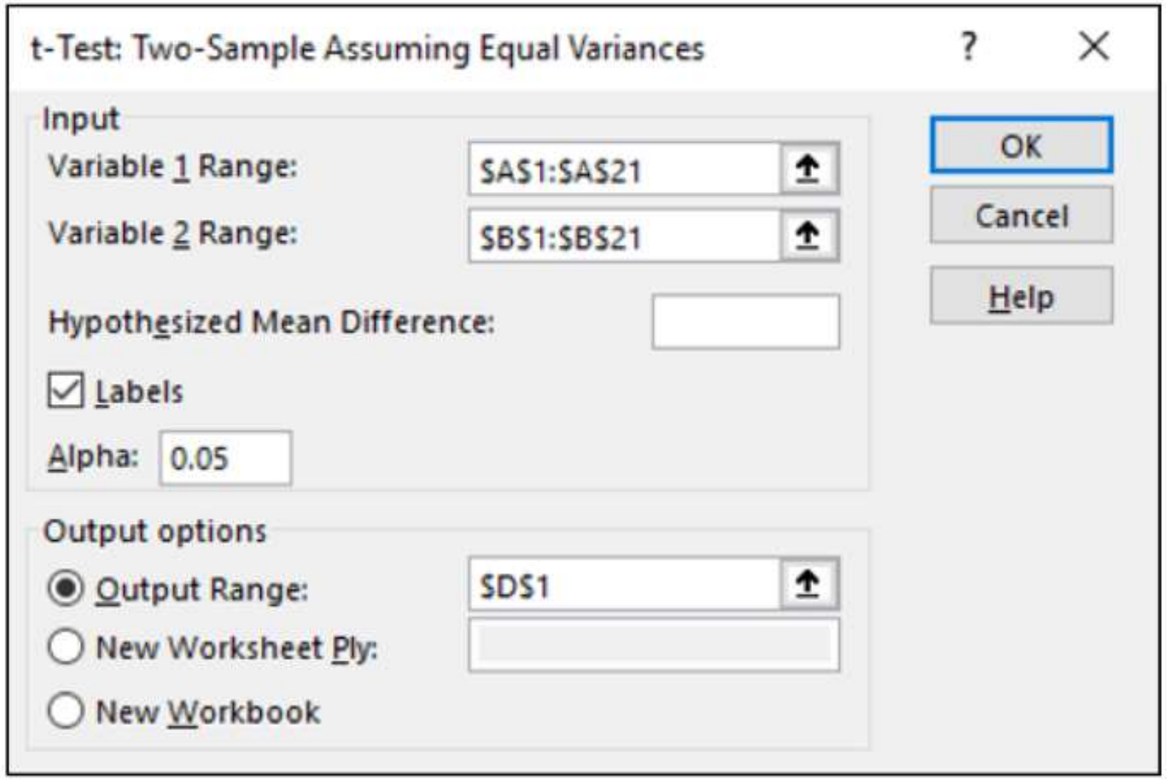


FIGURE 13-5: The t-Test: Two-Sample Assuming Equal Variances dialog box.

Using the t-Test Tools

4. Use the Hypothesized Mean Difference text box to indicate whether you hypothesize that the means are equal.
5. In the Alpha text box, state the confidence level for your t-test calculation.
6. In the Output Options section, indicate where the t-test tool results should be stored.
7. Click OK.

	A	B	C	D	E	F
1	Sample 1	Sample 2		t-Test: Two-Sample Assuming Equal	Variances	
2	0.390639	0.597253				
3	0.960314	0.247645			*Sample 1*	*Sample 2*
4	0.002978	0.76919		Mean	0.496319	0.443844
5	0.073425	0.83317		Variance	0.138465	0.097277
6	0.311795	0.450877		Observations	20	20
7	0.451693	0.08733		Pooled Variance	0.117871	
8	0.989853	0.247164		Hypothesized Mean Difference	0	
9	0.946743	0.036413		df	38	
10	0.88257	0.591507		t Stat	0.48333	
11	0.846565	0.475535		P(T<=t) one-tail	0.315817	
12	0.817594	0.06112		t Critical one-tail	1.685954	
13	0.933039	0.703724		P(T<=t) two-tail	0.631635	
14	0.013688	0.003346		t Critical two-tail	2.024394	
15	0.08753	0.887344				
16	0.017276	0.11998				
17	0.642356	0.393307				
18	0.782696	0.070239				
19	0.391383	0.837355				
20	0.142597	0.707126				
21	0.241643	0.757264				

FIGURE 13-6: The results of a t-test.

Calculating the Covariance

The use the Covariance tool follow these steps:

1. Choose Data ➪ Data Analysis.
2. Use the Analysis Tools list to select the Covariance tool and then click OK.
3. Identify the range of X and Y values that you want to analyze.
4. Select an output location.
   - Figure 13-13 shows a completed version of the dialog box.
5. Click OK after you select the output options.
   - Figure 13-14 shows the covariance results for the list price and units sold data.

FIGURE 13-13: The completed Covariance dialog box.

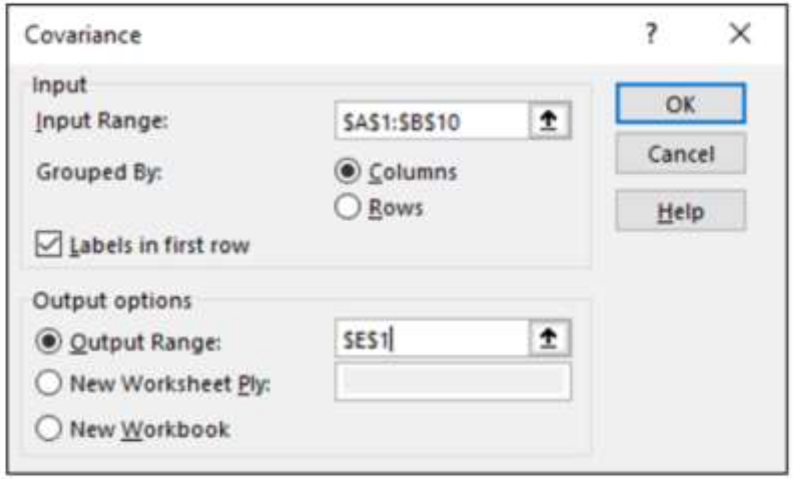


FIGURE 13-14: The worksheet showing the covariance results for the list price and units sold data.

	A	B	C	D	E	F	G
1	List Price	Units Sold				List Price	Units Sold
2	$6.99	9,985			List Price	6.666667	
3	$7.99	8,110			Units Sold	-6053.56	5882524
4	$8.99	6,039					
5	$9.99	5,721					
6	$10.99	4,555					
7	$11.99	4,075					
8	$12.99	3,290					
9	$13.99	2,592					
10	$14.99	2,289					

Main Reference

- **Chapter 3 (sections 3.2 to 3.6)**: “Nature of Data, Statistical Modeling, and Visualization” from “Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”.
- **Chapter 3, 5, 11 & 13:** “Getting Started with Data Analysis” from “Excel Data Analysis For Dummies”

Week self-review exercises

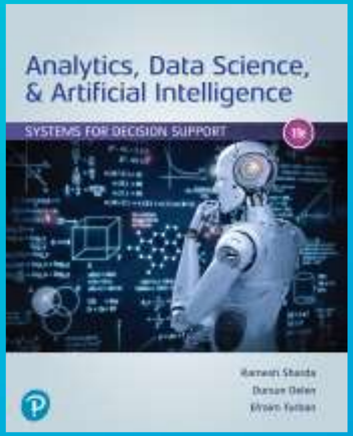
- **Application Case 3.1 to Application Case 3.4** from “Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support”.
- **Practice examples of chapter Ch 3, 5, 11 and 13** from Excel Data Analysis For Dummies using Excel

Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

IT445
Decision Support Systems

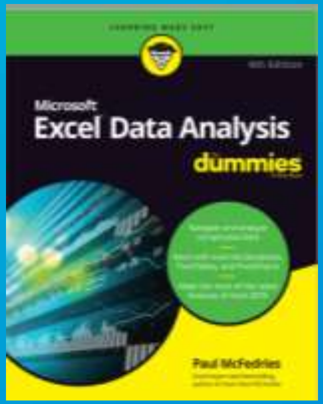
College of Computing and Informatics



Week 5

Chapter 3 – Part 2 (sections 3.7 to 3.11): Nature of Data, Statistical Modeling, and Visualization

Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support



Chapter 6-9: Analysing Data with PivotTables and PivotCharts & Dashboard development

Excel Data Analysis For Dummies

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Define business reporting and understand its historical evolution
2. Understand the importance of data/information visualization
3. Learn different types of visualization techniques
4. Appreciate the value that visual analytics brings to business analytics
5. Know the capabilities and limitations of dashboards
6. Demonstrate the ability in utilizing Excel to perform data preprocessing and data analysis

Required Reading

- **Chapter 3**: "Nature of Data, Statistical Modeling, and Visualization" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*".
- ***Chapter 6, 7, 8 & 9 - (sections 3.7 to 3.11)****: "Getting Started with Data Analysis"* from *"Excel Data Analysis For Dummies"*

Recommended Reading

- Excel Charts Tutorial
  https://www.tutorialspoint.com/excel_charts/index.htm

Recommended Video

- **10 Advanced Excel Charts**
  https://www.youtube.com/watch?v=lUjZa4YBixI
- **Introduction to Pivot Tables, Charts, and Dashboards in Excel (Part 1)**
  https://www.youtube.com/watch?v=9NUjHBNWe9M
- **Introduction to Pivot Tables, Charts, and Dashboards (Part 2)**
  https://www.youtube.com/watch?v=g530cnFfk8Y

Chapter 3 – Part 2 (sections 3.7 to 3.11): Nature of Data, Statistical Modeling, and Visualization

Analytics, Data Science, & Artificial Intelligence Systems For Decision Support

3.7 Business Reporting

- Overview

Overview

- Decision makers need information to make accurate and timely decisions.
- Information is usually provided to decision makers in the form of a written report that contains organized information referring to specific time periods.
- **Business reports can fulfill many different functions:**
    - To ensure that all departments are functioning properly
    - To provide information
    - To provide the results of an analysis
    - To persuade others to act
    - To create an organizational memory

Overview (cont.)

- **Business reporting** (also called OLAP or BI) is an essential part of the larger drive toward improved, evidence-based, optimal managerial decision making.
- The foundation of these business reports is various sources of data coming from both inside and outside the organization (OLTP systems).
- Creation of these reports involves extract, transform, and load (ETL) procedures in coordination with a data warehouse and reporting tools.
- This reporting process involves querying structured data sources, which were created using different logical data models and data dictionaries, to produce a human-readable, easily digestible report.

Overview (cont.)

- These types of business reports allow managers to stay informed and involved, review options and alternatives, and make informed decisions.
- **Metric Management Reports:** In many organizations, business performance is managed through outcome-oriented metrics. For external groups, these are service-level agreements. For internal management, they are key performance indicators (KPIs).
- **Dashboard-type Reports:** A popular idea in business reporting has been to present a range of different performance indicators on 1 page like a dashboard.
- **Balanced Scorecard–type Reports:** This method attempts to present an integrated view of success in an organization. In addition to financial performance, balanced scorecard–type reports also include customer, business process, and learning and growth perspectives.

3.8 Data Visualization

- Overview
- History of Data Visualization

Overview

- **Data visualization** is “the use of visual representations to explore, make sense of, and communicate data”.
- **Information** is the aggregation, summarization, and contextualization of data, what is portrayed in visualizations is the information, not the data.
- Data visualization is closely related to the fields of information graphics, information visualization, scientific visualization, and statistical graphics.
- Until recently, the major forms of data visualization available in both BI applications have included charts and graphs as well as the other types of visual elements used to create scorecards and dashboards.

History of Data Visualization

- Although visualization has not been widely recognized as a discipline until fairly recently, today's most popular visual forms date back a few centuries.
- Companies and individuals are interested in data; that interest has in turn sparked a need for visual tools that help them understand it.
- Countless applications, tools, and code libraries help people collect, organize, visualize, understand data.
- The future of data visualization holds more 3D imaging, experience with multidimensional data, and holographs of information.

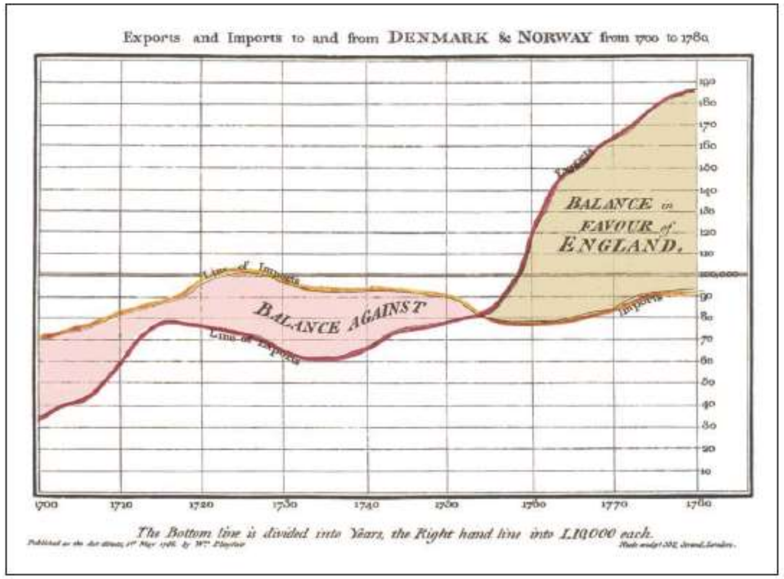


FIGURE 3.19 The First Time-Series Line Chart Created by William PlayFair in 1801.

3.9 Different Types of Charts and Graphs

- Basic Charts & Graphs
- Specialized Charts & Graphs
- Which Chart/Graph should you use?

Basic Charts & Graphs

- **Line Chart**: Line charts show the relationship between two variables; they are most often used to track changes or trends over time.
- **Bar Chart**: Effective when you have nominal/numerical data that splits into different categories so you can quickly see comparative results and trends.
- **Pie Chart**: Pie charts illustrate relative proportions of a specific measure. If the number of categories is more than a few consider using a bar chart.
- **Scatter Plot**: The scatter plot is often used to explore the relationship between two or three variables (in 2D or 3D visuals).
- **Bubble Chart**: The bubble chart is an enhanced scatter plot. By varying the circles, one can add additional data dimensions, offering more enriched data.

Specialized Charts & Graphs

- **Histogram:** Histograms are used to show the frequency distribution of 1+ variables. The x-axis shows categories, and the y-axis the frequencies.
- **Gantt Chart:** A special case of bar charts used to portray project timelines, project tasks/activity durations, and overlap among the tasks/activities.
- **PERT Chart**: developed primarily to simplify the scheduling of large/complex projects. A PERT chart shows relationships among project activities/tasks.
- **Geographic Map**: Used when the data set includes any kind of location data.

Specialized Charts & Graphs

- **Bullet**: Used to show progress towards a goal. It is a variation of a bar chart.
- **Heat Map**: A visual that illustrates the comparison of continuous values across two categories using color.
- **Highlight Table**: Two-dimensional tables with cells populated with numerical values and gradients of colors.
- **Tree Map**: A tree map displays hierarchical data as a set of nested rectangles. Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing subbranches.

Which Chart/Graph should you use?

- The capabilities of the charts helps select the proper chart for a specific task..

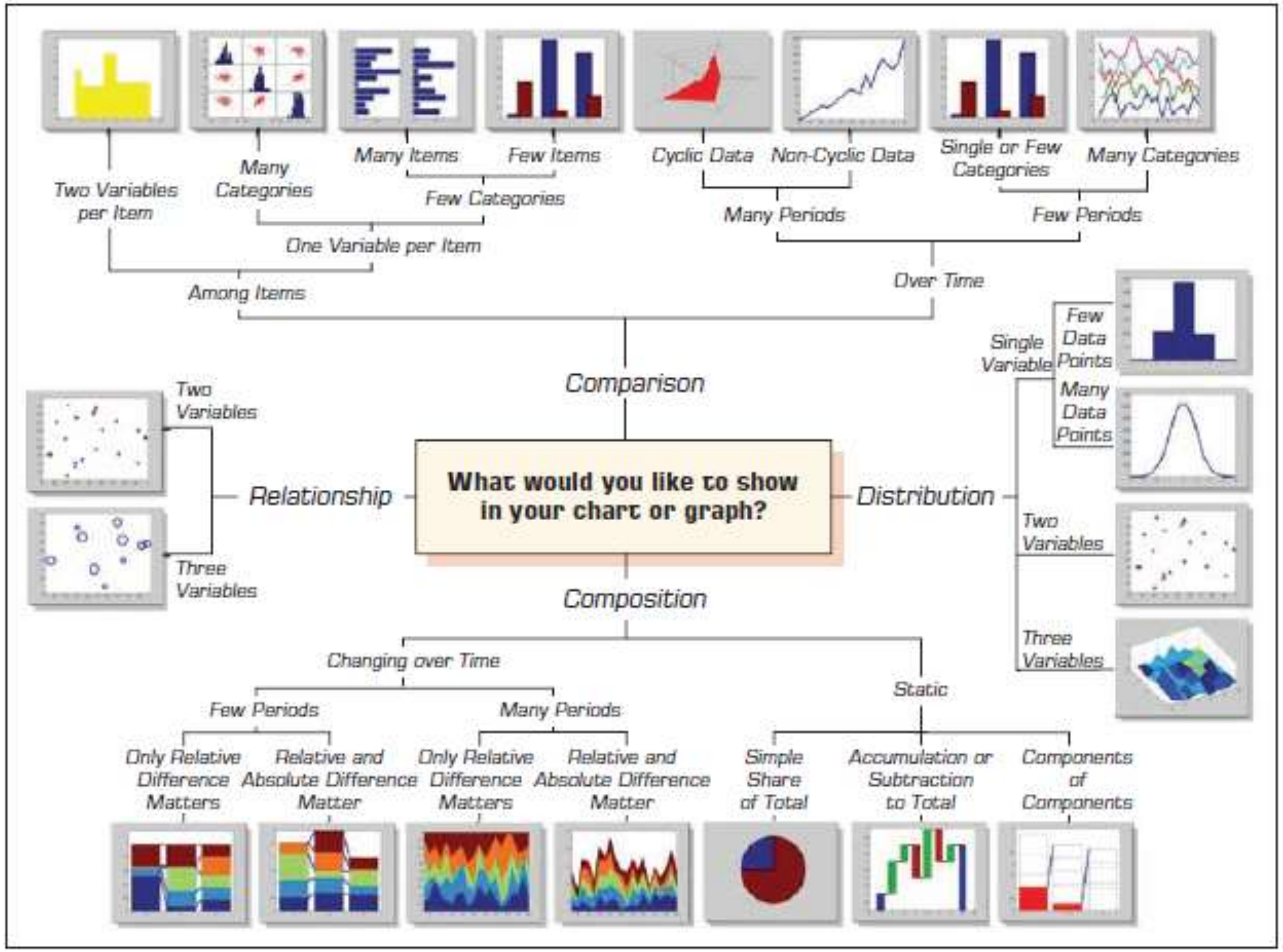


FIGURE 3.21 **A Taxonomy of Charts and Graphs.**
Source: Adapted from Abela, A. (2008). *Advanced*

Which Chart/Graph should you use?

- The **taxonomic structure** is organized around the **purpose** of the chart or graph.
- The taxonomy divides the purpose into **four** different types—relationship, comparison, distribution, and composition—and further divides the branches into subcategories based on the number of variables involved and time dependency of the visualization.
- The current trend is to combine and animate these charts for better-looking and more intuitive visualization of today's complex and volatile data sources.

3.10 Emergence of Visual Analytics

- Overview
- Visual Analytics
- What is a good story?
- Analysis as a Story
- High-Powered Visual Analytics Environments

Overview

- In BI and analytics, the key challenge of visualization is representation of large, complex data sets with multiple dimensions and measures.
- Typical charts, graphs, and visual elements used involve two dimensions, sometimes three, and fairly small subsets of data sets.
- In contrast, data in visual systems reside in a data warehouse. At a minimum, these warehouses involve a range of dimensions (e.g., product, location, organizational structure, time), a range of measures, and millions of data cells.
- In an effort to address these challenges, a number of researchers have developed a variety of new visualization techniques.

Visual Analytics

- What is meant by visual analytics is the combination of visualization and predictive analytics.
- **Information visualization** is aimed at answering “What happened?” and “What is happening?” and is closely associated with BI (routine reports, scorecards, and dashboards)
- **Visual analytics** is aimed at answering “Why is it happening?” and “What is more likely to happen?” and is usually associated with business analytics (forecasting, segmentation, correlation analysis).

What is a Good Story?

- Most people can easily remember a funny story because it contains certain characteristics and components, such as:
  - Good characters.
  - The character is faced with a challenge that is difficult but believable.
  - There are hurdles that the character overcomes.
  - The outcome or prognosis is clear by the end of the story.
  - The situation may not be resolved—but the story has a clear endpoint.

Analysis as a Story

- With story elements in place, write out the **storyboard**, that represents the structure and form of your story.
- The storyboard will help you think about the best analogies, clearly set up challenge or opportunity, and finally see the flow and transitions needed.
- By following the **best practices**, you can get people to focus on your message:

- Think of your analysis as a story—use a story structure.
- Be authentic—your story will flow.
- Be visual—think of yourself as a film editor.
- Make it easy for your audience and you.
- Invite and direct discussion.

High-Powered Visual Analytics Environments

- There is a movement towards highly efficient visualization systems. SAS Visual Analytics, is a very high-performance computing, in-memory solution for exploring massive amounts of data in a very short time.

- **Benefits proposed by the SAS analytics platform are the following:**

  - Empowers users with data exploration techniques and approachable analytics to drive improved decision making.
  - Has easy-to-use, interactive interfaces that broaden the audience for analytics.
  - Improves information sharing and collaboration.
  - Liberates IT by giving users a new way to access the information they need.
  - Provides room to grow at a self-determined pace.

E. High-Powered Visual Analytics Environments

FIGURE 3.26 A Screenshot from SAS Visual Analytics. Source: Copyright © SAS Institute, Inc. Used with permission.

FIGURE 3.25 An Overview of SAS Visual Analytics Architecture. Source: Copyright © SAS Institute, Inc. Used with permission.

3.11 Information Dashboards

- Overview
- Dashboard Design
- Best Practices in Dashboard Design
- Benchmark KPIs with Industry Standards
- Wrap the Dashboard Metrics with Contextual Metadata
- Validate the Dashboard Design by a Usability Specialist
- Prioritize and Rank Alerts/Exceptions Streamed to the Dashboard
- Enrich the Dashboard with Business-User Comments
- Present Information in Three Different Levels
- Pick the Right Visual Construct Using Dashboard Design Principles
- Provide for Guided Analytics

Overview

- **Information dashboards** are common components of BI platforms, business performance management systems, and measurement software suites.
- Dashboards provide visual displays of important information arranged on a single screen so that the information can be digested at a single glance.
- This executive dashboard shows functional groups surrounding the products intended to give executives a quick and accurate idea of what is going on within the organization.

FIGURE 3.27 A Sample Executive Dashboard. *Source:* A Sample Executive Dashboard from Dundas Data Visualization, Inc., www.dundas.com, reprinted with permission.

Dashboard Design

- Today, it would be rather unusual to see a large company using a BI system that does not employ some sort of performance dashboards.
- The most distinctive feature of a dashboard is its **three layers** of information:
  - Monitoring: Graphical, abstracted data to monitor key performance metrics
  - Analysis: Summarized dimensional data to analyze the root cause of problems
  - Management: Data that identifies actions needed to resolve a problem
- "The fundamental challenge of dashboard design is to display all the required information on a single screen."
- Because of these layers, dashboards pack a large amount of information into a single screen

Best Practices in Dashboard Design

- **Data** is one of the most important things to focus on in dashboard design.
- Even if a dashboard's appearance looks professional, is aesthetically pleasing, and includes graphs and tables, it is also important to ask about the data:

  - Are they reliable?
  - Are they timely?
  - Are any data missing?
  - Are they consistent across all dashboards?

Benchmark KPIs with Industry Standards

- Many customers want to know if the metrics they are measuring are the right metrics to monitor.
- Sometimes customers have found that the metrics they are tracking are not the right ones to track.
- Doing a gap assessment with industry benchmarks aligns you with industry best practices.

E. Wrap the Dashboard Metrics with Contextual Metadata

- When a report or a visual dashboard/scorecard is presented to business users, questions remain unanswered. The following are some examples:

- Where did you source these data?
- While loading the data warehouse, what percentage of the data was rejected/ encountered data quality problems?
- Is the dashboard presenting "fresh" information or "stale" information?
- When was the data warehouse last refreshed?
- When is it going to be refreshed next?
- Were any high-value transactions that would skew the overall trends rejected as a part of the loading process?

Validate the Dashboard Design by a Usability Specialist

- In most dashboard environments, the **dashboard** is designed by a tool specialist without giving consideration to usability principles.
- Even though it is a well-engineered data warehouse that can perform well, many business users do not use the dashboard.
- It is perceived as not being user friendly, leading to poor adoption of the infrastructure and change management issues.
- Up-front validation of the dashboard design by a usability specialist can mitigate this risk.

Prioritize and Rank Alerts/Exceptions Streamed to the Dashboard

- There are tons of raw data, having a mechanism by which important exceptions are proactively pushed to the information consumers is important.
- A **business rule** can be codified, which detects the alert pattern of interest.
- It can be coded into a program, using database-stored procedures, which crawl through the fact tables and detect patterns needing immediate attention.
- This way, information finds the business user as opposed to the business user polling the fact tables for the occurrence of critical patterns.

H. Enrich the Dashboard with Business-User Comments

- In the event that dashboard information is presented to multiple business users, a small text box can be provided to capture the comments from an end user's perspective.
- This can often be tagged to the dashboard to put the information in context, adding perspective to the structured KPIs being rendered.

I. Present Information in Three Different Levels

- Information can be presented in **three** layers depending on the granularity of the information:
  - the visual dashboard level
  - the static report level
  - the self-service cube level
- When a user navigates the dashboard, a simple set of 8 to 12 KPIs can be presented, which would give a sense of what is going well and what is not.

J. Pick Right Visual Construct Using Dashboard Design Principles

- In a dashboard, some information is presented best with bar charts, or time-series line graphs.
- When presenting correlations, a scatter plot is useful. Sometimes merely rendering it as simple tables is effective.
- Once the dashboard design principles are explicitly documented, all the developers working on the front end can adhere to the same principles while rendering the reports and dashboard.

K. Provide for Guided Analytics

- In an organization, users can be at various levels of analytical maturity.
- The capability of the dashboard can be used to guide a business user to access the same navigational path as that of an analytically savvy business user.

Practical Portion

Chapter 6-9: Analyzing Data with PivotTables and Pivot Charts & Dashboard development

Excel Data Analysis For Dummies

Practical Portion

Chapter 6: Analyzing Table Data with Functions

Excel Data Analysis For Dummies

Chapter 6: Analyzing Table Data with Functions

- Terms
- DGET Function
- Other Functions

  (DMAX/DMIN, & DPRODUCT, DCOUNT & DAVERAGE, DSTDEV & DVAR)

Terms

- **Database**: The range of cells that make up the table you want to work with. You can use either the table name or the table range address. If you go with the table name, be sure to reference the entire table by using the syntax Table[#All]
- **Field**: A reference to the table column on which you want to perform the operation. You can use either the column header or the column number (where the leftmost column is 1, the next column is 2, and so on). If you use the column name, enclose it in quotation marks (for example, “Unit Price”).
- **Criteria**: The range of cells that hold the criteria you want to work with. You can use either a range name, if one is defined, or the range address.

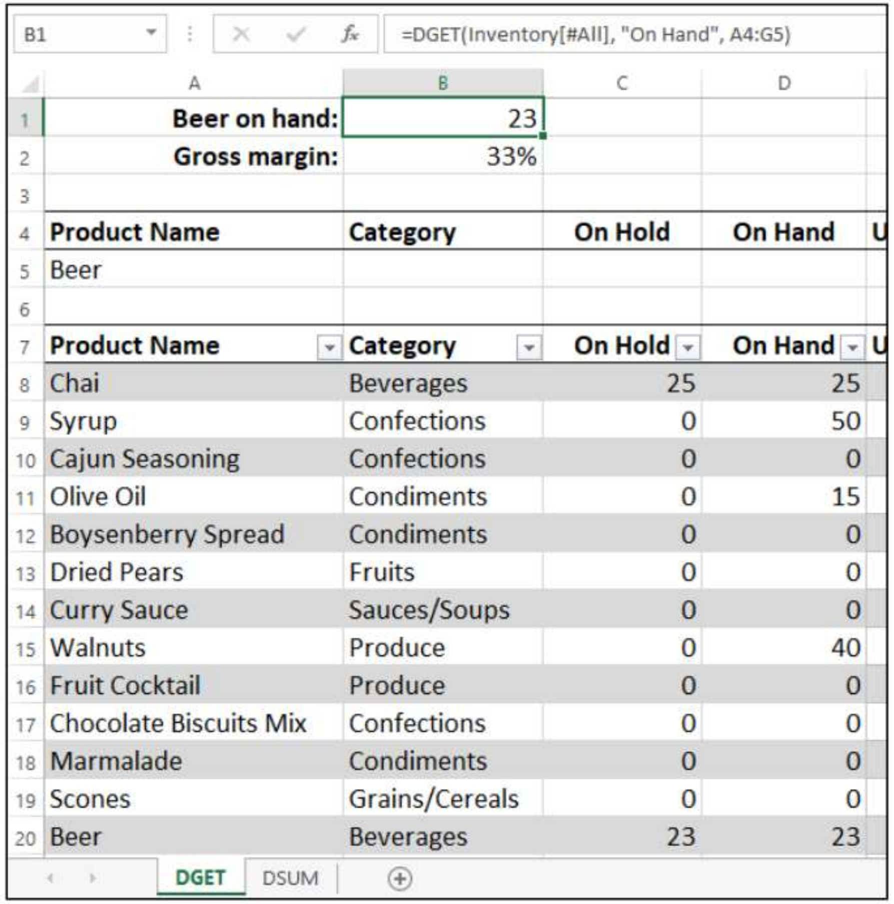
DGET Functions

- Take a look at the workbook DS498_week5-Ch6.
- The DGET function retrieves a value from a table according to the criteria you specify. Its syntax is:

  DGET(database, field, criteria)

- For example, suppose you want to know how many units are on hand of the product Beer.
- To set up the criteria range, you enter Beer below the Product Name field.
- With that done, you can build your DGET function (cell B1):

  DGET(Inventory[#All], "On Hand", A4:G5)



B1 =DGET(Inventory[#All], "On Hand", A4:G5)

	A	B	C	D
1	Beer on hand:	23		
2	Gross margin:	33%		
3				
4	Product Name	Category	On Hold	On Hand
5	Beer			
6				
7	Product Name	Category	On Hold	On Hand
8	Chai	Beverages	25	25
9	Syrup	Confections	0	50
10	Cajun Seasoning	Confections	0	0
11	Olive Oil	Condiments	0	15
12	Boysenberry Spread	Condiments	0	0
13	Dried Pears	Fruits	0	0
14	Curry Sauce	Sauces/Soups	0	0
15	Walnuts	Produce	0	40
16	Fruit Cocktail	Produce	0	0
17	Chocolate Biscuits Mix	Confections	0	0
18	Marmalade	Condiments	0	0
19	Scones	Grains/Cereals	0	0
20	Beer	Beverages	23	23

DGET DSUM

FIGURE 6-2: Use DGET to retrieve a value from a table based on your criteria.

Other Functions

- The **DSUM** function adds values from a table based on the criteria you specify
- The **DCOUNT** function counts numeric values in the field, based on a specified criteria
- The **DCOUNTA** function counts nonblank items in the field, based on a specified criteria
- The **DAVERAGE** function calculates an average for values that meet specified criteria
- The **DMAX** and **DMIN** functions find the largest and smallest values, respectively, in a table column for those rows that match specified criteria
- The **DPRODUCT** function is used to perform table multiplication
- The **DSTDEV** and **DSTDEVP** functions calculate the standard deviation
- The **DVAR** and **DVARP** functions calculate column variance

Practical Portion

Chapter 7: Creating and Using PivotTables

Excel Data Analysis For Dummies

Chapter 6: Analyzing Table Data with Functions

- Understanding PivotTables
- Steps of Building a PivotTable from an Excel Table
- Building a PivotTable from an Excel Table
- Creating a PivotTable from External Data
- Refreshing PivotTable Data
- Refreshing PivotTable Data Automatically
- Adding Multiple Fields to a PivotTable Area
- Pivoting a Field to a Different Area
- Grouping Numeric Values
- Grouping date and time values
- Grouping text values
- Applying a report filter
- Filtering row or column items
- Applying a Value Filter
- Filtering a PivotTable with a slicer

Understanding PivotTables

- Take a look at the workbook DS498_week5-Ch8.
- PivotTables take a large amount of information and condense that data into a report that tells you something useful or interesting.
- PivotTables help you analyze large amounts of data by performing three operations:
  - **Grouping**: automatically groups large amounts of data into smaller chunks.
  - **Summarizing**: Displays summary calculations for each group.
  - **Filtering**: enables you to view just a subset of the data.

SUMMER SALES PROMOTION - ORDERS

Date	Product	Quantity	Net $	Promotion	Advertisement
2019-06-01	Smartphone case	11	$119.70	1 Free with 10	Social media
2019-06-01	HDMI cable	6	$77.82	Extra Discount	Blog network
2019-06-01	USB car charger	15	$100.95	Extra Discount	Search
2019-06-01	HDMI cable	11	$149.71	1 Free with 10	Blog network
2019-06-02	USB car charger	22	$155.40	1 Free with 10	Blog network
2019-06-02	USB car charger	3	$20.19	Extra Discount	Search
2019-06-02	Earbuds	5	$33.65	Extra Discount	Social media
2019-06-02	Smartphone case	22	$239.36	1 Free with 10	Search
2019-06-02	HDMI cable	10	$129.70	Extra Discount	Blog network
2019-06-05	USB car charger	22	$155.40	1 Free with 10	Blog network
2019-06-05	Smartphone case	8	$82.96	Extra Discount	Social media
2019-06-05	Smartphone case	22	$239.40	1 Free with 10	Social media
2019-06-05	Earbuds	55	$388.50	1 Free with 10	Blog network
2019-06-05	USB car charger	25	$168.25	Extra Discount	Search
2019-06-05	HDMI cable	22	$299.42	1 Free with 10	Blog network
2019-06-06	USB car charger	33	$256.41	1 Free with 10	Blog network
2019-06-06	Smartphone case	11	$119.70	1 Free with 10	Blog network

FIGURE 7-1: Some great data, but how do you make sense of it?

Promotion	(All)			
Sum of Quantity	Column Labels			
Row Labels	Blog network	Search	Social media	Grand Total
Earbuds	555	562	322	1439
HDMI cable	719	587	402	1708
Smartphone case	546	460	338	1344
USB car charger	1596	1012	752	3360
Grand Total	3416	2621	1814	7851

FIGURE 7-2: The PivotTable creates order out of data chaos.

Understanding PivotTables

- **Row area**: Displays vertically the unique values from a field in your data.
- **Column area**: Displays horizontally the unique values from a field in your data.
- **Value area**: Displays results of the calculation applied to a numeric field in your data.
- **Row field header**: Identifies field contained in row area.
- **Column field header**: Identifies field contained in column area.
- **Value field header**: Specifies calculation and field used in the value area.
- **Filter area**: Displays a drop-down list that contains the unique values from a field.

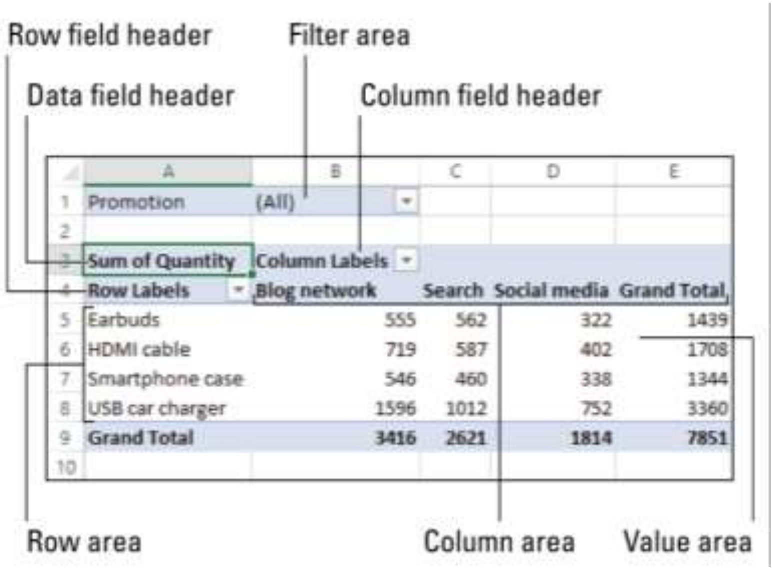


FIGURE 7-3: The features of a typical PivotTable.

Steps of Building a PivotTable from an Excel Table

1. Select a cell in the table that you want to use as the source data.
2. Choose Design ➪ Summarize with PivotTable.
3. Select the New Worksheet radio button.
4. Click OK.
5. Drag a text field and drop it inside the Rows area.
6. Drag a numeric field and drop it inside the Values area.
7. If desired, drag fields and drop them in the Columns area and the Filters area.

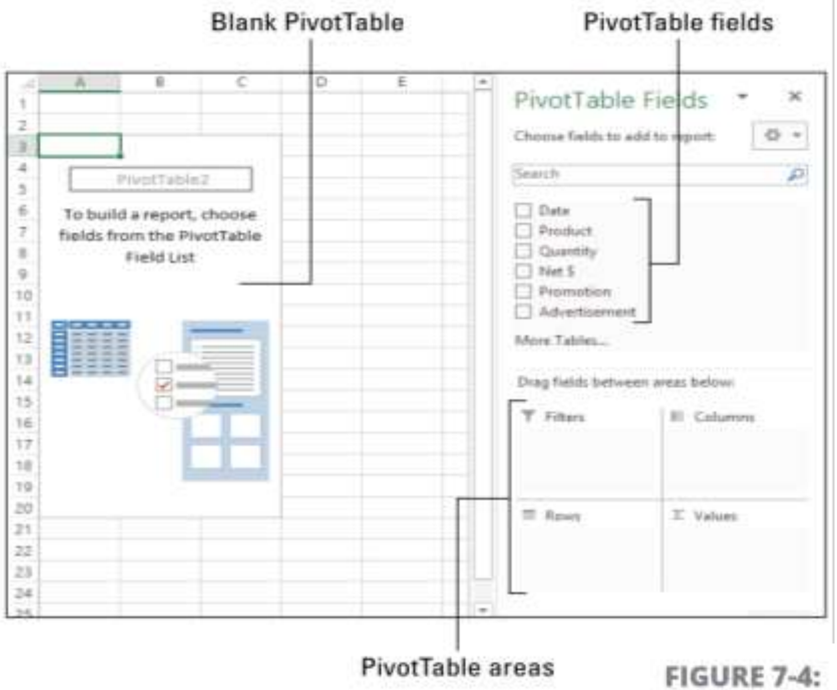


FIGURE 7-4: You start with a blank PivotTable and the PivotTable Fields pane.

Building a PivotTable from an Excel Table

- Figure 7-5 shows a complete PivotTable, with fields in all four areas.
- To remove a field from a PivotTable area:
  - Drag the field out of the PivotTable Fields pane.
  - Click the field button in the PivotTable Fields pane to drop down the field menu, and then click Remove Field.

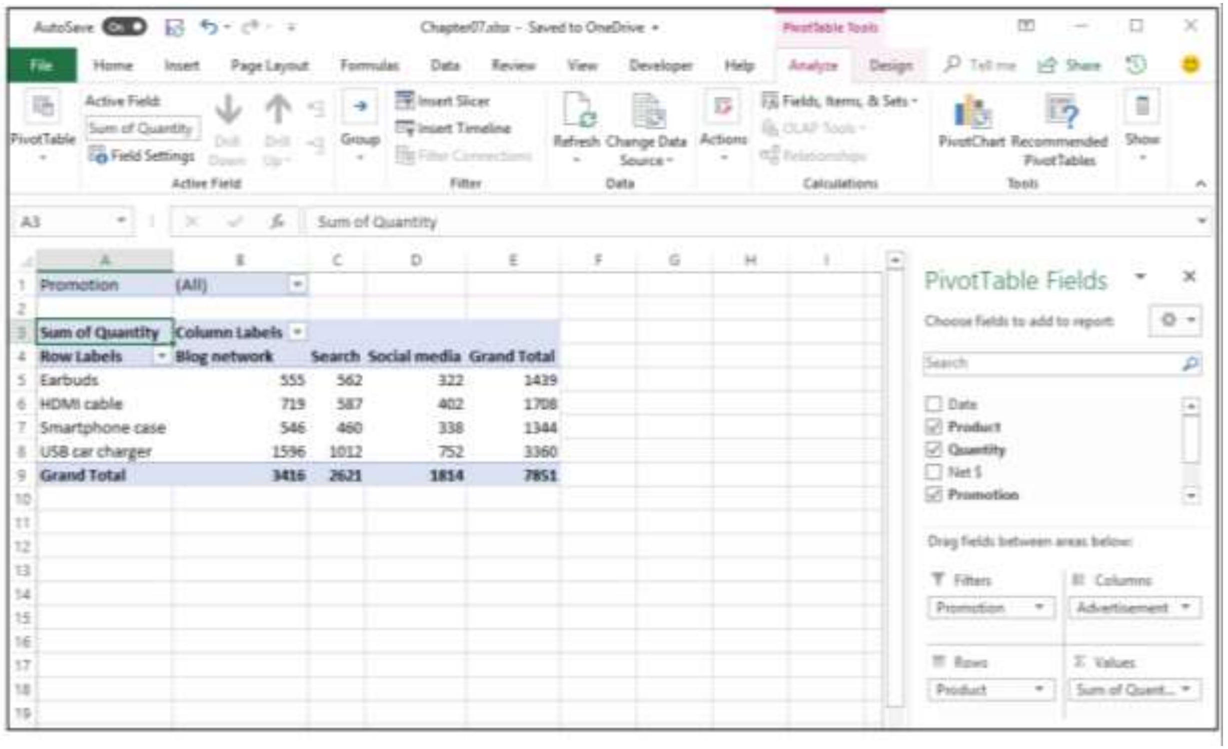


FIGURE 7-5: The features of a typical PivotTable.

Creating a PivotTable from External Data

Here are the steps to follow to build a PivotTable based on an external data source:

1. Choose Data ➪ Get Data ➪ From Other Sources ➪ From Microsoft Query.
2. Select the Use the Query Wizard to Create/Edit Queries check box.
3. On the Databases tab, select the database you want to query and then click OK.
4. Add the columns you want to work with to the Columns in Your Query list and then select Next button.
5. (Optional) Filter your data based on multiple filters by selecting the And or Or radio buttons; then select Next.

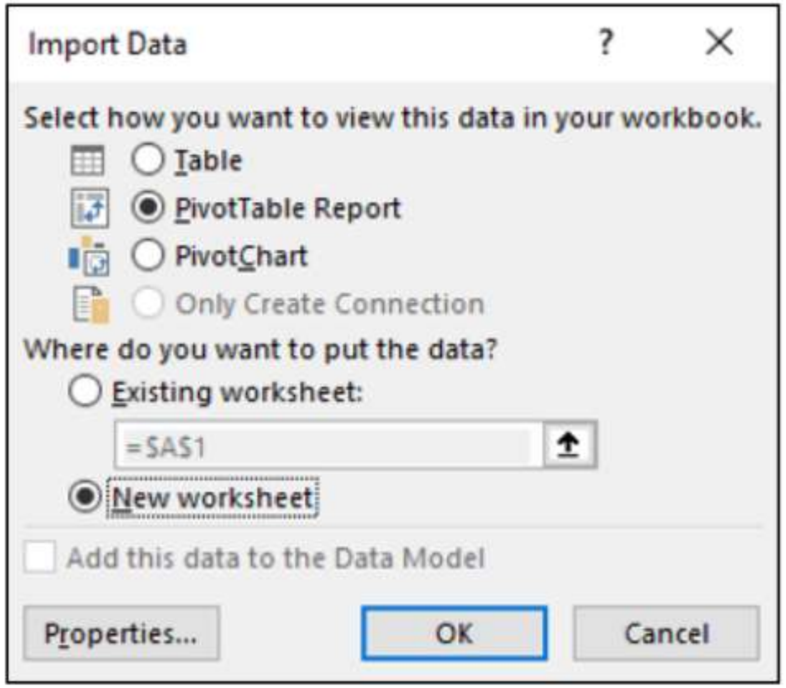


FIGURE 7-6: Import the external data to a PivotTable Report.

Creating a PivotTable from External Data

6. (Optional) Choose a sort order for the query result data from the Query Wizard - Sort Order dialog box and then click Next.
7. Select the Return Data to Microsoft Excel radio button and then click the Finish button.
8. Select the PivotTable Report radio button, select the New Worksheet radio button, and then click OK.
9. Drag a text field and drop it inside the Rows area.
10. Drag a numeric field and drop it inside the Values area.
11. If desired, drag fields and drop them in the Columns area and the Filters area.

Refreshing PivotTable Data

- Refreshing a PivotTable ensures that the data analysis represented is up-to-date.
- Excel offers two methods for refreshing a PivotTable:
    - A manual refresh is one that you perform
    - An automatic refresh is one that Excel handles for you
- To refresh your PivotTable data manually, you have two choices:
- Update a single PivotTable: Select any cell inside the PivotTable and then choose Analyze ➪ Refresh (or Alt+F5)
- Update every PivotTable in the workbook: Select a cell inside any PivotTable and then choose Analyze ➪ Refresh ➪ Refresh All (or Ctrl+Alt+F5)

Refreshing PivotTable data automatically

Refreshing the PivotTable automatically:

1. Select any cell inside the PivotTable.
2. Choose Analyze ➪ PivotTable ➪ Options.
3. Click the Data tab.
4. Select the Refresh Data When Opening the File. Click OK.

If the PivotTable is based on external data, you can set up an automatic refresh at specified intervals:

1. Select any cell inside the PivotTable. Choose Analyze ➪ Refresh ➪ Connection Properties.
2. Select the Refresh Every check box, use the spin buttons to specify the refresh interval, in minutes. Click OK.

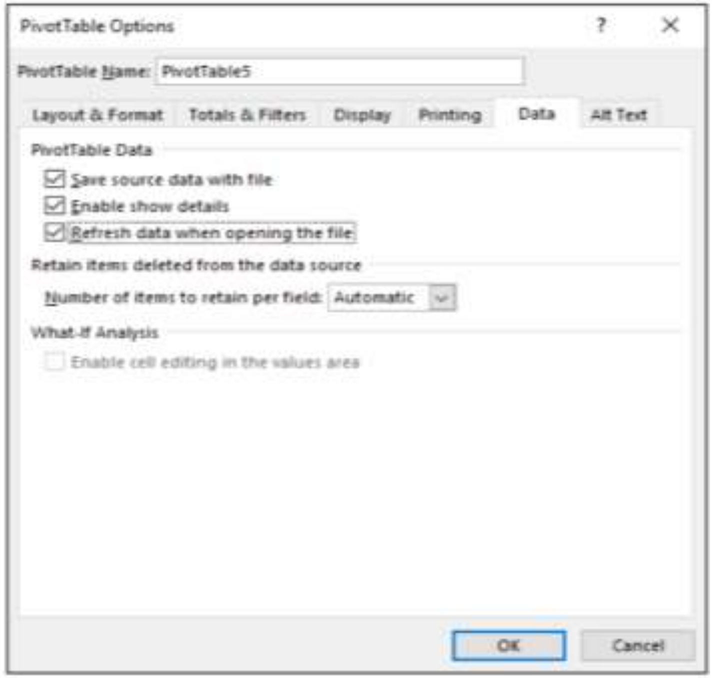


FIGURE 7-7: Import the external data to a PivotTable Report.

Adding Multiple Fields to a PivotTable Area

- You can add two or more fields to any of the PivotTable areas.
- Select a cell within the PivotTable, and use any of the following techniques to add another field to the PivotTable area:
    - **Add a field to the Rows area**: In the PivotTable Fields pane, select the check box of the text or date field that you want to add.
    - **Add a field to the Value area:** In the PivotTable Fields pane, select the check box of the numeric field that you want to add.
    - **Add a field to any area**: In the PivotTable Fields pane, drag the field and drop it inside the box of the area where you want the field to appear.
- After you add a second field to the row or column area, you can change the field positions to change the PivotTable view.
- In the PivotTable Fields pane, drag the button of the field you want to move out of the Rows/Columns box and drop the field above or below an existing field button.

Pivoting a Field to a Different Area

- You can move a PivotTable's fields from one area of the PivotTable to another. This enables you to view your data from different perspectives.
- One way to pivot the data is to move fields between the row and column areas.
- You can also pivot data by moving a row or column field to the filter area.
- Either way, you perform the pivot by dragging the field from its current box in the PivotTable Fields pane and then dropping it inside the area where you want it moved.

Grouping numeric values

- To make a PivotTable easier to work with, you can group items using the following steps:

1. Select any item in the numeric field you want to group.
2. Choose Analyze ➪ Group ➪ Group Field. The Grouping dialog box appears.
3. Use the Starting At text box to enter the starting numeric value.
4. Use the Ending At text box to enter the ending numeric value.
5. In the By text box, enter the size you want to use for each grouping.
6. Click OK.

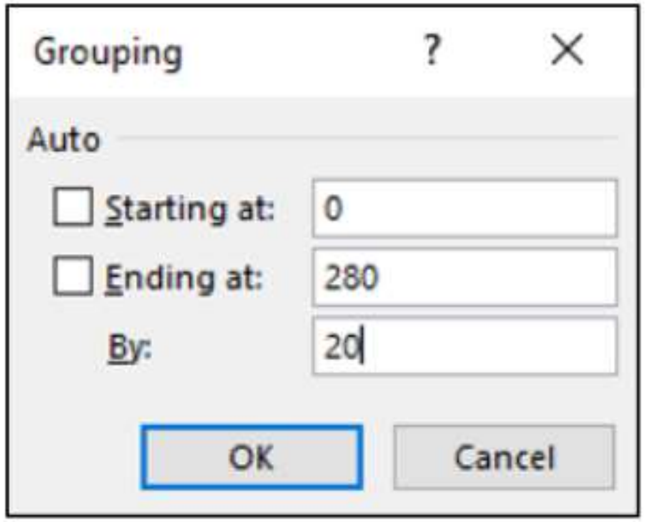


FIGURE 7-8: The Grouping dialog box.

Grouping date and time values

If your PivotTable includes a field with date/time, you can use the grouping feature to consolidate that data into more manageable or useful groups. Follow these steps:

1. Select any item in the date or time field you want to group.
2. Choose Analyze ➪ Group ➪ Group Field.
3. In the Starting At text box, enter the starting date or time.
4. Use the Ending At text box to enter the ending date or time.
5. From the By list, select the grouping you want, such as Months for dates or Hours for times.
6. Click OK.

Grouping text values

- One common problem that arises when you work with PivotTables is that you often need to consolidate items, but you have no corresponding field in the data.
- The solution in both cases is to use the Grouping feature to create custom groups by:

1. Select the items that you want to include in the group.
2. Choose Analyze ➪ Group ➪ Group Selection.
3. Select the cell that contains the group label, type a new name for the group, and then press Enter.
4. Repeat Steps 1 to 3 for the other items in the field until you have created all groups.

Applying a report filter

- Each PivotTable report displays a summary for all the records in your source data.
- You may have situations in which you need to focus more closely on some aspect of the data and can do so by taking advantage of PivotTable's report filter field.
- Follow these steps to apply a PivotTable report filter:

1. Select the filter field's drop-down arrow.
2. Select the report filter you want to view.
3. In Figure 7-9, I use the report filter list to select Canada.
4. Select OK.

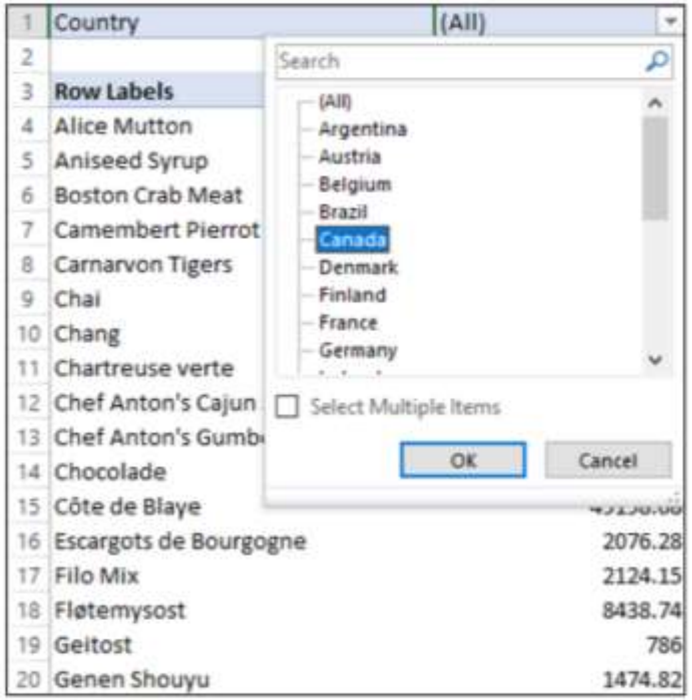


FIGURE 7-9: From the filter field's drop-down list, select a report filter.

Filtering row or column items

- When you modify a PivotTable report to show only a subset of the row or column items, you are applying a label filter to the report.

- Follow these steps to apply a label filter to row or column items:

  1. Click the drop-down arrow in the header of the field you want to filter.

  2. Select Label Filters and the filter type you want to apply, such Begins With.

  3. Type the filter criteria and then click OK.

- To remove a row or column label filter, click the drop-down arrow in the field's header and then select Clear Filter from Field, where Field is the name of the filtered field.

Applying a Value Filter

Follow these steps to apply a value filter to your PivotTable:

1. Click the drop-down arrow in the header of any row or column field.
2. Select Value Filters and then select the filter type you want to apply, such as Top 10.
3. Type the filter criteria and then click OK.

To remove a value filter, click the drop-down arrow in the header of the filtered field and then choose Value Filters ➪ Clear Filter.

Filtering a PivotTable with a slicer

A slicer is similar to a report filter, except that it's independent of any PivotTable. This means that you can use a single slicer to filter multiple PivotTables.

First, here are the steps to follow to create a slicer to filter a PivotTable:

1. Select a cell inside your PivotTable.
2. Choose Analyze ➪ Insert Slicer.
3. Select the check box beside each field for which you want to create a slicer; then click OK.
4. Select a field item that you want to include in your filter.

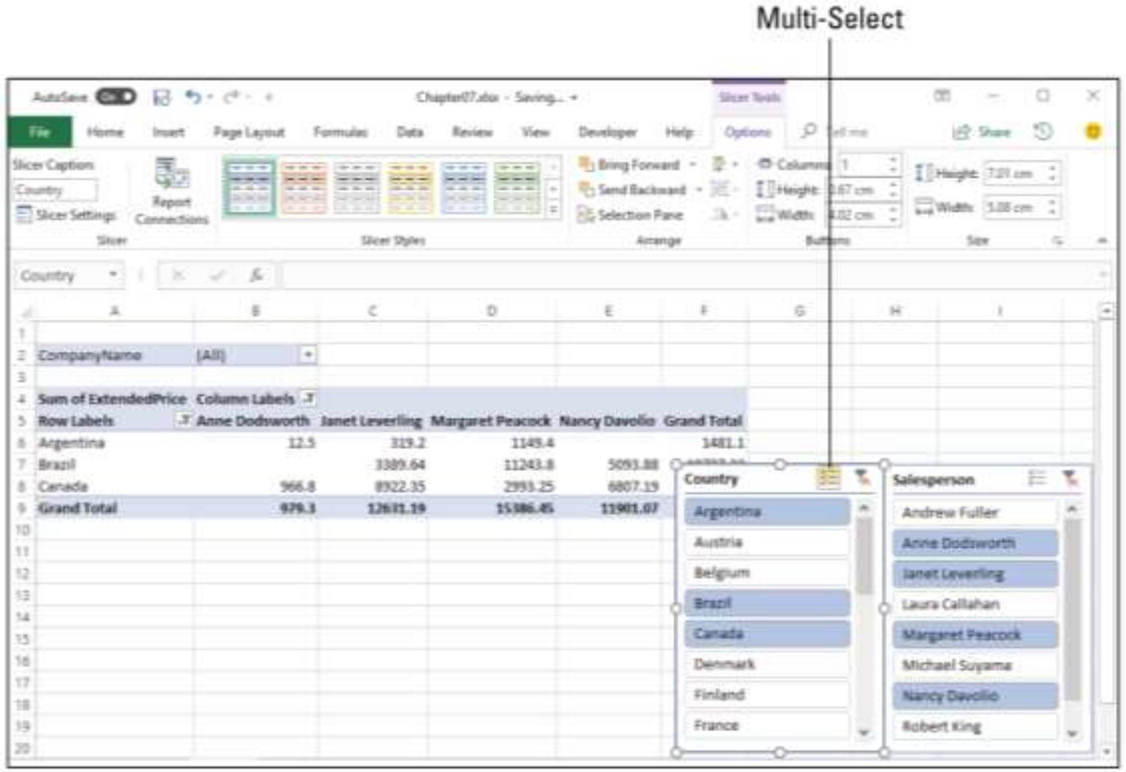


FIGURE 7-10: Excel filters the PivotTable to show just the selected items in each slicer.

Practical Portion

Chapter 8: Performing PivotTable Calculations

Excel Data Analysis For Dummies

Chapter 8: Performing PivotTable Calculations

- Messing around with PivotTable Summary Calculations
- Summary Calculations
- Changing the PivotTable summary calculation
- Trying out the difference summary calculation
- Applying a percentage summary calculation
- Adding a running total summary calculation
- Creating an index summary calculation
- Turning off subtotals for a field
- Displaying multiple subtotals for a field
- Introducing Custom Calculations
- Understanding custom calculation limitations
- Inserting a custom calculation limitations
- Inserting a custom calculation field
- Inserting a custom calculation item
- Editing a Custom Calculation
- Deleting a Custom Calculation

Messing around with PivotTable Summary Calculations

- The calculation that Excel uses to populate the PivotTable data area is called the summary calculation.
- Most of the time, the default Sum calculation will get the job done, but Excel offers lots of options for a higher analytical level.
- The default summary calculation depends on the type of field you add to the data area:
    - If you add a numeric field to the data area, Excel uses Sum as the default summary calculation.
    - If you use a text field in the data area, Excel uses Count as the default summary calculation.

Summary Calculations

Here's a list of the summary calculations:

- **Average**: Calculates the mean value in a numeric field.
- **Count**: Displays the total number of cells in the source field.
- **Count Numbers**: Displays the total number of numeric values in the source field.
- **Max**: Displays the largest value in a numeric field.
- **Min**: Displays the smallest value in a numeric field.
- **Product**: Multiplies the values in a numeric field.
- **StdDev**: Calculates the standard deviation of a population sample, which tells you how much the values in the source field vary with respect to the average.
- **StdDevp**: Calculates the standard deviation when the values in the data field represent the entire population. » Sum: Adds the values in a numeric field.
- **Var**: Calculates the variance of a population sample; the square of the standard deviation.
- **Varp**: Calculates the variance when values in the data field represent the population.

Changing the PivotTable summary calculation

Take a look at the workbook DS498_week5-Ch8a.

Here are the steps to follow to try a different summary calculation:

1. Select any cell in the data field.
2. Choose Analyze ➪ Field Settings. The Value Field Settings dialog box appears.
3. In the Summarize Value Field By list, select the summary calculation you want to use.
4. Click OK.

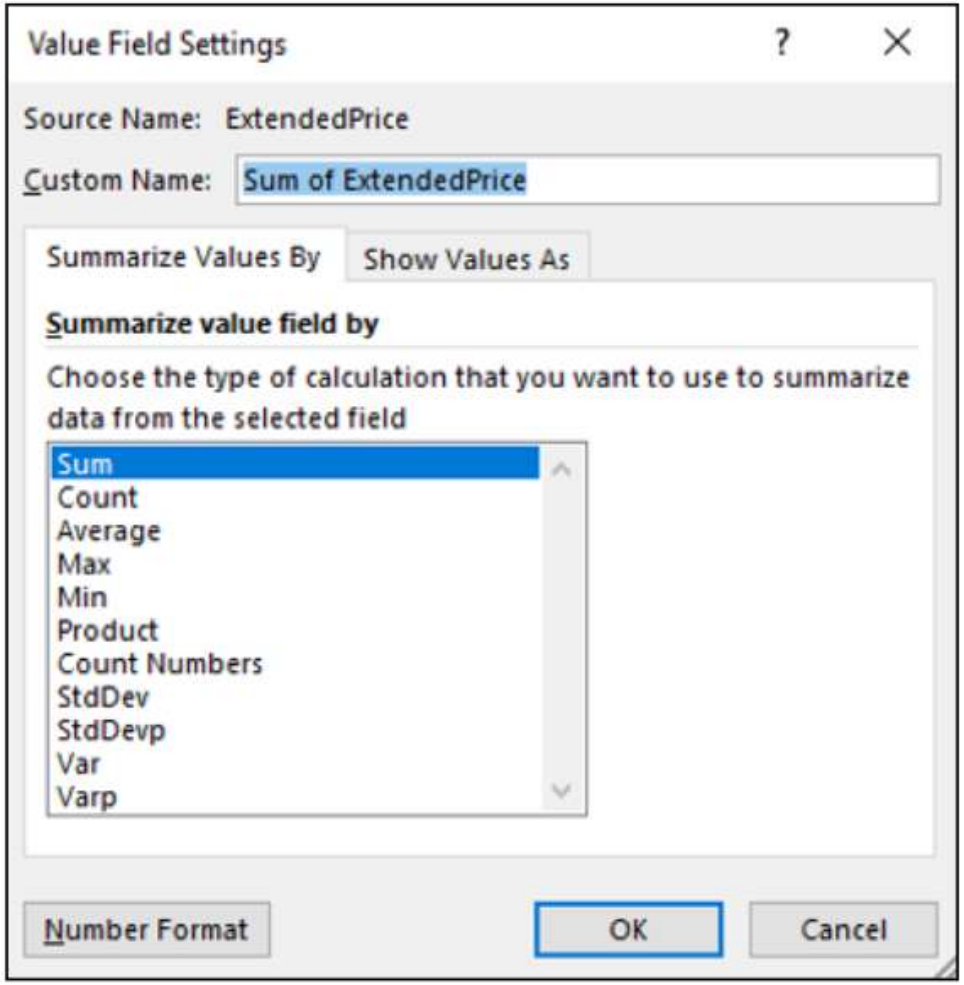


FIGURE 8-1: Use the Value Field Settings dialog box to choose a summary calculation.

Trying out the difference summary calculation

- The built-in summary calculations apply over an entire field. However, a major part of data analysis involves comparing one item with another.
- Excel offers two difference calculations that can help with this kind of analysis:
  - **Difference From:** Compares one numeric item with another and returns the difference between them
  - **% Difference From:** Compares one numeric item with another and returns the percentage difference between them
- You need to decide which field in your PivotTable to use as the base field, and which item within that field to use as the basis for all the comparisons (base item).

Sum of Extended Price	Order Date		
Customer	2018	2019	Grand Total
Alfreds Futterkiste	$2,250.50	$2,022.50	$4,273.00
Ana Trujillo Emparedados y helados	$603.20	$799.75	$1,402.95
Antonio Moreno Taquería	$1,063.20	$5,960.77	$7,023.97
Around the Horn	$6,983.75	$6,406.90	$13,390.65
Berglunds snabbköp	$11,078.57	$13,849.01	$24,927.58
Blauer See Delikatessen	$2,160.00	$1,079.80	$3,239.80
Blondel père et fils	$10,716.20	$7,817.88	$18,534.08
Bólido Comidas preparadas	$1,206.00	$3,026.85	$4,232.85
Bon app'	$10,754.89	$11,208.35	$21,963.24
Bottom-Dollar Markets	$13,171.35	$7,630.25	$20,801.60
B's Beverages	$2,910.40	$3,179.50	$6,089.90
Cactus Comidas para llevar	$1,576.80	$238.00	$1,814.80
Centro comercial Moctezuma	$100.80		$100.80
Chop-suey Chinese	$5,832.48	$6,516.40	$12,348.88

... PT-Invoices 2018 vs 2019 | Sales | PT-Sale ...

FIGURE 8-2: A PivotTable that shows sales in two years: 2018 and 2019.

Trying out the difference summary calculation

Here are the steps to apply a difference summary calculation to a PivotTable:

1. Select any cell in the value field.
2. Choose Analyze ➪ Field Settings.
3. Click the Show Values As tab.
4. In the Show Values As list, select Difference From.
5. In the Base Field list, select the field from which you want Excel to calculate the difference.
6. In the Base Item list, select a base item.
7. Click OK.

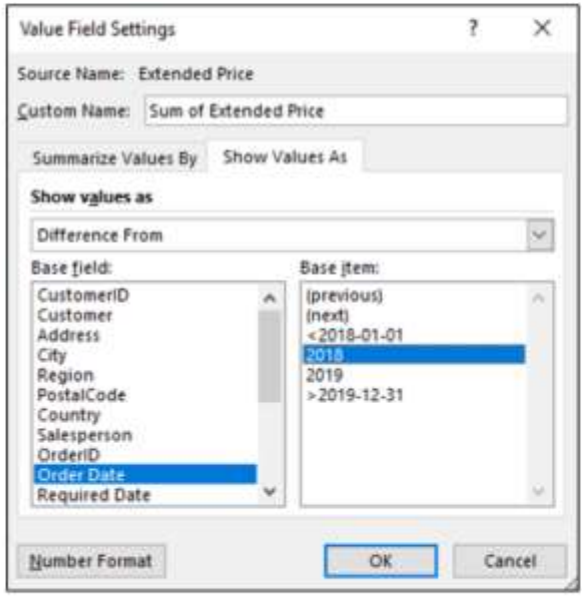


FIGURE 8-3: Use the Value Field Settings dialog box to choose a summary calculation.

Sum of Extended Price	Order Date		
Customer	2018	2019	Grand Total
Alfreds Futterkiste		-$228.00	
Ana Trujillo Emparedados y helados		$196.55	
Antonio Moreno Taquería		$4,897.57	
Around the Horn		-$576.85	
Berglunds snabbköp		$2,770.44	
Blauer See Delikatessen		-$1,080.20	
Blondel père et fils		-$2,898.32	
Bólido Comidas preparadas		$1,820.85	
Bon app'		$453.46	
Bottom-Dollar Markets		-$5,541.10	
B's Beverages		$269.10	
Cactus Comidas para llevar		-$1,338.80	
Centro comercial Moctezuma		-$100.80	
Chop-suey Chinese		$683.92	

FIGURE 8-4: The PivotTable from Figure 8-2 is now using the Difference From calculation.

Applying a percentage summary calculation

- When analyzing data, comparing two or more items as a percentage is often helpful because percentage calculations enable you to make apples-to-apples comparisons between values (Figure 8-5).

Excel offers seven percentage calculations that can help you perform this kind of analysis:

- **% Of:** Returns the percentage of each value with respect to a selected base item.
- **% of Row Total:** Returns the percentage that each value in a row represents of the total value of the row.

Sum of Sales	Quarter				
Region	1st	2nd	3rd	4th	Grand Total
East	$377,568	$343,706	$368,121	$374,260	$1,463,655
Midwest	$321,220	$307,992	$365,790	$370,213	$1,365,215
South	$346,345	$330,999	$376,658	$355,542	$1,409,544
West	$411,647	$390,493	$361,091	$314,653	$1,477,884
Grand Total	**$1,456,780**	**$1,373,190**	**$1,471,660**	**$1,414,668**	**$5,716,298**

... PT-Sales By Region & Qtr | Invoices 2019 ...

FIGURE 8-5: A PivotTable that shows quarterly sales by region.

Applying a percentage summary calculation

- **% of Column Total:** Returns the percentage that each value in a column represents of the total value of the column.
- **% of Grand Total:** Returns the percentage that each value represents of the PivotTable grand total.
- **% of Parent Row Total:** If you have multiple fields in the row area, this calculation returns the percentage that each value in an inner row represents with respect to the total of the parent item in the outer row.
- **% of Parent Column Total:** If you have multiple fields in the column area, this calculation returns the percentage that each value in an inner column represents with respect to the total of the parent item in the outer column.
- **% of Parent Total:** If you have multiple fields in the row or column area, this calculation returns the percentage of each value with respect to a selected base field in the outer row or column.

Applying a percentage summary calculation

Here are the steps to follow to apply a percentage summary calculation:

1. Select any cell in the value field.
2. Choose Analyze ➪ Field Settings.
3. Click the Show Values As tab.
4. In the Show Values As list, select the percentage calculation you want to use.
5. In the Base Field list, select the field from which you want Excel to calculate the percentages.
6. In the Base Item list, select a base item.
7. Click OK.

Sum of Sales	Quarter				
Region	**1st**	**2nd**	**3rd**	**4th**	**Grand Total**
East	100.00%	91.03%	97.50%	99.12%	
Midwest	100.00%	95.88%	113.88%	115.25%	
South	100.00%	95.57%	108.75%	102.66%	
West	100.00%	94.86%	87.72%	76.44%	
Grand Total	**100.00%**	**94.26%**	**101.02%**	**97.11%**	

PT-Sales By Region & Qtr | Invoices 2019

FIGURE 8-6: The PivotTable from Figure 8-5, now using the % Of calculation.

Adding a running total summary calculation

- A running total is the cumulative sum of the values that appear in a given set of data.
- You use a running total in data analysis when you need to see a snapshot of the overall data at various points.
- Excel offers a built-in Running Total In summary calculation that you can apply to your PivotTable results.
- The Running Total In summary applies to not just the Sum calculation but also related calculations, such as Count and Average.
- Before you configure your PivotTable to use a Running Total In summary calculation, you must choose the field on which to base the accumulation, called the base field.
- This field will most often be a date field, but you can also create running totals based on other fields, such as customer, division, product, and so on.

Adding a running total summary calculation

Here are the steps to apply a Running Total In summary calculation to a PivotTable:

1. Select any cell in the value field.
2. Choose Analyze ➪ Field Settings.
3. Click the Show Values As tab.
4. In the Show Values As list, select Running Total In.
5. In the Base Field list, select the field from which you want Excel to accumulate the running totals.
6. Click OK.

OrderDate	Sum of ExtendedPrice
Jan	$53,981.61
Feb	$42,386.08
Mar	$41,921.21
Apr	$53,032.95
May	$50,506.46
Jun	$39,637.61
Jul	$44,811.34
Aug	$53,497.15
Sep	$55,629.24
Oct	$65,354.48
Nov	$44,928.54
Dec	$71,398.41
Grand Total	**$617,085.08**

FIGURE 8-7: A PivotTable showing monthly order totals.

OrderDate	Sum of ExtendedPrice
Jan	$53,981.61
Feb	$96,367.69
Mar	$138,288.90
Apr	$191,321.85
May	$241,828.31
Jun	$281,465.92
Jul	$326,277.26
Aug	$379,774.41
Sep	$435,403.65
Oct	$500,758.13
Nov	$545,686.67
Dec	$617,085.08
Grand Total	

FIGURE 8-8: The PivotTable from Figure 8-7, with the Running Total In calculation applied.

Creating an index summary calculation

- This determination of the relative importance of the results of your calculations is vital in a PivotTable.
- Excel offers the built-in Index calculation, which determines the weighted average of each cell in the PivotTable results. The formula Excel used is:

  (Cell Value) * (Grand Total) / (Row Total) * (Column Total)

- In the Index calculation results, the higher the value, the more important the cell is in the overall PivotTable report.

Sum of Quantity	Region			
CategoryName	Idaho	Oregon	Washington	Grand Total
Beverages	301	71	182	554
Condiments	240	32	99	371
Confections	359	79	72	510
Dairy Products	343	86	54	483
Grains/Cereals	247	42	25	314
Meat/Poultry	337	6	96	439
Produce	110	30	5	145
Seafood	637	35	85	757
Grand Total	**2574**	**381**	**618**	**3573**

... PT-Units By Category & Region PT-Custor ...

FIGURE 8-9: A PivotTable showing units sold by category and region.

Creating an index summary calculation

Follow these steps to apply the Index summary calculation to a PivotTable:

1. Select any cell in the value field.
2. Choose Analyze ➪ Field Settings. The Value Field Settings dialog box appears with the Summarize Values By tab displayed.
3. Click the Show Values As tab.
4. In the Show Values As list, select Index.
5. Click OK.

- To format the data field to show just two decimal places.
- Select any cell in the value field, choose Analyze ➪ Field Settings, click Number Format, select Number in the Category list, and then click OK.

Sum of Quantity	Region			
CategoryName	Idaho	Oregon	Washington	Grand Total
Beverages	0.75	1.20	1.90	1.00
Condiments	0.90	0.81	1.54	1.00
Confections	0.98	1.45	0.82	1.00
Dairy Products	0.99	1.67	0.65	1.00
Grains/Cereals	1.09	1.25	0.46	1.00
Meat/Poultry	1.07	0.13	1.26	1.00
Produce	1.05	1.94	0.20	1.00
Seafood	1.17	0.43	0.65	1.00
Grand Total	**1.00**	**1.00**	**1.00**	**1.00**

... PT-Units By Category & Region | PT-Custor ...

FIGURE 8-10: The PivotTable from Figure 8-9, with the Index calculation applied.

Turning off subtotals for a field

- Excel displays two sets of subtotals: one for the second (middle) field and one for the first (outer) field. And for every extra field you add to the row or column area, Excel mindlessly adds yet another set of subtotals.
- To reduce the complexity of the PivotTable layout turn off the subtotals by:
  1. Select any cell in the field you want to work with.
  2. Choose Analyze ➪ Field Settings.
  3. In the Subtotals group, select the None radio button.
  4. Click OK.

Displaying multiple subtotals for a field

- A common data analysis task is to view items from several different points of view.
- You can view multiple subtotals for each field, with each using a different summary calculation by:

1. Select any cell in the field you want to mess with.
2. Choose Analyze ➪ Field Settings.
3. In the Subtotals group, select the Custom radio button.
4. In the list below the Custom options, select each calculation that you want to appear as a subtotal.
5. Click OK.

CompanyName	OrderID	Sum of ExtendedPrice
Around the Horn		
	10453	$407.70
	10558	$2,142.90
	10707	$1,641.00
	10741	$228.00
	10743	$319.20
	10768	$1,477.00
	10793	$191.10
Around the Horn Sum		$6,406.90
Around the Horn Average		$355.94
Around the Horn Max		$1,060.00
Around the Horn Min		$45.00

... PT-Customer Orders | Orders

FIGURE 8-11: A PivotTable with multiple subtotals.

Introducing Custom Calculations

- A custom calculation is a formula that you define yourself to produce PivotTable values that wouldn't otherwise appear in the report.
- Custom calculations are formulas with certain restrictions imposed. A custom calculation formula begins with an equals sign, followed by 1+ operands and operators:
  - **Operands**: The values that the formula uses as the raw material for the calculation. The operands can be numbers, worksheet functions, or fields from your data source.
  - **Operators**: The symbols that the formula uses to perform the calculation.
- When building a custom calculation for a PivotTable, Excel offers two types:
  - **Calculated field:** A new data field in which the values are the result of a custom calculation formula.
  - **Calculated item:** A new item in a row or column field in which the values are the result of a custom calculation.

Understanding custom calculation limitations

Excel imposes a number of limitations on custom formulas including:

- The major limitation is with the exception of constant values such as numbers, you can't reference anything outside the Pivot-Table's source data:
- You can't use a cell reference, range address, or range name as an operand in a custom calculation formula.
- You can't use any worksheet function that requires a cell reference, range, or defined name.
- You can't use the PivotTable's subtotals, row totals, column totals, or Grand Total as an operand in a custom calculation formula.
- Field references: When you reference a PivotTable field in your formula, Excel interprets this reference as the sum of that field's values.
- Field reference problems: The fact that Excel defaults to a Sum calculation when you reference another field in your custom calculation can lead to problems.

Understanding custom calculation limitations

- A formula for a calculated item can’t reference items from any field except the one in which the calculated item resides.
- You can’t insert a calculated item into a PivotTable that has at least one grouped field. You must ungroup all the PivotTable fields before you can insert a calculated item.
- You can’t group a field in a PivotTable that has at least one calculated item.
- You can’t insert a calculated item into a filter field. Also, you can’t move a row or column field that has a calculated item into the filter area.
- You can’t insert a calculated item into a PivotTable in which a field has been used more than once.
- You can’t insert a calculated item into a PivotTable that uses the Average, StdDev, StdDevp, Var, or Varp summary calculations.

Inserting a Custom Calculated Field

Excel offers the Calculated Field feature, which provides a dialog box for you to name the field and construct the formula.

Here are the steps to follow to insert a custom calculated field into a PivotTable:

1. Select any cell inside the PivotTable’s value area.
2. Choose Analyze ➪ Fields, Items, & Sets ➪ Calculated Field.
3. In the Name text box, enter a name for the calculated field.
4. In the Formula text box, start the formula.

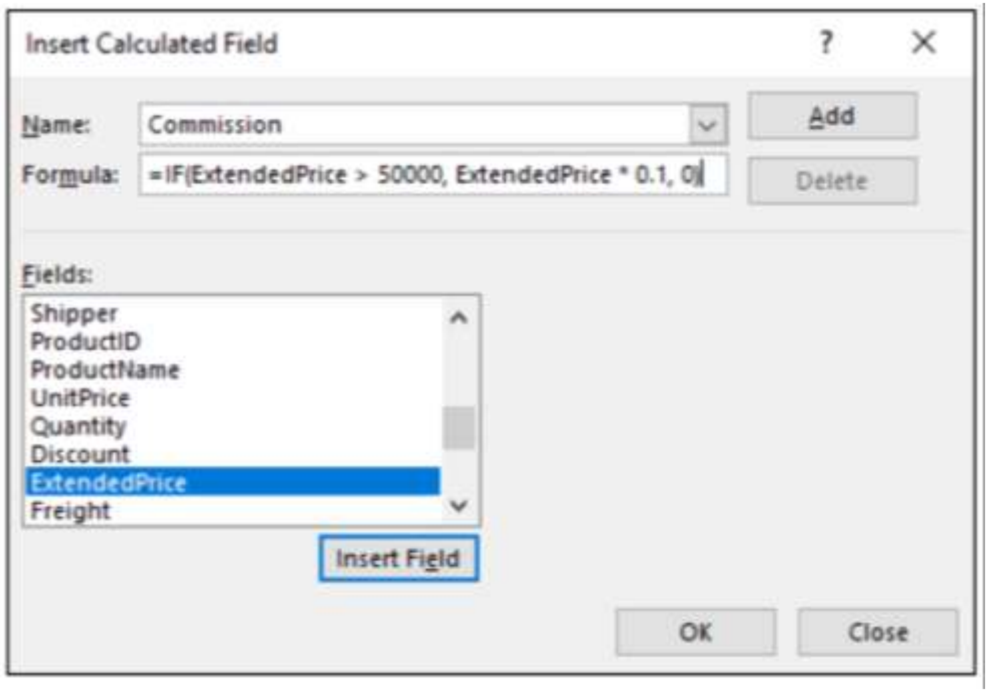


FIGURE 8-12: A custom calculated field, ready for insertion into the PivotTable.

Inserting a Custom Calculated Field

5. When you get to the point in your formula at which you need to add a field, select a field in the Fields list and then click Insert Field.
6. Keep building your formula, repeating Step 5 to add fields as needed.
7. When the formula is complete, click Add.
8. Click OK.

Row Labels	Sum of ExtendedPrice	Sum of Commission
Andrew Fuller	$70,444.14	$ 7,044.41
Anne Dodsworth	$26,310.39	$ -
Janet Leverling	$108,026.13	$ 10,802.61
Laura Callahan	$56,032.60	$ 5,603.26
Margaret Peacock	$128,809.78	$ 12,880.98
Michael Suyama	$43,126.37	$ -
Nancy Davolio	$93,148.04	$ 9,314.80
Robert King	$60,471.19	$ 6,047.12
Steven Buchanan	$30,716.44	$ -
Grand Total	**$617,085.08**	**$ 61,708.51**

... PT - Employee Total Sales

FIGURE 8-13: The custom calculated field in action.

Inserting a Custom Calculated Item

Excel offers the Calculated Item command, which displays a dialog box in which you name the item and construct the formula.

Here are the steps to follow to insert a custom calculated item into a PivotTable:

1. Select any cell inside the field to which you want to insert the item.
2. Choose Analyze ➪ Fields, Items, & Sets ➪ Calculated Item.
3. In the Name text box, enter a name for the calculated field.
4. In the Formula text box, start the formula.

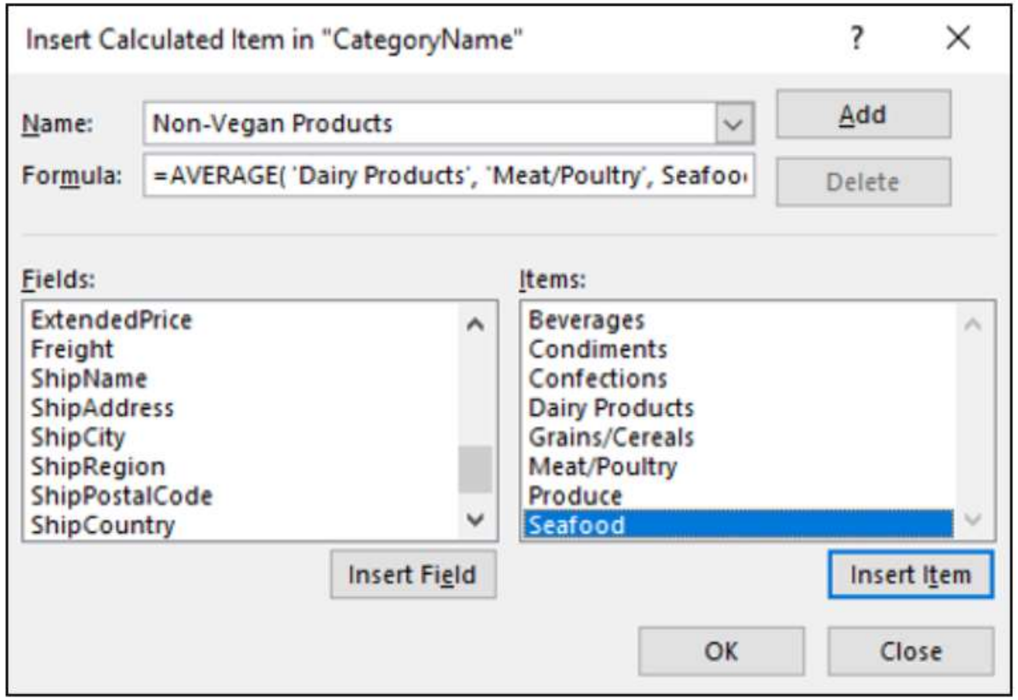
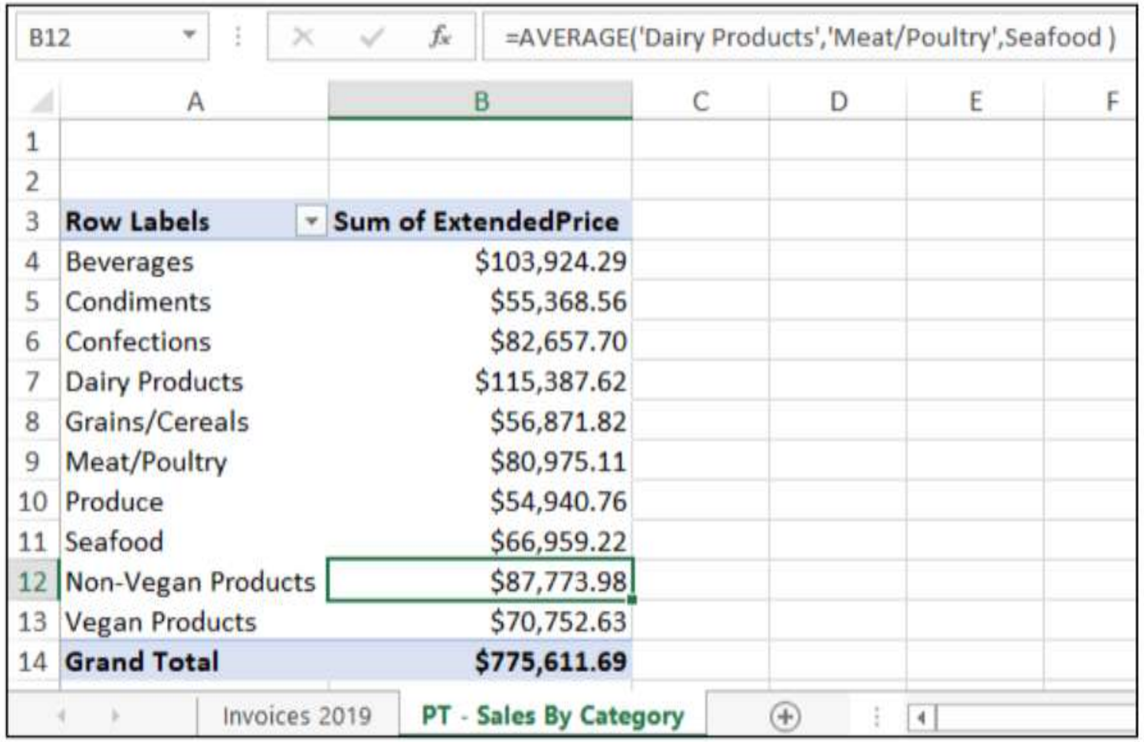


FIGURE 8-14: A custom calculated item, ready for action.

Inserting a Custom Calculated Item

5. When you get to the point in your formula at which you need to add a field, select the field in the Fields list and then click Insert Field.
6. When you get to the point in your formula at which you need to add an item, select the item in the Items list and then click Insert Item.
7. Keep building your formula, repeating Steps 5 and 6 to add fields and items as needed.
8. When formula is complete, click Add.
9. Click OK.



Row Labels	Sum of ExtendedPrice
Beverages	$103,924.29
Condiments	$55,368.56
Confections	$82,657.70
Dairy Products	$115,387.62
Grains/Cereals	$56,871.82
Meat/Poultry	$80,975.11
Produce	$54,940.76
Seafood	$66,959.22
Non-Vegan Products	$87,773.98
Vegan Products	$70,752.63
Grand Total	**$775,611.69**

FIGURE 8-15: Two custom calculated items added to the row area.

Editing a Custom Calculation

Whether your custom calculation contains an error or your data analysis needs have changed, Excel enables you to edit the formula to produce the result you want by:

1. To edit a calculated field, select any cell inside the PivotTable's data area.
2. Choose Analyze ➪ Fields, Items, & Sets ➪ Calculated Field.
3. In the Name list, select the calculation that you want to delete.
4. Edit the formula.
5. Click Modify.
6. Click OK.

Deleting a Custom Calculation

- Excel enables you to delete those calculated fields or items that you no longer need by following these steps:

1. Select any cell in the PivotTable.
2. To delete a calculated field, choose Analyze ➪ Fields, Items, & Sets ➪ Calculated Field.
3. In the Name list, select the calculation that you want to delete.
4. Click Delete.
5. Click OK.

Practical Portion

Chapter 9: Building PivotCharts

Excel Data Analysis For Dummies

Chapter 9: Building PivotCharts

- Introducing the PivotChart
- Taking a PivotChart tour
- Understanding PivotChart limitations
- Embedding a PivotChart on a PivotTable’s worksheet
- Moving PivotCharts to another sheet
- Changing the PivotChart type
- Adding data labels to your PivotChart
- Manually Sorting Data Series Or Categories
- Adding PivotChart titles
- Displaying a data table with the PivotChart

Introducing the PivotChart

The following list explains these differences and introduces important PivotChart concepts:

- **Chart Categories (X-Axis):** groups data into smaller, more manageable groups.
- **Chart Data Series**: breaks down your data in terms of a second field.
- **Chart Values (Y-Axis)**: When you add a numeric field for the summary calculation, Excel displays the results as chart values (Y-axis).
- **Dynamic PivotCharts**: is a dynamic object that you can reconfigure as needed
- **Filtering**: using unique values in another field to filter the results that appear.

- On the positive side, a PivotChart is a powerful data-analysis tool because it combines the strengths of Excel's charting capabilities with the features of a PivotTable.
- On the negative side, a PivotChart can quickly become extremely confusing when you have multiple Category fields or Data Series fields.

Taking a PivotChart tour

PivotCharts also use a number of unique terms:

- **Field buttons**: Buttons that display a drop-down list with unique values from the field that you use to filter data.
- **Category items**: Unique field values that define chart categories.
- **Category axis**: A chart's X-axis displays the category items.
- **Data series items**: The unique field values that define the chart's data series.
- **Data series axis**: The chart's Y-axis displays the values of the data series items.
- **Value area**: Displays the charted results of the calculation that Excel applied to a numeric field in your data.

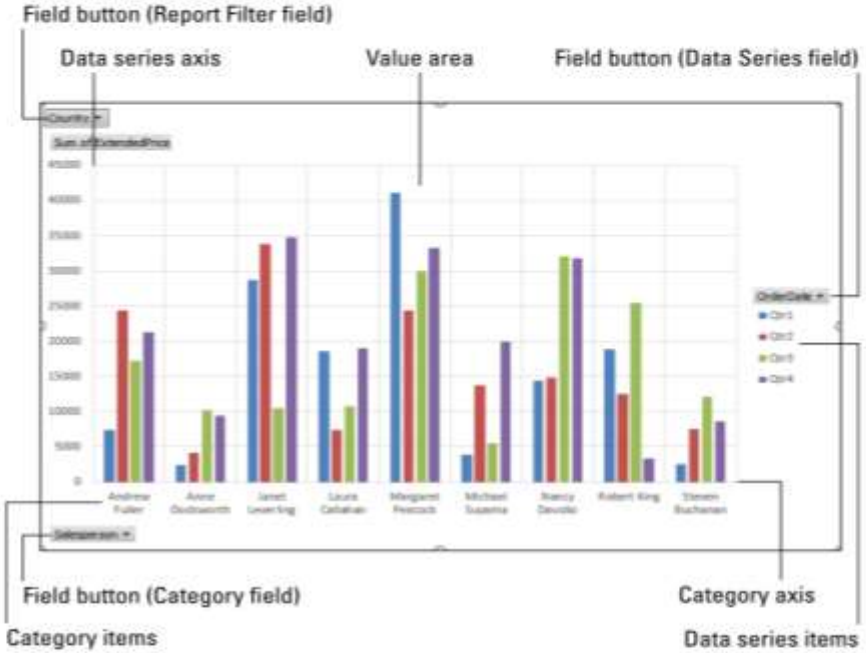


FIGURE 9-1: The major sights to see in the PivotChart landscape.

Taking a PivotChart tour

Understanding how Excel maps PivotTable's row and column areas is important:

- **Row area versus category axis**: In a PivotTable, the row area contains the unique values that Excel has extracted from a field in the source data. That is, each unique value from the source field has a corresponding category axis value.

- **Column area versus series axis**: In a PivotTable, the column area contains the unique values that Excel has extracted from a field in the source data. The PivotChart equivalent is the series axis, which corresponds to the chart's Y-axis. That is, each unique value from the source field has a corresponding data series.

- Take a look at the workbook DS498_week5-Ch9.

Understanding PivotChart limitations

- You also face a number of limitations that control the types of charts you can make and the formatting options you can apply:
  - **Chart types**: You can’t apply the types Bubble, XY (Scatter), and Stock to a PivotChart.
  - **Adding and removing fields**: After you create the PivotChart, as long as you’re working with the chart itself, you can’t add or remove fields. If you want to reconfigure the PivotChart’s fields, you have to add or remove the fields using the underlying PivotTable.
  - **Pivoting fields**: You can’t pivot the fields from one part of the PivotChart to another. If you want to pivot a field, you have to use the underlying PivotTable.
- Convincing Excel to make that PivotChart requires two steps:
  1. Select any cell in the PivotTable.
  2. Press F11.

Embedding a PivotChart on a PivotTable's worksheet

- Creating a new PivotChart on the same worksheet as an existing PivotTable is called embedding the PivotChart. Here are the steps to follow to embed a PivotChart:

    1. Select any cell in the PivotTable.
    2. Choose Analyze ➪ PivotChart.
    3. In the list of chart types on the left side of the Insert Chart dialog box, select the chart type you want.
    4. On the right side of the Insert Chart dialog box, select the chart subtype you want.
    5. Click OK.

- To move the PivotChart, move the mouse pointer over an empty part of the chart area and then drag the chart object to the new position.
- To resize the PivotChart, select the chart, move the mouse pointer over any of the of the selection handles that appear on the chart corner/sides. Drag to the required size.

Creating a PivotChart from an Excel table

- To create a PivotTable from an excel table:

1. Select a cell in the table that you want to use as the source data.
2. Choose Insert ➪ PivotChart.
3. Select the New Worksheet radio button.
4. Click OK.
5. Drag a text field and drop it inside the Axis (Categories) area.
6. Drag a numeric field and drop it inside the Values area.
7. If desired, drag fields and drop them in the Legend (Series) area and the Filters area.

FIGURE 9-2: Excel kicks things off with a blank PivotTable and PivotChart, and the PivotChart Fields pane.

PivotCharts & Columns

- To get a better understanding of a chart, you might want to know what data is represented by specific columns.
- You can find the specifics related to each column by moving the mouse pointer over the column in the plot area. Excel then displays a banner with data in the following format:

Series "Series Item" Point "Category Item" Value: Value

- For example, if the Shipper field has an item named United Package, the Salesperson field has an item named Steven Buchanan, and the value is 488, the banner shows the following:

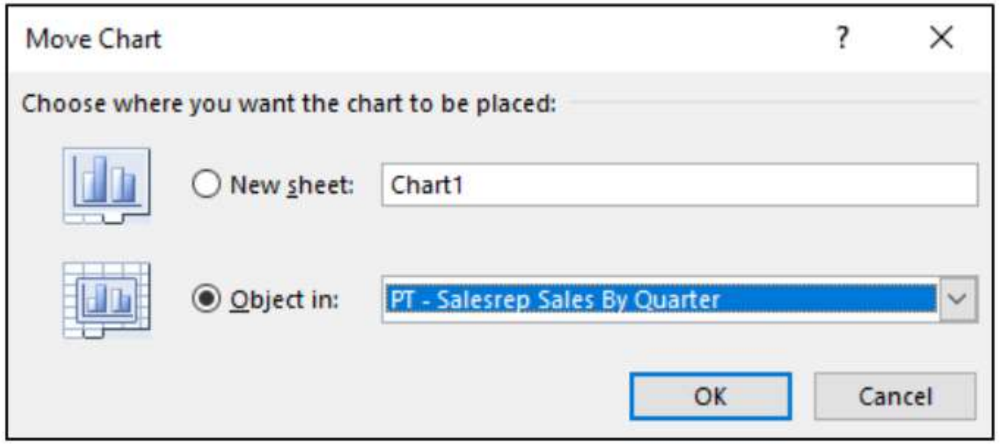
Series "United Package" Point "Steven Buchanan" Value: 488

Moving PivotCharts to another sheet

- To move your PivotChart or PivotCharts to the worksheet you prefer, follow these steps:

1. Select the PivotChart you want to move.
2. Choose Design ➪ Move Chart.
3. Select the Object In radio button and open its drop-down list to select the sheet where you want the PivotChart moved (see Figure 9-4).
4. Click OK.

FIGURE 9-4: Use the Move Chart dialog box to move a PivotChart to another worksheet.



Changing the PivotChart type

- Excel enables you to change the default PivotChart type to any of the following types: Column, Bar, Line, Pie, Area, Doughnut, Radar, Surface, Cylinder, Cone, or Pyramid.

- Follow these steps to change your PivotChart's type:

1. Select the PivotChart.
2. Choose Design ➪ Change Chart Type.
3. In the list of chart types on the left side of the Insert Chart dialog box, select the chart type you want to use.
4. On the right side of the Insert Chart dialog box, select the Chart subtype you want to use.
5. Click OK.

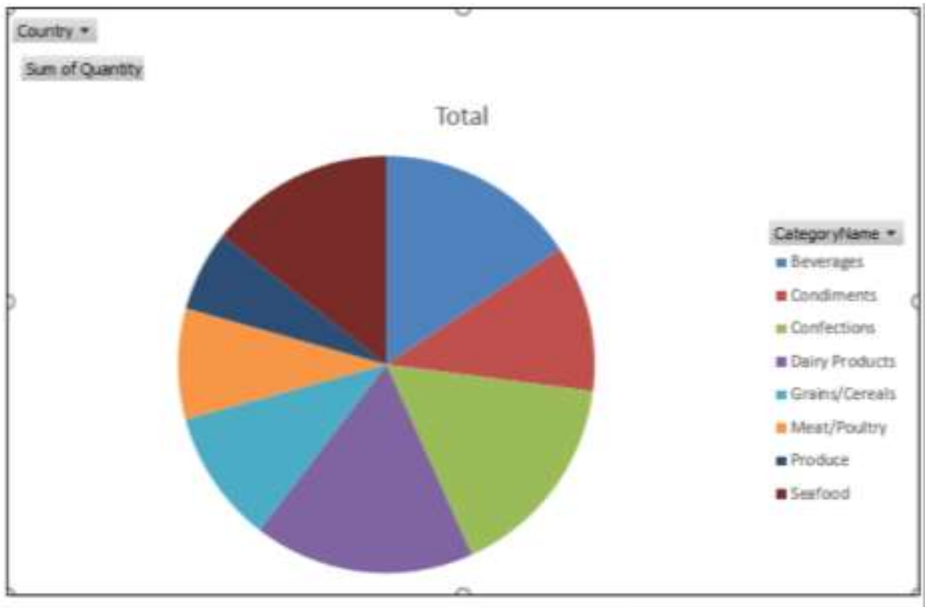


FIGURE 9-6: The PivotChart from Figure 9-5 now displayed as a pie chart.

Adding data labels to your PivotChart

- You can augment your chart with the actual values from a report (add labels) by:
  1. Select the chart.
  2. Choose Design ➪ Add Chart Element ➪ Data Labels.
  3. Select the data label position you want to use: Center, Inside End, Outside End, Best Fit (available only with certain chart types), or Data Callout.

- You can also display categories in a custom order (sort data series items) by:
  1. Select the PivotChart.
  2. Select the field button for the PivotChart’s Data Series field.
  3. Select the sort order you want to use.

Manually Sorting Data Series Or Categories

- Excel allows you to sort the data series items or categories manually.
- **To manually sort the data series items:** Select the PivotChart's underlying PivotTable , then select the label of the column field item you want to move. Move the mouse pointer to the right edge of the label cell and drag the label left/right to a new position.
- **To manually sort the categories**: Display the PivotChart's underlying PivotTable and select the label of the row field item you want to move. Move the mouse pointer to the bottom edge of the label cell and then drag the label up or down to the new position.
- When you return to the PivotChart a new sort order will have appeared
- Follow these steps to sort the category items:
    1. Select the PivotChart.
    2. Select the field button for the PivotChart's Category field.
    3. Select the sort order you want to use.

Adding PivotChart titles

- You can add three type of titles to a PivotChart to make it more comprehensible:
  - An overall chart title that is above the chart's plot area or overlaid on the plot area.
  - A category (X) axis title that sits below the category items.
  - A data series (Y) axis title that sits to the left of the data series axis labels.
- To add a chart title:
  1. Select the PivotChart.
  2. Choose Design ➪ Add Chart Element ➪ Chart Title.
  3. Select the type of chart title you want to add.
  4. Select the chart title.
  5. Enter the title you want to use.
  6. Click or tap outside the chart title to set it.

Adding PivotChart titles

Here are the steps to follow to add an axis title:

1. Select the PivotChart.
2. Choose Design ➪ Add Chart Element ➪ Axis Titles.
3. Select the type of chart title you want to add.

   You have two choices:

   » **Primary Horizontal:** Adds a category (X) axis title.

   » Primary Vertical: Adds a data series (Y) axis title.
4. Select the title.
5. Enter the title you want to use.
6. Click or tap outside the chart title to set it.

Adding PivotChart titles

- To format a chart title, select the title and then choose Format ➪ Format Selection. The pane that appears offers two tabs:
  - **Title Options:** with three subtabs Fill & Line, Effects, and Size & Properties
  - **Text Options**: with three subtabs: Text Fill & Outline, Text Effects, and Textbox.
- Excel gives you three methods for removing a title from a PivotChart:
  - Follow the steps in this section, and in the menu of title options, select None.
  - Right-click the title you want to remove and then choose the Delete command.
  - Select the title you want to remove and then press the Delete key.

Moving the PivotChart legend

- Excel enables you to move the legend to one of five positions with respect to the plot area. Follow these steps to set the position of the PivotChart legend:
  1. Select the PivotChart.
  2. Choose Design ➪ Add Chart Element ➪ Legend.
  3. Select More Legend Options.
  4. Click the Legend Options subtab and select the radio button for the legend position you want.

- In some cases, you might prefer to not display the legend at all. To do this, follow the preceding Steps 1 and 2 to display the Legend menu and then select None. Alternatively, right-click the legend and then choose the Delete command

Displaying a data table with the PivotChart

A Pivot-Chart data table is a table that displays the chart's categories as columns and its data series as rows, with the cells filled with the actual data values.

Display a Pivot-Chart data table by:

1. Select the PivotChart.
2. Choose Design ➪ Add Chart Element ➪ Data Table.
    - Select a predefined data table:
        - **With Legend Keys**: Displays the data table with the same colored squares that appear in the legend to identify each series.
        - **No Legend Keys**: Displays the data table without the colored squares.

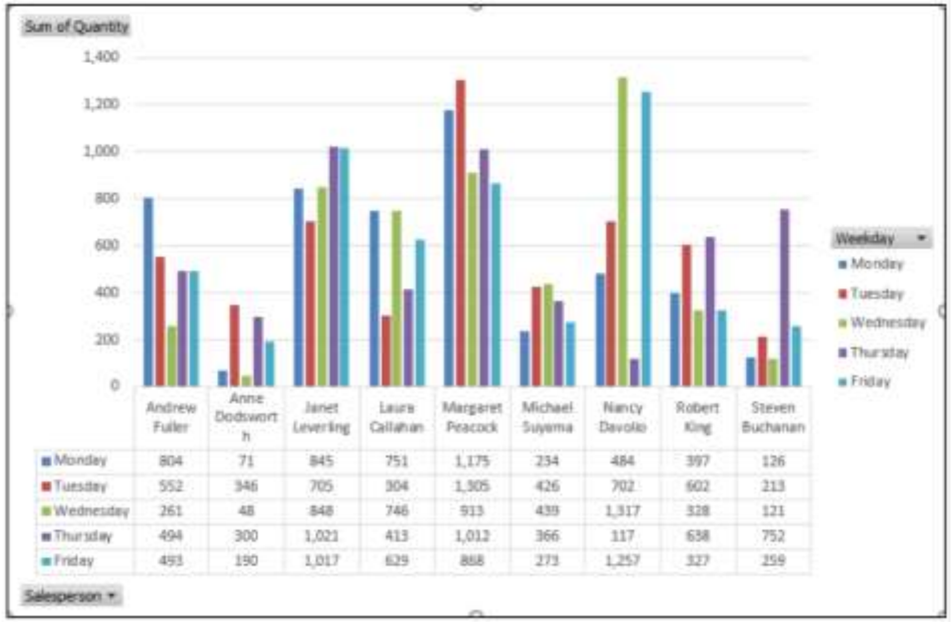


FIGURE 9-7: A data table shown below a PivotChart.

Main Reference

- **Chapter 3**: "Nature of Data, Statistical Modeling, and Visualization" from "Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".

- **Chapter 6, 7, 8 & 9 - (sections 3.7 to 3.11):** "Getting Started with Data Analysis" from "Excel Data Analysis For Dummies"

Week self-review exercises

- **Application Case 3.5 to Application Case 3.8** from ""Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".

Thank You

الجامعة السعودية الإلكترونية
SAUDI ELECTRONIC UNIVERSITY
2011-1432

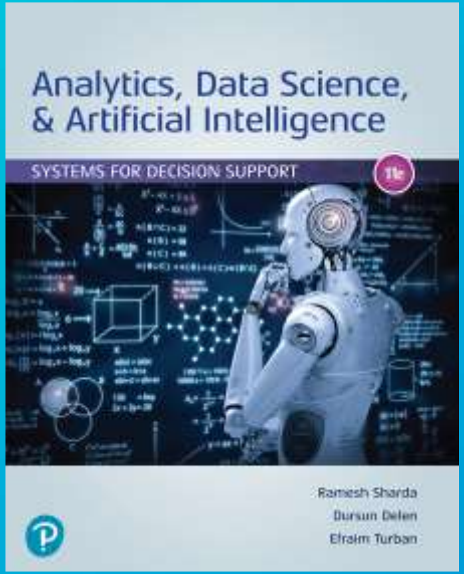
IT445
Decision Support Systems

College of Computing and Informatics

Week 13

Chapter 9:Big Data, Cloud Computing, and Location Analytics: Concepts and Tools

Chapter 12:Knowledge Systems: Expert Systems, Recommenders, Chatbots, Virtual Personal Assistants, and Robo Advisors



Analytics, Data Science, & Artificial Intelligence
Systems For Decision Support

This Presentation is mainly dependent on this textbook

Contents

Weekly Learning Outcomes

1. Learn what Big Data is and how it is changing the world of analytics and understand the motivation for and business drivers of Big Data analytics
2. Become familiar with the wide range of enabling technologies for Big Data analytics including Hadoop, MapReduce, and NoSQL.
3. Compare and contrast the complementary uses of data warehousing and Big Data technologies
4. Become familiar with in-memory analytics and Spark applications
5. Become familiar with selected Big Data platforms and services
6. Understand the need for and appreciate the capabilities of stream analytics and their applications.
7. Describe the current and future use of cloud computing in business analytics
8. Describe how geospatial and location-based analytics are assisting organizations
9. Describe recommendation systems, expert systems and chatbots
10. Understand the drivers and capabilities of chatbots and their use
11. Describe virtual personal assistants and their benefits and describe the use of chatbots as advisors
12. Discuss the major issues related to the implementation of chatbots

Required Reading

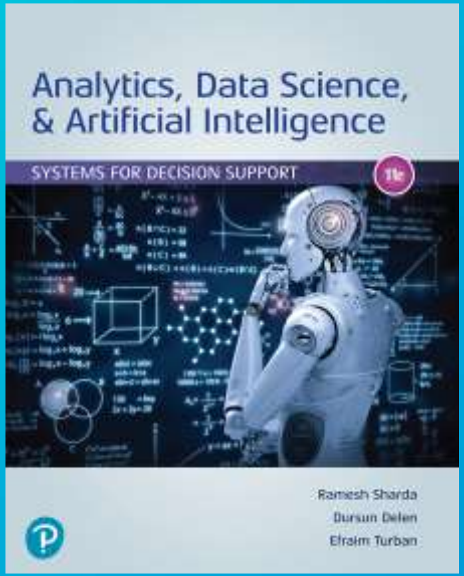
- **Chapter 9**: *"Big Data, Cloud Computing, and Location Analytics: Concepts and Tools"* from *"Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".*
- **Chapter 12**: "*Knowledge Systems: Expert Systems, Recommenders, Chatbots, Virtual Personal Assistants, and Robo Advisors* " from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".*

Recommended Reading

- Big Data: What it is and why it matters. SAS. https://www.sas.com/en_sa/insights/big-data/what-is-big-data.html

Recommended Video

- The future of cloud data analytics (2020, Oct 8). [Video]. YouTube https://www.youtube.com/watch?v=1mQ8VtxIAjk



Week 13 Part 1

Chapter 9:Big Data, Cloud Computing, and Location Analytics: Concepts and Tools

9.2 Definition of Big Data

Definition and concepts

- The term Big Data means different things to people with different backgrounds and interests.
- Big Data, traditionally, has been used to describe the massive volumes of data
  - E.g., the one analyzed by huge organizations like Google, NASA..
- The relative term: “Big” depends on an organization’s size for most businesses.
- **Where does Big Data come from?**
  - A simple answer is “everywhere.” It may come from Web logs, radio-frequency identification (RFID), global positioning systems (GPS), sensor networks, social networks, …
  - The prime use of such data is generating insights through analytics, sometimes the term Big Data is expanded as Big Data analytics. *More on that later.*

Definition and concepts

Popular description of Big Data:

- Big Data exceeds the reach of commonly used hardware environments and/or capabilities of software tools to capture, manage, and process it within a tolerable time span.
- Describe the exponential growth, availability, and use of information, both structured and unstructured.

- "Big Data" is a misnomer → Big Data is not just "big."

The "V"s That Define Big Data

Volume - most common trait of Big Data.

- Come from exponential increase in transaction-based data, text data constantly streaming in from social media, increasing amounts of sensor data being collected…

Variety - data come in all types of format.

- E.g., text documents, e-mail, XML, video, audio, and stock ticker data...

Velocity - how fast data is being produced and processed.

Veracity - conformity to facts: accuracy, quality, truthfulness, or trustworthiness of the data.

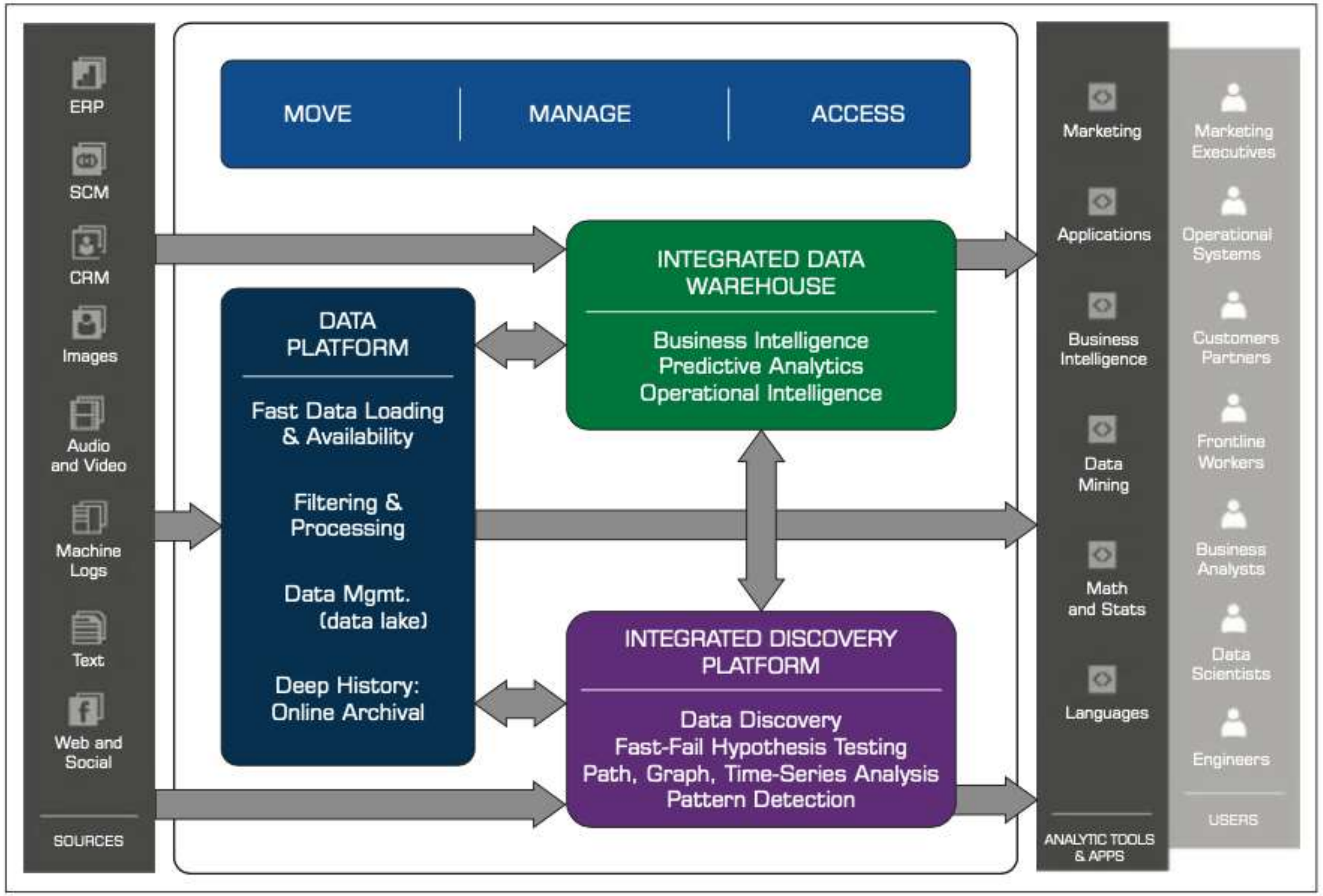
Variability - data flows can be highly inconsistent with periodic peaks.

- Daily, seasonal, and event-triggered…

Value Proposition - contains more patterns and interesting anomalies than "small" data.

A High-Level Conceptual Architecture for Big Data Solutions

- Big Data is converted to business insight
- Through the use of a combination of advanced analytics and delivered to a variety of different users/roles for faster/better decision making.



9.3 Fundamentals of Big Data Analytics

- Big Data considerations
- Critical Success Factors for Big Data Analytics
- Big Data Analytics Enablers
- Challenges of Big Data Analytics
- Business Problems Addressed by Big Data Analytics

Fundamentals of Big Data Analytics

- Big Data by itself, regardless of the size, type, or speed, is worthless.

- To delivers value big data is combined with “big” analytics.

With the value proposition, Big Data also brought about big challenges for organizations:

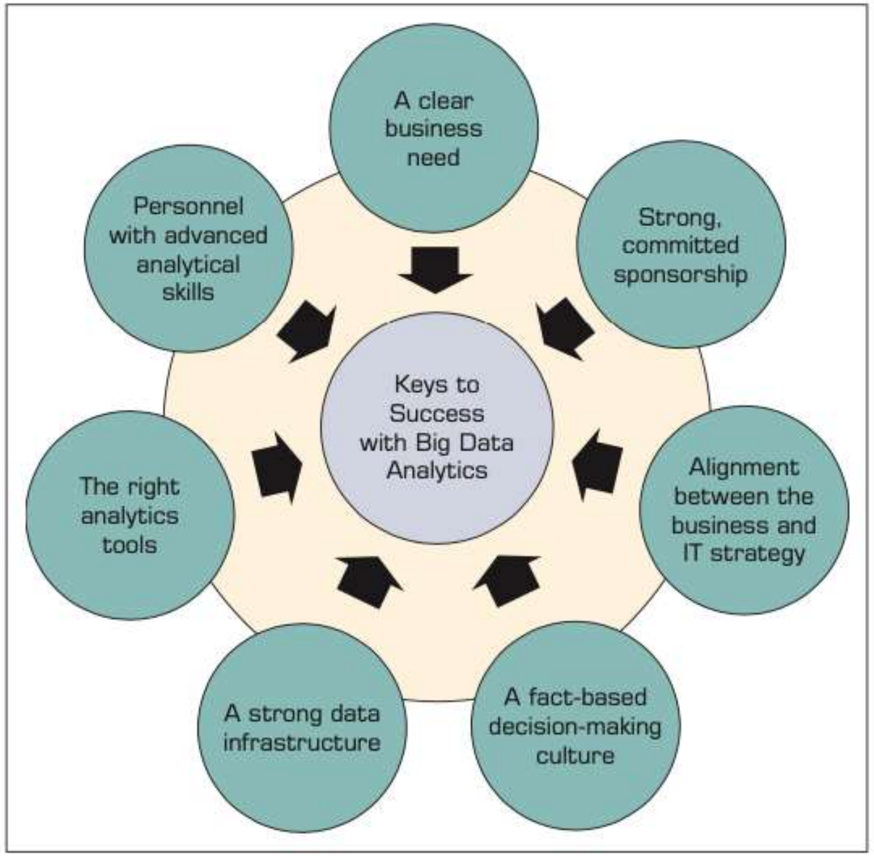
  - The traditional means for capturing, storing, and analyzing data are not capable of dealing with Big Data effectively and efficiently.
  - New breeds of technologies need to be developed (or purchased/hired/outsourced).

Big Data considerations

- You can’t process the amount of data that you want to because of the limitations posed by your current platform or environment.
- You want to involve new/contemporary data sources (e.g., social media, RFID, sensory, Web, GPS, textual data) into your analytics platform, but you can’t because it does not comply with the data storage schema-defined rows and columns without sacrificing fidelity or the richness of the new data.
- You need to (or want to) integrate data as quickly as possible to be current on your analysis.
- You want to work with a schema-on-demand data storage paradigm because the nature of the new data may not be known, or there may not be enough time to determine it and develop a schema for it.
- The data is arriving so fast at your organization’s doorstep that your traditional analytics platform cannot handle it.

Critical Success Factors for Big Data Analytics

- A clear business need (alignment with the vision and the strategy).
- Strong, committed sponsorship (executive champion).
- Alignment between the business and IT strategy.
- A fact-based decision-making culture.
- A strong data infrastructure.
- The right analytics tools and personnel with advance analytical skills.



Big Data Analytics Enablers

A number of innovative computational techniques and platforms have been developed to enable Big data Data Analytics (a.k.a High-performance computing):

- **In-memory analytics:** allow Big Data to be processed in-memory and distributed across a dedicated set of nodes.
- **In-database analytics:** enable data integration and analytic functions inside the database.
- **Grid computing:** Promotes efficiency, lower cost, and better performance by processing jobs in a shared, centrally managed pool of IT resources.
- **Appliances:** Brings together hardware and software in a physical unit that is not only fast but also scalable on an as needed basis.

Challenges of Big Data Analytics

Data volume: The ability to capture, store, and process a huge volume of data at an acceptable speed.
Data integration: The ability to combine data from different source quickly at reasonable cost.
Processing capabilities: The ability to process data quickly, as it is captured.
Data governance: The ability to keep up with the security, privacy, ownership, and quality issues
Skills availability: People with skills to do the job (often called data scientists).
Solution cost: Experimentation and discovery is taking place to determine the patterns that matter and the insights that turn to value.

Business Problems Addressed by Big Data Analytics

- Process efficiency and cost reduction
- Brand management
- Revenue maximization, cross-selling, and up-selling Enhanced customer experience
- Churn identification, customer recruiting
- Improved customer service
- Identifying new products and market opportunities Risk management
- Regulatory compliance
- Enhanced security capabilities

9.4 BIG DATA TECHNOLOGIES

- MapReduce
- Hadoop
- NoSQL

BIG DATA TECHNOLOGIES

- There are a number of technologies for processing and analyzing Big Data
  - MapReduce, Hadoop, Hive, Pig, NoSQL, Hbase, Flume, Oozie, Ambari, Avro, Mahout, Sqoop, Hcatalog…

The three Big Data technologies that will transform the business analytics and data management markets are:

- MapReduce
- Hadoop
- NoSQL

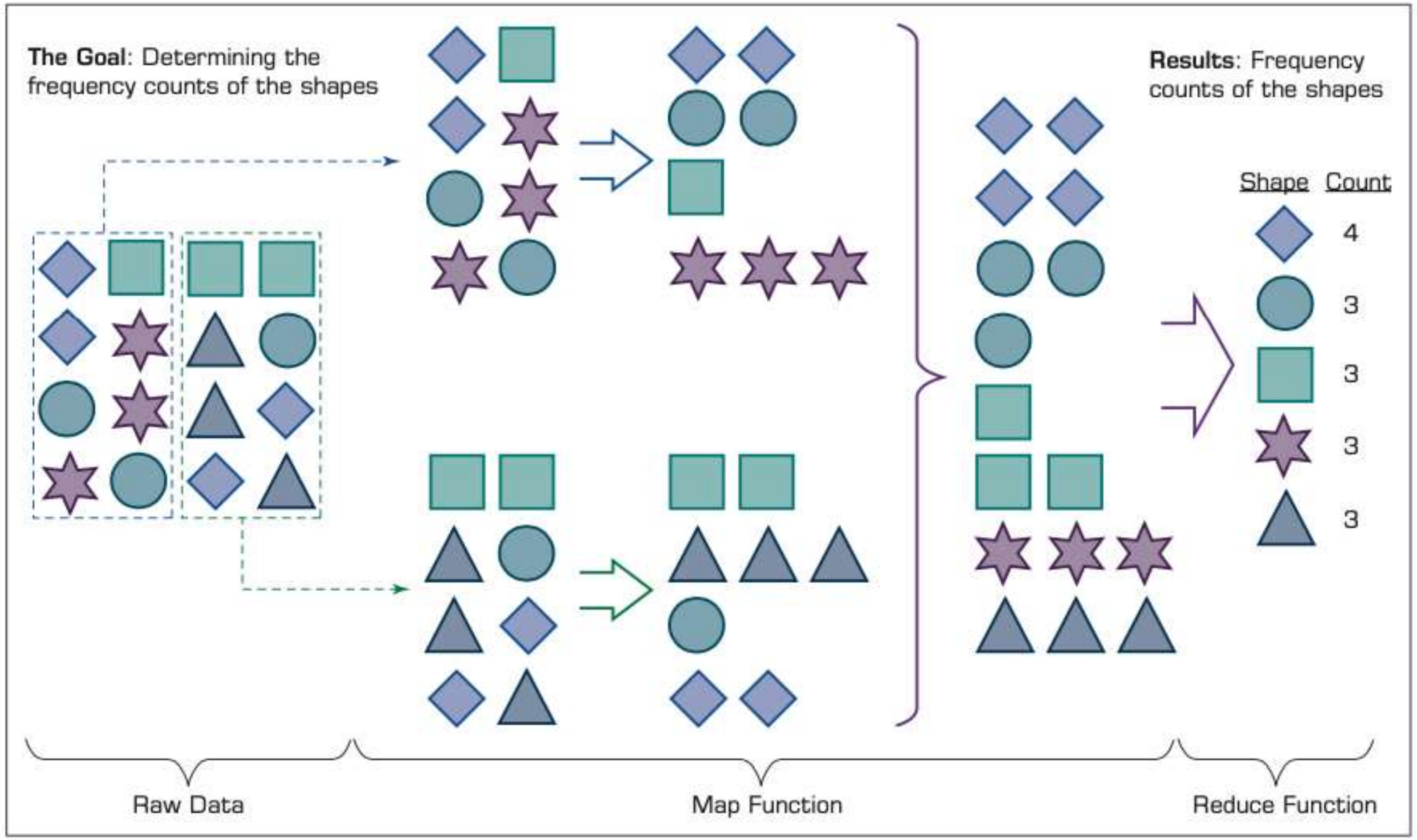
MapReduce

MapReduce distributes the processing of very large multistructured data files across a large cluster of machines.

- High performance is achieved by breaking the processing into small units of work that can be run in parallel across thousands of nodes in the cluster.
- Technique popularized by Google.
- MapReduce is a programming model not a programming language.
- Aids organizations in processing and analyzing large volumes of multistruc-tured data.
  - E.g., indexing and search, graph analysis, text analysis, machine learning, data transformation…

MapReduce

- The MapReduce system first reads the input file and splits it into multiple pieces.
- There are two splits in this example, but in a real-life scenario, the number of splits would typically be much higher.



Hadoop

Hadoop is an open source framework for processing, storing, and analyzing massive amounts of distributed, unstructured data.

- Originally created by Doug Cutting at Yahoo!
- Designed to handle petabytes and exabytes of data distributed over multiple nodes in parallel.
- Hadoop clusters run on inexpensive commodity hardware so projects can scale-out cost effectively.
- Hadoop breaks up Big Data into multiple parts so each part can be processed and analyzed at the same time.
- Open source under Apache Software Foundation
  - hundreds of contributors continuously improve the core technology.

How Does Hadoop Work?

- Accesses unstructured and semistructured data from sources (E.g., log files, social media feeds, and internal data stores).
- It breaks the data up into “parts,” which are then loaded into a file system (HDFS) made up of multiple nodes running on commodity hardware.
- Each “part” is replicated multiple times and loaded into the file system so that if a node fails, another node has a copy of the data contained on the failed node.
- A node acts as the Facilitator and another as Job Tracker.
- Facilitator node communicate back to the client information.
  - E.g., which nodes are available, where in the cluster certain data resides, and which nodes have failed.
- Once the data is loaded into the cluster, it is ready to be analyzed via the MapReduce framework.
  - A “Map” job is submitted to Job Tracker node to determine which data it needs to access to complete the job and where in the cluster that data is located.
  - Once the jobs are completed the results are collected and aggregated using MapReduce.

Cont...

- Once the MapReduce phase is complete, the processed data is ready for further analysis by data scientists and others with advanced data analytics skills.

- The data can also be modeled and transferred from Hadoop clusters into existing relational databases, data warehouses, and other traditional IT systems for further analysis to support transactional processing.

Data scientists

- Can manipulate and analyze the data using any of a number of tools for any number of uses, including searching for hidden insights and patterns, or use as the foundation for building user-facing analytic applications.

Hadoop Technical Components

- **Hadoop Distributed File System (HDFS):** The default storage layer in any given Hadoop cluster.

- **Name Node:** provides the client information on where in the cluster particular data is stored and if any nodes fail.

- **Secondary Node:** A backup to the Name Node, it periodically replicates and stores data from the Name Node should it fail.

- **Job Tracker:** initiates and coordinates MapReduce jobs or the processing of the data.

- **Slave Nodes:** store data and take direction to process it from the Job Tracker.

Hadoop: The Pros and Cons

- Hadoop consists of multiple products.
- Hadoop is open source but available from vendors, too.
- Hadoop is an ecosystem, not a single product.
- Hadoop is primarily a distributed file system and lacks capabilities we would associate with a DBMS.
- Hive resembles SQL but is not standard SQL.
- Hadoop and MapReduce are related but not require each other.
- MapReduce provides control for analytics, not analytics.
- Hadoop is about data diversity, not just data volume.
- Hadoop complements a DW; it's rarely a replacement.
- Hadoop enables many types of analytics, not just Web analytics.

NoSQL

- NoSQL (stands for Not Only SQL) is a new style of database.

- Goal - serving up discrete data stored among large volumes of multistructured data to end-user and automated Big Data applications.

- Hbase: is a popular NoSQL database
  - modeled after Google BigTable that is often deployed on top of HDFS
  - provide low-latency and quick lookups in Hadoop.

- NoSQL databases downside is they trade ACID (atomicity, consistency, isolation, durability) compliance for performance and scalability.

9.5 Big Data and Data Warehousing

Hadoop versus Data Warehousing (DW)

- Two paradigms that are seemingly competing for the same job which one will prevail?
  - both can run in parallel, scale-up to enormous data volumes.
  - But they are not interchangeable!

Use Cases for Hadoop

- Hadoop as the repository and refinery
- Hadoop as the active archive

Use Cases for Data Warehousing

- Data warehouse performance
- Integrating data that provides business value
- Interactive BI tools

Hadoop versus DW

When to Use Which Platform

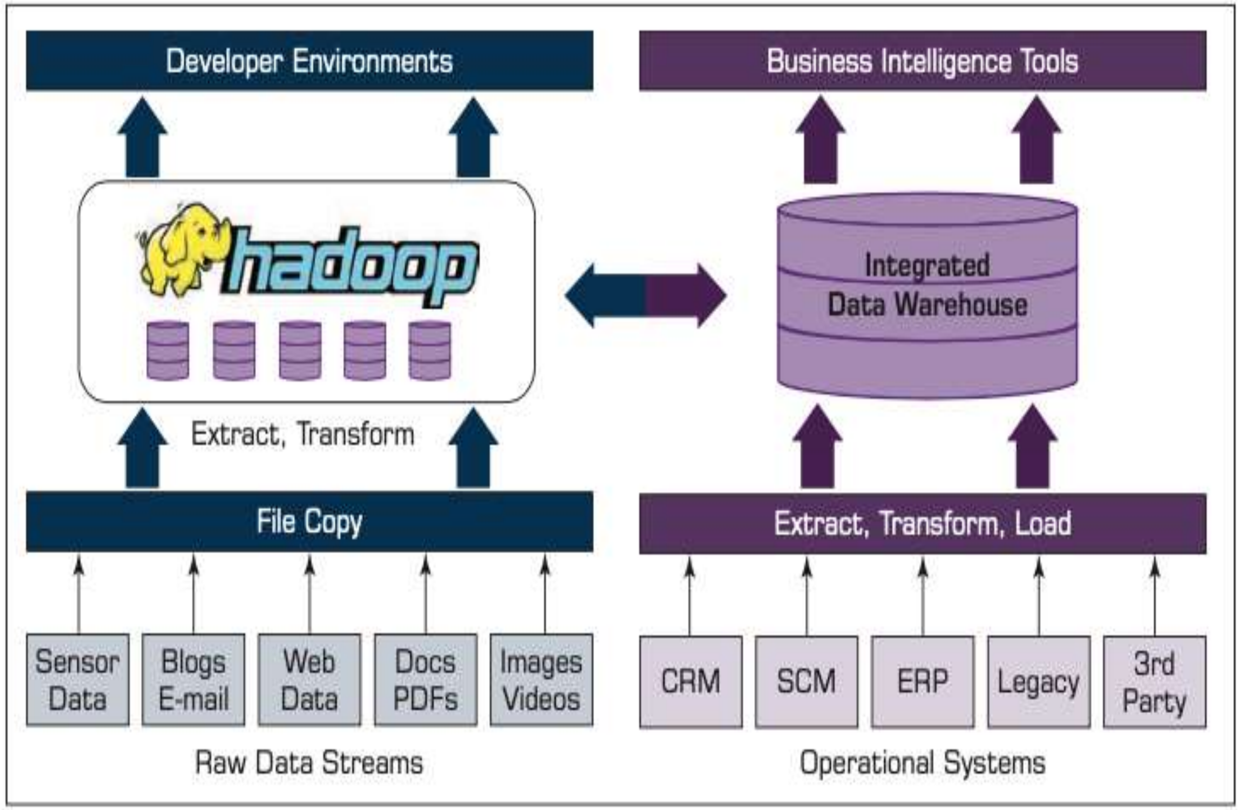
- Depends on the requirements and the preferences of the organization.
- Often Hadoop and the data warehouse work together in an information supply chain

TABLE 9.1 When to Use Which Platform—Hadoop versus DW

Requirement	Data Warehouse	Hadoop
Low latency, interactive reports, and OLAP	☑	
ANSI 2003 SQL compliance is required	☑	☑
Preprocessing or exploration of raw unstructured data		☑
Online archives alternative to tape		☑
High-quality cleansed and consistent data	☑	☑
100s to 1,000s of concurrent users	☑	☑
Discover unknown relationships in the data		☑
Parallel complex process logic	☑	☑
CPU intense analysis	☑	
System, users, and data governance		☑
Many flexible programming languages running in parallel		☑
Unrestricted, ungoverned sandbox explorations		☑
Analysis of provisional data	☑	
Extensive security and regulatory compliance	☑	☑

Coexistence of Hadoop and Data Warehouse

- Use Hadoop for storing and archiving multistructured data.
- Use Hadoop for filtering, transforming, and/or consolidating multistructured data.
- Use Hadoop to analyze large volumes of multistructured data and publish the analytical results.
- Use a relational DBMS that provides MapReduce capabilities as an investigative computing platform.
- Use a front-end query tool to access and analyze data.



9.6 Big Data and In-memory Analytics and Apache SparkTM

- In-memory analytics
- Apache Spark
- Component of Apache Spark

In-memory analytics

- Emerging processing technique to analyse data stored in in-memory databases.

- In-memory processing is more efficient than the batch processing.
  - Which allows for the analytics of streaming data in real-time.

- Real-time applications
  - build real-time dashboards for better insights and faster decision making. understanding customer behaviour and engagement
  - forecasting stock price
  - optimizing airfare
  - predicting fraud…
- Apache Spark is most popular tool supporting the in-memory processing.

Apache Spark

- A unified analytics engine that can execute both batch and streaming data.

- Originally developed at University of California, Berkeley in 2009.

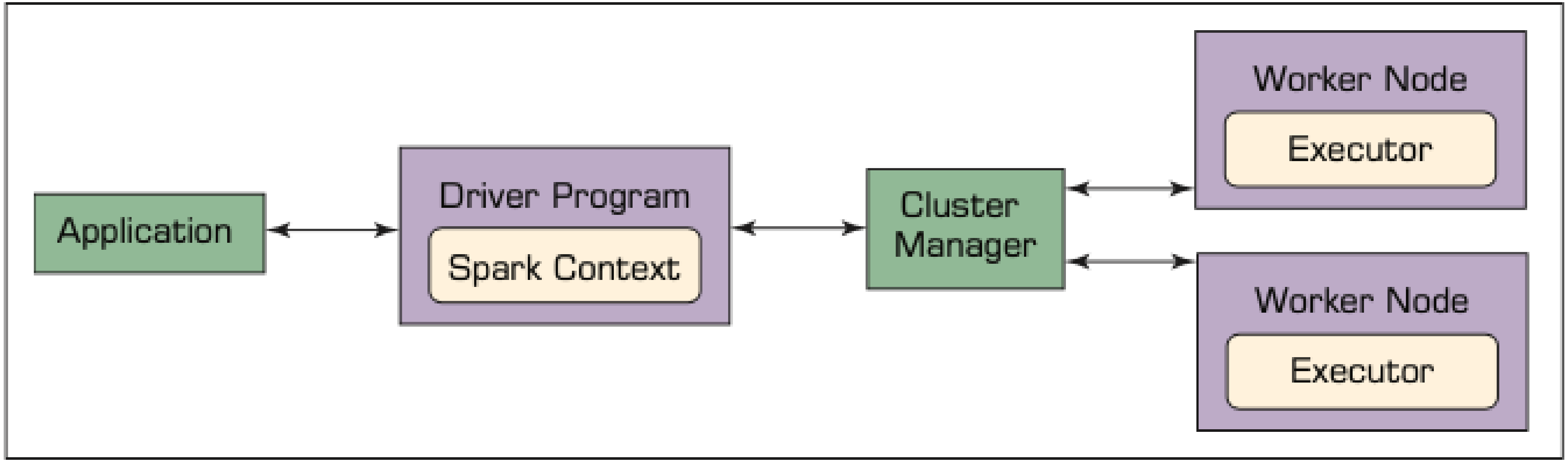
- In-memory computation to achieve high performance on large-scale data processing.

- Apache SparkTM runs faster than the traditional Apache Hadoop.
  - It can can run on Apache Hadoop, Apache Mesos, Kubernetes, standalone, or in the cloud.

- Several big market players in the analytics sector have adopted Apache Spark for real-time analytics.
  - E.g., Uber, Pinterest, Netflix, Yahoo, and eBay….

Apache Spark

Architecture of Apache Spark



Component of Apache Spark

Resilient Distributed Dataset (RDD)

- handles lineage, memory management, fault tolerance, and data partitioning across all nodes in a cluster.
- provides several transformation functions like map, filter, and join.

- All transformations in Spark are lazy in nature.
  - Spark does not execute operations until any action function is performed on data.
  - The action functions (e.g., count, reduce) print or return value after an execution.
  - **So called Lazy Evaluation**

- In Spark Streaming, Dstream, are utilized to process streaming data.

9.7 Big Data and Stream Analytics

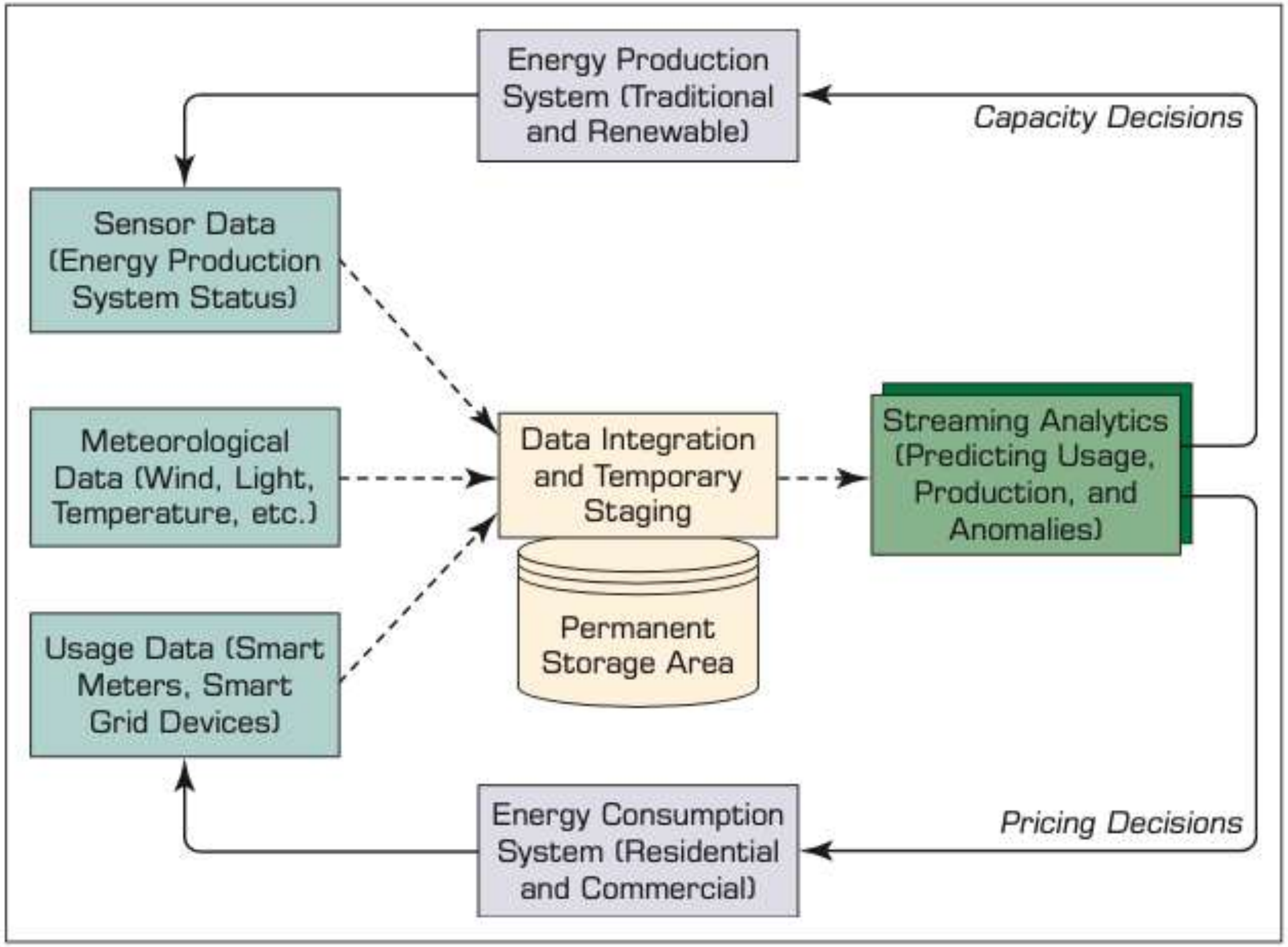
Stream Analytics

- One of the Vs in Big Data is velocity → speed at which the data is created and streamed into the analytics environment.

- Stream analytics process of extracting actionable information from continuously flowing/streaming data.
  - a.k.a data-in-motion analytics and real-time data analytics.

- Stream analytics is more popular because
  - time-to-action has become an ever-decreasing value.
  - technological means are exist to capture and process the data while it is created.

Stream Analytics



A Use Case of Streaming Analytics in the Energy Industry

Stream Analytics

Stream analytics versus perpetual analytics

- Streaming analytics involves applying transaction-level logic to real-time observations.
- Perpetual analytics evaluates every incoming observation against all prior observations to enables the discovery of real-time insight.

- **Critical event processing** is a method of capturing, tracking, and analyzing streams of data to detect events (out of normal happenings) of certain types that are worthy of the effort.
  - goal is to take rapid actions to prevent /mitigate these events (e.g., fraud or network intrusion) from occurring.

- **Data stream mining**, is the process of extracting novel patterns and knowledge structures from continuous, rapid data records.

Applications of Stream Analytics

- E-Commerce
- Telecommunication
- Law Enforcement and Cyber Security
- Power Industry
- Financial Services
- Health Services
- Government

9.8 Big Data Vendors and Platforms

Big Data vendor

- The Big Data vendor landscape is developing very rapidly.

Broad categories of Big Data providers:

1. Infrastructure Services Providers
   - E.g., Amazon Web Services, Microsoft Azure, Google Cloud, IBM Cloud Cloudera,…
2. Analytics Solution Providers
   - E.g., Dell EMC, IBM Big Insights (now part of Watson), Microsoft Analytics, SAP's Hanna, Oracle Big Data, and Teradata,…
3. Legacy BI Providers Moving to Big Data
   - E.g., SAS, Microstrategy,…

9.9 Cloud Computing and Business Analytics

Cloud Computing

Cloud Computing - a style of computing in which dynamically scalable and often virtualized resources are provided over the Internet.

- Users need not have knowledge of, experience in, or control over the technology infrastructures in the cloud that supports them.
- Cloud computing is a new name for: utility computing, application service provider grid computing, on-demand computing, software as a service (SaaS), and centralized computing with dumb terminals…
  - reference to the Internet as a "cloud" and represents an evolution of all of the previously shared/ centralized computing trends.

Cloud Computing

Examples

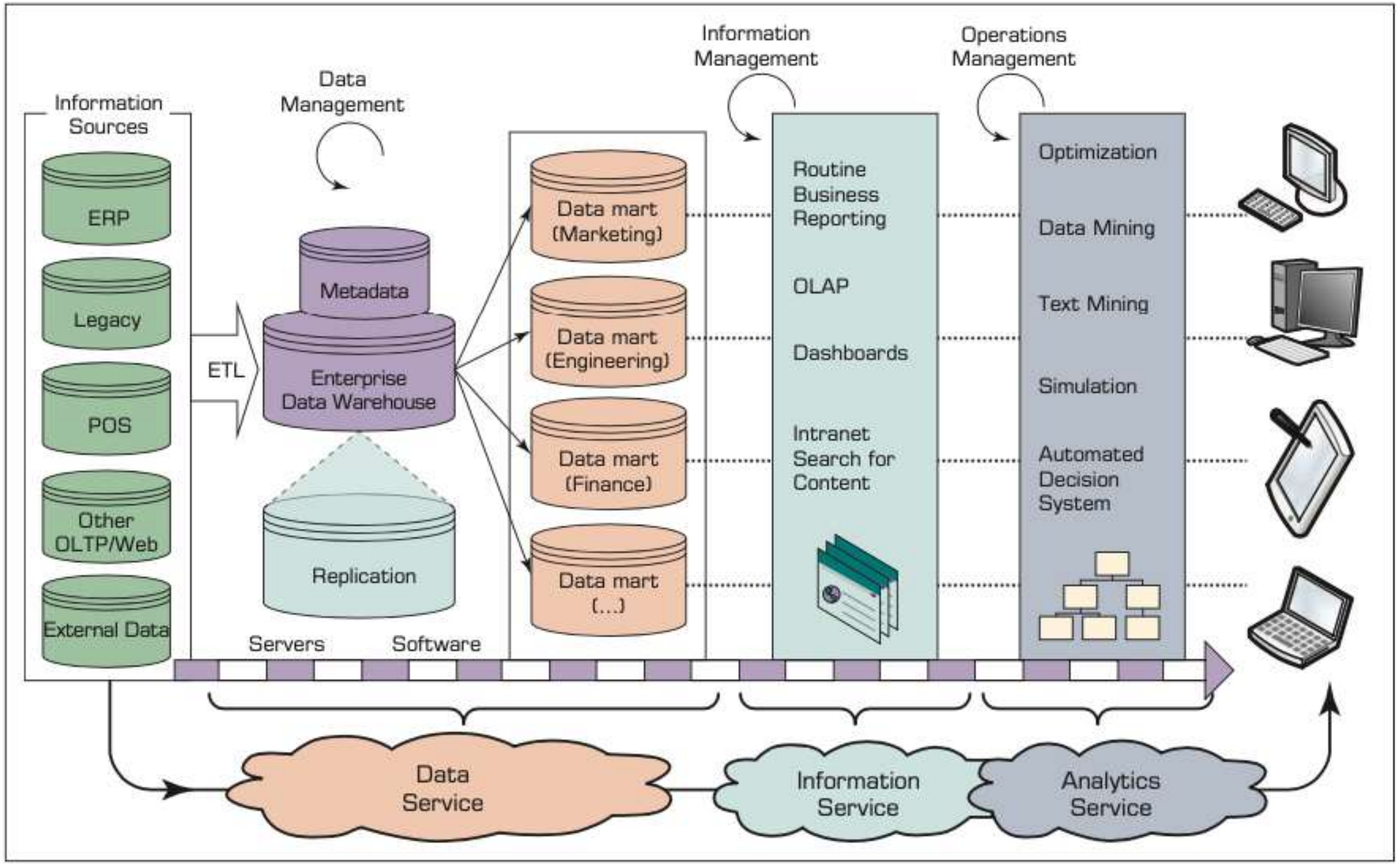
- Web-based e-mail - a basic cloud application (e.g., Gmail)
  - stores the data (e-mail messages)
  - the software (e-mail programs that process and manage e-mails).
  - e-mail provider supplies the hardware/software and all of the basic infrastructure.
  - access the e-mail application from anywhere in the cloud via web browser.

- General cloud application – E.g., Google Docs and Spreadsheets
  - This application allows a user to create text documents or spreadsheets that are stored on Google's servers and are available to the users anywhere they have access to the Internet.

- Social networking Web sites - E.g., Facebook, Twitter, and LinkedIn...

- Business cloud application: Amazon Web services (AWS), Salesforce.com, IBM Cloud...

Cloud Computing



Conceptual Architecture of a Cloud-Oriented Support System

Cloud Computing and Service-Oriented

- Cloud computing is a combination of several IT components as services.

Data as a Service (DaaS)

- Access data wherever it resides.
- Ensure data quality with centralization.
- Access the data via open standards (e.g., SQL, XQuery, and XML).
- An extension of DaaS is information as a service (IaaS)

Software as a Service (SaaS)

- Consumers can use software that run on distant computers in the cloud infrastructure.
- Need not worry about managing underlying cloud infrastructure.
- Pay for the use of software only.
- Access from anywhere using a Web browser!

Cloud Computing and Service-Oriented

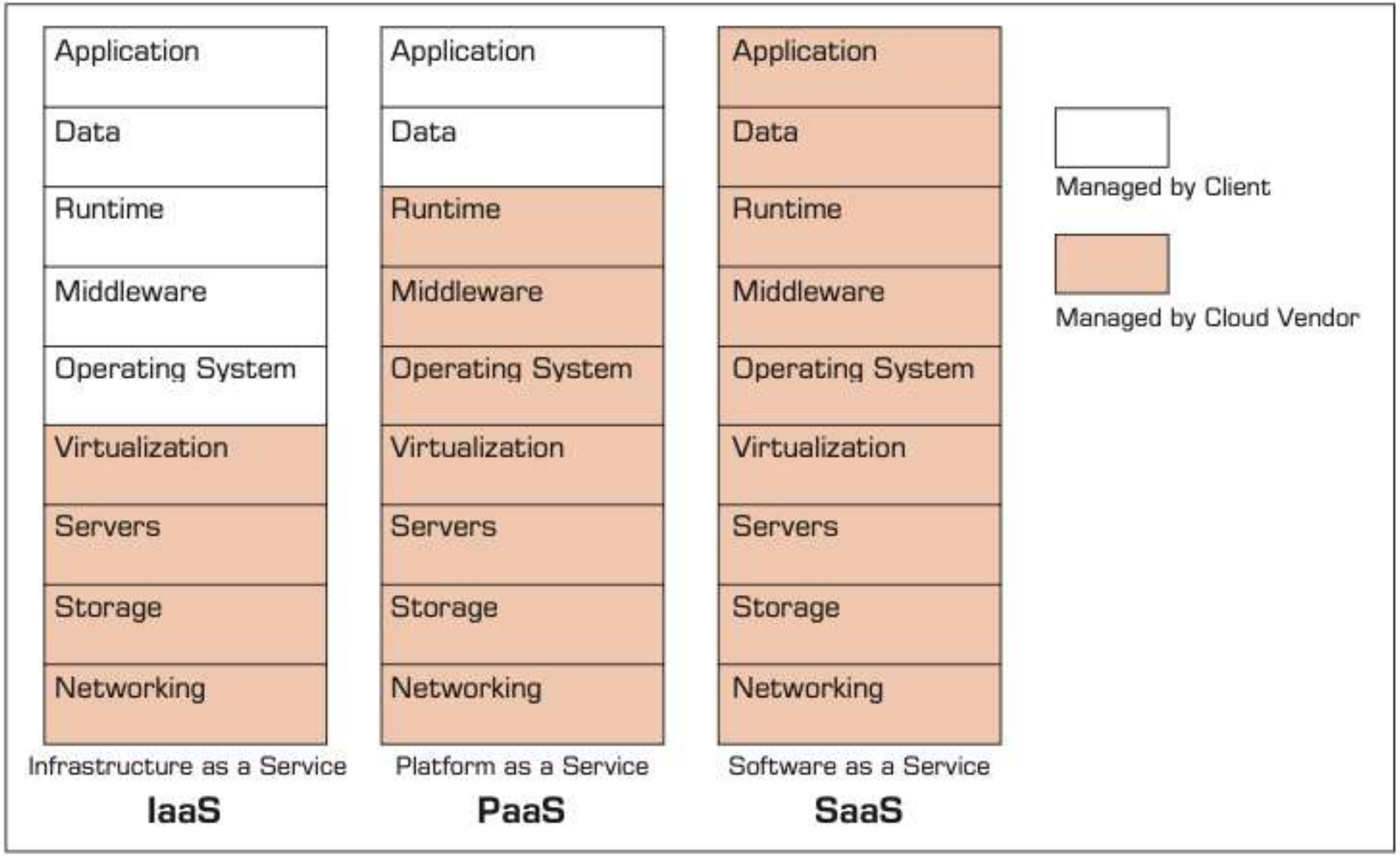
Platform as a Service (PaaS)

- Companies deploy their applications in the cloud.
- No need to manage resources (e.g., networks, servers, storage, or operating systems).
- Saves time for setting up the infrastructure.
- Reduces the cost of maintaining underlying infrastructure for running software.
- Companies to focus on their business rather than focusing on managing infrastructure.
- E.g., Microsoft Azure, Amazon EC2, and Google App Engine.

Infrastructure as a Service (IaaS)

- Infrastructure resources like networks, storage, servers, and other computing resources are provided to clients.
- Clients can have administrative rights to use these resources but do not manage underlying infrastructure.
- They have to pay for usage of infrastructure.
- E.g., Amazon Web services.

Cloud Computing and Service-Oriented



The level of service subscriptions a client uses in each of the three major types of cloud offerings

Cloud Computing

Cloud Deployment Models

Public cloud

- Subscriber uses the resources offered by service providers over the Internet.
- Advantage: saving time and money in setting up hardware/software required to run the business.
- E.g., Microsoft Azure, Google Cloud Platform, and Amazon AWS.

Private cloud:

- Allow direct control over the data and applications.
- Disadvantage: is the cost of maintaining and managing the cloud, as on-premise IT staff are responsible for managing it.

Hybrid cloud:

- Gives businesses great flexibility by moving workloads between private and public clouds.

Cloud Computing

Essential Technologies for Cloud Computing

- **Virtualization -** creation of a virtual version of something like an operating system or server.
  - Network virtualization
  - Storage virtualization
  - Server virtualization

Major cloud platform providers in analytics

  - Amazon Elastic Beanstalk
  - IBM Cloud
  - Microsoft Azure:
  - Google App Engine
  - Openshift

Cloud Computing

Analytics as a Service (AaaS)

- Provides an agile model for reporting and analytics to businesses.
- AaaS in the cloud has economies of scale, better scalability and higher cost savings.
- Combines cloud computing with Big Data analytics and empowers data scientists to access centrally managed information data sets.
- Reduce costs and compliance risk, while increasing productivity of users.
- Data and text mining is another very promising application of AaaS.

- Representative Analytics as a Service Offerings: IBM cloud, minemytext.com, SAS viya, Tableau, Snowflake…

Cloud Computing

Illustrative analytics applications employing the cloud infrastructure

- Using Azure IOT, Stream Analytics, and Machine Learning to Improve Mobile Health Care Services
- Gulf Air Uses Big Data to Get Deeper Customer Insight
- Chime Enhances Customer Experience Using Snowflake

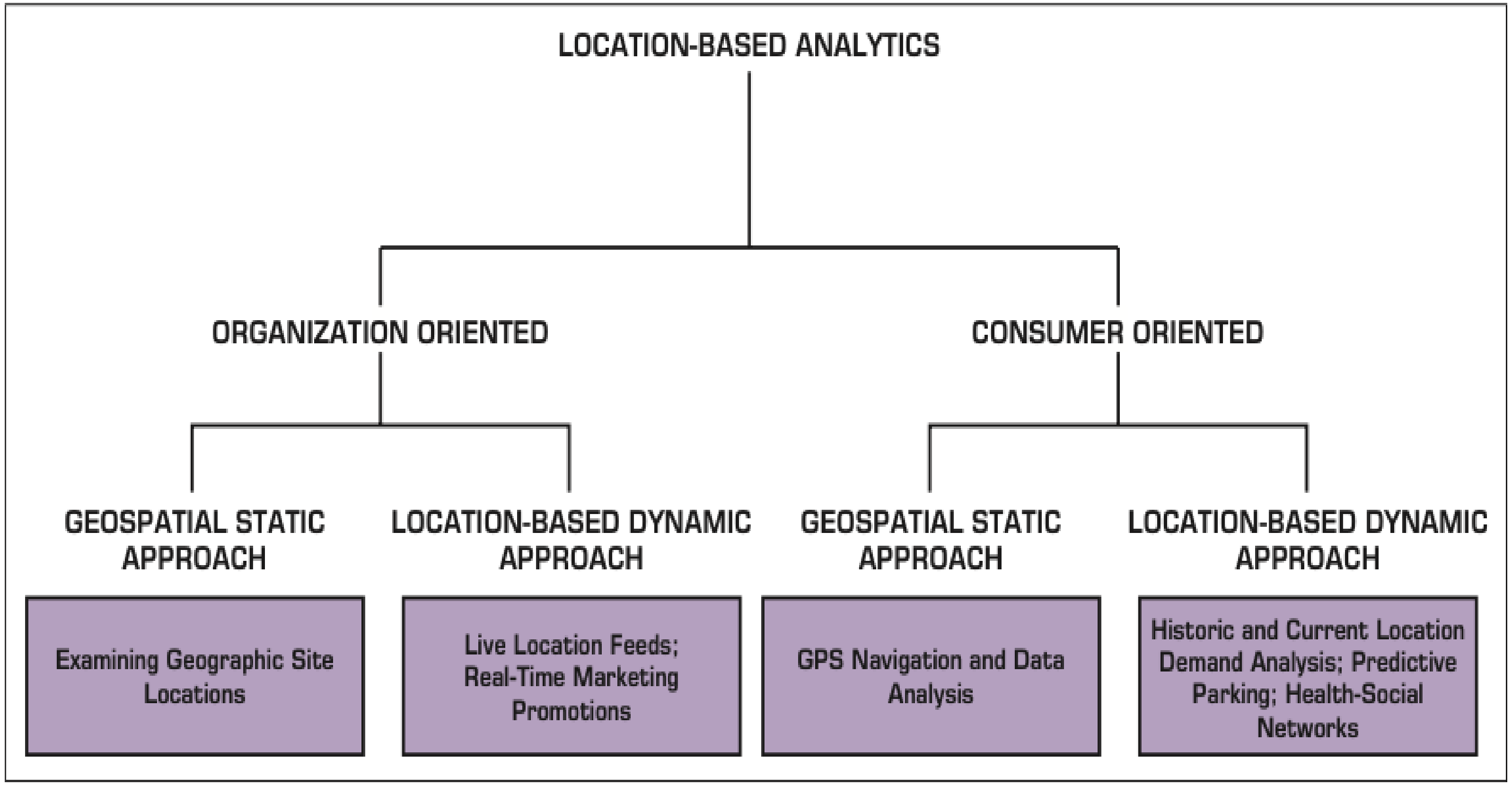
9.10 Location-Based Analytics for Organizations

- Geospatial Analytics
- Uses of Geospatial Analytics
- Real-Time Location Intelligence
- Analytics Applications for Consumers

Geospatial Analytics

- Incorporation of location data can enhance analytics applications.

Classification of location-based analytic applications



Geospatial Analytics

Location-based analytic techniques use geocoding

- Visual maps
- Postal codes
- Latitude & Longitude

- Map-based visualizations have enabled organizations to view the aggregated data and get more meaningful location-based insights.
- Locations based on postal codes offer an aggregate view of a large geographic area.
- Location components based on latitudinal and longitudinal enables organizations to add a new dimension of “where” to traditional business analyses.

Geospatial Analytics

Geographic Information Systems (GIS)

- GIS used to capture, store, analyze, and manage data linked to a location using integrated sensor technologies, global positioning systems installed in smartphones, or through RFID deployments in the retail and healthcare industries.

Location intelligence

- Interactive maps that further drill down to details about any location.
- Enable analysts to pinpoint trends and patterns in revenue, sales, and profitability across geographical areas.
- Enable organizations to gain critical insights and make better decisions by optimizing important processes and applications.

- ESRI company is is the market leader in providing GIS data.

Uses of Geospatial Analytics

Retailers

- Can determine how sales vary by population level and proximity to other competitors.
- They can assess the demand and efficiency of supply chain operations.
- Consumer product companies can identify the specific needs of customers and customer complaint locations and easily trace them back to the products.
- Sales reps can better target their prospects by analyzing their geography.

Agricultural applications

- Very precise irrigation and fertilizer applications can be planned, by combining location, weather, soil, and crop-related data.

Crime analysis

- Superimposition of crime data including date, time, and type of crime onto the GIS data can provide significant insights into crime patterns and police staffing.

Disease spread prediction

- Track and then predict outbreaks of diseases, such as the flu, using GIS data.

Uses of Geospatial Analytics

Telecommunication companies

- Identify the geographic areas experiencing a large number of failed connection attempts of voice, data, text, or Internet.
- Analytics can help determine the exact causes based on location and drill down to an individual customer to provide better customer service.

Benefit of combining geographic information with other data being generated by an organization:

- better identify the customer churn.
- help in formulating strategies specific to locations for increasing operational efficiency, quality of service, and revenue.

Real-Time Location Intelligence

- Many devices in use by consumers and professionals are constantly sending out their location information
  - E.g., Cars, buses, taxis, mobile phones, cameras, navigation systems...
  - Resulted in a massive database of historical and real-time streaming location information.
- Reality mining
  - Real-time location information of users can be mined to develop promotion campaigns that are targeted at a specific user in real time.
- Location information from mobile phones can be used to create profiles of user behavior and movement.
  - Such location information can enable users to find other people with similar interests and advertisers to customize their promotions.

Real-Time Location Intelligence

- Augmented reality

Pokémon GO - a location-sensing augmented reality-based game

- It encourages users to claim virtual items from select geographic locations.
- The user can start anywhere in a city and follow markers on the app to reach a specific item.
- Virtual items are visible through the app when the user points a phone’s camera toward the virtual item.
- The user can then claim this item.

Candybar - Business applications

- Allows businesses to place these virtual items on a map using Google Maps.
- The placement of this item can be fine-tuned using Google’s Street View.
- Once all virtual items have been configured with the information and location, the business can submit items, which are then visible to the user in real time.
- Provides usage analytics to the business to enable better targeting of virtual items.

Analytics Applications for Consumers

The explosive growth of the apps industry for smartphone platforms

- iOS, Android, Windows, …
- meant for direct use by a consumer
- Aim to mine a consumer’s usage/purchase data to create a profile for marketing specific products or services.
- these apps are meant for enabling consumers to make better decisions by employing specific analytics.

Examples

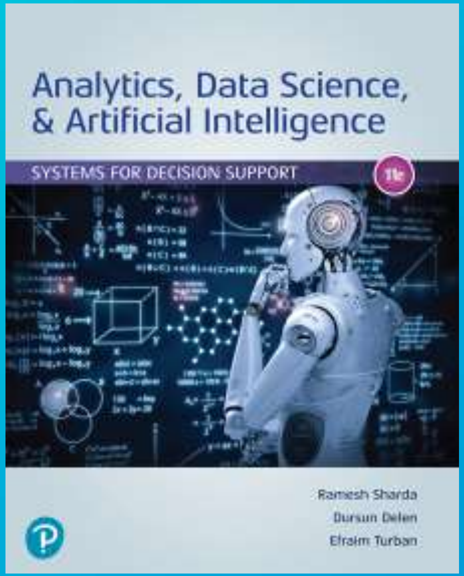
- Waze: a social Web app that assists users in identifying a navigation path and alerts users about potential issues such as accidents, police checkpoints, speed traps, …
- ParkPGH: directs drivers to parking lots in areas where parking is available.
  - It includes predictive capabilities to estimate parking availability.

Analytics Applications for Consumers

- Analytics-based applications are emerging not just for fun and health
- Productivity
  - E.g., Google's e-mail app analyzes billions of e-mail transactions and develops automated responses for e-mails.

- The growth of consumer-oriented analytic applications will continue and create many entrepreneurial opportunities.

- One key concern in employing these technologies is the loss of privacy.



Week 13 Part 2

Chapter 12:Knowledge Systems: Expert Systems, Recommenders, Chatbots, Virtual Personal Assistants, and Robo Advisors

12.2 Expert Systems and Recommenders

- Basic Concepts of Expert Systems (ES)
- Structure and Process of ES
- Limitations of ES
- New generation of ES
- Recommendation Systems

Basic Concepts of Expert Systems (ES)

- A category of autonomous decision systems (the earliest applications of AI).
- Started in 1960s by research institutions (e.g., Stanford University, IBM) and was adopted commercially in 1980s.
- Expert Systems - is a computer-based system that emulates decision making and/or problem solving of human experts. These decisions and problems are in complex areas (e.g., diagnosing a problems)
- ES objective is to transfer the expertise from experts to a machine → to enable nonexperts to make decisions and solve problems that usually require expertise.
- Focus on decisions and problems in narrowly defined domains that require expertise to solve
  - E.g. making small loans, providing tax advice, analyzing reasons for machine failure
- Classical ES use “what-if-then” rules for their reasoning.

Basic Concepts of Expert Systems (ES)

EXPERTS - person who has the special knowledge, judgment, experience, and skills to provide sound advice and solve complex problems in a narrowly defined area.

- Decision performance and the level of knowledge a person has are typical criteria used to determine whether a particular person is an expert as related to ES.

Human experts are typically capable of doing the following:

- Recognizing and formulating a problem.
- Solving a problem quickly and correctly.
- Explaining a solution.
- Learning from experience.
- Restructuring knowledge.
- Breaking rules (i.e., going outside the general norms) if necessary.
- Determining relevance and associations.

Basic Concepts of Expert Systems (ES)

EXPERTISE - The extensive, task-specific knowledge that experts possess. It is associated with a high degree of intelligence and learning from past successes and mistakes.

- The level of expertise determines the success of a decision made by an expert.
- Expertise is often acquired through training, learning, and experience in practice. This includes:
  - explicit knowledge: such as theories learned from a textbook or a classroom
  - implicit knowledge: gained from experience.

Types of experts' knowledge used in ES applications:

- Theories about the problem domain.
- Rules and procedures regarding the general problem domain.
- Heuristics about what to do in a given problem situation.
- Global strategies for solving of problems amenable to expert systems.
- Meta-knowledge (i.e., knowledge about knowledge).
- Facts about the problem area.

Basic Concepts of Expert Systems (ES)

Expertise often includes the following characteristics

- It is usually associated with a high degree of intelligence, but it is not always associated with the smartest person.
- It is usually associated with a vast quantity of knowledge.
- It is based on learning from past successes and mistakes.
- It is based on knowledge that is well stored, organized, and quickly retrievable
- from an expert who has excellent recall of patterns from previous experiences.

Characteristics and Benefits of ES

Benefits of Expert Systems (ES)

- Perform routine tasks (e.g., diagnosis, candidate screening, credit analysis) that require expertise much faster than humans.
- Reduce the cost of operations.
- Improve consistency and quality of work (e.g., reduce human errors).
- Speed up decision making and make consistent decisions.
- May motivate employees to increase productivity.
- Preserve scarce expertise of retiring employees.
- Help transfer and reuse knowledge.
- Reduce employee training cost by using self-training.
- Solve complex problems without experts and solve them faster.
- See things that even experts sometimes miss.
- Combine expertise of several experts.
- Centralize decision making (e.g., by using the “cloud”).
- Facilitate knowledge sharing.

Characteristics and Benefits of ES

Typical Areas for ES Applications

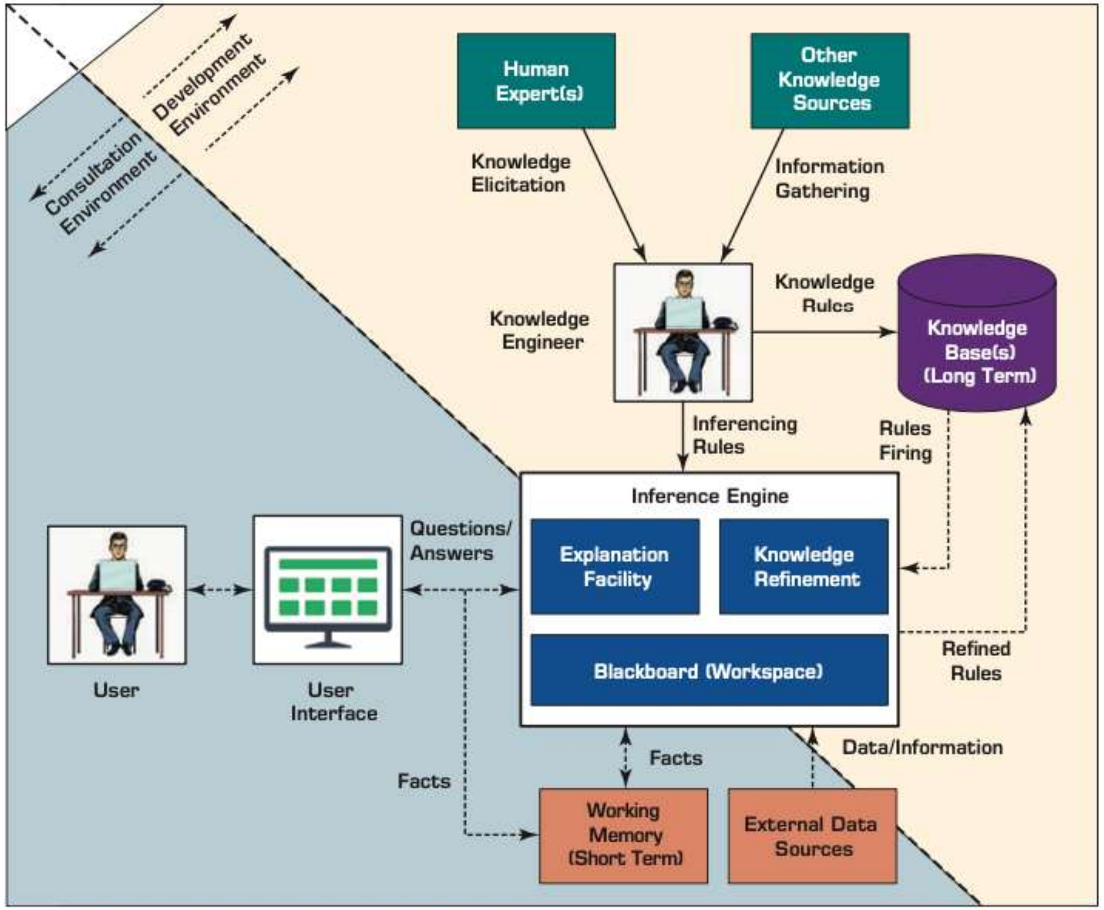
- Finance (investments, credit, and financial reports, …).
- Data processing (planning, equipment selection, equipment maintenance,…).
- Marketing (customer relationship management, market research and analysis,…).
- Human resources (planning, performance evaluation, staff scheduling, ...).
- Manufacturing (e.g., production planning, complex product configuration, …).
- Homeland security (e.g., terrorist threat assessment).
- Business process automation (desk automation, call center management, …).
- Healthcare management.
- Regulatory and compliance requirements.
- Web site design.

Structure and Process of ES

- Development Environment
  - ES developer loads the knowledge-base with appropriate representation of expert knowledge.
- Consultation Environment
  - Used by nonexpert to obtain advice and solve problems using the expert knowledge embedded into the system

Major Components of ES

- Knowledge acquisition
- Knowledge Base (a.k.a repository)
- Knowledge representation
- Inference Engine
- User Interface
- Explanation subsystem and knowledge-refining system.



General Architecture of Expert Systems

Limitations of ES

Classical type of ES is disappearing despite benefits, because:

1. knowledge acquisition from human experts very expensive
2. Acquired knowledge needed to be updated frequently at a high cost.
3. The rule-based foundation are not robust and not too reliable or flexible and could have too many exceptions to the rules.
4. The rule-based user-interface needed to be supplemented (e.g., by voice communication, image maps). This could make ES too cumbersome.
5. The reasoning capability of rule-based technology is limited compared to use of newer mechanisms such as those used in machine learning.

New generation of ES based on machine learning algorithms and other AI technologies are deployed to create better system

New generation of ES

Three major AI types of applications:

- Chatbots.
- Virtual personal assistants.
- Robo advisors.

Other AI technologies that perform similar activities:

- IBM Watson (some of its advising capabilities are similar to those of ES but are much superior).

- Recommendation system (a newer variations that use machine learning and IBM Watson Analytics).

Recommendation Systems

- knowledge system for recommending one-to-one targeted products or services (a.k.a recommendation engine).
  - It tries to predict the importance (rating or preference) that a user will attach to a product or service. Once the rating is known, a vendor knows users' tastes and preferences and can match and recommend a product or service to the user.
  - The recommendations are typically given in rank order.
- Recommendation system uses several AI technologies to provide personalized recommendations.

- Online recommendations are preferred by many people over regular searches, which are less personalized, slower, and sometimes less accurate.

- Top applications areas: movies, music, and books, systems for travel, restaurants, insurance,…

Benefits Of Recommendation Systems

Using recommendation systems may result in substantial benefits both to buyers and sellers.

Benefits to customers	
Personalization	They receive recommendations that are very close to fulfilling what they like or need. This depends, of course, on the quality of the method used.
Discovery	They may receive recommendations for products that they did not even know existed but were what they really need.
Customer satisfaction	With repeated recommendations tends to increase.
Reports	Some recommenders provide reports and others provide explanations about the selected products.
Increased dialog with sellers	Because recommendations may come with explanations, buyers may want more interactions with the sellers.
Benefits to sellers	
Higher conversion rate	With personalized product recommendations, buyers tend to buy more.
Increased cross-sell	Recommendation systems can suggest additional products. E.g., **Amazon.com** shows other products that “people bought together with the product you ordered.”
Increased customer loyalty	As benefits to customers increase, their loyalty to the seller increases.
Enabling of mass customization	This provides more information on potential customized orders.

Recommendation Systems

- Two methods for building recommendation systems are are collaborative filtering and content-based filtering.

Collaborative filtering

- Builds a model that summarizes the past behavior of shoppers (browsing, purchase, rating).
- Considers what shoppers with similar profiles bought and how they rated their purchases.
- Uses AI algorithms to predict the preference of both old and new customers. Then, the computer program makes a recommendation.

Content-based filtering

- Allows vendors to identify preferences by the attributes of the product(s) that customers have bought or intend to buy.
- The vendor recommends to customers products with similar attributes (Knowing product preferences).
- E.g., the system may recommend a text-mining book to a customer who has shown interest in data mining, or action movies after a consumer has rented one in this category.

- Several other filtering methods exist (e.g., rule-based filtering and activity-based filtering).

12.3 CONCEPTS, DRIVERS, AND BENEFITS OF CHATBOTS

Chatbots

- The world is now infested with chatbots – 60% of millennials have already used them (Knight, 2017).
- Chatbot - A computerized service that enables easy conversations between humans and humanlike computerized (robots, image characters, or sometimes over the Internet) in writing, voice or images. Conversations frequently involve short questions and answers and are executed in a natural language.
- Used primarily for information search, communication and collaboration, and rendering advice in limited, specific domains.

Chatbot Evolution

- Chatbots originated decades ago.
- Eliza first Q&A chatbot, it is simple ES that enabled machines to answer questions posted by users.
- Since 2000, more and more capable AI machines for Q&A dialogs have been developed.
- Around 2010, conversational AI machines were named chatbots and later were developed into virtual personal assistants, championed by Amazon's Alexa.

Chatbots

- Chatbots contain a knowledge-base (e.g., rule- based) and a natural language understanding capability, in order to converse with a human.
  - Knowledge bases are updated today in the “cloud” in a central location.
  - The stored knowledge is matched with questions asked by users.
- To understand unstructured dialog intelligent chatbots are equipped with NLPs.
  - Advanced chatbots can also understand human gestures, cues, and voice variations.
- Learning chatbots gain more knowledge with their accumulated experience.
- Chatbots are used for many different tasks (e.g., education, banking, insurance, retail, travel, healthcare, and customer experience,…).
  - The service is often available on messaging services such as Facebook Messenger or WeChat, and on Twitter.

Chatbots

Types of Bots – according to their capabilities:

- **Regular bots:** conversational intelligent agents that do simple, repetitive, tasks (e.g., showing customers their bank's debits)
- **Chatbots:** stimulate conversations with people.
- **Intelligent bots:** have a knowledge base that is improving with experience (e.g., Alexa, some robo advisors).

Drivers use of Chatbots and Benefits

Drivers

- The need to cut costs.
- The increasing capabilities of AI, especially NLP and voice technologies.
- The ability to converse in different languages.
- The increased quality and capability of captured knowledge.
- The push of devices by vendors (e.g., virtual personal assistants such as Alexa and Google Assistant).

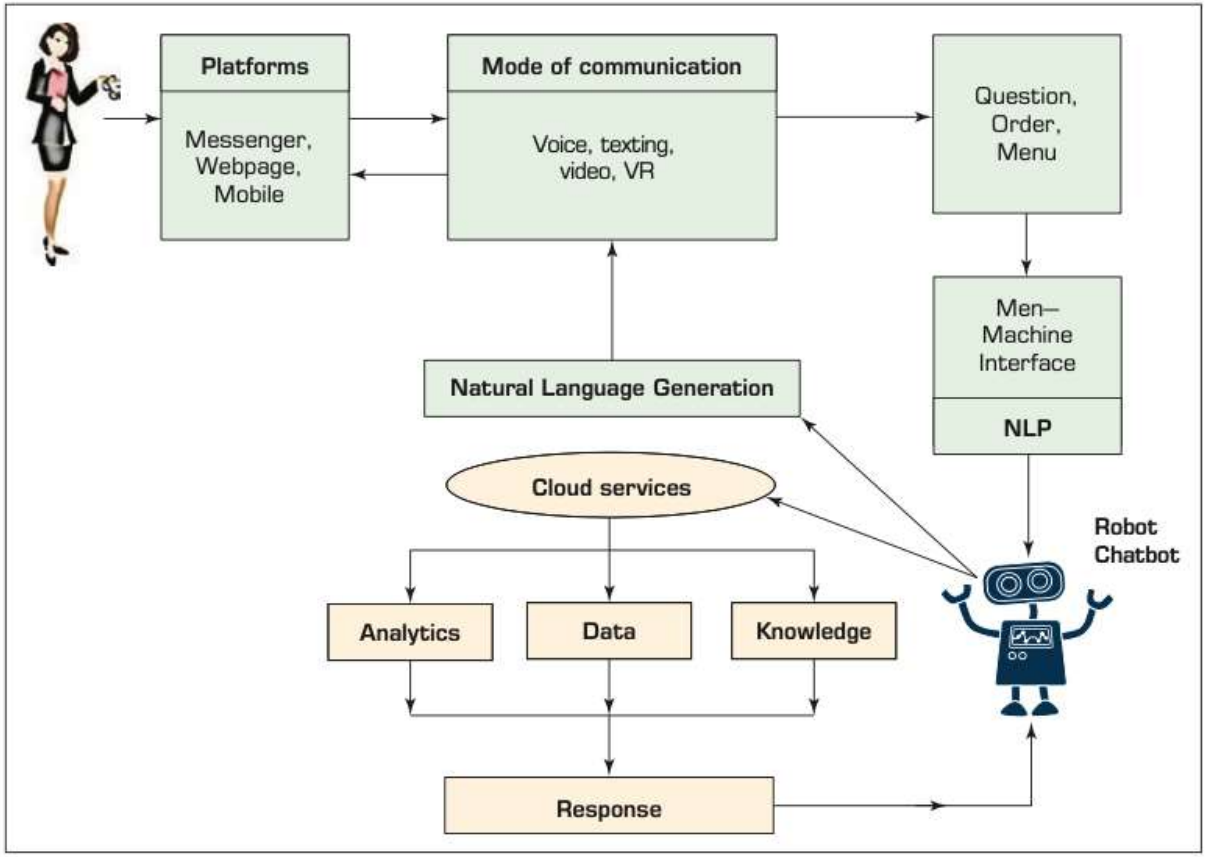
Benefit

- Providing superb and economic customer service and conducting market research.
- Its use for text and image recognition.
- To facilitate shopping.
- Its support of decision making.

Components of Chatbots and the Process of Their Use

The major components of chatbots:

- A person (client).
- A computer, avatar, or robot (the AI machine).
- A knowledge base that can be embedded in the machine or available and connected to the "cloud."
- A human-computer interface that provides the dialog for written or voice modes.
- An NLP that enables the machine to understand natural language.



The process of chatting with chatbots

Chatbots

Representative Chatbots from around the world

- **RoboCoke:** This is a party and music recommendation bot created for Coca-Cola
- **Kip:** This shopping helper is available on Slack (a messaging platform).
- **Zoom:** an automated virtual assistant, is for everyone in the workplace.
- **… and more in the book.**

Major categories of chatbots' applications

- Chatbots for enterprise activities, including communication, collaboration, customer service, and sales.
- Chatbots that act as personal assistants (virtual personal assistants).
- Chatbots that act as advisors (a.k.a robo advisors), mostly on finance-related topics.

12.4 Enterprise Chatbot

12.5 Virtual Personal Assistants

12.6 Chatbots as Professional Advisors (Robo Advisors)

Enterprise Chatbots

- Chatbots can fundamentally change the way that business is done both in external and internal applications.
- The benefits of chatbots to enterprises:
  - making dialog less expensive and more consistent.
  - interact with customers and business partners more efficiently,
  - available anytime 24/27
  - can be reached from anywhere.

Reasons for using enterprise bots (Beaver, 2016):

- “AI has reached a stage in which chatbots can have increasingly engaging and human conversations, allowing businesses to leverage the inexpensive and wide- reaching technology to engage with more consumers.

- Chatbots are particularly well suited for mobile-perhaps more so than apps. Messaging is at the heart of the mobile experience, as the rapid adoption of chat app demonstrates’’
- … more on the book.

Examples of Enterprise Chatbots

Marketing and customer experience:

- Very useful in providing marketing and customer services (e.g., obtaining sales leads, persuading customers to buy products and services, providing critical information to potential buyers, optimizing advertising campaigns).
- Using voice and texting can provide personalization as well as superb customer experience.
- Can enable vendors to improve personal relationships with customers and run marketing campaigns.
- E.g., **LinkedIn** is introducing chatbots that conduct tasks such as comparing the calendars of people participating in meetings and suggesting meeting times and places.

Financial Services

- Banking Chatbots can use predictive analytics and cognitive messaging to perform tasks such as making payments, inform customers about personalized deals.
- Banks' credit cards can be advertised via chatbots on Facebook Messenger.
- E.g., Citi Bank Chatbot it can answer FAQs about people's accounts in a natural language.

Examples of Enterprise Chatbots

Service Industries

Healthcare

- Robot receptionists direct patients to departments in hospitals.
- Several chatbots are chatty companions for people who are elderly and sick.
- Chatbots are used in telemedicine; patients converse with doctors and healthcare professionals who are in different locations.
- Chatbots can connect patients quickly and easily with information they need.

Education

- Chatbot tutors are used in several countries to teach subjects ranging from English to mathematics.
- The chatbot treats all students equally.
- Machine translation of languages will enable students to take online classes in languages other than their own.
- chatbots can be used as private tutors.

Government

- Chatbots are spreading in government as a new dialog tool for use by the public (Lacheca, 2017).
- E.g., providing access to government information and answering government-related questions.

Travel and Hospitality

- Chatbots are working as tour guides in several countries (e.g., Norway).
- They act as concierges, providing information and personalized recommendations (e.g., about restaurants), arrange reservations for hotel rooms, meals, and events.

Chatbots Inside Enterprises

- Companies lately have started to use chatbots to automate tasks for supporting internal communication, collaboration, and business processes.

Benefit (Newlands (2017a):

  - Chatbots can support decision-making activities.
  - cut costs
  - increase productivity
  - assist working groups
  - foster relationships with business partners.

Representative examples of chatbot tasks are inside enterprises:

  - Help with project management, Handle data entry, Conduct scheduling, Streamline payments with partners, Advise on authorization of funds, Monitor work and workers, Analyze internal Big Data, Find discounted and less expensive products. Simplify interactions, Facilitate data-driven strategy, Use machine learning, Facilitate and manage personal finance.

- Chatbots' Platform Providers: ChattyPeople, Kudi, Twyla, BM Watson, Microsoft's Bot Framework.

Knowledge for Enterprise Chatbots

- Knowledge for chatbots depends on their tasks.
- Enterprise chatbots operate very similarly to ES except that the interface occurs in a natural language and frequently by voice.
  - For example, the knowledge of Sephora’s bot is specific to that company and its products and is organized in a Q&A format.
- Most marketing and customer care bots require proprietary knowledge, which is usually generated and maintained in-house. This knowledge is similar to that of ES.
- Chatbots that are used within the enterprise (e.g., to train employees) may not be company specific.
  - A company can buy the knowledge and modify it to fit local situations and its specific needs (as is done in ES)
  - Newer chatbots use machine learning to extract knowledge from data.

Virtual Personal Assistants (VPA)

- An emerging type of chatbot is designed as a virtual personal assistant for both individuals and organizations.
- Major objective is to help people improve their work, assist in decision making, and facilitate their lifestyle.
  - Examples of such assistants are, Apple's Siri, Google's Assistant, and Microsoft Cortana.
  - A well-known VPA is Amazon's Alexa that is accessed via a smart speaker called Echo (or other smart speakers).

- Assistant for Information Search help users conduct a search by voice for information.
  - In business situations delegating the search to a machine may save sellers considerable money and make customers happy by not having to wait for the service.

Knowledge for Virtual Personal Assistants

- knowledge for virtual personal assistants is kept and centrally maintained in the "cloud".
- So, it become available to millions of users, and need to provide dynamic, updated information (e.g., weather conditions, news, stock prices).
- Knowledge usually universal and disseminated via a Q&A dialog.

Chatbots as Professional Advisors (Robo Advisors)

- A special category of virtual personal assistants designed to provide personalized professional advice in specific domain. Mainly they operate in investment and portfolio management.

Robo Financial Advisors

- Online providers that offer automated, low-cost, personalized investment advisory services.
- They use AI algorithms that allocate, deploy, rebalance, and trade investment products.

Robo Advisors 2.0: several of the fully automated advisors started to add a human option (a.k.a the human touch)

- Quality of advice provided by robo advisors depends on their knowledge, the type of investments involved, the inference engine of the AI machine.

Chatbots as Professional Advisors (Robo Advisors)

Other Professional Advisors

- **Computer operations:** To cut costs, major computer vendors (hardware and soft- ware) try to provide users with self-guides to solve encountered problems.
- **Travel:** Several companies provide advice on planning future national and international trips.
- **Medical and health advisors:** A large number of health and medical care advisors operate in many countries.
  - E.g, Health Tap acts like a medical doctor by providing a solution to common symptoms provided by patients.
  - Florence is a personal nurse available on Facebook Messenger.
- **Shopping advisors:** act as shopping advisors
  - E.g., shopadvisor.com/our-platform, a comprehensive platform that includes three components: Product intelligence, Context intelligence, Shopper intelligence to help companies attract customers.
- **IBM Watson: p**robably the most knowledgeable virtual advisor.

12.7 Implementation Issues

- Disadvantages and Limitations of Bots
- Quality of Chatbots
- Constructing Bots

Disadvantages and Limitations of Bots

- Some bots provide inferior performance, at least during their initiation, making users frustrated.
- Some bots do not properly represent their brand. Poor design may result in poor representation.
- The quality of AI-based bots depends on the use of complex algorithms that are expensive to build and use.
- Some bots are not convenient to use.
- Some bots operate in an inconsistent manner.
- Enterprise chatbots pose great security and integration challenges.

Disadvantages and Limitations of Bots

Technology issues

- Virtual personal assistants have imperfect voice recognition.
  - No good feedback system yet for voice recognition systems to understands the users.
  - Voice recognition systems may not know when to do a current task and need to ask for human intervention.

- Connecting to an NLP system
  - Security and connectivity difficulties…

- Multilingual chatbots need to be connected to a machine language translator.

Quality of Chatbots

- The quality of most Chatbots is not perfect, but improving over time.
- Chatbots that retrieve information for users and are properly programmed can do a perfect job.
- Bots that serve a large number of people, such as Alexa and Google Assistant, exhibit an increasing level of accuracy.

Quality of robo advisor

- Major issue when engaging bots is the potential loss of human touch.
- Bots cannot bring empathy or a sense of friendship.
- Solution:
  1. Bots should perform only tasks that they are suited to do.
  2. They should provide a visible benefit to the customer.
  3. Because the bots face customers, the interactions must be fully planned to make sure the customers are happy.

- **Microsoft's Tay:** Twitter-based chatbot that failed and was discontinued by Microsoft.

Constructing Bots

Setting Up Alexa’s Smart Home System

- Alexa is useful in controlling smart homes.
- A six-step process for how to use Alexa in smart homes (Crist ,2017):
  1. Get a speaker (e.g., Echo).
  2. Think about the location of the speaker.
  3. Set up the smart home devices.
  4. Sync related gadgets with Alexa.
  5. Set up group and scene.
  6. Fine-tune during the process.

Constructing Bots

Guide for creating a Facebook Messenger bot

1. Give it a unique name.
2. Give customers guides on how to build a bot and how to converse with it.
3. Experiment in making a natural conversation flow.
4. Make the bot sound smart, but use simple terminology.
5. Do not deploy all features at the same time.
6. Optimize and maintain the bot to constantly improve its performance.

Constructing Bots

Using Microsoft's azure bot service

- Azure is a comprehensive but not a very complex bot builder.
- Its Bot Service provides five templates for quick and easy creation of bots.

TABLE 12.1 Azure's Templates

Template	Description
Basic	Creates a bot that uses dialogues to respond to user input.
Form	Creates a bot that collects input from users via a guided conversation that is created using Form Flow.
Language understanding	Creates a bot that uses natural language models (LUIS) to understand user intent.
Proactive	Creates a bot that uses Azure Functions to alert users of events.
Question & Answer	Creates a bot that uses a knowledge base to answer users' questions.

Main Reference

- **Chapter 9**: "*Big Data, Cloud Computing, and Location Analytics: Concepts and Tools*" from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".*

- **Chapter 12**: "*Knowledge Systems: Expert Systems, Recommenders, Chatbots, Virtual Personal Assistants, and Robo Advisors* " from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".*

Week self-review exercises

- **Application case 9.1 – 9.11** from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".*

- **Application case 12.1 – 12.7** from "*Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support".*

- Getting started with Watson Assistant: https://cloud.ibm.com/docs/assistant?topic=assistant-getting-started

Thank You