Data Mining - Presentation

Dr. Jean-Michel RICHER

FACULTÉ
DES SCIENCES
Unité de formation
et de recherche
DÉPARTEMENT
INFORMATIQUE

2018

jean-michel.richer@univ-angers.fr

Outline

1. Introduction

Flood of data

With the flood of data available nowdays, companies are turning to **analytics** solutions to extract **meaningful and valuable information** in order to help improve **decision making**

Flood of data

With the flood of data available nowdays, companies are turning to **analytics** solutions to extract **meaningful and valuable information** in order to help improve **decision making**

Analysis and analysts

Those companies need capabilities (tools, human resources)

- to **analyze** historical data
- **forecast** what might happen in the future

Analysts

The people who are doing this job are called mathematicians, statisticians, business analysts or **data scientists**

Analysts

The people who are doing this job are called mathematicians, statisticians, business analysts or **data scientists**

Research area

- Expert System ('70)
- Knowledge Discovery in Database ('80)
- Machine Learning, Adaptative Learning
- Business Intelligence, Decision Making
- Data Mining, **Big Data**

A word on analytics

Analysis vs Analytics (Wikipedia)

Analytics is the discovery, interpretation, and communication of meaningful patterns in data.
It generally refers to the **methodology**

Analytics is multidisciplinary

- extensive use of mathematics and statistics
- analyze massive, complex data sets
- use of software to store / collect, organize, select, process data

The Walmart 2004 example (1/4)

Walmart Database

- in 2004 a series of hurricanes crossed the state of Florida
- after the first hurricane : *what customers really wanted to buy prior to the arrival of a hurricane ?*
- in **Walmart** Retail Transaction Database one particular item that increased in sales by a factor of 7 over normal shopping days was found
- can you guess what it was ?

The Walmart 2004 example (2/4)

Why would people buy strawberry pop tarts ?

The Walmart 2004 example (3/4)

Why do people buy this article

- do not require refrigeration
- do not need to be cooked
- come in individually wrapped portions
- have a long shelf life
- are a snack / breakfast food
- everybody love them

The Walmart 2004 example (4/4)

a win-win partnership

- Walmart stocked their stores with tons of strawberry pop tarts prior to the next hurricanes
- sold them out
- Walmart wins by making the sell
- customers win by getting the product that they most want

2. Analytics : what, why, future, action

Kinds of analysis (1/2)

What, Why, Future and Action

- **Descriptive** *What has happened ?* insight into the past or current state
- **Diagnostic** *Why something happened ?*
- **Predictive** *What could happen ?* predict the future
- **Prescriptive** *Actions required to influence a particular outcome*

Descriptive, Predictive and Prescriptive analytics, are inter-related solutions helping companies make the most out of the big data that they have

Kinds of analysis (2/2)

What, Why, Future and Action

- **Descriptive:** profit per store, per region
- **Diagnostic:** why did sales go down in particular region
- **Predictive:** which products are likely to perform better in next quarter based on past data
- **Prescriptive:** which customer segment shall be targeted next quarter to improve profitability

Kinds of analysis (2/2)

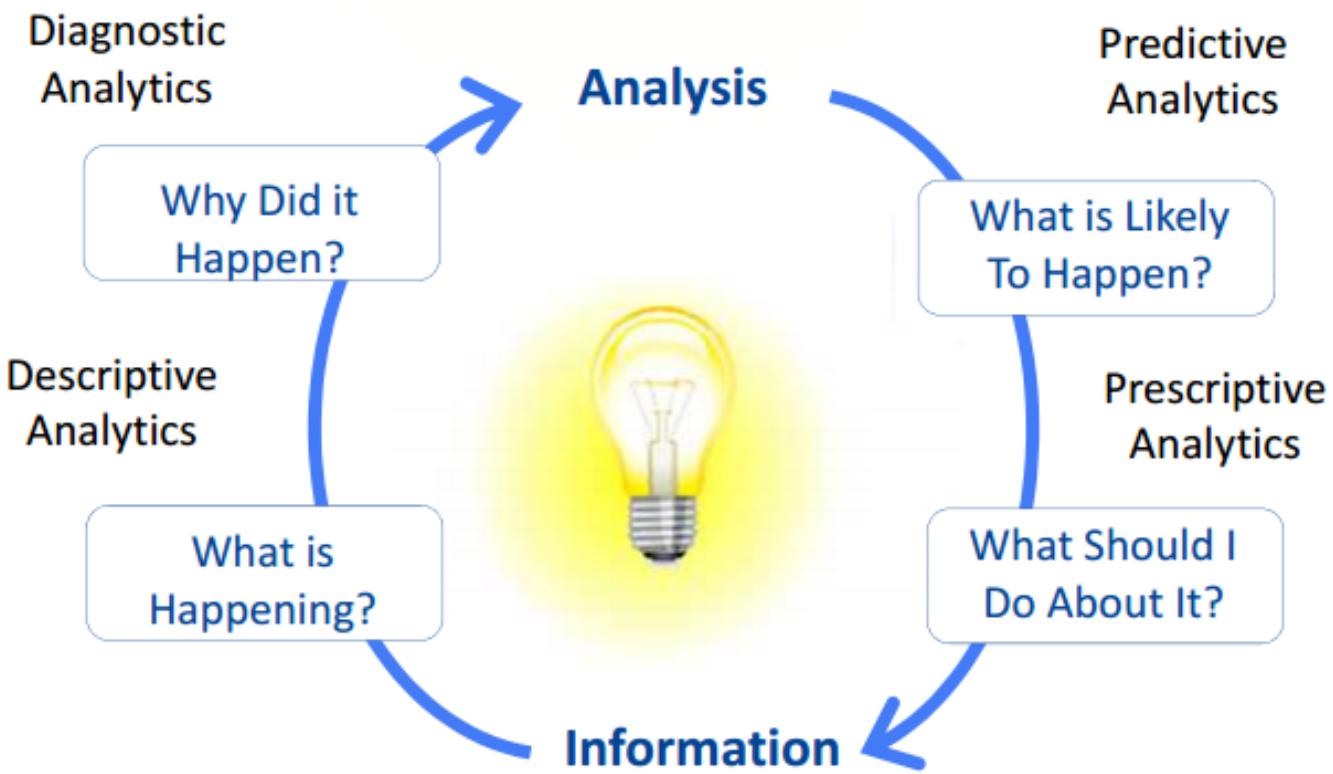
What, Why, Future and Action

- **Descriptive:** profit per store, per region
- **Diagnostic:** why did sales go down in particular region
- **Predictive:** which products are likely to perform better in next quarter based on past data
- **Prescriptive:** which customer segment shall be targeted next quarter to improve profitability

My vision

- **past:** Descriptive and Diagnostic
- **future:** Predictive and Prescriptive

The four kinds of analysis



Descriptive Analysis - BI and DM

- describe or summarize data
- make it something that is interpretable by humans
- nearly 80% of the data mining workload
- example: reports that provide historical insights regarding the company's production, financials, operations, sales, finance, inventory and customers (Dash board)
- is it reporting ?

Predictive analysis (1/2)

Predictive Analysis - Forecasting

- forecast (estimate) the future (not 100% true)
- set realistic goals for the business, effective planning and restraining expectations
- use various statistical and machine learning algorithms
- take existing data and fill in the missing data with best possible guesses

Predictive analysis (2/2)

Can be further categorized as

- forecasting (what if the trends continue)
- predictive modelling (what will happen if ...)
- root cause analysis (why it happens)
- Monte-Carlo simulation (what could happen)
- Pattern identification and alerts (when to correct a process)

Prescriptive analysis (1/2)

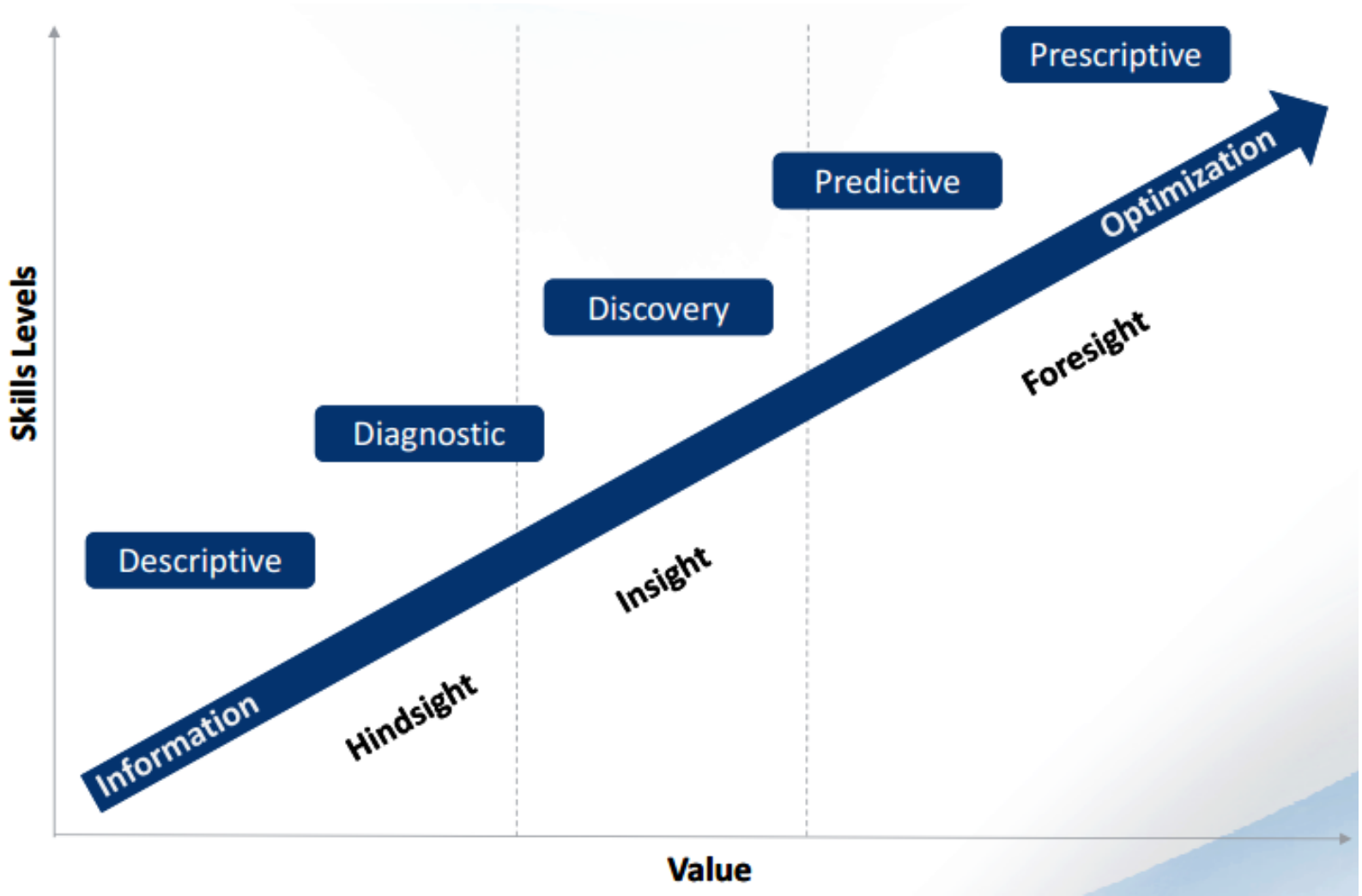
Prescriptive Analysis

Advanced analysis based on

- with too many choices, which one is the best ?
- stochastic optimizations to identify data uncertainties
- a combination of data, mathematical models and various business rules

3. Broaden your knowledge

The five kinds of analytics

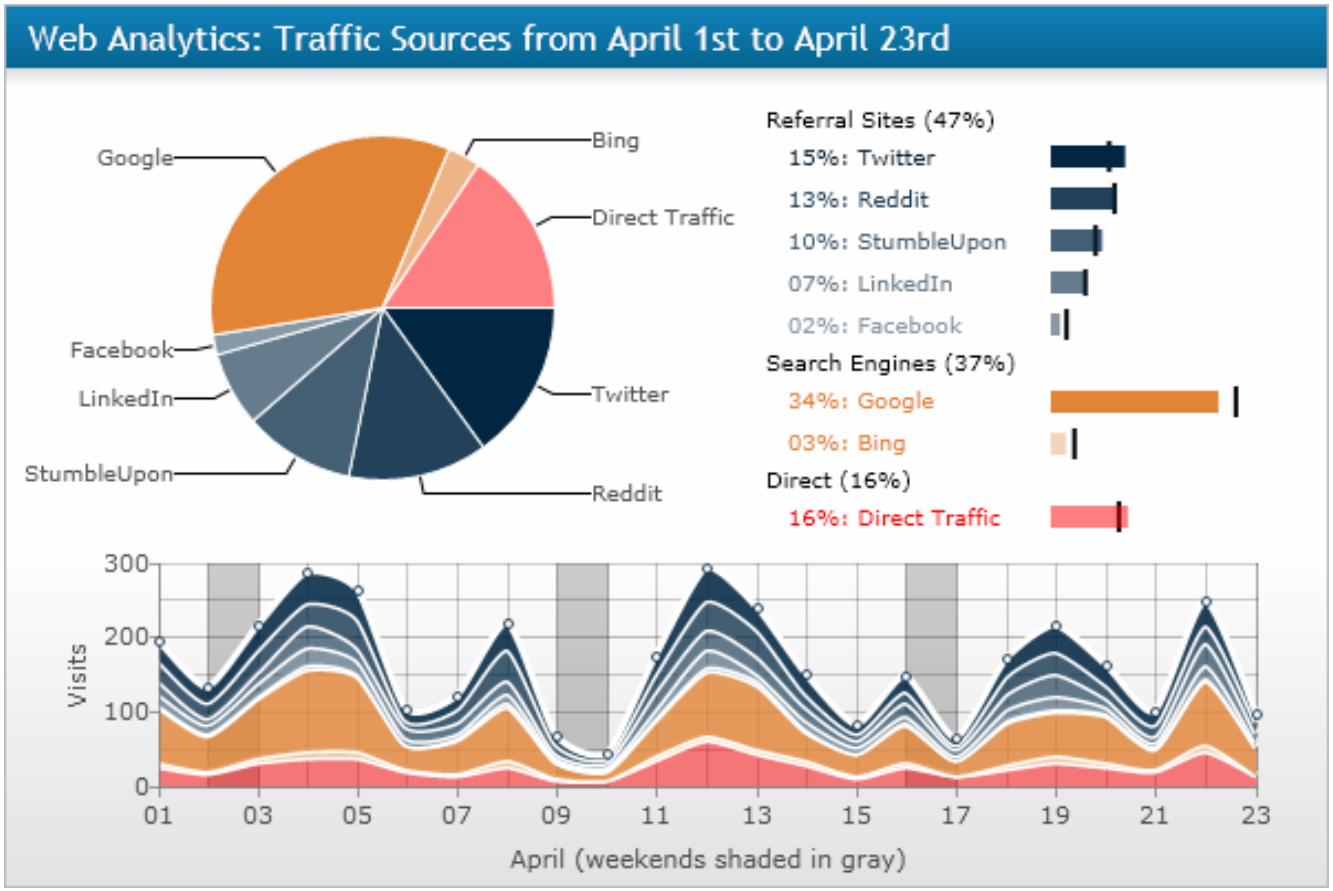


Tools and analysis

A good combination

- each type of analytics (analysis) has its methods
- analysts buy tools to perform the analytics in discrete steps
- but many decisions require a combination of data analysis and human experience

Use graphics



Think of what to do

Exercise: Products sale

- Imagine you are a supermarket and you want to sell / promote five different products
- what would you do ?
- how would you measure if your strategy worked ?

4. Skills requirements

Skills

- be able to define the different kinds of analysis of data mining
- be able to explain a particular analysis (like the descriptive or predictive analysis)

4. End

FACULTÉ
DES SCIENCES
Unité de formation
et de recherche
DÉPARTEMENT
INFORMATIQUE

UA - Angers
2 Boulevard Lavoisier 49045 Angers
Cedex 01
Tel: (+33) (0)2-41-73-50-72

université
angers
FACULTÉ
DES SCIENCES
Unité de formation
et de recherche
DÉPARTEMENT
INFORMATIQUE