Data Mining, Data Science, Business Analytics

José Luís Borges

jlborges@fe.up.pt

(adaptado de materiais cedidos por Carlos Soares csoares@fe.up.pt)

Introduction

- Analytics in business is booming
- Businesses collect data about many processes
- There are several external sources of data available
- Companies want to exploit the data for competitive advantages
  - It is not enough to build reports and create dashboards
  - Moving the focus of interest from the past to the future

Introduction

- From
  - How many customers did we lose last year?

  (descriptive)
- To
  - Who will most likely churn within the next 10 days and what can we do about it?

  (predictive)

Introduction

- From
  - What campaign was the most successful in the past?
- To
  - What will be the next action to trigger a purchase action for each of the prospects?

What is data mining?

- “non-trivial process of identifying valid, novel and potentially useful and ultimately understandable **patterns in data**”

- (Fayyad, piatetsky-shapiro and smyth, 1996)

knowledge discovery process involves: data cleaning, data integration, data selection, data transformation, *data mining*, evaluation, presentation

Application example

- Germany won the 2014 **world cup** against Argentina, having Big Data on its side!
- Partnered SAP to create a custom match analysis tool that collects and analyzes player performance data
  - data captured by video cameras around the pitch
  - performance indicators for individual players
  - team virtual “defensive shadows” that show how much area a player can protect with his own body. That can help visualize and exploit weak links in an opponent’s setup
  - ...

- There were eight cameras covering each pitch in Brazil and data was available to all the teams
- only Germany made use of this type of big data analytics.
- Oliver Bierhoff, assisting coach, said:
  - “We had a lot of qualitative data for the opposition available. Jerome Boateng asked to look at the way Cristiano Ronaldo moves in the box, for example.
  - And before the game against France, we saw that the French were very concentrated in the middle but left spaces on the flanks because their full-backs didn’t push up properly. So we targeted those areas.”

16 JUN 2014 - 13:00 Local time **GROUP G** Arena Fonte Nova Salvador	GERMANY	FULL-TIME 4-0	PORTUGAL	
21 JUN 2014 - 16:00 Local time **GROUP G** Estadio Castelao Fortaleza	GERMANY	FULL-TIME 2-2	GHANA	
26 JUN 2014 - 13:00 Local time **GROUP G** Arena Pernambuco Recife	USA	FULL-TIME 0-1	GERMANY	
30 JUN 2014 - 17:00 Local time **ROUND OF 16** Estadio Beira-Rio Porto Alegre	GERMANY	FULL-TIME 2-1 Germany win after extra time	ALGERIA	
04 JUL 2014 - 13:00 Local time **QUARTER-FINALS** Estadio do Maracana Rio De Janeiro	FRANCE	FULL-TIME 0-1	GERMANY	
08 JUL 2014 - 17:00 Local time **SEMI-FINALS** Estadio Mineirao Belo Horizonte	BRAZIL	FULL-TIME 1-7	GERMANY	
13 JUL 2014 - 16:00 Local time **FINAL** Estadio do Maracana Rio De Janeiro	GERMANY	FULL-TIME 1-0 Germany win after extra time	ARGENTINA	

A simple example

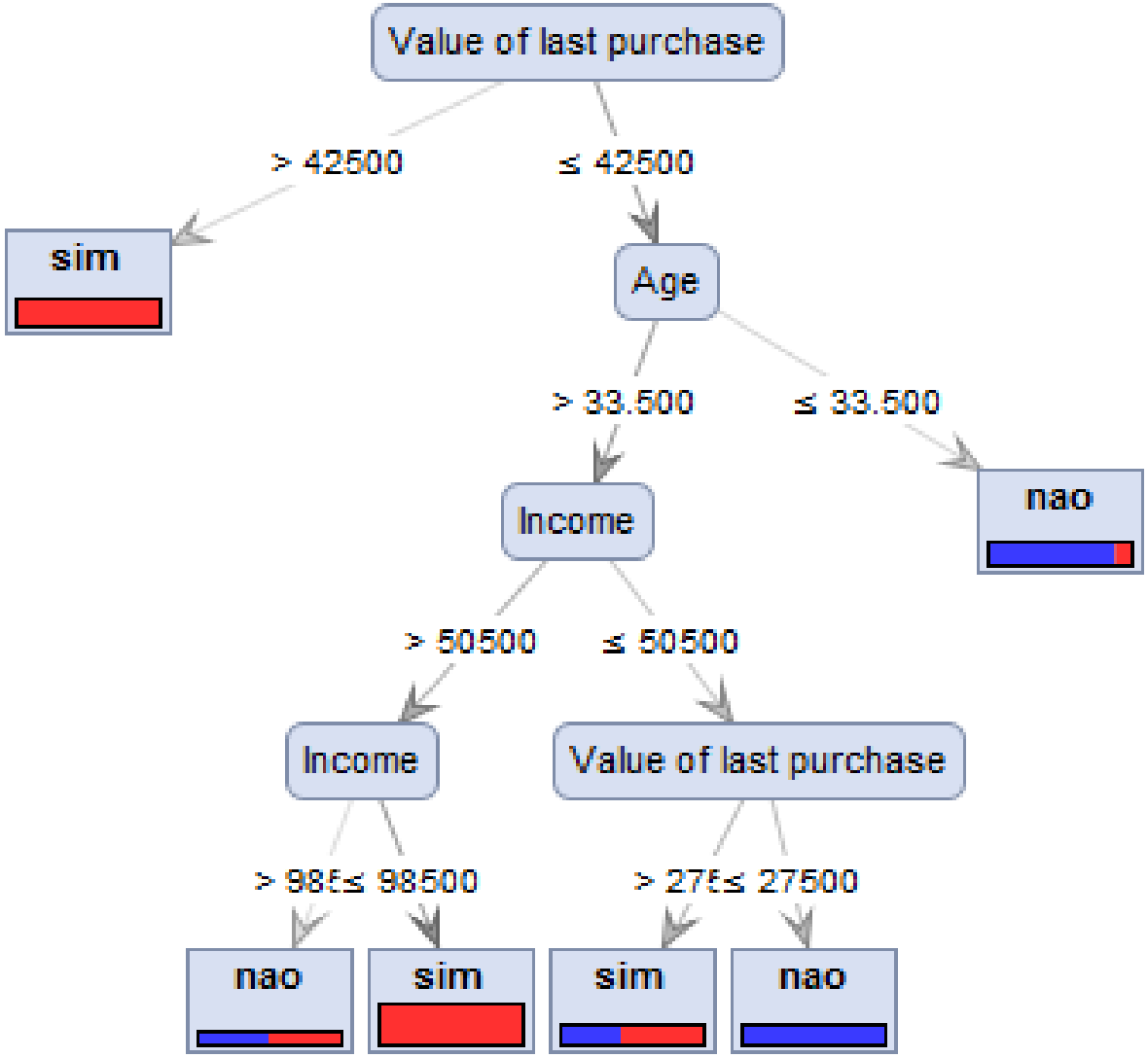
Who should I send the promotion to?

Bought?	Age	Income	Family size	Cars bought previously	Value of last purchase
	41	50000	2	1	0
	39	68000	2	0	30000
	58	61000	4	0	0
	26	25000	3	0	0
	21	50000	1	1	20000
	38	43000	2	0	0
	44	43000	4	1	47000
	27	47000	2	1	21000
	70	23000	2	0	25000

Maybe I could learn something from previous promotions...

Bought?	Age	Income	Family size	Cars bought previously	Value of last purchase
nao	37	49000	2	1	42000
sim	43	68000	3	0	0
sim	42	61000	4	0	0
sim	26	52000	2	0	0
sim	40	64000	1	1	21000
sim	38	52000	1	0	0
sim	45	43000	4	1	47000
sim	35	45000	2	1	34000
nao	39	43000	2	0	0
sim	31	55000	3	1	46000
sim	34	57000	3	1	52000
nao	38	44000	4	0	0
nao	34	68000	2	1	33000
sim	30	45000	2	1	44000
sim	38	41000	3	1	47000
sim	40	62000	3	0	0
sim	43	69000	2	0	0
nao	26	45000	3	0	0
sim	35	66000	4	1	17000
...	...	...	...	...	...

a model relating attributes to result



and use it to make predictions

prediction...	confidence(nao)	confidence(sim)	Age	Income	Family si...	Cars bough...	Value of last...
nao	1	0	41	50000	2	1	0
sim	0	1	39	68000	2	0	30000
sim	0	1	58	61000	4	0	0
nao	0.889	0.111	26	25000	3	0	0
nao	0.889	0.111	21	50000	1	1	20000
nao	1	0	38	43000	2	0	0
sim	0	1	44	43000	4	1	47000
nao	0.889	0.111	27	47000	2	1	21000
nao	1	0	70	23000	2	0	25000

Other names for Data Mining

- Data science
  - Involves principles, processes, and techniques for **understanding phenomena via the analysis of data**
- Business analytics
  - Refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive **business** planning. Focuses on developing **new insights and understanding of business performance based on data and statistical** methods

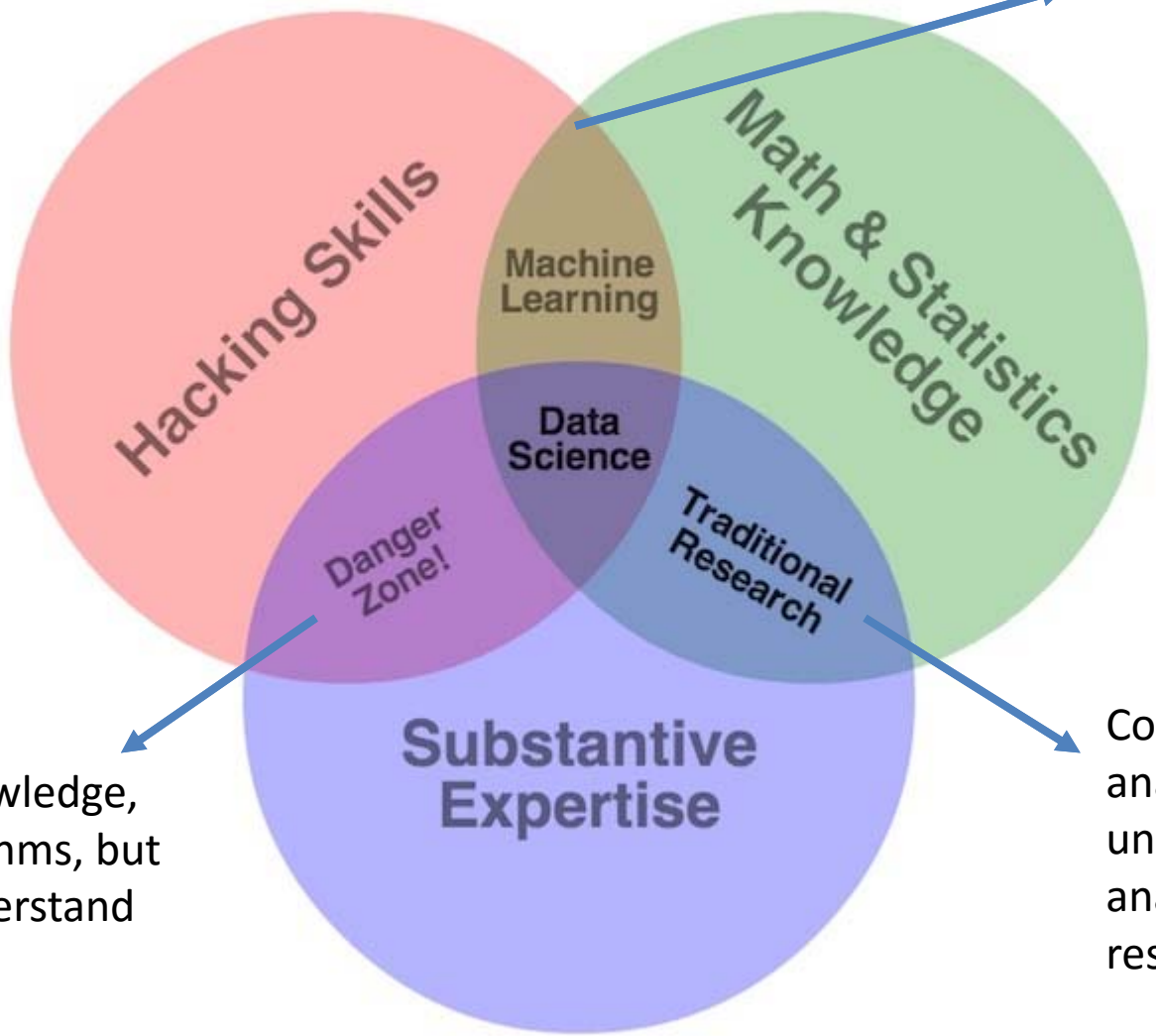
Learning from data

- Data mining explores the study and construction of algorithms that can learn from data
- Basic Idea:
  - Instead of trying to create a very complex program to do X
  - Use a (relatively) simple program that can learn to do X
- Example:
  - Instead of trying to program a car to drive (If light(red) && NOT(pedestrian) || speed(X) <= 12 && ..),
  - create a program that watches human drive, and learns how to drive

Why learning from data

- It is often much cheaper, faster and more accurate
- It may be possible to teach a computer something that we are not sure how to program.
  - We could explicitly write a program to tell if a person is obese
    - If (weight_kg /(height_m × height_m)) > 30, printf("Obese")
  - It is hard to write a program to tell is a person is sad
    We can easily obtain a 1,000 photographs of sad/not sad people, and ask an algorithm to learn to tell them apart

The data scientist Venn diagram (1)

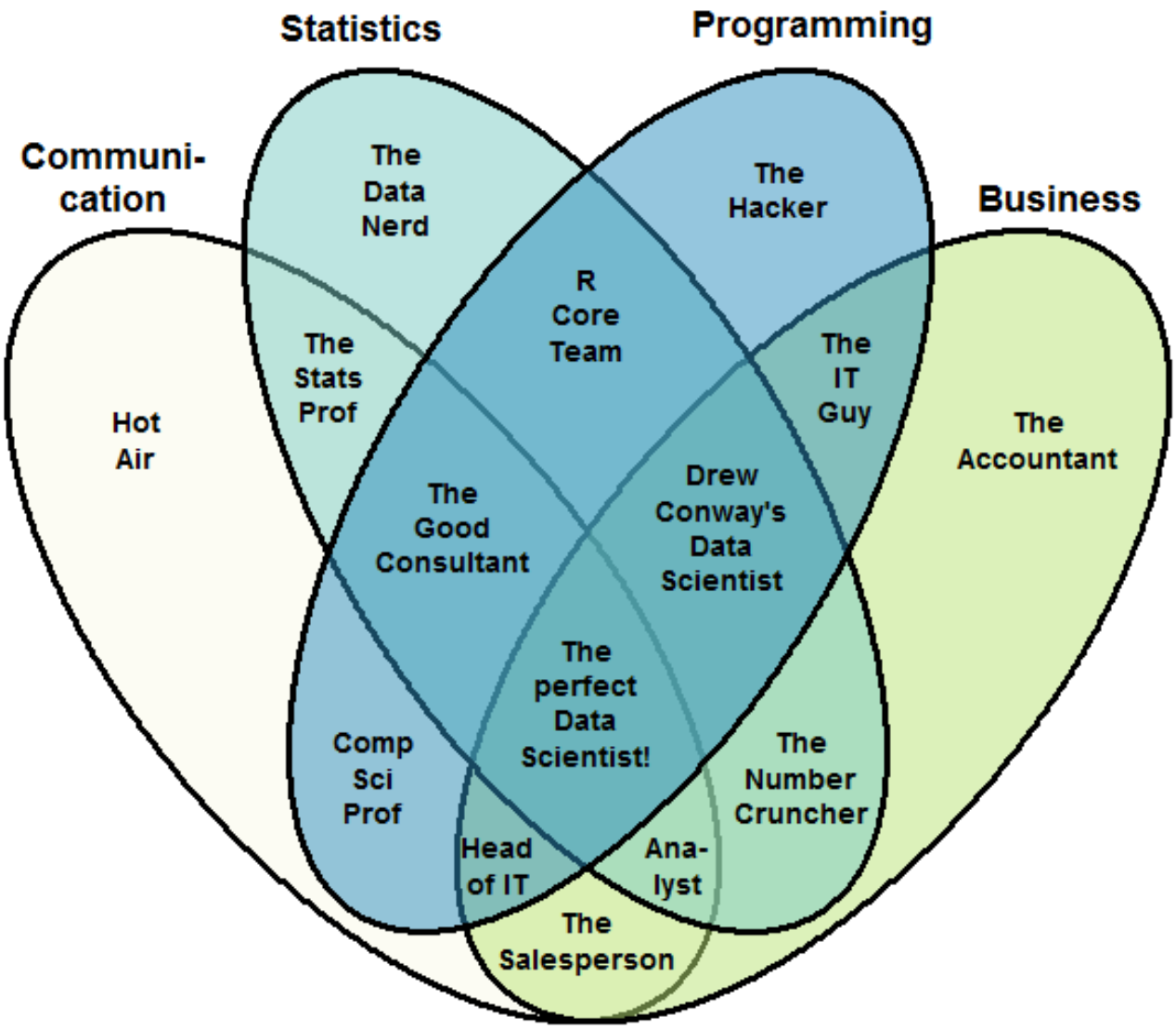


Apply algorithms to large amounts of data, and understand both the algorithms and the results

Collect knowledge, apply algorithms, but do not understand

Collect data, analyze it and understand the analytical process and results

The data scientist Venn diagram (2)



Hot topic

fortune.com/2012/01/06/the-hot-tech-gig-of-2022-data-scientist/

FORTUNE

The hot tech gig of 2022: D scientist

Jessi Hempel
Updated: Jan 06, 2012 10:00 AM GMT

By the end of the decade 50 billion devices wi emitting information nonstop. Data scientist help manage it all.

FORTUNE -- A decade from now the smart techies who decided to become app developers may wish they had taken an applied-mathematics class or two. The coming deluge of data

https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

Harvard Business Review

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

SUMMARY | SAVE | SHARE | COMMENT 9 | TEXT SIZE | PRINT | $8.95 BUY COPIES

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million

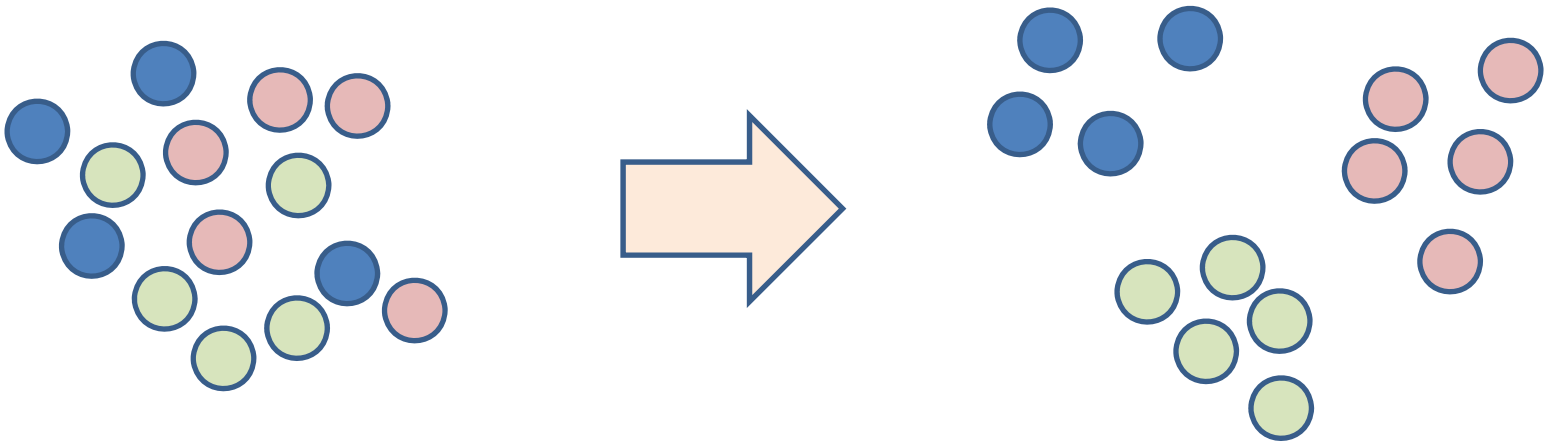
> "Data scientists have tended to come from two different disciplines, computer science and statistics, but the best data science involves both disciplines. One of the dangers is statisticians not picking up on some of the new ideas that are coming out of machine learning, or computer scientists just not knowing enough classical statistics to know the pitfalls."

Rob Hyndman

https://blog.stitchdata.com/5-things-you-should-know-before-getting-a-degree-in-data-science-40cddf44aac3

What for?

- **Clustering** - group elements
  - customers according to credit card spending
  - customers according to call behavior
  - shops according to customers segments
  - doctors according to prescription patterns
  - http://yippy.com



Yippy - data mining

yippy.com/search/?v%3Aproject=clusty-new&query=data+mining&xtoken=30071668358f4d32b88558

yippy

data mining

Sources Sites Time Topics

Top 558 Results remix

- + Knowledge, Discovery (52)
- + Data warehousing (36)
- + Learning, Machine (40)
- + Education (29)
- + Artificial intelligence (13)
- + Data Mining Tools (22)
- + Healthcare (16)
- + SAS (9)
- + Similar data gathering and extraction (13)
- + Blog (15)
- + Image (18)
- + Conference, International (16)
- + Data Mining Techniques (13)
- + Predictive modeling (15)
- + Canada, CBC (7)
- + Tutorials (12)

Data mining - Wikipedia new window preview

Data mining is the computing process of discovering patterns in large **data** sets involving machine ...

https://en.wikipedia.org/wiki/**Data_mining** - - Yippy Index V

Data Mining: What is **Data Mining**? - UCLA Anderson School ... new windo

Overview Generally, **data mining** (sometimes called **data** or knowledge discovery) is the p summarizing it into ...

www.anderson.ucla.edu/.../technologies/palace/datamining.htm - - Yippy Index V

What is **data mining**? | SAS new window preview

Data Mining History and Current Advances. The process of digging through **data** to disco long history.

https://www.sas.com/en_us/insights/analytics/**data-mining**.html - - Yippy Index V

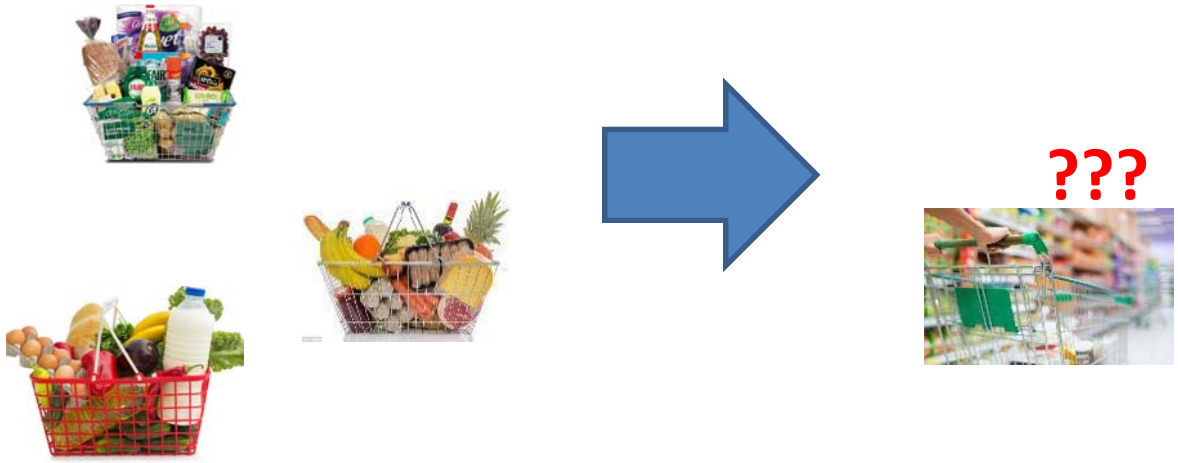
Data Mining Definition | Investopedia new window preview

Data mining is a process used by companies to turn raw **data** into useful information. By u **data**, businesses can ...

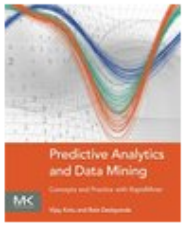
www.investopedia.com/terms/d/datamining.asp - - Yippy Index V

What for?

- **Association** - co-occurrence
  - cross-selling product
  - recommend books (Amazon.co.uk,...)
  - music playlists



Frequently Bought Together



+

Total price: £54.48

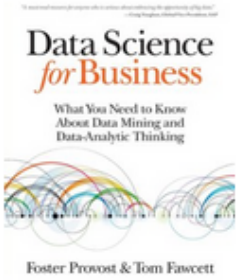
Add both to Basket

☑ **This item:** Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner by Vijay Kotu Paperback £30.49

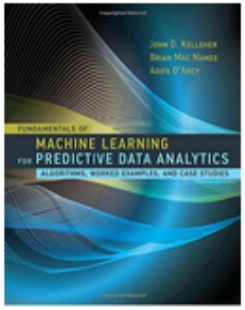
☑ Exploring Data with RapidMiner by Andrew Chisholm Paperback £23.99

Customers Who Bought This Item Also Bought

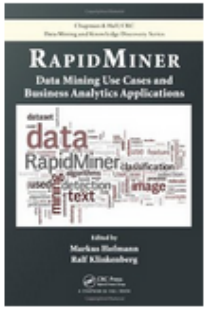
Exploring Data with RapidMiner
Andrew Chisholm
Paperback
£23.99 Prime



Data Science for Business: What you need to know about data mining and...
› Foster Provost
4.4 12
Paperback
£28.50 Prime



Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms,...
› John D. Kelleher
5.0 1
Hardcover
£59.95 Prime

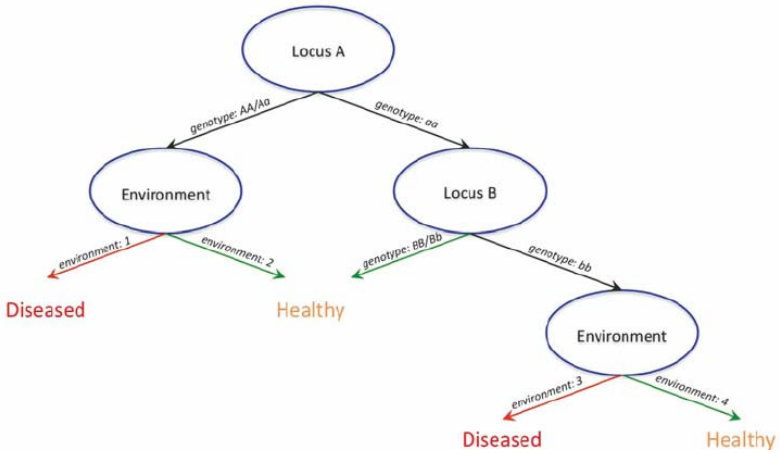


RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC...
Markus Hofmann
Hardcover
£61.99 Prime

Data Smart: Using Data Science to Transform Information into Insight
› John W. Foreman
4.4 15
Paperback
£18.79 Prime

What for?

- **Classification** - assign to a know set of categories
  - classify customers into known segments (good credit, bad credit, grey area)
  - classify customers as churners/non churners
  - diagnostic according to symptoms
  - http://news.google.pt/



https://news.google.pt/news/section?cf=all&pz=1&ned=pt-PT_pt&geo=detect_metro_area&siidp=60324cacb87eae8be32ddd7a5a

Um lembrete de privacidade da Google | Rever agora | Ler mais tarde

Google

Notícias

Edição Portugal | Moderno

- Notícias principais
- Porto, Porto
- Mundo
- Portugal
- Negócios
- Ciência
- Entretenimento
- Desporto
- Saúde

Porto, Porto

Jornal de Notícias

Cobertura tempo real

Mala suspeita cortou trânsito no centro do Porto - Jornal de Notícias

Jornal de Notícias - há 16 horas

Uma mala suspeita foi detetada, este domingo à noite, na dependência da Caixa Geral de Depósitos de

Mala suspeita lançou alerta no centro histórico do Porto Correio da Manhã

Correio da ...

Jornal Eco...

Universidade do Porto "muda-se" para Palácio de Cristal

Jornal Económico (liberação de imprensa) - há 8 horas

As faculdades e laboratórios desta universidade vão promover experiências científicas para toda a família dura
A mostra da Universidade do Porto regressa ao Palácio de Cristal para dar a ...

Correio da ...

Ryanair pede a passageiros do Porto que cheguem três horas antes do voo

Correio da Manhã - há 19 horas

A companhia aérea Ryanair está este domingo a avisar os clientes que vão viajar, a partir do Aeroporto do Po
horário previsto do voo, por causa da greve parcial de seguranças. "Fomos avisados ...

Correio da ...

Mala suspeita lança alerta no centro histórico do Porto

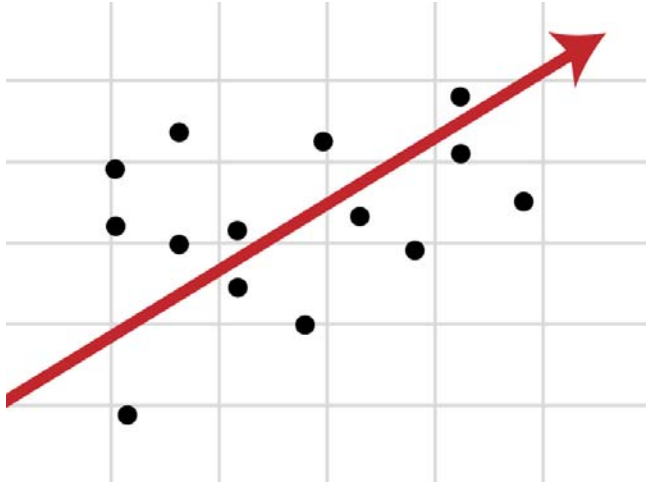
Correio da Manhã - há 15 horas

Uma mala de viagem suspeita foi encontrada na noite deste domingo no interior dependência bancária da Cai
Porto. A PSP montou um perímetro de segurança e cortou o trânsito entre os ...

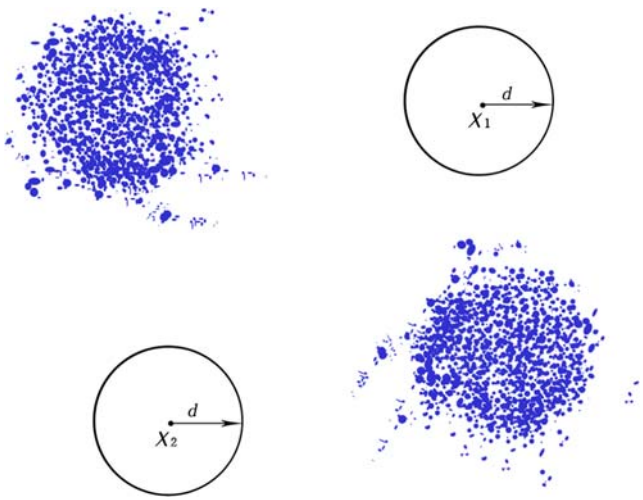
Bombeiros do Porto são os mais acionados pelo INEM no norte

What for?

- **Regression/Forecasting**
  - surgery duration
  - travel time
  - retail sales



- **Outlier/Fraud**
  - credit card
  - prescriptions
  - telecom
  - spam



Analytics and politics

https://www.publico.pt/2016/12/11/politica/noticia/a-politica-na-era-dos-robos-1754224

Foi aqui que a campanha de Donald Trump fez a diferença, "na sua capacidade de construir robôs, para que a eficácia das mensagens fosse potenciada, constituindo nichos específicos de eleitores, que recebem a comunicação certa".

Estes robôs, explica Maurício, são algoritmos, programas informáticos que varrem milhares de milhões de *posts*, comentários e conversas nas redes sociais, para definir grupos e respectivas características, e bombardeá-los com as mensagens adequadas. O robô que Maurício usa é um servidor localizado em Hong Kong que pode recolher informação maciça sobre eleitores americanos, ou sobre adeptos do Benfica em Lisboa, e, entre estes, os que estão descontentes com a arbitragem do último jogo. Ou ainda sobre os portugueses que se queixam das obras na capital, dividindo-os entre os que acreditam e os que não acreditam que a sua situação vai melhorar depois das obras.

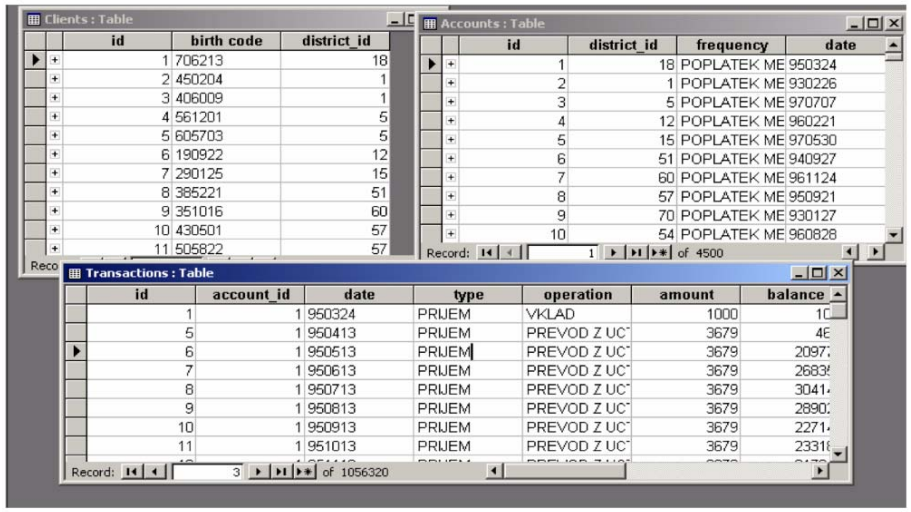
https://motherboard.vice.com/en_us/article/how-our-likes-helped-trump-win

Types of Data

Bought?	Age	Income	Family size	Cars bought previously	Value of last purchase
nao	37	49000	2	1	42000
sim	43	68000	3	0	0
sim	42	61000	4	0	0
sim	26	52000	2	0	0
sim	40	64000	1	1	21000
sim	38	52000	1	0	0
sim	45	43000	4	1	47000
sim	35	45000	2	1	34000
nao	39	43000	2	0	0
sim	31	55000	3	1	46000
sim	34	57000	3	1	52000
nao	38	44000	4	0	0
nao	34	68000	2	1	33000
sim	30	45000	2	1	44000
sim	38	41000	3	1	47000
sim	40	62000	3	0	0
sim	43	69000	2	0	0
nao	26	45000	3	0	0
sim	35	66000	4	1	17000
...	...	...	...	...	...

Types of data

Relational data



Clients : Table

id	birth code	district_id
1	706213	18
2	450204	1
3	406009	1
4	561201	5
5	605703	5
6	190922	12
7	290125	15
8	385221	51
9	351016	60
10	430501	57
11	505822	57

Accounts : Table

id	district_id	frequency	date
1	18	POPLATEK ME	950324
2	1	POPLATEK ME	930226
3	5	POPLATEK ME	970707
4	12	POPLATEK ME	960221
5	15	POPLATEK ME	970530
6	51	POPLATEK ME	940927
7	60	POPLATEK ME	961124
8	57	POPLATEK ME	950921
9	70	POPLATEK ME	930127
10	54	POPLATEK ME	960828

Record: 1 of 4500

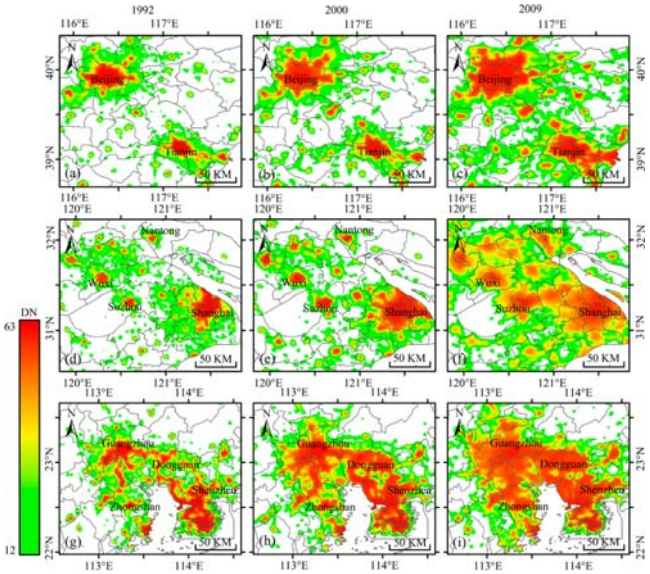
Transactions : Table

id	account_id	date	type	operation	amount	balance
1	1	950324	PRIJEM	VKLAD	1000	1
5	1	950413	PRIJEM	PREVOD Z UC	3679	4
6	1	950513	PRIJEM	PREVOD Z UC	3679	2097
7	1	950613	PRIJEM	PREVOD Z UC	3679	2683
8	1	950713	PRIJEM	PREVOD Z UC	3679	3041
9	1	950813	PRIJEM	PREVOD Z UC	3679	2890
10	1	950913	PRIJEM	PREVOD Z UC	3679	2271
11	1	951013	PRIJEM	PREVOD Z UC	3679	2331

Record: 3 of 1056320

Text Documents

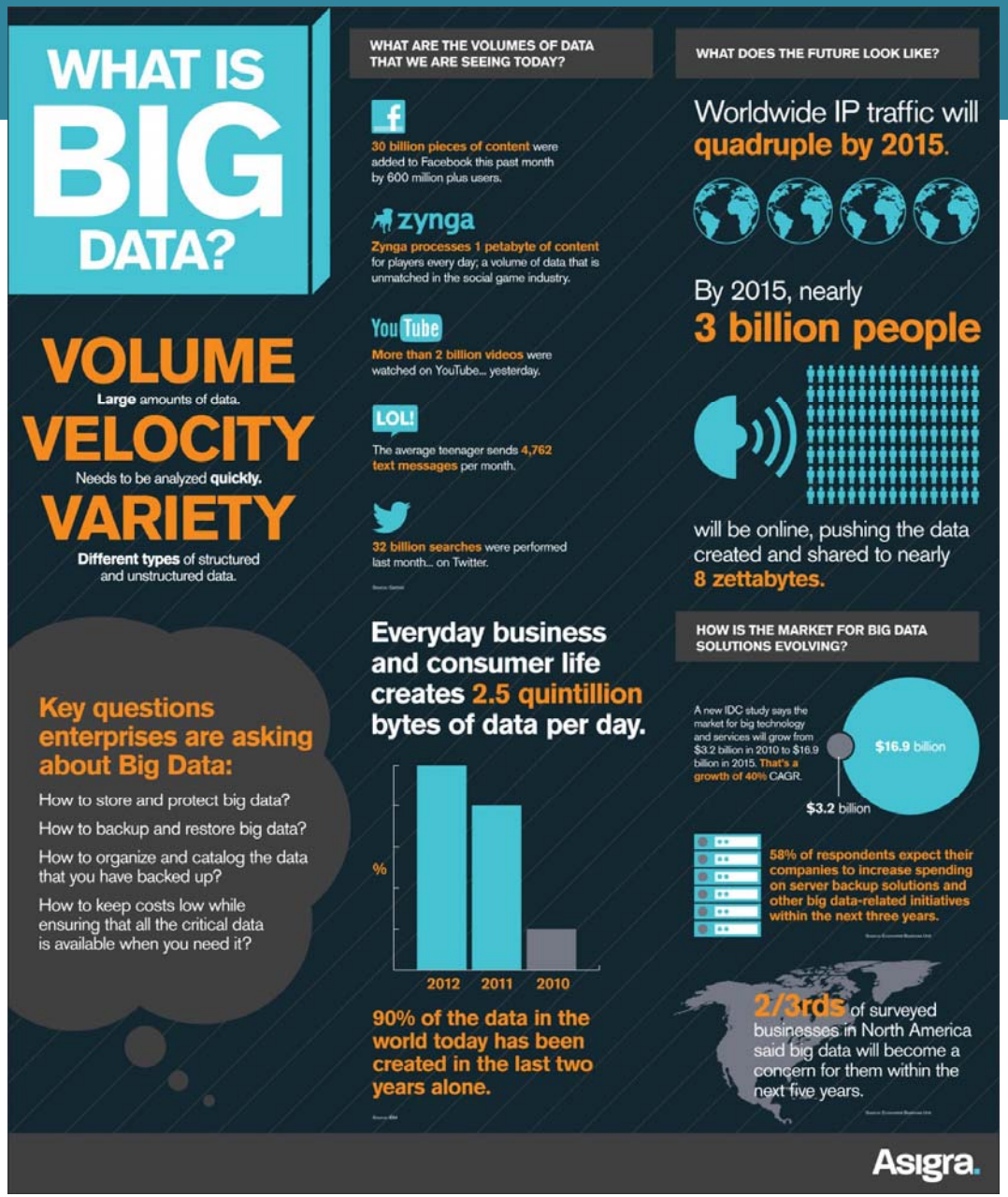
Spatial temporal



Graphs

images
video
streaming
...

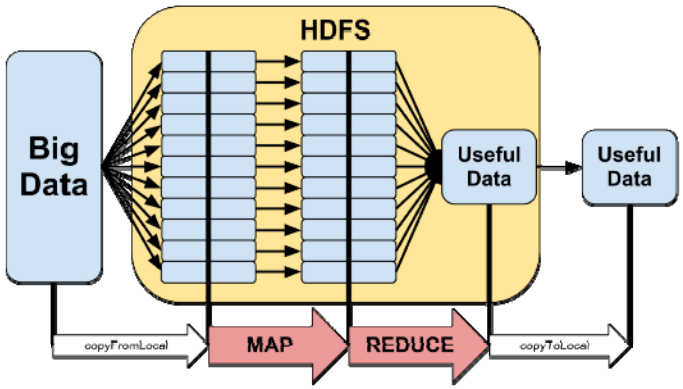
Challenges



http://visual.ly/what-big-data

Volume

- Some companies are generating huge amounts of data
- Traditional technologies
  - *(just)* read 1 TB of data
    - 90 MB/s -> 3.23 hours
    - 350 MB/s -> 50 minutes



Map Reduce can **sort** 1000 TB of data in 33 minutes using 8000 machines

http://www.glennklockwood.com/data-intensive/hadoop/overview.html

Be aware that more data does not always mean better models

- The 1936 election: the literary digest poll
- Candidates: Democrat FD Roosevelt and Republican Alfred Landon
- Sample Size: 2.3 million people
- Prediction: Landon to win with 57% of the vote
- Outcome: Landon lost with only 38% of the vote
- Literary Digest went bankrupt soon after

SAMPLING

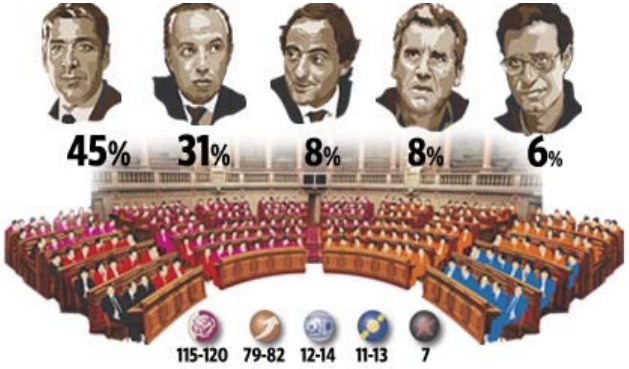
Legislativas 2002

DN e TSF -> Marktest

819 entrevistas e apresenta um erro de amostragem para um intervalo de confiança de 95 por cento, de mais ou menos 3,42 por cento.

EXPRESSO-SIC-Renascença -> Eurosondagem

2057 entrevistas validadas e apresenta um erro de amostra de 2,16 por cento para um grau de probabilidade de 95 por cento.

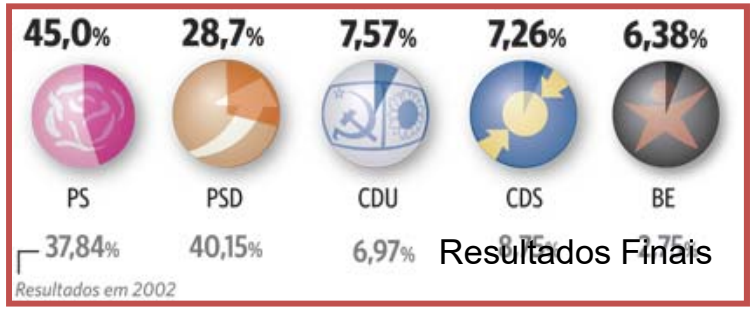


Independente -> Instituto de Pesquisa de Opinião e Mercado (IPOM)

997 entrevistas validadas e apresenta um erro de amostragem, para um nível de confiança de 95,5 por cento, de mais ou menos 3,1 pontos percentuais.

JN -> Intercampus

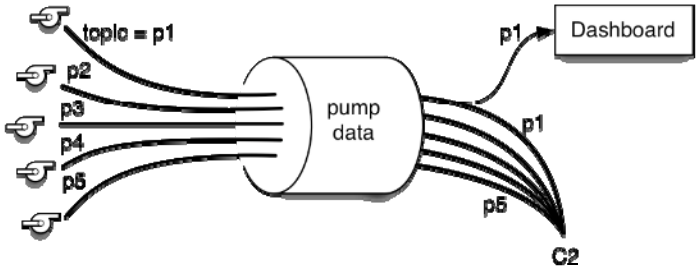
1015 entrevistas, e apresenta um erro de amostragem, para um intervalo de confiança de 95 por cento, de mais ou menos 3,1 por cento.



Resultados Finais

Velocity

- Data arriving at high velocity
- Needs to be analyzed quickly
  - Traditional algorithms analyze each example multiple times
- Streaming analytics
  - New algorithms process each example once and store sufficient statistics



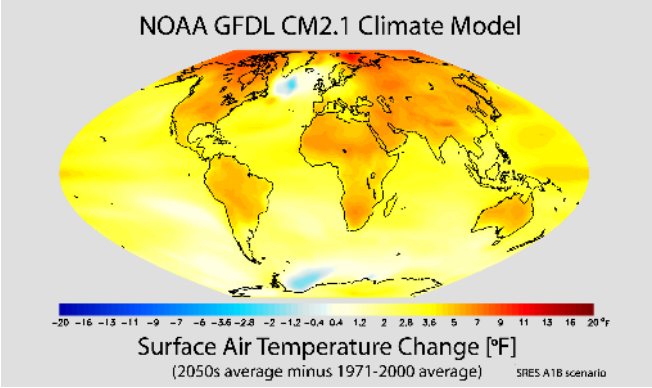
Variety

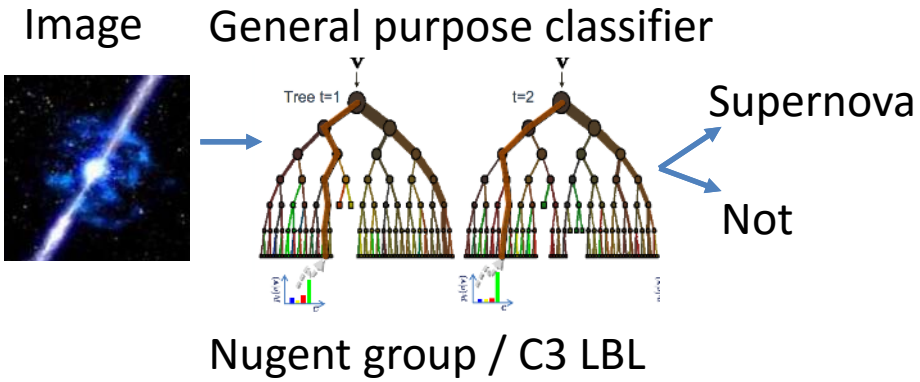
- Different types of structured and unstructured data
- New methods to deal with new data types

	Databases	Data Science
Data Value	“Precious”	“Cheap”
Data Volume	Modest	Massive
Examples	Bank records, Personnel records, Census, Medical records	Online clicks, GPS logs, Tweets, Building sensor readings
Priorities	Consistency, Error recovery, Auditability	Speed, Availability, Query richness
Structured	Strongly (Schema)	Weakly or none (Text)
Properties	Transactions, ACID*	CAP* theorem (2/3), eventual consistency
Realizations	SQL	NoSQL: Riak, Memcached, Apache River, MongoDB, CouchDB, Hbase, Cassandra,…

ACID = Atomicity, Consistency, Isolation and Durability CAP = Consistency, Availability, Partition Tolerance

https://bcourses.berkeley.edu/courses/1377158/pages/cs-194-16-introduction-to-data-science-fall-2015



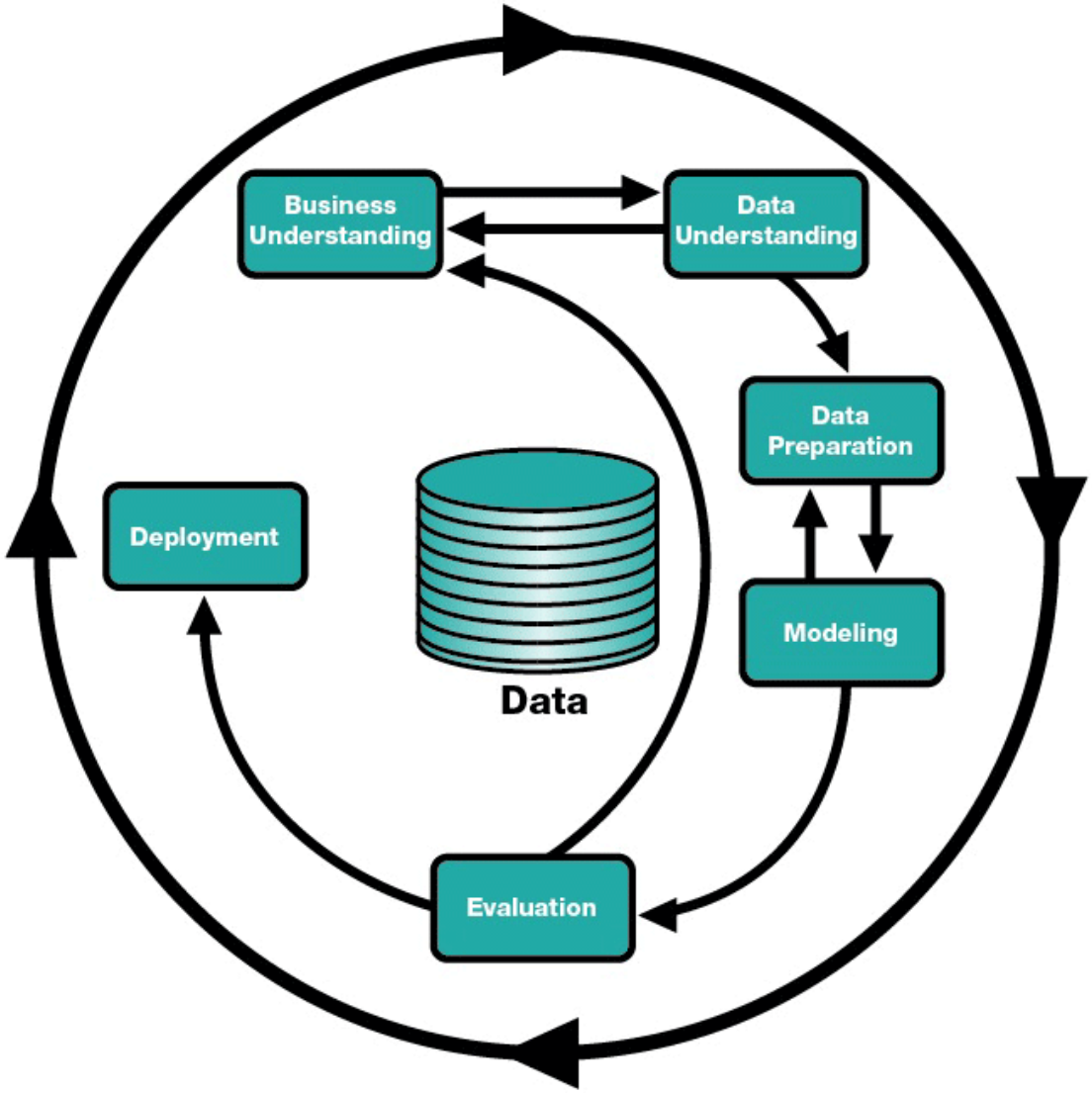


Scientific Modeling
Physics-based models
Problem-Structured
Mostly deterministic, precise
Run on Supercomputer or High-end Computing Cluster

Data-Driven Approach
General inference engine replaces model
Structure not related to problem
Statistical models handle true randomness, and **unmodeled complexity**.
Run on cheaper computer Clusters (EC2)

https://bcourses.berkeley.edu/courses/1377158/pages/cs-194-16-introduction-to-data-science-fall-2015

Methodology CRISP-DM



DM Methodologies

- Framework for recording experience
  - Allows projects to be replicated
- Aid to project planning and management
  - “Comfort factor” for new adopters
- Demonstrates maturity of Data Mining
  - Reduces dependency on “stars”
- Encourage best practices and help to obtain better results

CRISP tasks

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives *Background* *Business Objectives* *Business Success Criteria* **Assess Situation** *Inventory of Resources* *Requirements, Assumptions, and Constraints* *Risks and Contingencies* *Terminology* *Costs and Benefits* **Determine Data Mining Goals** *Data Mining Goals* *Data Mining Success Criteria* **Produce Project Plan** *Project Plan* *Initial Assessment of Tools and Techniques*	**Collect Initial Data** *Initial Data Collection Report* **Describe Data** *Data Description Report* **Explore Data** *Data Exploration Report* **Verify Data Quality** *Data Quality Report*	**Select Data** *Rationale for Inclusion/ Exclusion* **Clean Data** *Data Cleaning Report* **Construct Data** *Derived Attributes* *Generated Records* **Integrate Data** *Merged Data* **Format Data** *Reformatted Data* *Dataset* *Dataset Description*	**Select Modeling Techniques** *Modeling Technique* *Modeling Assumptions* **Generate Test Design** *Test Design* **Build Model** *Parameter Settings* *Models* *Model Descriptions* **Assess Model** *Model Assessment* *Revised Parameter Settings*	**Evaluate Results** *Assessment of Data Mining Results w.r.t. Business Success Criteria* *Approved Models* **Review Process** *Review of Process* **Determine Next Steps** *List of Possible Actions* *Decision*	**Plan Deployment** *Deployment Plan* **Plan Monitoring and Maintenance** *Monitoring and Maintenance Plan* **Produce Final Report** *Final Report* *Final Presentation* **Review Project** *Experience Documentation*

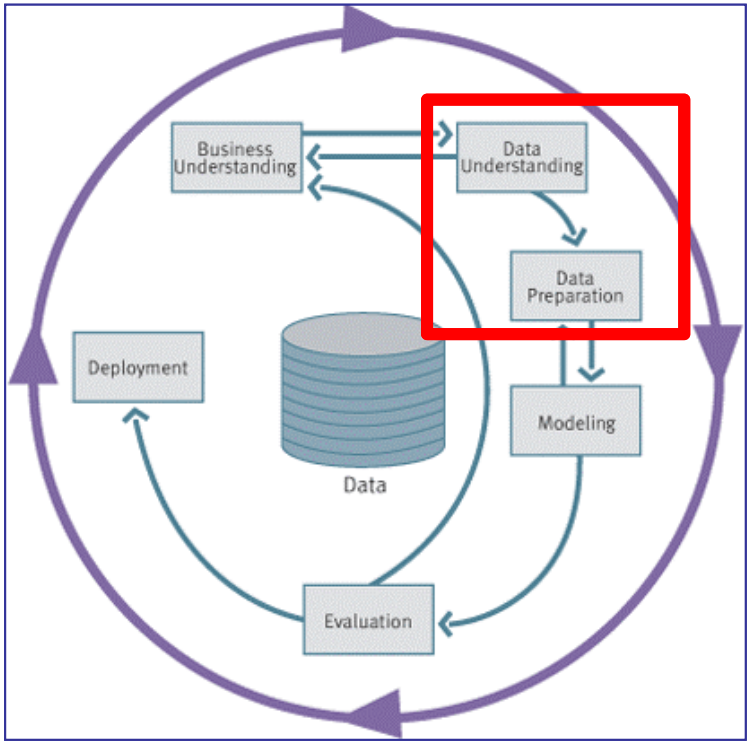
Figure 3: Generic tasks (bold) and outputs (italic) of the CRISP-DM reference model

Keep in mind that

- A data mining project should always start with an analysis of the data with traditional query tools
- 80% of the interesting information can be extracted using SQL
  - how many transactions per month include item number 15?
  - show me all the items purchased by Sandy Smith
- 20% of hidden information requires more advanced techniques
  - which items are frequently purchased together by my customers?
  - how should I classify my customers in order to decide whether future loan applicants will be given a loan or not?

Be aware that

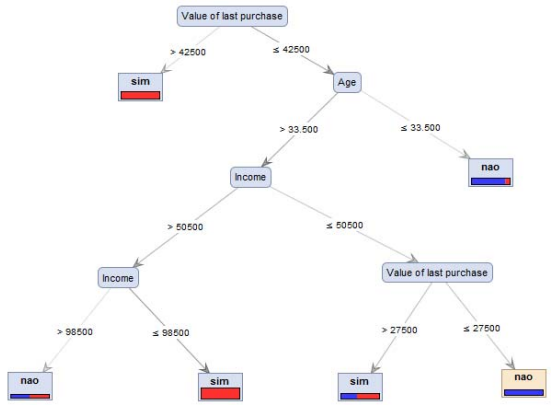
- 60 to 80% of the effort in a Data Mining project is about preparing the data and the remaining 20% is about mining



How NOT to estimate performance

1. Induce model

Bought?	Age	Income	Family size	Cars bought previously	Value of last purchase
nao	37	49000	2	1	42000
sim	43	68000	3	0	0
sim	42	61000	4	0	0
sim	26	52000	2	0	0
sim	40	64000	1	1	21000
sim	38	52000	1	0	0
sim	45	43000	4	1	47000
sim	35	45000	2	1	34000
nao	39	48000	2	0	0
sim	31	55000	3	1	46000
sim	34	57000	3	1	52000
nao	38	44000	4	0	0
nao	34	68000	2	1	33000
sim	30	45000	2	1	44000
sim	38	41000	3	1	47000
sim	40	62000	3	0	0
sim	42	68000	2	0	0
nao	26	45000	3	0	0
sim	35	66000	4	1	17000
...	...	...	...	...	...



2. Evaluate prediction

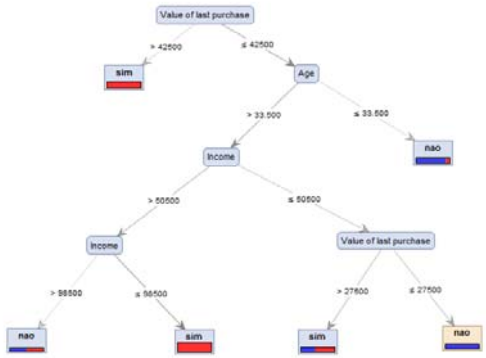
How to estimate performance

training set

Bought?	Age	Income	Family size	Cars bought previously	Value of last purchase
nao	37	49000	2	1	42000
sim	43	68000	3	0	0
sim	42	60000	4	0	0
sim	26	52000	2	0	0
sim	48	64000	1	1	21000
sim	50	52000	1	0	0
sim	45	49000	4	1	47000
sim	35	45000	2	1	34000
nao	38	48000	2	0	0
sim	51	38000	3	1	48000
sim	34	57000	3	1	54000
nao	39	44000	4	0	0
nao	34	58000	2	1	28000
sim	50	45000	2	1	44000
sim	38	43000	3	1	47000
sim	40	52000	5	0	0
sim	45	67000	2	0	0
nao	36	46000	3	0	0
sim	35	60000	4	1	17000
...	...	...	...	...	...

1. randomly split

2. Induce model



test set

Bought?	Age	Income	Family size	Cars bought previously	Value of last purchase
nao	37	49000	2	1	42000
sim	43	68000	3	0	0
sim	42	60000	4	0	0
sim	26	52000	2	0	0
sim	48	64000	1	1	21000
sim	50	52000	1	0	0
sim	45	49000	4	1	47000
sim	35	45000	2	1	34000

3. Evaluate prediction

Privacy and ethics

- Spotify asks for your photos, location and contacts in privacy code revamp
  http://www.telegraph.co.uk/technology/news/11815778/spotify-privacy-policy-asks-for-photos-and-contacts.html

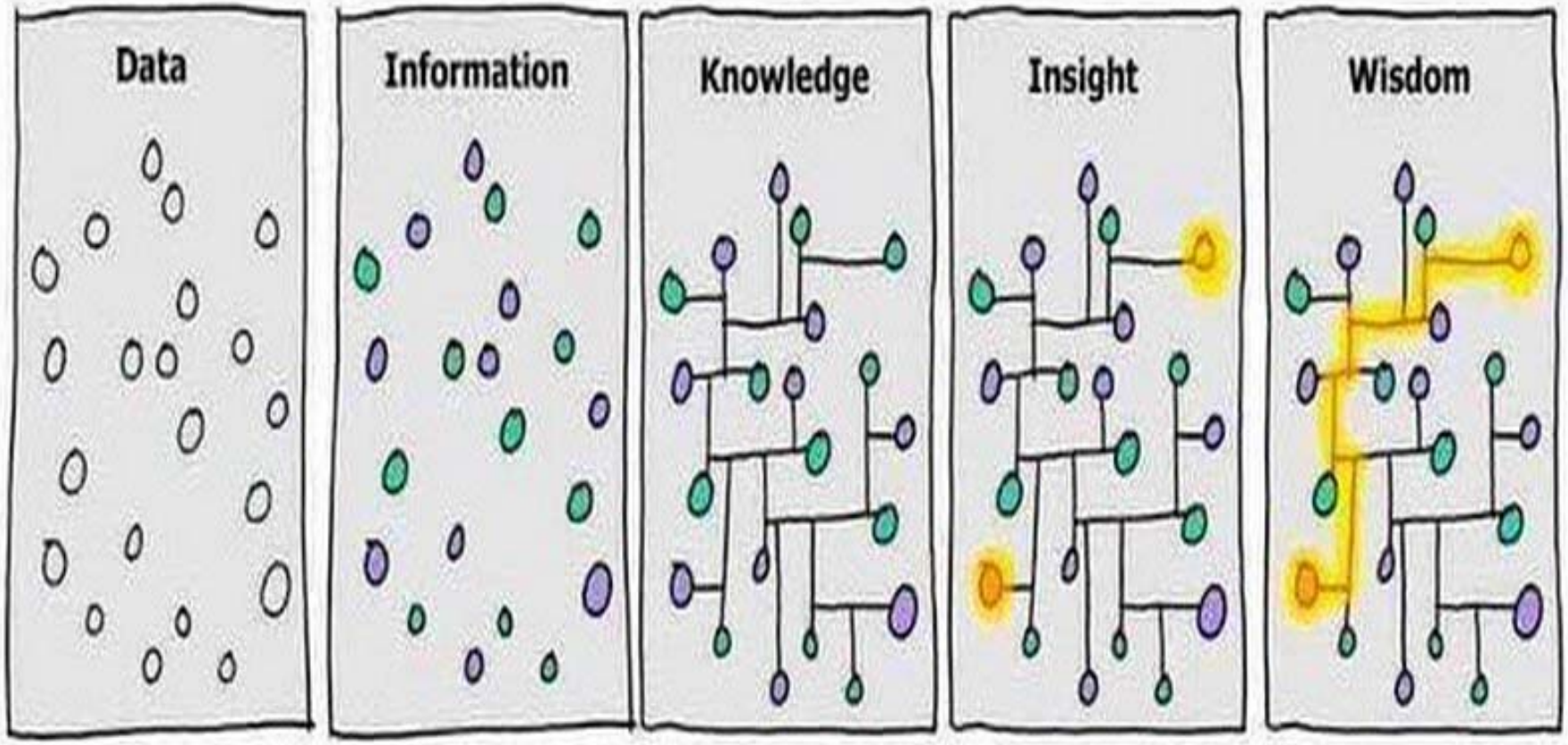
- What google knows about you...

- *"These databases will grow to connect every individual to at least one closely guarded secret. This might be a secret about a medical condition, family history, or personal preference...."*
  https://hbr.org/2012/08/dont-build-a-database-of-ruin

- How Companies Learn Your Secrets
  "If we wanted to figure out if a customer is pregnant, even if she didn't want us to know, can you do that? "
  http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=1

Data-Analytic Thinking

- Data-Driven Decision Making
  - the practice of basing decisions on the analysis of data, rather than on intuition
- When facing a business problem
  - ability to assess whether and how data can improve performance
- Understanding the fundamental concepts
  - will help to envision opportunities for improving data-driven decision-making


Data
Information
Knowledge
Insight
Wisdom

Application examples

- Retail
  - Target retailer used DS methods to predict that customers were expecting a baby
  - Retailers' coupons targeting
    and human resource management to anticipate impact of promotions
- Spotify sells data about customers listening habits
  - Bands' tours
- Online adds
  - Google

Applications

- Contact centers
  - Customer profiling
  - Match between operator and costumer needs
- Recommender Systems
  - Long tail - help users find what they want
- Churn
  - Will the customer leave?
- Manufacturing

Questions?