Business Intelligence & Process Modelling

Frank Takes

Universiteit Leiden

Lecture 1 — Introduction
BIPM (in short)

- Business Intelligence

BI
Business Intelligence

Converting raw company data into information that contributes to (knowledge on) business development
BIPM (in short)

- **Business Intelligence (BI)**
  Converting raw company data into information that contributes to (knowledge on) business development

- **Process Modelling (PM)**
BIPM (in short)

- **Business Intelligence**
  Converting raw company data into information that contributes to (knowledge on) business development

- **Process Modelling**
  Methods and techniques to systematically map and analyze business processes
Course Introduction (Dutch)
Vakinformatie

- Voertaal: Nederlands (materiaal: grotendeels Engels)
- 9 feb 2018 t/m 2 jun 2018 (niet op 16 en 30 mrt, 27 apr en 11 mei)
- Hoorcollege: vrijdag van 11.00 tot 12.45 in 412
- Werkcollege: vrijdag van 13:30 tot 15:15 in 302/304
- Vakinformatie: http://liacs.leidenuniv.nl/~takesfw/BIPM
- Docent: dr. Frank Takes
  - f.w.takes@liacs.leidenuniv.nl
  - Snellius kamer 157b, aanwezig op:
    (dinsdag of woensdag), donderdag en vrijdag
- Assistenten:
  - Gerrit-Jan de Bruin MSc, kamer 150,
    g.j.de.bruin@liacs.leidenuniv.nl
  - Jasper van Nijhuis BSc, j.c.van.nijhuis@umail.leidenuniv.nl
Format

- Materiaal: boek, slides en wellicht wetenschappelijke artikelen
- Schriftelijk tentamen (8 juni 2018, 14.00–17.00) 60%
- Practicum met deliverables (7 mrt, 19 apr, 31 mei) 3 × 13,3%
- Elk onderdeel moet voldoende (> 5.5) zijn.
- Eindcijfer afgerond op basis van participatie naar dichtsbijzijnde element in {1, 2, 3, 4, 5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10}
- Hertentamen: 5 juli 2018
- Practicum herkansingsdeadline: 30 juni 2018
- 6 ECTS

These slides are partially based on the slides of the (previous edition of the) book.
Disclaimer

- New course format since 2015, larger adjustments in 2018
- Definitions and abbreviations . . .
- Pictures and schemas . . .
- Participation
- Feedback always welcome!
Topics

- February: Business Intelligence
- March: Business Intelligence 2.0
- April/May: Process Modelling
Assignments

1 Business Intelligence: BI in a gaming industry context
2 Business Intelligence 2.0: BI meets data science
3 Process Modelling: BPMN and Petri net tools
Introduction to Business Intelligence
Business Intelligence

- **Business Intelligence**: techniques and tools for the transformation of raw data into meaningful and useful information for business analysis purposes (Wikipedia.org)

- **Business Intelligence**: a data analysis process aimed at boosting business performance by helping corporate executives and other end users make more informed decisions (TechTarget.com)

- **Business Intelligence**: anything that aims at providing actionable information that can be used to support decision making (v.d. Aalst)
Business Analysis

- **Business analysis**: identifying business needs and determining solutions to business problems
  - Business architecture analysis
  - Data warehouse analysis
  - Business process analysis
- Role of IT: Extraction, Transformation and Loading (ETL)
- Identifying changes that are required to achieve strategic goals
- Business Analyst
Business Analytics

- **Business Analytics**: skills, technologies and practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning
  - **Descriptive analytics**: gain insight from historical data
  - **Predictive analytics**: predictive modeling using statistical and machine learning techniques
  - **Prescriptive analytics**: recommend decisions using optimization, simulation, etc.

- Domains: marketing, pricing, sales, accounting, supply chains, communication, ...
Data

- **Data**: facts, measurements or text collected for reference or analysis (Oxford dictionary)
  - **Unstructured data**: data that does not fit a certain data structure (text, a list of numeric measurements)
  - **Structured data**: data that fits a certain data structure (table, tree, graph/network, etc.)
Moore’s Law & Transistors

WHAT IS MOORE’S LAW?
First postulated in a 1965 paper by Intel co-founder Gordon E. Moore, Moore’s Law states that the transistor count in integrated circuits is roughly every 2 years.

Although Moore’s Law is only an observation rather than an actual law, it’s held approximately true ever since.

THE LAW describes an exponential growth in the number of transistors due to improved design and manufacturing methods. An original Intel 8008 processor from 1974 contained around 4,500 transistors. In 1993 the first Pentium contained 3.1 million transistors, and today a high-end PC processor might have around 3 billion transistors.

Moore’s Law & The Future
Moore himself acknowledged that manufacturers would eventually reach the atomic level after which transistors as we understand them can’t get any smaller.

TRANSISTORS, KEY PROCESSORS, YEARS

<table>
<thead>
<tr>
<th>Year</th>
<th>Transistors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978</td>
<td>29,000</td>
</tr>
<tr>
<td>1982</td>
<td>134,000</td>
</tr>
<tr>
<td>1985</td>
<td>275,000</td>
</tr>
<tr>
<td>1989</td>
<td>1.2 million</td>
</tr>
<tr>
<td>1993</td>
<td>3.1 million</td>
</tr>
<tr>
<td>1997</td>
<td>7.5 million</td>
</tr>
<tr>
<td>2000</td>
<td>42 million</td>
</tr>
<tr>
<td>2006</td>
<td>291 million</td>
</tr>
<tr>
<td>2008</td>
<td>731 million</td>
</tr>
<tr>
<td>2013</td>
<td>2.3 billion</td>
</tr>
</tbody>
</table>

WHAT DO WE EXPECT OF THE FUTURE?

2015...

http://visual.ly
Moore’s Law & CPU’s

Transistors vs Year graph showing the exponential growth of transistors over time for different CPU generations such as 8008, 8085, 8086, 80286, 80386, and more advanced models like Sandy Bridge, Nehalem, Ivy, G6, G5, G4, G3, Pentium M, Pentium 4, Pentium III, Pentium Pro, Pentium MMX, and Pentium.
<table>
<thead>
<tr>
<th>Unit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Megabyte</td>
<td>1,000,000 bytes</td>
</tr>
<tr>
<td>Gigabyte</td>
<td>1,000,000,000 bytes</td>
</tr>
<tr>
<td>Terabyte</td>
<td>1,000,000,000,000 bytes</td>
</tr>
<tr>
<td>Petabyte</td>
<td>1,000,000,000,000,000 bytes</td>
</tr>
<tr>
<td>Exabyte</td>
<td>1,000,000,000,000,000,000 bytes</td>
</tr>
<tr>
<td>Zettabyte</td>
<td>1,000,000,000,000,000,000,000 bytes</td>
</tr>
<tr>
<td>Yottabyte</td>
<td>1,000,000,000,000,000,000,000,000 bytes</td>
</tr>
</tbody>
</table>
Storage: Digital vs. Analog

The World’s Technological Capacity to Store, Communicate, and Compute Information by Martin Hilbert and Priscilla López (DOI 10.1126/science.1200970)

The world’s storage capacity has shifted from analog to digital over time. In 1986, the storage capacity was primarily analog, with 2.62 billion gigabytes. By 2000, the storage capacity had increased significantly, and by 2007, it had reached 18.86 billion gigabytes, with 93.8% digital and 6.2% analog.

**Computing Power**

In 1986, pocket calculators accounted for much of the world’s data-processing power. The percentage of available processing power by device was as follows:

- **Pocket calculators**: 41%
- **Personal computers**: 33%
- **Video game consoles**: 9%
- **Servers, mainframes**: 17%

By 2007:
- **Pocket calculators**: 66%
- **Personal computers**: 25%
- **Video game consoles**: 3%
- **Servers, mainframes**: 6%

Digital storage formats include:

- **CDs and minidisks**: 6.8%
- **Computer servers and mainframe hard disks**: 8.9%
- **Digital tape**: 11.8%
- **Paper, film, audiotape and vinyl**: 6.2%
- **Analog videotapes**: 93.8%
- **Other digital media**: 0.8%
- **Portable media players, flash drives**: 2.4%
- **Portable hard disks**: 2.4%
- **DVD/Blu-ray**: 22.8%
- **PC hard disks**: 44.5%

*Other includes chip cards, memory cards, floppy disks, mobile phones/PDAs, cameras/camcorders, video games*
Moore’s Law & Data

Figure: Zettabytes produced per year

Source: http://www1.unece.org/stat/platform/display/msis/Big+Data
Big Data

Volume
(data size)

Velocity
(speed of change)

Variety
(different types/sources of data)

Veracity
(uncertainty of data)
Big Data: The four Vs
Volume, Velocity, Variety and Value

- **Volume**: Large amounts of data
- **Velocity**: Need to be analysed quickly
- **Variety**: Different types of unstructured and structured data
- **Value**: Extracting business insights and revenue from data

© World Newsmedia Network 2013
Data Mining

- Data explosion
- “We are drowning in data, but starving for knowledge!”
- Interpretation
- Understanding
- Learning
- Acting
Data Mining

- Data $\rightarrow$ Information $\rightarrow$ Knowledge
- Knowledge Discovery
- Machine Learning
- Data Mining
  - Descriptive data mining: clustering, pattern mining, etc.
  - Predictive data mining: classification, prediction, etc.
- Big Data
- Data Science
The DIKW Pyramid illustrates the hierarchy of data, information, knowledge, and wisdom from raw data at the base to wisdom at the top. The diagram shows:

- **Data**
  - Raw: Red, 192.234.235.245.678, v2.0

- **Information**
  - Meaning: South facing traffic light on corner of Pitt and George Streets has turned red

- **Knowledge**
  - Context: The traffic light I am driving towards has turned red

- **Wisdom**
  - Applied: I better stop the car!
DIKWD Pyramid (2)

- **DATA**
  - signals, know-nothing
- **INFORMATION**
  - useful, organized, structured
- **KNOWLEDGE**
  - contextual, synthesized, learning
- **WISDOM**
  - understanding, integrated, actionable
- **DECISIONS**
  - change, movement

**FUTURE**
- WHAT ACTION?
  - reveals direction
- WHAT IS BEST?
  - reveals principles

**PAST**
- WHY?
  - reveals patterns
- WHAT?
  - reveals relationships
Data Science

- Computer Science
- Maths & Statistics
  - Machine Learning
  - Data Science
  - Data Analysis
- Business / Domain Expertise
- Traditional Software
Data Science

- Data mining
- Machine learning
- Predictive analytics
- Databases
- Algorithms
- Statistics
- Privacy, security, law & ethics
- Behavioral/social science
- Business models & marketing
- Visualization & visual analytics
- Distributed systems
Break?
Introduction to Process Modelling
Processes

- **Process**: a set of related actions and transactions to achieve a certain objective
  - Service process: providing a value transfer
  - Production process: providing a value creation

- Various domains: organizations, environments, biological systems, economies, ...
Business Processes

- **Business process**: a sequence of activities aimed at producing something of value for the business (Morgan)

- **Business process**: collection of related, structured activities or tasks that produce a specific service or product (serve a particular goal) for a particular customer or customers (Wikipedia)

- **Business Process Management**: the discipline that combines knowledge from information technology and knowledge from management sciences and applies this to operational business processes (v.d. Aalst)
Process Modelling

"The Real Thing"

The Model

The Reason

Representa-tion

Viewpoint

At a point in time

Modeler

Analyze, design, and optimize business processes.

MANAGERS, BUSINESS PEOPLE

Derive requirements on IT landscape and software.

IT EXPERTS

Demonstrate compliance (e.g. SOX) or get certification (e.g. ISO).

AUDITORS
Types of Business Processes

- **Management processes**: the processes that govern the operation of a system, e.g., corporate governance and strategic management.

- **Operational processes**: processes that constitute core business and create the primary value stream, e.g., purchasing, manufacturing, advertising, marketing and sales.

- **Supporting processes**: processes that support the core business, e.g., accounting, recruitment, HR, customer support and ICT support.
Business Process Model

- **Model**: abstract or schematic representation of reality

- Modelling (task): creating a model, for example based on real observations

- **Business Process Model**: abstract representation of business processes

- Process model functions:
  - Descriptive: what is actually happening?
  - Prescriptive: what should be happening?
  - Explanatory: why is the process designed this way?
Classical BPM Lifecycle

adjustment

enactment/monitoring

data

models

(re)design

diagnosis/requirements

performance analysis

insight
discussion
animation

verification
specification

configuration

configuration

documentation
Process Model Goals

- **Insight**: while making a model, the modeler is triggered to view the process from various angles
- **Discussion**: the stakeholders use models to structure discussions
- **Documentation**: processes are documented for instructing people or certification purposes
- **Verification**: process models are analyzed to find errors in systems or procedures (e.g., potential deadlocks)
Process Model Goals (2)

- **Performance analysis**: techniques like simulation can be used to understand the factors influencing response times, service levels, etc.

- **Animation**: models enable end users to play out different scenarios and provide feedback to the designer

- **Specification**: models can be used to describe a Process-Aware Information System (PAIS) before it is implemented and can hence serve as a “contract” between the developer and the end user/management

- **Configuration**: models can be used to configure a system
Process Modelling

- In practice: **formalize** and **visualize** business processes
- Business Process Languages (BPLs):
  - Petri Nets
  - Business Process Model Notation (BPMN)
- Process discovery: derive the process from a description of activities
  - Manual
  - Automated: **Process Mining**
Process Science

- optimization
- operations management & research
- process mining
- business process management
- business process improvement
- process automation & workflow management
- formal methods & concurrency theory
- stochastics
Process Mining

- Executable models may be used to force people to work in a particular manner
- Models are possibly not well-aligned with reality
- Hand-made models may be disconnected from reality and provide an idealized view on the processes at hand: “paper tigers”
- An abundance of event logs (data) is usually available
- **Process Mining** is the task of converting event data into process models
Process Discovery Example (Dutch)

- **Postorderbedrijf acties:**
  - A = het plaatsen van een bestelling
  - B = de betaling van deze bestelling
  - C = de levering van de bestelling
  - D = het versturen van een dankbrief
  - E = klant op de hoogte brengen bij niet-leverbare order

- **Verzamelde procespaden:**
  - Pad ABCD is 543 keer gevolgd
  - Pad ACBD is 378 keer gevolgd
  - Pad AED is 45 keer gevolgd
Petri Net for Example (Dutch)

`start`  \(\rightarrow\)  `A`  \(\rightarrow\)  `p1`  \(\rightarrow\)  `B`  \(\rightarrow\)  `E`  \(\rightarrow\)  `p3`  \(\rightarrow\)  `D`  \(\rightarrow\)  `end`

- `A`: `plaats bestelling`
- `B`: `betalen`
- `E`: `informeer`
- `D`: `verstuur dankbrief`
- `C`: `levering`
Process Mining

- **Data-oriented** business information analysis
- Model-based process analysis
- Event logs can be replayed on the model
- Compliance, auditing, internal review, etc.
Process Mining

“world”
people
machines
components
organizations

support/
controls

software
system

models
analyzes

specifies
configures
implements
analyzes

records
events, e.g.,
messages,
transactions,
etc.

(process)
model
discovery
conformance
enhancement

event
logs
Process Mining tasks

- **Discovery**: how can we automatically construct a model based on event logs?
- **Conformance**: given an event log and a model, what are the differences between the model and reality?
- **Enhancement**: how can the model be extended using knowledge inferred from event logs?
Process Mining

Data science

Process mining

Machine learning

Optimization

Operations management & research

Predictive analytics

Business process management

Distributed systems

Visualization & visual analytics

Business models & marketing

Behavioral/social science

Privacy, security, law & ethics

Statistics

Algorithms

Data mining
Assignment 1

- Gaming industry context
- Sales log spanning roughly 4 years of sales
- Apply and compare BI techniques
- Inspect, visualize, aggregate, . . .
- Deliverables:
  1. Web-based BI dashboard
  2. Answers to BI-related questions
- Format: short assignment report in \LaTeX
Werkcollege vandaag

- ULCN-account bruikbaar in Snellius zaal 302/304?
- Toegang tot webserver liacs.leidenuniv.nl mogelijk?
- MySQL-wachtwoord bekend?
- Verfris uw HTML/CSS/Javascript/jQuery/PHP/MySQL-kennis
- Speel een (Facebook)spel (bijv. Jelly Splash)
- Onderzoek het verdienmodel van dit spel
- Specifiek: waarvoor en op welke manieren kan men betalen?
- Hoe werken micropayments? Wat zijn payment service providers?
- Hoe kun je online betalen in Frankrijk?
Credits