Analytics in Banking: State of the Art

Prof. dr. Bart Baesens

Department of Decision Sciences and Information Management
KU Leuven (Belgium)

School of Management
University of Southampton (United Kingdom)
Bart.Baesens@kuleuven.be

YouTube/Facebook/Twitter: DataMiningApps

Presenter: Bart Baesens

- Studied at KU Leuven (Belgium)
  – Business Engineer in Management Informatics, 1998
  – PhD. in Applied Economic Sciences, 2003
- PhD. : Developing Intelligent Systems for Credit Scoring Using Machine Learning Techniques
- Professor of Big Data & Analytics at KU Leuven, Belgium
- Lecturer at the University of Southampton, UK
- Research: Big Data & Analytics, Credit Risk, Fraud, Marketing, …
- YouTube/Facebook/Twitter: DataMiningApps
- www.dataminingapps.com
- Bart.Baesens@kuleuven.be

Books
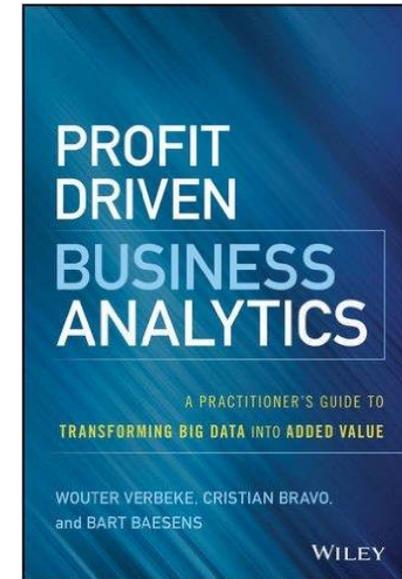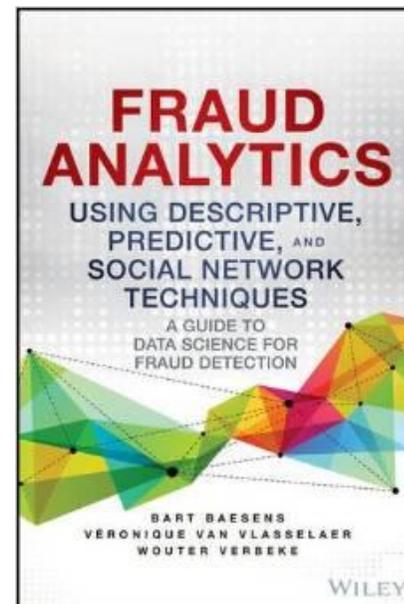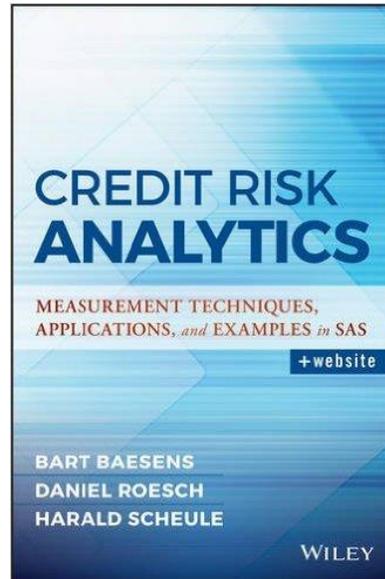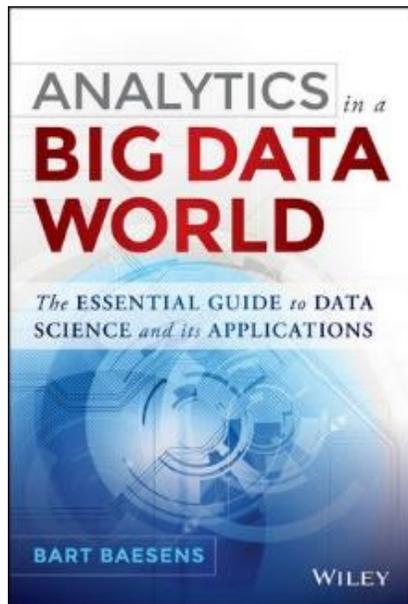
Overview

- Introduction
- NoSQL versus SQL
- Risk Analytics
- Marketing Analytics
- HR Analytics
- Organizational perspective
- In- versus Outsourcing
- Commercial versus Open-source Software

NoSQL versus SQL

- NoSQL: umbrella term
  - Key-Value stores, Tuple/Document stores, Column Oriented databases, **Graph-based databases (!)**
- ACID (Atomicity, Consistency, Isolation and Durability) versus BASE (Basically available, Soft State, and Eventual Consistency)
- Do we really have Big Data?
  - Volume, Variety, Velocity
- "Think Small, Start Big" Strategic Myopia
- Avoid meaningless data pumping
- E.g. Twitter, HealthCare.gov, …

Risk Analytics

- Key application arreas:
  - Credit Risk
  - Operational Risk
  - Market Risk
  - Insurance Risk

- Regulatory frameworks (e.g. Basel II/III, IFRS 9)

- Shift in analytical model focus:
  - Through the Cycle (TTC) versus Point In Time (PIT)
  - Acquisition Cost versus Fair Value (e.g. haircuts)

Characteristic Name	Attribute	Scorecard Points
AGE 1	Up to 26	100
AGE 2	26 - 35	120
AGE 3	35 - 37	185
AGE 4	37+	225
GENDER 1	Male	90
GENDER 2	Female	180
SALARY 1	Up to 500	120
SALARY 2	501-1000	140
SALARY 3	1001-1500	160
SALARY 4	1501-2000	200
SALARY 5	2001+	240

Risk Analytics

- Analytical models (descriptive + predictive) used for provisioning, buffering, (risk-based) pricing, loan approaval, securitization, stress testing

- Analytical model errors directly affect profitability, solvency, shareholder value, macro-economy, …, society as a whole!

- Internal validation by audit teams (Chinese wall principle)

- External validation by bank regulators (e.g. OCC, Federal Reserve, …)

- Maturity level: **HIGH!**

Marketing Analytics

- Key application areas:
  - Response modeling
  - Churn prediction
  - Customer Segmentation
  - Customer Lifetime Value modeling
  - Recommender systems
  - Customer journeys
- No regulatory frameworks
- Talk the talk, but not walk the walk!

Marketing Analytics

- Data: The 4 W problem
  - What, Where, When, Who
- Business users: education needed!
- Take into account lessons learned in risk management
  - model development, model validation, model deployment and model monitoring
- Maturity level : **MEDIUM!**

HR Analytics

- Key applications:
  - Employee absenteeism
  - Employee churn
  - Employee Lifetime Value
  - Employee Engagement
  - Collaboration patterns

- Regulatory frameworks (Unions, Privacy, …)
  - "The robots are coming to take me away, Ha-Haaa!"

- Success stories limited to descriptive analytics only

- Maturity level : **<u>LOW!</u>**

FIRING

What to Do Before You Fire a Pivotal Employee

by Bart Baesens, Sophie De Winne, and Luc Sels

JANUARY 28, 2016
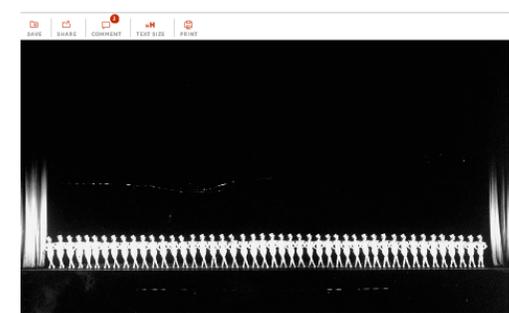
These days, the majority of a company's employees are networked together on social media sites like Facebook, LinkedIn, and Instagram. They no doubt also have important vendors, customers, and future customers among their contacts. That's why understanding, modeling, and measuring your employee network should be a key ingredient in your company's strategic HR decisions.

As noted in research by Paul Adler and Seok-Woo Kwon at the University of Southern California, a well-designed employee network essentially makes up the "social capital" of a company, due to all the assets or resources that can be mobilized through the network. Your company's decisions can have intended, or unintended, consequences that quickly ripple through the network. Let's take the case of firing, be it on an individual or collective basis. Say your analysis has determined that a person whom you want to fire is very well connected to a few key customers or highly influential employees. How do you get ahead of the situation to manage any ripple effects that may come from the firing?

Organizational Perspective

- Centralized versus De-Centralized
- Need for a corporate wide Analytical Centre of Excellence
  - platform, regulatory guidelines, privacy standards
- Small teams of data scientists (e.g. 3-5) embedded in the respective business units
- Augment with model audit teams where necessary
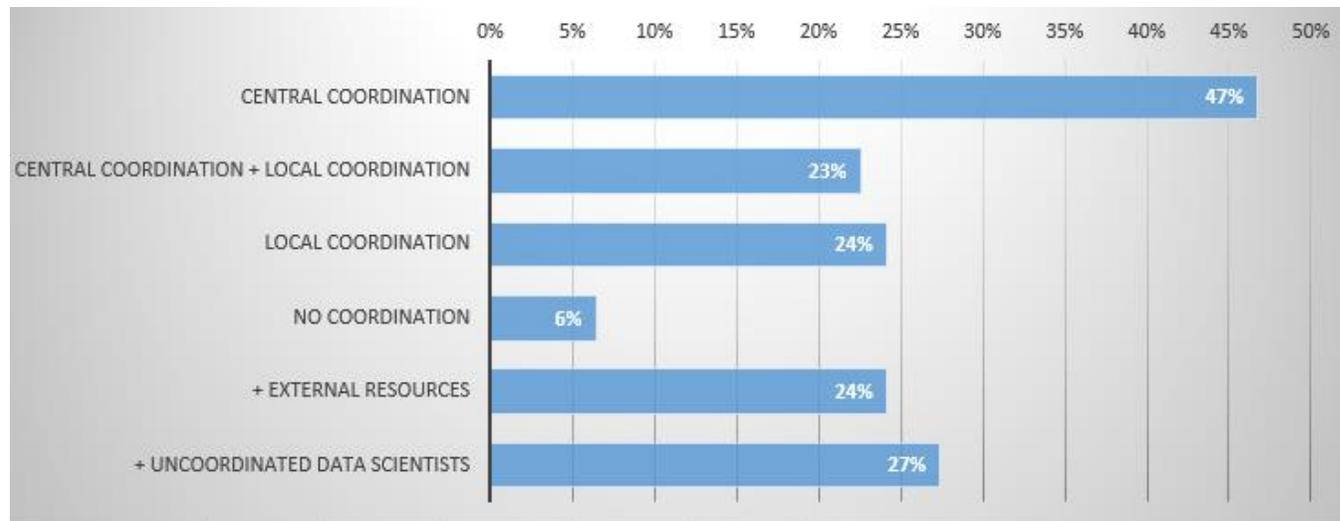
Lismont et al., 2016

In- versus Outsourcing

- Example sourcing activities: heavy lifting data grunt work, analytical platforms (hardware and software), education, model construction, visualization, evaluation, monitoring, …

- Difference with ICT services is that analytics concerns a company's front end strategy

- Data is a company's DNA
  - avoid dilution of competitive advantage by outsourcing

- Hard to outsource creativity!

Commercial versus Open-Source

- **<u>Commercial</u>**
  - E.g., SAS, IBM, Microsoft, Oracle, Matlab
  - Black box
  - Well-engineered business focused solutions (end-to-end)
  - Extensive quality control and help facilities
  - Business continuity

- **<u>Open Source</u>**
  - E.g., R, Python, Weka, Rapidminer
  - White box
  - Free
  - No quality control (regulatory compliance!)
  - Low SME hurdle

Key Lessons Learned

- Avoid NoSQL herd behavior
- Maturity scale (high to low)
  - Risk Management, Marketing Analytics, HR analytics
- Integrated (aka Corporate) Analytics still far off!
- Analytical Centre of Excellence combined with local data scientists
- Do not outsource your core analytical activities
- Commercial and Open Source each have their merits

More Info

datamining apps

search here …   Go

Home   Research   Partnerships and Chairs   Staff   Books   Articles   Videos   Contact Information   Subscribe to our Newsletter   中文

Data Science & Analytics @ LIRIS, KU Leuven

This website contains information about the Data Mining, Data Science and Analytics Research conducted in the research team chaired by prof. dr. Bart Baesens and dr. Seppe vanden Broucke at KU Leuven (Belgium).

Current topics of interest include:

- Marketing Analytics
- Credit Risk Analytics
- Fraud Analytics
- Process Analytics
- Human Resource Analytics

→ Meet our research team

More info

Self-Paced E-learning course: Advanced Analytics in a Big Data World

https://support.sas.com/edu/schedules.html?ctry=us&id=2169

The E-learning course starts by refreshing the basic concepts of the analytics process model: data preprocessing, analytics and post processing. We then discuss decision trees and ensemble methods (bagging, boosting, random forests), neural networks, support vector machines (SVMs), Bayesian networks, survival analysis, social networks, monitoring and backtesting analytical models. Throughout the course, we extensively refer to our industry and research experience. Various business examples (e.g. credit scoring, churn prediction, fraud detection, customer segmentation, etc.) and small case studies are also included for further clarification. The E-learning course consists of more than 20 hours of movies, each 5 minutes on average. Quizzes are included to facilitate the understanding of the material. Upon registration, you will get an access code which gives you unlimited access to all course material (movies, quizzes, scripts, ...) during 1 year. The E-learning course focusses on the concepts and modeling methodologies and not on the SAS software. To access the course material, you only need a laptop, iPad, iPhone with a web browser. No SAS software is needed. See https://support.sas.com/edu/schedules.html?ctry=us&id=2169

More info

Self-Paced E-learning course: Credit Risk Modeling

https://support.sas.com/edu/schedules.html?ctry=us&id=2455

The E-learning course covers both the basic as well some more advanced ways of modeling, validating and stress testing Probability of Default (PD), Loss Given Default (LGD) and Exposure At Default (EAD) models. Throughout the course, we extensively refer to our industry and research experience. Various business examples and small case studies in both retail and corporate credit are also included for further clarification. The E-learning course consists of more than 20 hours of movies, each 5 minutes on average. Quizzes are included to facilitate the understanding of the material. Upon registration, you will get an access code which gives you unlimited access to all course material (movies, quizzes, scripts, ...) during 1 year. The course focusses on the concepts and modeling methodologies and not on the SAS software. To access the course material, you only need a laptop, iPad, iPhone with a web browser. No SAS software is needed. See https://support.sas.com/edu/schedules.html?ctry=us&id=2455

Analytics in Banking: State of the Art

Prof. dr. Bart Baesens

Department of Decision Sciences and Information Management
KU Leuven (Belgium)

School of Management
University of Southampton (United Kingdom)
Bart.Baesens@kuleuven.be

YouTube/Facebook/Twitter: DataMiningApps