WEBINAR@LUNCHTIME
ROCKING ANALYTICS IN A DATA FLOODED WORLD:
CHALLENGES AND OPPORTUNITIES
BY PROF. BART BAESENS
PROF. DR. BART BAESENS

• Studied at KU Leuven (Belgium)
  • Business Engineer in MIS, 1998
  • PhD. in Applied Economic Sciences, 2003
• PhD. Title: Developing Intelligent Systems for Credit Scoring Using Machine Learning Techniques
• Professor at KU Leuven, Belgium
• Lecturer at the University of Southampton, UK
• Research: analytics, credit risk, fraud, marketing, …
• YouTube/Facebook/Twitter: DataMiningApps
  • www.dataminingapps.com
  • Bart.Baesens@kuleuven.be
WEBINAR@LUNCHTIME  LIVING IN A DATA FLOODED WORLD!

Customers

Web/email

Call center

Survey

Corporate data

Partners

Analytics
THE ANALYTICS PROCESS MODEL

Identify Business Problem
Identify Data Sources
Select the Data
Clean the Data
Transform the Data
Analyze the Data
Interpret, Evaluate, and Deploy the Model
Preprocessing
Analytics
Post-processing
WEBINAR@LUNCHTIME  FEEL THE VIBE!

APPLICATIONS

- Fraud Detection
- Market Basket Analysis
- Social Network Analytics
- Customer Lifetime Value
- Web Analytics
- Response Modeling
- Customer Segmentation
- Churn Prediction

WEBINAR@LUNCHTIME  FEEL THE VIBE!
### EXAMPLE: MARKETING CONTEXT

<table>
<thead>
<tr>
<th>Customer</th>
<th>Age</th>
<th>Recency</th>
<th>Frequency</th>
<th>Monetary</th>
<th>Churn</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>35</td>
<td>5</td>
<td>6</td>
<td>100</td>
<td>Yes</td>
</tr>
<tr>
<td>Sophie</td>
<td>18</td>
<td>10</td>
<td>2</td>
<td>150</td>
<td>No</td>
</tr>
<tr>
<td>Victor</td>
<td>38</td>
<td>28</td>
<td>8</td>
<td>20</td>
<td>No</td>
</tr>
<tr>
<td>Laura</td>
<td>44</td>
<td>12</td>
<td>4</td>
<td>280</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Decision Tree**

- **Age < 40**
  - **Recency < 10**
    - Yes: Churn
    - No: No Churn
  - **Recency ≥ 10**
    - No: No Churn
  - **Frequency < 5**
    - Yes: Churn
    - No: No Churn
  - **Frequency ≥ 5**
    - No Churn

---

Copyright © 2012, SAS Institute Inc. All rights reserved.
More than ever before, analytical models steer strategic risk decisions of financial institutions!

Minimum equity (buffer capital) and provisions a financial institution holds are directly determined, a.o., by:
- credit risk analytics
- market risk analytics
- operational risk analytics
- insurance risk analytics
- …

Business analytics is typically used to build all these models!

Often subject to regulation (e.g. Basel II, Basel III, Solvency II, …)! Model errors directly affect profitability, solvency, shareholder value, macro-economy, …, society as a whole!
• Database/Datawarehouse administrator
• Business expert (e.g. marketeer, credit risk analyst, …)
• Legal expert
• Data scientist/data miner
• Software/tool vendors
• **A multidisciplinary team needs to be set up!**
• A data scientist should be a good programmer!
• A data scientist should have solid quantitative skills!
• A data scientist should excel in communication and visualization skills!
• A data scientist should have a solid business understanding!
• A data scientist should be creative!
• Term often used interchangeably with data mining, knowledge discovery, predictive/descriptive modeling, …
• Essentially refers to extracting useful business patterns and/or mathematical decision models from a preprocessed data set

• **Predictive analytics**
  – Predict the future based on patterns learnt from past data
  – Classification (churn, response) versus regression (CLV)

• **Descriptive analytics**
  – Describe patterns in data
  – Clustering, Association rules, Sequence rules
WEBINAR@LUNCHTIME  ANALYTIC MODEL REQUIREMENTS

- **Business relevance**
  - Solve a particular business problem

- **Statistical performance**
  - Statistical significance of model
  - Statistical prediction performance

- **Interpretability + Justifiability**
  - Very subjective (depends on decision maker), but CRUCIAL!
  - Often need to be balanced against statistical performance

- **Operational efficiency**
  - How can the analytical models be integrated with campaign management?

- **Economical cost**
  - What is the cost to gather the model inputs and evaluate the model?
  - Is it worthwhile buying external data and/or models?

- **Regulatory compliance**
  - In accordance with regulation and legislation
• Interpretation and validation of analytical models by business experts
  • Trivial versus unexpected (interesting?) patterns

• Sensitivity analysis
  • How sensitive is the model wrt sample characteristics, assumptions and/or technique parameters?

• Deploy analytical model into business setting
  • Represent model output in a user-friendly way
  • Integrate with campaign management tools and marketing decision engines

• Model monitoring and backtesting
  • Continuously monitor model output
  • Contrast model output with observed numbers
SOCIAL NETWORK ANALYTICS

- Networked data
  - Telephone calls
  - Facebook, Twitter, LinkedIn, ...
  - Web pages connected by hyperlinks
  - Research papers connected by citations
  - Terrorism networks

- Applications
  - Product recommendations
  - Churn detection
  - Web page classification
  - Fraud detection
  - Terrorism detection
• Traditional churn prediction models treat customers as isolated entities
• However, customers are strongly influenced by their social environment:
  – recommendations from peers, mouth-to-mouth publicity
  – social leader influence
  – promotional offers from operators to acquire groups of friends
  – reduced tariffs for intra-operator traffic

→ take into account the customers’ social network!
SOCIAL NETWORKS FOR FRAUD DETECTION

GOTCHA!
Baesens, Van Vlasselaer, Verbeke, 2015
• Employee churn
• Employee performance
• Employee absence
• Employee satisfaction
• Employee Lifetime Value
• …
So, a new employee needs to be scored:

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Attribute</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Up to 26</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>26-35</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>35-37</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>37+</td>
<td>225</td>
</tr>
<tr>
<td>Function</td>
<td>No-manager</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Manager</td>
<td>180</td>
</tr>
<tr>
<td>Department</td>
<td>HR</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Marketing</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>Finance</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>Production</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>IT</td>
<td>240</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Name</th>
<th>Attribute</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>32</td>
<td></td>
<td>120</td>
</tr>
<tr>
<td>Function</td>
<td>Manager</td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>Department</td>
<td>Finance</td>
<td></td>
<td>160</td>
</tr>
</tbody>
</table>

Total 460 points

Let cutoff = 500

No Absenteeism!
What to Do Before You Fire a Pivotal Employee

by Sant Bassens, Sophie De Winne, and Luc Dele

June 25, 2012

These days, the majority of a company’s employees are networked together on social media sites like Facebook, LinkedIn, and Instagram. There 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 Zach Moskowitz 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 and 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?
• Extracting knowledge from event logs of information systems
  – Control flow perspective
  – Organizational perspective
  – Information perspective

• **Advanced Analytics in a Big Data World**
  May 23-25, Heidelberg (Germany)
  [https://support.sas.com/edu/schedules.html?id=2169&ctry=DE&showAll=y&eventId=13074](https://support.sas.com/edu/schedules.html?id=2169&ctry=DE&showAll=y&eventId=13074)

• **Credit Risk Modeling**
  May 23-25, Heidelberg, (Germany)
  [https://support.sas.com/edu/schedules.html?id=2394&ctry=DE&showAll=y&eventId=13075](https://support.sas.com/edu/schedules.html?id=2394&ctry=DE&showAll=y&eventId=13075)
Self-Paced E-learning course: Advanced Analytics in a Big Data World

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

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 (random forests), neural networks, 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.
Self-Paced E-learning course: Credit Risk Modeling


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.
NEXT WEBINAR@LUNCHTIME: 25. APRIL 2016

SAS® CERTIFIED BASE PROGRAMMER FOR SAS®9 – DIE VORBEREITUNG BRINGT’S!
DOWNLOAD SLIDES AND RECORDING
WWW.SAS.DE/LUNCHTIME

HOW DID YOU LIKE OUR WEBINAR?