

BIG DATA IN BANKING AND INSURANCE

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

Twitter/Facebook/Youtube: DataMiningApps

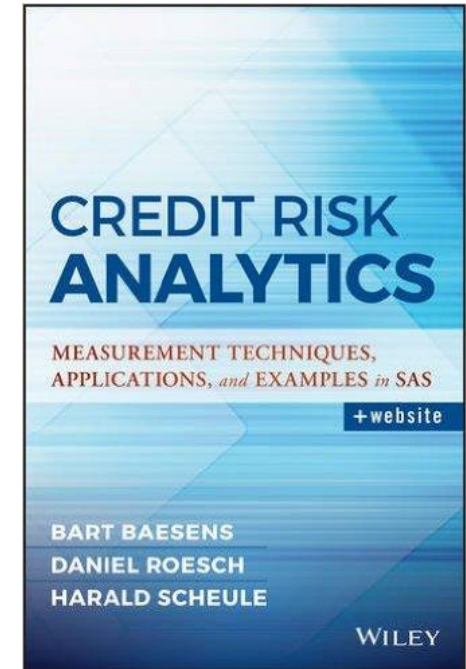
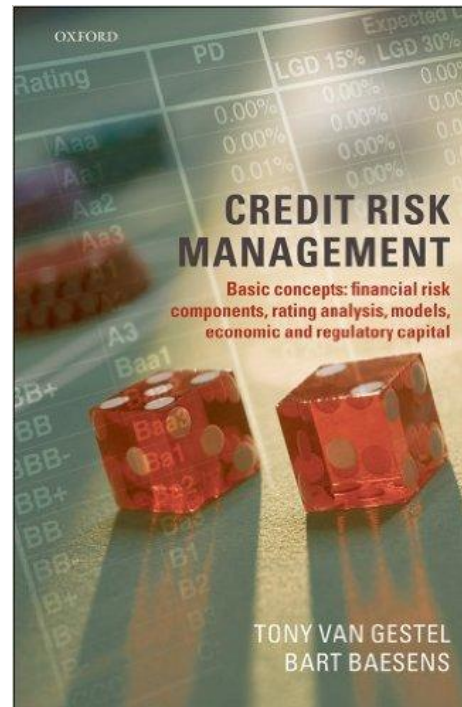
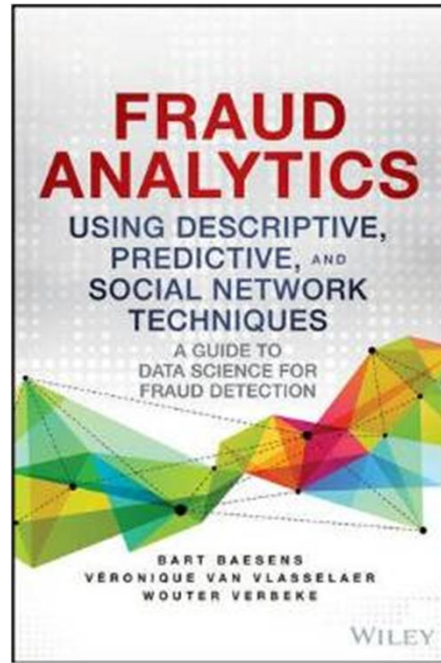
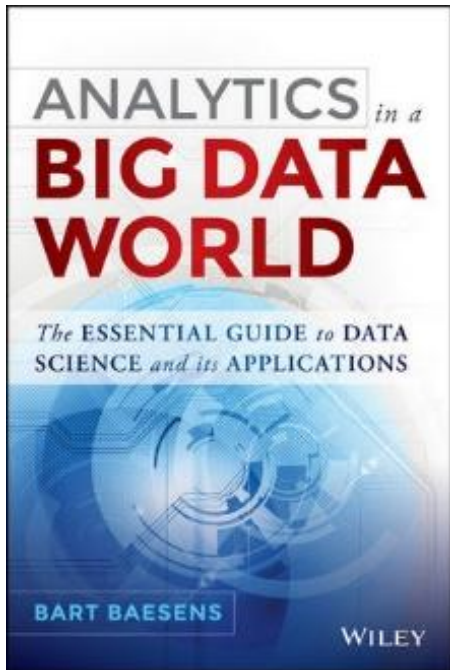
www.dataminingapps.com

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 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

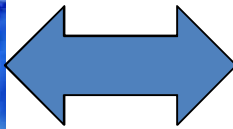


Books



Living in a Data Flooded World!

Customers



Web/email



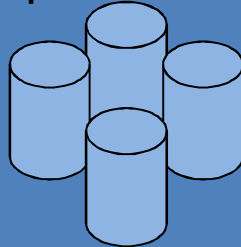
Call center



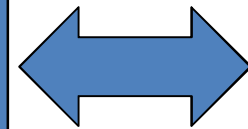
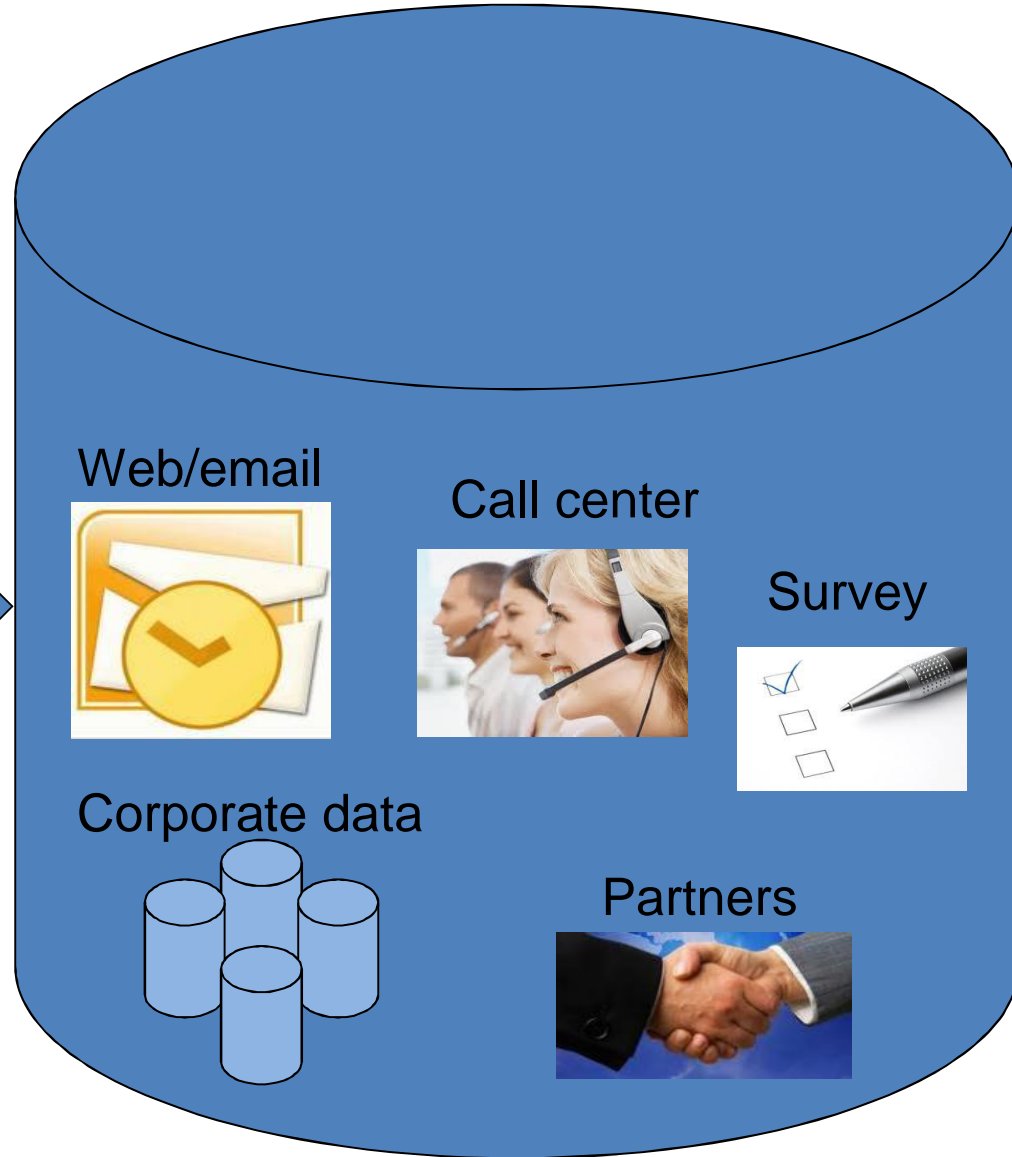
Survey



Corporate data



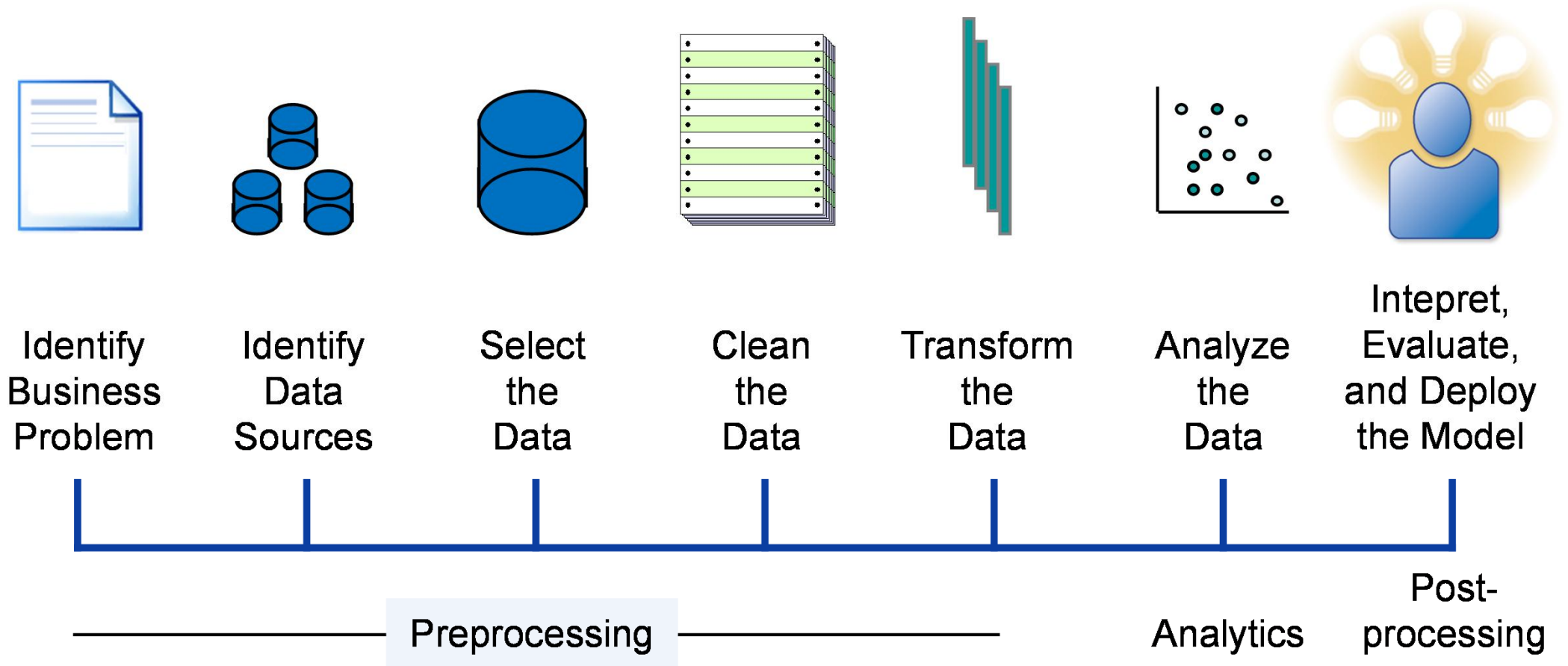
Partners



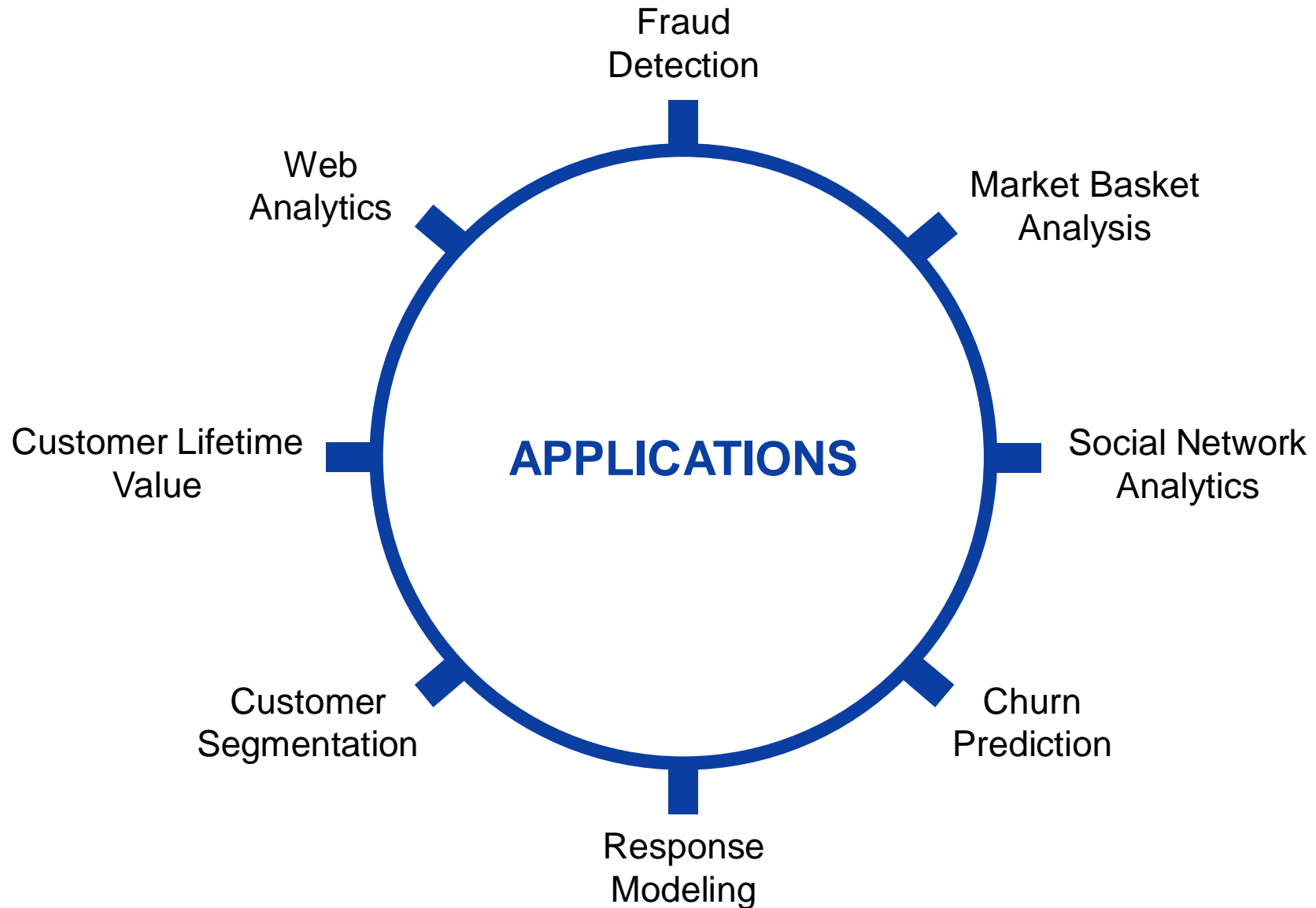
Analytics



The Analytics Process Model

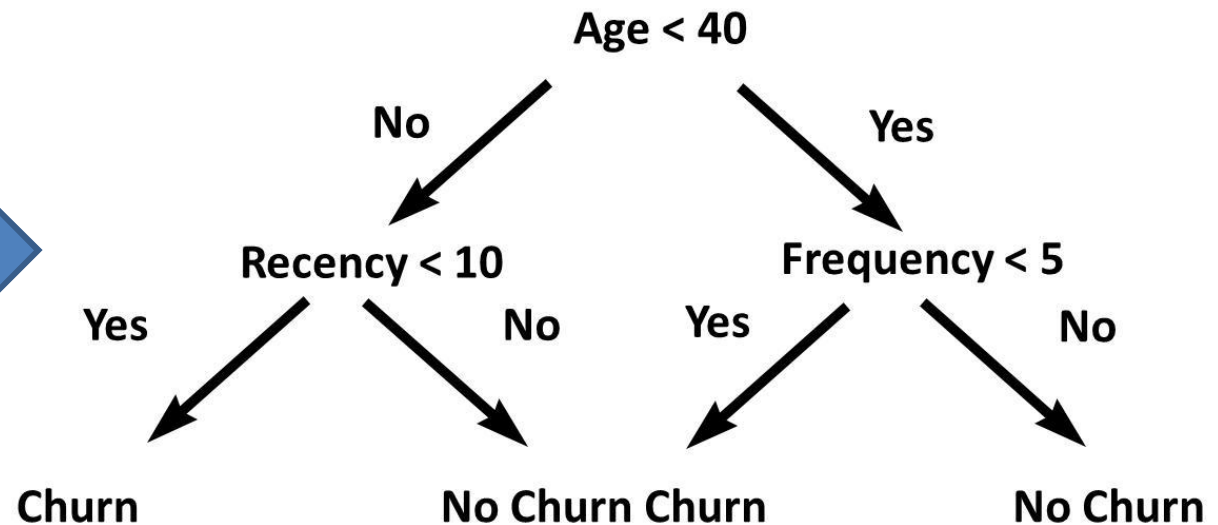
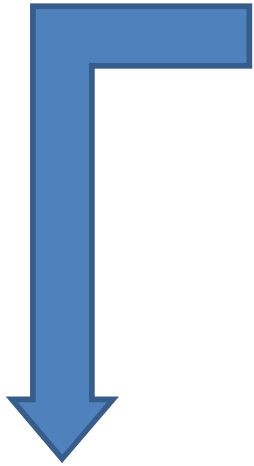


Feel the vibe!



Example: marketing context

Customer	Age	Recency	Frequency	Monetary	Churn
John	35	5	6	100	Yes
Sophie	18	10	2	150	No
Victor	38	28	8	20	No
Laura	44	12	4	280	Yes



Applications in Banking and Insurance

- Credit Risk Modeling
- Fraud Detection
- Telematics
- Marketing Analytics
- Employee Analytics

Example: risk management context

- 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!
-

Basel Accords



Pillar 1: Minimum Capital Requirement

- Credit Risk
- Operational Risk
- Market Risk

Pillar 2: Supervisory Review Process

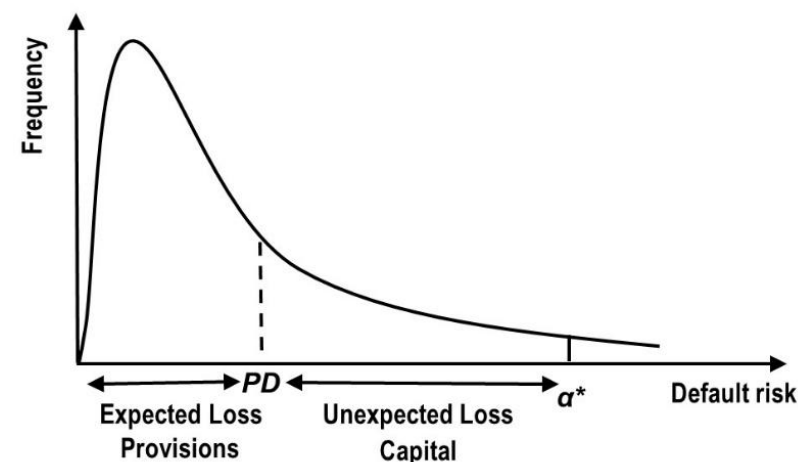
- Sound internal processes to evaluate risk (ICAAP)
- Supervisory monitoring

Pillar 3: Market Discipline and Public Disclosure

- Disclosure of:
 - Bank's risk profile
 - Qualitative and Quantitative information
 - Risk management processes
 - Risk management Strategy
-

Credit Risk Components

- Probability of default (PD) (decimal): probability of default of a counterparty over a one year period (Art. 4, EU)
- Exposure at default (EAD) (currency): amount outstanding
- Loss given default (LGD) (decimal): ratio of the loss on an exposure due to default of a counterparty to the amount outstanding (Art. 4, EU)
- Expected loss = $PD \times LGD \times EAD$
- Unexpected loss = $f(PD, LGD, EAD)$



Example Application Scorecard

Let cutoff = 500

So, a new customer applies for credit:

AGE	32	120 points
GENDER	Female	180 points
SALARY	\$1,150	160 points
<hr/>		
Total		460 points

REFUSE CREDIT

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	2000+	240

Fraud Detection

Credit Card Fraud

Anti-Money Laundering

Insurance Fraud

Tax evasion

Corruption

Counterfeit

Click Fraud

Healthcare Fraud

Product warranty Fraud

Identity theft

Telecommunications Fraud

Definition of Fraud (Baesens, Van Vlasselaer, Verbeke, 2015)

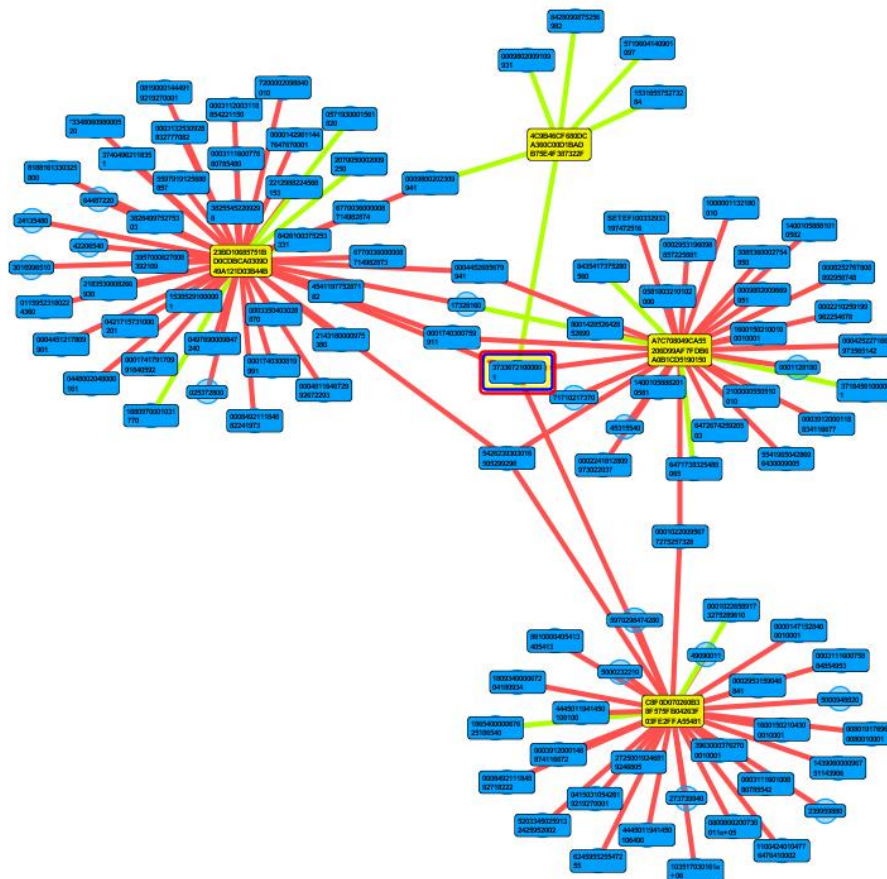
“Fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types and forms.”

Key Challenges in Fraud Detection

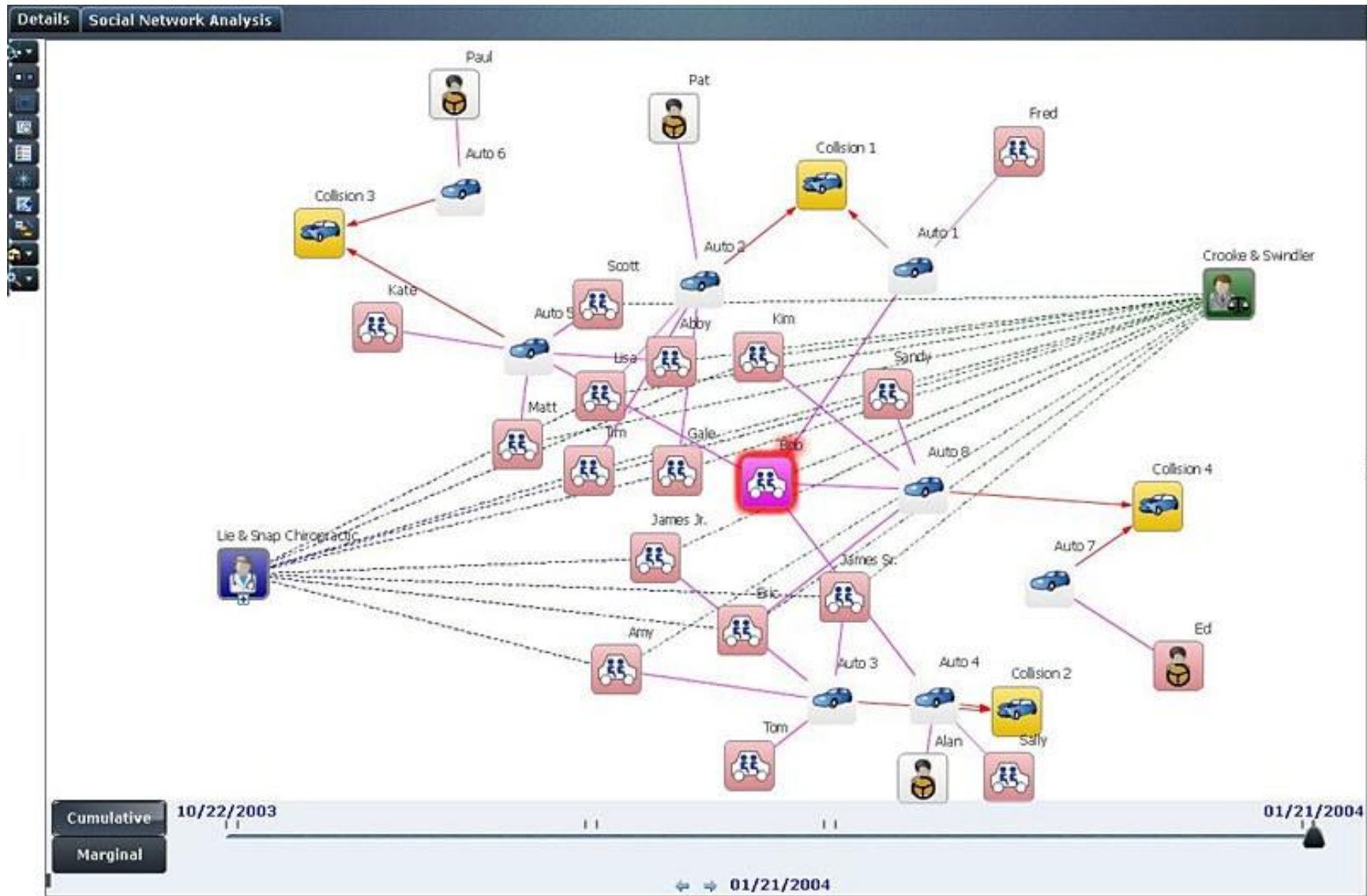
- Fraudsters try to constantly outsmart analytical models
 - Cost of missing fraud is huge
 - Avoid harassing good customers by blocking their accounts
 - Needle in a haystack problem
 - Very skewed data sets
 - E.g. in credit card fraud $< 0.5\%$ frauds typically
 - Operational efficiency is key
 - E.g. in credit card fraud < 8 seconds decision time
 - Huge volumes of data to be processed!
-

Credit Card Transaction Fraud

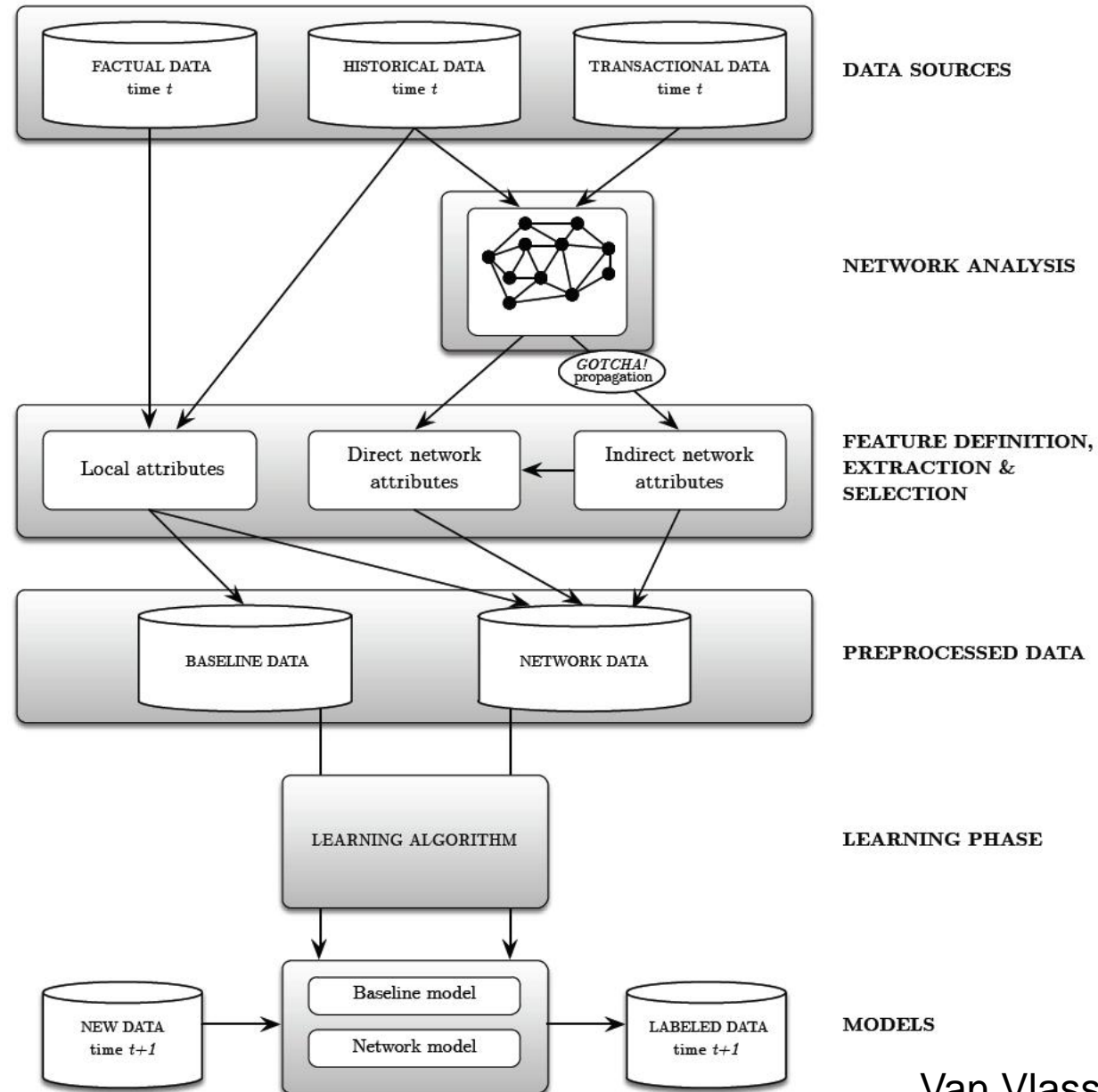
- Credit card transaction fraud:
- Stolen credit cards (yellow nodes) are often used in the same stores (blue nodes)
- Store itself also processes *legitimate* transactions to cover their fraudulent activities



Insurance Claim Fraud



Gotcha!



Telematics

- Telematics is the integrated use of telecommunications and informatics
- Black-box device is installed in the vehicle to monitor real driving behavior
- Allows for better risk assessment and personalized premiums based on individual driving data
- Drives down the cost for low-mileage clients and good drivers
- May fundamentally change the car insurance industry
- AKA usage-based insurance (UBI); pay-as-you-drive (PAYD); pay-how-you-drive (PHYD)

Telematics



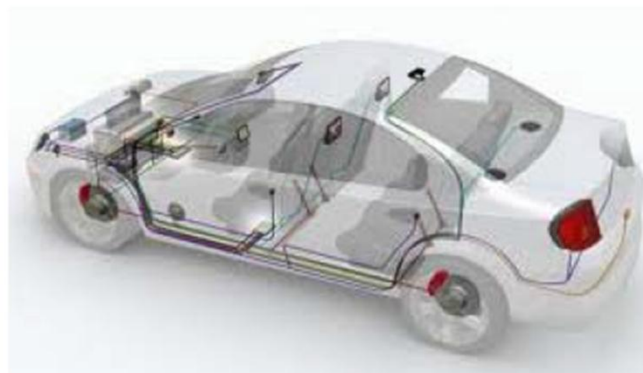
Professionally installed



Self-installed



Smart phone app




OEM embedded system


Telematics

TRADITIONAL DATA

1 VEHICLE DETAILS | 2 DRIVER DETAILS | 3 DISCOUNTS

VEHICLE 1

Number of Vehicles on this Quote: 1
[+ Add another vehicle](#) 

 Your information is secure and will not be sold.

Vehicle year

Vehicle [make](#)

Vehicle [model](#)

Is this vehicle leased?

Purchase or lease date

[Primary use](#) of this vehicle

TELEMATICS DATA

- the time of day
- how long you have been driving
- the location
- the speed
- harsh or smooth breaking
- aggressive acceleration or deceleration
- your cornering and parking skills

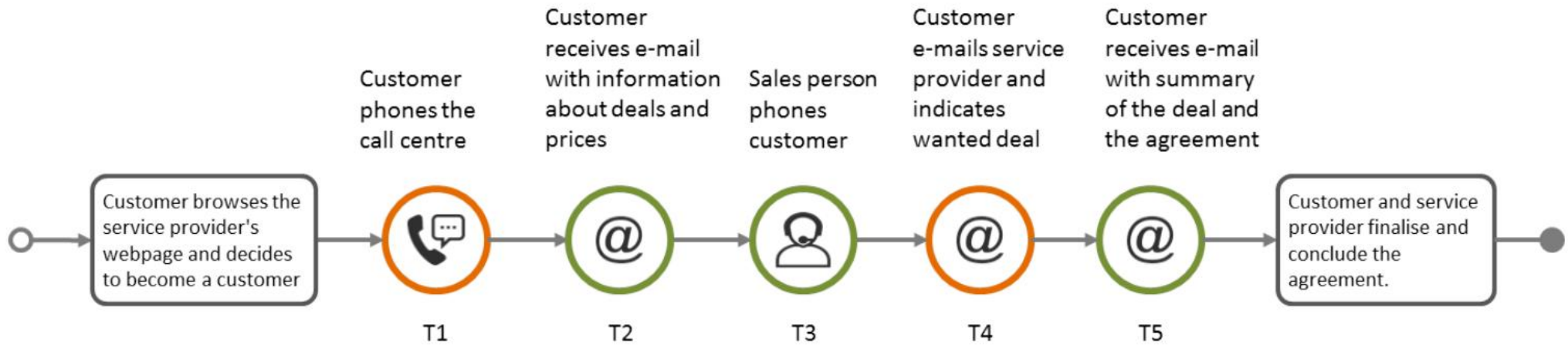
EXTERNAL DATA

- road maps
- weather traffic
- information

Marketing Analytics

- Churn prediction
 - Predict who will leave your firm
 - Response modeling
 - Predict who will respond to a marketing campaign
 - Recommendation systems
 - Predict what products customers will purchase next
 - Customer Lifetime Value Modeling
 - Predict value of customers during specific time frame
 - Customer Journey Modeling
 - Predict the process of customer interaction
-

Customer Journey Modeling



HR analytics

- Employee churn
 - Employee performance
 - Employee absence
 - Employee satisfaction
 - Employee Lifetime Value
 - ...
-

Example Absenteeism scorecard

So, a new employee needs to be scored:

Age	32	120 points
Function	Manager	180 points
Department	Finance	160 points
Total		460 points

Let cutoff = 500

No Absenteeism!

Characteristic Name	Attribute	Points
Age	Up to 26	100
	26-35	120
	35-37	185
	37+	225
Function	No-manager	90
	Manager	180
Department	HR	120
	Marketing	140
	Finance	160
	Production	200
	IT	240

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
-

Conclusions

- Data integration
 - Data quality
 - Hiring the right profiles (data scientists!)
 - Regulatory compliance
 - Economic value
 - Privacy
 - Backtesting
-

Corporate Research Partners



More Information

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.

More information

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> for more details.
