Big Data Analytics @ Munich Re
SAS Global Forum Executive Program - Orlando

Wolfgang Hauner
Chief Data Officer, Munich Re

Marc Wewers
IT Architect, Munich Re

http://dx.doi.org/10.7927/H4X63V7C
Agenda

1. Data Analytics Framework
2. Technology
3. Method Example: AI
4. Case Study: Cross Selling
Big Data in Trend Radar

- Loc-based services
- Smart Home Telematics
- Virtual Assistant Systems
- Haptic Technologies
- Integrated Systems
- Autonomous Systems and Devices
- Automated Decision Taking
- 3D Printing
- Loc-based services
- Telematics
- Smart Home
- Wearable Devices
- Robotics/Drones
- Big Data
- Digitization
- Internet of Things
- 3D Printing
- Augmented and virtual worlds
- Digitalization
- User Centered Design
- On-Demand-Everything
- Digital Identity
- Risk-based Security
- Cybersecurity
- Cloud/Client Architecture
- Software-defined Anything
- New Payment Models
- Web-Scale IT
- Web 4.0
- Context-aware Computing
- Open Data
- Collaborative Consumption
- Mobile Health Services
- Virtual Assistant Systems
- Integrated Systems
- Autonomous Systems
- Automated Decision Taking
- Haptic Technologies
- Big Data
- Digitization
- Internet of Things
- Big Data
- +
- Digitization
- +
- Internet of Things
When does it become BIG Data

- 43 zettabytes of data will probably be generated by 2020
- 300 times the volume in 2005

Source: IBM
Big Data Analytics is a Combination of Methods, Technology, Data and People

- **Methods**
  - Regression Models
  - Machine Learning Models
  - Text Mining

- **Technology**
  - Hardware (Compute power)
  - Software (SAS, R, Spark, …)

- **Data**
  - Internal Data
  - External Data
  - Structured Data
  - Unstructured Data

- **People**
  - Data Scientists
  - Data Engineers
  - Business People
  - IT Architects
Building the Team, and the Environment

- Business-Units
  - Business-/ Domain knowledge
  - Story-telling
- IT
  - System Implementation
  - DB Administration
- Visualization
- Programming
- Maths
- Data Storage
- Statistics
- Modelling
Building the infrastructure

BI Lab

Production

A2P

Data Lake (HDFS)

Long term unstructured and structured data
Which topics drive our clients?

- Textmining
- Data Sources
- Churn Analysis
- Supply Chain
- Predictive UW
- Up-/Cross-Selling
- Fraud Detection
- Social Media Analysis
- Big Data Technology
- Telematics
- Geospatial
- Sensor Data/IoT
- Geospatial
Big Data use cases in insurance

Make the uninsurable insurable
- Diabetics
- Wind Energy

Consolidate the information and process
- Automated underwriting
- Risk management platform

Artificial Intelligence supported workflow
- Early Loss Detection
- Visual Loss Adjustment
Agenda

1. Data Analytics Framework
2. Technology
3. Method Example: AI
4. Case Study: Cross Selling
Design principles for Big Data & Analytics Platform

- SAS & Hortonworks
- Self-Service
- Multi Tenancy
- Automation
- One Central Datalake
- DevOps
- Continuous improvement
- On-Prem & Cloud
Roadmap to Production via Lab environment

Q2 - 2015

- Setup of first BI-Lab Hadoop Cluster

Q3 - 2015

- Setup of new BI-Lab Hadoop Cluster
- Stabilization of BI-Lab Hadoop cluster Authentication & Security Automation

Q4 - 2015

- On-boarding & support of Big Data & Analytics pilots
- New BI-Lab Hadoop Cluster available
  - Large shared cluster
  - Dedicated clusters
  - Single-Node cluster
- Pilot SAS – Hadoop Integration
Building the Big Data & Analytics Platform
Production Environment

Design Setup / Build Run Enhance Optimize Enhance / Optimize

- SAS 9.4 M3
- SAS Visual Analytics (VA)
- Self-Service Data Upload
- SAS Embedded Process for Hadoop
- HDP 2.3
- Hue
- Hive
- Ambari
- Ranger with LDAP
- SAS Enterprise Guide (EG)
- SAS MS Office Add-in
- Data Access to SAP HANA, Oracle & MS SQL-Server
- Sqoop
- Pig
- Spark 1.4
- Oozie
- SAS Enterprise Miner
- SAS Contextual Analysis
- SAS Mobile BI iOS App
- HDP 2.4.2
- Ambari Views
- Spark 1.6
- Solr Cloud
- Tesseract
- SAS VA Row Level Security
- SAS HA
- HDP 2.5
- Atlas
- Zeppelin
- Data Catalogue Tool
- Data Lineage
- Compliance & Security

Stability & Performance:
- Technical staging (SBX, INT, PRD)
- HDP Hotfixes
- SAS Hotfixes (> #100)

2016
- 12 01 02 03 04 05 06 07 08 09 10 11 12
- Start setup platform
- Release v1.0
- Release v2.0
- Release v3.0

2017
- 08 09 10 11 12
- 2-week iterations with Rolling Upgrades
- Release v4.0
Big Data & Analytics Production Environments
IT and Business Deployment

Production

Integration

Sandbox

Self-Service
Ad-hoc Analytics

Scheduled
Analytics

Business Deployment

IT Deployment

SAS

Hortonworks
Big Data & Analytics Production Environments

Scalability

Production

Integration

Sandbox
“Simplified“ Server Architecture SAS and HDP

SAS Mgmt & Metadata
SAS In-Memory LASR
SAS EP
Spark
Solr
Hive
YARN
HDFS
SAS In-Memory LASR
SAS EP
Spark
Solr
Hive
YARN
HDFS
SAS In-Memory LASR
SAS EP
Spark
Solr
Hive
YARN
HDFS
SAS In-Memory LASR
SAS EP
Spark
Solr
Hive
YARN
HDFS
SAS EP
Spark
Solr
Hive
YARN
HDFS
SAS EP
Spark
Solr
Hive
YARN
HDFS
SAS EP
Spark
Solr
Hive
YARN
HDFS
SAS EP
Spark
Solr
Hive
YARN
HDFS

EP = Embedded Process "bring calculation to data"

Hadoop Mgmt & Metadata
Hadoop Frontend

Ambari Views, Zeppelin, …

HDFS
YARN
Spark
Solr
Hive

x < y

SAS and HDP share the same technical IT-resources!
Lessons learned

1. Make use of Lab environment
2. Enable Self-Service
3. Agile IT-Project management
4. Automatization
5. Security
6. YARN queue management
Agenda

1. Data Analytics Framework
2. Technology
3. Method Example: AI
4. Case Study: Cross Selling
Artificial Intelligence (AI) is coming …

1st Machine Age

Automation of physical tasks

2nd Machine Age

Automation of cognitive tasks
AI Evolution

1997
- IBM’s deep blue defeats world chess champion
- Purely rule based

2011
- IBM’s Watson AI system wins Jeopardy match against human players
- Mixed machine learning and rule based

2016
- Google’s DeepMind defeats top ranked Go player (Lee Se-dol)
- Purely machine learning based
Insurance specific AI

Munich Re as industry leader in insurance-specific AI

Google, Facebook, Microsoft, Open AI, IBM, Siemens, GE, Bosch, …

Stanford, Oxford, MIT, Open-Source community, …

General AI
Methods

Neural Network

- System of interconnected nodes, exchanging information
- Weights of connections can be adjusted by supervised/unsupervised “learning”
- Pros: Accuracy usually high, prediction fast
- Cons: “Black box” – acquired knowledge not easily comprehensible, training effort high, appropriate data needed
- Application areas, e.g., speech recognition, computer vision, medical diagnosis, automated trading, game-playing (AlphaGo)
AI perspective in insurance

**Insurance value chain**

**AI trends**

**Automated advisory**
- Complement face-to-face interaction
- Enable innovative distribution (e.g., P2P)
- Serve low volume, high frequency segment

**Automated underwriting**
- Leverage all data and resources
- Improve workflow and customer experience
- Serve low volume, high frequency segment

**Claims**
- Enable fast response
- Obtain more detailed information
- Handle low volume, high frequency claims
Use case example
Automated detection of triggers for insurance demand (for marketing campaigns)

German Geo Analytics Project

Project overview

- Using machine learning from geo data for insurance
- Implementation of algorithms for automated detection of triggers for insurance demand from high resolution aerial images, e.g., solar panels

Achievements and outlook

- Algorithm for automated detection of solar panels and satellite dishes implemented (high recognition performance)
- More detection capabilities (e.g., cars) and scaling to all of Germany easily possible and planned for next project phase
- First regional pilot currently in discussion
Agenda

1. Data Analytics Framework
2. Technology
3. Method Example: AI
4. Case Study: Cross Selling
Next Generation Sales Analytics Solutions:

- Data Processing & Data Enrichment
- Machine Learning
- Scoring & Result Visualisation

Digital companies established sales analytics solutions to improve online-sales.

Munich Re transfers this expertise to the insurance sector in order to predict new cross-selling opportunities.
Our solution includes four main steps to improve cross-selling for our client’s portfolio:

1. **Data Processing & Validation**
   - Client data is transmitted, cleansed, validated and enriched (using external and MR data).

2. **Machine Learning**
   - Using machine learning methods like “Random Forrest”, hundreds of decision trees will be generated simulating customer characteristics.

3. **Discuss Results with Client**
   - Results are discussed with client. Data Analysis is modified, if necessary, and a detailed report on the findings is sent to the client.

4. **Campaign Design and Execution**
   - Campaign is designed and tested on a small sample. Findings of this test are used to optimise the strategy. Campaign is carried out.
Afterwards, we match insureds that do not have the target product yet with the clusters calculated in the first step. This provides us with the individual probability of buying the target product for a single insured.

Using machine learning methods, we identify typical clusters of insureds that bought target product, e.g. personal accident:

- **€75 average premium**
  - 3 policies
  - 36 years old
  - 6 years with client
  - Resides in N. Germany
  - Probability of Buying: 37%

- **€85 average premium**
  - 2 policies
  - 32 years old
  - 4 years with client
  - Resides in N. Germany

We make use of the methodology: “Clients who bought this, also bought that!”

**How does it work?** The portfolio is clustered based on comparable characteristics.
The clusters are the result of a data-driven decision tree.

Clusters calculated are the leaf nodes of a decision tree.
Each and every insured is now assigned to a certain leaf node according to his or her attributes.
The buying probability of a customer is determined by which leaf node he or she is assigned.
“Random Forest” is the predominant machine learning method in our sales analytics approach.
The clusters are visualised in an interactive report which can be examined further.

Each leaf node corresponds to one combined tile diagram:

- By selecting a single tile, more insight in the composition of the cluster is available.
- Again, the darker the color of each tile, the higher the profit potential.
- The same concept applies for size. Tile size corresponds to the number of insureds within that cluster.
The selected cluster is visualised and compared to the whole portfolio

Portfolio visualisation:

- The output shows the distribution of different characteristics within the portfolio compared to the chosen cluster.
- A comparison of cluster characteristics against the average portfolio reveals new insights on targeted customers.
Additional interesting findings can be made by comparing the current hot spots of a product with the potential hot spots.
Define Campaigns from Analytics Results:

To ensure smooth transition and deployment of the analysis steps, MR supports your sales campaign:

- Discuss and set-up campaign strategy
- Integrate existing campaign management
- Select appropriate sales channels for campaign (preferred: online channels)
- Provide A/B-Testing approach to clarify on ideal customer approach
- Compare with recent campaign set-ups and optimise first tests

A/B-Testing

Select top-scored insureds with highest potential according to …

Sample A

Compare hit ratio

… existing approach

… Munich Re’s approach

Sample B

We support campaign design & execution to ensure full leverage of our Sales Analytics Solution