

Moderne Methoden des Data-Mining in der Lebensversicherung

W. Olbricht, Universität Bayreuth & R. Krüger, Swiss Re
DVfVW-Tagung in Köln, 16. November 2016

Einleitung

Big Data – ein unscharfer Begriff:

(Big Data) ... „steht dabei grundsätzlich für große digitale Datenmengen, aber auch für deren Analyse, Nutzung, Sammlung, Verwertung und Vermarktung.“

Reichert, R. (2014), Seite 10

Big Data ist nichts Neues in der Lebensversicherung ...

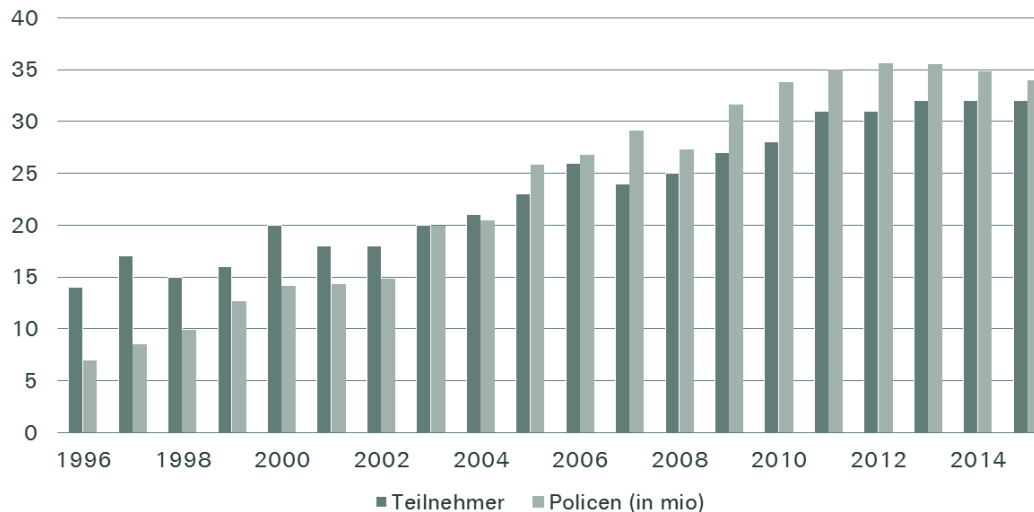
... aber:

Was können wir aus der Anwendung von Data Mining-Methoden lernen?

Bestandsmonitoring

Hintergrund und Entwicklung

Anzahl Teilnehmer und Policen



# Datensätze der wichtigsten Produkte 2013	(in Mio.)
Risikolebensversicherung	4,5
Kapitallebensversicherung	8,4
Fondsgebundene Vers.	2,0
Rentenversicherung	10,4
BU Zusatzversicherung	4,6
Selbständige BU	1,9

1996

- 14 Lebensversicherer
- Produkte: KLV, RLV, BUZ, Rente
- 6 Mio. Policen

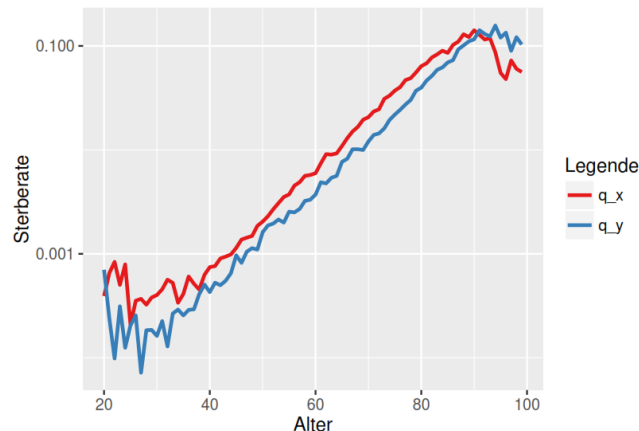
2015

- mehr als 30 Lebensversicherer (ein Drittel des Marktes)
- 20 Produkte (... + FLV, FRV, SBU, LTC etc.)
- 34 Mio. Policen



Datengrundlage für die Fallstudie

- Daten von Todesfalltarifen (KLV, RLV, FLV) von fünf Gesellschaften (A – E) der Jahre 2010 – 2015
- Insgesamt 2,8 Millionen Policen mit rund 43.000 Todesfällen
- 13 Variablen
Gesellschaft, Tarif, Geschlecht, Alter, Selektionsjahr, Vertrieb, Summe, Raucher/Nichtraucher, Versicherungsdauer, Risikoprüfung, Zuschlag, Bundesland, Bevölkerung
- Aufteilung in Trainings- (2010) und Testdatensätze (2011– 2015)
- Ein erster Blick:



Quelle: Klenner (2016),
Abbildung 3.3

Steckbrief: “Klassisches” Verfahren

Grundidee:

Anpassung der beobachteten Sterblichkeit an eine marktübliche Referenzsterbetafel mittels eines Faktors:

$$\hat{q}_x = \alpha \cdot q_x^{Ref}$$

Üblicher Faktor:

$$\alpha = \frac{A(ctual)}{E(xpected)} = \frac{\sum_x T_x}{\sum_x L_x \cdot q_x^{Ref}},$$

mit T_x : Anzahl der Toten und L_x : Exposure des Bestandes

Anmerkungen:

- Referenztafel hier: DAV 2008 T
- Anpassung erfolgt getrennt für KLV, RLV, FLV und Männer/Frauen

Steckbrief: Logistische Regression

Grundidee:

$$Y_i = \begin{cases} 1 & \text{falls } \beta_0 + \beta_1 x_{1,i} + \dots + \beta_m x_{m,i} + \varepsilon > 0 \\ 0 & \text{sonst} \end{cases}, \text{ wobei } \varepsilon \sim \text{Logistisch}(0,1)$$

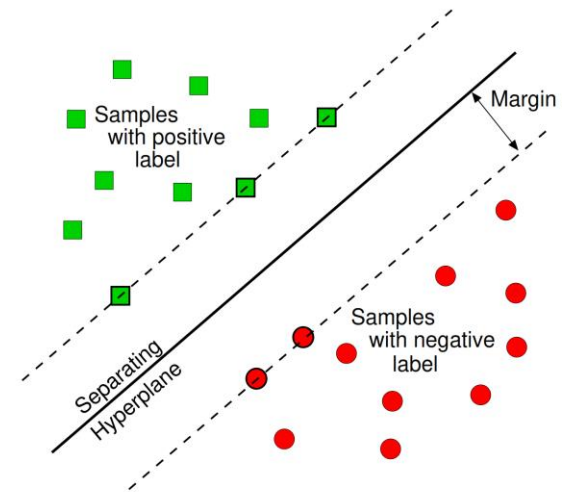
Anmerkungen:

- Gesucht: $p_i = P(Y_i = 1 | x_{1,i}, \dots, x_{m,i})$
- Alternative Formulierung als Generalized Linear Model (mit logit als Link-Funktion)
- Enge Verwandschaft zum probit-Modell
- Schätzung mittels Maximum Likelihood

Steckbrief: Support Vector Machines

Grundidee:

Finde eine möglichst gute Hyperebene (d.h. mit maximaler Trennspanne (Margin)), die die Daten in zwei Klassen separiert.



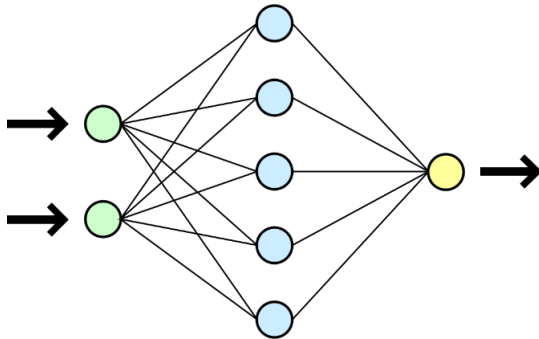
Quelle: Markowetz (2003),
Seite 12

Anmerkungen:

- Flexibler und Vermeidung von Überanpassung durch Einführung von Schlupfvariablen
- Kernel-Trick für den Fall nichtlinear trennbarer Daten

Steckbrief: Künstliche Neuronale Netze

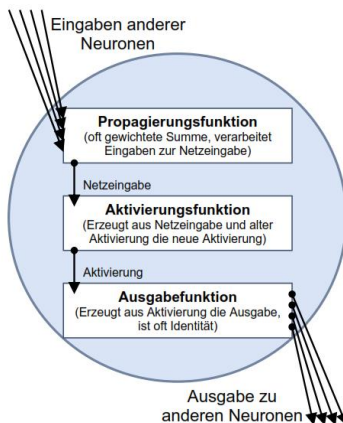
Grundidee:



Quelle: Wikipedia (2016)

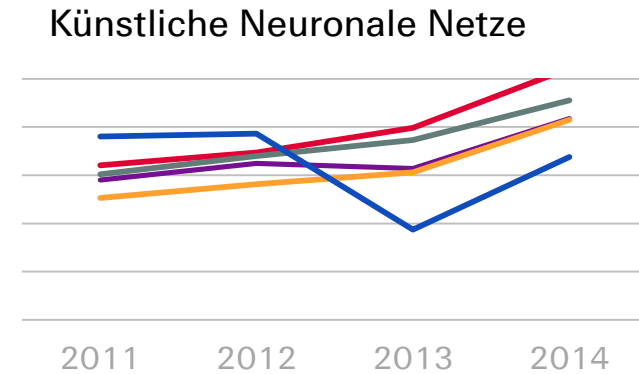
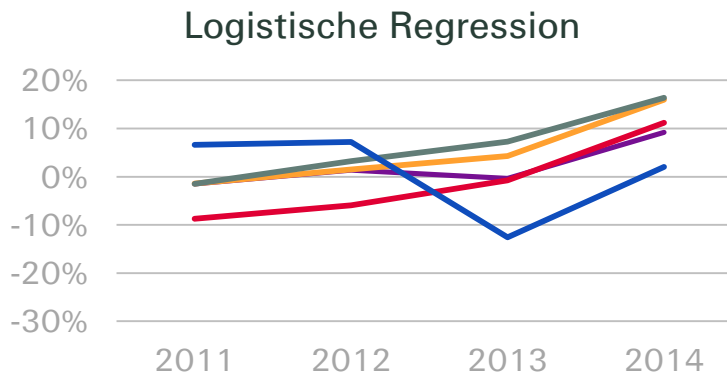
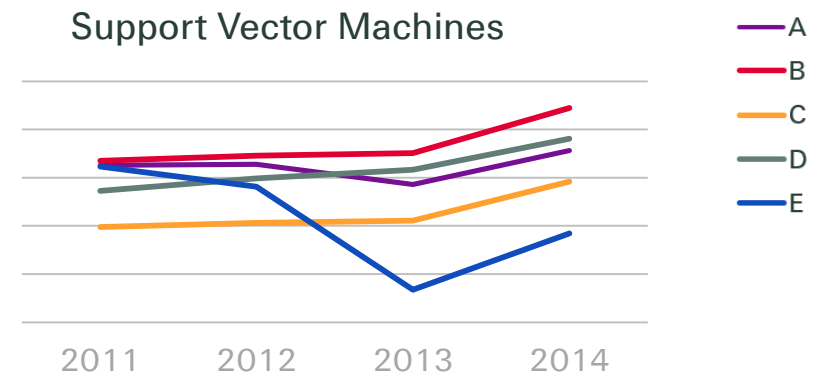
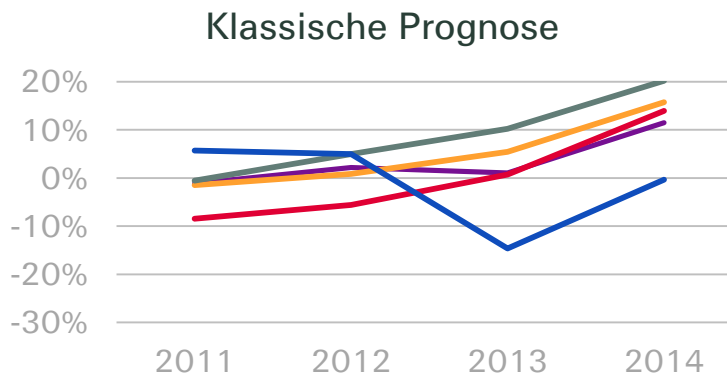
Anmerkungen:

- Ein Künstliches Neuronales Netz besteht aus:
 - Neuronen
 - Topologie des Netzes
 - Dynamik der Neuronen untereinander
- Einfachste Form: Feed-Forward-Netz mit einer verborgenen Schicht
- Lernen durch Backpropagation



Quelle: Kriesel (2007),
Abbildung 3.1

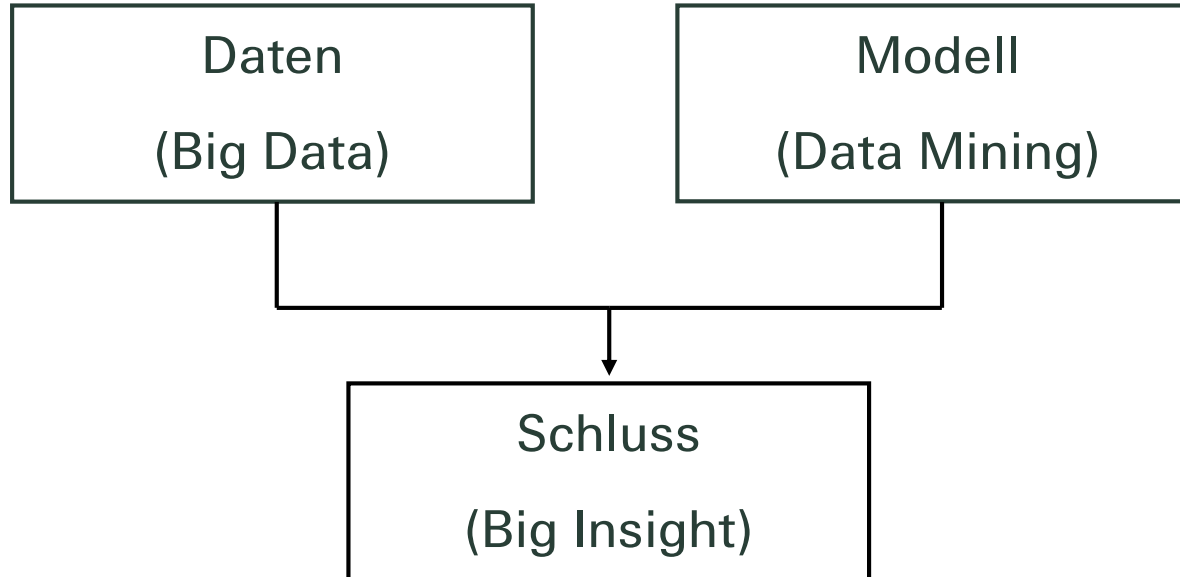
Vergleich der Ergebnisse: % Abweichung des prognostizierten Wertes



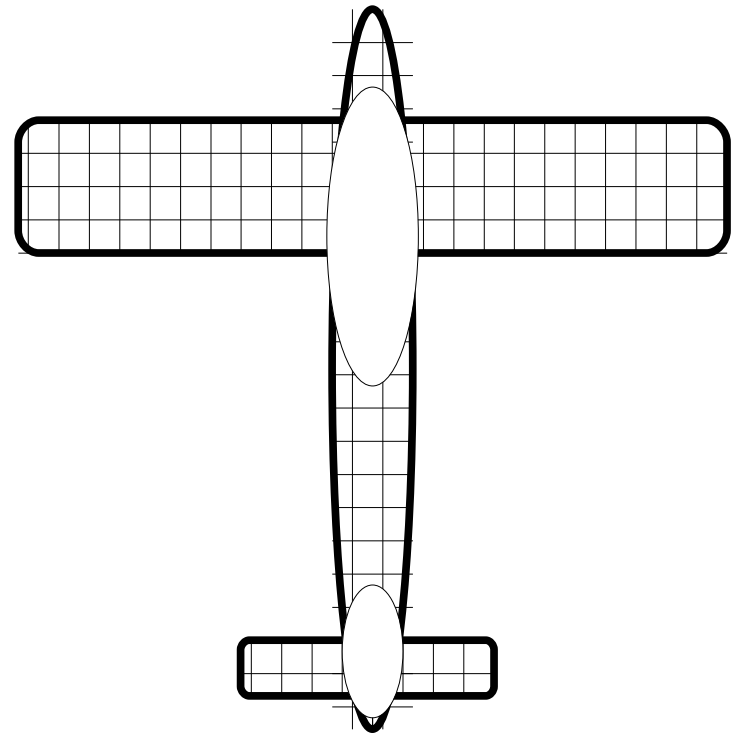
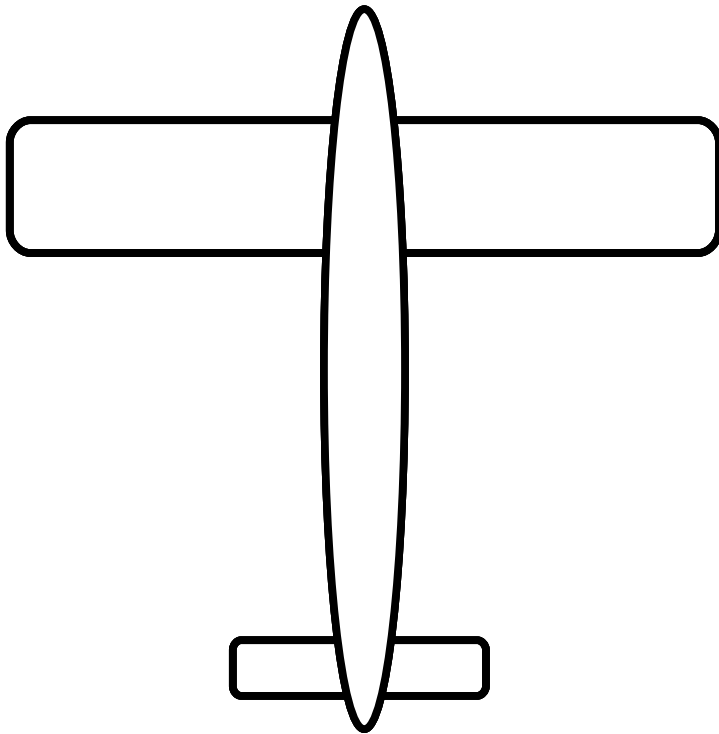
Insgesamt sehr ähnliche Ergebnisse (kleine Abstriche bei SVM)

Theoretische Aspekte

Grundmodell der Statistik:



Flugzeugpanzerung



„Big Data“

sind in der Regel

- sehr groß und nur mit Hilfe von Computern zu verarbeiten
- aus verschiedenen Quellen
- vor allem: **Datensammlungen, keine Zufallsstichproben**

Wissenschaftliche Analyse des Unterschieds: Xiao-Li Meng (2016)

Betrachtung des MSE (mean squared error):

Zufallsstichprobe: kein Bias, MSE hängt vom Stichprobenumfang ab

Datensammlung : Bias, MSE hängt vom “Datensammlungsbruchteil” ab

„Big Data“

Numerisches Beispiel (Meng 2016): Unter gewissen Annahmen entspricht eine Datensammlung vom Umfang **160 Mio** (Datensammlungsbruchteil = $\frac{1}{2}$ für USA) einer Zufallsstichprobe von ... **400**.

Illustration: Alkoholbestimmung, Politbarometer.

Historisches Beispiel: Literary Digest Prognose der Präsidentschaftswahl 1936, Prozentanteil Roosevelt:

	LD	Gallup für LD	Gallup	Resultat
Umfang	2,4 Mio	3000	50000	alle Wähler
Prognose	43	44	56	63

(LD stellte sein Erscheinen ein...)

„Big Data“

Zwischenfazit:

- Es zählt Qualität, nicht Quantität
- Bei Datensammlungen kommt es auf den Datensammlungsbruchteil an.
- Die klassischen Verfahren (z. B. Konfidenzintervalle) sind bei Big Data fast sicher falsch.

„Data Mining“ (Modelle)

sind in der Regel

- Computeralgorithmen
- mit vielen Einstellparametern versehen
- relativ wenig transparent und schwer durch Eingriffe steuerbar (black box)

Wissenschaftliche Analyse der Konsequenzen: David Hand (2006)

Zwei Punkte:

Abnehmender Grenznutzen: Einfache Modelle erfassen schon den **Großteil der Information**.

Fehlende Stabilität in einem **sich ändernden** Umfeld: Einfache Methoden konzentrieren sich auf die stabileren Aspekte.

„Data Mining“ (Modelle)

Numerisches Beispiel (Hand 2006): Regression bei Äquikorrelation

Illustration: s. u.

Historisches Beispiel: Google Flu Trends (Lazer u.a. 2014, Harford 2014)

Zunächst Enthusiasmus (besser, schneller und einfacher als CDC), dann Ernüchterung (laufende Überschätzung z. T. auf das Doppelte).

Wichtiger Grund: Der Google-Algorithmus identifizierte 50 Suchbegriffe, die zwar für 2009 sehr gut waren, aber den Sachverhalt übermodellierten und später – auch aufgrund von Veränderungen im Suchverhalten – nicht mehr richtig funktionierten.

(Google Flu Trends wurde inzwischen eingestellt...)

„Data Mining“ (Modelle)

Zwischenfazit:

- Komplexität ist problematisch.
- Gerade die rasante Veränderung bei Big Data legt Konzentration auf das Wesentliche nahe.

Blick auf die LV

- Hier ist Big Data eigentlich gar **nichts Neues!**

(O & K, 2010:

Datennatur: Im aktuariellen Bereich geht es oft eher um Substrukturen in (Quasi-)Totalerhebungen als um Zufallsfehler bei „echten“ Stichproben.

Cox (1972, S. 187): *In other words, the applications are more likely to be in industrial reliability studies and in medical statistics than in actuarial science.)*

Blick auf die LV

- Hier ist gibt es einen großen (bisher nicht angesprochenen) Aktivposten: **Kontextwissen.**

(O & K, 2010:

Die Intransparenz vieler Methoden erschwert Plausibilitätsbetrachtungen und das Einbringen von Kontextwissen.)

Illustration

Beispiel: durchschnittliche Schadenhöhe für vier gleich große Untergruppen mit hoher Besetzung

	Raucher	Nichtraucher
Männer	1000 950 1000	700 750 700
Frauen	500 550 450	400 350 450

- Rohdaten
- lineares Modell
- nichtlinearer Baum

Illustration

Lineares Modell:

$$\text{SCHADENHÖHE} = \mu + \theta_1 \text{MÄNNER} + \theta_2 \text{RAUCHER}$$

μ = Grundniveau (für nichtrauchende Frauen) ≈ 350

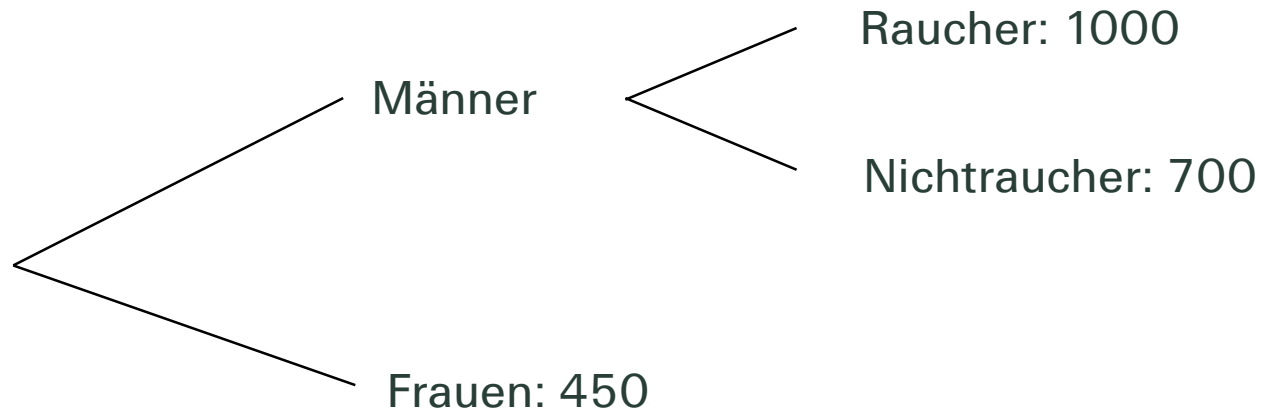
θ_1 = additiver Zuschlag für Männer ≈ 400

θ_2 = additiver Zuschlag für Raucher ≈ 200

(Die Zahlen sind die KQ-Schätzungen.)

Illustration

Nichtlinearer Baum:



- Bessere Anpassung
- Einbringen von Hintergrundwissen und Plausibilitätsbetrachtungen möglich
- Man kann in diesem Modell „denken“.

Illustration

Zielvorstellung:

Ein Ansatz mit folgenden Eigenschaften:

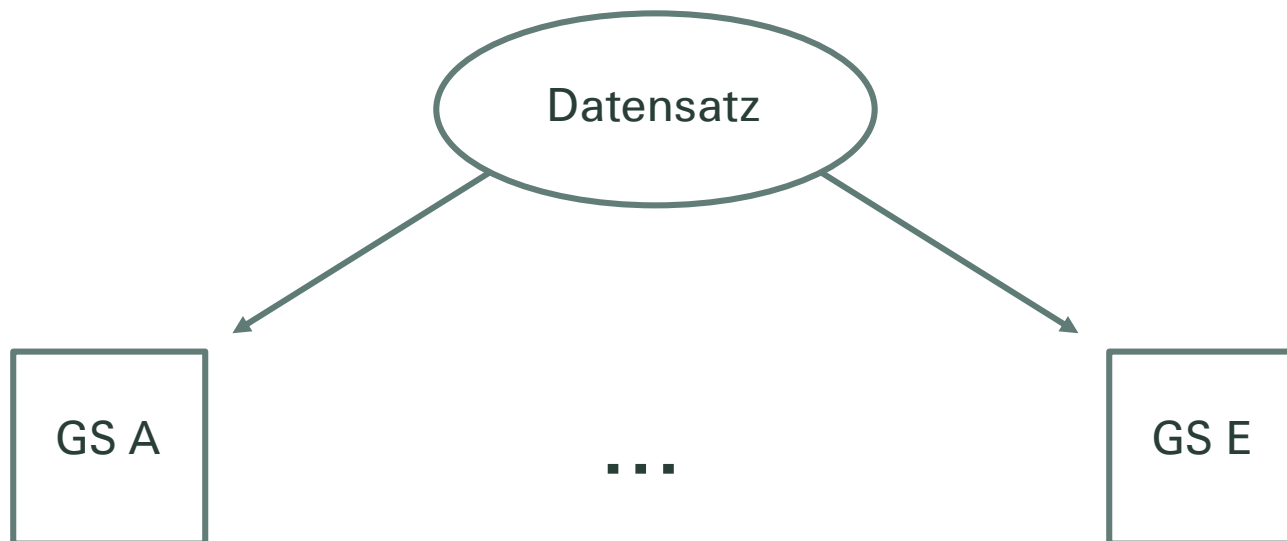
- möglichst datennah
- im Prinzip nichtlinear
- transparent und verständlich
- idealerweise dem „natürlichen“ verfeinernden Denk- und Analyseprozess entsprechend

Lösungsvorschlag: baumbasierte Methoden

Anwendung auf die Studie

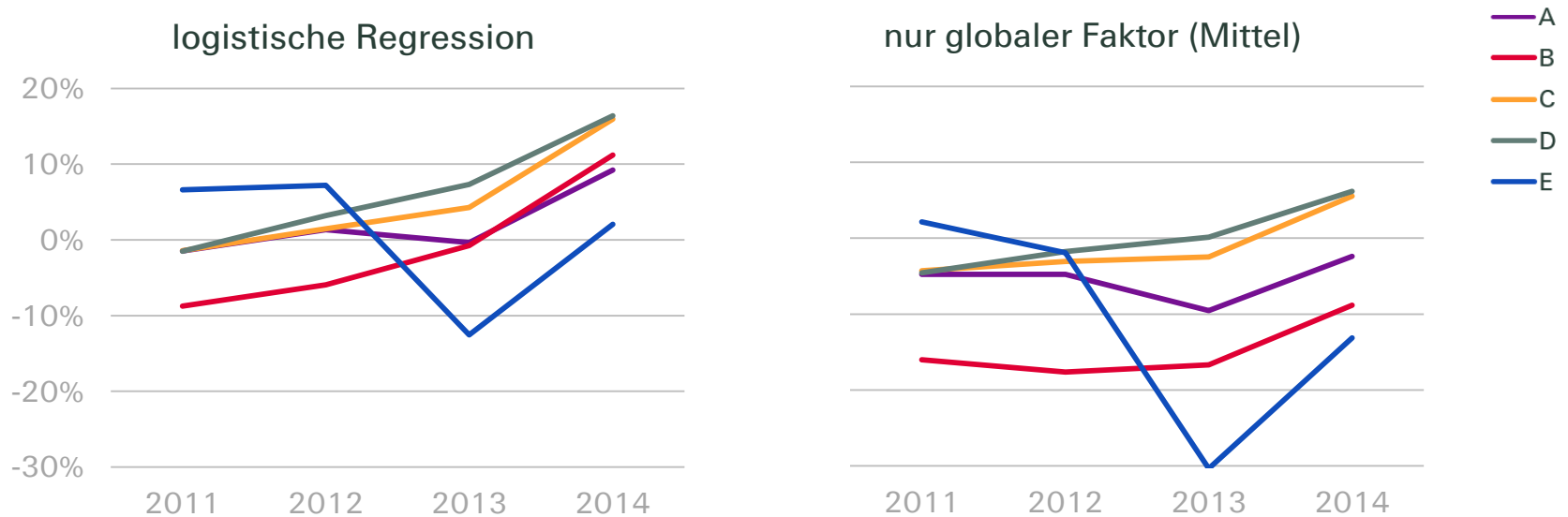
Die Untergliederung nach Unternehmen ist bereits ...

eine vom Kontextwissen getragene Baumstruktur:



Anwendung auf die Studie

Das gilt auch für jedes andere “handfeste” Ergebnis, etwa:



Zugleich Illustration der obigen Punkte von Hand (2006)

Anwendung auf die Studie

Zwei wichtige Fragen aus Sicht der Praxis:

- Ein Black-Box-Algorithmus kennt nur die “Zahlen”.
Ist da wirklich alles Relevante enthalten?
- Ein summarischer Vergleich gilt nur “im Allgemeinen”.
Ist es wirklich das, was man braucht?

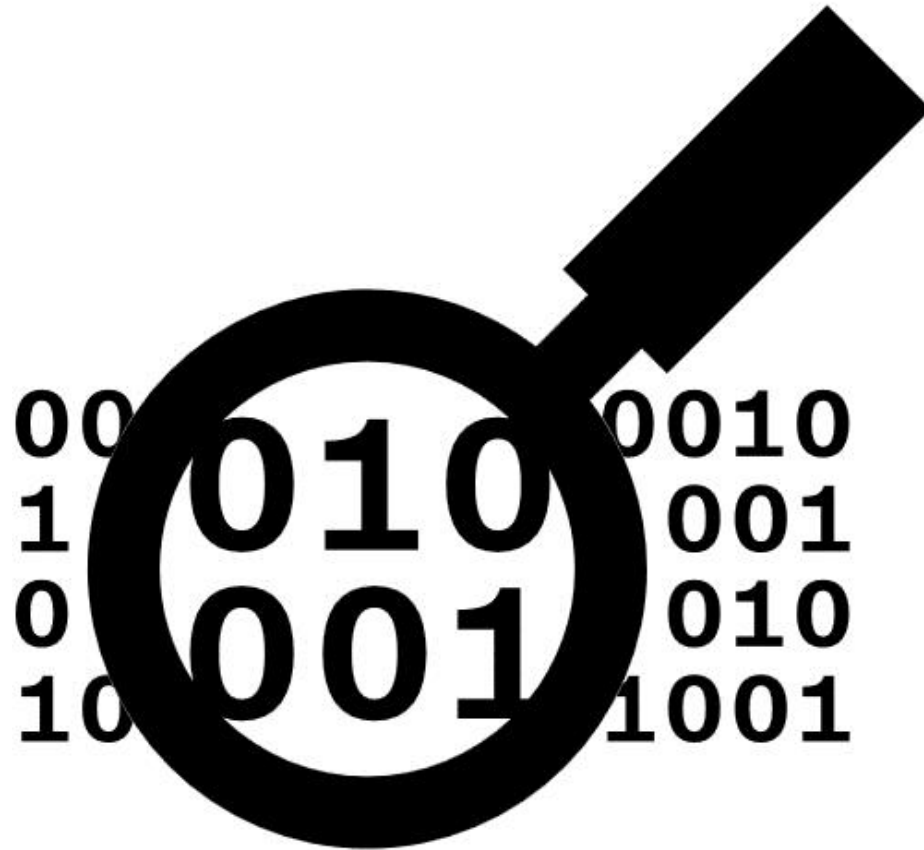
Einordnung

		Situation	
		einfach	komplex
Methoden	einfach	suboptimal	angemessen
	komplex	angemessen	gefährlich!

Fazit

- „“Big Data” has arrived, but big insights have not. The challenge now is to solve new problems and gain new answers – without making the same old statistical mistakes on a grander scale than ever.” (Harford 2014, S. 19)
- problembezogene Analyse, Einbringen von Kontextwissen und einfache (z. B. baumbasierte) Methoden können helfen

Wir freuen uns auf Ihre Fragen



Literatur

- Bauer M, Krüger R, Olbricht W (2013) Tree-based methods: an application to disability probabilities, EurActuarJ 3:491-513
- Cox DR (1972) Regression models and life-tables, JRSS B 34(2): 187-220
- Hand DJ (2006) Classifier Technology and the Illusion of Progress, Statistical Science 21:1-14
- Harford T (2014) Big data: are we making a big mistake? Significance, volume 11, december 2014:14-19
- Klenner F (2016) Ausgewählte Methoden des Data Mining in Anwendung auf Todesfallversicherungen, Masterarbeit Uni Ulm
- Kriesel D (2007) Ein kleiner Überblick über Neuronale Netze, <http://www.dkriesel.com/media/science/neuronalenetze-de-zeta2-2col-dkrieselcom.pdf> (abgerufen am 5.11.2016)

Literatur

- Lazar D, Kennedy R, King G, Vespignani A (2014) The Parable of Google Flu: Traps in Big Data Analysis, Science 343:1203-1205
- Markowetz F (2003) Klassifikation mit Support Vector Machines, Genomische Datenanalyse SoSe 2002, http://lectures.molgen.mpg.de/statistik03/docs/Kapitel_16.pdf (abgerufen am 6.11.2016)
- Meng X-L (2016) Statistical paradises and paradoxes in Big Data, RSS conference, 6 September 2016, <https://www.youtube.com/watch?v=8YLdIDOMEZs> (abgerufen am 10.11.2016)
- Olbricht W, Krüger R (2010) Neue aktuarielle Ansätze in der Berufsunfähigkeitsversicherung, Vortrag auf der Herbsttagung 2010 der DAV-LEBENS-Gruppe

Literatur

- Olbricht W (2012) Tree-based methods: a useful tool for life insurance, EurActuarJ 2:129-147
- Reichert R (2014) Big Data: Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie, transcript Verlag Bielefeld
- Wikipedia (2016) Künstliches Neuronales Netz, https://de.wikipedia.org/w/index.php?title=K%C3%BCnstliches_neuronales_Netz&printable=yes#/media/File:Neural_network.svg (abgerufen am 5.11.2016)



Legal notice

©2016 Swiss Re. All rights reserved. You are not permitted to create any modifications or derivative works of this presentation or to use it for commercial or other public purposes without the prior written permission of Swiss Re.

The information and opinions contained in the presentation are provided as at the date of the presentation and are subject to change without notice. Although the information used was taken from reliable sources, Swiss Re does not accept any responsibility for the accuracy or comprehensiveness of the details given. All liability for the accuracy and completeness thereof or for any damage or loss resulting from the use of the information contained in this presentation is expressly excluded. Under no circumstances shall Swiss Re or its Group companies be liable for any financial or consequential loss relating to this presentation.