



BIG DATA für Versicherungen

Horror, Hype oder Heilsbringer?

Prof. Dr. Maria Heep-Altiner

10. Juni 2016

Inhalt

Vorbemerkungen

Begriffsbildung / Begriffsabgrenzung

Big Data unter verschiedenen Aspekten

Fazit / Ausblick

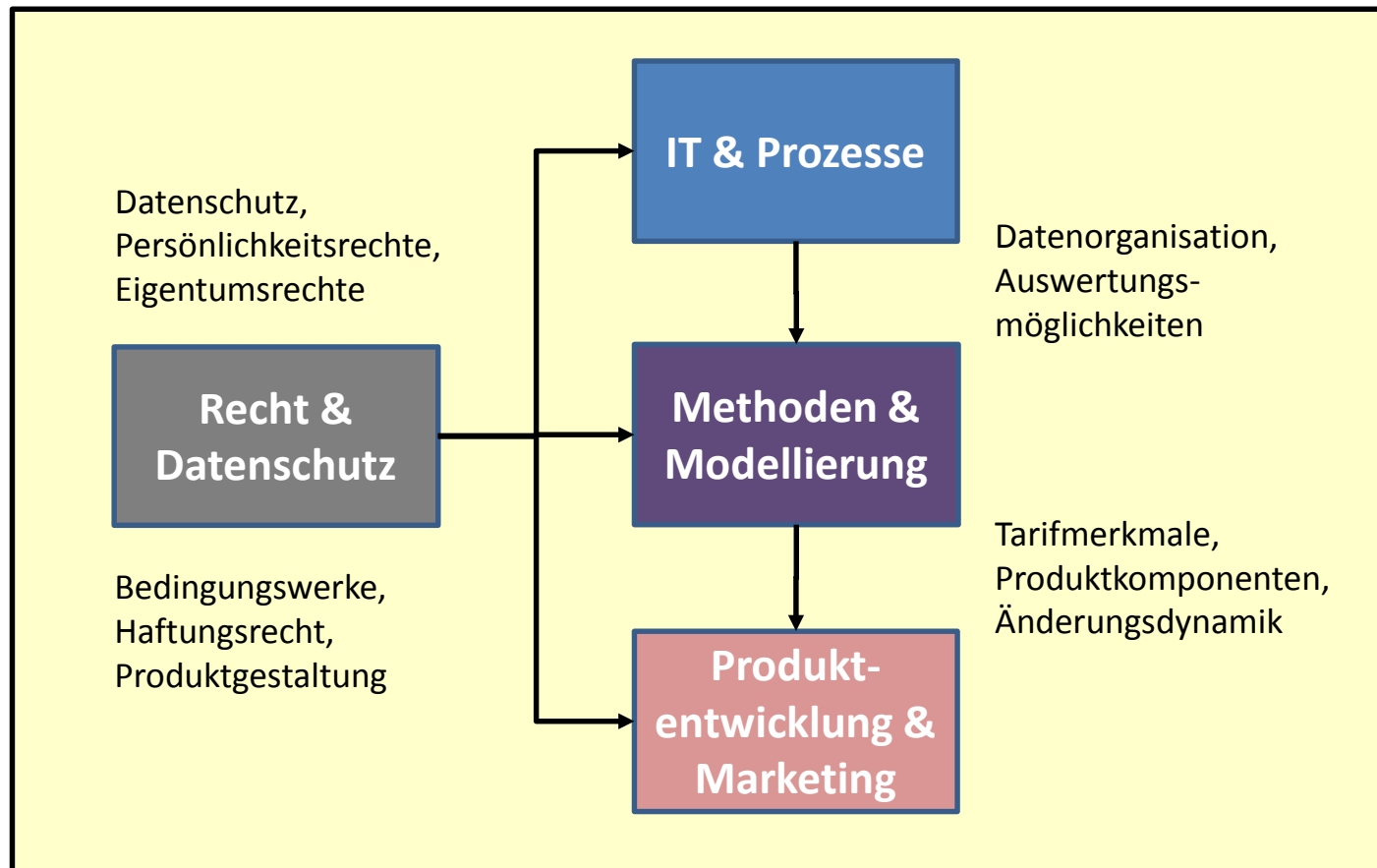
Vorbemerkungen

Big Data als Forschungsschwerpunkt des ivwKöln in 2016

- Übergreifendes Forschungsthema über die vier Forschungsbereiche
 - Aktuarielles Risikomanagement
 - Rückversicherung
 - Versicherungsmarkt
 - Versicherungsrecht
- Möglichst breite Abdeckung des Themengebietes im Hinblick auf die Aspekte
 - IT & Prozesse
 - Methoden & Modellierung
 - Produktentwicklung & Marketing
 - Recht & Datenschutz

Vorbemerkungen

Die wichtigsten Zusammenhänge



Begriffsbildung / Begriffsabgrenzung

DBC: Digitalisierung / Big Data / Cloud Computing

Ad Hoc Definition

Erweiterte Definition

Digitali- sierung

Überführung analoger
Größen in diskrete Größen

Gesamtheit der Technologien und
Methoden zum Aufbereiten und
Speichern analoger Informationen auf
digitalen Medien

Big Data

Große und komplexe
Datenmengen

Gesamtheit der Technologien und
Methoden zum Sammeln und
Auswerten großer Datenmengen

Cloud Computing

Rechnen in einer „Wolke“

Gesamtheit der Technologien und
Methoden zum dezentralen Speichern
und Ausführen von Daten und
Programmen

**Die Ad Hoc Definitionen
sind trivial, die erweiterten
Definitionen unscharf und
schwer sauber abgrenzbar!**

Begriffsbildung / Begriffsabgrenzung

DBC: Digitalisierung / Big Data / Cloud Computing

Klassische Sichtweise

- Technische Restriktionen
- Geringe Datenvolumina (Small Data)
- Nur strukturierte Daten
- Stand Alone Verarbeitung
- Klassische statistische Verfahren
- Einfache Produktstrukturen (z. B. Tarife auf Basis einfach strukturierter Merkmale)
- Datenschutzproblematik als rechtliches Handlungsfeld

Moderne Sichtweise

- Technische Verbesserungen
- Hohe & komplexe Datenvolumina (Big Data)
- Auch unstrukturierte Daten
- Vernetzte Verarbeitung
- Neue Verfahren für unstrukturierte Daten
- Komplexe Produktstrukturen (z. B. Telematiktarife in KH, Vitality Tarife in der KV)
- Viele zusätzliche rechtliche Handlungsfelder

Big Data unter verschiedenen Aspekten

Strukturierte vs. unstrukturierte Daten

IT & Prozesse

Lfd. Nr	Vorname	Name	Geschl.	Geburts-jahr	Gewicht in kg	Größe in cm
1	Heinz	Müller	M	1960	86	183
2	Maria	Huber	W	1974	65	172
3	Anita	Meier	W	1991	56	167
4	Wilhelm	Mayer	M	1979	78	179
5	Jana	Berger	W	1995	63	169
6	Andreas	Bergmann	M	1980	87	185
7	Uwe	Heller	M	1982	73	173
8	Holger	Weber	M	1973	69	177
9	Udo	Wagner	M	1967	93	185
10	Anna	Winter	W	1957	71	170

Strukturierte Daten:

Klare Datensatzformate

Unstrukturierte Daten:

Unsystematische Textformate,
Sprach- und Bildinformationen

Was machst Du gerade??



sitze in der vorlesung von
heep-altiner und langweile
mich. wirklich total öde.

Geht mir mit meiner
Vorlesung genau so. Mathe
ist nichts für mich. Gehen
wir gleich zusammen in die
Mensa?

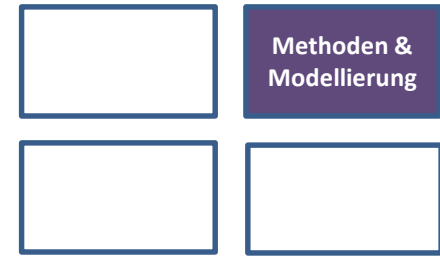
muss jetzt abschalten wegen
regel nr. 3.

Regel Nr. 3:

Während der Vorlesung
dürfen mobile Geräte
nicht genutzt werden.

Big Data unter verschiedenen Aspekten

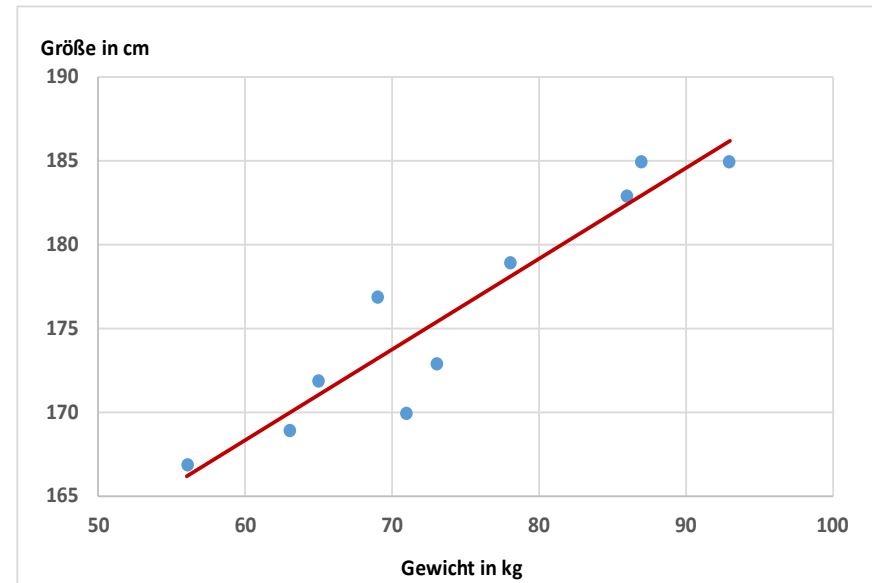
Klassische Statistik bei strukturierten Daten



Originäre Daten

Lfd. Nr	Vorname	Name	Geschl.	Geburts- jahr	Gewicht in kg	Größe in cm
1	Heinz	Müller	M	1960	86	183
2	Maria	Huber	W	1974	65	172
3	Anita	Meier	W	1991	56	167
4	Wilhelm	Mayer	M	1979	78	179
5	Jana	Berger	W	1995	63	169
6	Andreas	Bergmann	M	1980	87	185
7	Uwe	Heller	M	1982	73	173
8	Holger	Weber	M	1973	69	177
9	Udo	Wagner	M	1967	93	185
10	Anna	Winter	W	1957	71	170
Mittelwert					74,1	176,0
Stand. Abw.					11,8	6,8

Beziehungen zwischen Daten



Lageparameter: Mittelwert, Median, Modalwerte, Quantile, etc.

Streuparameter: Varianz, Standardabweichung, Range, Quartilsabstand, etc.

Abhängigkeitsmaße: Korrelationskoeffizient: 94,0%
Bestimmtheitsmaß: 88,4%

Anwendungsgebiet: Substitution eines Merkmals, Tarifierung, etc.

Big Data unter verschiedenen Aspekten

Kontextanalyse bei unstrukturierten Daten

Methoden & Modellierung

Was machst Du gerade??

😊😊😊❤❤❤

sitze in der vorlesung von heep-altiner und langweile mich. wirklich total öde

Geht mir mit meiner Vorlesung genau so. Mathe ist nichts für mich. Gehen wir gleich zusammen in die Mensa?

muss jetzt abschalten wegen regel nr. 3.

Klassische Statistik:

Weiß: 23 Wörter bei 2 Statements
= 11,5 Wörter pro Statement

Grün: 19 Wörter bei 2 Statements
= 9,5 Wörter pro Statement

➔ *Relativ sinnlos*

Kontextanalyse

Schlüsselwörter: langweile, öde
Kontext: Vorlesung, Heep-Altiner

Ein User: Geringe Relevanz
Viele User: Hohe Relevanz
➔ Heep-Altiner sollte die Vorlesung verbessern **oder**
➔ Studenten müssen sich besser konzentrieren.

Big Data unter verschiedenen Aspekten

Umfang von Digitalisierung und Big Data

IT & Prozesse

Methoden & Modellierung

Ausbildungsmodul	Inhalte gemäß Ausbildungsmodul des Fraunhofer Instituts
Basic Analytics	Datentypen / Datenaufbereitung, explorative Analyse, Validierung von statistischen Modellen, Klassifikation, Clustering und Decisiontrees. <u>IT-Basis:</u> RapidMiner, statistische Programmiersprache R.
Big Data Architecture	Speicherung großer Datenmengen in verteilten Dateisystemen, Map Reduce-Verfahren, Echtzeitanalyse in Streaming-Daten, Verteilung und Skalierbarkeit, parallele Datenworkflows und NoSQL-Datenbanken.
Big Data Analytics	Sampling bei großen Datenmengen, Analyse großer Datenmengen mit SQL auf Hadoop, Integration von Quellen mit Lingual, Modellerstellung mit Cascading und Analyse von Datenströmen mit Storm. <u>Anwendungsbeispiele:</u> Online- Empfehlungssysteme und Ranking nach Popularität.
Linked Enterprise Information Integration	Potenzialträchtige Einsatzmöglichkeiten für semantisch verknüpfte Daten im Unternehmen identifizieren, z. B. Linked-Data-Grundlagen, Web-of-Data-Wissensbasen, Abfragesprache SPARQL sowie Mapping und Verlinkung von Daten.

Quelle: Feilmeier, Fühaupter: Digitalisierung und Big Data für die DAV.

Big Data unter verschiedenen Aspekten

Umfang von Digitalisierung und Big Data

IT & Prozesse

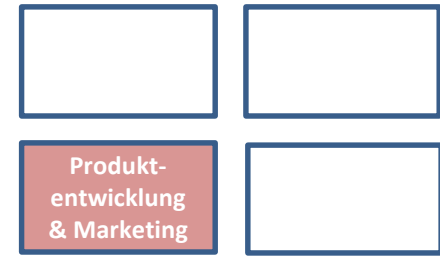
Methoden & Modellierung

Ausbildungs-modul	Inhalte gemäß Ausbildungsmodul des Fraunhofer Instituts
Visual Business Analytics	<p>Visual Analytics vs. Informationsvisualisierung und BI, Grundlagen der Visualisierung, Trends und Werkzeuge.</p> <p><u>Beispiele:</u> Visualisierung im CRISP-Prozess des Data-Mining, Datenvorverarbeitung und visuelle Exploration, visuelles Debugging von statistischen Modellen sowie visuelle Auswertung von Modellergebnissen.</p>
Social Media Analytics	<p>Szenarien / Probleme / Aufgaben / aktueller Stand, Crawling und Monitoring, Repositories und Vorverarbeitung, Analyse von kompletten Social-Media-Beiträgen, semantische Ähnlichkeit von Begriffen, Erkennung von Namen, Produkten und Firmen.</p> <p><u>Beispiel:</u> Marketing im Automotive-Bereich.</p>
Multimedia Analytics	<p>Automatische Spracherkennung, Audio Fingerprinting, Optische Zeichenerkennung in Videos und Logo-Erkennung.</p>

Quelle: Feilmeier, Fürhaupter: Digitalisierung und Big Data für die DAV.

Big Data unter verschiedenen Aspekten

Versicherungsprodukte & Big Data



- Telematik Tarife in Kraftfahrt – Pay as you Drive
- Vitality Tarife in der Krankenversicherung – Pay as you Live
- Business Continuity Management
- Logistik / Supply Chain / Risikomanagement
- Transport / GPS Daten
- Gebäudeversicherung / GPS Daten

Big Data unter verschiedenen Aspekten

Juristische Themenfelder



- Datenschutz
 - Allgemeine Persönlichkeitsrechte
 - Betrifft nur **personenbezogene** Daten (künftig EU einheitlich geregelt).
 - Umfasst nicht **objektbezogene** Daten, z. B. Daten zu einem Maschinenstatus etc. ¹⁾
- Eigentumsrechte
 - Wem gehören z. B. Daten aus technischen Systemen oder Maschinen, auf die mehrere Nutzer zugreifen oder
 - die mehrere Parteien betreiben. ¹⁾
- Haftungsrecht
 - Z. B. Haftung bei autonomen Systemen wie selbstfahrende Pkw oder
 - sonstige Systeme wie etwa Roboter
 - Transformation in das Produkthaftungsrecht?
- Versicherungsrecht

Vgl.: Grillo: Neue Regeln für Big Data. Handelsblatt, 28. Januar 2016.

Fazit / Ausblick

Big Data für Versicherungen

- + Bessere Einschätzung von objektiven Risiken durch Zusatzinformationen wie
 - GPS Daten, Google Maps, Tanker Routen, Supply Chains, etc.

- + Beeinflussung von subjektivem Risikoverhalten durch
 - Pay as you drive, pay as you live, etc.

- + Zur Verringerung der asymmetrischen Informationssituation zwischen VU und VN
 - Bei Vertragsbeginn → Adverse Selektion
 - Während der Vertragslaufzeit → Moral Hazard

Fazit / Ausblick

Big Data für Versicherungen

- Im Extremfall könnte eine absolute Individualisierung der Versicherungsprodukte die Folge sein.
- Dies führt aber den Versicherungsgedanken – nämlich eine „*Ökonomisierung von Risiken durch Kollektivierung*“ – ad Absurdum.
- **Denn:** Der risikoadäquateste Tarif besteht natürlich darin, dass jeder seine Schäden gefälligst selbst zahlt. Das spart dann auch die Verwaltungskosten eines VU.

Die Branche sollte also im eigenen Interesse hier eine gesunde Balance finden, wenn sie Verwerfungen vermeiden will.

Fazit / Ausblick

Big Data: Horror, Hype oder Heilsbringer

- **Horror?**

Eher nein, zumindest wenn man kein Hundertjähriger ist.

- **Hype?**

In einigen Bereichen mit übertriebenen Vorstellungen sicherlich, aber der größte Teil wird dauerhaft Bestand haben. Fortschritt hat sich noch nie in der Geschichte zurückdrehen lassen.

- **Heilsbringer?**

Nur bei verantwortungsbewusstem und vernünftigem Umgang. Die Versicherungsbranche insbesondere wird wohl auch darauf achten müssen, dass sich ihre Produkte nicht „in Luft auflösen“.