

Aus dem Maschinenraum der AG „Q“

Neues vom CADS & mehr

Dr. Stefan Nörtemann, msg life central europe gmbh

DAV Jahrestagung, Fachgruppe ADS, 30. April 2021

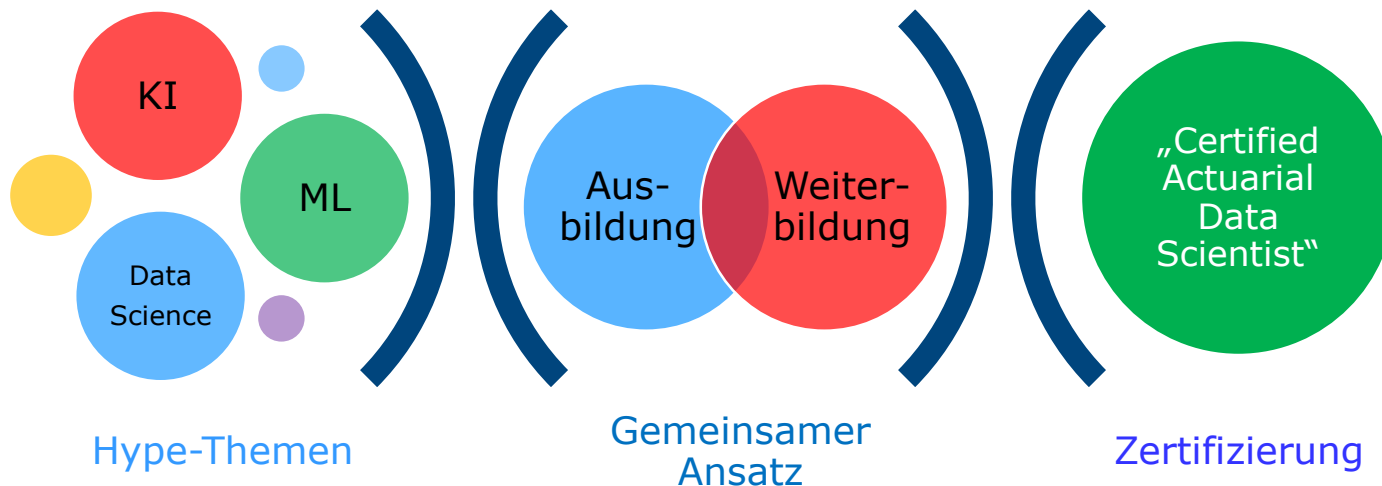


DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

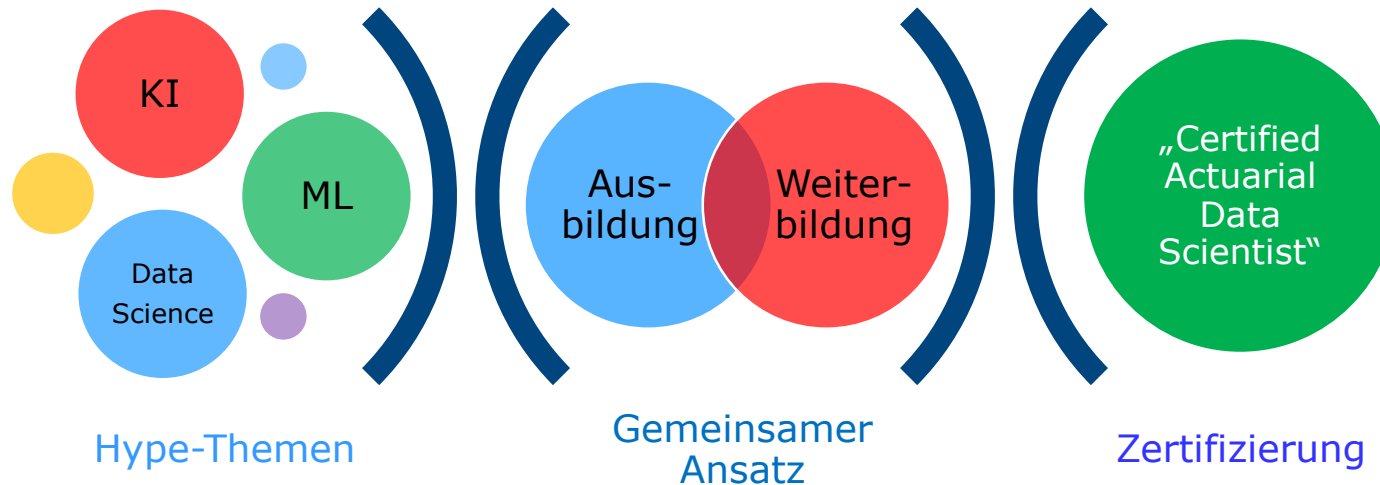
Agenda

1. AG „Q“ – Arbeitsgruppe Qualifizierung
2. CADS – Certified Actuarial Data Scientist



Agenda

1. AG „Q“ – Arbeitsgruppe Qualifizierung
2. CADS – Certified Actuarial Data Scientist



AG Qualifizierung

Was machen die?

- Qualitätssicherung und Freigabe der Prüfungen in
 - ⌘ ADS Basic
 - ⌘ ADS Advanced
 - ⌘ ADS Immersion
 - ⌘ ADS Completion
- in enger Zusammenarbeit mit den beiden Prüfungskommissionen
 - ⌘ ADS Spezialwissen (Leitung: Dr. René Külheim)
 - ⌘ ADS Vertiefungswissen (Leitung: Friedrich Loser)
- Überprüfung und Weiterentwicklung der Lernziele
- Vertretung im Ausschuss für Ausbildung und Prüfung (APA) & im Weiterbildungsausschuss
- Koordination & Unterstützung der beiden Dozententeams
 - ⌘ ADS Basic / ADS Advanced
 - ⌘ ADS Immersion / ADS Completion
- Koordination des Weiterbildungsangebotes in ADS Abstimmung mit der DAA

AG Qualifizierung

Was machen die?

- Qualitätssicherung und Freigabe der Prüfungen in
 - ⌘ ADS Basic
 - ⌘ ADS Advanced
 - ⌘ ADS Immersion
 - ⌘ ADS Completion
- in enger Zusammenarbeit mit den beiden Prüfungskommissionen
 - ⌘ ADS Spezialwissen (Leitung: Dr. René ...)
 - ⌘ ADS Vertiefungswissen (Leitung: Friedrich Loser)
- Koordination & Unterstützung der beiden Dozententeams
 - ⌘ ADS Basic / ADS Advanced
 - ⌘ ADS Immersion / ADS Completion
- Koordination des Weiterbildungsangebotes in ADS Abstimmung mit ...
- Überprüfung und Weiterentwicklung der Lernziele
- Vertretung im Ausschuss für Ausbildung und Prüfung (APA)

AKTUELL: Umfängliche Revision und Vereinheitlichung der Lernziele über alle vier Fächer hinweg!

AG Qualifizierung

Was machen die?

- Qualitätssicherung und Freigabe der Prüfungen in
 - ⌘ ADS Basic
 - ⌘ ADS Advanced
 - ⌘ ADS Immersion
 - ⌘ ADS Completion
- in enger Zusammenarbeit mit den beiden Prüfungskommissionen
 - ⌘ ADS Spezialwissen (Leitung: Dr. René Külheim)
 - ⌘ ADS Vertiefungswissen (Leitung: Friedrich Loser)
- Überprüfung und Weiterentwicklung der Lernziele
- Vertretung im Ausschuss für Ausbildung und Prüfung (APA)

AG „Q“ – Wer ist das?

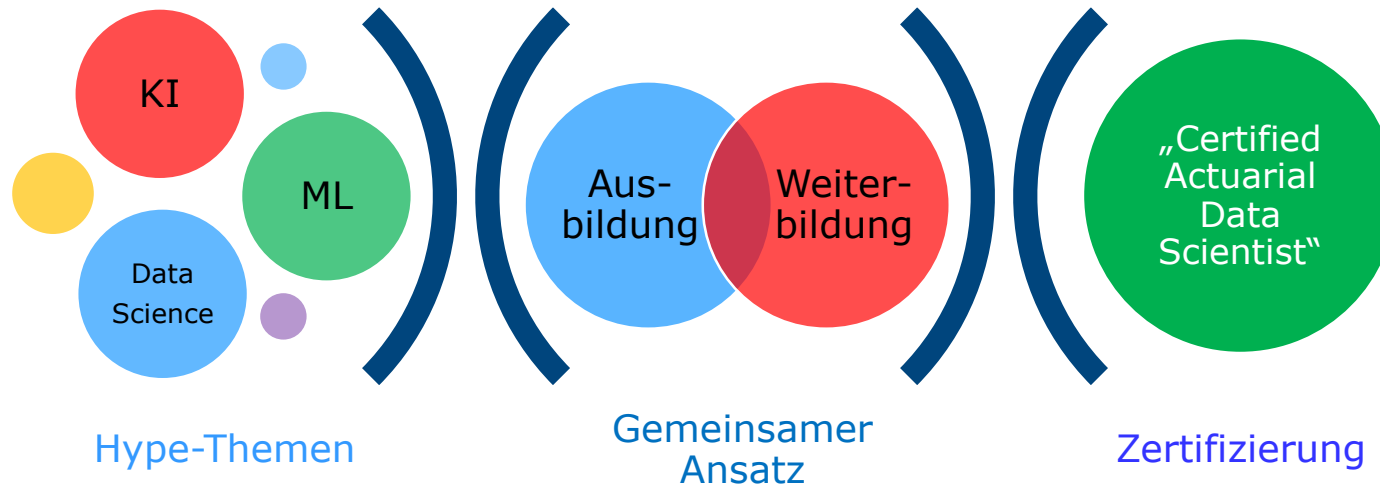
- Dr. Axel Kaiser
- Dr. René Külheim
- Dr. Zoran Nikolic
- Dr. Stefan Nörtemann (Leiter)
- Dariush Sadeghi-Yam
- Dr. Wiltrud Weidner (stv. Leiterin)

Spezialwissen ADS Basic & Advanced (& Immersion)

- Mit der Prüfungsordnung PO 4 wurden zwei neue Fächer mit Wahlpflichtfächer eingeführt:
 - ⌘ **Actuarial Data Science Basic**
 - ⌘ **Actuarial Data Science Advanced**
- Dazu gibt es (seit 2019) je ein dreitägiges Seminar, in dem die Inhalte der Lernziele vermittelt werden.
- Mit der PO 4.1 wurde die Möglichkeit geschaffen, ein drittes Fach Spezialwissenwahlfach zu wählen:
 - ⌘ **Actuarial Data Science Immersion**
- Wie die Anmeldezahlen belegen, wird davon wird seit Beginn an reger Gebrauch gemacht.

Agenda

1. AG „Q“ – Arbeitsgruppe Qualifizierung
2. CADS – Certified Actuarial Data Scientist





CADS

spricht: "C-A-D-S" oder "Katz"

- Strukturierte Weiterbildung und zusätzliche Qualifikation in **Actuarial Data Scientist** soll einerseits die neuesten spezifischen Anforderungen im Bereich Data Science für Aktuare erlernbar machen ...
- ... und andererseits über den Titel / Brand des CADS der Öffentlichkeit den Anspruch der Aktuare deutlich machen, diese „neuen“ Felder auch über die aktuariellen Kernfelder hinaus zu besetzen.
- Bereits seit 2019 Teilnahme von Aktuarinnen und Aktuaren an den Seminaren **ADS Basic** und **ADS Advanced**.
- Seit 2021 Seminare **ADS Immersion** (Februar, März) und **ADS Completion** (September 2021)
- Beschluss der Einführung des Titels bei der Mitgliederversammlung im November 2020

„Certified Actuarial Data Scientist (CADS)“

Das Spezialwissen

Tabelle 1: Die Fächer des Spezialwissens

Fach	Basic	Advanced
ADS Grundlagen & Umfeld	- Grundlagen - Digitalisierung - Gesellschaftliches Umfeld und Ethik 1 - Datenschutz 1	- Gesellschaftliches Umfeld und Ethik 2 - Datenschutz 2
Informationstechnologie	- Datenmanagement 1 - Datenverarbeitungstechnologien 1 - Informationsverarbeitung in Versicherungen	- Datenmanagement 2 - Datenverarbeitungstechnologien 2 - Cloud Computing
Mathematik / Statistik	- Regressions- und Clustermethoden 1 - Datenvisualisierung - Datenaufbereitung zur Modellerstellung 1 - Modellselektion und Regularisierung 1	- Regressions- und Clustermethoden 2 - Datenaufbereitung zur Modellerstellung 2 - Modellselektion und Regularisierung 2
Insurance Analytics	- Data Mining 1 - Analytics 1 - Innovative Produkte 1	- Data Mining 2 - Analytics 2 - Innovative Produkte 2
Case Studies	Basic Case Studies	Advanced Case Studies

Die beiden Seminare geben einen Einblick in die zentralen Fragestellungen, Begriffe und Anwendungen von Data Science im Umfeld der Versicherungswirtschaft. Diese werden durch Anwendungen in den verschiedenen Sparten ergänzt.

Das Vertiefungswissen

Im Vertiefungswissen werden die Inhalte und Schwerpunkte des Spezialwissens erweitert und insbesondere das eigenständige Erarbeiten von Data Science Anwendungen und Lösungen erlernt.

Tabelle 2: Die Fächer des Vertiefungswissens

Fach	Immersion	Completion
ADS Umfeld	- Berufsständisches & Business Skills - Digitalisierung 2 - Data Science Community	
Informationstechnologie	- Informationstheoretische Grundlagen - Systemarchitekturen	- Entwicklungsmethoden - Kodierungstheorie
State-of-the-Art Tools & -Programme	- Einführung & Überblick - Spartenübergreifende Data Science Tools - Data Science z. B. mit R und Python	- Entwicklungsumgebungen für Data Science - Big Data Analytics z. B. mit Apache Spark
Mathematik / Statistik	- Überwachtes und unüberwachtes Lernen - Deep Learning 1: Grundlagen künstlicher neuronaler Netze	- Korrelation & kausale Inferenz - Deep Learning 2: Spezielle Anwendungsfälle tiefer neuronaler Netze
Insurance Analytics	- Data-Mining-Vertiefung - Visualisierung	- Textmining - Anomaly Detection - Interpretation (von Modellen und Ergebnissen)
Case Studies	eigenständige & interaktive Übungen aus verschiedenen Sparten	umfassende praktische Übungen

Überblick

Zielgruppe	DAV-Mitglieder, die ihre Kenntnisse im Bereich Data Science erweitern möchten			
Titel	Certified Actuarial Data Scientist (CADS)			
Prüfungen	Spezialwissen		Vertiefungswissen	
	Basic*	Advanced*	Immersion*	Completion
	Pflichtseminar (3 Tage)	Pflichtseminar (3 Tage)	Pflichtseminar (3 Tage)	Pflichtseminar (3 Tage)
	Klausur (180 Minuten)	Klausur (180 Minuten)	Projektarbeit	Projektarbeit
	* entfällt, wenn Spezialwissenfach ADS absolviert wurde; elementare Kenntnisse einer Data-Science-Programmiersprache werden vorausgesetzt			
Kosten	Die Kosten für die Seminare und Prüfungen entsprechen denen der DAV-Ausbildung.			



Einladung & Termine

Herzliche Einladung an Sie alle!

Seminare ADS Advanced

- im August & September 2021

Seminare ADS Completion

- im August & September 2021

Seminare ADS Basic

- Januar bis März 2022

Seminare ADS Immersion

- Februar & März 2022

Hinweis: Diese Seminare sind auch als Weiterbildungsseminare interessant!

Hinweis: Und wenn Sie dazu noch Lust auf ein spannendes Data Science-Projekt haben, wäre die Teilnahme an der Data Science Challenge 2021 für Sie interessant!

Dazu gibt es eine eigene Recorded Session!

Herzlichen Dank!



Haben Sie **Fragen**
oder **Anmerkungen**?

Dr. Stefan Nörtemann
msg life central europe gmbh
- Niederlassung Köln -
Tunisstraße 21
D-50667 Köln
E-mail: stefan.noertemann@msg-life.com

Weitere Informationen finden Sie auf www.akuar.de.



Informationen aus der AG Informatik/Tools

Dr. Axel Kaiser



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.



DGVFM

DEUTSCHE GESELLSCHAFT
FÜR VERSICHERUNGS-UND
FINANZMATHEMATIK e.V.

SIGNAL IDUNA Krankenversicherung a.G.

Mitglieder der AG

- Wolfgang Abele (stellvertretender Leiter)
- Carina Götzen
- Andreas Grotefeld
- Christof Husenbeth
- Dr. Axel Kaiser (Leiter)
- Peter Kauker
- Friedrich Loser
- Markus Maier
- Mathias Ott
- Norbert Quapp

Steckbriefe

- Die bisherige Hauptaktivität bestand in der Konzeption und Erstellung der ersten Steckbriefe.
- Sie finden sie im *Internen Bereich* der DAV-Homepage unter [Actuarial Data Science](#)

Komponenten eines Steckbriefes

Name des Tools

- Neben dem Namen des Tools kann hier auch angegeben sein, dass es es sich bei dem Steckbrief um eine Herstellerangabe handelt

Allgemeine Eigenschaften

Die allgemeinen Eigenschaften können umfassen

- Informationen, ob es sich um eine Programmiersprache oder eine Toolbox handelt
- Angaben des Programmierparadigmas
- Historische Informationen oder Referenzen zu anderen Tools

Komponenten eines Steckbriefes

Lizenzmodell

- Im Data Science-Bereich finden sich viele Open Source-Tools, aber auch solche mit kommerzieller Lizenz spielen eine Rolle

Systemvoraussetzungen

- Angabe, auf welchen Systemen bzw. Betriebssystemen das Tools verfügbar ist

Entwicklungsumgebungen

- Gibt es erwähnenswerte Entwicklungsumgebungen für das Tool?

Komponenten eines Steckbriefes

Besondere Eigenschaften in Bezug auf ADS

- Wahrscheinlich der wichtigste Block im Steckbrief, in dem beschrieben wird, inwiefern sich das betrachtete Tool für Actuarial Data Science eignet
- Gibt es besondere Bibliotheken oder Utilities für das Tool?
- Existieren spezielle ADS-Anwendungsfälle für das Tool?

Code-Samples

- Verweise auf aus ADS-Sicht interessante Code- oder Anwendungsbeispiele, die idealerweise einen einfachen Einblick in die Nutzung des Tools bieten

Komponenten eines Steckbriefes

Kompatibilitäten

- Werden bekannte Systeme oder Umgebungen unterstützt?
- Existieren besondere Anbindungen an Datenbanken oder andere Datenquellen?

Vorzüge/Nachteile/Pitfalls

Bewusst relativ offen und subjektiv gehaltene Liste von wahrgenommen Vor- und Nachteilen sowie von Punkten, an denen man aufpassen sollte;
Beispiele

- Interpretersprache oder kompiliert; schnell oder langsam
- Beginnen Indizes bei 0 oder bei 1?

Komponenten eines Steckbriefes

ADS-Bibliotheken / -Funktionalitäten

- Informationen über ADS-spezifische Erweiterungen

Quellen

- Homepage der Anwendung
- zusätzliche Informationen zu dem Tool

Redaktionsteam

Aktuell Konstituierung eines dreiköpfigen Redaktionsteams

Aufgaben des Redaktionsteams

Erster Ansprechpartner für Fragen und Anmerkungen zu Steckbriefen;
insbesondere

- Begutachtung und Freigabe neuer Steckbriefe
- Rückmeldungen aus der Community
- Schiedsstelle bei Meinungsverschiedenheiten

Mitarbeit

Die Steckbriefe sollen eine praxisorientierte Orientierungshilfe bieten. Hierzu ist es wichtig, dass einerseits viele Steckbriefe verfügbar sind und insbesondere, dass Rückmeldungen aus der "community", also von Ihnen!, kommen.

Rückmeldungen können Anmerkungen zu und Kritik an vorhandenen Steckbriefen umfassen, aber auch das Erstellen neuer Steckbriefe oder auch Anregungen und Wünsche zu zu bearbeitenden Fragestellungen.

Wir vertrauen hier auf Ihre Phantasie und freuen uns über Emails an

ads@aktuar.de

Vielen Dank für Ihre Aufmerksamkeit!

Fragen? / Anmerkungen?

Kontaktdata: axel.kaiser@aktuar-hamburg.de

AG Daten und Datenschutz

Ein aktueller Einblick

Dariush Sadeghi-Yam
viadico GmbH



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.



DGVFM

DEUTSCHE GESELLSCHAFT
FÜR VERSICHERUNGS-UND
FINANZMATHEMATIK e.V.

e-Jahrestagung 2021, 30. April 2021

Die AG Daten und Datenschutz



Mitglieder

Christian Bökenheide
Thomas Budzyn
Bartek Maciaga
Tobias Renner
Dariush Sadeghi-Yam
Dr. Dirk Wehrmann
Barbara Winter

AG Daten und Datenschutz

„Ziel der Arbeitsgruppe Daten und Datenschutz ist es, sich mit der Erarbeitung von Grundlagen zur Datengewinnung und -aufbereitung sowie dem Datenschutz auseinanderzusetzen.“

Neben Gesetzen und Richtlinien wird sich die Arbeitsgruppe auch mit Fragen der Ethik im Umgang mit Daten befassen.“

- Der Aktuar 01/2019 -



... und die zugehörigen Pools

Anonymisierung / Pseudonymisierung

*Aufbereitung von Grundlagen zur
Anonymisierung und Pseudonymisierung*

Dr. Stefan Karrmann
Dariush Sadeghi-Yam
Olga Schäfer
Dr. Felix Spangenberg



Mitglieder

Analyse Use Cases im Hinblick auf den Datenschutz

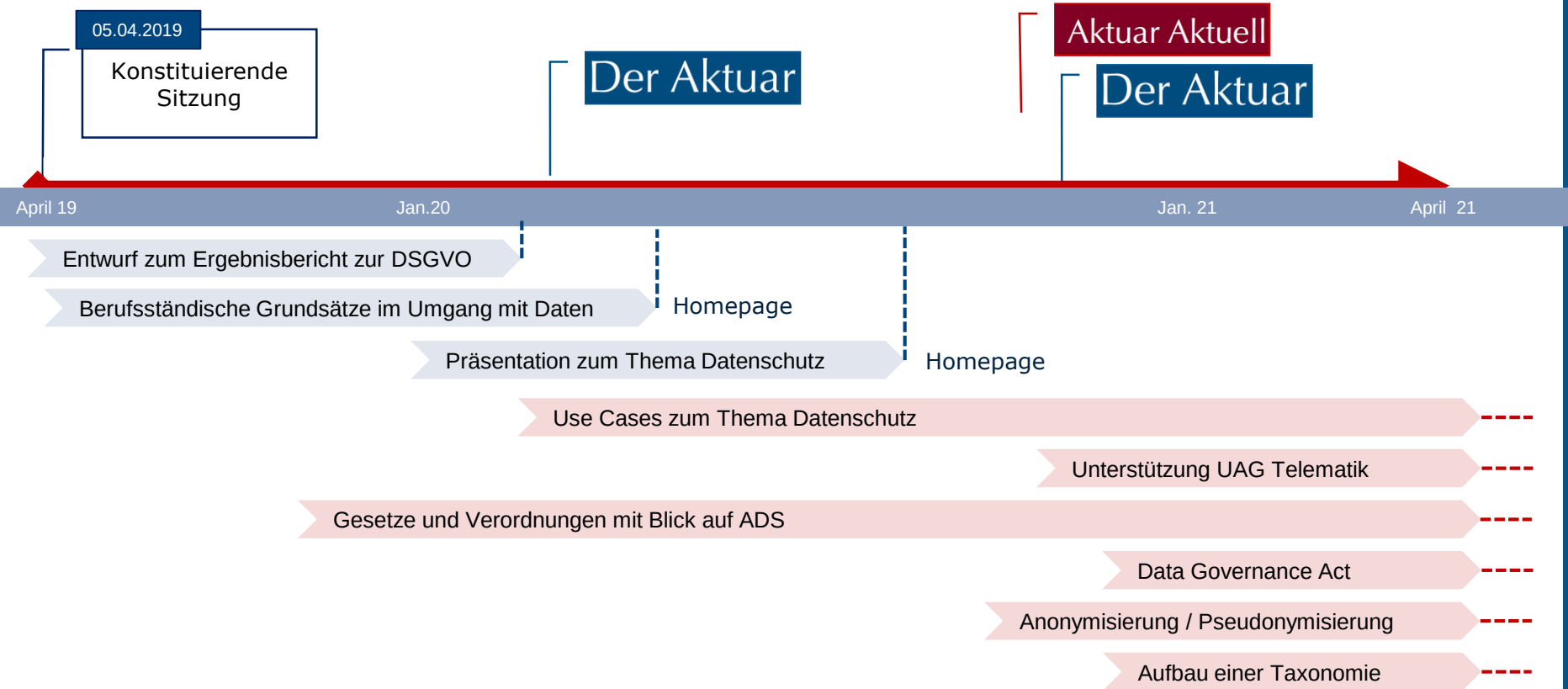
*Analyse von Use Cases aus verschiedenen
Sparten der Versicherungswirtschaft
hinsichtlich des Datenschutzes*

Dariush Sadeghi-Yam
Olga Schäfer
Dr. Christian Weiner

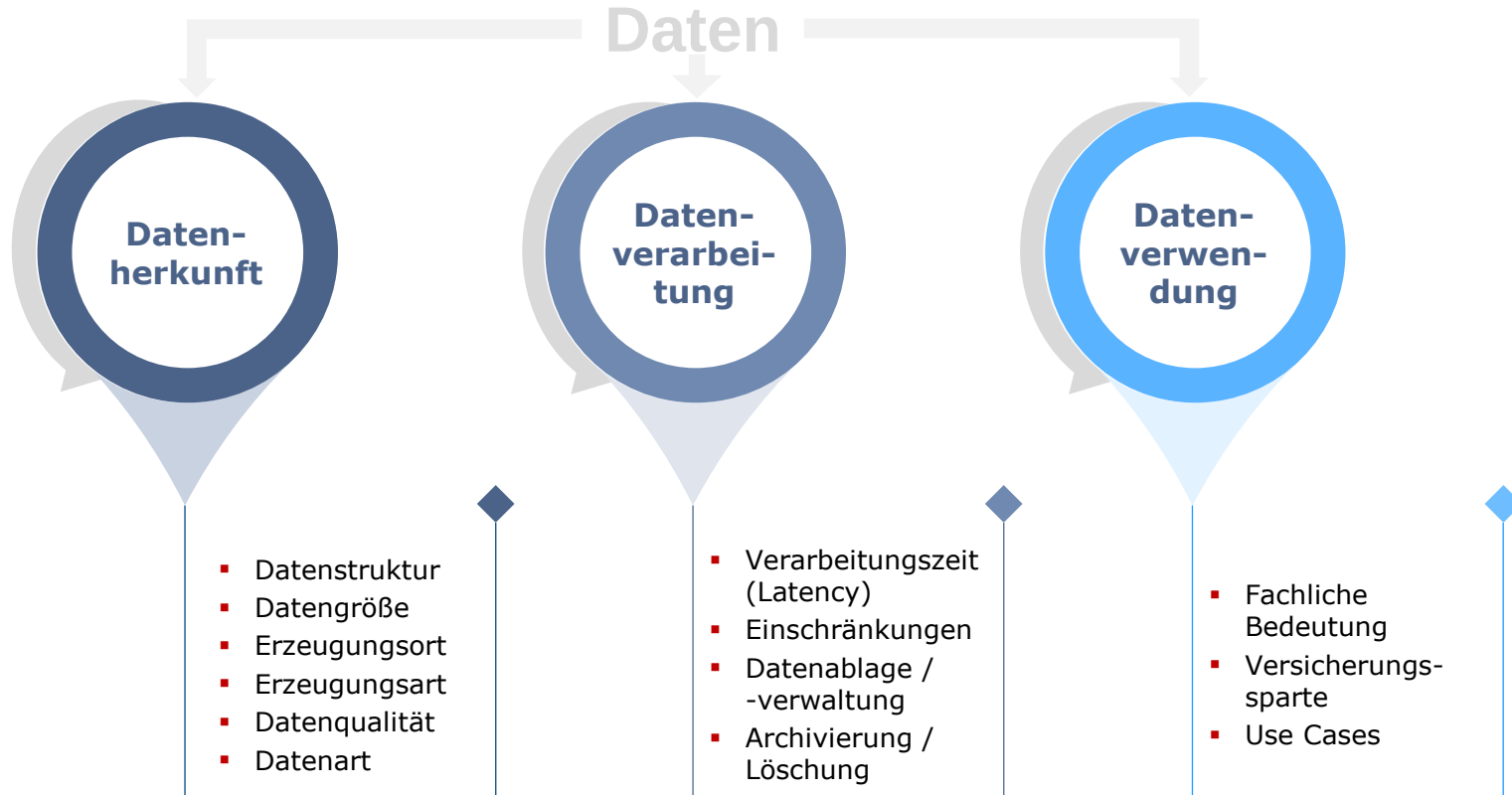
- ☐ Ergänzung von Ausbildungsunterlagen
- ☐ Erstellung von Schulungsunterlagen
- ☐ Veröffentlichung in der Zeitschrift „Der Aktuar“

- ☐ Veröffentlichung in der Zeitschrift „Der Aktuar“

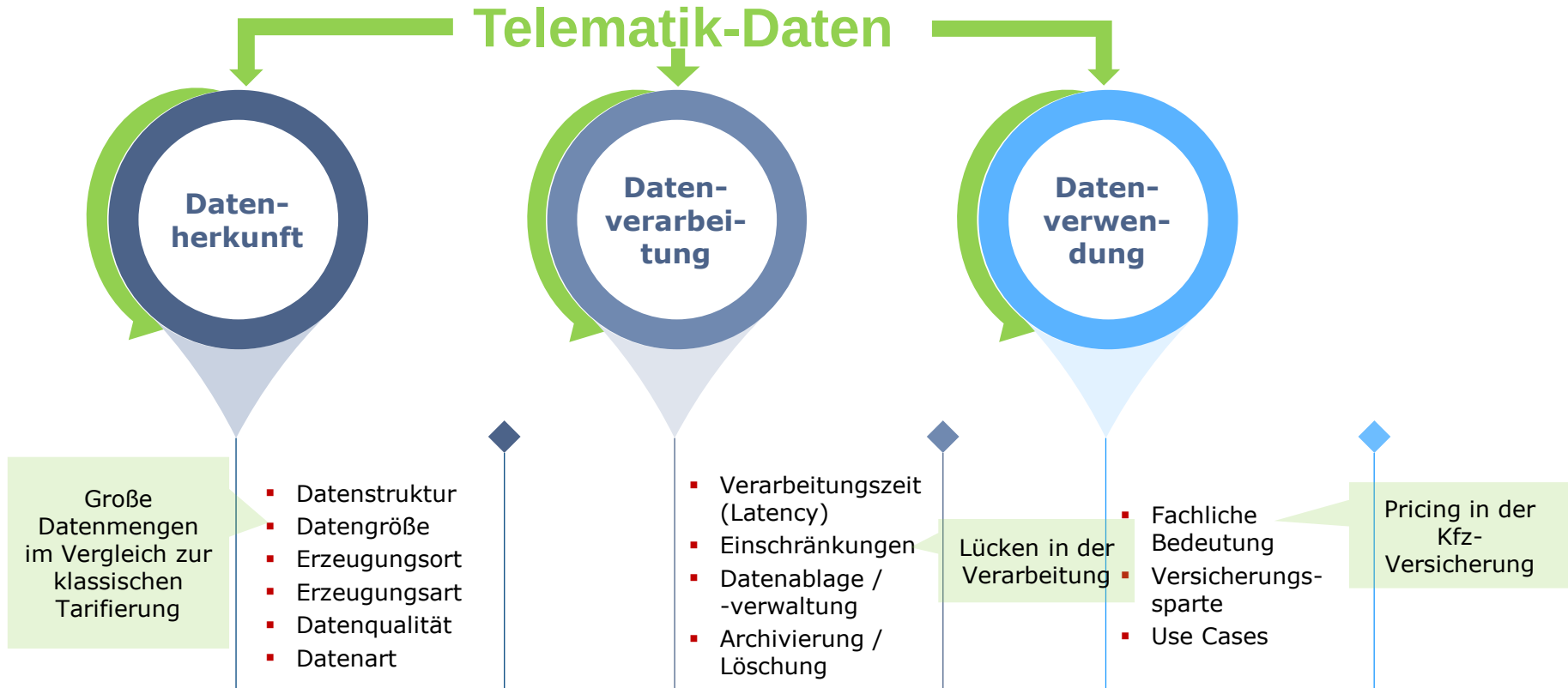
Wo kommen wir her, wo geht es hin?



Aufbau einer Taxonomie zu Daten



Aufbau einer Taxonomie zu Daten - Beispiel



Use Cases – Unser Analysemodell

Quellen

Vertragsdaten

Risikodaten

Personendaten

Bankdaten

Fitnessdaten

Schadendaten

Bonitätsdaten

Kundenkontakt-
historie

IOT-Daten

Social Media

...

Bewertungsdimensionen

Nutzen



Datenmenge



Sicherheit



Qualität



Verwaltungsanforderung



Ethik

DSGVO



I. Zweckbindung

II. Transparenz

III. Datenminimierung

IV. Speicherbegrenzung

V. Richtigkeit der Daten

VI. Rechenschaftspflicht

VII. Integrität und
Vertraulichkeit

Art der Datenverarbeitung

- ☒ Warum werden die Daten benötigt?
- ☒ Wie werden sie verwendet?
- ☒ Handelt es sich um einen besonderen Umstand?

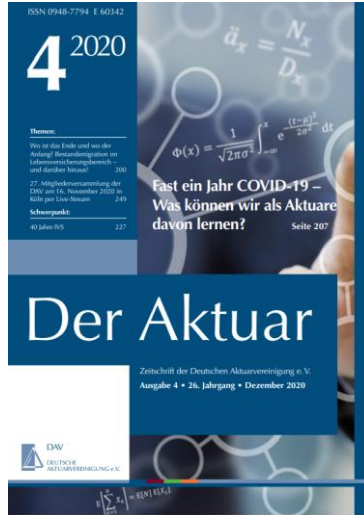
Datenkategorie

- ☒ Anonymisiert?
- ☒ Pseudonymisiert?
- ☒ Personenbezogen?
- ☒ Besondere Daten?

Rechtmäßigkeit / Dauer

- ☒ Einwilligung
- ☒ Vertragserfüllung
- ☒ Rechtliche Verpflichtung
- ☒ Berechtigtes Interesse
- ☒ Dauer des Erlaubnistatbestands

Use Cases – Beispiel



- Veröffentlichung in der „**Der Aktuar**“ (Ausgabe 4/2020)
- Titel: „**Anwendungsfall Daten und Datenschutz: Pricing in der Lebensversicherung**“
- Vorstellung der Analyse zu **Bankdaten** anhand der Bewertungsdimensionen
- Darstellung einer fiktiven Diskussion zwischen Datenschutzbeauftragten und Leitung der Produktentwicklung
- Aufgreifen verschiedener **Fragen**:
 - ? *Kann in der Produktentwicklung weitestgehend auf anonymisierte Daten des Banking-Partners ausgewichen werden?*
 - ? *An welcher Stelle wird Zugang zu sensiblen personenbezogenen Daten bei der Bank benötigt?*
 - ? *Wird eine Einwilligung der betroffenen Person benötigt?*
 - ? *Wie lange müssen die personenbezogenen Daten bearbeitet werden?*
 - ? *Müssen die Kunden über die Aufbewahrungsfristen aufgeklärt werden?*

Anonymisierung / Pseudonymisierung

Anonymisierung

- Verarbeitung personenbezogener Daten
- Personenbezug kann nicht oder nur unter unverhältnismäßigem Aufwand wiederhergestellt werden

Pseudonymisierung

- Verarbeitung personenbezogener Daten
- ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zuordenbar
- zusätzliche Informationen werden gesondert aufbewahrt und unterliegen technischen und organisatorischen Maßnahmen

Kennzahlen

- k-Anonymität
- l-Diversität
- t-Ähnlichkeit
- Differentieller Datenschutz

Methoden

- Maskierung (z.B. Ersetzung, Nichtangabe)
- Varianzmethode
- Mischung
- Kryptographische Methoden

Risiko der Re-Identifizierung

- Identifikation von Personen, z.B. durch Verknüpfen von Zusatzinformationen
- Adversarial Machine Learning

Ein exemplarischer Einblick in die Anonymisierung

Fiktiver Bestand

PLZ	Alter	Geschlecht	Diagnose
81223	35	Männlich	Demenz
81227	31	Weiblich	Demenz
81350	21	Männlich	Herzkrankheit
81788	19	Divers	Herzkrankheit
82966	61	Männlich	Demenz
82935	63	Männlich	Schizophrenie
82992	61	Weiblich	ADHS
82991	65		Depression
85123	59		Demenz
85127	57		Demenz
85198	54		Demenz
85288	51	Weiblich	Demenz

Quasi-
Identifikatoren

Sensitives
Attribut

Anonymisierter Bestand

PLZ	Alter	Geschlecht	Krankheit
81**	<40	*	Demenz
81**	<40	*	Demenz
81**	<40	*	Herzkrankheit
81**	<40	*	Herzkrankheit
82**	≥60	*	Demenz
82**	≥60	*	Schizophrenie
82**	≥60	*	ADHS
82**	≥60	*	Depression
85**	5*	*	Demenz
85**	5*	*	Demenz
85**	5*	*	Demenz
85**	5*	*	Demenz

3
Äquivalenz-
klassen

4-Divers,
aber alles
psychische
Krankheiten

Gruppieren

Nichtangabe

⇒ **4-Anonym**



Vielen Dank für Ihre
Aufmerksamkeit!

Machine-Learning-Anwendungsfälle im Aktuariat

Zoran Nikolić
AG Statistische Methoden ADS



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.



DGVFM

DEUTSCHE GESELLSCHAFT
FÜR VERSICHERUNGS-UND
FINANZMATHEMATIK e.V.

ADS Fachgruppe: Machine-Learning im Aktuariat
30. April 2021

Deloitte.



Agenda

- **Einführung:** Arbeitsgruppe Statistische Methoden ADS
- **Anwendungsfall 1:** GLM, Deep Learning und Gradient Boosting in der Schadentarifizierung, Teil 1: Schadenhäufigkeit
- **Anwendungsfall 2:** Use (this Solvency II) case! Neuronale Netze treffen auf Least Squares Monte Carlo
- **Anwendungsfall 3:** Neuronale Netze treffen auf Mortalitätsprognose
- Was beschäftigt derzeit die AG „Statistische Methoden“: **Neue Anwendungsfälle**
- **Zusammenfassung**



Einführung: Arbeitsgruppe Statistische Methoden ADS

Arbeitsgruppe Statistische Methoden



Mitglieder:

Carina Götzen
Stephanie Heitmüller
Martin Hüttemann
Christian Jonen
Florian Ketterer
Friedrich Loser
Sven Rehmann
Fabian Transchel
Corinna Walk



Hauptziel:

Ausarbeitung und
Veröffentlichung von
**konkreten
aktuariellen
Anwendungsfällen**
des maschinellen
Lernens in Form von
**Notebooks und
Daten**



Heute: Kurze

Vorstellung der **drei**
bereits veröffentlichten
Anwendungsfälle,
keine detaillierte
Beschreibung einzelner
Machine-Learning-
Methoden



Anwendungsfall 1: GLM, Deep Learning und Gradient Boosting in der Schadentarifizierung, Teil 1: Schadenhäufigkeit

ADS-Use-Case 1: Zusammenarbeit Ausschuss Schaden

- Veröffentlicht von **Friedrich Loser** (Techniker Krankenkasse) und **Daniel König** (Gen Re Consulting)
- Quelle für die Slides: Vortrag Daniel König max.99 in September 2020
- **Datenquelle** Use-Case: R-Package CASdatasets, freMTPL2freq
- **Schadenanzahl** Autohaftpflicht Frankreich
- Insgesamt 678.013 Policen
- **Neun Merkmale**, davon 2×Fahrer, 3×Region, 4×Fahrzeug
- Gewichtung: Exposure

GLM, Deep Learning und Gradient Boosting in der
Schadentarifizierung, Teil 1

Ausschuss Actuarial Data Science in Zusammenarbeit mit der Arbeitsgruppe
Tarifizierungsmethodik des Ausschusses Schadenversicherung, 23.04.2020

ADS-Use-Case 1: Jupyter-Notebook auf GitHub

Die erste Veröffentlichung auf der GitHub-Seite der DAV:

<https://github.com/DeutscheAktuarvereinigung/>

Table of Contents

1. [Data Exploration and Preparations](#)
2. [Classics: Generalized Linear Models \(GLM\)](#)
3. [Classics, Mixed: GLM and GLMM](#)
4. [Machine learning improved GLMs](#)
5. [Cross Validation and Boxplots](#)
6. [Deep Learning Approach with Embeddings and the "CANN"](#)
7. [Gradient Boosting Approach with XGBoost](#)
8. [Model Comparison, Monotonic Constraints and Remarks](#)
9. [Results and Summary](#)

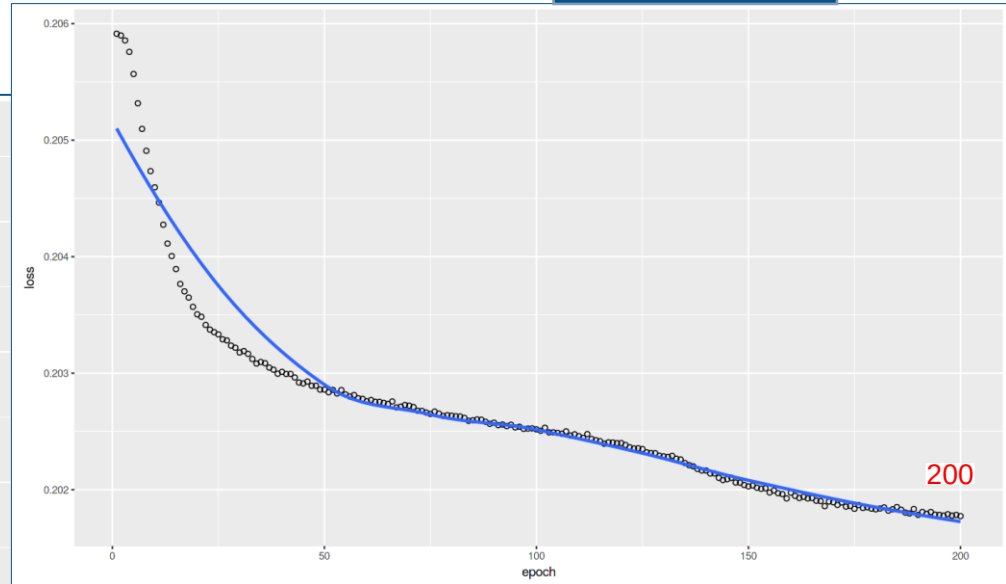
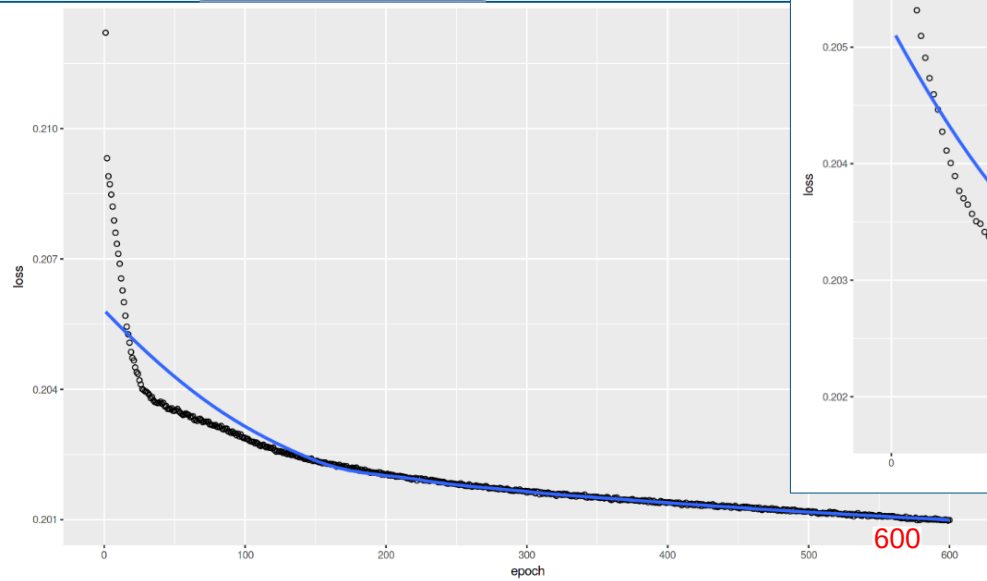
9

Beispiel: Performance

Das **Combined Actuarial Neural Net** „NNGLM“ konvergiert dank der „GLM skip connection“ wesentlich schneller.

NNGLM

NNemb





Anwendungsfall 2: Use (this Solvency II) case! Neuronale Netze treffen auf Least Squares Monte Carlo

ADS-Use-Case 2: Use (this) Solvency II Case!



Ein weiterer Anwendungsfall der DAV, die Aufgabenstellung stammt aus dem **Risikomanagement-Bereich**



Fokus: Approximation der Own Funds nach Solvency II für Lebens- und Krankenversicherungen

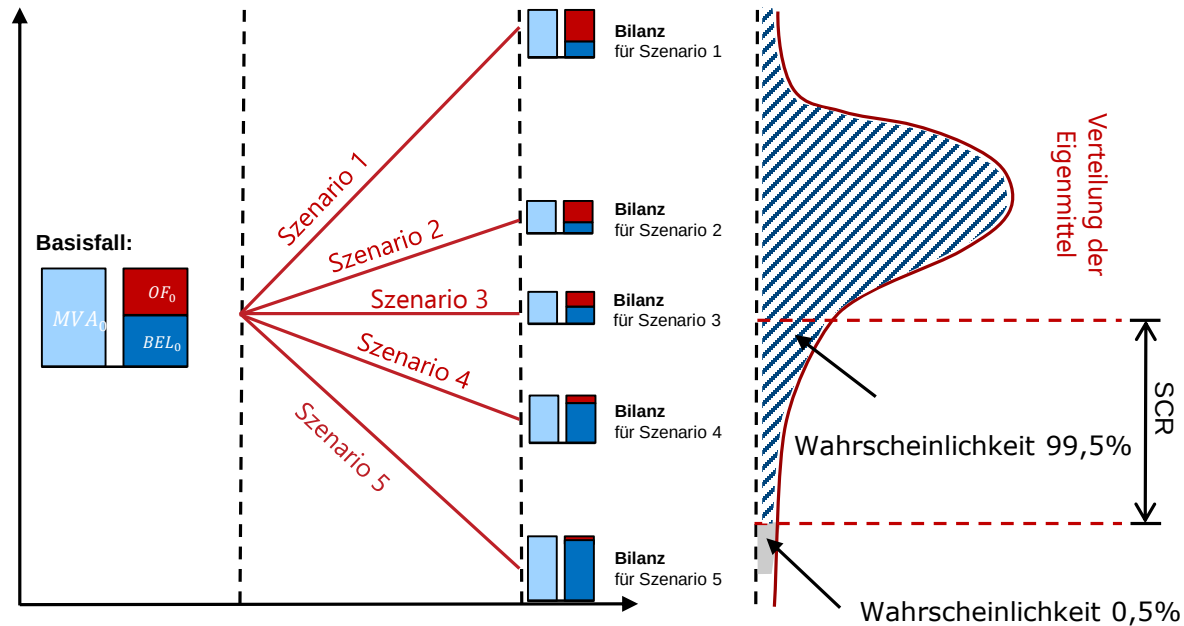


Insgesamt 13 bzw. 12 Risikofaktoren als Merkmale für das Training eines Machine-Learning-Modells; es gibt lediglich eine Output-Größe: Own Funds bzw. Eigenmittel



Quelle für die Folien: Vorträge **Christian Jonen** und **Tamino Meyhöfer** bei max.99 im September 2020 sowie Herbsttagung 2020

Motivation: SCR-Berechnung im Internen Modell

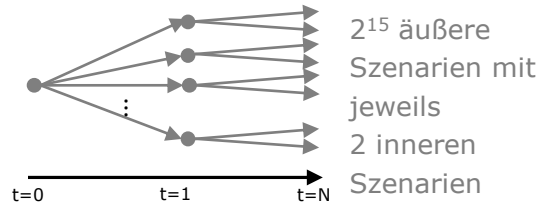


Hoher Rechenaufwand, Proxy-Funktionen erforderlich

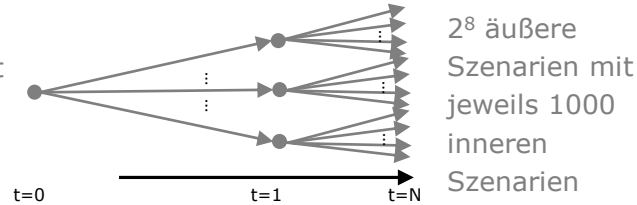
- Insgesamt **12 bzw. 13 Risikofaktoren** (z.B. Langlebigkeit, Storno, Zins, Aktien,...) im einjährigen Zeithorizont, diese dienen als Input-Größen bzw. als Features (Merkmale)
- Die Bestimmung der Positionen der Marktwertbilanz erfordert aufgrund komplexer Abhängigkeiten und Asymmetrien umfangreiche stochastische Simulationen von Projektionsmodellen
 - **Monte Carlo Auswertung**
- Hieraus resultiert ein „Nested Stochastic“ Problem mit explodierendem Rechenaufwand (Simulation in einer Simulation)
- Um dies zu umgehen, werden **Proxy-Funktionen** hergeleitet

Der Datensatz

1. Das Trainingsset

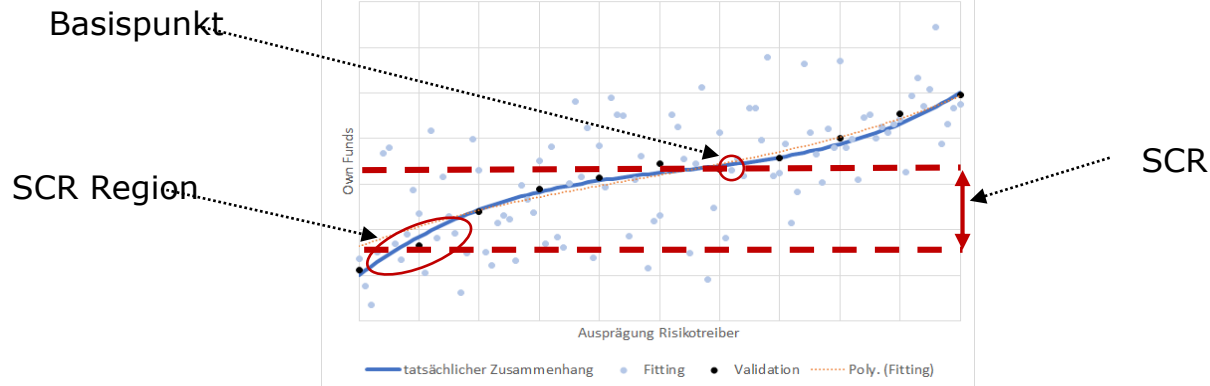


2. Das Validierungsset



3. Basis und die VaR-Region

- Basispunkt mit 16k Simulationen
- 129 bzw. 50 äußere Szenarien rund um das 0.5%-Quantilsszenario (SCR Region) mit jeweils 4k Simulationen





Anwendungsfall 3: Neuronale Netze treffen auf Mortalitätsprognose

ADS-Use-Case 3: NN treffen auf Mortalitätsprognose

- ADS-Use-Case zur Vorhersage der Sterblichkeit mithilfe von neuronalen Netzen
- **Fachlicher Hintergrund:**
 - Historische Sterblichkeiten als Datenbasis
 - Unterscheidung nach Jahr, Land, Geschlecht, Alter
 - Auswahl von neun Ländern für die gemeinsame Prognose
- Die im Anwendungsfall eingesetzten feedforward neuronalen Netze erreichen **noch kein zufriedenstellendes Ergebnis:**
 - Niedrige Stabilität bzgl. der variierten Hyperparameter
 - Güte der Vorhersage wie beim klassischen Lee-Carter-Modell
- **Quelle:** Vortrag von Corinna Walk und Huifeng Li, max.99 in 2020

Jupyter-Notebook

- Auf der DAV-Seite* ist ein Link zum **GitHub-Konto der DAV** mit dem Notebook und den Daten
- Das **Notebook**: „DAV Use Case Mortality Modeling-Final_V3.ipynb“ enthält die Einleseroutinen, den Code zum Training eines neuronalen Netzes einschließlich Hyperparameter-Auswahl, das Fitting des Lee-Carter-Modells sowie einige Visualisierungen und Vergleiche zwischen den neuronalen Netzen und dem Lee-Carter-Modell
- Die **Daten**: Die Datei „mortality.csv“ enthält die aufbereiteten Sterblichkeitsdaten für neun Länder

* <https://aktuar.de/unsere-themen/big-data/anwendungsfaelle/Seiten/anwendungsfall3.aspx>

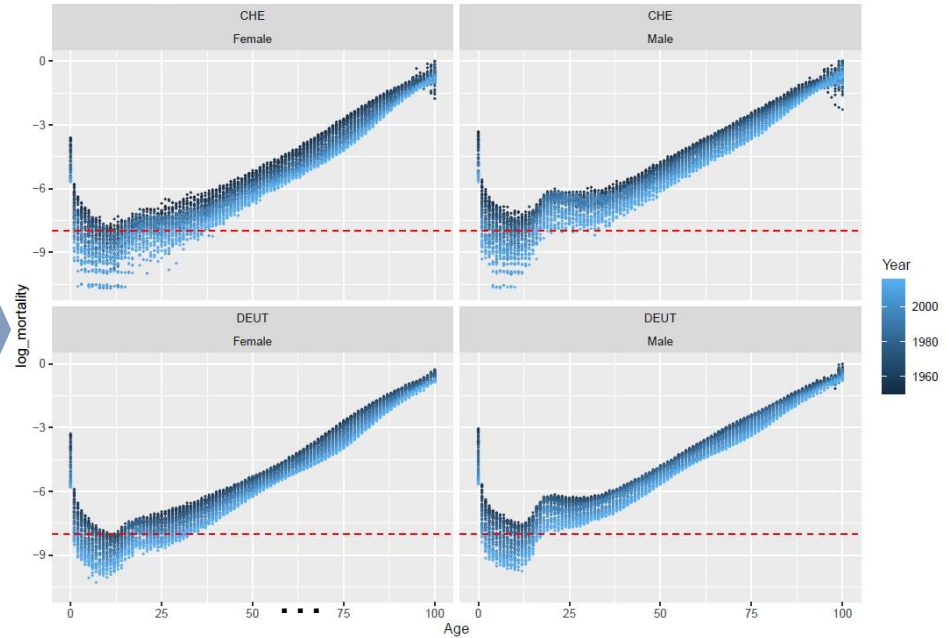
Datenbasis

Human Mortality Database (HMD)¹

- Bestands- und Todesfallzahlen für 41 Länder
- Aufgetrennt nach Geschlecht, Alter, Jahr, Land



Wandel der Log-Mortalitätsraten von 1950 bis 2016 in Deutschland und Schweiz



¹ <https://www.mortality.org/>

Datenaufbereitung: Von HMD zu mortality.csv

Berechnung logarithmierten Mortalitätsraten

- Im Alter von 0 bis 100 Jahre
- Von 1956 bis 2016
- Für Männer und Frauen getrennt

Sonderfall Deutschland

- Daten für Deutschland Gesamt erst ab 1990
- Daten liegen ab 1956 getrennt für Ost und West vor
- Zusammenführung der Daten zurück bis 1956
- Aber auch getrennte Betrachtung Ost/West bis 2016 möglich

Weitere Aufbereitungen

- Fehlende Werte durch Mittelwert ersetzt
- Sterblichkeit > 1 (Artefakte bei sehr kleinen Zahlen) durch 1 ersetzt

Insgesamt 143.218 Datenpunkte:
9 verschiedene Länder¹, getrennt nach Geschlecht
→ Vorhersage für 18 Kombinationen

¹ Deutschland, Schweiz, Polen, Dänemark, Frankreich, Spanien, Italien, USA, Japan



Was beschäftigt derzeit die AG „Statistische Methoden“: Neue Anwendungsfälle

Dimensionsreduktion

- Dimensionsreduktion häufig Vorarbeit für Machine-Learning-Modelle
 - Denn in der Praxis oft zu viele mögliche Einflussparameter
 - Unser Vorhaben: **ICD-10-Codes**

ICD-10-GM-2021

Kapitel I	A00-B99	Bestimmte infektiöse und parasitäre Krankheiten
Kapitel II	C00-D48	Neubildungen
Kapitel III	D50-D90	Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems
Kapitel IV	E00-E90	Endokrine, Ernährungs- und Stoffwechselkrankheiten
Kapitel V	F00-F99	Psychische und Verhaltensstörungen
Kapitel VI	G00-G99	Krankheiten des Nervensystems
Kapitel VII	H00-H59	Krankheiten des Auges und der Augenanhangsgebilde
Kapitel VIII	H60-H95	Krankheiten des Ohres und des Warzenfortsatzes
Kapitel IX	I00-I99	Krankheiten des Kreislaufsystems
Kapitel X	J00-J99	Krankheiten des Atmungssystems
Kapitel XI	K00-K93	Krankheiten des Verdauungssystems
Kapitel XII	L00-L99	Krankheiten der Haut und der Unterhaut
Kapitel XIII	M00-M99	Krankheiten des Muskel-Skelett-Systems und des Bindegewebes

Einzelschadenreservierung

- Traditionell werden **Schadendreiecke** für die Reservierung eingesetzt
 - Verwendung von Zahlungs- und Aufwandsinformationen
 - Stark aggregierte Dreiecksdaten
 - Anpassungen mit „Expert Judgement“, d. h. „nach Gefühl“
- Unser Plan: **Machine-Learning-Methoden** sollen ermöglichen, die Reservebeträge auf **Einzelpolicenebene** zu schätzen
 - Heranziehung weiterer Merkmale
 - Allerdings: Schwierig echte Unternehmensdaten für die geplante AG-Veröffentlichung zu erhalten
- Zusammenarbeit mit der **AG Schadenreservierung** des Ausschusses Schadenversicherung

Approximation künftiger Zahlungsströme

- Im Anwendungsfall 2 Approximation von Barwerten (Own Funds)
 - Jährliche Zahlungsströme nicht enthalten
- **Unser Vorhaben:** Approximation der künftigen Zahlungsströme aus einem Projektionstool mit neuronalen Netzen
 - Jährliche Schätzung von Beiträgen, Leistungen, Kosten usw.
- **Erklärende Variablen:** Grundsätzlich jeder der vielen Inputs der Projektionstools
 - Z. B. Kosten- oder Sterblichkeitsannahmen
 - Oder jährliche Realisierungen von stochastischen Kapitalmarktpfaden
- **Hauptproblem:** Verfügbarkeit von realistischen Trainingsdaten
 - Einfache LV-Modelle nicht realistisch
 - Anonymisierung von realistischen Modellen mühsam



Zusammenfassung

Zusammenfassung

- AG Statistische Methoden hat sich vorgenommen, konkrete **Notebooks** mit **Code** und den zugehörigen **Daten** zu veröffentlichen
- Bereits **drei veröffentlichte Anwendungsfälle**, viele weitere geplant
- Die ML-Methoden **häufig** aber **nicht immer besser**
 - „The Master Algorithm“ (P. Domingos) noch nicht gefunden
- Machine-Learning bringt Fortschritte und ermöglicht Anwendungen in **allen Sparten und allen Tätigkeitsgebieten im Aktuariat**
- Grundsätzlich: Steht man vor einer **Regressionsaufgabe**, kann man maschinelles Lernen anstelle der traditionellen Techniken anwenden
- Anwendung in der Regel **schnell** und **aufwandsarm** möglich:
 - Performante Libraries stehen kostenlos zur Verfügung



Vielen Dank!