

Prüfungsordnung 5.0

Lernziele im Spezialwissen *Actuarial Data Science Advanced*

Inhalt

1	ADS Grundlagen & Umfeld	2
1.1	Gesellschaftliches Umfeld & Ethik 2.....	2
1.2	Datenschutz 2.....	2
2	Informationstechnologie.....	3
2.1	Datenmanagement 2	3
3	Insurance Analytics	3
3.1	Data Mining 2.....	3
3.2	Analytics 2	4
3.3	Visualisierung 1	4
3.4	Innovative Produkte 1	5
4	Mathematik / Statistik.....	6
4.1	Überwachtes Lernen 2.....	6
4.2	Deep Learning 2	6
4.3	Korrelation & Kausale Inferenz.....	7
4.4	Datenaufbereitung zur Modellerstellung.....	7
5	Tools & Programme	8
5.1	Programmiersprachen für Data Science	8
6	Use Cases.....	8
6.1	Use Case.....	8

1 ADS Grundlagen & Umfeld

1.1 Gesellschaftliches Umfeld & Ethik 2

Zielsetzung: Die Kandidaten kennen die wichtigsten Reputationsrisiken im Zusammenhang mit der Anwendung von Data Science und sind mit Gefahren für das Geschäftsmodell der Versicherungen vertraut. Sie kennen die berufsständischen Grundsätze der DAV im Bereich Data Science und sind mit den regulatorischen Anforderungen für den KI-Einsatz vertraut.

- 1.1.1. Erläutern Sie den Begriff der Disruption anhand von Beispielen aus anderen Branchen. **(B2)**
- 1.1.2. Analysieren und diskutieren Sie konkrete Disruptionsgefahren für die Versicherungsbranche. **(C4)**
- 1.1.3. Benennen und erläutern Sie die wichtigsten berufsständischen Grundsätze der DAV im Bereich Data Science. **(B2)**
- 1.1.4. Analysieren Sie für konkrete Anwendungen im Bereich Data Science die Konformität mit den berufsständischen Grundsätzen der DAV. **(B4)**
- 1.1.5. Erläutern Sie die regulatorischen Anforderungen, die sich aus den einschlägigen europäischen Verordnungen, wie zum Beispiel dem Artificial Intelligence Act, für die Versicherungsbranche ergeben und wenden diese auf konkrete Anwendungsfälle an. **(C4)**

1.2 Datenschutz 2

Zielsetzung: Die Kandidaten sind mit den Grundprinzipien der zentralen Rechtsquellen für den Datenschutz in Deutschland und der Europäischen Union vertraut und können diese fallweise anwenden. Die Inhomogenität der Datenschutzregelungen weltweit ist bekannt und die Auswirkungen auf international tätige Unternehmen können eingeschätzt werden.

- 1.2.1 Erläutern Sie die Stellung und Umsetzung des Datenschutzes im Versicherungswesen im weltweiten Vergleich. Geben Sie Beispiele für unterschiedliche Datenschutzbestimmungen an. **(B2)**
- 1.2.2 Beurteilen Sie die möglichen Auswirkungen für international tätige Versicherungsunternehmen. **(C5)**
- 1.2.3 Benennen und erläutern Sie die Grundprinzipien der zentralen Rechtsquellen für den Datenschutz in Versicherungsunternehmen in Deutschland und der Europäischen Union. **(B2)**
- 1.2.4 Entscheiden Sie an konkreten Business Cases, ob sie erlaubt oder verboten sind bzw. welche Maßnahmen zu ergreifen sind, damit diese datenschutzkonform umgesetzt werden können. **(B5)**

2 Informationstechnologie

2.1 Datenmanagement 2

Zielsetzung: Die Kandidaten sind mit dem Konzept von relationalen Datenbanken und Grundzügen der Structured Query Language (SQL) vertraut. Sie kennen wichtige Typen/Vertreter aus dem Bereich der NoSQL-Datenbanken. Sie verstehen die Notwendigkeit dispositiver Datenhaltung für Analysezwecke und kennen die Konzepte Datawarehouse und Data-Lake.

- 2.1.1 Erklären Sie die gebräuchlichsten Konzepte aus dem Bereich der relationalen Datenbanksysteme und wenden Sie diese in Beispielfällen an. **(B3)**
- 2.1.2 Interpretieren Sie ein Entity-Relationship-Diagramm. **(B2)**
- 2.1.3 Unterscheiden Sie zwischen den verschiedenen Typen von SQL-Statements, erklären Sie grundlegende Operationen und wenden Sie diese auf Beispielfälle an. **(B3)**
- 2.1.4 Erläutern Sie das Konzept der Transaktion sowie die Grundprinzipien der ACID-Compliance. **(C2)**
- 2.1.5 Definieren Sie die Begriffe operative und dispositive Datenhaltung und erläutern Sie, warum beides im Versicherungsumfeld benötigt wird. **(A2)**
- 2.1.6 Erklären Sie Aufbau und Struktur eines Data-Warehouse. Diskutieren Sie Vor- und Nachteile im Vergleich zur traditionellen Normalform in relationalen Datenbankverwaltungssystemen. **(B2)**
- 2.1.7 Erläutern Sie die Prinzipien des CAP-Theorems. **(C2)**
- 2.1.8 Erklären Sie weitere wichtige Typen von Datenbanken und diskutieren Sie spezifische Einsatzmöglichkeiten. **(B2)**
- 2.1.9 Erklären Sie das Konzept des Data Lakes. Grenzen Sie es von traditionellen relationalen Datenbankverwaltungssystemen ab (schema on read vs. schema on write) und ziehen Sie Verbindungen zu NoSQL-Technologien. **(B3)**
- 2.1.10 Erklären Sie das Konzept von Distributed File Systems z. B. anhand des Hadoop Distributed File Systems. **(B2)**

3 Insurance Analytics

3.1 Data Mining 2

Zielsetzung: Die Kandidaten sind mit dem Data Mining-Prozess und den Ergebnissen vertraut. Sie können die Ergebnisse des Data Minings interpretieren und anwenden und Daten innerhalb des Data Mining-Prozesses visualisieren.

- 3.1.1 Benennen Sie verschiedene Prozess-Modelle des Data Minings und skizzieren Sie die einzelnen Schritte der Prozess-Modelle. **(B1)**
- 3.1.2 Erläutern Sie die Schritte und Abläufe der Prozess-Modelle des Data Minings. Übertragen Sie die Schritte des Data Mining-Prozesses auf die Prozessabläufe typischer Anwendungen im Versicherungsumfeld. **(B3)**
- 3.1.3 Beschreiben Sie, wie Ergebnisse des Data Minings interpretiert und evaluiert werden können, und erläutern Sie dies anhand konkreter Anwendungen in der Versicherung. **(C2)**

- 3.1.4 Erläutern Sie beispielhaft, wie Modellergebnisse in die operative Anwendung integriert werden können. **(C2)**

3.2 Analytics 2

Zielsetzung: Die Kandidaten sind mit (vertieften) Methoden zur Modellerstellung vertraut und können die Anwendbarkeit der Methoden und Modelle beurteilen. Sie können Verfahren zur Erkennung und Vermeidung von Overfitting anwenden.

- 3.2.1 Erläutern Sie die Funktionsweise, die Anwendungsmöglichkeiten und die praktische Anwendung von (Advanced-) Methoden des Maschinellen Lernens und des Statistischen Lernens. **(B2)**
- 3.2.2 Wenden Sie (Advanced-) Methoden des Maschinellen Lernens und des Statistischen Lernens auf Daten aus dem Versicherungsumfeld an. **(C3)**
- 3.2.3 Stellen Sie (Advanced-) Methoden des Maschinellen Lernens und des Statistischen Lernens gegenüber und erläutern Sie die Vor- und Nachteile der einzelnen Methoden hinsichtlich Anwendbarkeit, Voraussetzungen, Ergebnisse und Rechenaufwand. **(B4)**
- 3.2.4 Beurteilen Sie die Anwendbarkeit von Modellierungsmethoden auf aktuarielle Problemstellungen. **(B5)**
- 3.2.5 Erläutern Sie die Begriffe Underfitting, Overfitting, Generalisierung und Modellkomplexität. **(B2)**
- 3.2.6 Beschreiben Sie, wie Overfitting bei der Modellerstellung erkannt werden kann. **(C2)**
- 3.2.7 Erklären Sie Konzepte und praktische Anwendung von verschiedenen Methoden (wie z. B. Kreuzvalidierung, Prognose-Ist-Vergleich auf Validierungsdaten, Bootstrap-Stichproben, Parallelentwicklungen, Backtesting) zur Erkennung und Vermeidung von Overfitting. **(B2)**
- 3.2.8 Interpretieren Sie die Ergebnisse von „Fitting Graphen“ und „Learning Graphen“ bei der Erstellung und der Auswahl von Modellen. **(B4)**

3.3 Visualisierung 1

Zielsetzung: Aufbauend auf dem Grundwissenfach Angewandte Stochastik können die Kandidaten Werkzeuge der Datenvisualisierung einsetzen, um Informationen zum besseren Verständnis aufzubereiten und zu analysieren. Die Kandidaten können Methoden der Datenvisualisierung im Vorfeld einer Modellbildung zur Beurteilung der Datenqualität und zur Aufbereitung der Daten verwenden. Die Kandidaten können grafische Methoden zur Überprüfung der Modellgüte und Trennungseigenschaft von Modellen verwenden.

- 3.3.1 Wenden Sie graphische Methoden an, um Zusammenhänge zwischen Kovariablen und Zielvariablen zu erkennen (z. B. Mosaikplots, Effect Plots etc.) und erläutern Sie ihre Bedeutung für die Modellbildung (Variablenselektion, stetige vs. kategoriale Merkmale, Klassifikation/Gruppierung von Kovariablen). **(C3)**
- 3.3.2 Beurteilen Sie die Voraussetzungen und die Anpassungsgüte von Modellen hinsichtlich Schätzgüte und Trennungseigenschaft durch Anwendung von quantitativen (Interpretation von Kenngrößen, wie z. B. Bestimmtheitsmaße) und graphischen Methoden (z. B. Residuenplots, Lift-Kurven oder Lorenzkurven) auf Trainings- und Validierungsdaten. **(C3)**
- 3.3.3 Erläutern Sie grundlegende, explorative Verfahren zur grafischen Analyse von Kovariablen und Zielvariablen, die eine Einschätzung hinsichtlich extremer Werte, Ausreißern und falscher Werte ermöglichen. **(B2)**

3.4 Innovative Produkte 1

Zielsetzung: Die Kandidaten sind mit den Anforderungen der Datenerhebung und -verarbeitung von innovativen Produkten vertraut und kennen die Vorteile und Möglichkeiten von innovativen Produkten.

- 3.4.1 Erläutern Sie die Anwendung von Data Science in der Entwicklung von „klassischen“ Versicherungsprodukten in verschiedenen Versicherungssparten. **(C2)**
- 3.4.2 Beschreiben Sie die Bedeutung von Data Science für die Entwicklung von „innovativen“ Produkten, die über die Anwendung von Verfahren und Methoden der „klassischen“ Produktentwicklung hinausgehen. **(B2)**
- 3.4.3 Benennen Sie mögliche Vorteile von innovativen Versicherungsprodukten. **(B1)**
- 3.4.4 Erläutern Sie die Vorteile und Möglichkeiten der Anwendung von Data Science in der Produktentwicklung exemplarisch an verschiedenen Beispielen wie der Risikodifferenzierung. **(B2)**
- 3.4.5 Vergleichen Sie für unterschiedliche Datengrundlagen die Möglichkeit der Risikodifferenzierung und beurteilen Sie aufbauend darauf die Verbesserungsmöglichkeiten zur Selektion von Risiken. **(B5)**
- 3.4.6 Erläutern Sie den möglichen Einfluss von Kundensichtmodellen auf Design und Marketing von innovativen Produkten. **(B2)**
- 3.4.7 Benennen Sie unterschiedliche Datenquellen als Grundlage der Produktentwicklung und beschreiben Sie die Charakteristika der Daten. **(B2)**
- 3.4.8 Erläutern Sie die technischen Möglichkeiten und Anforderungen zur Erhebung und Verarbeitung von (individuellen) Risikodaten. Beurteilen Sie hierbei die Auswirkungen und Anforderungen der Verarbeitung von großen, heterogenen Datenmengen und heterogenen Daten auf den Produktentwicklungsprozess. **(B4)**
- 3.4.9 Beschreiben Sie die Anforderungen zur Integration und Verknüpfung von externen Daten mit (unternehmens-) internen Daten. Verknüpfen Sie externe mit internen Daten als Grundlage der Produktentwicklung eines Versicherungsunternehmens für exemplarische Anwendungsfälle. **(B3)**
- 3.4.10 Erläutern Sie die notwendigen Datengrundlagen und die prinzipielle Vorgehensweise zur Bildung einer spartenübergreifenden Kundensicht und zum Management der Kundenbeziehung. **(B2)**

4 Mathematik / Statistik

4.1 Überwachtes Lernen 2

Zielsetzung: Aufbauend auf den ADS-Basic-Modulen Überwachtes Lernen 1 und Unüberwachtes Lernen 1 vertiefen die Kandidaten ihre Kenntnisse in den Verfahren der linearen und verallgemeinerten linearen Modelle, die auch additiv oder gemischt sein können.

- 4.1.1 Beschreiben Sie die grundlegenden Konzepte der multiplen linearen Regression. **(B2)**
- 4.1.2 Erläutern Sie die Erweiterung der Regressionsverfahren zu den verallgemeinerten additiven Modellen mit Spline-Schätzfunktionen. **(B2)**
- 4.1.3 Nennen Sie Erweiterungen von Regressionsverfahren wie z. B. Generalized Estimating Equations, Mixed Models oder nichtparametrische Regressionsverfahren und deren Anwendungsmöglichkeiten. **(B1)**
- 4.1.4 Erläutern Sie die Schätz- und Testtheorie spezieller verallgemeinerter linearer Modelle, die in der Risikomodellierung und bei der Analyse dichotomer Zielgrößen zur Anwendung kommen. **(B2)**

4.2 Deep Learning 2

Zielsetzung: Die Kandidaten verstehen die erweiterten Konzepte neuronaler Netze und sind in der Lage, neuronale Netze auf Klassifikations- und Regressionsprobleme mit strukturierten Daten sowie auch Bildern anzuwenden.

- 4.2.1 Skizzieren Sie die Anwendung eines Konvolutions-Filters auf eine zweidimensionale Matrix. **(B2)**
- 4.2.2 Nennen und beschreiben Sie zwei Regularisierungsmethoden für tiefe neuronale Netze. **(B2)**
- 4.2.3 Beurteilen Sie, in welchen Anwendungsszenarien der Einsatz neuronaler Netze besonders vorteilhaft sein kann. **(C5)**
- 4.2.4 Gestalten Sie „Deep-Learning-Pipelines“ für Problemstellungen im Versicherungsumfeld. Welche Netzarchitektur ist geeignet? Kann ein vortrainiertes Modell verwendet werden (transfer learning)? Wie werden die Hyperparameter bestimmt? Wie verläuft das „Lernen“ der gegebenen Problemstellung? Wie lässt sich die Modellgüte beurteilen? **(C5)**

4.3 Korrelation & Kausale Inferenz

Zielsetzung: Die Kandidaten lernen den Unterschied zwischen rein empirisch beobachteten und tatsächlich kausalen Zusammenhängen kennen und können diese in Anwendungsbeispielen voneinander abgrenzen.

- 4.3.1 Nennen und beschreiben Sie verschiedene Techniken des probabilistischen Schließens. Geben Sie ein konkretes Anwendungsbeispiel einer der beiden Techniken in der Versicherungsmathematik. **(C2)**
- 4.3.2 Stellen Sie Probleme dar, die bei der Arbeit mit beobachteten Daten auftreten. **(B2)**
- 4.3.3 Erklären Sie die Begriffe „Survival Bias“, „Outcome Bias“, „Omitted-Variable Bias“ und „Alternative Blindness“ jeweils anhand eines konkreten Beispiels. **(C2)**
- 4.3.4 Grenzen Sie ausgehend von den philosophischen Grundlagen die Begriffe „Kausalität“ und „Korrelation“ voneinander ab. **(B2)**
- 4.3.5 Erläutern Sie die grundlegenden Prinzipien einer Kausalordnung und beschreiben Sie Grenzen der Kausalität. **(C2)**
- 4.3.6 Beschreiben Sie den modernen mathematischen Ansatz nach Pearl zur Beschreibung von Kausalität. **(C2)**
- 4.3.7 Erklären Sie den Begriff eines „Causal Experiments“ und nennen Sie Techniken, die zum Design eines solchen Experiments verwendet werden können. **(C2)**
- 4.3.8 Interpretieren und analysieren Sie die Modellierungsergebnisse der klassischen Verfahren. **(D2)**

4.4 Datenaufbereitung zur Modellerstellung

Zielsetzung: Die Kandidaten kennen die wesentlichen Techniken der Datenaufbereitung, die in der Modellbildung verwendet werden. Sie können die Notwendigkeit und die Funktionsweise der Techniken interpretieren und einschätzen.

- 4.4.1 Erläutern Sie Methoden zur Ersetzung von fehlenden oder falschen Werten und Vorgehensweisen bei der Modellbildung mit fehlenden Kovariablenwerten. **(B2)**
- 4.4.2 Beschreiben Sie das Konzept der Dummyvariablen und setzen Sie es in Anwendungsfällen um. **(C3)**
- 4.4.3 Interpretieren Sie die Ergebnisse von Datenvisualisierungen und beurteilen Sie, ob Datenableitungen (wie Klassenbildung, Interaktion, Missing-Kategorie) sinnvoll sind. **(C4)**
- 4.4.4 Beschreiben Sie die Funktionsweise und die praktische Anwendung der Selektion von informativen Attributen, unter der exemplarischen Erläuterung verschiedener Konzepte (wie z. B. AIC und BIC). **(B2)**
- 4.4.5 Erläutern Sie die gängigen Datentransformationen (polynomial, log, Box-Cox, etc.) und erkennen Sie, sofern eine Datentransformation für die Modellierung notwendig ist, welche dafür geeignet ist. **(C4)**
- 4.4.6 Nennen Sie die grundlegenden Eigenschaften und die Funktionsweise von Resampling-Methoden (z. B. Bootstrap) und beschreiben Sie deren Einsatz in den Verfahren der Regression und der Modellvalidierung (z. B. Kreuzvalidierung). **(B4)**

5 Tools & Programme

5.1 Programmiersprachen für Data Science

Zielsetzung: Die Kandidaten sind in der Lage, mit den gängigen Sprachen eigenständig Data Science-Projekte durchzuführen, und kennen deren wichtigste Sprachelemente und Bibliotheken.

- 5.1.1 Benennen und erläutern Sie die wichtigsten Sprachelemente von Python. **(B2)**
- 5.1.2 Benennen und erläutern Sie wichtige Methoden aus der Bibliothek scikit-learn. **(B2)**
- 5.1.3 Entwerfen Sie für eine Fragestellung im Versicherungsumfeld einen Analyseablauf des überwachten Lernens mit scikit-learn. **(C5)**
- 5.1.4 Benennen und erläutern Sie die wichtigsten Python-Bibliotheken zur Datenaufbereitung und Visualisierung. **(B2)**
- 5.1.5 Erstellen Sie Notebooks in Python zur Analyse von Datenbeständen aus dem Versicherungsumfeld unter Beachtung üblicher Programmierstandards (Wartung, Zuverlässigkeit, Effizienz, Benutzerfreundlichkeit). **(C5)**

6 Use Cases

6.1 Use Case

Zielsetzung: Die Kandidaten sind in der Lage, Data Science-Analysen sowie Anwendungen des Maschinellen Lernens selbstständig durchzuführen.

- 6.1.1 Basierend auf einer realistischen Fragestellung und einem gegebenen Datenbestand führen Sie eine Data Science-Analyse selbstständig durch. Dabei durchlaufen Sie alle Phasen eines Data Mining-Prozesses und dokumentieren den Prozess und das Ergebnis in geeigneter Form, z. B. in einem Notebook unter Beachtung üblicher Programmierstandards (Wartung, Zuverlässigkeit, Effizienz, Benutzerfreundlichkeit). Sie wenden Modelle an, interpretieren und beurteilen die Ergebnisse und präsentieren diese zielgruppengerecht. **(C5)**