

## Prüfungsordnung 5.0

### Lernziele im Spezialwissen *Actuarial Data Science Basic*

#### Inhalt

<b>1</b>	<b>ADS Grundlagen &amp; Umfeld .....</b>	<b>2</b>
1.1	Grundlagen .....	2
1.2	Digitalisierung .....	2
1.3	Gesellschaftliches Umfeld & Ethik 1 .....	3
1.4	Datenschutz 1 .....	3
<b>2</b>	<b>Informationstechnologie.....</b>	<b>4</b>
2.1	Datenmanagement 1 .....	4
2.2	Datenverarbeitungstechnologien 1 .....	4
2.3	Informationsverarbeitung in Versicherungen .....	4
<b>3</b>	<b>Insurance Analytics .....</b>	<b>5</b>
3.1	Data Mining 1 .....	5
3.2	Analytics 1 .....	6
<b>4</b>	<b>Mathematik / Statistik.....</b>	<b>7</b>
4.1	Überwachtes Lernen 1 .....	7
4.2	Unüberwachtes Lernen 1 .....	7
4.3	Deep Learning 1 .....	7
<b>5</b>	<b>Tools &amp; Programme .....</b>	<b>8</b>
5.1	Einführung und Überblick.....	8
5.2	Sprachübergreifende Data Science-Tools.....	8
5.3	Programmiersprachen für Data Science .....	8
<b>6</b>	<b>Use Cases.....</b>	<b>9</b>
6.1	Use Case .....	9

## 1 ADS Grundlagen & Umfeld

### 1.1 Grundlagen

Zielsetzung: Die Kandidaten sind mit den Grundbegriffen im Gebiet Actuarial Data Science vertraut und können die zugrundeliegenden Daten geeignet kategorisieren.

- 1.1.1. Erläutern Sie die zentralen Begriffe im Kontext Data Science, wie zum Beispiel Data Engineering, Data Mining, Digitalisierung, Big Data, Maschinelles Lernen sowie Künstliche Intelligenz, unter Berücksichtigung ihrer historischen Entwicklung und grenzen Sie diese voneinander ab. **(B2)**
- 1.1.2. Erklären Sie die Idee der Künstlichen Intelligenz. Diskutieren Sie allgemeine Einsatzmöglichkeiten in der Versicherungswirtschaft. **(B2)**
- 1.1.3. Erläutern Sie die Idee des Maschinellen Lernens sowie die grundlegenden Unterschiede zur klassischen Programmierung. **(C3)**
- 1.1.4. Erläutern Sie die wichtigsten Kategorien des Maschinellen Lernens und grenzen Sie diese voneinander ab. **(C3)**
- 1.1.5. Beschreiben Sie die Auswirkungen der Entwicklung von Actuarial Data Science auf die Versicherungswirtschaft insgesamt, auf ein Versicherungsunternehmen und auf das Berufsbild der Aktuarin / des Aktuars im Speziellen. **(C3)**
- 1.1.6. Erläutern Sie die Unterschiede von strukturierten, unstrukturierten und semistrukturierten Daten sowie die Bedeutung dieser Unterscheidung im Kontext von Data Science. **(B2)**

### 1.2 Digitalisierung

Zielsetzung: Die Kandidaten sind mit den zentralen Themen der Digitalisierung vertraut und können die Auswirkungen auf die Versicherungsbranche einschätzen und bewerten.

- 1.2.1. Benennen und erläutern Sie die zentralen Begriffe, Themen und Technologien im Kontext Digitalisierung im Versicherungsumfeld. **(B2)**
- 1.2.2. Benennen und erläutern Sie Einsatzmöglichkeiten von Data Science und Maschinellem Lernen in der Versicherung. Nennen Sie für jede Sparte Beispiele und erläutern Sie das Vorgehen. **(B2)**
- 1.2.3. Erläutern Sie die Auswirkungen und die Anforderungen der Digitalisierung auf die Anwendungslandschaft und die Datenverarbeitung in einem Versicherungsunternehmen. **(B2)**
- 1.2.4. Analysieren und diskutieren Sie anhand konkreter Beispiele die Auswirkungen der Digitalisierung auf das Geschäftsmodell und die Geschäftsprozesse der Versicherungsbranche und leiten Sie daraus mögliche Auswirkungen auf die Aufgaben der Aktuarin / des Aktuars ab. **(D5)**
- 1.2.5. Erklären Sie die Bedeutung und die Aufgaben von Verbänden (zum Beispiel GDV, DAV) im Umfeld der Digitalisierung und des Einsatzes Künstlicher Intelligenz. **(B2)**

### 1.3 Gesellschaftliches Umfeld & Ethik 1

Zielsetzung: Die Kandidaten kennen die wichtigsten Argumente und Vorbehalte gegen Data Science und sind für ethische Fragen im Kontext Data Science sensibilisiert.

- 1.3.1. Erläutern Sie anhand geeigneter Beispiele, wie sich die Anwendung von Data Science positiv oder negativ auf das Marketing auswirken kann. **(B2)**
- 1.3.2. Erläutern Sie anhand eines geeigneten Beispiels, inwiefern die Anwendung von Data Science ein Reputationsrisiko bergen kann. **(B2)**
- 1.3.3. Erläutern Sie die Systematik der Regulatorik im Zusammenhang mit dem Einsatz Künstlicher Intelligenz in der Versicherungsbranche. Nennen Sie die relevanten Rechtsquellen. **(B2)**
- 1.3.4. Erläutern Sie die ethischen Probleme, die sich durch Anwendungen von Verfahren der Künstlichen Intelligenz in der Versicherungswirtschaft ergeben können. Erläutern Sie darüber hinaus die möglichen Auswirkungen für die Gesellschaft. **(B2)**
- 1.3.5. Entscheiden Sie fallweise, ob eine Actuarial Data Science-Anwendung ethische Probleme in sich birgt und beurteilen Sie sie mit Blick auf regulatorische Anforderungen. **(B4)**

### 1.4 Datenschutz 1

Zielsetzung: Die Kandidaten sind mit der Rechtssystematik und den wichtigsten Regelungen der Datenschutzbestimmungen in Deutschland und der Europäischen Union vertraut und können diese auf einfache Fallbeispiele anwenden bzw. einschätzen, wann weitergehender juristischer Rat einzuholen ist.

- 1.4.1 Erläutern Sie die Rechtssystematik in Deutschland im Umfeld Datenschutz. Benennen und erläutern Sie die einschlägigen Rechtsquellen in Deutschland und der Europäischen Union. **(B2)**
- 1.4.2 Benennen und erläutern Sie die Prinzipien der Datenschutzgrundverordnung der EU und setzen Sie diese in Beziehung zu den anderen Rechtsquellen. **(B3)**
- 1.4.3 Benennen und erläutern Sie die zentralen Grundsätze des Code of Conduct-Datenschutzes für das Versicherungswesen. Wenden Sie diese Grundsätze auf die Verarbeitung von Daten in Versicherungsunternehmen an. **(B3)**
- 1.4.4 Wenden Sie die deutschen und europäischen Datenschutzbestimmungen in konkreten Fallbeispielen an. **(B3)**
- 1.4.5 Entscheiden Sie an konkreten Business Cases, ob sie verboten, erlaubt oder ein Grenzfall sind. **(B5)**

## 2 Informationstechnologie

### 2.1 Datenmanagement 1

Zielsetzung: Die Kandidaten kennen die wichtigsten Konzepte, die zur Speicherung von Daten in Filesystemen und zum Transfer verwendet werden. Sie können fallbezogen die Vor- und Nachteile von verschiedenen Optionen bewerten.

- 2.1.1 Geben Sie einen Überblick über die wichtigsten Aspekte und Technologien der Ablage von Daten und erläutern Sie deren Vor- und Nachteile. Diskutieren Sie die Anwendbarkeit der Technologien in verschiedenen Szenarien. **(B2)**
- 2.1.2 Nennen und definieren Sie die gebräuchlichsten (offenen) Datenformate zur Speicherung von Daten in Dateien und diskutieren Sie deren spezifische Einsatzmöglichkeiten. Wenden Sie Datentransformationen in konkreten Beispielfällen an. **(C3)**

### 2.2 Datenverarbeitungstechnologien 1

Zielsetzung: Die Kandidaten kennen wichtige Standards, Tools und Praktiken, die zu verlässlicher und effizienter Codeentwicklung im Analytics-Bereich beitragen.

- 2.2.1 Erklären Sie die Qualitätssicherungs- und die Dokumentationsaufgabe von Softwaretests und erläutern Sie das Paradigma des „Test Driven Development“. Beschreiben Sie, wie die Ausführung von Tests von Software-Plattformen unterstützt wird (z. B. JUnit, Test-Runner) und erarbeiten Sie Kriterien, die eine effektive Testbarkeit von Codes unterstützen. **(C2)**
- 2.2.2 Benennen Sie Kern-Funktionalitäten, die Data Science-Plattformen zur Verfügung stellen sollten. **(A1)**
- 2.2.3 Erklären Sie die wichtigsten Verarbeitungsschritte der Datenaufbereitung und demonstrieren Sie sie anhand von Beispielen. **(B3)**
- 2.2.4 Definieren Sie die Begriffe horizontale und vertikale Skalierbarkeit. Erläutern, klassifizieren und grenzen Sie verschiedene Konzepte zur Umsetzung von paralleler Datenverarbeitung in Data Science-Anwendungen wie z. B. Multiprocessing/-threading, Parallel Collections, Message-basierte Systeme, GPU Computing voneinander ab. **(C2)**
- 2.2.5 Erläutern Sie das Konzept der „Reproducible Research“ und erklären Sie die Umsetzung am Beispiel von sogenannten „Notebook-Documents“. **(B2)**

### 2.3 Informationsverarbeitung in Versicherungen

Zielsetzung: Die Kandidaten kennen die Kernprozesse eines Versicherungsunternehmens und die Bestandteile der Systemlandschaft, in denen sie umgesetzt sind. Sie kennen die wesentlichen Kriterien zur Beurteilung der Datenqualität und können sie anwenden.

- 2.3.1 Benennen Sie die wichtigsten Geschäftsprozesse in einem Versicherungsunternehmen. Diskutieren Sie, für welche Geschäftsprozesse Data Science-Anwendungen sinnvoll sind. Nennen Sie Beispiele für Data Science-Anwendungen und ordnen Sie diese einem Geschäftsprozess zu. **(C2)**
- 2.3.2 Benennen Sie die wichtigsten Anwendungssysteme in einem Versicherungsunternehmen und erläutern Sie die Abbildung von Prozessen auf die Anwendungssysteme. Gehen Sie hierbei auf die Bestandteile einer Gesamtarchitektur (Geschäftsprozesse, Fachliche Architektur, Softwarearchitektur, Softwarekomponenten) ein. **(C2)**

- 2.3.3 Erklären Sie typischerweise vorhandene Softwarearchitekturen in Versicherungsunternehmen und die typischen Anforderungen wie Revisionssicherheit, Wartbarkeit, Skalierbarkeit, Flexibilität und Bezahlbarkeit. **(A2)**
- 2.3.4 Erläutern Sie die Bedeutung und die Struktur der Bestandsverwaltung sowie deren Einbindung in die Systemlandschaft eines Versicherungsunternehmens. **(B2)**
- 2.3.5 Erläutern Sie die Bedeutung der Bestandsverwaltung als Datenquelle für Data Science-Anwendungen. Nennen Sie weitere Systeme in einem Versicherungsunternehmen, die als Datenquelle für Data Science-Anwendungen wichtig sein können. **(C2)**
- 2.3.6 Definieren Sie den Begriff „Datenqualität“. Nennen und erläutern Sie Kriterien zur Prüfung der Datenqualität und analysieren Sie mit ihnen die Datenqualität in konkreten Data Science-Anwendungen. **(B4)**
- 2.3.7 Benennen Sie Maßnahmen zur Verbesserung der Datenqualität und wenden Sie sie in konkreten Beispielen an. **(B4)**

### 3 Insurance Analytics

#### 3.1 Data Mining 1

Zielsetzung: Die Kandidaten sind mit den Grundbegriffen und den grundlegenden Methoden des Data Minings vertraut und können Daten als Grundlage des Data Mining- Prozesses aufbereiten und analysieren.

- 3.1.1 Erläutern Sie den Begriff Data Mining im Allgemeinen und im Umfeld der Versicherungswirtschaft. **(B2)**
- 3.1.2 Beschreiben Sie die Eigenschaften und Anwendungsmöglichkeiten von deskriptiven, prädiktiven und präskriptiven Analysen im Data Mining. **(C2)**
- 3.1.3 Wenden Sie deskriptive, prädiktive und präskriptive Analysen beispielhaft auf Problemstellungen im Versicherungsumfeld an. **(C3)**
- 3.1.4 Erklären Sie die Merkmale und Unterschiede von überwachtem und unüberwachtem Lernen. Nennen Sie beispielhafte Anwendungen in Versicherungsunternehmen. **(C2)**
- 3.1.5 Benennen und erläutern Sie die grundlegenden Anwendungsklassen des Data Minings. **(B2)**
- 3.1.6 Nennen und klassifizieren Sie die Merkmale von Daten und kategorisieren Sie Daten nach Art und Typ. **(B4)**
- 3.1.7 Erklären Sie die Merkmale und Unterschiede von abhängigen und unabhängigen Variablen (Ziel- und Kovariablen) und kategorisieren Sie Daten nach den Arten der Variablen. **(B4)**
- 3.1.8 Skizzieren Sie die Auswahl und Integration von unterschiedlichen Datenquellen zur Bereitstellung einer Datenbasis für Data Mining. Beschreiben Sie die hierbei auftretenden möglichen Probleme und erläutern Sie Lösungsansätze dafür. **(C2)**
- 3.1.9 Erläutern Sie die Ursachen und die Notwendigkeit zur Datenbereinigung, Datenreduzierung und Datentransformation als Grundlage des Data Minings. **(B2)**
- 3.1.10 Beschreiben Sie Methoden zur Bereinigung von Daten, zur Reduktion von Daten und zur Transformation von Daten. **(C2)**

## 3.2 Analytics 1

Zielsetzung: Die Kandidaten sind mit den Konzepten und den (grundlegenden) Methoden zur Modellerstellung vertraut und können Methoden und Verfahren zur Erzeugung von Modellen auswählen und anwenden.

- 3.2.1 Erläutern Sie die Funktionsweise und die Anwendungsmöglichkeiten von (Basic-) Methoden des Maschinellen Lernens und des Statistischen Lernens. **(B2)**
- 3.2.2 Erläutern Sie die Eigenschaften der Modellerstellung zur Prognose von Entscheidungen, Rangfolgen oder zur Bestimmung von Schätzwerten. **(B2)**
- 3.2.3 Beschreiben Sie für Aufgabenstellungen aus der Versicherungswirtschaft ein Vorgehen mit (analytischen) Modellen und bestimmen Sie die Ziele der Modell- und Prognoseerstellung. **(B2)**
- 3.2.4 Erläutern Sie die Notwendigkeit und den Nutzen der Selektion von informativen Attributen für die Modellbildung. **(B2)**
- 3.2.5 Erklären Sie, wie redundante und irrelevante Attribute für die Erstellung eines prädiktiven Modells ermittelt und ausgeschlossen werden. **(B2)**
- 3.2.6 Bestimmen Sie für exemplarische Anwendungsfälle Kenngrößen zur Prädiktionsgüte der Attribute und wählen Sie auf dieser Basis informative Attribute aus. **(B4)**
- 3.2.7 Erläutern Sie den Prozess des Model-Fittings. **(C2)**
- 3.2.8 Beschreiben Sie verschiedene Verfahren aus dem Maschinellen und Statistischen Lernen zur Bestimmung von Modellparametern bzw. Modellstrukturen. **(B2)**
- 3.2.9 Wenden Sie Verfahren zur Erstellung von prädiktiven Modellen an. **(B3)**
- 3.2.10 Erläutern Sie das Konzept (Techniken und Ziele) zur Trennung der Daten in Trainings-, Test- und Validierungsdaten. **(B2)**
- 3.2.11 Erklären Sie die Möglichkeiten und Konzepte zur Bewertung der Anpassungsgüte von Modellen, der Modellgüte und der Überprüfung der Modellvoraussetzungen. **(B2)**

## 4 Mathematik / Statistik

### 4.1 Überwachtes Lernen 1

Zielsetzung: Die Kandidaten verstehen die grundlegenden Konzepte des überwachten Maschinellen Lernens, haben einen Überblick über die wichtigsten Verfahren und können diese insbesondere für Klassifikations- und Regressionsfragestellungen sachgerecht anwenden.

- 4.1.1 Beschreiben Sie die algorithmische Funktionsweise von Entscheidungsbäumen (Klassifikations- und Regressionsbäume) und unterschiedliche Split-Kriterien. **(B2)**
- 4.1.2 Erläutern Sie die Methodik bei erweiterten Baumverfahren wie z. B. Random Forest und Bagging. **(B2)**
- 4.1.3 Erläutern Sie das Prinzip des (stochastic) Gradient Boostings. **(B2)**

### 4.2 Unüberwachtes Lernen 1

Zielsetzung: Aufbauend auf dem Grundwissenfach Angewandte Stochastik erwerben die Kandidaten erste Kenntnisse in den Verfahren unüberwachten Lernens. Sie werden dadurch insbesondere in die Lage versetzt, die Ergebnisse von gängigen Data Mining- Softwarepaketen zu interpretieren und die Modellbildung kritisch zu prüfen.

- 4.2.1 Beschreiben Sie grundlegende Verfahren der Clusteranalyse. **(B2)**
- 4.2.2 Geben Sie einen Überblick über im unüberwachten Lernen verwendete Methoden und beschreiben Sie den prinzipiellen Lösungsansatz des jeweiligen Algorithmus. **(C2)**

### 4.3 Deep Learning 1

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

- 4.3.1 Definieren Sie den Begriff „Künstliches Neuronales Netz“ mathematisch. Beschreiben Sie die grundlegende Struktur eines vorwärts gerichteten neuronalen Netzes. **(B2)**
- 4.3.2 Stellen Sie mögliche Aktivierungsfunktionen für Neuronen formal und graphisch dar. Analysieren und diskutieren Sie ihre Vor- und Nachteile. **(B4)**
- 4.3.3 Erläutern Sie die Funktionsweise und Intuition hinter einem sog. Bias-Neuron. **(B2)**
- 4.3.4 Beschreiben Sie die grundlegende Idee und Vorgehensweise des Backpropagation-Algorithmus und wenden Sie diese in einem einfachen neuronalen Netz an. **(C3)**
- 4.3.5 Erläutern Sie, wie ein vorwärts gerichtetes Künstliches Neuronales Netz trainiert wird bzw. wie es „lernt“. **(B2)**

## 5 Tools & Programme

### 5.1 Einführung und Überblick

Zielsetzung: Die Kandidaten erhalten einen Überblick über die wichtigsten Data Science- Tools und - Programme.

- 5.1.1 Nennen Sie die wichtigsten Programmiersprachen und Programmbibliotheken für Data Science. **(A1)**
- 5.1.2 Erläutern Sie die Funktionsweise von Notebooks. Diskutieren Sie die Vorteile beim Einsatz in Data Science-Projekten. **(B5)**
- 5.1.3 Nennen Sie verbreitete Tools und Software zur Unterstützung von Data Science-Anwendungen. **(A4)**

### 5.2 Sprachübergreifende Data Science-Tools

Zielsetzung: Die Kandidaten lernen codefreie sowie sprachübergreifende Tools kennen.

- 5.2.1 Nennen Sie verschiedene sprachübergreifende Gradient Boosting-Tools, beschreiben Sie die Unterschiede, nennen Sie die bevorzugte Wahl und begründen Sie diese. **(B3)**
- 5.2.2 Beschreiben Sie, wie sich Deep Learning-Methoden mit State of the Art- Frameworks umsetzen lassen. **(B2)**
- 5.2.3 Nennen und beschreiben Sie Tools, die über eine Benutzeroberfläche bedient werden, und inwieweit in diese Code (insbes. Python, R) eingebunden werden kann. **(B2)**
- 5.2.4 Nennen und beschreiben Sie Tools, die sowohl über eine Benutzeroberfläche als auch über R- und Python-Application Programming Interface angewendet werden können. **(B2)**

### 5.3 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.

- 5.3.1 Benennen und erläutern Sie die wichtigsten Sprachelemente von R. **(B2)**
- 5.3.2 Benennen und erläutern Sie die wichtigsten R-Bibliotheken zur Datenaufbereitung und Visualisierung. **(B2)**
- 5.3.3 Erstellen Sie Notebooks in R 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 einfachen 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)**