

## Prüfungsordnung 5.0

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

#### Inhalt

<b>1</b>	<b>Informationstechnologie.....</b>	<b>2</b>
1.1	Kodierungstheorie.....	2
1.2	Cloud Computing.....	2
1.3	Entwicklungsmethoden.....	3
<b>2</b>	<b>Insurance Analytics .....</b>	<b>3</b>
2.1	Anomaly Detection.....	3
2.2	Interpretation (von Modellen und Ergebnissen).....	3
<b>3</b>	<b>Mathematik / Statistik.....</b>	<b>4</b>
3.1	Deep Learning 4 .....	4
3.2	Anonymisierung / Pseudonymisierung 2 .....	4
3.3	Quantencomputing.....	5
<b>4</b>	<b>Use Cases.....</b>	<b>5</b>
4.1	Use Case.....	5

## 1 Informationstechnologie

### 1.1 Kodierungstheorie

Zielsetzung: Die Kandidaten verstehen die Grundzüge der Kodierungstheorie, haben Einblick in die wesentlichen Funktionsweisen sicheren Datenaustauschs sowie in ihre Anwendbarkeit im Versicherungskontext.

- 1.1.1. Erläutern Sie jeweils ein Beispiel für reversible bzw. irreversible (sog. Hashing) Verschlüsselungsalgorithmen und ihren Nutzen im Versicherungskontext. **(C2)**
- 1.1.2. Analysieren Sie die rechtlichen und technischen Rahmenbedingungen, innerhalb derer Verschlüsselung zu Datensicherung und -austausch verwendet wird, und welche, beispielsweise aus der Komplexitätstheorie erwachsenden, Abwägungen zwischen Sicherheit und Praktikabilität zu treffen sind. **(D4)**
- 1.1.3. Skizzieren und analysieren Sie ein Protokoll zur sicheren Übertragung einer binären Information zwischen zwei nicht vertrauenswürdigen Parteien (sog. Bit Commitment). **(C4)**
- 1.1.4. Skizzieren und analysieren Sie ein Protokoll zur quantensicheren Übertragung von Information (sog. Quantenkryptographie). **(C4)**
- 1.1.5. Nennen Sie Beispiele in der Versicherungsbranche, deren Geschäftsmodell ohne wirksame Verschlüsselung aus regulatorischen, formalen oder faktischen Gründen hinfällig wäre. **(C1)**
- 1.1.6. Geben Sie eine Übersicht über die konzeptuellen Grundlagen der Blockchain- Technologie, also Kettenlängen, Hashing, Verifizierbarkeit etc. **(C2)**

### 1.2 Cloud Computing

Zielsetzung: Die Kandidaten kennen die gängigen Service- und Bereitstellungsmodelle der verschiedenen Anbieter und kennen Kriterien, die für eine Umsetzung einer Anwendung in der Cloud berücksichtigt werden sollten.

- 1.2.1 Erklären Sie verschiedene Servicemodelle wie z. B. IaaS, PaaS und SaaS. **(B2)**
- 1.2.2 Benennen Sie Bereitstellungsmodelle wie z. B. Private Cloud, Hybrid Cloud, Community Cloud, Public Cloud und grenzen Sie diese voneinander ab. **(B2)**
- 1.2.3 Diskutieren Sie die Vor- und Nachteile der Umsetzung von Prozessen und Berechnungen in der Cloud. **(C4)**
- 1.2.4 Erläutern Sie die Möglichkeiten und die Einschränkungen der Verwendung von Cloud Computing für actuarielle Anwendungsfälle. **(B2)**
- 1.2.5 Nennen Sie die verbreitetsten Ansätze zur verteilten Verarbeitung und Analyse großer Datenbestände und deren wesentliche Konzepte. **(C1)**
- 1.2.6 Skizzieren Sie die Funktionsweise des Map/Reduce-Algorithmus und seiner Vor- und Nachteile. **(C5)**

### 1.3 Entwicklungsmethoden

Zielsetzung: Die Kandidaten wissen, was sich hinter dem Begriff „Agilität“ verbirgt und kennen die wichtigsten agilen Vorgehensmodelle.

- 1.3.1 Benennen Sie verschiedene Organisationsmethoden in der IT-Entwicklung, analysieren Sie diese und grenzen Sie diese gegeneinander ab. Methoden sind z. B. die klassische Wasserfall-Methode und Scrum in der Projektorganisation sowie der Design-Thinking-Prozess. **(D4)**
- 1.3.2 Benennen Sie die unterschiedlichen Ebenen des Testens von Anwendungssystemen. **(C1)**

## 2 Insurance Analytics

### 2.1 Anomaly Detection

Zielsetzung: Die verschiedenen Methoden zur Erkennung von Ausreißern können nach Art der vorliegenden Daten benannt und ihre Funktionsweise mathematisch beschrieben werden. Aktuelle Beispiele aus der Versicherungswirtschaft und Lösungsansätze für diese können skizziert werden.

- 2.1.1 Erklären Sie die Begriffe Anomalie und Ausreißer und grenzen Sie diese voneinander ab. **(B2)**
- 2.1.2 Erläutern Sie, wie man durch eine Analyse von einfachen Ausreißern im Vorfeld der Modellierung bereits Aussagen zur Qualität der Daten und der Auswirkung auf die Analysen machen kann. **(B2)**
- 2.1.3 Grenzen Sie die Methoden des überwachten und unüberwachten Lernens im Kontext der Anomaly Detection voneinander ab. Nennen Sie Anwendungen aus der Versicherungswirtschaft. **(B2)**
- 2.1.4 Erklären Sie im Zusammenhang der Anomaly Detection den Begriff des Rauschens und unterscheiden Sie zwischen dem statistischen und systematischen Rauschen. Erklären Sie dabei den Einfluss auf die Analyse von Ausreißern. **(B2)**
- 2.1.5 Nennen und vergleichen Sie zwei Verfahren des unüberwachten Lernens im Kontext der Anomaly Detection. **(B4)**
- 2.1.6 Nennen Sie überwachte Verfahren (z. B. Hidden Markov Models, Support Vector Machines und Time Series Analytics) zur Erkennung anormaler Zusammenhänge und beschreiben Sie, wie diese genutzt werden können. Stellen Sie die Modelle und deren Annahmen vor. **(C2)**

### 2.2 Interpretation (von Modellen und Ergebnissen)

Zielsetzung: Die Kandidaten sind in der Lage, Modelle und Ergebnisse unterschiedlicher Komplexität zu interpretieren. Hierbei können Sie anwendungsfallbezogen bewerten und einschätzen, wie hoch die Komplexität und Interpretationsfähigkeit von Modellen sein soll.

- 2.2.1 Beschreiben Sie regulatorische und operative Anforderungen bezüglich der Interpretation und Nachvollziehbarkeit von Modellen innerhalb der Versicherungswirtschaft. **(B2)**
- 2.2.2 Erläutern Sie die unterschiedlichen Komplexitätsgrade in der Erstellung und der Interpretation von Modellen. **(B2)**
- 2.2.3 Erklären Sie die Bedeutung von Akzeptanz und dem Verständnis für Modelle in der (operativen) Anwendung von komplexen Modellen. **(B2)**

- 2.2.4 Erläutern Sie das Kosten-Nutzen-Verhältnis von steigender Komplexität in der Erstellung von Modellen und der Interpretation von Modellergebnissen. **(B2)**
- 2.2.5 Benennen und erläutern Sie Methoden (z. B. Surrogate Model, LIME, Maximum Activation Model, Variable Importance Measure, Shapley Value Explanation, localGLMnet, ICEnet, MACQ) zur Veranschaulichung und Erläuterung von Verfahren und Ergebnissen aus komplexen Modellen. **(B2)**
- 2.2.6 Beurteilen Sie für Anwendungsfälle aus der Versicherungswirtschaft, welche Modelle mit welchem Komplexitätsgrad anzuwenden sind, um ein optimales Verhältnis zwischen Genauigkeit und Interpretationsfähigkeit zu erhalten. **(B5)**

### 3 Mathematik / Statistik

#### 3.1 Deep Learning 4

Zielsetzung: Die Kandidaten kennen analytische Modelle des Text Mining und können die Vorgehensweise zur Analyse von Texten beschreiben. Es können Anwendungsfälle aus der Versicherungswirtschaft und Lösungsansätze beschrieben werden. Darüber hinaus sollen Text Mining-Tools bekannt und deren Umgang exemplarisch geübt worden sein.

- 3.1.1 Beschreiben Sie typische Text Mining-Anwendungen im Versicherungsunternehmen und zeigen Sie Möglichkeiten und Grenzen auf. **(B4)**
- 3.1.2 Benennen Sie ein klassisches Relevanzmaß für Textdokumente bezüglich einer Suchanfrage. Vergleichen Sie klassische Ranking-Metriken mit Ergebnissen anderer Methoden des NLP (natural language processing). **(C2)**
- 3.1.3 Erläutern Sie die grundlegende Funktionsweise von Word-Vektoren. Nennen Sie Einschränkungen und Ansätze, diese zu überwinden. **(C2)**
- 3.1.4 Erläutern Sie ein Text Mining-Verfahren, dem ein Unsupervised Learning- Verfahren zugrunde liegt. **(A2)**
- 3.1.5 Beschreiben Sie die üblichen Verarbeitungsschritte eines Textes, bevor er mittels NLP-Verfahren verarbeitet werden kann. **(C2)**
- 3.1.6 Erläutern Sie die Unterschiede zwischen Rückkopplung (RNN), Faltung (CNN) und Attention bezüglich des betrachteten Datenausschnittes und erörtern Sie die Vorteile und Nachteile von Attention gegenüber anderen Methoden. **(C2)**
- 3.1.7 Skizzieren Sie den Aufbau der Transformer-Architektur und erläutern Sie, welche Teile für die Sprachkodierung bzw. für die Sprachgenerierung benötigt werden. Benennen Sie Anwendungsfälle innerhalb und außerhalb der Sprachverarbeitung. Ordnen Sie Vor- und Nachteile vor dem Hintergrund von Stichworten wie ökologischem Fußabdruck, Halluzinationen und Data Leakage ein. **(C3)**

#### 3.2 Anonymisierung / Pseudonymisierung 2

Zielsetzung: Die Kandidaten kennen Möglichkeiten, die Anonymität von Daten und Methoden zu messen und zu erhöhen. Diese können anhand von Beispielen ermittelt und Schwachstellen aufgezeigt werden.

- 3.2.1 Erläutern und motivieren Sie Methoden zur Messung der Anonymität von Daten (wie die k-Anonymität oder l-Diversität). **(B2)**

- 3.2.2 Wenden Sie Methoden zur Messung von Anonymität von Daten auf Beispiele mit Versicherungsbezug an. **(B3)**
- 3.2.3 Nennen Sie Schwachstellen der Methoden und erklären Sie diese anhand von Beispielen. **(B5)**
- 3.2.4 Erklären Sie Differential Privacy und gehen Sie dabei auf verschiedene Verteilungsannahmen ein. **(B2)**
- 3.2.5 Zeigen Sie den Nutzen von Differential Privacy anhand eines Beispiels auf. Gehen Sie dabei auf die Wahl der Parameter ein. **(B5)**
- 3.2.6 Diskutieren Sie den Begriff der Anonymität kritisch. Nennen Sie bekannte Beispiele aus der Praxis, bei denen es zum Datenbruch gekommen ist. **(B5)**

### 3.3 Quantencomputing

Zielsetzung: Die Kandidaten sind mit den Besonderheiten und Vorteilen der quantenmechanischen Datenverarbeitung in Bezug auf Maschinelles Lernen vertraut und können diese zielgerichtet zur Optimierung von Modellstrukturen und -vorhersagen anwenden.

- 3.3.1 Beschreiben Sie die qualitativen und quantitativen Unterschiede zwischen klassischer Datenverarbeitung und quantenbasierter Datenverarbeitung, insbesondere in Bezug auf verschränkte Zustände der Qubit-Informationseinheiten. **(B4)**
- 3.3.2 Beschreiben Sie die Entsprechungen im Quanten-Gatter-System im Vergleich zur klassischen Gatter-Theorie. Geben Sie insbesondere Beispiele für die nicht-klassische Funktionsweise des Hardamard-Gatters und des Phase-Shift-Gatters. **(C3)**
- 3.3.3 Erläutern Sie die Funktionsweise des Shor-Algorithmus und des Grover- Algorithmus, deren Einordnung in die Komplexitätsklasse BQP und die prinzipielle Möglichkeit des Einsatzes für Machine Learning. **(D3)**
- 3.3.4 Erläutern Sie die Funktionsweise von Quantum-Enhanced Support Vector Machines. **(C3)**
- 3.3.5 Erläutern Sie die Funktionsweise von Quantum-Enhanced Neural Nets. **(C3)**
- 3.3.6 Zeigen Sie anhand eines Beispiels die praktische Durchführung von Quantumassisted Machine Learning. **(D5)**

## 4 Use Cases

### 4.1 Use Case

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

- 4.1.1 Basierend auf einer komplexen Fragestellung und einem gegebenen Datenbestand führen Sie eine Data Science-Analyse selbstständig durch. Dabei durchlaufen Sie alle Phasen eines Data Mining-Prozesses und erstellen ein Notebook in Python, R oder einer anderen geeigneten Sprache bzw. nutzen ein Framework, wobei Sie die üblichen Programmierstandards (Wartung, Zuverlässigkeit, Effizienz, Benutzerfreundlichkeit) beachten. Sie wenden Modelle an, interpretieren und beurteilen die Ergebnisse, erklären die Wirkungsweise des Modells und präsentieren diese Erkenntnisse zielgruppengerecht. **(C5)**