

FRAUNHOFER-GESELLSCHAFT

ZERTIFIZIERUNGSHANDBUCH UND PRÜFUNGSORDNUNG

Personenzertifizierungen im Bereich
Data Science
(Normatives Dokument)

Revision 14

Gültig ab Januar 2019

Fraunhofer-Personenzertifizierungsstelle
Schloss Birlinghoven
53757 Sankt Augustin

ZERTIFIZIERUNGSHANDBUCH UND PRÜFUNGSORDNUNG

Personenzertifizierungen im Bereich
Data Science

Dorothea Kugelmeier

Leiterin der Fraunhofer-Personenzertifizierungsstelle
angesiedelt am

Fraunhofer-Institut für Angewandte Informationstechnik FIT
Schloss Birlinghoven
53757 Sankt Augustin

Inhalt

1	VORWORT	6
2	ANWENDUNGSBEREICH	7
3	ALLGEMEINGÜLTIGE BEGRIFFE	9
4	VORGABEN FÜR DAS ZERTIFIZIERUNGSVERFAHREN	11
4.1	Ziel.....	11
4.2	Antragstellung	11
4.3	Prüfungsdurchführung	11
4.3.1	Zusammenstellung und Bereitstellung der Prüfungsunterlagen und Beauftragung der Prüfungsbeauftragten	12
4.3.2	Durchführung von schriftlichen Prüfungen (theoretisch (und praktisch)).....	12
4.3.3	Durchführung mündlicher Prüfungen (theoretisch und praktisch).....	12
4.4	Prüfungsfragen und -aufgaben	13
4.5	Auswertung und Bewertung von Prüfungen	13
4.6	Zertifizierung.....	13
5	RECHTE UND PFLICHTEN (Stand Oktober 2018).....	14
5.1	Bekanntmachung	14
5.2	Rechte.....	14
5.3	Pflichten.....	14
5.3.1	Gewissenhaftigkeit und Fortbildung	14
5.3.2	Unabhängigkeit	15
5.3.3	Persönliche Aufgabenerfüllung	15
5.3.4	Zulässige Verwendung von Zertifikaten	15
5.3.5	Verwendung des Fraunhofer-Logos.....	16
5.3.6	Anzeigepflicht.....	16
5.3.7	Auskunftspflicht.....	16
5.4	Verstoß gegen die Pflichten als zertifikatstragende Person	16
	ANLAGE A: »CERTIFIED DATA SCIENTIST BASIC LEVEL«.....	17
A 1	Verweis auf andere Normen und Dokumente.....	17
A 2	Anforderungsprofil.....	17
A 2.1	Bestimmung des Anforderungsprofils.....	17
A 2.2	Zugangsvoraussetzungen	17
A 2.2.1	Vorbildungen	17
A 2.2.2	Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten	18
A 2.2.3	Persönliche Voraussetzungen.....	18
A 2.3	Geforderte Kompetenzen (Lernziele).....	19
	ANLAGE B: »CERTIFIED DATA SCIENTIST SPECIALIZED IN DATA MANAGEMENT«	27
B 1	Verweis auf andere Normen und Dokumente.....	27
B 2	Anforderungsprofil.....	27
B 2.1	Bestimmung des Anforderungsprofils.....	27
B 2.2	Zugangsvoraussetzungen	27
B 2.2.1	Vorbildungen	27
B 2.2.2	Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten	28
B 2.2.3	Persönliche Voraussetzungen.....	28
B 2.3	Geforderte Kompetenzen (Lernziele).....	29

ANLAGE C: »DATA SCIENTIST SPECIALIZED IN DATA ANALYTICS«	34
C 1 Verweis auf andere Normen und Dokumente	34
C 2 Anforderungsprofil	34
C 2.1 Bestimmung des Anforderungsprofils.....	34
C 2.2 Zugangsvoraussetzungen	34
C 2.2.1 Vorbildungen	34
C 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten	35
C 2.2.3 Persönliche Voraussetzungen	35
C 2.3 Geforderte Kompetenzen (Lernziele).....	36
ANLAGE D: »DATA SCIENTIST SPECIALIZED IN BIG DATA ANALYTICS«	40
D 1 Verweis auf andere Normen und Dokumente	40
D 2 Anforderungsprofil	40
D 2.1 Bestimmung des Anforderungsprofils.....	40
D 2.2 Zugangsvoraussetzungen	41
D 2.2.1 Vorbildungen	41
D 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten	41
D 2.2.3 Persönliche Voraussetzungen	41
D 2.3 Geforderte Kompetenzen (Lernziele).....	42
ANLAGE E: »DATA SCIENTIST SPECIALIZED DEEP LEARNING«	46
E 1 Verweis auf andere Normen und Dokumente	46
E 2 Anforderungsprofil	46
E 2.1 Bestimmung des Anforderungsprofils.....	46
E 2.2 Zugangsvoraussetzungen	48
E 2.2.1 Vorbildungen	48
E 2.2.2 Zusätzliche Ausbildungen und praktische Anforderungen.....	49
E 2.2.3 Persönliche Voraussetzungen	49
E 2.3 Geforderte Kompetenzen (Lernziele).....	50
E 2.3.1 Applied Deep Learning	50
E 2.3.2 Vertiefungsbereich: Cognitive Robotics.....	55
E 2.3.3 Vertiefungsbereich: Image and Video Understanding	58
E 2.3.4 Vertiefungsbereich: Cognitive Cyber Security.....	62
E 2.3.5 Vertiefungsbereich: Skalierbare Lernsysteme.....	65
E 2.3.6 Vertiefungsbereich: Deep Learning for Textmining.....	67
E 2.3.7 Vertiefungsbereich: Zeitreihenanalyse	71
ANLAGE F: »DATA SCIENTIST ADVANCED LEVEL«	74
F 1 Verweis auf andere Normen und Dokumente	74
F 2 Anforderungsprofil	74
F 2.1 Bestimmung des Anforderungsprofils.....	74
F 2.2 Zugangsvoraussetzungen	74
F 2.2.1 Vorbildungen	74
F 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten	76
F 2.2.3 Persönliche Voraussetzungen	76
F 2.3 Geforderte Kompetenzen (Lernziele).....	76

ANLAGE G: »SENIOR DATA SCIENTIST«	77
G 1 Verweis auf andere Normen und Dokumente.....	77
G 2 Anforderungsprofil.....	77
G 2.1 Bestimmung des Anforderungsprofils.....	77
G 2.2 Zugangsvoraussetzungen	77
G 2.2.1 Vorbildungen	77
G 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten	79
G 2.2.3 Persönliche Voraussetzungen.....	79
G 2.3 Geforderte Kompetenzen (Lernziele).....	80

1 VORWORT

Im Folgenden wird das Verfahren für Personenzertifizierungen im Bereich »Data Science« in Anlehnung an die Vorgaben der EN ISO 17024 »Allgemeine Kriterien für Stellen, die Personen zertifizieren« beschrieben und damit ein einheitliches Zertifizierungssystem vorgegeben. Gleichzeitig dient dieses Zertifizierungshandbuch als Prüfungsordnung.

Der Anwendungsbereich des vorliegenden Zertifizierungshandbuchs erstreckt sich auf die Personenzertifizierungen im Bereich »Data Science« durch die Fraunhofer-Personenzertifizierungsstelle.

Die Personenzertifizierungen im Bereich »Data Science« beziehen sich auf folgende Zertifizierungsprofile:

- Level 1 (Basic Level):
 - Certified Data Scientist Basic Level
 - Certified Data Scientist Specialized in Data Management
 - Certified Data Scientist Specialized in Data Analytics
 - Certified Data Scientist Specialized in Machine Learning
- Level 2 (Advanced Level): Certified Data Scientist Advanced Level
- Level 3 (Senior Level): Certified Senior Data Scientist

Die verschiedenen Zertifizierungsprofile bauen wie folgt aufeinander auf:

Auf Level 1 (Basic Level) werden der Titel »Certified Data Scientist Basic Level« vergeben sowie Zertifikate in Spezialgebieten mit dem Titel »Certified Data Scientist Specialized in »Name der Spezialisierung««.

Das Zertifikat »Certified Data Scientist Basic Level« in Kombination mit einem der Zertifikate zum »Certified Data Scientist Specialized in ...« beinhaltet eine grundlegende Qualifikation in den wesentlichen inhaltlichen und methodischen Aspekten des Fachgebiets. Die Reihenfolge ist frei wählbar. Die Zugangsvoraussetzungen richten sich nach dem jeweiligen Spezialgebiet.

Level 2 (Advanced Level), wird mit dem Titel »Certified Data Scientist Advanced Level« abgeschlossen und beinhaltet die Anwendung von Methoden aus dem Bereich Data Science im Arbeitsalltag. Das Zertifikat wird vergeben nach Erwerb des Zertifikats »Certified Data Scientist Basic Level« plus eines der Zertifikate zum »Certified Data Scientist Specialized in...« sowie dem Nachweis über eine mindestens einjährige Berufserfahrung im Bereich Data Science.

Auf Level 3 (Senior Level) wird der Titel »Certified Senior Data Scientist« vergeben, wenn neben mehrjähriger Berufserfahrung, das Zertifikat »Certified Data Scientist Advanced Level« nachgewiesen werden kann und eine Studienarbeit angefertigt wurde. Nachfolgende Abbildung stellt die Zusammenhänge zwischen den einzelnen Zertifizierungsprofilen dar.

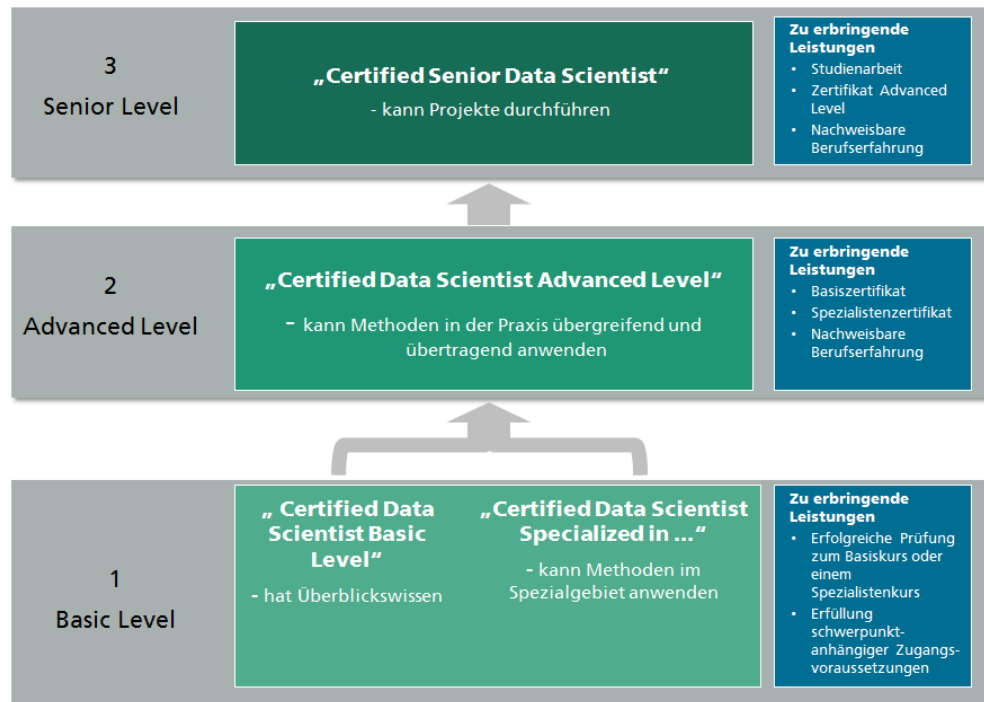


Abbildung 1: Zusammenhänge zwischen den Personenzertifizierungen im Bereich Data Science

Die Anforderungen der konkreten Zertifizierungsprofile sind in den Anlagen des vorliegenden Dokuments aufgeführt und sind Bestandteil der jeweiligen Personenzertifizierung.

■ Fraunhofer-Personenzertifizierungsstelle

Stelle in der Fraunhofer-Gesellschaft, die Zertifizierungen der Konformität von normativen Vorgaben und der tatsächlichen Personenqualifikation durchführt.

■ Prüfungsbeauftragte (PB)

Fachkräfte, die im Auftrag der Fraunhofer-Personenzertifizierungsstelle tätig werden, um Personen zu prüfen. Sie sind in der Wahrnehmung ihrer Prüfungsaufgaben fachlich unabhängig. Es ist sichergestellt, dass die Prüfungsbeauftragten die vollständigen Ausbildungsinhalte kennen.

■ Prüfungsbeisitzer (PBei)

Personen, die im Auftrag der Fraunhofer-Personenzertifizierungsstelle tätig werden, um die Prüfungsbeauftragten bei der Abnahme von Prüfungen zu unterstützen. Sie sind den Prüfungsbeauftragten fachlich unterstellt.

■ Fachausschüsse (FA)

Von der Fraunhofer-Personenzertifizierungsstelle berufene Gremien von Fachkräften, welche Prüfungsinhalte verifizieren und validieren, Prüfungsaufgaben erstellen, für Fachanfragen zuständig sind sowie die Fraunhofer-Personenzertifizierungsstelle hinsichtlich der fachlichen Qualität der Prüfungsbeauftragten beraten. Näheres zu Aufgaben und Befugnissen findet sich in der »Geschäftsordnung des Fachausschusses«. Für jedes Zertifizierungsprofil wird jeweils ein eigener Fachausschuss gebildet.

■ Begriff »kennen«

Befindet sich nach der Bloom'schen Lernzieltaxonomie (*Taxonomie von Lernzielen im kognitiven Bereich*. (Taxonomy of educational objectives, 1974). 5. Auflage. Beltz Verlag, Weinheim 1976) auf der ersten und zweiten Stufe der sechststufigen Skala. Kennzeichnend dafür ist die Wiedergabe aus dem Gedächtnis auf Abruf durch Stichworte. Die dafür ausgeprägten Fertigkeiten sind Wissen, Erkennen und Nachahmen.

Das Ziel »kennen« in Prüfungen im Bereich Data Science beinhaltet für jedes Zertifizierungsprofil unterschiedliche Inhalte. Diese werden in den Anhängen zu diesem Dokument beschrieben.

■ Begriff »anwenden«

Ist ein synonym verwendeter Begriff für die dritte und vierte Lernzielstufe der Bloom'schen Lernzieltaxonomie.

Kennzeichnend dafür ist die eigene Verarbeitung und Anordnung des Gelernten. Die dafür ausgeprägten Fertigkeiten sind Verstehen, Reagieren und Üben.

Das Ziel »anwenden« in Prüfungen im Bereich Data Science beinhaltet für jedes Zertifizierungsprofil unterschiedliche Inhalte. Diese werden in den Anhängen zu diesem Dokument beschrieben.

■ **Begriff »beurteilen«**

Ist ein synonym verwendeter Begriff für die Lernzielstufe »Transfer« und »Problemlösendes Denken«. Ist ein synonym verwendeter Begriff für die fünfte und sechste Lernzielstufe der Bloom'schen Lernzieltaxonomie.

Kennzeichnend dafür ist die Übertragung der Grundprinzipien auf neue, ähnliche Aufgaben bzw. auf für die Lernenden neue Leistungen. Die dafür ausgeprägten Fertigkeiten sind Anwenden, Werten, Koordinieren bzw. Problemlösen, Werte leben, Automatisieren.

Das Ziel »beurteilen« in Prüfungen im Bereich Data Science beinhaltet für die Zertifizierungsprofile unterschiedliche Inhalte. Diese werden in den Anhängen zu diesem Dokument beschrieben.

4 VORGABEN FÜR DAS ZERTIFIZIERUNGSVERFAHREN

Nachfolgend werden Vorgaben für das Zertifizierungsverfahren beschrieben.

4.1 Ziel

Durch Zertifizierungen werden anhand von definierten Anforderungsprofilen Qualifikationsmerkmale geprüft und deren Qualität durch ein Kompetenzzertifikat attestiert.

4.2 Antragstellung

Zertifiziert werden können Personen, die eine Prüfung der Fraunhofer-Personenzertifizierungsstelle im Bereich Data Science erfolgreich bestehen und die definierten Zugangsvoraussetzungen entsprechend den Anlagen dieses Zertifizierungshandbuchs erfüllen.

Personen, welche an der Zertifizierungsprüfung / Wiederholungsprüfung teilnehmen möchten, haben bei der Fraunhofer-Personenzertifizierungsstelle dazu einen schriftlichen Antrag zu stellen. Dieser Antrag muss folgende Angaben des Prüfungsteilnehmenden enthalten:

- Name, Geburtsdatum und private Postanschrift
- Arbeitsstelle mit Anschrift (wenn Arbeitsstelle vorhanden)
- Tätigkeit
- Zu zertifizierendes Zertifizierungsprofil
- Angabe, ob es sich um eine Erstprüfung oder Wiederholungsprüfung handelt.

Die Prüfungstermine werden von der Fraunhofer-Personenzertifizierungsstelle festgelegt.

4.3 Prüfungsdurchführung

Nachfolgend wird die Prüfungsdurchführung beschrieben.

Die Prüfungen auf Level 1 (Basic Level) zum »Certified Data Scientist Basic Level« erfolgen schriftlich und bestehen aus einem theoretischen Prüfungsteil. Die Prüfungen in den Spezialgebieten mit dem Titel »Certified Data Scientist Specialized in ...« erfolgen ebenfalls schriftlich und bestehen aus einem theoretischen Prüfungsteil mit praktischen Anteilen, die schriftlich bearbeitet werden.

Für die Zertifizierung auf Level 2 erfolgt keine gesonderte Prüfung. Allerdings müssen das Zertifikat »Data Scientist Basic Level« sowie mindestens eins der Zertifikate in den Spezialgebieten und Berufserfahrung nachgewiesen werden (siehe Anlage).

Für die Zertifizierung auf Stufe 3 erfolgt eine mündliche Prüfung zur Studienarbeit (Siehe Anlage).

4.3.1 Zusammenstellung und Bereitstellung der Prüfungsunterlagen und Beauftragung der Prüfungsbeauftragten

Die Fraunhofer-Personenzertifizierungsstelle stellt die Prüfungsfragen für die theoretische (schriftliche oder mündliche) Zertifizierungsprüfung aus einem vom zuständigen Fachausschuss bestätigten Fragenkatalog für das jeweilige Zertifizierungsprofil zusammen.

Die Bereitstellung der Prüfungsaufgaben muss zeitnah zur Prüfung erfolgen, damit die Prüfungsbeauftragten die Prüfung zum festgesetzten Termin durchführen können. Zudem muss die Bereitstellung der Prüfungsaufgaben geschützt vor unbefugtem Zugriff erfolgen.

Die Leitung der Fraunhofer-Personenzertifizierungsstelle beauftragt die Prüfungsbeauftragten mit der Abnahme und Korrektur der Prüfung.

4.3.2 Durchführung von schriftlichen Prüfungen (theoretisch (und praktisch))

Schriftliche Prüfungen finden an einem von der Fraunhofer-Personenzertifizierungsstelle abgenommenen Ort statt, der die von der Fraunhofer-Personenzertifizierungsstelle festgelegten Bedingungen erfüllt.

Die Prüfungsfragen /-aufgaben einer schriftlichen Prüfung sind handschriftlich zu beantworten. Es wird sichergestellt, dass für die Beantwortung der Fragen der theoretischen Prüfung ausreichend Zeit zur Verfügung steht. Hierzu wird bereits bei der Konzeption der Fragen vom zuständigen Fachausschuss überprüft, wie viel Zeit die Beantwortung der Fragen ungefähr in Anspruch nimmt.

Hilfsmittel sind grundsätzlich keine zugelassen.

Für Teilnehmende, die die Prüfung aufgrund einer Beeinträchtigung nicht in der vorgesehenen Form durchführen können, sind individuelle Ausnahmeregelungen vorgesehen.

4.3.3 Durchführung mündlicher Prüfungen (theoretisch und praktisch)

Mündliche Prüfungen finden in Form von Einzelprüfungen statt und werden von mindestens zwei Prüfungsbeauftragten durchgeführt.

Die Prüfungsfragen und -aufgaben werden dem Prüfungsfragenkatalog entnommen, der vom Fachausschuss Data Science zusammengestellt wurde.

Mündliche Prüfungen werden im Zertifizierungsprofil »Senior Data Scientist« durchgeführt. Im Rahmen eines Fachgesprächs wird der Prüfungsteilnehmende aufgefordert, seine Studienarbeit (Zulassungsvoraussetzung) zu präsentieren und Fachfragen zu der Arbeit sowie weiterführende Fragen im Bereich Data Science zu beantworten.

Das Fachgespräch ist in drei Abschnitte unterteilt

1. Präsentation der Projektarbeit (10 Min)
2. Fragen zur Projektarbeit mit Bezug zu Data Science im Allgemeinen und der nachgewiesenen Spezialisierung im Besonderen (15 Min)
3. Weiterführende Fragen im Bereich Data Science (15 Min)

Hilfsmittel sind grundsätzlich keine zugelassen. Allerdings erfolgt die Präsentation der Ergebnisse anhand einer im Vorfeld einzureichenden Power Point-Präsentation.

Für Teilnehmende, die die Prüfung aufgrund einer Beeinträchtigung nicht in der vorgesehenen Form durchführen können, sind individuelle Ausnahmeregelungen vorgesehen.

4.4 Prüfungsfragen und -aufgaben

Der Prüfungsfragenkatalog unterscheidet sich je nach Zertifizierungsprofil. Gleiches gilt für die Anzahl der Fragen pro Themenkomplex.

Der Prüfungsfragenkatalog beinhaltet rein theoretische Prüfungsfragen und Prüfungsfragen mit Praxisbezug.

Die Fragen sind eindeutig den Zertifizierungsprofilen und Themenbereichen zugeordnet. Jedem Prüfungsteilnehmenden dürfen nur Fragen und Aufgaben gestellt werden, die seinem fachlichen Anforderungsprofil entsprechen.

4.5 Auswertung und Bewertung von Prüfungen

Die Prüfungsteilnehmenden müssen einen Mindesterfüllungsgrad ihrer Aufgaben von 67% erreichen.

Bei Abweichungen unter dem Mindesterfüllungsgrad wird kein Zertifikat erteilt.

Bei Nichtbestehen kann die Prüfung maximal zweimal wiederholt werden.

Für jede Frage und Aufgabe werden den Prüfungsbeauftragten Musterlösungen vorgegeben, die als Richtlinie für die Beurteilung der Frage verwendet werden. Zusätzlich wird für jede Frage und Aufgabe die zu erreichende Punktzahl durch den zuständigen Fachausschuss vorgegeben.

4.6 Zertifizierung

Nach erfolgreich abgelegter Prüfung und Erfüllung der Zugangsvoraussetzungen wird dem Prüfungsteilnehmenden von der Fraunhofer-Personenzertifizierungsstelle das für das jeweilige Zertifizierungsprofil vorgesehene Zertifikat ausgehändigt.

Die Prüfungsteilnehmenden haben die Möglichkeit, fehlende Berufserfahrung innerhalb von einem Jahr nach Ablegen der jeweiligen Zertifizierungsprüfung (Zertifizierungsprofile in den Anhängen) nachzuweisen. Die Zertifikatserteilung erfolgt, sobald die Berufserfahrung nachgewiesen wurde. Die Zertifikatserteilung muss spätestens ein Jahr nach Ablegen der letzten Prüfung erfolgen.

Zertifikate im Zertifizierungsbereich »Data Science« (unabhängig vom Zertifizierungsprofil) sind unbegrenzt gültig.

5 RECHTE UND PFLICHTEN (Stand Oktober 2018)

Die Erteilung des Zertifikats ist mit einigen Rechten und Pflichten verbunden, auf die wir bereits im Vorfeld hinweisen möchten. Diese Regelungen werden Ihnen mit der späteren Erteilung des Zertifikats nochmals ausgehändigt.

5.1 Bekanntmachung

Die Fraunhofer-Personenzertifizierungsstelle darf auf schriftliche *Anfrage*, (z.B. von potentiellen Auftraggebern einer zertifikatstragenden Person) unter Angabe der Zertifikatsnummer Auskunft darüber erteilen, ob diese Person das Zertifikat rechtmäßig trägt. Zur Identifikation der zertifikatstragenden Person werden deren Name, Geburtsdatum und Geburtsort gespeichert. Mit der Anmeldung erklären Teilnehmende durch ihre Unterschrift ihre Absicht, diese Regelungen im Falle der Erteilung des Zertifikats zu akzeptieren. Die Fraunhofer-Personenzertifizierungsstelle ist an die Bestimmungen des deutschen Bundesdatenschutzgesetzes gebunden.

5.2 Rechte

Die zertifikatstragende Person ist berechtigt, im Rahmen ihrer Tätigkeit im Bereich »Data Science«:

- auf persönlichen Briefbögen, in sonstigen Drucksachen in Zusammenhang mit ihrer Person sowie im Internet im Zusammenhang mit ihrer Person auf ihre Zertifizierung wie folgt hinzuweisen: »zertifizierter NAME DES ZERTIFIKATS, geprüft durch die Fraunhofer-Personenzertifizierungsstelle« oder »zertifizierter »NAME DES ZERTIFIKATS« (z.B. »zertifizierter Data Scientist (Basic Level)« oder »zertifizierter Data Scientist specialized in Data Analytics«). Bei Verwendung der Variante 1 ist darauf zu achten, dass die Bezeichnung »geprüft durch die Fraunhofer-Personenzertifizierungsstelle« nicht größer ist als der zugehörige Name der Person.
- die ausgehändigte Zertifizierungs-Urkunde zu verwenden, allerdings nur im Ganzen.
- das Zertifizierungshandbuch »Personenzertifizierungen im Bereich Data Science« einzusehen, welches das Zertifizierungssystem im Bereich Data Science der Fraunhofer-Personenzertifizierungsstelle erläutert.

Näheres ist unter den Pflichten in 3.4 geregelt.

5.3 Pflichten

Folgende Pflichten sind bei der Ausübung der Aufgaben im Bereich »Data Science« von der zertifikatstragenden Person einzuhalten:

5.3.1 Gewissenhaftigkeit und Fortbildung

Die zertifikatstragende Person hat die in ihrem zertifizierten Profil genannten Tätigkeiten unter Berücksichtigung des Standes der anerkannten Regeln im Bereich Data Science zu erledigen.

Das Handeln der zertifikatstragenden Person ist von dem Grundsatz geprägt, dass stets die Gebrauchstauglichkeit des Produktes im Vordergrund steht.

Sie ist verpflichtet, die Zertifizierung nicht in einer missbräuchlichen Art und Weise zu verwenden und keinerlei Aussagen zu treffen, die von der Fraunhofer-Personenzertifizierungsstelle als irreführend oder unbefugt betrachtet werden müssen.

5.3.2 Unabhängigkeit

Die zertifikatstragende Person hat insbesondere darauf zu achten, dass sie ihr Handeln ohne Rücksicht auf dienstliche Beziehungen im Unternehmen, die übrigen Beschäftigten und / oder deren Ergebniswünschen ausrichtet (persönliche Unabhängigkeit).

5.3.3 Persönliche Aufgabenerfüllung

Die zertifikatstragende Person hat die von ihr geforderten Leistungen bei der Vorbereitung, Durchführung und Bewertung von Data-Science-Projekten persönlich zu erbringen bzw. zu überwachen. Sie darf ihre Zertifizierungsurkunde nicht in missbräuchlicher Weise verwenden.

5.3.4 Zulässige Verwendung von Zertifikaten

Folgende Regelungen gelten bezüglich der Verwendung von Zertifikaten:

- Das Zertifikat wird zwar der jeweiligen zertifikatstragenden Person erteilt; die Zertifikatsurkunde bleibt jedoch Eigentum der Fraunhofer-Personenzertifizierungsstelle.
- Es dürfen nur gültige Zertifikate verwendet werden.
- Das Zertifikat darf nicht missbräuchlich verwendet werden.
- Die Zertifizierungs-Urkunde darf nicht verändert werden und nur im Ganzen verwendet werden.
- Das Zertifikat ist der Fraunhofer-Personenzertifizierungsstelle unverzüglich zurückzugeben, nachdem das Zertifikat ausgelaufen ist, oder sobald die zertifikatstragende Person durch die Fraunhofer-Personenzertifizierungsstelle über den Entzug des Zertifikats informiert wurde
- Bei Aussetzung, Erlöschen oder Entzug von Zertifikaten ist die Verwendung des Zertifikats unverzüglich einzustellen; etwaige Hinweise auf das Zertifikat und die Fraunhofer-Personenzertifizierungsstelle sind unverzüglich zu löschen. Etwaige noch vorhandene Briefbögen und sonstige Drucksachen sind, im Falle der Aussetzung für deren Dauer nicht zu verwenden, ansonsten sind sie zu vernichten.
- Die Nutzung des Zertifikats bzw. Hinweise auf das Zertifikat sind nur im Geltungsbereich des Zertifikats gestattet.
- Das Zertifikat darf ausschließlich im Zusammenhang mit der darin zertifizierten Person verwendet werden.
- Die Verwendung des Zertifikats und Hinweise auf das Zertifikat sind nur zulässig, wenn für den Betrachter eindeutig erkennbar ist, welche Person in welchem Bereich geprüft und zertifiziert wurde.
- Durch die Verwendung des Zertifikats und Hinweise auf das Zertifikat darf nicht der Eindruck entstehen, dass die zertifizierte Person zum Personal der Fraunhofer-Gesellschaft gehört oder sie in ihrem Auftrag handelt.
- Die zertifikatstragende Person ist für die korrekte Verwendung des Zertifikats verantwortlich; etwaige Zweifel gehen zu ihren Lasten.

5.3.5 Verwendung des Fraunhofer-Logos

Das Zertifikat der Fraunhofer-Personenzertifizierungsstelle enthält auch das Fraunhofer-Logo. Das Logo darf ausschließlich als Teil des Zertifikats verwendet werden und zwar dergestalt, dass die Zertifizierungs-Urkunde im Ganzen als Nachweis der ausstellenden Fraunhofer-Personenzertifizierungsstelle für z. B. Kunden oder Arbeitgeber kopiert bzw. im Internet eingestellt werden kann. Jedwede darüber hinaus gehende Nutzung des Fraunhofer-Logos oder die markenmäßige Verwendung des Namens Fraunhofer ist ausdrücklich untersagt und kann im Falle von Zuwiderhandlungen Schadensersatzansprüche der Fraunhofer-Gesellschaft nach sich ziehen.

5.3.6 Anzeigepflicht

Die zertifikatstragende Person hat der Fraunhofer-Personenzertifizierungsstelle unverzüglich schriftlich anzuzeigen:

- Namensänderung (z. B. durch Hochzeit),
- die Änderung ihres Wohnsitzes,
- den Verlust des Zertifikates.

5.3.7 Auskunftspflicht

Die zertifikatstragende Person hat auf Verlangen der Fraunhofer-Personenzertifizierungsstelle die Einhaltung ihrer Pflichten erforderlichen Auskünfte (mündlich / schriftlich) innerhalb der gesetzten Fristen und unentgeltlich zu erteilen sowie angeforderte Unterlagen auf ihre Kosten vorzulegen.

Sie kann die Auskunft auf solche Fragen verweigern, deren Beantwortung sie selbst oder einen ihrer Angehörigen der Gefahr strafrechtlicher Verfolgung oder eines Verfahrens nach dem Gesetz über Ordnungswidrigkeiten aussetzen würde.

5.4 Verstoß gegen die Pflichten als zertifikatstragende Person

Ein Verstoß gegen die unter Punkt 5.3.1 bis 5.3.7 aufgeführten Pflichten führt je nach Schwere zur Aussetzung oder zum Entzug der Zertifizierung, welche der zertifikatstragenden Person schriftlich mitgeteilt wird. Für die Dauer der Aussetzung bzw. nach erfolgtem Entzug der Zertifizierung ist es der zertifikatstragenden Person untersagt, auf die Zertifizierung und die Fraunhofer-Personenzertifizierungsstelle hinzuweisen.

A 1 Verweis auf andere Normen und Dokumente

- EN ISO 17024

A 2 Anforderungsprofil

A 2.1 Bestimmung des Anforderungsprofils

Das Anforderungsprofil eines »Certified Data Scientist Basic Level« ergibt sich aus der Charakteristik und Beschreibung seines Tätigkeitsfeldes.

Ein zertifizierter »Data Scientist Basic Level«

- ist informiert über alle Ebenen der Data Science Wertschöpfungskette,
- ist in der Lage aus großen Datenmengen Informationen abzuleiten, die für das Unternehmen nutzbringend verwendet werden können,
- verwendet dabei Methoden u.a. aus der Informatik, der Mathematik, der Statistik, des maschinellen Lernens und der Mustererkennung,
- verbindet analytische Fähigkeiten mit technischem Verständnis für verarbeitende Software-Architekturen und Geschäftsverständnis und
- stellt ein Bindeglied zwischen verschiedenen Ebenen eines Unternehmens dar.

Abgrenzungskriterien des »Certified Data Scientist Basic Level« gegenüber anderen Profilen im Bereich Data Science sind im Zertifizierungshandbuch dokumentiert.

Die Bezeichnung lautet: »Certified Data Scientist Basic Level«

A 2.2 Zugangsvoraussetzungen

A 2.2.1 Vorbildungen

Ein zertifizierter »Data Scientist Basic Level« muss nachweisen:

Ein erfolgreich abgeschlossenes Studium an

- einer deutschen wissenschaftlichen Hochschule,
- einer deutschen staatlichen oder staatlich anerkannten Fachhochschule oder
- einer von der zuständigen Stelle des Landes als gleichwertig anerkannten ausländischen Hochschule

oder

- eine mindestens einjährige Tätigkeit im Zusammenhang mit der Analyse von Daten im Unternehmensumfeld etwa im Bereich Business Intelligence
- oder eine mindestens einjährige Tätigkeit in Übereinstimmung mit dem Anforderungsprofil in Abschnitt A2.

Anmerkung:

Im zu prüfenden Einzelfall hat die antragstellende Person die Möglichkeit, fehlende Zugangsvoraussetzungen innerhalb von einem Jahr nach Ablegen der Prüfung nachzuweisen.

Nach Prüfung der eingereichten Unterlagen entscheidet die Fraunhofer-Personenzertifizierungsstelle über die Voraussetzung. Sollten Zugangsvoraussetzungen nicht erfüllt sein, teilt die Fraunhofer-Personenzertifizierungsstelle dies der antragstellenden Person unverzüglich über das Sekretariat der Fraunhofer-Personenzertifizierungsstelle mit.

Grundsätzlich kann die Fraunhofer-Personenzertifizierungsstelle in begründeten Ausnahmefällen davon abweichende Nachweise akzeptieren. Diese Nachweise und die Entscheidung der Fraunhofer-Personenzertifizierungsstelle sind zu dokumentieren.

A 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten

Ein »Certified Data Scientist Basic Level« muss keine zusätzlichen Ausbildungen, Erfahrungen und Berechtigungen nachweisen.

A 2.2.3 Persönliche Voraussetzungen

Keine.

A 2.3 Geforderte Kompetenzen (Lernziele)

Grundlage für die Prüfung zum »Certified Data Scientist Basic Level« sind folgende Kompetenzen (Lernziele):

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden
Analysis of big data potentials			
Herausforderungen und Gefahren bei der Auswahl geeigneter Big Data Lösungen (Challenges and traps of selecting an appropriate big data solution)	<ul style="list-style-type: none"> ▪ Den Ansatz zum geschäftszielorientierten Big Data und seine Ziele benennen können. ▪ Den Ansatz zum datengetriebenen Big Data und seine Ziele benennen können. 	x	-/-
Vorgehen zur Potentialanalyse von Big Data für ein spezifisches Unternehmen (Analysis of big data potentials for a specific organization)	<ul style="list-style-type: none"> ▪ Die Kernelemente der Potentialanalyse von Big Data und ihre Ziele benennen können. ▪ Die fünf Kernelemente der Potenzialanalyse sind: Umfang und aktuelle Situation (Scope & Current Situation), Geschäftslösung (Business Solution), Technische Lösung (Technical Solution), Evaluation und Bereitschaft (Evaluation & Readiness) und Geschäftsszenario (Business Case). 	x	-/-
Geschäftsmodell (Business Model)	<ul style="list-style-type: none"> ▪ Die Kernelemente des Geschäftsmodells und ihre Bedeutung erläutern (benennen) können. ▪ Die neun Kernelemente des Geschäftsmodells sind: Nutzenversprechen (Value Proposition), Kundensegmente (Customer Segments), Kundenbeziehungen (Customer Relationships), Kanäle (Channels), Einnahmequellen (Revenue Streams), Schlüsselaktivitäten (Key Activities), Schlüsselressourcen (Key Resources), Schlüsselpartner (Key Partners), Kostenstrukturen (Cost Structures). 	x	-/-

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden
Geschäftsszenario (Business Case)	<ul style="list-style-type: none"> ▪ Den Unterschied zwischen Zieloption (Zieleszenario) und Alternativoptionen (Alternativszenarien) benennen können. ▪ Die Kernelemente einer Lösungsoption (Lösungsszenario) und ihre Ziele benennen können. Die fünf Kernelemente eines Lösungsszenarios sind: ▪ Annahmen (Assumptions), Nutzen (Benefit), Kosten (Cost), Risiken (Risks), und Zeitrahmen (Timeframe). 	x	-/-
Vorgehen zur Evaluation und Verbesserung der potenziellen Lösungen	<ul style="list-style-type: none"> ▪ Die Kernkriterien zur Evaluation einer Big Data Lösungsidee und die Zeile benennen können. ▪ Die drei Evaluationskriterien sind: Attraktivität (Desirability), Profitabilität (Profitability), Machbarkeit (Feasibility) Erfolgskriterien 	x	-/-
Vorgehen zur Analyse der Leistungsfähigkeit und Bereitschaft eines Unternehmens bzgl. Big Data (Analyzing organizational capability and readiness w.r.t. big data)	<ul style="list-style-type: none"> ▪ Bestandteile eines Big-Data-Lösungskonzepts (Solution Concept) und ihre Ziele benennen können. ▪ Die sieben Bestandteile eines Big-Data-Lösungskonzepts sind: Konkretes Zielszenario (Concrete Target Scenario), Spezifische Informationsbedürfnisse (Specific Information Needs), Lösungsidee und Annahmen (Solution Idea and Assumptions), Datenanalyseansatz (Data Analysis Approach), Bewertungskriterien und -ansatz (Evaluation Approach), Benötigte Daten und Datenqualitätsbedürfnisse (Required Data and Quality Needs), Architektur-/Infrastrukturvorschlag (Architecture / Infrastructure Proposal). 	x	-/-
Bereitschaftsanalyse (Readiness Analysis)	<ul style="list-style-type: none"> ▪ Zweck und Hauptschritte der Bereitschaftsanalyse (Readiness Analysis) benennen können. 	x	-/-

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden
Vorgehen zur gestuften Umsetzung und Anpassung eines Big Data Lösungskonzepts (Staged realization of big data solution concept)	<ul style="list-style-type: none"> ▪ Die Stufen des Umsetzungsansatzes für ein Big-Data-Konzept und die entsprechenden Ziele benennen können. ▪ Die fünf Stufen sind: Potentialanalyse (Potential Analysis), Laborstudie-klein(Lab Small), Laborstudie-umfänglich (Lab Full), Feldstudie (Field Study), Markteinführung (Roll-Out) 	x	-/-
Data Management			
Grundaufgaben des Daten-managements (What is data management)	<ul style="list-style-type: none"> ▪ Ziele des Datenmanagements beschreiben und in CRISP einordnen können. 	x	-/-
Data Lake Architekturen	<ul style="list-style-type: none"> ▪ Eine Data Lake Architektur und die Funktionen verschiedener Schichten beschreiben können. 	x	-/-
Datenmodellierung	<ul style="list-style-type: none"> ▪ Grundbegriffe der Datenmodellierung erklären und voneinander abgrenzen können. Die Schritte zur Erstellung eines Konzeptuellen Datenmodells darstellen können. 	x	-/-
Datenintegration	<ul style="list-style-type: none"> ▪ Allgemeine Schritte der Datenintegration beschreiben können. Verschiedene Datenintegrations-architekturen wie materialisierte oder virtuelle erläutern können. 	x	-/-
Datenaufbereitung	<ul style="list-style-type: none"> ▪ Verschiedene Verfahren der Datenaufbereitung benennen und erklären können. Das richtige Verfahren in einem Beispielszenario ermitteln können. 	x	-/-

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden
Datenqualität	<ul style="list-style-type: none"> ▪ Eine Definition für Datenqualität ausdrücken können. Die Begriffe Datenqualitätsziel, -charakteristik, -messmethode und -beobachtung voneinander abgrenzen können. Die Phasen der TDQM-Methodik beschreiben können. 	x	-/-
Data Analytics			
Definition von Data Mining und verwandten Konzepten (Definition of data mining and related concepts)	<ul style="list-style-type: none"> ▪ Definition von Datenanalyse und Unterschied zu Reporting formulieren können. ▪ Notwendige Skills für Datenanalysten (Statistik-Programmieren-Domänenwissen) benennen können. 	x	-/-
Branchen-übergreifende Standardprozesse zum Data Mining (Cross industry standard process for data mining)	<ul style="list-style-type: none"> ▪ CRISP-DM mit seinen Phasen business understanding, data understanding, data preparation, modeling, evaluation, deployment darstellen können 	x	-/-
Grundsätzliche Zwecke und Modellklassen der Datenanalyse (Main purposes and types of data analysis)	<ul style="list-style-type: none"> ▪ Grundsätzliche Zwecke und Modellklassen der Datenanalyse erklären können. 	x	-/-
Attribute und Datentypen (Attributes and data types)	<ul style="list-style-type: none"> ▪ Arten von Attributen (id. Kategorisch vs. Numerisch, Label, abgeleitete Attribute erklären können. 	x	-/-
Ansätze zur Datenaufbereitung (Data preparation approaches)	<ul style="list-style-type: none"> ▪ Transformation, Reduktion, Merkmalsauswahl, Normalisierung und Feature Design erklären können. 	x	-/-
Aufgabenstellungen und Modelle des statistischen Lernens (Models of statistical learning)	<ul style="list-style-type: none"> ▪ Überwachtes und unüberwachtes Lernen beschreiben können ▪ Regression, Klassifikation und Clusteranalyse erklären können ▪ Gängige Lernverfahren nennen können 	x	-/-

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden
Evaluation und Optimierung statistischer Modelle (Model evaluation)	<ul style="list-style-type: none"> ▪ Performanz-Maße wie Accuracy, Precision, Recall, RMS absolute mean error, ROC Kurve, Silhouetten-Kurven kennen ▪ Kreuzvalidierung erklären können ▪ Konfusionsmatrizen erklären können ▪ Overfitting erläutern können ▪ Schildern, wie Overfitting erkannt werden kann 	x	-/-
Visualisierung			
Visualisierungszwecke (Visualization purposes)	<ul style="list-style-type: none"> ▪ Exploration & Bestätigung, Erklärung 	x	-/-
Visualisierungsziele (Visualization Goals)	<ul style="list-style-type: none"> ▪ Den Verwendungszweck von Visualisierungen unterscheiden können. Beschreiben können, dass man Visualisierung einsetzt, um 1) eigene Analysen durchzuführen und zu verstehen, und 2) um die gefundenen Ergebnisse zur Erklärung zielgruppengerecht visuell aufzubereiten. 	x	x
Herausforderungen bei Big Data Visualisierung (Big Data Visualization and Analysis: Challenges of visualizing big data)	<ul style="list-style-type: none"> ▪ Die wichtigsten Herausforderungen bei Big Data Visualisierung kennen und verstehen. ▪ Beschreiben können, warum große Datenmengen zunächst durch filtern, aggregieren, sampeln etc. so zu verkleinern sind, dass existierende Werkzeuge für die Visualisierung und interaktive Exploration der Daten verwendet werden kann. 	x	x

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden
Grundwissen Big Data Visualisierung (Big Data Visualization and Analysis: Basics)	<ul style="list-style-type: none"> ▪ Notwendige Grundlagen bzgl. Visualisierungen kennen und verstehen. ▪ Beschreiben können, wo die Grenzen menschlicher Wahrnehmung liegen. ▪ Wichtige Visualisierungsbeispiele und ihre Einsatzmöglichkeiten kennen. ▪ Zumindest die vier wichtigsten Grundprinzipien Effizienz, Verständlichkeit, Konsistenz und Korrektheit beim Einsatz von Visualisierungen kennen und beschreiben können. ▪ Einige Visualisierungswerkzeuge kennen. Erläutern können, aus was man bei der Auswahl eines Werkzeugs achten sollte. 	x	x
Big Data Visualisierungsansätze (Approaches to visualize big data)	<ul style="list-style-type: none"> ▪ Datentransformation, Datenreduktion, Interaktive Visualisierung, Verdichtete (muster-) Diagramme, Multidiagramme, Hybride Diagramme, Animation 		
Big Data Systems			
Grundaspekte von Verarbeitungs- und Speichersystemen für Big Data (Big data processing vs. big data storage systems)	<ul style="list-style-type: none"> ▪ Die Bedeutung von Grundbegriffen wie Horizontal scalability, ▪ fault tolerance kennen und erklären können. ▪ Grundverfahren zur Verteilung von Daten und Datenverarbeitung beschreiben können, Grundeigenschaften von Big data databases und NoSQL Datenbanken nennen können. 	x	-/-
CAP-Theorem über verteilte Systeme (CAP Theorem for distributed systems)	<ul style="list-style-type: none"> ▪ Die Bedeutung des CAP Theorems sowie der verwendete Begrifflichkeit wie Data Replication, Availability, Eventual consistency erklären können. 	x	-/-

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden
Lambda-Architekturen, Verarbeitung in Batch- und Speed-Layer (Lambda architecture, batch and speed processing)	<ul style="list-style-type: none"> ▪ Den Aufbau und Grundprinzipien der Lambda-Architektur wie All-Data Principle ▪ Precomputed Views, ▪ Organisation in Batch- and Speed processing erklären können. 	x	-/-
Konzept von MapReduce (The concept of MapReduce)	<ul style="list-style-type: none"> ▪ Den Aufbau des Map Reduce Prinzips, Beispiele für Map Reduce Anwendungen (Word Count), sowie die Aufgabe und Leistung des Fault-Tolerant Distributed File System (HDFS) beschreiben können. 	x	-/-
Konzepte für Speed-Layer und Verarbeitung von Datenströmen (Speed layers and stream processing)	<ul style="list-style-type: none"> ▪ Grundprinzipien und Herausforderungen der Verarbeitung von Datenströmen beschreiben können. 	x	-/-
Konzepte und Methoden zur Datenanalyse in Big Data Systemen (Concepts and methods for data analysis in Big Data System)	<ul style="list-style-type: none"> ▪ Die prinzipielle Vorgehensweise zur Modellbildung in Big Data Systemen beschreiben können, Beispiele von Data Analytics Tools (R, Spark) kennen und ihre Leistungsfähigkeit bzgl. Big Data Analytics einschätzen können. 	X	-/-
Security and Privacy for Big Data			
Security Herausforderungen (Challenges)	<ul style="list-style-type: none"> ▪ Die Schutzziele der Informationssicherheit nennen und beschreiben können. ▪ Sicherheitsherausforderungen für heterogene, verteilte Systeme angeben können. ▪ Sicherheitsrisiken und -maßnahmen im CRISP-DM-Prozess anführen können. 	x	-/-
Zugangskontrolle (Access control)	<ul style="list-style-type: none"> ▪ Authentifizierung und Autorisierung unterscheiden und erklären können ▪ Die Grundzüge des Kerberos-Protokolls darstellen können. ▪ Die Security Tokens von Hadoop bezeichnen können. 	x	-/-

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden
EU-Privacy Richtlinie (EU privacy regulation)	<ul style="list-style-type: none"> ▪ Die Begriffe »Datenschutz« und »personenbezogene Daten« deuten können. ▪ Die Grundsätze der EU-DSGVO und erläutern können. ▪ Das Konzept Privacy by Design beschreiben können. 	x	-/-
Technologien für den Schutz von Privacy (Technologies for privacy protection)	<ul style="list-style-type: none"> ▪ Anonymisierung und Pseudonymisierung definieren können. ▪ Privatsphärenrisiken (vermeintlich) anonymisierter Daten benennen können. ▪ Das Konzept der k-Anonymität anwenden und die Limitationen davon benennen können. 	x	x

Der »Certified Data Scientist Basic Level« umfasst noch keine Themenbereiche, die er zwingend beurteilen können muss.

ANLAGE B: »CERTIFIED DATA SCIENTIST SPECIALIZED IN DATA MANAGEMENT«

ANLAGE B: »CERTIFIED DATA
SCIENTIST SPECIALIZED IN DATA
MANAGEMENT«

B 1 Verweis auf andere Normen und Dokumente

- EN ISO 17024

B 2 Anforderungsprofil

B 2.1 Bestimmung des Anforderungsprofils

Das Anforderungsprofil eines »Certified Data Scientist Specialized in Data Management« ergibt sich aus der Charakteristik und Beschreibung seines Tätigkeitsfeldes.

Ein zertifizierter »Data Scientist Specialized in Data Management«

- ist informiert über gute Datenmanagement-Praxis und setzt diese um,
- ist informiert über alle Ebenen der Datenmanagement Wertschöpfungskette,
- kennt Prozeduren für die Beschreibung, Transformation und qualitative Aufbereitung von Daten,
- kennt Software für das Management von Daten,
- kennt Verfahren zur statistischen und visuellen Datenanalyse,
- verfasst Anweisungen, Richtlinien und Pläne für ein nachhaltiges Datenmanagement in einer Organisation.

Abgrenzungskriterien des »Certified Data Scientist Specialized in Data Management« gegenüber anderen Profilen im Bereich Data Science sind im Zertifizierungshandbuch dokumentiert.

Die Bezeichnung lautet: »Certified Data Scientist Specialized in Data Management«
Kurzbezeichnung: »Data Manager«

B 2.2 Zugangsvoraussetzungen

B 2.2.1 Vorbildungen

Ein zertifizierter »Data Scientist Specialized in Data Management« muss nachweisen:

Ein erfolgreich abgeschlossenes Studium an

- einer deutschen wissenschaftlichen Hochschule,
- einer deutschen staatlichen oder staatlich anerkannten Fachhochschule oder
- einer von der zuständigen Stelle des Landes als gleichwertig anerkannten ausländischen Hochschule

oder

- eine mindestens einjährige Tätigkeit im Zusammenhang mit der Analyse von Daten im Unternehmensumfeld etwa im Bereich Business Intelligence

- oder eine mindestens einjährige Tätigkeit in Übereinstimmung mit dem Anforderungsprofil in Abschnitt B2.

Anmerkung:

Im zu prüfenden Einzelfall hat die antragstellende Person die Möglichkeit, fehlende Zugangsvoraussetzungen innerhalb von einem Jahr nach Ablegen der Prüfung nachzuweisen.

Der Hoch-, Fachhoch- bzw. Fachschulabschluss sowie der Nachweis der Berufserfahrung erfolgt über eine Selbstauskunft. Die Fraunhofer-Personenzertifizierungsstelle behält sich vor die Selbstauskünfte zu überprüfen. Nach Prüfung der eingereichten Unterlagen entscheidet die Fraunhofer-Personenzertifizierungsstelle über die Erfüllung der Voraussetzung. Sollten Zugangsvoraussetzungen nicht erfüllt sein, teilt die Fraunhofer-Personenzertifizierungsstelle dies dem Antragsteller unverzüglich über das Sekretariat der Fraunhofer-Personenzertifizierungsstelle mit.

Grundsätzlich kann die Fraunhofer-Personenzertifizierungsstelle in begründeten Ausnahmefällen davon abweichende Nachweise akzeptieren. Diese Nachweise und die Entscheidung der Fraunhofer-Personenzertifizierungsstelle sind zu dokumentieren.

B 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten

Ein »Certified Data Scientist Specialized in Data Management« muss keine zusätzlichen Ausbildungen, Erfahrungen und Berechtigungen nachweisen.

B 2.2.3 Persönliche Voraussetzungen

Keine.

B 2.3 Geforderte Kompetenzen (Lernziele)

Grundlage für die Prüfung zum »Certified Data Scientist Specialized in Data Management« sind folgende Kompetenzen (Lernziele):

ANLAGE B: »CERTIFIED DATA
SCIENTIST SPECIALIZED IN DATA
MANAGEMENT«

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden	beurteilen
Metadata Management				
Modellierung von Daten und Metadaten	<ul style="list-style-type: none"> ▪ UML-Klassendiagramme vervollständigen können. ▪ Verschiedene Modellierungsmethoden nennen und kurz erklären können. ▪ 	x	x	
	<ul style="list-style-type: none"> ▪ Beschreiben können, welche Vorteile die Verwendung von Metadaten-Standards hat. ▪ Die Vollständigkeit von konzeptuellen Datenmodellen in Bezug auf die Lösung eines Geschäftsproblems beurteilen können. ▪ 	x		x
Semantische Datenbeschreibung	<ul style="list-style-type: none"> ▪ Beschreiben können, wie man ein Datenmodell (Schema) mit Standards annotiert. ▪ Unterschied von »klassischem« Web der Dokumente zum Semantic Web (Web der Daten) darstellen können. ▪ Das RDF-Datenmodell erklären können ▪ Vorteile und Grenzen von Vokabularen und Ontologien nennen können. ▪ Unterschied zwischen Vokabularen und Ontologien anhand von Beispielen aus verschiedenen Forschungsbereichen darstellen können. ▪ Beschreiben können, wie man Ontologien zur Annotation von Daten und Datenmodellen nutzen kann ▪ Grundlagen des Semantic Web und Linked Open Data nennen können 	x		

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden	beurteilen
Metadaten-Extraktion	<ul style="list-style-type: none"> ▪ Vorgehensweisen für die Extraktion von Metadaten aus Dateien und Informationssystemen beschreiben können. ▪ Beschreiben können, was ein regulärer Ausdruck ist und wie er für die Extraktion von Metadaten verwendet werden kann. ▪ Vor- und Nachteile für die Kodierung von Metadaten in Dateinamen benennen können und mit der strukturierten Ablage von Metadaten vergleichen können. ▪ Vorgehensweise für die Metadatenextraktion mit einem Datenintegrationswerkzeug (wie KNIME) umsetzen können. ▪ Vor- und Nachteile unterschiedlicher Vorgehensweisen der strukturierten Metadatenablage bewerten (beurteilen) können. 	x	x	x
Datenmanagement Architekturen				
Datenbank-Management-Systeme	<ul style="list-style-type: none"> ▪ Die Architektur von Data Warehouse- und Data Lake-Systemen beschreiben können. ▪ Die Ziele eines Data Lakes beschreiben können. ▪ Die Unterschiede von relationalen Datenbanksystemen und NoSQL Datenbanksystemen beschreiben können. ▪ Die grundsätzlichen Eigenschaften von NoSQL-Datenbanksystemen erklären können. ▪ Geeignete Anwendungsgebiete von relationalen und NoSQL-Datenbankmanagement-Systemen beurteilen können. 	x	x	x

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden	beurteilen
Data Lakes	<ul style="list-style-type: none"> ▪ Verschiedene Werkzeuge für die Data Ingestion nach ihren Einsatzgebieten kategorisieren können. ▪ Vorgehensweisen für die Data Ingestion von Rohdaten mit einem Datenintegrationswerkzeug (wie KNIME) umsetzen können. ▪ Vor- und Nachteile einer Data Ingestion von Rohdaten in einem Data Lake unterscheiden können. 	x	x	
Daten-abfragen	<ul style="list-style-type: none"> ▪ Den Unterschied zwischen XML- und JSON- Dokumenten erklären können. ▪ Methoden für die Abfrage von Daten aus XML- und JSON- Dokumenten benennen können. ▪ Einfache SQL-Abfragen interpretieren können. ▪ Einfache Abfragen für dokumenten-orientierte NoSQL-Systeme interpretieren können. ▪ Darstellen können, wie verschiedene Abfragekonzepte in unterschiedlichen Datenbanktypen abgebildet werden. ▪ Einfache Abfragen in der Abfragesprache von MongoDB formulieren können. ▪ Einfache Abfragen mit SQL formulieren können 	x	x	
Data Integration				
Grundlagen der Datenintegration	<ul style="list-style-type: none"> ▪ Den Unterschied zwischen materialisierter und virtueller Datenintegration beschreiben können. 	x		
	<ul style="list-style-type: none"> ▪ Grundlegende Operationen der Datenintegration beschreiben und in einem einfachen Beispiel anwenden können. 	x	x	

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden	beurteilen
Datenaufbereitung	<ul style="list-style-type: none"> ▪ Verfahren der Datenaufbereitung benennen können. ▪ Den Bedarf für die Harmonisierung von Daten erkennen können und erklären können wie diese durchgeführt wird. ▪ Vorgehensweise (Prozess) zur Überprüfung von Daten auf Fehlerfreiheit darstellen können. ▪ Die wichtigsten Verfahren/Methoden zur Erkennung von fehlerhaften Daten benennen und auf unterschiedliche Datentypen anwenden können. ▪ Die Grenzen zur Überprüfung von Daten auf mögliche Fehlerhaftigkeit benennen können. 	x	x	
	<ul style="list-style-type: none"> ▪ Vor/Nachteile von rein berechneten Verfahren (z.B. über statistische Größen) versus a priori Wissen zur Bereinigung von Datensätzen an Beispielen von Experimentaldaten darstellen können. 	x	x	
Dateninterpretation				
Visuelle Datenanalyse (VA)	<ul style="list-style-type: none"> ▪ Anwendung der VA benennen und beschreiben können. ▪ Grafische Darstellung verschiedener Datentypen zur qualitativen Beurteilung von Datensätzen einsetzen können. ▪ Am Beispiel einer Fragestellung eine geeignete Visualisierungsart auswählen und die Auswahl begründen können. ▪ 	x	x	
Statistische Datenanalyse	<ul style="list-style-type: none"> ▪ Den Unterschied zwischen Korrelation und Kausalität erklären können. 	x	x	
Vor- und Nachteile der VA bzw. statistischer Methoden	<ul style="list-style-type: none"> ▪ Grenzen der VA und der statistischen Datenanalyse erklären können. 	x	x	

Themenbereich	Kompetenzen (Lernziele)	kennen	anwenden	beurteilen
Data Governance				
Grundlagen der Data Governance	<ul style="list-style-type: none"> ▪ Aufgaben der Data Governance beschreiben können. ▪ Die wesentlichen Phasen in einem Data Mining Projekt beschreiben und miteinander in Beziehung setzen können. ▪ Die Hauptphasen des CRISP modellieren können. ▪ Den Begriff und einzelne Phasen des Data Life Cycle erklären können. ▪ Abweichungen vom typischen Lebenszyklus für Forschungsdaten benennen können. 	x	x	
Datenmanagement-Planung	<ul style="list-style-type: none"> ▪ Risiken im Datenmanagement anhand eines Szenarios identifizieren können und korrigierende Maßnahmen beschreiben können. ▪ Vorteile von Datenmanagement Plänen benennen können. ▪ Richtlinien für das Datenmanagement verfassen können. 	x	x	
Datenmanagement budgetieren	<ul style="list-style-type: none"> ▪ Beschreiben können, wie Ressourcen für das Datenmanagement geplant werden sollten. ▪ Kostenstrukturen verschiedener Datenmanagement-Systeme nennen können. 	x	x	

Der »Certified Data Scientist Specialized in Data Management« umfasst noch keine Themenbereiche, die er zwingend beurteilen können muss.

ANLAGE C: »DATA SCIENTIST SPECIALIZED IN DATA ANALYTICS«

C 1 Verweis auf andere Normen und Dokumente

- EN ISO 17024

C 2 Anforderungsprofil

C 2.1 Bestimmung des Anforderungsprofils

Das Anforderungsprofil eines »Data Scientist Specialized in Data Analytics« ergibt sich aus der Charakteristik und Beschreibung seines Tätigkeitsfeldes.

Ein zertifizierter »Data Scientist Specialized in Data Analytics«

- ist informiert über gängige Datenanalyseprozesse und wendet diese an,
- kennt Verfahren für die Transformation und Aufbereitung von Daten,
- kennt Methoden der explorativen Datenanalyse und maschinelle Lernverfahren,
- kennt gängige Strategien und Verfahren zur Bewertung von Analysen und deren Ergebnisse,
- und kennt Software-Lösungen für die Aufbereitung und Analyse von Daten und kann diese einsetzen.

Abgrenzungskriterien des »Certified Data Scientist Specialized in Data Analytics« gegenüber anderen Profilen im Bereich Data Science sind im Zertifizierungshandbuch dokumentiert.

Die Bezeichnung lautet: »Certified Data Scientist Specialized in Data Analytics«
Kurzbezeichnung: »Data Analyst«

C 2.2 Zugangsvoraussetzungen

C 2.2.1 Vorbildungen

Ein zertifizierter »Data Scientist Specialized in Data Analytics« muss nachweisen:

Ein erfolgreich abgeschlossenes Studium an

- einer deutschen wissenschaftlichen Hochschule,
- einer deutschen staatlichen oder staatlich anerkannten Fachhochschule oder
- einer von der zuständigen Stelle des Landes als gleichwertig anerkannten ausländischen Hochschule

oder

- eine mindestens einjährige Tätigkeit im Zusammenhang mit der Analyse von Daten im Unternehmensumfeld etwa im Bereich Machine Learning
- oder eine mindestens einjährige Tätigkeit in Übereinstimmung mit dem Anforderungsprofil in Abschnitt C2.

Anmerkung:

Im zu prüfenden Einzelfall hat die antragstellende Person die Möglichkeit, fehlende Zugangsvoraussetzungen innerhalb von einem Jahr nach Ablegen der Prüfung nachzuweisen.

ANLAGE C: »DATA SCIENTIST
SPECIALIZED IN DATA
ANALYTICS«

Der Hoch-, Fachhoch- bzw. Fachschulabschluss sowie der Nachweis der Berufserfahrung erfolgt über eine Selbstauskunft. Die Fraunhofer-Personenzertifizierungsstelle behält sich vor die Selbstauskünfte zu überprüfen. Nach Prüfung der eingereichten Unterlagen entscheidet die Fraunhofer-Personenzertifizierungsstelle über die Erfüllung der Voraussetzung. Sollten Zugangsvoraussetzungen nicht erfüllt sein, teilt die Fraunhofer-Personenzertifizierungsstelle dies dem Antragsteller unverzüglich über das Sekretariat der Fraunhofer-Personenzertifizierungsstelle mit.

Grundsätzlich kann die Fraunhofer-Personenzertifizierungsstelle in begründeten Ausnahmefällen davon abweichende Nachweise akzeptieren. Diese Nachweise und die Entscheidung der Fraunhofer-Personenzertifizierungsstelle sind zu dokumentieren.

C 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten

Ein »Data Scientist Specialized in Data Analytics« muss keine zusätzlichen Ausbildungen, Erfahrungen und Berechtigungen nachweisen.

C 2.2.3 Persönliche Voraussetzungen

Keine.

C 2.3 Geforderte Kompetenzen (Lernziele)

Grundlage für die Prüfung zum »Data Scientist Specialized in Data Analytics« sind folgende Kompetenzen (Lernziele):

Wissensgebiet	Kompetenzen (Lernziele)	ken- nen	an- wenden	beur- teilen
Grundlagen der Datenanalyse				
Grundlagen der Daten-Modellierung	<ul style="list-style-type: none"> ▪ Überwachte und unüberwachte Lernaufgaben unterscheiden können. ▪ Korrelation und Kausalität beschreiben können. ▪ Den Unterschied zwischen Korrelation und Kausalität herausstellen können. ▪ Klassifikations-, Regressions- und Clusteraufgaben unterscheiden können. ▪ Das Prozessmodell CRISP-DM und die einzelnen Phasen beschreiben können. 	x	x	
Daten-Vorverarbei-tung	<ul style="list-style-type: none"> ▪ Datentypen und Skalen benennen und voneinander unterscheiden können. ▪ Cleaning, Integration, Transformation, Anreicherung und Reduzierung beschreiben und wesentliche Verfahren anwenden können. ▪ Datentransformationen anwenden können. 	x	x	
Daten-analysen mit graphischen Tools	<ul style="list-style-type: none"> ▪ Das Arbeiten mit Workflows und Operatoren beschreiben können. ▪ Workflows beschreiben und erklären können. ▪ Analyseworkflows erstellen können. 	x	x	
Datenanalyse mit Skript-sprachen	<ul style="list-style-type: none"> ▪ Vor- und Nachteile der Datenanalyse mit grafischen Tools im Vergleich zu Skriptsprachen beschreiben können. ▪ Funktion von Code-Snippets beschreiben können. 	x		

Wissensgebiet	Kompetenzen (Lernziele)	ken- nen	an- wenden	beur- teilen
Grundlegende Analyse- methoden				
Visuelle Datenanalyse	<ul style="list-style-type: none"> ▪ Rollen der Visualisierung in der Datenanalyse erklären können. ▪ Den Begriff der visuellen Länge kennen und erklären können. ▪ Grundlegende Visualisierungen benennen können. ▪ Visualisierungen erstellen können. ▪ Vorteile interaktiver Plots benennen können. 	x	x	
Klassifizierung	<ul style="list-style-type: none"> ▪ Verfahren der Klassifizierung benennen können. ▪ Den Unterschied zwischen erklärbaren Modellen und nicht erklärbaren Modellen beschreiben und passende Verfahren benennen können. ▪ Vor- und Nachteile verschiedener Lernverfahren beschreiben können. ▪ Entscheidungsbäume erstellen können. ▪ Regeln ermitteln können. ▪ Support-Vektor-Machines vergleichen können. ▪ K-Nächste Nachbarn berechnen können. ▪ Metriken zur Evaluation von Klassifikationsmodellen anwenden können. 	x	x	
Regression	<ul style="list-style-type: none"> ▪ Verfahren der Regressionsanalyse benennen können. ▪ Unterschiede zwischen Regressionsverfahren herausstellen können. ▪ Regressionsverfahren anwenden können. ▪ Metriken zur Evaluation von Regressionsmodellen ermitteln können. 	x	x	

Wissensgebiet	Kompetenzen (Lernziele)	ken- nen	an- wenden	beur- teilen
Clusteranalyse	<ul style="list-style-type: none"> ▪ K-Means anwenden können. ▪ DB-SCAN anwenden können ▪ Hierarchisches Clustern anwenden können. ▪ Unterschiede von Clusterverfahren herausstellen können. ▪ Metriken zur Validierung von Clustermodellen ermitteln können. 	x	x	
Fort- geschrittene Analyse- methoden				
Verfahren zur Modell- optimierung	<ul style="list-style-type: none"> ▪ Bias und Varianz beschreiben können. ▪ Over- und Underfitting unterscheiden und beschreiben können. ▪ Verfahren zur Merkmalsauswahl beschreiben können. ▪ Optimierungsstrategien implementieren können. ▪ Verfahren der Kreuzvalidierung anwenden können. ▪ Gängige Ensemble-Methoden benennen können. 	x	x	
Zeitreihen- analyse	<ul style="list-style-type: none"> ▪ Grundlegende Komponenten/Zerlegung einer Zeitreihe anwenden können. ▪ Autoregression erklären können. ▪ Autoregressionsmodelle erklären können. ▪ Dynamic Time Warping anwenden können. 	x	x	

Wissensgebiet	Kompetenzen (Lernziele)	ken- nen	an- wenden	beur- teilen
Neuronale Netze	<ul style="list-style-type: none"> ▪ Funktionsweise eines Perceptrons beschreiben können. ▪ Aktivierungsfunktionen kennen. ▪ Komponenten von neuronalen Netzen benennen können. ▪ Eigenschaften gängiger Architekturen benennen können. ▪ Trainingsverfahren erklären können. ▪ Neuronale Netze erstellen können. 	x	X	
Ausreißer-Erkennung	<ul style="list-style-type: none"> ▪ Annahme und Ziel der Ausreißer-Erkennung beschreiben können. ▪ Ausreißer-Erkennung mittels grundlegender Analysemethoden anwenden können. ▪ LOF anwenden können. 	x	x	
Daten-Vorverarbei-tung	<ul style="list-style-type: none"> ▪ Erklären von typischen Code-Auszügen zur Daten-Vorverarbeitung. ▪ Schritte der Datenvorverarbeitung anwenden können. 	x	x	
Explorative Datenanalyse	<ul style="list-style-type: none"> ▪ Verfahren zur explorativen Analyse benennen können. ▪ Verfahren zur explorativen Analyse anwenden können. 	x	x	
Statistische Datenanalyse	<ul style="list-style-type: none"> ▪ Erklären von typischen Code-Auszügen zur Datenmodellierung. ▪ Datenmodellierungsverfahren anwenden können. 	x	x	
Modell-optimierung und -auswahl	<ul style="list-style-type: none"> ▪ Erklären von typischen Code-Auszügen zur Modelloptimierung. ▪ Verfahren zur Modelloptimierung anwenden können. 	x	x	

ANLAGE C: »DATA SCIENTIST
SPECIALIZED IN DATA
ANALYTICS«

Der »Data Scientist Specialized in Data Analytics« umfasst noch keine Themenbereiche, die er zwingend beurteilen können muss.

ANLAGE D: »DATA SCIENTIST SPECIALIZED IN BIG DATA ANALYTICS«

D 1 Verweis auf andere Normen und Dokumente

- EN ISO 17024

D 2 Anforderungsprofil

D 2.1 Bestimmung des Anforderungsprofils

Das Anforderungsprofil eines »Data Scientist Specialized in Big Data Analytics« ergibt sich aus der Charakteristik und Beschreibung seines Tätigkeitsfeldes.

Ein zertifizierter »Data Scientist Specialized in Big Data Analytics«

- Ist informiert darüber, wie Algorithmen zur Datenanalyse in einer skalierbaren Big-Data-Architektur implementiert werden können
- kennt den Einsatz von Batch- und Streaming Verfahren und Systemen für die Datenanalyse auf großen Datenmengen,
- kennt Methoden des maschinellen Lernens und der Anwendung in Tools für Big Data Systeme wie z.B. Spark und kann diese einsetzen,
- kennt Strategien und Verfahren zum Aufbau und Einsatz von Systemen zur Datenanalyse unter Echtzeitbedingungen sowie für das Deployment von Analysemodellen,
- und kennt fortgeschrittene Analysetechniken und Algorithmen sowie deren Umsetzung in Big Data Systemen.

Abgrenzungskriterien des »Certified Data Scientist Specialized in Big Data Analytics« gegenüber anderen Profilen im Bereich Data Science sind im Zertifizierungshandbuch dokumentiert.

Die Bezeichnung lautet: »Certified Data Scientist Specialized in Big Data Analytics«
Kurzbezeichnung: »Big Data Analyst«

D 2.2 Zugangsvoraussetzungen

D 2.2.1 Vorbildungen

Ein zertifizierter »Data Scientist Specialized in Big Data Analytics« muss nachweisen:

Ein erfolgreich abgeschlossenes Studium an

- einer deutschen wissenschaftlichen Hochschule,
- einer deutschen staatlichen oder staatlich anerkannten Fachhochschule oder
- einer von der zuständigen Stelle des Landes als gleichwertig anerkannten ausländischen Hochschule

oder

- eine mindestens einjährige Tätigkeit im Zusammenhang mit der Analyse von Daten im Unternehmensumfeld etwa im Bereich Machine Learning
- oder eine mindestens einjährige Tätigkeit in Übereinstimmung mit dem Anforderungsprofil in Abschnitt D2.

Anmerkung:

Im zu prüfenden Einzelfall hat die antragstellende Person die Möglichkeit, fehlende Zugangsvoraussetzungen innerhalb von einem Jahr nach Ablegen der Prüfung nachzuweisen.

Der Hoch-, Fachhoch- bzw. Fachschulabschluss sowie der Nachweis der Berufserfahrung erfolgt über eine Selbstauskunft. Die Fraunhofer-Personenzertifizierungsstelle behält sich vor die Selbstauskünfte zu überprüfen. Nach Prüfung der eingereichten Unterlagen entscheidet die Fraunhofer-Personenzertifizierungsstelle über die Erfüllung der Voraussetzung. Sollten Zugangsvoraussetzungen nicht erfüllt sein, teilt die Fraunhofer-Personenzertifizierungsstelle dies dem Antragsteller unverzüglich über das Sekretariat der Fraunhofer-Personenzertifizierungsstelle mit.

Grundsätzlich kann die Fraunhofer-Personenzertifizierungsstelle in begründeten Ausnahmefällen davon abweichende Nachweise akzeptieren. Diese Nachweise und die Entscheidung der Fraunhofer-Personenzertifizierungsstelle sind zu dokumentieren.

D 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten

Ein »Data Scientist Specialized in Big Data Analytics« muss keine zusätzlichen Ausbildungen, Erfahrungen und Berechtigungen nachweisen.

D 2.2.3 Persönliche Voraussetzungen

Keine.

D 2.3 Geforderte Kompetenzen (Lernziele)

Grundlage für die Prüfung zum »Data Scientist Specialized in Big Data Analytics« sind folgende Kompetenzen (Lernziele):

Wissensgebiet	Kompetenzen (Lernziele)	ken- nen	an- wenden	beur- teilen
Einsatz von Big Data Systemen zur Datenanalyse				
Grundlagen der Datenspeicherung und -verarbeitung in Big Data Systemen	<ul style="list-style-type: none"> ▪ Die Prinzipien der verteilten Datenspeicherung und -verarbeitung in Big Data Systemen erklären können. ▪ Herausforderungen und Möglichkeiten bei der Implementierung von Machine Learning Algorithmen in Big Data Systemen beschreiben können. ▪ SQL basierte Big-Data Tools kennen und anwenden können, sowie den Unterschied zu Datenbanksystemen erklären können. ▪ Den hierarchischen Aufbau von Big Data Systemen zum Einsatz in der Datenanalyse erklären können. 	x	x	
Machine Learning Algorithmen in Big Data Tools	<ul style="list-style-type: none"> ▪ Bibliotheken zur Datenanalyse in Big Data Systemen kennen und anwenden können. ▪ Möglichkeiten der verteilten Datenvorverarbeitung und Feature-Erzeugung kennen und mit Machine Learning Algorithmen aus den Bibliotheken verbinden können. ▪ Gängige Machine Learning Algorithmen wie Clustering, Regression und Klassifikation kennen und in Big Data Systemen anwenden können. 	x	x	

Wissensgebiet	Kompetenzen (Lernziele)	ken- nen	an- wenden	beur- teilen
	<ul style="list-style-type: none"> ▪ Beispiele der Parallelisierung von Machine-Learning Algorithmen auf Big Data Systemen erklären können. ▪ Zusammenhang zwischen der Parallelisierung von Machine-Learning Algorithmen und den Big Data Basisoperatoren wie Map/Reduce erklären können. 	x	x	
Daten-analysen mit graphischen Tools	<ul style="list-style-type: none"> ▪ Das Arbeiten mit Workflows und Operatoren beschreiben können. ▪ Analyseworkflows erklären können ▪ Den Zusammenhang zwischen Workflow-Tools und Analysen auf einem Big-Data-Cluster erklären können. 	x	x	
Massiv parallele Datenbanken	<ul style="list-style-type: none"> ▪ Den Aufbau und den Einsatz von massiv parallelen Datenbanken zur Datenanalyse beschreiben können. ▪ Massive Parallele Datenbanken von Big Data Systemen abgrenzen können. 	x		
Datenströme, Echtzeitfähigkeit und Deployment				
Streaming und Datenanalyse	<ul style="list-style-type: none"> ▪ Die Herausforderungen bei der Analyse von Datenströmen beschreiben können. Techniken zur Analyse von Datenströmen kennen. 	x		
Kopplung von Batch und Stream Processing	<ul style="list-style-type: none"> ▪ Die Schritte zur Modellerstellung aus dem CRISP-Modell in die Lambda-Architektur abbilden können. ▪ Zwischen Feature-Erzeugung, Modellbildung und Modellanwendung unterscheiden und auf Batch- und Stream-Processing abbilden können. 	x	x	

Wissensgebiet	Kompetenzen (Lernziele)	ken- nen	an- wenden	beur- teilen
Deployment und Echtzeitfähigkeit	<ul style="list-style-type: none"> ▪ Verfahren und Software-Lösungen zur Übertragung von Analysemodellen kennen und beschreiben können. ▪ Herausforderungen der Echtzeitfähigkeit einordnen und auf Analyse-Systeme abbilden können. 	X	x	
Complex Event Processing	<ul style="list-style-type: none"> ▪ Verfahren des Complex Event Processing kennen und in eine Big Data Analyse einordnen können. ▪ Die Aufgaben von Complex Event Processing im Kontext von Deployment von Analysemodellen beschreiben können. 	x		
Fort- geschrittene Analyse- methoden für Big Data				
Aufbau von Empfehlungssystemen	<ul style="list-style-type: none"> ▪ Aufbau und Funktionsweise der Utility Matrix im Kontext von Collaborative Filtering erklären können. ▪ Abstandsmaße kennen und anwenden können. ▪ Zerlegungen der Utility Matrix und Anwendung der Faktormatizen kennen und erklären können. ▪ Möglichkeiten und Grenzen der Parallelisierung von Collaborative Filtering kennen. 	x	x	
Verfahren zur Modell-optimierung	<ul style="list-style-type: none"> ▪ Verfahren zur Merkmalsauswahl beschreiben können. ▪ Optimierungsstrategien im Kontext von Big Data erklären können. ▪ Typische Big Data Code-Auszüge zur Modelloptimierung erklären können. ▪ Workflow-Konzepte für Machine Learning im Kontext von Big Data kennen und erklären können. 	x	x	

Wissensgebiet	Kompetenzen (Lernziele)	kennen	anwenden	beurteilen
Big Data Algorithmen	<ul style="list-style-type: none"> ▪ Problematik und Herangehensweise zur Dimensionsreduktion erklären können. ▪ Beispiele von probabilistischen Algorithmen zur Dimensionsreduktion wie z.B. MinHashing kennen und erklären können 	x	X	
Datenvorverarbeitung	<ul style="list-style-type: none"> ▪ Typische Big Data Code-Auszügen zur Datenvorverarbeitung erklären können. ▪ Schritte der Datenvorverarbeitung anwenden können. 	x	x	
Statistische Datenanalyse	<ul style="list-style-type: none"> ▪ Typische Big Data Code-Auszüge zur Datenmodellierung erklären können. ▪ Datenmodellierungsverfahren anwenden können. 	x	x	

ANLAGE D: »DATA SCIENTIST
SPECIALIZED IN BIG DATA
ANALYTICS«

Der »Data Scientist Specialized in Big Data Analytics« umfasst noch keine Themenbereiche, die er zwingend beurteilen können muss.

ANLAGE E: »DATA SCIENTIST SPECIALIZED DEEP LEARNING«

E 1 Verweis auf andere Normen und Dokumente

- EN ISO 17024

E 2 Anforderungsprofil

E 2.1 Bestimmung des Anforderungsprofils

Das Anforderungsprofil eines »Data Scientist Specialized in Deep Learning« ergibt sich aus der Charakteristik und Beschreibung seines Tätigkeitsfeldes.

Ein zertifizierter »Data Scientist Specialized in Deep Learning «

- Kennt die Grundlagen und Methoden des Deep Learning und wendet diese exemplarisch an. Dies beinhaltet
 - Kennt die Grundlagen des Designs und Trainings tiefer neuronaler Netze (Deep Learning) und wendet diese an.
 - Kennt und wendet die theoretischen Grundlagen exemplarisch an von
 - Faltungsnetzen (Convolutional Neural Networks [CNN])
 - Rekurrenten neuronalen Netzen [RNN]
 - Generativen neuronalen Netzen [GAN]
 - Reinforcement Learning
 - Recomender Systemen
 - Unüberwachten Lernansätzen
- Kennt die Prinzipien und Methoden seines Vertiefungsbereichs und setzt sie exemplarisch um.

Im Zertifizierungsprofil »Data Scientist Specialized in Deep Learning« ist die Besonderheit, dass aufgrund der Vielfältigkeit des Aufgabenfeldes neben den Grundlagen und Methoden des Deep Learning die zertifizierte Person zusätzliche Kompetenzen in einem Vertiefungsbereich nachweisen muss. Der Nachweis erfolgt durch den Besuch einer von der Fraunhofer-Personenzertifizierungsstelle anerkannten Weiterbildungsveranstaltung, welche die vorgegebenen Lernziele abdeckt.

Entsprechend ergibt sich für den Data Scientist Specialized in Deep Learning folgende zusätzliche Beschreibung des Tätigkeitsfeldes in dem von ihm gewählten Vertiefungsbereich:

Vertiefungsbereich: Cognitive Robotics

- Kennt Anwendungsmöglichkeiten von ML auf Industrieroboter (z.B. Computer Vision für Objekterkennung, Reinforcement Learning für Navigation und Greifen von Objekten),
- Kennt Tool Chain zu Entwurf Lösungskonzept, Datengenerierung, Design und Training eines neuronalen Netzes, Ausführung auf Robotersteuerung,
- Kennt Transfer Learning und weiß, wann dieses sinnvoll eingesetzt werden kann,
- Kennt Vor- und Nachteile von simulierten Lernumgebungen und kann eine virtuelle Umgebung zur Datengenerierung konzeptionell entwerfen,

Vertiefungsbereich: Image and Video Understanding

- ist informiert über Lösungsansätze zu Detektions-, Klassifikations- und Segmentierungsaufgaben,
- kann ein geeignetes Vorgehen für diese Aufgaben zusammenstellen aus Feature Engineering- und Deep-Learning-Methoden,
- kennt weiterführende Deep-Learning-Methoden der Bildverarbeitung und kann sie auf eigene Problemstellungen anwenden und gegebenenfalls anpassen,
- und kennt Verfahren aus dem Feature-Engineering- und Deep-Learning-Bereich, um Informationen aus Videos zu extrahieren und kann sie geeignet kombinieren, um die Genauigkeits- und Geschwindigkeitsanforderungen einer gegebenen Aufgabe zu erfüllen.

Vertiefungsbereich: Cognitive Cyber Security

- Kennt Anwendungsgebiete von ML im Bereich Cyber Security: (Anomalie-Erkennung, Authentisierung, Natural-Language-Processing, Malware Erkennung, Adversarial Machine Learning)
- Kennt die Vorteile von ML basierten Anwendungen in diesen Bereichen gegenüber herkömmlichen Methoden
- Kann mit Datensätzen aus diesem Bereich umgehen und grundlegende ML-basierte Lösungen konzeptionieren

Vertiefungsbereich: Skalierbare Lernsysteme

- Kennt die Grundlagen paralleler Algorithmen und deren Komplexität
- Kennt die Grundlagen Verteilter Rechnersysteme
- Kann gängige Parallelisierungsansätze auf Machine Learning Verfahren anwenden
- Kennt aktuelle Softwareumgebungen für Verteilte Big Data und High Performance Computing (HPC) Anwendungen
- Kann Anwendungen aktueller Machine Learning Algorithmen in einer verteilten Big Data und HPC Softwareumgebung umsetzen

Vertiefungsbereich: Deep Learning for Textmining

- Weiß über die Grundlagen von Textverstehen im Kontext von ML Bescheid.
- Weiß, wie Daten in dem Kontext vorverarbeitet werden müssen und kann selbst Klassifikatoren zur Textverarbeitung entwerfen und umsetzen.
- Kennt sich mit der Semantischen Ähnlichkeit aus

Vertiefungsbereich: Zeitreihenanalyse

- Kennt die Besonderheiten von Zeitreihen
- Weiß, was State Space ist
- Kennt sich mit Recurrent Neural Networks aus
- Kann Modelle zum Forecasting anwenden und die Grenzen von Forecasting erläutern
- kann klassische Data Mining Verfahren auf Zeitreihen anwenden

Abgrenzungskriterien des »Certified Data Scientist Specialized in Deep Learning« gegenüber anderen Profilen im Bereich Data Science sind im Zertifizierungshandbuch dokumentiert.

Die Bezeichnung lautet: »Certified Data Scientist Specialized in Deep Learning«
Kurzbezeichnung: »Specialist Deep Learning«

Auf dem Zertifikat wird zusätzlich zu den nachgewiesenen Kompetenzen im Bereich „Applied Deep Learning“ auch der jeweilig nachgewiesene Vertiefungsbereich genannt.

E 2.2 Zugangsvoraussetzungen

E 2.2.1 Vorbildungen

Ein zertifizierter »Data Scientist Specialized in Deep Learning « muss nachweisen:

Ein erfolgreich abgeschlossenes Studium an

- einer deutschen wissenschaftlichen Hochschule,
- einer deutschen staatlichen oder staatlich anerkannten Fachhochschule oder
- einer von der zuständigen Stelle des Landes als gleichwertig anerkannten ausländischen Hochschule

oder

- eine mindestens einjährige Tätigkeit im Zusammenhang mit der Analyse von Daten im Unternehmensumfeld etwa im Bereich Deep Learning
- oder eine mindestens einjährige Tätigkeit in Übereinstimmung mit dem Anforderungsprofil in Abschnitt E2.

Nachweis der Zulassungsvoraussetzungen:

Der Hoch-, Fachhoch- bzw. Fachschulabschluss wird nachgewiesen durch eine Kopie der Urkunde. Der Nachweis der Berufserfahrung sowie der zusätzlich geforderten Kenntnisse erfolgt durch die Bescheinigung des Arbeitgebers oder bei Selbstständigen durch eine eidesstattliche Erklärung.

Die Fraunhofer-Personenzertifizierungsstelle behält sich vor die Nachweise zu überprüfen. Nach Prüfung der eingereichten Unterlagen entscheidet die Fraunhofer-Personenzertifizierungsstelle über die Erfüllung der Voraussetzung. Sollten Zugangsvoraussetzungen nicht erfüllt sein, teilt die Fraunhofer-Personenzertifizierungsstelle dies dem Antragsteller unverzüglich über das Sekretariat der Fraunhofer-Personenzertifizierungsstelle mit.

Grundsätzlich kann die Fraunhofer-Personenzertifizierungsstelle in begründeten Ausnahmefällen davon abweichende Nachweise akzeptieren. Diese Nachweise und die Entscheidung der Fraunhofer-Personenzertifizierungsstelle sind zu dokumentieren.

Anmerkung:

Im zu prüfenden Einzelfall hat die antragstellende Person die Möglichkeit, fehlende Zugangsvoraussetzungen innerhalb von einem Jahr nach Ablegen der Prüfung nachzuweisen.

E 2.2.2 Zusätzliche Ausbildungen und praktische Anforderungen

ANLAGE E: »DATA SCIENTIST
SPECIALIZED DEEP LEARNING«

Ein »Data Scientist Specialized in Deep Learning« muss in mindestens einem Vertiefungsbereich nachweisen, dass er an einer, durch die Fraunhofer-Personenzertifizierungsstelle anerkannten Weiterbildung teilgenommen hat. Die Weiterbildung muss praktische Übungen in dem jeweiligen Themenfeld im Umfang von mindestens 1,5 Personentagen beinhalten.

Die Weiterbildungsmaßnahme im Vertiefungsbereich sollte zeitlich nach der Prüfung im Bereich „Applied Deep Learning“ erfolgen.

Nachweis:

Der Nachweis erfolgt durch eine Teilnahmebescheinigung des von der Fraunhofer-Personenzertifizierungsstelle anerkannten Weiterbildners. Aus der Teilnahmebescheinigung muss neben den Inhalten der besuchten Weiterbildung hervorgehen, dass die Person die praktische Übungsaufgabe erfolgreich absolviert hat.

Anmerkung:

In begründeten Ausnahmefällen kann die Fraunhofer-Personenzertifizierungsstelle auch solche Nachweise akzeptieren, die eine Weiterbildung bescheinigen, die zeitlich vor dem Prüfungstermin im Bereich „Applied Deep Learning“ liegen.

E 2.2.3 Persönliche Voraussetzungen

Keine.

E 2.3 Geforderte Kompetenzen (Lernziele)

Die Grundlage für die **Prüfung** zum »Data Scientist Specialized in Deep Learning« sind **ausschließlich** die unter E 2.3.1 aufgeführten Kompetenzen (Lernziele) im Bereich **»Applied Deep Learning«** und müssen durch eine schriftliche Prüfung nachgewiesen werden:

Zusätzlich zu der bestandenen Prüfung ist für die Zertifizierung zum »Data Scientist Specialized in Deep Learning« der Nachweis über die Teilnahme an einer Weiterbildung in einem Vertiefungsbereich erforderlich, in der die unten (E 2.3.2 ff) aufgeführten Kompetenzen vermittelt wurden.

E 2.3.1 Applied Deep Learning

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Maschinelles Lernen	Maschinelles Lernen und dessen unterschiedlichen Arten erläutern können.	x		
	Grundlagen zur Evaluierung von Modellen (Kreuzvalidierung) und zugehörige Maße (Genauigkeit, Precision, Recall, AUC) erläutern können.	x		
	Begriffe wie Probabilistische Modelle, Maximum Likelihood, Overfitting oder Regularisierung erklären können.	x		
	verschiedene Methoden wie z.B. Hyperparameter-Optimierung, Ensemblemethoden anwenden können.		x	
Programmieren	Die grundlegende Architektur eines aktuellen Deep Learning	x		
	einen Überblick über aktuelle ML Frameworks (Tensorflow, Scikit-Learn, pyTorch, Caffe,...)) geben und deren Vor- und Nachteile für praktische Anwendungen beurteilen können.	x	x	
	den grundlegenden Ablauf des Modelltrainings (Einlesen, Vorverarbeitung, Optimierung, Modell speichern) erläutern können.	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	den grundlegenden Ablauf der Modellanwendung (Einlesen, Vorverarbeitung, Modell laden, Modell anwenden) erläutern können.	X		
	die Möglichkeiten zur parallelen Ausführung von Machine Learning Algorithmen nennen können.	x		
	DNN Quellcode eines aktuellen Deep Learning Frameworks lesen können.		x	
	DNN Code selbständig erweitern können.		x	
	DNN-Code debuggen können.		x	
	eigene Daten für ein Deep Learning Framework zugänglich machen können.		x	
	einfache, eigene Deep Learning Projekte konzipieren und umsetzen können.		x	
Deep Learning Grundlagen	die Kernkomponenten und Strukturen neuronaler Netze benennen können.	x		
	Backpropagation als verbreitetes Verfahren für das Trainieren künstlicher neuronaler Netze beschreiben können. (trainieren)	x		
	SGD (stochastic gradient descent) und Varianten wie (Averaging, AdaGrad, RMSProp, Adam usw.) beschreiben können.	x		
	den Unterschied zwischen Deep und Shallow Learning erklären können.	x	x	
	typische Architekturen wie AutoEncoder, CNN, RNN, LSTM und GAN. erklären können.	x		
	einer gegebenen Fragestellung eine Grundarchitektur zuordnen können.	x	x	

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	zu einer gegebenen Fragestellung Komplexität und Bedarf an Daten abschätzen können.		x	
	Möglichkeiten zur gleichzeitigen Verarbeitung heterogener Datentypen erläutern können.	x		
	mögliche Grenzen der Anwendung von Maschinellen Lernverfahren erläutern können.	x		
Grundlagen Unüberwachtes Lernen	die Aufgaben von Clustering und Embeddings erklären können.	x		
	Clustering- und Embedding-Architekturen (K-means, Word2vec, doc2vec) beschreiben können.	x		
	Anwendungsszenarien für Datenmuster, Clustering und Embeddings. erläutern können.	x		
	Beispiele Embedding-Modelle für einfache Textanalyseanwendungen nennen können,		x	
Grundlagen CNNs	die theoretischen Grundlagen von CNNs wie Faltung und Weight Sharing erklären können.	x		
	Anwendungsszenarien für CNNs benennen können (z.B.)	x		
	moderne CNN-Architekturen aufzählen können.	x		
	moderne CNN-Architekturen auf Bilddaten anwenden können (z.B. AlexNet, InceptionNets, ResNet, Fully Convolutional Netz).		x	
	CNN für einfache Bildverarbeitungsanwendungen nennen (Beispiele), entwerfen und in einen Deep Learning Framework umsetzen können.		x	

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Grundlagen RNNs	die theoretischen Grundlagen von RNNs erklären können.	x		
	Anwendungsszenarien für RNNs (z.B. Sprachmodelle und Prognosen von Worteigenschaften) erklären können.	x		
	moderne RNN-Architekturen auf Textdaten anwenden können (Embeddings, Noise Contrastive Estimation).		x	
	RNNs für einfache Textanalyseanwendungen entwerfen und in einen Deep Learning Framework umsetzen können.		x	
Zeitreihenanalyse und Sequence to Sequence	Anwendungsszenarien für ZRA (Kursprognose, predictive Maintenance) und S2S (Übersetzung, Frage und Antwort) benennen können.	x		
	Aufgaben der ZRA und von s2S beschreiben können (z.B. predictive maintenance).	x		
	moderne S2S-Architekturen und deren Anwendungen (z.B. business data, Übersetzung, Spracherkennung) erläutern können.	x		
	ZRA und S2S für einfache Anwendungen entwerfen und in einem Deep Learning Framework umsetzen können.		x	
Grundlagen GANs	die theoretischen Grundlagen von GAN (Generative Adversarial Network) wie Spieltheorie und Nullsummerspiel erläutern können.	x		
	Anwendungsszenarien für GANs benennen können.	x		
	GAN-Architekturen erläutern können.	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	GANs für einfache Anwendungen entwerfen und in einem Deep Learning Framework umsetzen können.		x	
Grundlagen Reinforcement Learning	die theoretischen Grundlagen von RL wie Q-Learning und Policy Optimization erläutern können.	x		
	ein einfaches Beispiel adaptieren und anwenden können.		x	
Grundlagen Memory-Netze	Memory Netze beim Zusammenführen vieler Informationen nutzen können.	x		
	ein einfaches Beispiel adaptieren und anwenden können.		x	
Grundlagen Recommender Systeme	die typische Aufgabenstellung für RS beschreiben können.	x		
	verwendbare Inputs (user groups, item properties,...) benennen können.	x		
	die theoretischen Grundlagen von RS beschreiben können.	x		
	Deep Learning Ansätze für RS beschreiben können.	x		
	einer gegebenen Fragestellung eine Grundarchitektur zuordnen können.	x		
	ein einfaches RS-Problem umsetzen können.		x	

E 2.3.2 Vertiefungsbereich: Cognitive Robotics

ANLAGE E: »DATA SCIENTIST
SPECIALIZED DEEP LEARNING«

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wen- den	beur- teilen
Einführung	die Motivation für die Verwendung von ML in der Robotik erläutern können.	x		
	typische Herausforderungen in der Robotik in Bezug auf ML beschreiben können.	x		
	Anwendungsbeispiele von ML in der Robotik kurz beschreiben können.	x		
Methoden- baukasten für ML und Robotik	verschiedene Sensor-Techniken bis hin zur 3D-Bildverarbeitung beschreiben können (Sensorfusion, Sensordatenströme).	x		
	grundlegende Roboterkinematiken mit deren Arbeitsraum nennen können (z.B. mobile Roboter, stationäre Roboter).	x		
Robot Vision	typische Computer Vision Aufgaben, z.B. Klassifizierung, Lokalisierung, Segmentierung nennen.	x		
	Arten von Sensoren (1D, 2D, 3D), beschreiben könne	x		
	State-of-the-art Algorithmen für die Computer Vision Aufgaben nennen.	x		
	den Unterschied zwischen Semantic Segmentation und Instance Segmentation und deren Anwendung in der Robotik erläutern.	x		
	Repräsentationen von Rotationen (z.B. Rotationsmatrix, Euler-Winkel) kennen.	x		
	die grundlegende Funktionsweise des Algorithmus Mask R-CNN beschreiben.	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wen- den	beur- teilen
	die grundlegende Funktionsweise des Algorithmus YOLO beschreiben.	x		
	potentielle Anwendungsmöglichkeiten der Computer Vision Algorithmen in der Robotik erläutern können darstellen.		x	
	eine Faltungsoperation und Max Pooling auf ein gegebenes Beispiel anwenden.		x	
	für spezifische Roboterapplikationen entscheiden, ob maschinelle Lernverfahren zur Objekterkennung geeignet sind.		x	
Deep Reinforcement Learning für Robotik	Den Ablaufzyklus von Reinforcement Learning skizzieren.	x		
	typische Herausforderungen im Bereich RL für die Anwendung in der Robotik erläutern.	x		
	"Reward Shaping" für eine Roboter Aufgabe erläutern.	x		
	die Begriffe Exploration und Exploitation beschreiben und unterscheiden.	x		
	die Grundidee von Hindsight Experience Replay (HER) erläutern können.		x	
	potentielle Anwendungsmöglichkeiten der aufgeführten Algorithmen darstellen.		x	
	bestehende Anwendungsbeispiele von RL in der Robotik erläutern.		x	
Robotic Grasp Detection	Verfahren zur Greifplanung kategorisieren können.	x		
	Vor- und Nachteile der thematisierten Verfahren nennen können.	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wen- den	beur- teilen
Simulationsumgebung und Sim-to-Real Transfer	Vor- und Nachteile der Verwendung einer Simulationsumgebung nennen können.		x	
	Beschreiben wann der Einsatz von Transfer Learning sinnvoll ist.		x	
	Die Idee von Domain Adaptation beschreiben können.		x	
	Die Idee von Domain Randomization beschreiben können.			
	Die zwei Ausprägungen von Domain Randomization inklusive Beispiele nennen können			
Greifpunkterkennung und -auswahl	die Motivation für einen ML basierten Ansatz zur Greifpunktbestimmung beschreiben können.	x		

E 2.3.3 Vertiefungsbereich: Image and Video Understanding

Wissensgebiet	Kompetenzen/Lernziele	kennen	anwenden	beurteilen
Bilder und Videos	die Definitionen von Bildern und Videos wiedergeben können.	x		
	den Unterschied zwischen Bildern und Videos erklären können.	x		
	je ein oder mehrere gängige Bild- und Videoformate aus den Kategorien verlustbehaftete Kompressionsverfahren, verlustfreie Kompressionsverfahren, Video-Kompressionsverfahren, Video-Containerformate nennen können.	x		
	Bilder und Videos zur Verarbeitung vorbereiten können (konvertieren, in verschiedenen, für Lernverfahren geeigneten Formaten abspeichern und diese für Lernverfahren laden).		x	
Feature Engineering	Bildmerkmale aus den Kategorien statistische Bildmerkmale, formbeschreibende Merkmale, texturbeschreibende Merkmale, und Feature-Deskriptoren nennen können.	x		
	den Wert von Bildmerkmalen unterscheiden können (für statistische Bildmerkmale, formbeschreibende Merkmale, texturbeschreibende Merkmale, Feature-Deskriptoren).		x	
	Bildmerkmale der genannten Kategorien extrahieren können.		x	
	den generellen Aufbau der Bildverarbeitungspipeline beschreiben und programmieren können.	x	x	
	Klassifikatoren anwenden können (Naive Bayes, Random Forests, SVN).		x	

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	den Zusammenhang zwischen Segmentierung und Klassifikation erläutern können.	x		
	Detektion und Segmentierung voneinander abgrenzen können.	x		
	einfache Segmentierungsaufgaben mit Feature Engineering lösen können.		x	
	Detektionsaufgaben mit statistischen Feature-Deskriptoren lösen können.		x	
	Klassifikationsaufgaben mit Feature-Ansatz lösen können.		x	
	die Grenzen von Feature Engineering-Ansätzen nennen und beschreiben können (curse of dimensionality, Redundanz der features).	x		
Validierung von Klassifikations-, Detektions- und Segmentierungsalgorithmen	Spezielle Metriken in der Bildanalyse kennen und implementieren können, wie Dice-Score und Jaccard-Koeffizient als voxel-basierte Metriken, und volumen-basiert Metriken wie Oberflächen-Distanz und Volumetrische Überlappung.	x	x	
	Detektionsalgorithmen auswerten können.		x	
Segmentierung mit CNNs	Architekturen für Segmentierung erklären können.	x		
	das Konzept des rezeptiven Feldes kennen und die Größe abschätzen können.	x	x	
	Fully Convolutional Networks (FCN) beschreiben, die praktische Bedeutung von FCN erläutern und ein Netzwerk in ein FCN konvertieren können..	x	x	
	das U-Net/V-Net effizient implementieren können.		x	

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	das Overlapping Tile-Konzept erklären können.	x		
	die Segmentierung mehrerer Strukturen zugleich vergleichen und bewerten können (überlappend, nicht-überlappend, Losses für multi-class-Segmentierung).	x		
	Regularisierungsarten beschreiben und beurteilen können (Dropout, Batch Normalization, L1/L2-Normalisierung).	x	x	
	Aktivierungsfunktionen und Losses benennen und diese geeignet wählen können.	x	x	
	Segmentierungsnetzwerke konstruieren, trainieren und Trainingserfolge überwachen und beurteilen können.	x	x	
Detektion mit Deep Networks	die Detektion als Spezialfall von Voxel- oder Regionen-Klassifikation erläutern können.	x		
	Architekturen wie zum Beispiel Faster R-CNN und aktuelle Netzwerk-Topologien (zum Beispiel YOLO, RetinaNet) erläutern können.	x		
	vortrainierte Netze anpassen können.		x	
Klassifikation mit Deep Networks	Architekturen wie DenseNet und Capsule Networks beschreiben können.	x		
	den Umgang mit seltenen Klassen erläutern können.	x		
Explainability	Ansätze beschreiben können, um probabilistische Ergebnisse zu erhalten statt binärer Masken.	x		
	Ansätze benennen können, um die Bildeigenschaften, die für die Klassifikation entscheidend waren, zu bestimmen.	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	bestehende Netze um Explainability-Komponenten erweitern können.		x	
Weiterführende Bildbearbeitung mit tiefen Netzen	die Bedeutung der Loss-Funktion im Zusammenhang mit generativen Modellen erläutern können.	x		
	Architekturen für generative Netze kennen (z.B. Cycle GAN, Wasserstein GAN, DC-GAN)	x		
	Ein gegebenes GAN an eine gestellte Aufgabe anpassen können.		x	
	Image to Image Translation erklären können.	x		
	Bildinhalte trennen können (Style Transfer Learning).		x	
Videoanalyse	die Unterschiede von Bild- und Videoverarbeitung und deren Übertragung erläutern können.	x		
	Gründe für und das Vorgehen zur Verbindung von Feature-Engineering- und Deep-Learning-Methoden beschreiben können (zum Beispiel Einhaltung von Echtzeit-Constraints, Verbindung von Expertenwissen und statistischen Modellen).	x		
	den allgemeinen Aufbau von RNN (Recurrent Neural Networks) beschreiben können.	x		
	CNN und RNN kombinieren können (ConvLSTM-Layer).	x	x	
	charakteristische Sequenzen in Filmen detektieren können.		x	
	den Optical Flow aus Videosequenzen ermitteln können.		x	

E 2.3.4 Vertiefungsbereich: Cognitive Cyber Security

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Cybersicherheit	<p>Anwendungsfelder von ML für die Cybersicherheit kennen:</p> <ul style="list-style-type: none"> • Anomalie-Erkennung, • Authentisierung, • Natural-Language-Processing, • Malware Erkennung, • Adversarial Machine Learning. <p>Erläutern können, wie ML in diesen Feldern die Beschränkungen bereits bestehender Methoden überwinden kann.</p>	x		
	<p>Einführung in typische Datensätze und Datentypen in der Cybersicherheit:</p> <ul style="list-style-type: none"> • Netzwerkverkehr-Mitschnitte (PCAP) • Source Code • Binary Code • Logfiles <p>Diese Datentypen kennen, deren grundlegende Struktur verstehen und in der Lage sein diese für das Trainieren von ML Algorithmen entsprechend aufzubereiten können.</p>	x		
	<p>Das offensive Potential von Machine Learning kennen:</p> <ul style="list-style-type: none"> • Side-channel Analysis • Attacken auf PUFs (Physical-Uncloneable-Function) 	x		
Anomalie-Erkennung	<p>Verstehen der grundsätzlichen Funktionsweise eines Intrusion Detection Systems.</p>	x		
	<p>Wichtige ML Algorithmen für Anomalieerkennung kennen und anwenden können:</p> <ul style="list-style-type: none"> • Density Estimation, • PCA • One-Class SVM 	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Authentisierung	<ul style="list-style-type: none"> Die wichtigsten Konzepte und Methoden der ML-basierten, kontinuierlichen Authentisierung nennen können: <ul style="list-style-type: none"> Verhaltensanalyse für Shell/Bash, Modellierung der Benutzung der Maus/Tastatur, Touch Dynamics 	x		
	Kontinuierliche Authentisierung mit ML anhand eines Beispiels umsetzen können: <ul style="list-style-type: none"> Benutzeridentifikation aus Keystroke Dynamics 		x	
Natural Language Processing	Zentrale Anwendungsgebiete für NLP im Bereich Cybersicherheit kennen und beschreiben können: <ul style="list-style-type: none"> Spamfilter mittels Naive Bayes Quellcode Analyse mittels tf-idf und multi-class SVM 	x		
	Mittels NLP sicherheitsrelevante Dokumente analysieren können: <ul style="list-style-type: none"> Einen ML-basierten Spamfilter entwerfen können 		x	
	Cryptoanalyse mit Machine Learning: <ul style="list-style-type: none"> Mittels Hidden Markov Modellen klassische Kryptoalgorithmen angreifen 	x		
Malware-erkennung	Grundlagen des Deep Learning für Malware Klassifikation verstehen: <ul style="list-style-type: none"> Feature Extraction aus Malware Binaries Malware Klassifikation 	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Adversarial Machine Learning	<p>Schwachstellen von ML-Methoden beschreiben können:</p> <ul style="list-style-type: none"> • Grundlagen des Adversarial ML verstehen, insb. Causative und Evasive Attacks. 	x		
	<p>Taxonomie je einer der folgenden Angriffe beschreiben können:</p> <ul style="list-style-type: none"> • Label Poisoning, • Boundary Attack auf Neuronale Netzwerke 	x		
	<p>Pentesting Angriffe auf ML Modelle durchführen zu können:</p> <ul style="list-style-type: none"> • Adversarial Attacks auf Bilderkennung anhand eines Opensource Tools wie z.B. Cleverhans, DeepFool 	x		

E 2.3.5 Vertiefungsbereich: Skalierbare Lernsysteme

ANLAGE E: »DATA SCIENTIST
SPECIALIZED DEEP LEARNING«

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Komplexität und Laufzeit-analyse von Lern-algorithmen	die Komplexität gängiger ML Algorithmen erklären können.	x		
	die Grundlagen der Laufzeitanalyse erläutern können.	x		
	die Grundlagen gängiger numerischer Optimierungsverfahren in ML beschreiben können.	x		
	die theoretischen Grundlagen paralleler Algorithmen einordnen können.		x	
	Map Reduce in ML anwenden können.		x	
	die Vorteile der Parallelisierung von numerischen Optimierungsverfahren erläutern können.	x		
Implemen- tierung von Lernalgorithmen auf GPUs, Mehrkern- und verteilten Systemen	technische Grundlagen gängiger Hardwarebeschleuniger benennen können (GPUs, Xeon Phi, TPU).	x		
	gängige hardwarenahe Softwarebibliotheken für ML Algorithmen (cuDNN, MKL) anwenden können.		x	
Hardware- Architektur für ML-Systeme	Performance-Anforderungen von ML-Systemen beschreiben können (Rechenoperationen, I/O, Storage).	x		
	gängige ML System-Architekturen erläutern können.	x		
High Performance Computing für maschinelles Lernen	die Grundlagen von HPC-Systemen darstellen können (Netzwerk, IO/Storage, Batch-Verarbeitung).	x		
	Batch-Systeme nutzen können.		x	
	verteilte Parallelisierung einsetzen können (mit dem Message Passing Interface (MPI), Global Address Space Programm Interface (GPI)).		x	

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Implementierung von ML-Systemen in der Cloud / Big Data	die Architektur von Cloud Services am Beispiel AWS erläutern können.	x		
	die Grundlagen von Hadoop und Spark erklären können.	x		
	ML Bibliotheken in Hadoop und Spark einsetzen können (Mahout, MLlib).		x	

E 2.3.6 Vertiefungsbereich: Deep Learning for Textmining

ANLAGE E: »DATA SCIENTIST
SPECIALIZED DEEP LEARNING«

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Grundlagen der Textanalyse	die Aufgabenstellung des Textverstehens erläutern können.	x		
	die Repräsentation von Text beschreiben können.	x		
	die möglichen Vorteile des maschinellen Lernens für das Textverstehen erklären können.	x		
	Phasen der Textanalyse benennen können	x		
	Verfahren zur Textdatenbeschaffung benennen und erklären können	x		
	die Inhalte der wichtigsten Bibliotheken erläutern und diese in Python-Programmen exemplarisch verwenden können.		x	
Textklassifikation	die Aufgabenstellung der Klassifizierung anhand der Bestandteile Ein-/Ausgabedaten, Modellfunktionen und Kostenfunktionen anwenden können.	x	x	
	Repräsentation textueller Daten und Nutzung für Klassifikationsverfahren verstehen.	x		
	notwendige Vorverarbeitungsschritte erläutern können (Tokenisierung, BPE, Selektion des Vokabulars und der Merkmale, dünnbesetzte Datenmatrizen).	x		
	eine Vorverarbeitung in scikit.learn / TF implementieren können.		x	
die wichtigsten Klassifikationsmodelle (klassisch: logistische Regression, random forest, SVM; Deep Learning: MLP, CNN, RNN) erläutern können.	x			

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	Klassifikatoren zur Textverarbeitung für einfache Anwendungen entwerfen und in scikit.learn und TF umsetzen können.		x	
	Auswirkung der Anzahl der Merkmale (Features) auf das Verhalten der Klassifikationsmodelle und Vorgehen bei der geeigneten Auswahl an Merkmalen verstehen.	x		
	Konzepte von Ensemble Methoden verstehen.	x		
	Nutzung von Klassifikatoren in libraries wie sklearn und tensorflow verstehen und den Output solcher libraries interpretieren können.	x		
Semantische Ähnlichkeit	die Grundlagen der semantischen Ähnlichkeit erklären können.	x		
	Anwendungsszenarien für Embeddings (Exploration, Retrieval) beschreiben können.	x		
	die grundlegenden Konzepte des Topic Modelings verstehen.	x		
	Topic-Modelle für einfache Anwendungen entwerfen und umsetzen können.		x	
	Grundlagen des LDA-Modells, die dahinterliegende Theorie und den Output verstehen.	x		
	Grundlagen von Embeddingverfahren wie Word2Vec, FastText, Starspace verstehen	x		
	Embedding-Modelle für Textanalyseanwendungen entwerfen und in TF umsetzen können.			x
Entdecken von Entitäten	die Aufgaben der Entitätenentdeckung erklären können (POS, NER, shallow parsing, supersenses).	x		
	Architekturen der Entitätenentdeckung (CRFs, RNNs, CNNs) erläutern können.	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	Annotation von Entitäten und Relationen beschreiben können.	x		
	Datenschemata definieren und vergleichen können.		x	
	Daten mit geeigneten Tools annotieren/korrigieren können (BRAT, Stanford).		x	
	Disambiguierung von Entitäten und Entity Linking beschreiben können.	x		
	Entitätenentdeckung mit CRF für einfache Textanalyseanwendungen entwerfen und umsetzen können.		x	
	Entitätenentdeckung mit RNN für einfache Textanalyseanwendungen entwerfen und umsetzen können.		x	
	weiterführende Verfahren wie ELMO, BERT und Temporal Convolutional Networks kennen.	x		
Transformation von Wortfolgen / Sequenzen	die Bestandteile und Aufgaben der Transformation von Wortfolgen erläutern können (Sprachmodell, Übersetzung, Spracherkennung, Parsing).	x		
	moderne Architekturen zur Transformation von Wortfolgen (LSTM, seq2seq, WaveNet, attention-only) benennen können.	x		
	Anwendungsdetails (Kostenfunktionen, Performanzmaße, Attention, Beam Search) beschreiben können.	x		
	seq2seq für Textanwendungen entwerfen und in TF umsetzen können.		x	
Extraktion von Relationen	die Aufgaben REX (Relationsextraktion) erklären können (Relationen, Semantic Role Labeling, Parsing).	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	die Ansätze zur Relationsextraktion nennen und zuordnen können (überwacht, semi-supervised, distant supervision/open IE).	x	x	
	Modelle für überwachte Relationsextraktion entwerfen und in TF umsetzen können.		x	
Verstehen natürlicher Sprache	die Teilaufgaben des NLU (natural language understanding) nennen und erläutern können (Speech Recognition, Question Answering (QA), Semantic Parsing (SP), Answer Generation).	x		
	moderne Architekturen für chatbots, QA und SP beschreiben können.	x		
	ein vorgegebenes Modell für Question-Answering anwenden und interpretieren können.		x	

E 2.3.7 Vertiefungsbereich: Zeitreihenanalyse

ANLAGE E: »DATA SCIENTIST
SPECIALIZED DEEP LEARNING«

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Grundlegende Besonderheiten von Zeitreihen	den Unterschied von Zeitreihen zu anderen Daten erläutern können.	X		
	Beispiele von Zeitreihen-Daten nennen können.	X		
	die Definition von Zeitreihen Analyse benennen können.	X		
	den Unterschied zwischen den Modellklassen (State-Space Models vs. Modellen für die der Zustand mit einem Ausschnitt der Historie identisch ist) erklären können.	X		
Descriptive Analysis and Models	die Visualisierung von Zeitreihen anwenden können.		X	
	die Stationarität erklären können und wissen, wie man in der Praxis auf Stationarität prüft.	X	x	
	Beschreibende Statistiken von Zeitreihen erklären (rolling mean, autocorrelation, mean per intervall, mean per time part)	X		
	Verfahren zur Stationarisierung erklären	X		
	den Begriff Dekomposition (Sesionalität, Trend) erläutern können.	X	X	
	Spektral Analysis anwenden können (Fourier-Transformation).		X	
Modelle die direkt auf der Historie basieren (in Abgrenzung zu State Space Models)	AR-Modell anwenden können.	X	X	
	MA-Modell erläutern können.	X	X	
	ARMA und ARIMA erklären können.	X	X	
	Box Jenkins Verfahren zur Modellidentifikation anwenden können.	X		
	bei Makov-Ketten Übergangsmatrizen interpretieren können.	x		

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	die Darstellung von Markov Ketten als Graphische Modelle verstehen können.	x		
	die beiden Aufgabentypen, Parameter lernen und Wahrscheinlichkeit abschätzen, unterscheiden können.	x		
State Space Models	die drei Aufgabentypen, Parameter lernen, Wahrscheinlichkeit der Beobachtungen abschätzen und Hidden State schätzen erläutern können.	x		
	Bestandteile eines HMM Models erklären und Verbindungen zum Markov Chain Model beschreiben können.	x	X	
	beim Gaussian linear state-space die Aufgaben, Filtering, Smoothing, Model Learning unterscheiden können.	X	x	
Recurrent Neural Network (RNN)	das Prinzip der RNN beschreiben können.	X		
	LSTM anwenden können.		X	
	die Anwendung von Back Propagation through Time verstehen können.	x		
	die Architektur einer LSTM Zelle kennen.	x		
	Vorteil des „Constant Error Carousel“ verstehen können.	x		
	Beispiele für NN Architekturen, in denen LSTMs ein Baustein sind nennen können.	x		
	beim Echo state network die Architektur verstehen (kann Lernbare von zufällig initialisierten Gewichten unterscheiden) und die wesentlichen Vorteile von ESN benennen können und Bedeutung von Spektralradius und Anzahl der verborgenen Neuronen verstehen.	x	x	

Wissensgebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
Forecasting	das Prinzip von Vorhersagen Besonderheiten dabei (Train-test Split) erklären können.	x		
	Modelle zur Vorhersage anwenden können (z.B. ARMA, HMM, RNN).		X	
	Feature extraction für Vorhersage anwenden können.		x	
Klassifikation	<ul style="list-style-type: none"> - Das Prinzip der Klassifikation von Zeitreihen erklären können - Feature extraction für Klassifikation - Modelle zur Klassifikation anwenden. (z.B. RNN) 	x		

ANLAGE E: »DATA SCIENTIST
SPECIALIZED DEEP LEARNING«

Der »Data Scientist Specialized in Machine Learning« umfasst noch keine Themenbereiche, die er zwingend beurteilen können muss.

ANLAGE F: »DATA SCIENTIST ADVANCED LEVEL«

F 1 Verweis auf andere Normen und Dokumente

- EN ISO 17024

F 2 Anforderungsprofil

F 2.1 Bestimmung des Anforderungsprofils

Das Anforderungsprofil eines »Certified Data Scientist Advanced Level« ergibt sich aus der Charakteristik und Beschreibung seines Tätigkeitsfeldes.

Ein zertifizierter »Data Scientist Advanced Level«

- hat einen Überblick über die Data Science-Wertschöpfungskette,
- kann geeignete Methodik identifizieren, um Data Science Projekte im Allgemeinen und im Spezialgebiet durchzuführen,
- kann Methoden aus dem Bereich Data Science und in seinem Spezialgebiet anwenden,
- wendet Data Science-Methoden im beruflichen Kontext an.

Abgrenzungskriterien des »Certified Data Scientist Advanced Level« gegenüber anderen Profilen im Bereich Data Science sind im Zertifizierungshandbuch dokumentiert.

Die Bezeichnung lautet: »Certified Data Scientist Advanced Level«

F 2.2 Zugangsvoraussetzungen

F 2.2.1 Vorbildungen

Ein zertifizierter »Certified Data Scientist Advanced Level« muss nachweisen:

- Das Zertifikat als »Data Scientist Basic Level«
- Mindestens ein Zertifikat als »Data Scientist Specialized in...«,
- Mindestens ein Jahr Berufserfahrung mit der Durchführung von Data Science Projekten,
- Die Beschreibung der durchgeführten Tätigkeiten in einem mindestens zweimonatigen Data Science Projekt.

Art der Nachweise siehe unten.

Anmerkung:

Im zu prüfenden Einzelfall hat die antragstellende Person die Möglichkeit, fehlende Zugangsvoraussetzungen innerhalb von einem Jahr nach Ablegen der Prüfung nachzuweisen.

Der Nachweis der Berufserfahrung erfolgt

1.) über eine Bescheinigung des Arbeitgebers, die eine einjährige Tätigkeit im Bereich Data Science unter Angabe der durchgeführten Projekte und ausgeführten Tätigkeiten bestätigt oder bei Selbstständigen eine eidesstattliche Erklärung

und

2.) über eine Beschreibung der durchgeführten Tätigkeiten in einem mindestens zweimonatigen Data Science Projekt auf zwei bis drei DIN A 4 Seiten. Diese Tätigkeiten müssen vom Arbeitgeber bescheinigt werden.

Die Beschreibung sollte folgende Aspekte beinhalten:

- Kurze Zusammenfassung des Projektes mit der Beantwortung folgender Fragen (Maximal 0,5 Seiten):
 - Was war das Thema des Projekts?
 - Von wann bis wann lief das Projekt?
 - Was war der Business Case, was waren die Ziele des Projektes?
 - Welche Aktivitäten wurden in dem Projekt durchgeführt?
 - Was an dem Projekt war Data Science?
- Einordnung in die Data Science Landschaft (1 Seite):
(Falls folgende Punkte nicht anwendbar sind, begründen Sie bitte jeweils in einem Satz, warum nicht.)
 - Verortung der Aktivitäten im CRISP-DM mit der Beantwortung folgender Fragen:
 - Welche Aktivitäten wurden im Bereich Business Understanding durchgeführt und was waren die Ergebnisse?
 - Welche Aktivitäten wurden im Bereich Data Understanding durchgeführt und was waren die Ergebnisse?
 - Welche Aktivitäten wurden im Bereich Data Preparation durchgeführt und was waren die Ergebnisse?
 - Welche Aktivitäten wurden im Bereich Modeling durchgeführt und was waren die Ergebnisse?
 - Welche Aktivitäten wurden im Bereich Evaluation durchgeführt und was waren die Ergebnisse?
 - Welche Aktivitäten wurden im Bereich Deployment durchgeführt und was waren die Ergebnisse?
 - Welche Big-Data Systeme wurden verwendet? Begründen Sie Ihre Auswahl.
 - Welche Visualisierungen haben Sie verwendet und mit welchem Ziel?
 - Welche Security und Privacy Aspekte haben Sie in Ihrem Projekt adressiert?
- Beschreibung des Spezialgebietes (1,5 Seiten)
 - Was war die konkrete Aufgabenstellung im Spezialgebiet?
 - Darstellung der genauen Methodik mit erwogenen Alternativen und Begründung der finalen Auswahl.
 - Beschreibung der Umsetzung und (möglicher) aufgetretener Probleme
 - Beschreibung der Ergebnisse
 - Diskussion und Fazit

Die Fraunhofer-Personenzertifizierungsstelle behält sich vor die Angaben zu überprüfen. Nach Prüfung der eingereichten Unterlagen entscheidet die Fraunhofer-Personenzertifizierungsstelle über die Erfüllung der Voraussetzung. Sollten Zugangsvoraussetzungen nicht erfüllt sein, teilt die Fraunhofer-Personenzertifizierungsstelle dies dem Antragsteller unverzüglich über das Sekretariat der Fraunhofer-Personenzertifizierungsstelle mit.

Grundsätzlich kann die Fraunhofer-Personenzertifizierungsstelle in begründeten Ausnahmefällen davon abweichende Nachweise akzeptieren. Diese Nachweise und die Entscheidung der Fraunhofer-Personenzertifizierungsstelle sind zu dokumentieren.

F 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten

Ein »Certified Data Scientist Advanced Level« muss keine zusätzlichen Ausbildungen, Erfahrungen und Berechtigungen nachweisen.

F 2.2.3 Persönliche Voraussetzungen

Keine.

F 2.3 Geforderte Kompetenzen (Lernziele)

Im Zertifizierungsprofil »Data Scientist Advanced Level« entsprechen die geforderten Kompetenzen denen in den als Zulassungsvoraussetzung nachzuweisenden Zertifikaten. Zusätzlich wird über den Nachweis von Berufserfahrung u. a. im Rahmen der Beschreibung eines Projekts überprüft, ob die geforderten Kompetenzen in der Praxis Anwendung finden.

Das Zertifikat wird vergeben, wenn die Zulassungsvoraussetzungen erfüllt sind.

G 1 Verweis auf andere Normen und Dokumente

- EN ISO 17024

G 2 Anforderungsprofil

G 2.1 Bestimmung des Anforderungsprofils

Das Anforderungsprofil eines »Certified Senior Data Scientist« ergibt sich aus der Charakteristik und Beschreibung seines Tätigkeitsfeldes.

Ein zertifizierter »Senior Data Scientist«

- hat einen umfassenden Überblick über die Data Science-Wertschöpfungskette,
- beherrscht die Methoden des Data Science und in seinem Spezialgebiet,
- wendet die Data Science-Methoden im beruflichen Kontext zielgerichtet und effizient an,
- plant Data Science-Projekte,
- berät Stakeholder bei der Planung, Durchführung und Auswertung von Data Science-Projekten.

Abgrenzungskriterien des »Certified Senior Data Scientist« gegenüber anderen Profilen im Bereich Data Science sind im Zertifizierungshandbuch dokumentiert.

Die Bezeichnung lautet: »Certified Senior Data Scientist«

G 2.2 Zugangsvoraussetzungen

G 2.2.1 Vorbildungen

Ein zertifizierter »Senior Data Scientist« muss nachweisen:

- das Zertifikat als »Certified Data Scientist Advanced Level«
- mindestens zwei Jahre Berufserfahrung mit der Durchführung von Data Science Projekten
- eine eigenständig verfasste Studienarbeit über ein von der Fraunhofer-Personenzertifizierungsstelle zugelassenes Thema (Umfang ca. 40 DIN A 4 Seiten)

Anmerkung:

Im zu prüfenden Einzelfall hat die antragstellende Person die Möglichkeit, fehlende Berufserfahrung innerhalb von einem Jahr nach Ablegen der Prüfung nachzuweisen.

Der Nachweis der Berufserfahrung erfolgt über eine Bescheinigung des Arbeitgebers, die eine zweijährige Tätigkeit im Bereich Data Science unter Angabe der durchgeführten Projekte und ausgeführten Tätigkeiten bestätigt

Die Fraunhofer-Personenzertifizierungsstelle behält sich vor die Angaben zu überprüfen. Nach Prüfung der eingereichten Unterlagen entscheidet die Fraunhofer-Personenzertifizierungsstelle über die Erfüllung der Voraussetzung.

Sollten Zugangsvoraussetzungen nicht erfüllt sein, teilt die Fraunhofer-Personenzertifizierungsstelle dies dem Antragsteller unverzüglich über das Sekretariat der Fraunhofer-Personenzertifizierungsstelle mit.

Grundsätzlich kann die Fraunhofer-Personenzertifizierungsstelle in begründeten Ausnahmefällen davon abweichende Nachweise akzeptieren. Diese Nachweise und die Entscheidung der Fraunhofer-Personenzertifizierungsstelle sind zu dokumentieren.

Anforderungen an die Studienarbeit

Das Thema der Studienarbeit kann grundsätzlich frei aus dem Themenfeld Data Science gewählt werden, sofern es den unten stehenden Anforderungen genügt. Die Fraunhofer-Personenzertifizierungsstelle muss dem Thema zustimmen. Hierfür reicht der Teilnehmer ein Exposé ein. Dieses wird von der Fraunhofer-Personenzertifizierungsstelle geprüft.

Nach Zustimmung zum Exposé durch die Fraunhofer-Personenzertifizierungsstelle haben die Teilnehmer ein Jahr Zeit, das Thema zu bearbeiten und die Arbeit einzureichen. Es besteht jedoch auch die Möglichkeit ein Thema für eine Arbeit einzureichen, das der Teilnehmer bereits bearbeitet hat oder gerade bearbeitet. Bei Einreichung des Exposés darf der Abschluss der Arbeit allerdings nicht länger als ein Jahr her sein.

Eine Verlängerung der Einreichungsfrist ist nur in begründeten Ausnahmefällen mit Zustimmung der Leitung der Fraunhofer-Personenzertifizierungsstelle möglich.

Nach Einreichung der fertigen Arbeit prüft ein Prüfer der Fraunhofer-Personenzertifizierungsstelle, ob die Studienarbeit die Anforderungen erfüllt und spricht eine Empfehlung über die Anerkennung als Zugangsvoraussetzung für die mündliche Prüfung aus.

Insbesondere wird Wert darauf gelegt, dass es sich beim Verfassen der Arbeit um eine erkennbar eigenständige Leistung handelt. Plagiate werden nicht anerkannt. Die Begleitung der Projektarbeit durch einen Mentor ist gestattet.

Hat die Fraunhofer-Personenzertifizierungsstelle die Studienarbeit anerkannt und die übrigen Zulassungsvoraussetzungen sind ebenfalls erfüllt, wird der Teilnehmer zur Prüfung zugelassen.

Folgende **Inhalte und Gliederung** der ca. 40 Seiten umfassenden Arbeit werden erwartet:

1. Einleitung / Einführung des Themas

Aus der Einleitung sollte die Aufgabenstellung klar hervorgehen und gut motiviert werden. Sie sollte wenigstens folgende Punkte abdecken:

- Formulierung der Aufgabenstellung: Ausgangssituation und Ziele
- Motivation der Aufgabe: Herausforderungen und Neuigkeit, warum ist die Aufgabenstellung interessant und für wen? Welcher Wert wird der Lösung beigemessen?

2. Einordnung des speziellen Aufgabengebiets in den Kontext Data Science

In diesem Abschnitt sollen Aufgabe und Lösungsansätze in den Kontext existierender Ansätze im Bereich Data Science gestellt und auch alternative Lösungswege, die nicht beschränkt wurden, dargestellt werden. Zunächst soll der Zusammenhang zu den Themen der Zertifizierung dargestellt werden:

Wie und an welcher Stelle bestehen Zusammenhänge insbesondere zu folgenden Themengebieten?

ANLAGE G: »Senior Data Scientist«

- (Big) Data Analytics
- Big Data Systems
- Security and Privacy
- Data Management
- Big Data Potentials
- Visualisierung

Anschließend soll die Arbeit auch im Kontext des erworbenen Spezialisten-Zertifikats (z. B. „Data Scientist Specialized in Data Management“, „Data Scientist Specialized in Data Analytics“ oder „Data Scientist Specialized in Machine Learning“) im Hinblick auf die gewählte Vertiefung beleuchtet werden.

3. Herangehensweise an die Aufgabe/Fragestellung

Der Lösungsweg soll hier allgemeinverständlich und detailliert beschrieben werden.

Dabei soll auch die Verwendung bestehender Verfahren und Werkzeuge dokumentiert und in den Lösungsweg eingeordnet werden.

4. Evaluation

In diesem Abschnitt soll die Basis für die Bewertung der Güte der gefundenen Lösung gelegt werden. Dazu sollen geeignete Performance-Indikatoren ausgewählt, deren Auswahl begründet und die entsprechenden Ausprägungen bestimmt werden. Wo sinnvoll soll ein Vergleich zu alternativen Verfahren bzw. zu einer Baseline erfolgen.

5. Rekapitulation und Abschlussthese

Abschließend soll die Arbeit zusammenfassend dargestellt und bewertet werden.

Anmerkung: Bei der Erstellung der Arbeit sollte darauf geachtet werden, dass Formatierung und Zitierweise den Anforderungen wissenschaftlichen Arbeitens genügen. Der Inhalt der Arbeit muss explizit keinen wissenschaftlich-innovativen Charakter haben, wie dies beispielsweise bei einer Bachelor- oder Masterarbeit gefordert wird

G 2.2.2 Zusätzliche Ausbildungen/Berechtigungen und praktische Tätigkeiten

Ein »Certified Senior Data Scientist« muss keine zusätzlichen Ausbildungen, Erfahrungen und Berechtigungen nachweisen.

G 2.2.3 Persönliche Voraussetzungen

Keine.

G 2.3 Geforderte Kompetenzen (Lernziele)

Grundlage für die Bewertung der Projektarbeit sowie die mündliche Prüfung zum »Certified Senior Data Scientist« sind folgende nachzuweisende Kompetenzen (Lernziele):

Wissens- gebiet	Kompetenzen/Lernziele	ken- nen	an- wenden	beur- teilen
	Eine komplexe Data Science Aufgabenstellung unter Anwendung von Techniken aus dem Themenbereich des Zertifizierungsprogramms Data Science. lösen können	x	x	x
	Die Aufgabenstellung in den Kontext von Data Science einordnen können, insbesondere in die Themenfelder des Zertifikats „Data Scientist Basic Level“ und der gewählten Spezialisierung (z.B. Data Scientist Specialized in Data Management, Data Scientist Specialized in Data Analytics oder Data Scientist Specialized in Machine Learning“).	X	x	x
	Die Aufgabenstellung in einen (möglicherweise fiktiven) unternehmerischen Kontext einordnen können	x	x	X
	Den Lösungsweg für Personen mit einem Wissensstand auf dem Niveau eines „Zertifizierten Data Scientist Basic Level“ schriftlich darstellen können.	X	x	x
	Die Güte der gefundenen Lösung erläutern und bewerten können.	x	x	X