

SAS Data Scientist – Berufsbild mit Zukunft

Inhalt

Einleitung	3
Berufsbild mit Zukunft: Data Scientist	4
Die Bedeutung des Data Scientist	4
Aufgabengebiete des Data Scientist	5
Skills des Data Scientist	6
Die Basis: Fachliches Wissen, Methoden und Kreativität	6
Analytische Fähigkeiten	7
Kenntnisse im Datenmanagement	8
Programmierkenntnisse	9
Curriculum SAS® Data Scientists	10
SAS Ausbildung zum Data Scientist	11
• Warum sollte ein Data Scientist analytische Fähigkeiten haben?	11
• Warum sollte ein Data Scientist etwas über Datenaufbereitung und Datenqualität wissen?	12
• Warum muss ein Data Scientist programmieren können?	12
• Warum sollte der Data Scientist die Techniken der verständlichen Darstellung der Daten und Erkenntnisse beherrschen?	13
• Warum sollte der Data Scientist etwas von IT-Infrastruktur, IT-Architektur und neuen Techniken im Bereich Big Data verstehen?	13
Die Kompetenz von SAS Education	14
Ausbildungsberatung durch SAS Education Consultants	14

Einleitung

Daten gelten heute als einer der wichtigsten Rohstoffe überhaupt. In den meisten Unternehmen sind sie im Übermaß vorhanden. Allerdings kommt es darauf an, sie sinnvoll und wertschöpfend zu nutzen. Voraussetzung dafür ist es, sämtliche Daten zu erschließen und zu analysieren – sowohl die strukturierten in Form von Zahlen und Tabellen als auch unstrukturierte, beispielsweise Texte, Videos oder Website-Daten. Für diese umfassenden Analysen die richtigen Prozesse und Ressourcen zu schaffen, ist aktuell eine der drängendsten Herausforderungen für Unternehmen.

Big Data hat einen so starken Einfluss auf Geschäftsprozesse und -modelle, dass sich mittlerweile eine deutliche Entwicklung zu einer viel stärker datengetriebenen Wirtschaft abzeichnet. Unternehmen, die für das neue Datenzeitalter fit sein wollen, müssen in der Lage sein, jederzeit korrekte und fundierte Informationen aus der umfangreichen Datenmenge zu ziehen. Sie müssen die richtigen konkreten Schlüsse daraus ableiten und die Daten(-analysen) generell zur Verbesserung von Geschäftsprozessen und für die Entwicklung neuer Geschäftsmodelle im Unternehmen nutzen können.

Die neueste unabhängige Studie vom BARC-Institut im März 2014 belegt die wachsende Bedeutung von Big Data in Unternehmen. Hier einige Ergebnisse daraus:

- 30% der Unternehmen haben bereits Big-Data-Initiativen. Dabei geben 12% der Teilnehmer der Studie an, Big-Data-Initiativen seien fester Bestandteil ihrer Unternehmensprozesse, bei weiteren 18% befindet sich eine solche Initiative im Pilotstatus.
- 74% der Unternehmen wollen mit Big-Data-Initiativen neue und bessere Analysemöglichkeiten für ihre Daten erreichen.
- Die Big-Data-Analyse ist nützlicher als erwartet, aber Know-how-Engpässe bremsen die Umsetzung.
- Drei Viertel der Unternehmen verfügen durch Big Data Analytics über schnellere und detailliertere Analysemöglichkeiten, mehr als zwei Drittel können ihre operativen Prozesse besser steuern. 45% der Unternehmen mit Big-Data-Initiativen profitieren von gesenkten Prozesskosten. Fehlendes fachliches und technisches Know-how gehört zu den größten Herausforderungen in Big-Data-Projekten.
- 55% der Teilnehmer klagen über fehlendes fachliches und 47% über fehlendes technisches Know-how. Daher ist es für Unternehmen besonders wichtig, diesem Mangel mit geeigneten Schulungs- und Trainingsmaßnahmen entgegenzuwirken. Neue Berufsbilder wie der Data Scientist gehören dabei zu den attraktivsten Jobs in der IT-Branche.

Data Scientists arbeiten an der Schnittstelle zwischen Daten und Business. Ihre Aufgabe ist es, geschäftsrelevante Besonderheiten oder Zusammenhänge in den Daten eines Unternehmens zu identifizieren, zu analysieren und sie als Entscheidungsgrundlage für Management oder Fachabteilungen aufzubereiten. In jedem Unternehmen gibt es Mitarbeiter, die die Voraussetzungen mitbringen, sich zum Data Scientist weiterzubilden. Dafür sind Talent und Veranlagung, Flexibilität und Lust auf Neues erforderlich. Und eine Grundausstattung an technisch-methodischem Know-how.

Berufsbild mit Zukunft: Data Scientist

Daten liegen heute in einer breiten Vielfalt vor: in Text-Dateien, grafischen Dokumenten, Tabellen und Datenbanken genauso wie inzwischen auch auf Websites oder in sozialen Netzwerken, also in allen über das Internet erreichbaren Medien.

Für die Analyse und Aufbereitung der Daten gibt es heute extrem leistungsfähige Lösungen. Aber man braucht dafür Menschen, die es verstehen, Analysemodelle zu entwickeln, Auswertungen zu interpretieren und in einen Business-Kontext zu setzen. Das ist die Aufgabe des Data Scientist. Klar definiertes Ziel für diese Generalisten an der Schnittstelle von Daten und Geschäft: konkrete Optimierungs- oder Änderungsvorschläge für das Unternehmen zu formulieren.

Daraus ergibt sich für den Data Scientist ein sehr vielseitiges und abwechslungsreiches Anforderungsprofil: Er sollte sich mit System-Architekturen und Datenstrukturen auskennen und umfangreiches Wissen in Statistik und linearer Algebra mitbringen. Und er braucht Fachexpertise, Business-Kompetenz und die Fähigkeit, das Management im Unternehmen zu beraten und zu überzeugen. Kurz: Der Data Scientist verbindet gekonnt technisches Know-how mit Business-Kompetenz sowie Kommunikationsstärke und setzt diese Fähigkeiten im Sinne eines strategischen Mehrwerts für das Unternehmen ein.

Der Weg zum Data Scientist kann unterschiedlich verlaufen und hängt sowohl vom Unternehmen als auch vom jeweiligen Mitarbeiter ab: Manche kommen aus dem technisch-naturwissenschaftlichen Bereich und haben sich das notwendige Business-Know-how angeeignet. Andere starten mit Fachabteilungsperspektive, beispielsweise aus Marketing, Vertrieb oder Controlling, und bilden sich in den erforderlichen technischen Disziplinen weiter.

Worauf es ankommt, sind in jedem Fall mathematische Kenntnisse mit Schwerpunkt Statistik und fundiertes Informatikwissen – in Kombination mit Fachwissen.

Die Bedeutung des Data Scientist

Mit der steigenden Relevanz von Daten jeglicher Ausprägung als Basis für strategische Weichenstellungen nehmen auch Verantwortung sowie Breite und Tiefe des Tätigkeitsfelds des Data Scientist zu. Die Entscheidungen, die auf Basis seiner Arbeit fallen, betreffen alle Kernbereiche eines Unternehmens und seiner Strategie. Die Aufgabe des Data Scientist ist deshalb eingebunden in den strategischen Entscheidungsprozess des Unternehmens. Entsprechend wichtig ist es für Unternehmen, die richtigen Talente für diese Funktion zu identifizieren, zu fördern und weiterzuentwickeln. Keine leichte Aufgabe: Schon 2011 hat eine McKinsey-Studie den Mangel an Kandidaten für diese Spezialaufgabe als überaus kritisch bewertet. Und weiter stellt McKinsey fest, dass sich diese Situation in den nächsten Jahren weiter dramatisch verschärfen wird.

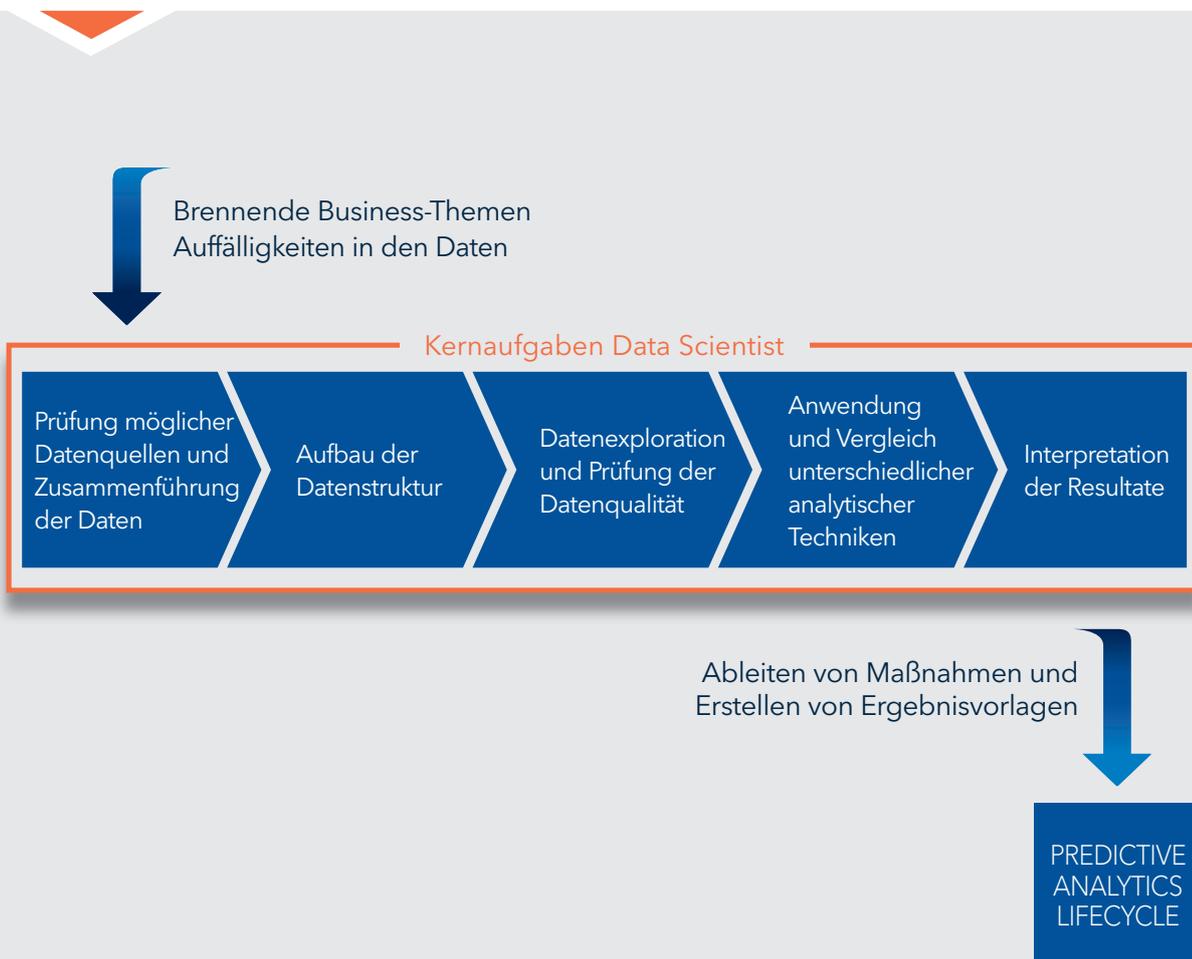
Aufgabengebiete des Data Scientist

Die primäre Aufgabe des Data Scientist ist das effiziente Entdecken von geschäftsrelevanten Potenzialen durch den Einsatz von modernen, kreativen Analysetechniken bei steigenden Datenmengen.

Der Data Scientist nutzt zwei Quellen, um Potenziale aufzudecken. Er hält einerseits engen Kontakt zu den Fachbereichen und stößt dabei auf die brennenden Themen und Fragestellungen. Andererseits nutzt er die Daten und kreative Analysemethoden, um Auffälligkeiten nachzugehen und das Management oder Fachabteilungen über mögliche geschäftsrelevante Potenziale zu informieren und Entscheidungsvorlagen zu entwerfen.

Es geht darum, schnell Ergebnisse zu erzielen und nachzuweisen, dass Spezifikationen oder Ideen tauglich sind. Das Vorgehen des Data Scientist ist daher vom explorativen Prototyping geprägt. Der Fokus liegt auf der Bestimmung der Anforderungen und der Beurteilung möglicher Problemlösungen. Erst wenn der Business-Nutzen nachgewiesen ist und vom Management getragen wird, kann die Funktionalität in ein produktives System überführt werden. Entwicklung und Betrieb der Modelle im Predictive Analytics Lifecycle wird dabei von anderen Rollen in der Organisation übernommen.

Für den Data Scientist bedeutet das, rasch seine Datenstruktur auszubauen, geeignete Analysetechniken zu wählen und aus den Ergebnissen unmittelbar Konsequenzen für das Business abzuleiten.



Zu den Fähigkeiten, die ein Data Scientist mitbringen sollte, gehört eine Kombination aus Programmierkenntnissen, analytischem Hintergrund und Business-relevantem Know-how. Die Schwerpunkte können je nach Unternehmen variieren und sind abhängig von den konkreten Aufgaben.

In jedem Fall aber ist die Rolle des Data Scientist sehr facettenreich, und er sollte über Talente in diesen drei Bereichen verfügen:

- Fachliches Wissen, Methoden und Kreativität
- Mathematik/Statistik/Operational Research/Forecasting
- Programmierung/Data Management

Skills des Data Scientist

Die Basis: Fachliches Wissen, Methoden und Kreativität

Die wesentliche Aufgabe eines Data Scientist ist es, durch seine Analysen einen konkreten Wert in das Unternehmen zu tragen. Dafür muss er die Aufgaben der Fachabteilungen verstehen und im Idealfall bei Auffälligkeiten in den Daten nützliche Fragestellungen eigenständig formulieren und mithilfe von Analyseergebnissen beantworten können. Bei der Identifikation der richtigen Fragestellungen hilft ihm die Technik der Informationsbedarfsanalyse.

In seinem Job geht es darum, über eine kreative Herangehensweise an die Analyse rasch zu geschäftsrelevanten Ergebnissen zu kommen und auf dieser Basis Entscheidungsvorlagen für das Management zu bauen. Dafür sind unter anderem Techniken wie das Rapid Prototyping aus dem Modellbau hilfreich. Dieses Prinzip auf die Tätigkeiten des Data Scientist übertragen, bedeutet: rasch und effizient aussagekräftige erste Analyseergebnisse zu liefern. Die daraus abgeleitete Entscheidungsvorlage erlaubt es dem Management, auf einen Blick den möglichen Wertbeitrag für das Unternehmen zu erkennen.

Der Data Scientist braucht zudem gute Kommunikationsfähigkeiten, um fachliche Themen zu diskutieren und die Ergebnisse auch Gesprächspartnern ohne Statistik-wissen zu vermitteln. Er nimmt vor diesem Hintergrund die Rolle eines „Übersetzers“ ein, der rein technische Analyseergebnisse aus der Welt der Daten und Methoden in Business-orientierte Entscheidungsvorlagen überträgt. Er muss auch die Ergebnisse mit den Fachexperten diskutieren und die Konsequenzen der Maßnahmen bzw. Entscheidungsvorlagen einschätzen und darlegen können.

Seine Aufgaben sind häufig abteilungsübergreifend, die Daten stammen aus unterschiedlichen Quellen und „gehören“ oftmals verschiedenen Bereichen. In diesem Zusammenhang geht es daher zunächst um die Informationsbeschaffung. Hier klärt der Data Scientist, welche Information zur Beantwortung einer Business-orientierten Frage erforderlich ist, woher die Daten bezogen werden können, in welcher Form sie vorliegen und ob ggf. eine Transformation der Daten erforderlich ist. Besondere Herausforderung bei einer solchen Informationsbeschaffung: Der Data Scientist muss nicht nur sämtliche Datenquellen im Unternehmen kennen, sondern auch die jeweils Datenverantwortlichen, um die Daten zeitnah und in ausreichender Qualität zu beschaffen. Dazu ist es hilfreich, mit den Quellsystemen vertraut zu sein und Grundkenntnisse in Datenbanken zu haben.

Eine weitere Aufgabe jedes Data Scientist ist der unternehmensweite sichere Umgang mit den Daten. So ergeben sich im Zuge der Informationsbeschaffung natürlich auch Fragen, die Datenschutz, Data Governance und Compliance betreffen. Zum Beispiel: Welche Daten dürfen aus datenschutzrechtlicher Sicht überhaupt verwendet werden? An welcher Stelle muss eine Anonymisierung erfolgen? Wer im Unternehmen darf welche Informationen einsehen? Die Analyseergebnisse sind häufig über viele Hierarchieebenen zu kommunizieren und somit für verschiedene Ebenen zu diskutieren und aufzubereiten. Ein Data Scientist, der seine Rolle optimal ausfüllt, muss deshalb auch ein guter Networker und begabter Projektmanager sein. Schließlich übernimmt er ja nicht alle Aufgaben selbst, beispielsweise die Beschaffung der Daten oder die Interpretation der Analyseergebnisse und das Ableiten der Maßnahmen mit den Fachabteilungen. All dies will aber koordiniert und organisiert sein.

Bereich	Aufgabe	Level
Business-Know-how	Expertise für die Fachbereiche und Branche	●●●
	Formulierung von Fragestellungen	●●●
	Erkennen von Auffälligkeiten	●●
	Übersetzung der statistischen Ergebnisse in fachbereichsrelevante Erkenntnisse	●●●
	Sensibilität der Ergebnistiefe für unterschiedliche Ebenen im Management	●●●
	Kontakt zum Vorstand	●●●
	Data Governance & Compliance	●●●
Methoden	Informationsbedarfsanalyse	●●●
	Rapid Prototyping	●●●
	Projektmanagement	●
	Präsentationstechniken	●●●
Soft Skills	Kreativität (Methode, Fragestellung, Daten, Visualisierung)	●●●
	Neugierde	●●●
	Verständnis für den Wertbeitrag der Datenanalyse	●●●
	Kommunikation	●●●
	Networking	●●●

● Grundkenntnisse ●● vertiefende Kenntnisse ●●● Expertenkenntnisse

Analytische Fähigkeiten

Die Aufgabe des Data Scientist ist es, große Datenmengen auf Business-relevante Besonderheiten oder Zusammenhänge zu untersuchen und diese für das Management aufzubereiten.

Der Data Scientist sollte daher die Fähigkeit besitzen, solche Besonderheiten in den Daten zu erkennen, daraus Hypothesen abzuleiten, diese statistisch zu prüfen und die Ergebnisse in geeignete Entscheidungsgrundlagen zu überführen. Je nach Aufgabenstellung und Entscheidung des Managements ist oft auch eine Umsetzung in ein Modell und die Übergabe der Analysetechniken (zum Beispiel: Scoring, Bildung von Clustern, etc.) an den IT-Betrieb notwendig. Die üblichen Aufgaben/Techniken dabei sind:

Bereich	Aufgabe	Level
Methoden	Data-Mining-Methodologie	•••
	Analyse von Big Data	••
Explorative Datenanalyse	Grundlegende grafische Methoden (Boxplot, Histogramm, Scatterplot, Mosaikplot)	•••
	Annahmen für statistische Inferenz einschätzen	•••
	Auswahl geeigneter statistischer Werkzeuge und Techniken	•••
	Aufdecken von Daten- und Abhängigkeitsstrukturen	•••
	Analyse und Beurteilung von Ausreißern	•••
Analytische Methoden	Testen von Hypothesen und ANOVA	••
	Multivariate Statistiken	••
	Zeitreihenanalyse	•••
	Regressionen, neuronale Netze	••
	Clustering, Segmentierung	•••
	Entscheidungsbäume	•••
	Assoziationsanalyse	••
	Optimierung	•
	Maschinelle Sprachverarbeitung	•
	Text Analytics	••
Präsentation	Visualisierung	••

• Grundkenntnisse •• vertiefende Kenntnisse ••• Expertenkenntnisse

Um all diese Techniken zu beherrschen, ist eine sehr gute statistische Ausbildung notwendig. Nicht alle werden in jedem Unternehmen eingesetzt - die Bandbreite an eingesetzten Techniken ist von individuellen Fragestellungen und Datenlandschaften abhängig. Querdenken und Kreativität sind gefragt. Die Anwendung etablierter statistischer Methoden aus ganz verschiedenen Bereichen - zum Beispiel Verfahren aus der Biometrie für die Beantwortung Business-relevanter Fragestellungen - kann zu ganz neuen Erkenntnissen führen. Oft ist der Data Scientist auch Mitglied eines Teams; Statistiker und Ökonometriker übernehmen in diesem Fall die komplexeren statistischen Methoden.

Kenntnisse im Datenmanagement

Auch für den Data Scientist gilt die Grundregel aller Business Analytics-Anwendungen: Ein Großteil des Aufwands entsteht für das Datenmanagement. Somit sind Kenntnisse aus allen Bereichen des Datenmanagements für den Data Scientist eine unabdingbare Notwendigkeit. Um schnelle Ergebnisse zu erzielen, wird häufig auf Programmiersprachen zurückgegriffen.

Bereich	Aufgabe	Level
Methoden	Datenmodellierung	••
	Master Data Management	•
	IT-Architekturen und IT-Infrastruktur	•
Daten-Akquirierung	Kenntnisse über Datenbanksysteme	•
	Zugriff auf Daten aus unterschiedlichen Quellen wie RDBMS, Hadoop usw.	••
	Beschaffung von Daten aus dem Internet	•
	Arbeiten mit unstrukturierten Daten	•••
	Umgang mit Big Data	••
	Zusammenführen und Aufbereiten der Daten	•••

Bereich	Aufgabe	Level
Daten-Aufbereitung	Umsetzung von Datenbereinigungsmaßnahmen	••
	Zusammenführen und Aufbereiten der Daten	•••
	Aufbau analytischer Datenstrukturen	•••
Explorative Daten-analyse	Grundlegende grafische Methoden, Prüfung von Annahmen, Aufdecken von Datenstrukturen und Abhängigkeiten, Identifizierung von Ausreißern	•••
	Visualisierung	••
Präsentation	Visualisierung	••

• Grundkenntnisse •• vertiefende Kenntnisse ••• Expertenkenntnisse

Auch hier gilt der Prototype-Ansatz, und erst nach der positiven Entscheidung des Managements über den Einsatz von Analytics zur Beantwortung eines Business-relevanten Themas kann der Lösungsansatz in die Produktion überführt werden. Viele Arbeitsschritte im Bereich Data Management, die der Data Scientist durch Einsatz von (schnellem) Programmcode löst, werden jetzt mit den etablierten ETL-Werkzeugen (Extract-Transform-Load) abgebildet und in den regelmäßigen IT-Betrieb übernommen. Um die Übergabe unterstützen zu können, sollte der Data Scientist zumindest Grundkenntnisse in der eingesetzten Data Management-Technologie haben. Während der Inbetriebnahme der analytischen Methode wird der Prototype in ein den „Predictive Analytics Lifecycle“ überführt.

Programmierkenntnisse

Ein effizientes Werkzeug zur Lösung der verschiedenen Schritte bei der Bearbeitung eines Problems sind nach wie vor Programmiersprachen. Gerade für die prototypische Herangehensweise sind sie bestens geeignet, um schnell und flexibel auf Veränderungen reagieren zu können. Beispiele dafür sind: das rasche Bilden neuer Variablen, die Hinzunahme neuer Datenquellen, das Testen verschiedener Transformationen, das Identifizieren von Ausreißern, das Aufdecken von Datenqualitätsproblemen, der Einsatz von Imputationstechniken, aber auch die Visualisierung im Bereich Datenexploration sowie unterschiedliche Analysetechniken auszuprobieren und zu vergleichen. Daher sind Programmierkenntnisse eines der drei wichtigsten Standbeine eines Data Scientist.

Für die obigen Aufgaben gibt es unterschiedliche Programmiersprachen und Techniken: Java, Perl, SQL, SAS, R, Microsoft Excel sind nur einige davon. Der Data Scientist muss nicht alle davon beherrschen. Die in seinem Unternehmen verwendeten Werkzeuge sollte er jedoch sehr gut kennen.

Curriculum SAS® Data Scientists

Die Aufgaben des Data Scientist sind vielfältig, herausfordernd und verlangen besondere Fähigkeiten. Die wesentlichen Fähigkeiten teilen sich grundsätzlich in drei Bereiche:

- Analytik: Statistik, Mathematik, Forecasting und Operation Research
- Programmierung, Data Management und
- ein Fundament von allgemeinen, übergreifenden Talenten (z.B. Kreativität, Networking, Querdenken)

Die Herangehensweise des Data Scientist gliedert sich in viele Schritte, bei denen häufig Teile daraus oder gar der gesamte Ablauf in Schleifen mehrmals durchlaufen werden müssen. In der nachfolgenden Grafik sind für jeden Arbeitsschritt die essenziellen Fähigkeiten aufgelistet.

Basis Skills des Data Scientist

- Kreativität beim Einsatz analytischer Verfahren
- Querdenken
- Rapid Prototyping
- Projektmanagement
- Verständnis für den Wertbeitrag für das Unternehmen

Brennende Business-Themen

- Kreativität und Networking
- Informationsbedarfsanalyse

Auffälligkeiten in den Daten

- SAS Programmierung ✓
- Arbeiten mit unstrukturierten Daten ✓
- Kreative und komplexe Abfragen ✓

Prüfung möglicher Datenquellen und Zusammenführung der Daten

- SAS Programmierung ✓
- Arbeiten mit unstrukturierten Daten ✓
- Kreative und komplexe Abfragen ✓
- Data Governance und Compliance

Aufbau der analytischen Datenstruktur

- SAS Programmierung ✓
- Komplexe Abfragen

Datenexploration und Prüfung Datenqualität

- SAS Programmierung ✓
- Kreative und komplexe Abfragen ✓
- Grundlegende Statistik und Daten Exploration ✓
- Visualisierung ✓

Anwendung und Vergleich unterschiedlicher analytischer Techniken

- Strategien und Konzepte für den Data Scientist ✓
- Grundlegende Statistik und Daten Exploration ✓
- Zeitreihenanalyse ✓
- Multivariate Statistiken ✓
- Regression, neuronale Netze, Entscheidungsbäume und Assoziationsanalyse ✓
- Clustering und Segmentierung ✓
- Text Analytics und Sentiment Mining ✓
- Modellierung von seltenen Ereignissen, fortgeschrittene Variablen, Selektion und Data Mining von zeitlichen Daten ✓
- Lösen von Optimierungsproblemen ✓
- High-Performance Analytics ✓
- Maschinelle Sprachverarbeitung
- Simulationen
- Kreative und komplexe Abfragen ✓
- Kreative Visualisierung der Analyse-Ergebnisse ✓

Interpretation der Resultate

- Kenntnisse der statistischen Methoden
- Fachliches Wissen
- Customer Relationship Management
- Vertrieb und Marketing
- Produktion
- Risikomanagement

Ableiten von Maßnahmen und Erstellen von Ergebnisvorlagen

- Verständnis für den Wertbeitrag für das Unternehmen
- Präsentationstechniken
- Kommunikationstechniken

Analytische Modelle in Produktion setzen

- IT-Architektur, SAS® Grid, Hadoop, SAS und SAP HANA® ✓
- Data Management ✓
- High-Performance Analytics ✓

Die durch ein ✓ gekennzeichneten Themen werden im Ausbildungsprogramm von SAS Education zum Data Scientist abgedeckt. Das modulare Programm ermöglicht dem Teilnehmer, seine bereits vorhandenen Kenntnisse zu ergänzen und so das Fähigkeitsprofil des Data Scientist zu vervollständigen.

Bleibt nur noch die Frage: Wie wird man Data Scientist?

SAS Ausbildung zum Data Scientist

SAS® hat ein umfassendes Ausbildungsprogramm für den SAS® Data Scientist zusammengestellt, das berücksichtigt, dass die Teilnehmer unterschiedliche Kenntnisse aus der IT bzw. Statistik mitbringen.

Warum sollte ein Data Scientist analytische Fähigkeiten haben?

Die Arbeitsweise des Data Scientist ist vergleichbar mit der von Christoph Kolumbus: immer auf der Suche nach neuen Erkenntnissen, die einen wertvollen Beitrag für das Unternehmen liefern und keine Scheu, dabei neue Wege zu bestreiten. Zum Aufbau grundlegender statistischer Kenntnisse und Techniken der Datenexploration empfiehlt SAS, abhängig von den jeweils eingesetzten SAS Produkten, folgende Kurse:

Kursname	WebCode 
SAS® Visual Analytics: Data Explorations und Reporting	YVAER71
Statistik 1: Varianzanalyse, Regression und logistische Regression	ST1131
Statistik 2: weiterführende Techniken der Varianzanalyse und Regression	ST2131
SAS® Enterprise Guide®: ANOVA, Regression und logistische Regression	EGBS61

Im Bereich der analytischen Methoden sowie statistischen Techniken ist es notwendig, die Vielfalt der Techniken und deren Einsatzmöglichkeiten zu kennen. Genau diese Informationen werden beispielhaft in folgenden Kursen vermittelt:

Kursname	WebCode 
Strategien und Konzepte für den Data Scientist und Business Analyst	BADSC
Analytics: Putting It All to Work	BAPUT
Advanced Analytics in a Big Data World	BDMC13
Data Scientist: Building Recommender Systems with SAS® and Hadoop®	BIDSBRS
Social Network Analysis for Business Applications	BSNA

Je nach Business-relevantem Thema im Unternehmen ist es für den Data Scientist vorteilhaft, detailliertere Kenntnisse über die speziellen statistischen Methoden zu erwerben. SAS bietet im Bereich Datenanalyse folgende vertiefende Ergänzungskurse:

Kursname	WebCode 
Komplexe Daten mit multivariater Statistik analysieren	MULT94
Angewandte Clustertechniken	CLUS94
Neural Network Modeling	DMNN71
Einführung in die Datenanalyse mit der SAS® Enterprise Miner™ Software	AAEM131
Weiterführende Analysen mit der SAS® Enterprise Miner™ Software	AAEM212
Vertiefende Modellierungsverfahren mit Entscheidungsbäumen	DMDT71
Customer Segmentation Using SAS® Enterprise Miner™	BEMC121
Text Analytics and Sentiment Mining Using SAS®	BTASM
SAS® Content Categorization Studio: Building Models	TACCIN
SAS® Sentiment Analysis Studio: Building Models	TASAS
Textanalyse mit der SAS® Text Miner™ Software	DMTX131
Statistische Prognosen mit der SAS® Software: ein Programmieransatz	FETSP
Formulieren und Lösen von Optimierungsmodellen mit SAS/OR®	OROP94
SAS® Visual Statistics: interaktive Modellanpassung	SVS071

 = Mit dem WebCode direkt zu Inhalten und Anmeldung: www.sas.de/education

Warum sollte ein Data Scientist etwas über Datenaufbereitung und Datenqualität wissen?

Die Suche nach brauchbaren, auswertbaren Daten ist eine wesentliche Aufgabe des Data Scientist. Das Prüfen der Daten auf Datenqualität ist dabei ein zentraler Schritt. Denn nur Daten, die den Qualitätsanforderungen entsprechen, können brauchbare Analyseergebnisse liefern, die eine Business-relevante Verallgemeinerung zulassen.

Soll aus einer Auffälligkeit auch ein prädiktives Modell entstehen, muss der Data Scientist für die geeignete Datenstruktur sorgen. Da diese Datenaufbereitungsschritte sehr zeitintensiv sind, ist es immens wichtig, dass der Data Scientist mit den entsprechenden Werkzeugen vertraut ist sowie effizient und schnell mit ihnen arbeiten kann.

Der Beurteilung der Daten hinsichtlich Datenqualität, deren Auswirkung auf die Modelle und dem effizienten Aufbau der geeigneten Datenstruktur widmen sich der Kurs:

Kursname	WebCode
Datenaufbereitung und Datenqualität für Analytics	BDADM2

Häufig kommen heute auch Visualisierungstechniken bei der Beurteilung der Datenqualität und der Aussagekraft von Daten zum Einsatz. Dafür ist es erforderlich, die Daten in einer bestimmten Struktur in die Visualisierungstools zu laden und die geeignete Datengrundlage zu schaffen. Die Fähigkeit, Daten effektiv in SAS Visual Analytics zu laden, vermittelt der Kurs:

Kursname	WebCode
SAS® Visual Analytics: Datengrundlagen erstellen	YVDB71

Warum muss ein Data Scientist programmieren können?

Der Data Scientist arbeitet direkt mit den Daten. Zu seinen Aufgaben zählt das Abziehen von Daten aus unterschiedlichen Quellen, das Zusammenführen von Daten, das Herstellen einer geeigneten Analysestruktur, das Aufdecken von Auffälligkeiten in den Daten, die Beurteilung und Verbesserung der Datenqualität sowie das Bereinigen von Daten. Alle diese Aufgaben sind sehr effizient und vor allem wiederverwendbar in Base SAS bzw. mittels der Makrosprache zu programmieren.

Kursname	WebCode
SAS® Programmierung 1: Grundlagen	PRG1
SAS® Programmierung 2: Datenmanagement im Data Step	PRG2
SAS® Makrosprache 1: Grundlagen	MAC1
SAS® SQL 1: Grundlagen	SQL1

Der SAS® Enterprise Guide ist der „Werkzeugkasten von SAS“. Mit diesem Tool lassen sich grundlegende Aufgaben aus dem Bereich Datenzugriff, Datenmanagement und Datenpräsentation effizient durchführen. Dafür leistet dieser Kurs Hilfestellung:

Kursname	WebCode
SAS® Enterprise Guide® 1: Abfragen und Berichte	EG171

 = Mit dem WebCode direkt zu Inhalten und Anmeldung: www.sas.de/education

Warum sollte der Data Scientist die Techniken der verständlichen Darstellung der Daten und Erkenntnisse beherrschen?

Eine der vielen Rollen des Data Scientist ist auch die des Übersetzers, der die analytischen Ergebnisse und Auffälligkeiten in Business-relevante Maßnahmen und Konsequenzen überführt. Zu seinen Aufgaben gehören einerseits die Interpretation der analytischen Ergebnisse, das Aufbereiten von Entscheidungsvorlagen für das Management und die Darstellung der Auswirkungen auf das Unternehmen sowie andererseits das Ableiten von Konsequenzen daraus. Dabei ist es hilfreich, die Ergebnisse beispielsweise einer logistischen Regression grafisch darzustellen, um eine klare Verständlichkeit zu erreichen. Dazu nutzt der Data Scientist neben den bekannten Techniken der Visualisierung (Plots, Histogramme etc.) auch neue Techniken, wie sie SAS Visual Analytics liefert.

Um die Techniken der Visualisierung und das Erstellen der notwendigen Datengrundlagen zu erlernen, stehen folgende Kurse zur Verfügung:

Kursname	WebCode
SAS® Visual Analytics: Datengrundlagen erstellen	YVDB71
SAS® Visual Analytics: Data Explorations und Reporting	YVAER71

Warum sollte der Data Scientist etwas von IT-Infrastruktur, IT-Architektur und neuen Techniken im Bereich Big Data verstehen?

Auch wenn der Aufbau und Betrieb einer IT-Infrastruktur und -Architektur in der Regel nicht zu den Kernaufgaben eines Data Scientist gehören, sollte er Bescheid wissen, welche Möglichkeiten in diesen Technologien beim Umgang mit großen Datenmengen stecken, und wie sich diese nutzen lassen. SAS bietet dazu folgende Kurse an:

Kursname	WebCode
Anwendungen und Architektur der SAS® Business Analytics Plattform im Überblick	AROV
Datenmodellierung und Informationsmanagement mit Big Data	BDADM1

Beim Einsatz von Hadoop stehen folgende Kurse zur Verfügung:

Kursname	WebCode
Einführung in die In-Memory Statistiken für Hadoop®	HADPIGHV
Introduction to SAS® and Hadoop®	STHAGS
Datenverarbeitung in Hadoop® mit Pig und Hive	DIACCHAD

Ergänzend/abhängig vom SAS® Portfolio

Kursname	WebCode
Using SAS® Client Applications with SAS® Grid Manager	PCCL
Predictive Modeling Using SAS® High-Performance Analytics Procedures	PMHP123
Vorhersagen mit der SAS® High-Performance Software	HPF123

 = Mit dem WebCode direkt zu Inhalten und Anmeldung: www.sas.de/education

Die Kompetenz von SAS Education

Das Berufsbild des Data Scientist in der beschriebenen Schnittstellenfunktion ist neu. Die damit verbundenen Aufgaben und Qualifikationen bilden jedoch seit Jahrzehnten das Kerngeschäft von SAS. Als Softwareanbieter, der Unternehmen laufend bei der Zusammenführung von Daten und Business begleitet, verfügen wir über fundierte Erfahrung in diesem Bereich. Unsere Experten kennen die Bedürfnisse von Fachabteilungen und Management und sind mit den spezifischen Anforderungen sämtlicher Branchen vertraut. Auf dieser Basis können wir Data Scientists so ausbilden, dass sie für die Praxis bestens gerüstet sind – und den analytischen Reifegrad eines Unternehmens auf ein neues Niveau heben können.

Das SAS Education Team betreut den Kunden in jeder Phase der Ausbildung und steht von der Planung und Organisation bis zur professionellen Durchführung der Kurse zur Verfügung. SAS Education verfügt über mehr als 30 Jahre Erfahrung in der Konzeption und Durchführung von SAS Schulungen. SAS schult weltweit jährlich mehr als 30.000 Anwender. In der deutschsprachigen Region werden mehr als 100 verschiedene Kurse angeboten. SAS Heidelberg wurde im International Training Center Rating von PerformNet mit „5 Sternen“ ausgezeichnet. Unsere Schulungen tragen das Siegel „SAS Accredited Training“ – es bescheinigt aktuellste Kursinhalte, akkreditierte Trainer, hochwertige Schulungsunterlagen und eine optimale Infrastruktur. 97 % der Kursteilnehmer würden die von ihnen besuchten Kurse weiterempfehlen.

Unsere akkreditierten Trainer stehen für höchste Qualität der Schulungen. Fachlich fundiert, didaktisch und methodisch versiert vermitteln sie Grundlagen und Spezialwissen, Theorie und Praxis – und Begeisterung für das Berufsbild des Data Scientist.

Ausbildungsberatung durch SAS Education Consultants

Tätigkeitsfelder des Data Scientist sind vielfältig und unterscheiden sich von Unternehmen zu Unternehmen. Kriterien, nach denen sich die spezifischen Aufgaben im Einzelfall bestimmen, sind zum Beispiel die tatsächlichen Business-relevanten Fragestellungen, die Aufgabenverteilung und die bereits vorhandenen Kompetenzen. Häufig gibt es schon einen Statistiker im Unternehmen, der die Aufgaben der Analyse, Modellierung, Testen etc. übernehmen kann. In diesem Fall ist der Data Scientist etwa für die Generierung von neuen fachlichen Fragestellungen, die Suche nach Auffälligkeiten in Daten, die bisher unentdeckt geblieben sind, die Erstellung von Entscheidungsvorlagen aufgrund der interpretierten Ergebnisse oder die Kommunikation mit den Entscheidungsträgern zuständig. Der Ausbildungsplan für einen angehenden Data Scientist ist insofern ganz individuell zu erstellen. Folgende Fragen sind dafür zu beantworten:

- Welche Aufgaben wird der Data Scientist im Unternehmen übernehmen?
- Welche IT-Architektur und Werkzeuge sind im Einsatz, welche Technologien kann der Data Scientist nutzen?
- Welche Schnittstellen zu bestehenden Abteilungen/Geschäftsbereichen gibt es?
Wie wird die Zusammenarbeit aussehen?
- Welche Vorkenntnisse hat der Mitarbeiter, der zum Data Scientist ausgebildet werden soll?
Wo sind die Lücken zum Anforderungsprofil? Welche Kurse sind erforderlich?

SAS bietet dazu eine Ausbildungsberatung an. Hier erarbeiten erfahrene SAS Education Consultants gemeinsam mit dem Kunden die Rolle des Data Scientist, identifizieren die Know-how-Lücken und generieren nach Vorgabe (Zeitplan, Budget, Verfügbarkeit der Mitarbeiter) gemeinsam einen effizienten Ausbildungsplan, der verschiedenste Trainings und Trainingsformate (E-Learning, Präsenztraining, Personal Training, Coaching etc.) umfassen kann.

„Data Scientist ist aufgrund der facettenreichen Aufgabe einer der spannendsten Jobs in der IT-Branche - der durch den erfolgreichen Einsatz eines Data Scientist generierte Wertbeitrag in Unternehmen ist enorm.“

Sie haben Fragen zum Thema?

Sprechen Sie mich an: Bernadette Fabits, Curriculum Manager Analytics
bernadette.fabits@sas.com



To contact your local SAS office: www.sas.com/germany
www.sas.com/austria
www.sas.com/switzerland

Phone: +49 6221 415-300
Phone: +43 1 252 42-0
Phone: +41 44 805 74-74

