

## Data Scientist

Start	Ende	Dauer	Ort	Teilnehmerzahl	Lehrgangsinhalte
19.04.2022	02.09.2022	20 Wochen	Krefeld	in der Regel 6 - 16, bis 25	<a href="#">Seite drucken</a>

Unterrichtszeiten: Montag bis Freitag von 08:30 bis 15:35 Uhr (in Wochen mit Feiertagen von 8:30 bis 17:10 Uhr)

### Zertifizierter Lehrgang. Maßnahmennummer 361/1191/21

19.04.2022 20 Wochen

#### Data Scientist

Durch die ständige Nutzung von Social Media, mobiler Navigation und digitaler Transaktionen werden täglich riesige Datenmengen produziert: Dieser Kurs zeigt, wie ein Data Scientist mit Hilfe dieser Daten strategische Entscheidungen für Unternehmen beeinflusst, indem er Daten in Informationen umwandelt. Neben der weit verbreiteten Programmiersprache Python erläutert der Lehrgang Grundkenntnisse in der Statistik und der empirischen Arbeit mit Daten. Sie lernen, wie Datenexperten das volle Geschäftspotenzial ausschöpfen können, indem sie Data Mining und statistische Analysen durchführen. Dafür nutzen sie u. a. SQL und NoSQL. Der Lehrgang führt weiter in alle wichtigen Kenntnisse ein, die für die jeweiligen Anforderungen, Strukturierung und Bereitstellung von relevanten Daten benötigt werden. Die Aufgaben bestehen dabei in der Auswahl der richtigen Soft- und ggf. auch Hardware-Architektur und im Aufbau von geeigneten Datenmodellen, um unterschiedliche Datentöpfe zu harmonisieren. Abgerundet wird der Lehrgang mit dem Thema Machine Learning: Beim Maschinellen Lernen wird künstliches Wissen aus Erfahrung generiert und Muster in bestehenden Datenbeständen mithilfe von Algorithmen identifiziert, anschließend in eigenständige Lösungen für Probleme umgewandelt.

**Abschlussprüfung:** Praxisbezogene Projektarbeiten mit Abschlusspräsentation

**Zertifikat:** alfatraining-Zertifikat „Data Scientist“

#### Lehrgangsinhalte

##### Programmierung mit Python

###### Grundlagen Python (ca. 1 Tag)

Geschichte, Konzepte  
Verwendung und Einsatzgebiete  
Syntax

###### Erste Schritte mit Python (ca. 5 Tage)

Zahlen  
Zeichenketten  
Datum und Zeit  
Standardeingabe und -ausgabe  
list, tuple dict, set  
Verzweigungen und Schleifen (if, for, while)

###### Funktionen (ca. 5 Tage)

Eigene Funktionen definieren  
Variablen  
Parameter, Rekursion  
Funktionale Programmierung

###### Fehlerbehebung (ca. 0,5 Tage)

try, except  
Programmunterbrechungen abfangen

###### Objektorientierte Programmierung (ca. 4,5 Tage)

Python-Klassen  
Methoden  
Unveränderliche Objekte  
Datenklasse  
Vererbung

###### Grafische Benutzeroberfläche (ca. 1 Tag)

Buttons und Textfelder  
grid-Layout  
Dateiauswahl

###### Projektarbeit (ca. 3 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

###### Big Data Analytics

###### Kurzeinführung Big Data (ca. 1 Tag)

Was ist Big Data?

Akzeptieren

Ablehnen

###### Grundlagen in Python (ca. 4 Tage)

Zur Analyse und Verbesserung unserer Website setzen wir mit Ihrer Einwilligung Cookies. Die Einwilligung ist freiwillig und kann abgelehnt oder mit Wirkung für die Zukunft widerrufen werden. Weitere Informationen zu Datenverarbeitung sowie zu Widerrufsmöglichkeiten finden Sie in unserer [Datenschutzerklärung](#)

Grundlagen SQL  
Vergleich von SQL und NoSQL Datenbanken

**Datenanalyse, Statistik und Data Mining** (ca. 5 Tage)

Explorative Datenanalyse mit Python  
Daten visualisieren in Python  
Statistische Grundlagen  
Data Mining Use Cases  
Data Mining Algorithmen in Python

**Big Data: Datenverarbeiten und Speichern** (ca. 2 Tage)

Der MapReduce-Ansatz  
Grundlagen des Cloud Computings

**Projektarbeit** (ca. 3 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

**Statistik**

**Statistische Grundlagen** (ca. 6 Tage)

Messtheoretische Grundlagen (Grundgesamtheit und Stichprobe, Stichprobenarten, Messung und Skalenniveaus)  
Univariate Deskriptivstatistik (Häufigkeitsverteilungen, Zentralmaße, Streuungsmaße, Standardwerte, Histogramme, Balkendiagramme, Kreisdiagramme, Liniendiagramme und Boxplots)  
Bivariate Deskriptivstatistik (Zusammenhangsmaße, Korrelationskoeffizienten, Kreuztabellen, Streudiagramme und gruppierte Balkendiagramme)  
Grundlagen der induktiven Inferenzstatistik (Wahrscheinlichkeitsverteilung, Normalverteilung, Mittelwertverteilung, Signifikanztest, Nullhypothesentest nach Fisher, Signifikanz, Effektgröße, Parameterschätzung, Konfidenzintervalle, Fehlerbalkendiagramme, Poweranalysen und Ermittlung des optimalen Stichprobenumfangs)

**Methoden zum Vergleich von zwei Gruppen** (ca. 5 Tage)

z- und t-Test für eine Stichprobe (Abweichung von einem vorgegebenen Wert)  
t-Test für den Mittelwertsunterschied von zwei unabhängigen/ verbundenen Stichproben  
Prüfung der Wirksamkeit von Aktionen, Maßnahmen, Interventionen und anderen Veränderungen mit t-Tests (Pretest-Posttest-Designs mit zwei Gruppen)  
Unterstützende Signifikanztests (Anderson-Darling-Test, Ryan-Joiner-Test, Levene-Test, Bonnet-Test, Signifikanztest für Korrelationen)  
Nonparametrische Verfahren (Wilcoxon-Test, Vorzeichentest, Mann-Whitney-Test)  
Kontingenzanalysen (Binomialtest, Exakter Test nach Fisher, Chi-Quadrat-Test, Kreuztabellen mit Assoziationsmaße)

**Methoden zum Mittelwertvergleich von mehreren Gruppen** (ca. 5 Tage)

Ein- und zweifaktorielle Varianzanalyse (einfache und balancierte ANOVA)  
Mehrfaktorielle Varianzanalyse (Allgemeines Lineares Modell)  
Feste, zufällige, gekreuzte und geschachtelte Faktoren  
Mehrfachvergleichsverfahren (Tukey-HSD, Dunnett, Hsu-MCB, Games-Howell)  
Interaktionsanalyse (Analyse von Wechselwirkungseffekten)  
Trennschärfe und Poweranalyse bei Varianzanalysen

**Einführung in die Versuchsplanung (DoE, Design of Experiments)** (ca. 1 Tag)

Voll- und teilfaktorielle Versuchspläne

**Projektarbeit** (ca. 3 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

**Data Engineer**

**Grundlagen Business Intelligence** (ca. 3 Tage)

CRISP-DM Referenzmodell  
Umgang mit Big Data - Volume, Variety, Velocity, Validity, Value  
Abgrenzungen und Aufgaben vom Data Engineer im Kontext zu den andern BI-Berufen  
Umgang und Verarbeitung von strukturierten, semi-strukturierten und unstrukturierten Daten im DWH  
OLAP  
OLTP

**Anforderung von Daten** (ca. 2 Tage)

Aufgaben, Ziele und Vorgehensweise in der Anforderungsanalyse  
Einführung/Modellierung in der UML

- Use-Case Analyse
- Klassendiagramme
- Aktivitätsdiagramme
- Modellierung mit ERM

**Datenbanken** (ca. 2 Tage)

Grundlagen von Datenbanksystemen  
ANSI/SPARC Referenzmodell  
Normalformen  
Architektur von Datenbankmanagementsystemen  
Praktische und theoretische Einführung in SQL

- DDL
- DML
- DSDL

Akzeptieren

Ablehnen

**Data Warehouse Modellierung** (ca. 4 Tage)

Star Schema

Zur Analyse und Verbesserung unserer Website setzen wir mit Ihrer Einwilligung Cookies. Die Einwilligung ist freiwillig und kann abgelehnt oder mit Wirkung für die Zukunft widerrufen werden.  
Weitere Informationen zu Datenverarbeitung sowie zu Widerrufsmöglichkeiten finden Sie in unserer [Datenschutzerklärung](#)

## ETL (ca. 6 Tage)

### Data Cleansing

- Null Values
- Aufbereitung von Daten
- Harmonisierung von Daten
- Anwendung von Regular Expressions

### Data Understanding

- Datenvalidierung
- Statistische Datenanalyse

### Praktischer Aufbau von ETL-Strecken

### Praktischer Aufbau eines Data Vault Modells - Business und Raw Vault

### Praktische Umsetzung von Hash-Verfahren

### Verwendung von verschiedenen DatenbanksqLite, postgresql, json, csv

## Projektarbeit (ca. 3 Tage)

### Zur Vertiefung der gelernten Inhalte

### Präsentation der Projektergebnisse

## Machine Learning

### Grundlagen (ca. 5 Tage)

#### Warum Machine Learning?

#### Anwendungsbeispiele

#### Überwachtes Lernen, Unüberwachtes Lernen, Teilüberwachtes Lernen, Reinforcement Lernen

#### Bibliotheken und Werkzeuge

#### Klassifizierung von Iris-Spezies

#### Daten kennenlernen

#### Trainings- und Testdaten

#### Daten sichten

#### Vorhersagen treffen

### Überwachtes Lernen (ca. 5 Tage)

#### Klassifikation und Regression

#### Verallgemeinerung, Overfitting und Underfitting

#### Größe des Datensatzes

#### Algorithmen zum überwachtem Lernen

#### Lineare Modelle

#### Bayes-Klassifikatoren

#### Entscheidungsbäume

#### k-nächste-Nachbarn

#### Vector Machines

#### Deep Learning

#### Entscheidungsfunktion

#### Wahrscheinlichkeiten

### Unüberwachtes Lernen (ca. 5 Tage)

#### Arten unüberwachten Lernens

#### Vorverarbeiten und Skalieren

#### Datentransformationen

#### Trainings- und Testdaten skalieren

#### Dimensionsreduktion

#### Extraktion von Merkmalen

#### Manifold Learning

#### Hauptkomponentenzerlegung (PCA)

#### Nicht-negative-Matrix-Faktorisierung (NMF)

#### Manifold Learning mit t-SNE

#### Clusteranalyse

#### k-Means-Clustering

#### Agglomeratives Clustering

#### DBSCAN

#### Clusteralgorithmen

### Evaluierung und Verbesserung (ca. 2 Tage)

#### Kreuzvalidierung

#### Gittersuche

#### Evaluationsmetriken

#### Klassifikation

### Projektarbeit (ca. 3 Tage)

#### Zur Vertiefung der gelernten Inhalte

#### Präsentation der Projektergebnisse

Änderungen möglich. Die Lehrgangsinhalte werden regelmäßig aktualisiert.

Akzeptieren

### Voraussetzungen:

Vorkenntnisse mit SQL-Datenbanken werden empfohlen.

Ablehnen

### Lehrgangsziel:

Sie verstehen, wie Sie Rohdaten in eine strukturierte Datenbasis umwandeln, diese analysieren und so eine Entscheidungsgrundlage für

Zur Analyse und Verbesserung unserer Website setzen wir mit Ihrer Einwilligung Cookies. Die

Einwilligung ist freiwillig und kann abgelehnt oder mit Wirkung für die Zukunft widerrufen werden.

Weitere Informationen zu Datenverarbeitung sowie zu Widerrufsmöglichkeiten finden Sie in unserer

**Datenschutzerklärung**

Data Scientists werden überall eingesetzt, wo große Datenmengen generiert werden, um Unternehmen dabei zu unterstützen, die bestehenden Prozesse zu optimieren und Kunden noch besser anzusprechen: Die Einsatzgebiete sind somit zahlreich: Onlineshops mit Algorithmen, personalisierte Werbung, automatisierte Spracherkennungsdienste, öffentliche Verkehrsbetriebe mit Fahrplanoptimierung, Logistikprozesse wie Retouren-Verringerung, Optimierungen in der Energieversorgung, Dienste im Gesundheitswesen und vieles mehr. Das aussagekräftige Zertifikat von alfatraining gibt Ihrem zukünftigen Arbeitgeber detaillierten Einblick in Ihre erworbenen Qualifikationen.

#### Konzept & Ausstattung:

Allen Teilnehmenden steht während des Lehrgangs ein ergonomischer Arbeitsplatz mit einem leistungsfähigen PC, Flachbildschirm und Internetzugang zur Verfügung. Unsere Dozenten/innen sind sowohl fachlich als auch didaktisch hoch qualifiziert und werden Sie vom ersten bis zum letzten Tag unterrichten (kein Selbstlernsystem).

#### Präsenzlehrgänge mit Videotechnik der neuesten Generation

Der Einsatz von Videokonferenzsystemen gehört zum Lehrgangskonzept von alfatraining. Diese Form von Unterricht trainiert das Arbeiten in einer vernetzten Arbeitsumgebung, wie sie in Industrie und Wirtschaft existiert. In der globalisierten Arbeitswelt arbeiten Firmen heutzutage sowohl firmenintern als auch mit anderen Unternehmen deutschlandweit, europaweit oder weltweit über moderne Kommunikationstechniken und Netzwerke zusammen. Sie lernen im Unterricht den Umgang und Einsatz dieser modernen Techniken kennen.

#### Wie funktioniert der Unterricht bei alfatraining?

Bei alfatraining findet der Unterricht via alfaview®, einer Videokonferenzsoftware, statt. Die Dozierenden und die Teilnehmenden können sich gegenseitig sehen, hören und miteinander sprechen – live, lippensynchron und in Fernsehqualität! Über den virtuellen Klassenraum alfaview® ist es möglich, dass Sie sowohl mit Dozierenden als auch mit allen Teilnehmerinnen und Teilnehmern Ihres Kurses sprechen und im Team standortübergreifend gemeinsam an Projekten arbeiten. Zusätzliche separate Online-Besprechungsräume eignen sich zur vertraulichen Kommunikation in Kleingruppen.

#### Lernen von zuhause: Kursteilnahme im Homeoffice

Über das Videokonferenzsystem alfaview® nehmen Sie von zuhause aus am Kurs teil.  
Mehr Informationen

#### Förderungsmöglichkeiten:

Unser Lehrgangskonzept ist von der Agentur für Arbeit anerkannt und nach der Zulassungsverordnung AZAV zertifiziert, eine Förderung mit Bildungsgutschein oder Aktivierungs- und Vermittlungsgutschein ist möglich.

Eine Förderung ist auch über den Europäischen Sozialfonds (ESF), die Deutsche Rentenversicherung (DRV) oder über regionale Förderprogramme möglich.

Für Zeitsoldaten besteht die Möglichkeit, Weiterbildungen über den Berufsförderungsdienst (BFD) zu besuchen.

Auch Firmen können ihre Mitarbeiter über eine Förderung der Agentur für Arbeit (Qualifizierungschancengesetz) qualifizieren lassen.

#### Haben Sie Fragen zu dem Lehrgang?

#### alfatraining Krefeld

Bitte rufen Sie uns an,  
wir beraten Sie gerne.

**0800 3456-500**

Montag bis Freitag von 8 bis 17 Uhr



**Kontaktformular**



**Rückrufbitte**



**E-Mail an uns**

#### Weiterbildung bei alfatraining



Lernen Sie das innovative Unterrichtskonzept von alfatraining kennen. Live-Unterricht mit Videotechnik.

**Mehr Informationen**

**Akzeptieren**

**Ablehnen**

Zur Analyse und Verbesserung unserer Website setzen wir mit Ihrer Einwilligung Cookies. Die Einwilligung ist freiwillig und kann abgelehnt oder mit Wirkung für die Zukunft widerrufen werden.  
Weitere Informationen zu Datenverarbeitung sowie zu Widerrufsmöglichkeiten finden Sie in unserer **Datenschutzerklärung**

## Lernen von zuhause

Weiterbildung mit Videotechnik



[Jetzt informieren](#)



© 2022 alfatraining Bildungszentrum GmbH. Alle Rechte vorbehalten.

Akzeptieren

Ablehnen

Zur Analyse und Verbesserung unserer Website setzen wir mit Ihrer Einwilligung Cookies. Die Einwilligung ist freiwillig und kann abgelehnt oder mit Wirkung für die Zukunft widerrufen werden. Weitere Informationen zu Datenverarbeitung sowie zu Widerrufsmöglichkeiten finden Sie in unserer [Datenschutzerklärung](#)