

Kursstart alle 4 Wochen

## Big Data und KI-Spezialist:in

Big Data Specialists verarbeiten und berechnen große Datenmengen. Der Kurs führt in Apache Frameworks ein, erläutert die MapReduce-Philosophie, die Themen NoSQL und HBase sowie die Datenvisualisierung. Begleitet wird dieses Thema durch eine Einführung in das Machine und das Deep Learning sowie in die Grundlagen Künstlicher Intelligenz (KI) in diesem Bereich.



### Abschlussart

Zertifikat „Big Data Specialist“  
Zertifikat „KI-Spezialist:in“



### Abschlussprüfung

Praxisbezogene Projektarbeiten mit Abschlusspräsentationen



### Dauer

12 Wochen



### Unterrichtszeiten

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



### Nächste Kursstarts

14.10.2024  
11.11.2024  
09.12.2024

## LEHRGANGSZIEL

Nach dem Lehrgang kannst du große, unstrukturierte Datenmengen mit Hilfe von branchenspezifischer Software verarbeiten. Du verfügst über Kenntnisse im Framework Apache und weißt, wie Daten ansprechend visualisiert werden.

Zudem besitzt du relevante Kenntnisse zu den Themen Machine Learning und Deep Learning. Du kennst die wichtigsten Gründe für die Verwendung des Machine Learning, Anwendungsgebiete sowie die verschiedenen Kategorien und Konzepte des Maschinellen Lernens. Auch verstehst du die Einsatzbereiche von Deep Learning und die Funktionsweisen neuronaler Netzwerke. Du bist in der Lage, maschinelles Lernen bereitzustellen und Prozesse zu dokumentieren.

## ZIELGRUPPE

Der Lehrgang richtet sich an Personen mit abgeschlossenem Studium in der Informatik, Wirtschaftsinformatik, Mathematik oder vergleichbarer Qualifikation.

## BERUFSAUSSICHTEN

Da Unternehmen für die Auswertung und Zielsetzung Ihrer Geschäftsprozesse immer größer werdende Datenmengen bewältigen und strukturieren müssen, sind Fähigkeiten in der Verarbeitung von Daten in allen Branchen nachgefragt.

Auch ist man als KI-Spezialist:in in den Fachbereichen Machine Learning und Deep Learning hochqualifiziert, kann branchenübergreifend eingesetzt werden und ist am Arbeitsmarkt entsprechend vielfach nachgefragt. Man kann große Datenmengen nach Mustern und Modellen untersuchen. Deep Learning kommt dabei häufig im Rahmen künstlicher Intelligenz für die Gesichts-, Objekt- oder Spracherkennung zum Einsatz.

Dein aussagekräftiges Zertifikat gibt detaillierten Einblick in deine erworbenen Qualifikationen und verbessert deine beruflichen Chancen.

## VORAUSSETZUNGEN

Programmierkenntnisse mit Python und Erfahrungen mit Datenbanken (SQL) werden vorausgesetzt, Vorkenntnisse im Bereich Data Analytics werden empfohlen.

## LEHRGANGSINHALTE

### BIG DATA SPECIALIST

#### Was ist Big Data? (ca. 1 Tag)

Volume, Velocity, Variety, Value, Veracity  
Chancen und Risiken großer Datenmengen  
Abgrenzung: Business Intelligence, Data Analytics, Data Science  
Was ist Data Mining?

#### Einführung in Apache Frameworks (ca. 2 Tage)

Big-Data-Lösungen in der Cloud  
Datenzugriffsmuster  
Datenspeicherung

#### MapReduce (ca. 3 Tage)

MapReduce Philosophie  
Hadoop Cluster  
Verketten von MapReduce Jobs

#### Künstliche Intelligenz (KI) im Arbeitsprozess

Vorstellung von konkreten KI-Technologien im beruflichen Umfeld  
Anwendungsmöglichkeiten und Praxis-Übungen

### Komponenten (ca. 3 Tage)

Kurzvorstellung von verschiedenen Tools  
Datenübertragung  
YARN-Anwendungen  
Hadoop JAVA-API  
Apache Spark

### NoSQL und HBase (ca. 3 Tage)

CAP-Theorem  
ACID und BASE  
Typen von Datenbanken  
HBase

### Big Data Visualisierung (ca. 3 Tage)

Theorien der Visualisierung  
Diagrammauswahl  
Neue Diagrammarten  
Werkzeuge zur Datenvisualisierung

### Projektarbeit (ca. 5 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

---

## MACHINE LEARNING

### Einführung in Machine Learning (ca. 5 Tage)

Warum Machine Learning?  
Anwendungsbeispiele  
Überwachtes Lernen, Unüberwachtes Lernen, Teilüberwachtes Lernen,  
Reinforcement Lernen  
Beispiele für Datenbestände  
Daten kennenlernen  
Trainings-, Validierungs- und Testdaten  
Daten sichten  
Vorhersagen treffen

### Überwachtes Lernen (ca. 5 Tage)

Klassifikation und Regression  
Verallgemeinerung, Overfitting und Underfitting  
Größe des Datensatzes  
Algorithmen zum überwachten Lernen  
Lineare Modelle  
Bayes-Klassifikatoren  
Entscheidungsbäume  
Random Forest  
Gradient Boosting  
k-nächste-Nachbarn  
Support Vector Machines  
Conditional Random Field  
Neuronale Netze und Deep Learning  
Wahrscheinlichkeiten

### Unüberwachtes Lernen (ca. 5 Tage)

Arten unüberwachten Lernens  
Vorverarbeiten und Skalieren  
Datentransformationen  
Trainings- und Testdaten skalieren  
Dimensionsreduktion  
Feature Engineering  
Manifold Learning  
Hauptkomponentenzerlegung (PCA)  
Nicht-negative-Matrix-Faktorisierung (NMF)  
Manifold Learning mit t-SNE  
Clusteranalyse  
k-Means-Clustering  
Agglomeratives Clustering  
Hierarchische Clusteranalyse  
DBSCAN  
Clusteralgorithmen

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

Modellauswahl und Modellevaluation  
Abstimmung der Hyperparameter eines Schätzers  
Kreuzvalidierung  
Gittersuche  
Evaluationsmetriken  
Klassifikation

### Projektarbeit (ca. 3 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

---

## DEEP LEARNING

### Einführung Deep Learning (ca. 1 Tag)

Deep Learning als eine Art von Machine Learning

### Grundlagen in neuronalen Netzen (ca. 4 Tage)

Perceptron  
Berechnung neuronaler Netze  
Optimierung der Modellparameter, Backpropagation  
Deep-Learning-Bibliotheken  
Regression vs. Klassifikation  
Lernkurven, Überanpassung und Regularisierung  
Hyperparameteroptimierung  
Stochastischer Gradientenabstieg (SGD)  
Momentum, Adam Optimizer  
Lernrate

### Convolutional Neural Network (CNN) (ca. 2 Tage)

Bildklassifizierung  
Convolutional-Schichten, Pooling-Schichten  
Reshaping-Schichten, Flatten, Global-Average-Pooling  
CNN-Architekturen ImageNet-Competition  
Tiefe neuronale Netze, Vanishing Gradients, Skip-Verbindungen, Batch-Normalization

### Transfer Learning (ca. 1 Tag)

Anpassen von Modellen  
Unüberwachtes Vortrainieren  
Image-Data-Augmentation, Explainable AI

### Regional CNN (ca. 1 Tag)

Objektlokalisierung  
Regressionsprobleme  
Verzweigte neuronale Netze

### Methoden der kreativen Bilderzeugung (ca. 1 Tag)

Generative Adversarial Networks (GAN)  
Deepfakes  
Diffusionsmodelle

### Recurrente neuronale Netze (ca. 2 Tage)

Sequenzanalyse  
Rekurrente Schichten  
Backpropagation through time (BPTT)  
Analyse von Zeitreihen  
Exploding und Vanishing Gradient Probleme  
LSTM (Long Short-Term Memory)  
GRU (Gated Recurrent Unit)  
Deep RNN  
Deep LSTM

### Textverarbeitung durch neuronale Netze (ca. 2 Tage)

Text-Preprocessing  
Embedding-Schichten  
Text-Klassifizierung  
Sentimentanalyse  
Transfer-Learning in NLP  
Übersetzungen  
Sequence-to-Sequence-Verfahren, Encoder-Decoder-Architektur

### Sprachmodelle (ca. 1 Tag)

BERT, GPT  
Attention-Schichten, Transformers  
Textgeneration-Pipelines  
Summarization  
Chatbots

### Deep Reinforcement Learning (ca. 1 Tag)

Steuerung dynamischer Systeme  
Agentensysteme  
Training durch Belohnungen  
Policy Gradients  
Deep-Q-Learning

### Bayes'sche neuronale Netze (ca. 1 Tag)

Unsicherheiten in neuronalen Netzen  
Statistische Bewertung von Prognosen  
Konfidenz, Standardabweichung  
Unbalancierte Daten  
Sampling-Methoden

### Projektarbeit (ca. 3 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

## UNTERRICHTSKONZEPT

### Didaktisches Konzept

Deine Dozierenden sind sowohl fachlich als auch didaktisch hoch qualifiziert und werden dich vom ersten bis zum letzten Tag unterrichten (kein Selbstlernsystem).

Du lernst in effektiven Kleingruppen. Die Kurse bestehen in der Regel aus 6 bis 25 Teilnehmenden. Der allgemeine Unterricht wird in allen Kursmodulen durch zahlreiche praxisbezogene Übungen ergänzt. Die Übungsphase ist ein wichtiger Bestandteil des Unterrichts, denn in dieser Zeit verarbeitest du das neu Erlernte und erlangst Sicherheit und Routine in der Anwendung. Im letzten Abschnitt des Lehrgangs findet eine Projektarbeit, eine Fallstudie oder eine Abschlussprüfung statt.

### Virtueller Klassenraum alfaview®

Der Unterricht findet über die moderne Videotechnik alfaview® statt - entweder bequem von zu Hause oder bei uns im Bildungszentrum. Über alfaview® kann sich der gesamte Kurs face-to-face sehen, in lippensynchroner Sprachqualität miteinander kommunizieren und an gemeinsamen Projekten arbeiten. Du kannst selbstverständlich auch deine zugeschalteten Trainer:innen jederzeit live sehen, mit diesen sprechen und du wirst während der gesamten Kursdauer von deinen Dozierenden in Echtzeit unterrichtet. Der Unterricht ist kein E-Learning, sondern echter Live-Präsenzunterricht über Videotechnik.

## FÖRDERMÖGLICHKEITEN

Alle Lehrgänge werden von der Agentur für Arbeit gefördert und sind nach der Zulassungsverordnung AZAV zertifiziert. Bei der Einreichung eines Bildungsgutscheines oder eines Aktivierungs- und Vermittlungsgutscheines werden in der Regel die gesamten Lehrgangskosten von Ihrer Förderstelle übernommen.

Eine Förderung ist auch über den Europäischen Sozialfonds (ESF), die Deutsche Rentenversicherung (DRV) oder über regionale Förderprogramme möglich. Als Zeitsoldat:in besteht die Möglichkeit, Weiterbildungen über den Berufsförderungsdienst (BFD) zu besuchen. Auch Firmen können ihre Mitarbeiter:innen über eine Förderung der Agentur für Arbeit (Qualifizierungschancengesetz) qualifizieren lassen.

- ① Änderungen möglich. Die Lehrgangsinhalte werden regelmäßig aktualisiert. Die aktuellen Lehrgangsinhalte findest Du immer unter [smartbuilding.alfatraining.de](https://smartbuilding.alfatraining.de).