

Career Track • Curriculum

Data Scientist – Fokus Python

Qualifizierung für die Jobrolle als **Data Scientist**





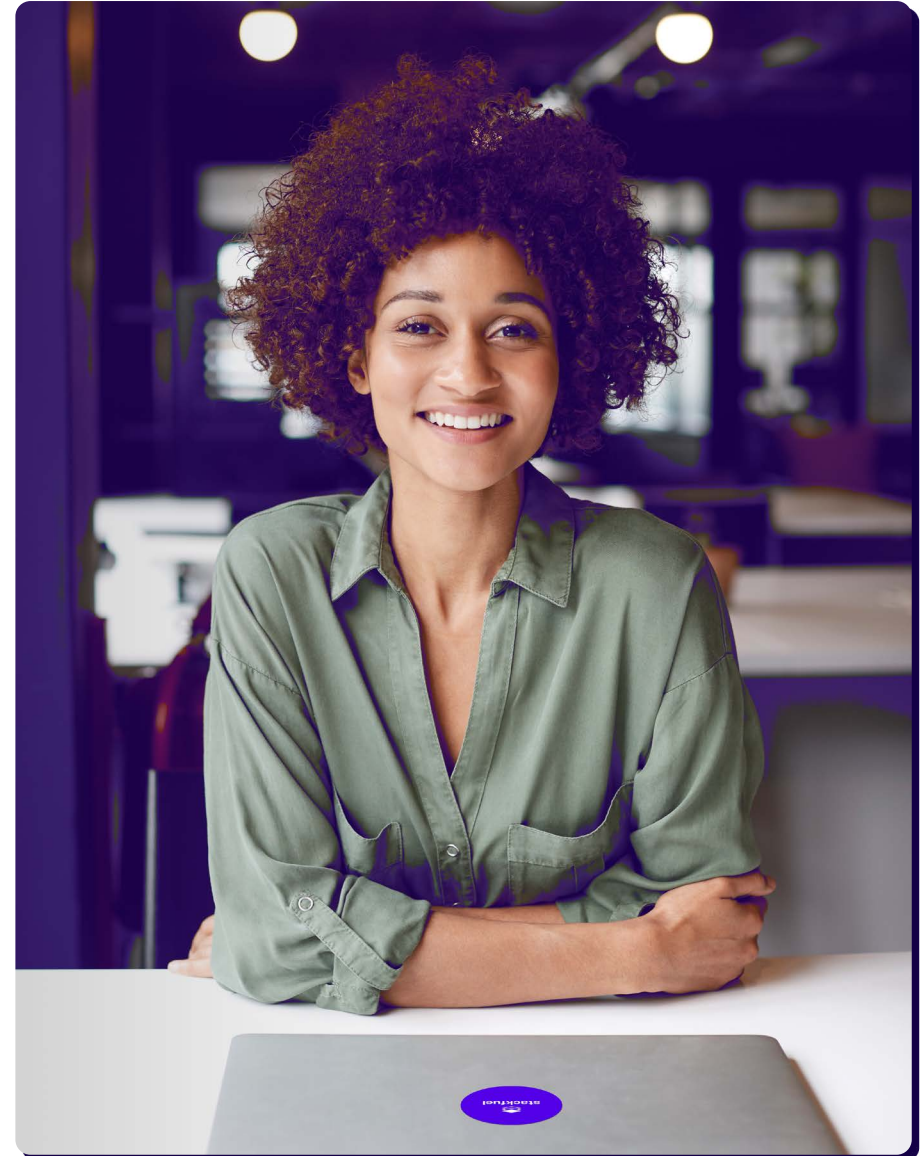
Warum dieser Kurs?

Data Science ist ein Berufs- und Themenfeld mit unglaublichem Potenzial, das auch in den nächsten Jahren weiter stark wachsen wird. Die Nachfrage nach qualifizierten Data Scientists ist sehr hoch, da in diesem Bereich großer Fachkräftemangel herrscht. Daher gilt die Jobrolle als Data Scientist branchenübergreifend als eine der erstrebenswertesten und nachgefragtesten Top-Karrieren.

Nicht zuletzt werden Data Scientist Weiterbildungen deshalb immer relevanter und nachgefragter. Die Fähigkeiten von Data Scientists sind im Zeitalter der Digitalisierung so wichtig wie begehrt: Fachkompetenzen in der Datenanalyse, Programmierkenntnisse in Python sowie die Anwendung von maschinellem Lernen und Deep Learning.

Die zertifizierte Online-Weiterbildung zum Data Scientist – Fokus Python befähigt Dich dazu, Prognosemodelle aus Daten abzuleiten, zu verifizieren und zu interpretieren, um die Modellergebnisse effizient zu kommunizieren. Der zusätzliche Aufbau von Fähigkeiten im Bereich des Machine Learnings qualifiziert Dich mit erfolgreichem Abschluss des Karrierepfades für die Jobrolle als Data Scientist oder eine andere analytische Jobrolle wie dem Business Intelligence Analyst oder Financial Analyst.

Mehr als **96.000** IT-Fachkräfte werden derzeit in Deutschland gesucht. Data Scientists bzw. Big Data Experts zählen dabei zu den meistgefragtesten Qualifikationen.



Lerninhalte.

Ziel des praxisorientierten Trainings ist es, Datenmodelle mit Python zur Vorhersage von unterschiedlichen Businessszenarien umzusetzen. Du lernst, überwachte und unüberwachte Machine-Learning Algorithmen zu implementieren und zu optimieren.

Ebenso baust Du Kompetenzen für gängige Visualisierungsmethoden auf, um Dich mit Abschluss des Data Scientist Trainings für die Jobrolle als Data Scientist zu qualifizieren.

Im Überblick:

- > Daten selbständig einlesen, bereinigen und filtern
- > Daten explorativ mithilfe von deskriptiver Statistik analysieren
- > Komplexe Prognosemodelle entwickeln und verifizieren
- > Python-Programmierkenntnisse vertiefen
- > Datenmodelle zur Vorhersage von Businessszenarien bauen
- > Machine-Learning-Algorithmen entwickeln
- > Best Practices für die effektive Datenvisualisierung
- > Methoden des Data Storytellings

Kursüberblick.

Dauer

108 Stunden (4,5 Monate)

Typ

Online-Training

Aufbau

4 Module + 1 Abschlussprojekt

Sprache

Deutsch, Englisch

Niveau

Fortgeschrittene

Abschluss

Abschlusszertifikat

Voraussetzungen

Für das Training werden Programmierkenntnisse in Python sowie Vorkenntnisse in den Python-Bibliotheken Pandas und Matplotlib vorausgesetzt.

Zielgruppe

Die Weiterbildung Data Scientist - Fokus Python eignet sich für alle, die Python als Programmiersprache einsetzen, Daten analysieren und auf Grundlage dieser Vorhersagen erstellen möchten, um datengetriebene Entscheidungen zu treffen.

Du solltest eine grundlegende Motivation für Statistik, logisches Denken und maschinelles Lernen mitbringen. Auch für Quereinsteiger:innen ist die Data Scientist Weiterbildung geeignet.



Modulübersicht.

Modul 0

Preparation

Ziel:

- > Auffrischung der Kenntnisse im Umgang mit Python sowie mathematischer Grundlagen Beschreibung

Beschreibung:

- > Teilnehmende führen Analysen und Datenmanipulationen in Python aus und nutzen dabei die Pakete Pandas und Matplotlib

Kapitel 1: Data Analytics with Python

Teilnehmende machen sich mit unserer interaktiven Programmierumgebung – dem Data Lab – vertraut und frischen die wichtigsten Programmier- und Python-Grundlagen zur Datenverarbeitung mit **Pandas**, Datenvisualisierung mit **Matplotlib** und **Seaborn** und Datenbankabfrage mit **SQLAlchemy** auf.

Kapitel 2: Linear Algebra

Teilnehmende machen sich mit dem mathematischen Hintergrund von **Data-Science-Algorithmen** vertraut und lernen die Grundbegriffe der **linearen Algebra** kennen. Unter Verwendung des Pakets **Numpy** rechnen die Teilnehmenden mit **Vektoren** und **Matrizen**.

Kapitel 3: Probability Distributions

Teilnehmende lernen mehr über den statistischen Hintergrund von **Data-Science-Algorithmen**. Sie beschäftigen sich mit wichtigen statistischen Konzepten und lernen **diskrete** und **kontinuierliche Verteilungen** kennen. Darüber hinaus erhalten Teilnehmende einen Einblick in die Versionierung von Code mit **Git**.



Modulübersicht.

Modul 1

Machine Learning Basics

Ziel:

- > Lösen von überwachten und unüberwachten Machine-Learning-Problemen mit sklearn

Beschreibung:

- > Teilnehmende erstellen Data-Science-Workflows mit sklearn, evaluieren ihre Modell-Performance anhand von geeigneten Metriken und werden für das Problem des Overfittings sensibilisiert

Kapitel 1: Supervised Learning (Regression)

Anhand der linearen Regression erlernen Teilnehmende den Umgang mit dem Python-Paket sklearn. Weiterhin beschäftigen sie sich mit den Annahmen des Regressionsmodells und der Evaluation der erzeugten Prognosen. In diesem Zuge werden auch der Bias-Variance Trade-Off, Konzepte der Regularisierung sowie verschiedene Maße der Modellgüte verdeutlicht.

Kapitel 2: Supervised Learning (Classification)

Teilnehmende werden in Klassifizierungsalgorithmen anhand des **k-Nearest-Neighbors-Algorithmus** eingeführt und lernen, den Algorithmus zu evaluieren und die Klassifizierungsperformance einzuschätzen. Sie optimieren die Parameter ihres Modells unter Beachtung der Aufteilung der Daten in Trainings- und Evaluationssets.

Kapitel 3: Unsupervised Learning (Clustering)

Teilnehmende lernen den **k-Means-Algorithmus** als Beispiel eines Algorithmus des unüberwachten Lernens kennen. Die Annahmen und Performance-Metriken des Algorithmus werden kritisch beleuchtet und ein kurzer Ausblick auf eine Alternative zum **k-Means-Clustering** geworfen.





Modulübersicht.

Modul 1

Machine Learning Basics

Kapitel 4: Unsupervised Learning (Dimensionality Reduction)

Teilnehmende lernen, wie sie mithilfe einer **Principal Component Analysis (PCA)** die Dimension der Daten verringern können und nutzen die PCA, um unkorrelierte Features aus den Ursprungsdaten zu erzeugen. In diesem Zusammenhang wird das Thema Feature Engineering näher betrachtet und aus den alten Features neue erzeugt.

Kapitel 5: Outlier Detection

Teilnehmende lernen verschiedene Ansätze kennen, um **Ausreißer** zu identifizieren und verstehen, mit diesen ungewöhnlichen Datenpunkten umzugehen. Sie nutzen robuste **Maße** und **Modelle**, um den Einfluss der Ausreißer zu minimieren.

Modul 2

Deep Dive Supervised Learning

Ziel:

- > Erweiterung des eigenen Data-Science-Toolkits

Beschreibung:

- > Teilnehmende intensivieren ihre Kenntnisse über Modelle zur Klassifikation von Daten. Dabei erweitern sie ihre Fähigkeiten im Sammeln und Aufbereiten von Daten

Kapitel 1: Data Gathering

Teilnehmende lernen, Daten zu sammeln, indem sie Webseiten und PDF-Dokumente auslesen. Mithilfe von **Regular Expressions** strukturieren sie gesammelte Textdaten so, dass sie diese zusammen mit bekannten Algorithmen verwenden können.

Kapitel 2: Logistic Regression

Teilnehmende lernen einen zweiten Klassifizierungsalgorithmus kennen: die **logistische Regression**. Sie nutzen neue Performance-Metriken zur Evaluation der Ergebnisse und erfahren, wie sie nicht-numerische Daten für ihre Modelle nutzbar machen.



Modulübersicht.

Modul 2

Deep Dive Supervised Learning

Kapitel 3: Decision Trees and Random Forests

Teilnehmende lernen den **Entscheidungsbaum** als leicht zu interpretierendes Modell kennen. Sie kombinieren mehrere Modelle zu einem Ensemble, um die Vorhersagen ihres Modells zu verbessern. Weiterhin erhalten sie Methoden zu unausgeglichene Kategorien an die Hand.

Kapitel 4: Support Vector Machines

Teilnehmende lernen einen letzten Klassifizierungsalgorithmus kennen – **Support Vector Machines (SVM)** und beleuchten das Verhalten verschiedener Kernel für die SVM. Außerdem erlernen sie die typischen Schritte des **Natural Language Processing (NLP)** und bearbeiten ein NLP-Szenario unter Verwendung von Bag-of-Words-Modellen.

Kapitel 5: Neural Networks

Teilnehmende werden in **künstliche neuronale Netze** eingeführt und lernen mehr über **Deep Learning**, um ein künstliches neuronales Netzwerk mit mehreren Schichten zu erzeugen und auf ein Datenset anzuwenden.

Modul 3

Advanced Topics in Data Science

Ziel:

- Selbstständiges Anwenden einfacher und komplexer Modellierungen

Beschreibung:

- Teilnehmende erlangen Souveränität im Lösen von Data-Science-Problemen und lernen, Ergebnisse kompetent zu kommunizieren

Kapitel 1: Visualization and Model Interpretation

Teilnehmende erlernen wichtige Methoden zur Interpretation und Visualisierung von **Machine-Learning-Modellen**. Durch die Verwendung **modelagnostischer Methoden** zur Interpretation lernen sie Erkenntnisse zur Funktionsweise ihrer Modelle abzuleiten und zu kommunizieren.

Kapitel 2: Spark

Teilnehmende erfahren, weshalb die Arbeit mit verteilten Speichersystemen relevant ist. Mit dem Python-Paket **PySpark** erlernen sie verteilte Datenbanken auszulesen, Big-Data-Analysen durchzuführen und bekannte Machine-Learning-Algorithmen auf verteilten Systemen zu nutzen.



Modulübersicht.

Modul 3

Advanced Topics in Data Science

Kapitel 3: Exercise Project

Teilnehmende bearbeiten ein **Prädiktionsproblem** mit Hilfe eines größeren Datensets und setzen ihre Data-Science-Fähigkeiten von der Reinigung des Datensets bis zur Interpretation des Modells eigenständig ein. In einer Projektbesprechung mit dem Mentorenteam von StackFuel erhalten Teilnehmende Feedback zu ihrem Lösungsansatz.

Kapitel 4: Final Project

Teilnehmende erhalten ein weiteres größeres Datenset, das sie selbstständig analysieren und im Vergleich zum Übungsprojekt mit weniger Hilfestellungen lösen müssen. In einer individuellen Projektbesprechung mit dem Mentoring Team von StackFuel erhalten Teilnehmende Feedback zu ihrem Lösungsansatz.

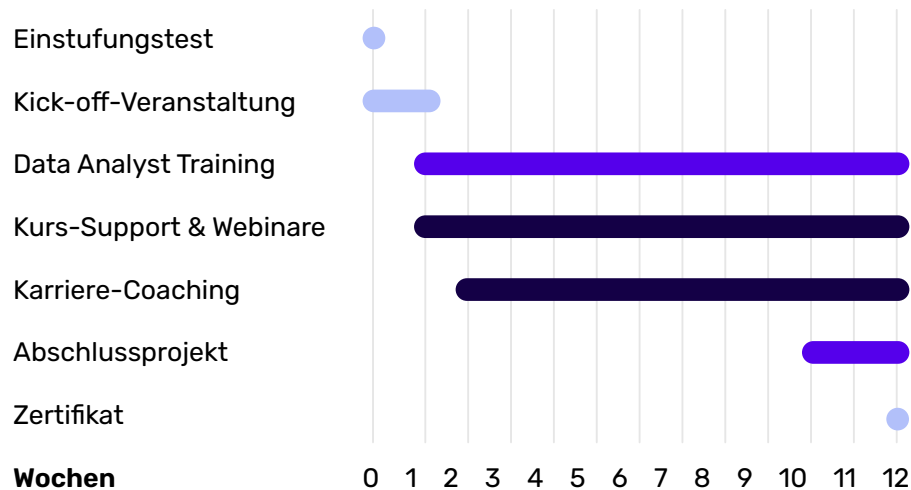


Über StackFuel.

Wir sind Dein strategischer Lernpartner, passend für jede Karrierestufe und berufliche Ausrichtung.

Ob Quereinsteiger:innen, Mitarbeitende oder Führungskräfte – mit unseren zertifizierten und staatlich geförderten Weiterbildungen im Bereich Daten, KI und Programmieren bleibst Du immer auf dem neusten Stand mit den aktuellen Technologien.

Trainingsablauf.



*Exemplarischer Trainingsablauf für das Data Analyst Training.

 Vor- & Nachbereitung  Trainingsinhalte  Mentoring & Support

Trainingsphilosophie.



Online & flexibel

Mache Deine Weiterbildung in Teilzeit oder Vollzeit und lerne 100 % online im Browser über den Laptop oder PC.



Praxisnah mit eigenen Projekten

Im Data Lab schreibst Du in Praxisprojekten und Coding Challenges selbst Algorithmen mit Industriedatensets.



Mentoring & Karriere-Coaching

Dein persönliches Mentoring Team begleitet Dich mit Beratung, Karriere-Coaching und in wöchentlichen Webinaren.



Zertifiziert & förderfähig

Als zertifizierter Träger kannst Du unsere staatlich geprüften Kurse über diverse Förderprogramme finanzieren lassen.

Fakten.

91%

Abschluss-
quote

80%

interaktive
Praxisinhalte

20%

Expertenvideos &
Textlektionen

+150.000

Lernstunden im Data Lab

4,3 von 5

Weiterempfehlungsrate



Fördermöglichkeiten.

Raten- oder Teilzahlung

Nutze unsere Raten- oder Teilzahlungsoptionen, um die Kosten für Deine Weiterbildung auf mehrere Monate zu verteilen und Deine finanzielle Flexibilität zu erhalten.

Bildungsgutschein

Mit einem **Bildungsgutschein** kannst Du Dir Deine Weiterbildung zu 100% durch das Jobcenter oder die Agentur für Arbeit finanzieren lassen, wenn Du aktuell arbeitslos oder arbeitssuchend registriert bist.

Qualifizierungschancengesetz

Wenn Du berufstätig bist, kannst Du Deine Weiterbildung vom Arbeitgeber teilweise oder ganz über das **Qualifizierungschancengesetz** finanzieren lassen – unabhängig von Qualifikation, Alter und Betriebsgröße.

StackFuel Stipendien

Wir vergeben regelmäßig diverse **Stipendien** für unsere Trainings, um für mehr Diversität in der Datenbranche zu sorgen und immer mehr Menschen für das Programmieren zu begeistern und Frauen für datengetriebene Jobs zu begeistern.



Support & Mentoring.

Coaching

- > Einstufungstest
- > Abschlussprojekt und -evaluation
- > 1:1 Projekt-Feedbackgespräche
- > Offizielles Abschlusszertifikat

Persönliche Betreuung

- > Kick-off-Veranstaltung
- > Webinare mit anderen Kursteilnehmenden
- > Support via E-Mail oder Telefon
- > Online-Forum mit Lernenden

Karriere-Services

- > Lebenslauf- und Bewerbungstraining
- > Talent-Pool und Karriere-Intros
- > Data Community
- > Karriere-Events



Neugierig geworden?

Melde Dich bei unseren **Berater:innen** und
baue die **Skills** auf, die Du brauchst!

Für Privatpersonen



Maria Schwenke

Weiterbildungsberaterin

Kostenlose Beratung buchen

Für Unternehmen



Ginesh Koottakara

Head of Sales

Kostenlosen Demotermin buchen

August 2023

Version: 23/08/21

Herausgeber: StackFuel GmbH

Web: www.stackfuel.com

E-Mail: info@stackfuel.com

Telefon: +49 (0)30 544 533 420

Irrtümer und Änderungen sind vorbehalten.

© 2023 StackFuel GmbH