


# Data Scientist Basis-Training mit Python und Spark MLlib

Dreitägiges Intensivseminar



Mehr als 25 Jahre Kompetenz und Erfahrung –  
lernen Sie uns kennen.  
Herzlich willkommen bei saracus!

Dieses dreitägige Grundlagen-Seminar gibt angehenden Data Scientists und Data Analysts neben der Modellbildung einen Einblick in die Bereiche Data Exploration, Feature Preparation und Visualisierung. Während des Seminars werden weit verbreitete Bibliotheken und Frameworks wie Pandas, Matplotlib, scikit-learn und Apache Spark vorgestellt und in einer Vielzahl von Übungen eingesetzt. Das Seminar vermittelt Ihnen sowohl die theoretischen Grundlagen im Bereich Advanced Analytics als auch deren Anwendung.

Die Referenten sind erfahrene Data Scientists mit didaktischen Fähigkeiten.

Alle Übungen werden auf einem Multi-Node-Hadoop-Cluster ausgeführt, welches in der Amazon EC2 Cloud installiert ist.

## An wen richtet sich das Seminar?

Das Seminar richtet sich an (angehende) Data Scientists und Data Analysts, die sich die wichtigsten Basiskenntnisse erarbeiten wollen, um valide Modelle mit Hilfe unterschiedlicher Verfahren des Machine Learnings entwickeln zu können.

## Voraussetzungen

Notwendig ist ein Basiswissen von Python und grundlegendes mathematisches und statistisches Verständnis.

## Seminar-Optionen

Der dritte Seminartag ist optional. Somit können Sie entweder die ersten beiden Seminartage buchen oder alle drei Seminartage.

Jeweils von 09:00 bis 17:00 Uhr

## Einführung in Data Mining (Tag 1)

- Vorstellung verschiedener Machine Learning Use Cases
- Allgemeines Vorgehen beim Data Mining
- Einführung in die Python-Bibliotheken Numpy und Pandas
- *Übung*: Praktischer Einsatz von Numpy und Pandas
- Techniken zur Data Exploration und Datenvisualisierung
- Einführung in die Python-Bibliotheken Matplotlib und Seaborn
- *Übung*: Datenvisualisierung mit Matplotlib und Seaborn
- Feature Engineering und Feature Preparation
- *Übung*: Datenaufbereitung und Feature Engineering mit Pandas

## Machine-Learning-Methoden (Tag 2)

- Detaillierte Besprechung wichtiger Machine Learning Algorithmen (Lineare und logistische Regression, Entscheidungsbäume, Random Forest, k-Nearest-Neighbor-Algorithmus, k-Means-Algorithmus)
- Einführung in die Python-Bibliothek scikit-learn
- *Übung*: Entwicklung einer Machine-Learning-Pipeline für Regression und binäre Klassifikation
- Vorstellung von Metriken zur Bewertung eines Machine-Learning-Modells
- *Übung*: Evaluierung der zuvor selbst erstellten Modelle

## Big Data Analysen und komplexe ML-Methoden (Tag 3, optional)

- Einführung in Spark und Spark SQL
- *Übung*: Data Preparation mit Spark SQL
- Einführung in MLlib
- *Übung*: Clustering mit Spark MLlib
- Advanced Machine Learning Topics:  
Curse of Dimensionality, Feature Selection, Imbalanced Datasets, etc.

Kaffee- und Teepausen finden nach Bedarf statt!

# Anmeldung zum Seminar

- Ja, ich nehme an dem Seminar  
„Data Scientist Basis-Training mit Python und Spark MLlib“ teil.**

Firma: \_\_\_\_\_

Name: \_\_\_\_\_ Vorname: \_\_\_\_\_

Funktion: \_\_\_\_\_

Straße: \_\_\_\_\_

PLZ/Ort: \_\_\_\_\_

Tel.-Nr.: \_\_\_\_\_ Fax-Nr.: \_\_\_\_\_

Email: \_\_\_\_\_

Datum: \_\_\_\_\_ Unterschrift: \_\_\_\_\_

**Die Seminargebühr beträgt pro Teilnehmer:**

- Bei Buchung von zwei Tagen: 1.750 Euro  
 Bei Buchung von drei Tagen: 2.610 Euro

**Gebuchter Seminartermin (bitte angeben):** \_\_\_\_\_

Ihre Anmeldung sollte spätestens 14 Tage vor dem Seminartermin erfolgen.

**Gebuchter Seminarort (bitte kreuzen Sie an):**

- Baden-Dättwil(CH)  Münster (D)  
 Ich benötige ein Hotel!

Nach Ablauf der Anmeldefrist erhalten Sie von uns eine Anmeldebestätigung und eine Rechnung sowie eine Anfahrtsskizze. Bitte überweisen Sie den Rechnungsbetrag vor Veranstaltungsbeginn.

**Bitte senden Sie diese Seite per Fax an +49 (0251) 98721-26  
oder per Post an: saracus consulting GmbH  
„Data Scientist Basis-Training mit Python und Spark MLlib“  
Hafenweg 46, D- 48155 Münster**

**Voraussetzung:**

Berücksichtigen Sie bitte, dass eine Mindestteilnehmerzahl von 4 Personen vorgesehen ist. Ansonsten kann das Seminar nicht stattfinden.

**Anmeldung:**

Bitte nehmen Sie Ihre Abmeldung schriftlich vor. Selbstverständlich ist eine Vertretung des angemeldeten Teilnehmers möglich. Alle genannten Preise verstehen sich zzgl. der gesetzl. MwSt.

**Stornierung:**

Bei Stornierung bis zu 14 Tage vor Veranstaltungsbeginn erheben wir eine Bearbeitungsgebühr von 50% der Gebühr. Bei späteren Absagen wird der gesamte Beitrag fällig.

**Seminarabsage:**

Sollten wir aus wichtigen Grund (z. B. Erkrankung des/der Referenten) gezwungen sein, den Kurs abzusagen, so teilen wir Ihnen dieses umgehend mit. Wir werden Ihnen in diesem Fall einen Ersatztermin anbieten. Passt dieser nicht zu Ihrer Terminplanung, erhalten Sie die bereits gezahlte Seminargebühr in voller Höhe zurück. Darüber hinausgehende Ansprüche bestehen nicht.



**Für Rückfragen steht Ihnen gerne das Team der Data & Analytics academy der saracus consulting unter +49 (0251) 98721-0 zur Verfügung.  
Email: [saracus@saracus.com](mailto:saracus@saracus.com)  
Internet: <https://www.saracus.com/know-how/data-and-analytics-academy/>**