

# Machine Learning Kurs mit Python







Training für Anfänger in scikit-learn



# Machine Learning Kurs mit Python

Training für Anfänger in scikit-learn

## Über den Kurs

-  **Dauer:** 2 Tage
-  **Gruppengröße:** 3-10
-  **Level:** Fortgeschritten mit Python-Erfahrung\*
-  **Anteil Coding:** 60%
-  **Sprache:** Python
-  **Bibliotheken:** scikit-learn, statsmodels, pandas

## Auf einen Blick

- ✓ Einführung in Machine Learning
- ✓ Supervised vs. Unsupervised Learning
- ✓ Umsetzung der Algorithmen in scikit-learn
- ✓ Interpretation der Ergebnisse
- ✓ Lineare Regression, logistische Regression
- ✓ Entscheidungsbaum, Ensemble-Methoden
- ✓ K-Nearest Neighbor, Einfaches Neuronales Netz
- ✓ K-Means Clustering und DBScan
- ✓ Hyperparametertuning, Kreuzvalidierung

## Überblick über das Seminar

Eine wesentliche Algorithmen Klasse der Künstlichen Intelligenz ist Maschinelles Lernen (ML). Machine Learning gliedert sich grob in zwei Bereiche: supervised und unsupervised Learning (überwachtes und unüberwachtes Lernen).

Supervised Learning umfasst all die Algorithmen, die durch (meist vom Menschen) gelabelte Daten trainiert werden, um gewisse "Aufgaben" zu erlernen. Bei gelabelten Daten gibt man dem Algorithmus die Zielvariable mit, die dieser erlernen soll. Beispiele hierfür sind die Klassifikation, welchen Fehler ein Bauteil aufweist oder die Vorhersage, wie viel Umsatz ein Kunde voraussichtlich generiert.

Unsupervised Algorithmen dagegen benötigen keine gelabelte Daten, sondern versuchen in den vorhandenen Daten beispielsweise Gruppen oder Muster zu erkennen, z.B. zur Kundengruppierung (verschiedene Kunden können dann anders im Marketing behandelt werden). Unsupervised Learning liefert Hinweise, welche Muster in Daten vorhanden sein könnten – es werden jedoch keine vorgegebenen Muster (wie Bauteil defekt – nicht defekt) erlernt.

Die im Seminar verwendete Programmiersprache Python ist sehr verbreitet im Maschinellen Lernen (und auch im Deep Learning). Das Modul scikit-learn umfasst zahlreiche Machine Learning Algorithmen.

## Inhalte des Seminars

Die Algorithmen werden vornehmlich mit dem Python Modul scikit-learn umgesetzt. Das Modul StatsModels wird auch für die logistische Regression verwendet, um einen Einblick in eine weitere bekannte Bibliothek für Machine Learning zu erhalten.

Der Fokus liegt auf supervised Algorithmen. Hierbei werden mehrere Maschinelle Lernen Algorithmen theoretisch erläutert und praktisch in Python umgesetzt: Lineare Regression, logistische Regression, Entscheidungsbaum, Ensemble-Methoden (Random-Forest und AdaBoost), K-Nearest Neighbor und Neuronale Netzwerke (Multi-Layer Perceptron, MLP).

\* Programmierkenntnisse in Python im Umfang des dreitägigen „Data Science in Python“ Kurses sind empfohlen. Siehe Abschnitt „Voraussetzungen“.

Das Splitting der Daten in Trainings und Test Daten und Cross Validation (Kreuzvalidierung) werden erläutert (beides Konzepte zum Erkennen von Overfitting). Zudem wird die Hyperparametersuche und die wichtigsten Metriken zur Bestimmung der Algorithmusgüte erläutert.

Im unsupervised Learning wird das Clustering mit dem K-means Algorithmus und DBSCAN behandelt.

Nach dem Seminar haben Sie die Struktur von Modul scikit-learn verstanden und können selbständig erste Machine Learning Aufgaben bearbeiten. Zudem haben Sie die Fähigkeit, Ihre Programmierkenntnisse in Python selbständig zu erweitern, um weitere Machine Learning Algorithmen, welche in scikit-learn vorhanden sind, umzusetzen.

### Wer sollte teilnehmen?

Dieser Kurs mit Python richtet sich an data scientists, angehende Machine Learning engineers, Datenanalysten, Business Intelligence Analysts, Data Analysts, o.ä. welche die Programmiersprache Python für Data Science/Data Mining vertiefen möchten. Vorhandene Programmierkenntnisse in Python werden erweitert, um Machine Learning mit scikit-learn eigenständig umsetzen zu können.

### Voraussetzungen

Dieser Python Kurs setzt Programmiererfahrung in Python voraus. Notwendig ist Vorerfahrung in der Arbeit mit Python und mit Daten. Die Inhalte des Kurses „Data Science mit Python“ geben einen Überblick, welche Vorerfahrung ein gutes Mitkommen im Seminar ermöglicht. Vor allem: pandas DataFrames (fehlende Werte ergänzen, Daten einlesen), Bibliotheken installieren und laden, eigene Funktionen schreiben, Kenntnis verschiedener Datenstrukturen (list, dictionary, tuple).

Notwendig sind außerdem grundlegende Vorkenntnisse im Bereich der Statistik (Begriffsdefinitionen wie bspw. Mittelwert, Median, Standardabweichung, Normalverteilung), Kenntnisse grundlegender mathematischer Symbole und Begriffe (Summenzeichen, Integral, Funktion, Ableitung, Exponentialfunktion) und Kenntnis der booleschen Algebra mit den logischen Operatoren (UND, ODER, NICHT).

Das Seminar wird auf Deutsch gehalten. Englischkenntnisse (lediglich im Verstehen von englischen Texten) sind sehr empfehlenswert, da die Programmiersprache, Fachbegriffe und die Dokumentationen im Internet auf Englisch sind. Aus diesem Grund sind auch die erstellten Folien in der Schulung auf Englisch.

### Methode des Seminars



Dieses Seminar ist sehr praxisorientiert. Die Teilnehmer arbeiten direkt und selbstständig mit der Programmiersprache Python in der Entwicklungsumgebung Spyder, so dass das Erlernte direkt geübt und vertieft werden kann. Der Trainer moderiert dabei verschiedene Aufgaben und begleitet die Teilnehmer durch die einzelnen Lehreinheiten.

## Technische Voraussetzungen der Teilnehmer (Laptop, etc.)

- ❑ Die Teilnehmer benötigen für die Übungsaufgaben Laptops. Wir empfehlen, Ihren eigenen Laptop mit der vorab installierten Software mitzubringen. Eine genaue Installationsanleitung für die Software wird Ihnen vor dem Seminar per E-mail zugesandt. Auf Anfrage stellen wir auch Schulungslaptops zur Verfügung.
- ❑ Bitte prüfen Sie, ob Ihr Firmenlaptop Zugangsbeschränkungen im Internet hat. Die digitalen Unterlagen (Skript, Code, Dateien) werden im Seminar online zum Download zur Verfügung gestellt. Sie erhalten vor dem Seminar per E-Mail den Link zu einer Testdatei zum Download, um dies überprüfen zu können.
- ❑ Sie sollten sich in firmenfremde WLAN-Netze registrieren können.
- ❑ Als Backup Lösung ist es möglich, dass der USB Port bei Ihrem Laptop freigeschaltet ist, um damit verwendete Dateien oder sonstige Unterlagen übertragen zu können.
- ❑ Im Seminar wird das Betriebssystem Windows verwendet. Der Umgang mit Ihrem verwendeten Betriebssystem und Laptop sollte bekannt sein. Insbesondere sollten Sie ohne Schwierigkeiten Sonderzeichen auf der Tastatur finden (insbesondere bei Apple Geräten werden auf manchen Tastaturen nicht immer runde, eckige bzw. geschweifte Klammern dargestellt).

**10.00-10.15**

## Begrüßung und Organisatorisches

- Vorstellungsrunde
- Erwartungen der Teilnehmer

**10.15-11.45**

## Grundlegendes über Maschinelles Lernen

- Einführung in Maschinelles Lernen
- Definition Supervised – Unsupervised Learning (überwachtes – unüberwachtes Lernen)
- Overfitting, Datasplitting (Training vs. Testdaten)

**11.45-12.00**

## Kaffeepause

**12.00-13.30**

## Lineare Regression

- Erläuterung des Algorithmus
- Datasplitting (Training und Test Daten)
- Umsetzung der linearen Regression in Python (scikit-learn)
- Ergebnisvalidierung

**13.30-14.30**

## Mittagspause

**14.30-16.00**

## Logistische Regression (mit StatsModels)

- Erläuterung des Algorithmus (von linearer zur logistischen Regression)
- Datasplitting (Training und Test Daten)
- Umsetzung der Logistische Regression in Python
- Ergebniss- und Modellvalidierung (AIC, BIC, Konfusionsmatrix)
- Interpretation der Koeffizienten (odds-ratio)
- ROC Curve und AUC (area under the curve)

**16.00-16.15**

## Kaffeepause

**16.15-18.00**

## Entscheidungsbaum

- Erläuterung des Algorithmus
- Datasplitting (Training und Test Daten)
- Umsetzung des Entscheidungsbaums in Python (scikit-learn)
- Anpassen der Hyperparameter
- Ergebnisvalidierung (Konfusionsmatrix, accuracy)
- Entscheidungsbaum für die Regression

**18.00**

## Ende

09.00-09.15

Rückblick und offene Fragen von Tag 1

09.15-10.45

## Ensemble Methods

- Ein Ensemble aus verschiedenen Algorithmen in scikit-learn umsetzen
- Random Forest zur Klassifizierung und zur Regression
- Ada Boost zur Klassifizierung und zur Regression
- Hyperparameter anpassen
- Ergebnisvalidierung

10.45-11.00

Kaffeepause

11.00-12.30

## Weitere wichtige Algorithmen

- K-Nearest Neighbor
- Einfaches neuronales Netz (Multi-Layer Perceptron, MLP)
- Umsetzung der Algorithmen in Python
- Ergebnisvalidierung

12.30-13.30

Mittagspause

13.30-15.15

## Hyperparameter Tuning und Kreuzvalidierung

- Automatisierte Hyperparametersuche in Algorithmen
- Kreuzvalidierung (Cross-Validation)
- Erklärung der beiden Konzepte und Umsetzung in scikit-learn

15.15-15.30

Kaffeepause

15.30-17.00

## Algorithmen zum Clustering

- Grundlagen von K-means Clustering und DBScan
- Cluster-Ergebnisse interpretieren
- Vergleich von Cluster-Ergebnissen verschiedener Algorithmen

17.00

Ende

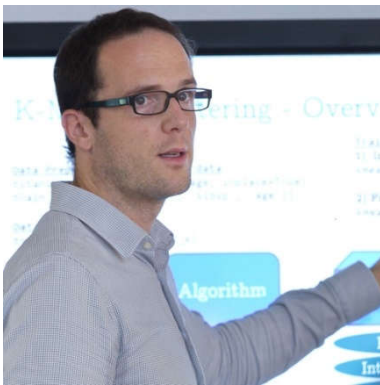
# Ihre Dozenten

Einer unserer folgenden Experten leitet das Seminar



## Dr. Rolf Köhler

Nach dem Studium der Mathematik und BWL promovierte er im Cyber Valley am Max-Planck Institut Tübingen. Sein Forschungsschwerpunkt war im Bereich Machine Learning und Bildverarbeitung. Seit 2015 arbeitet er bei der Robert Bosch GmbH im Bereich Deep Learning und implementiert und adaptiert verschiedene Algorithmen für industrielle Anwendungsfälle. Daraus sind mehrere Patentanmeldungen entstanden. Seit 7 Jahren verwendet er die Programmiersprache Python.



## Jan Köhler

Vom Hintergrund Statistiker und Wirtschaftsingenieur hat er über 7 Jahre an den neuesten Technologien in Machine Learning, Deep Learning und Data Science im Bosch Center for Artificial Intelligence (BCAI) gearbeitet, hat in der Praxis bei über 25 Patentanmeldungen (meist als Haupterfinder) beigetragen und ist Mitautor bei Veröffentlichungen auf Machine Learning Konferenzen neben Veröffentlichungen im Bereich der Medizinstatistik bzw. des Operations Research. In vielen Praxisprojekten unterstützte er bisher als Data Scientist und hat verschiedene Teilnehmer, vom Projektmitarbeiter bis zum Konzern-Vorstand geschult.

## Zusammenfassung

### € Preise

1190 € zzgl. MwSt.




### Termin und Ort

Termine und Orte finden Sie unter <https://enable-ai.de>

**Haben Sie Fragen? Wir helfen Ihnen. Versprochen.**

 Enable AI, Stuttgart

 0711 96881553

 [info@enable-ai.de](mailto:info@enable-ai.de)