

Datenanalyse auf Basis von KI-Methoden

Organisatorisches

Vorlesungen

17.02.2023	15:00 - 18:00 (Präsenz)
18.02.2023	10:00 - 13:00 (Online)
24.02.2023	15:00 - 18:00 (Präsenz)
25.02.2023	10:00 - 13:00 (Online)
03.03.2023	15:00 - 18:00 (Präsenz)
04.03.2023	10:00 - 13:00 (Online)
10.03.2023	15:00 - 18:00 (Präsenz)

Portfolio

Präsentation 1. Vorlesung

R + Python Skripte

Datenanalyse

Wichtigste Aufgabe der **Data Science** → **Datenanalyse**

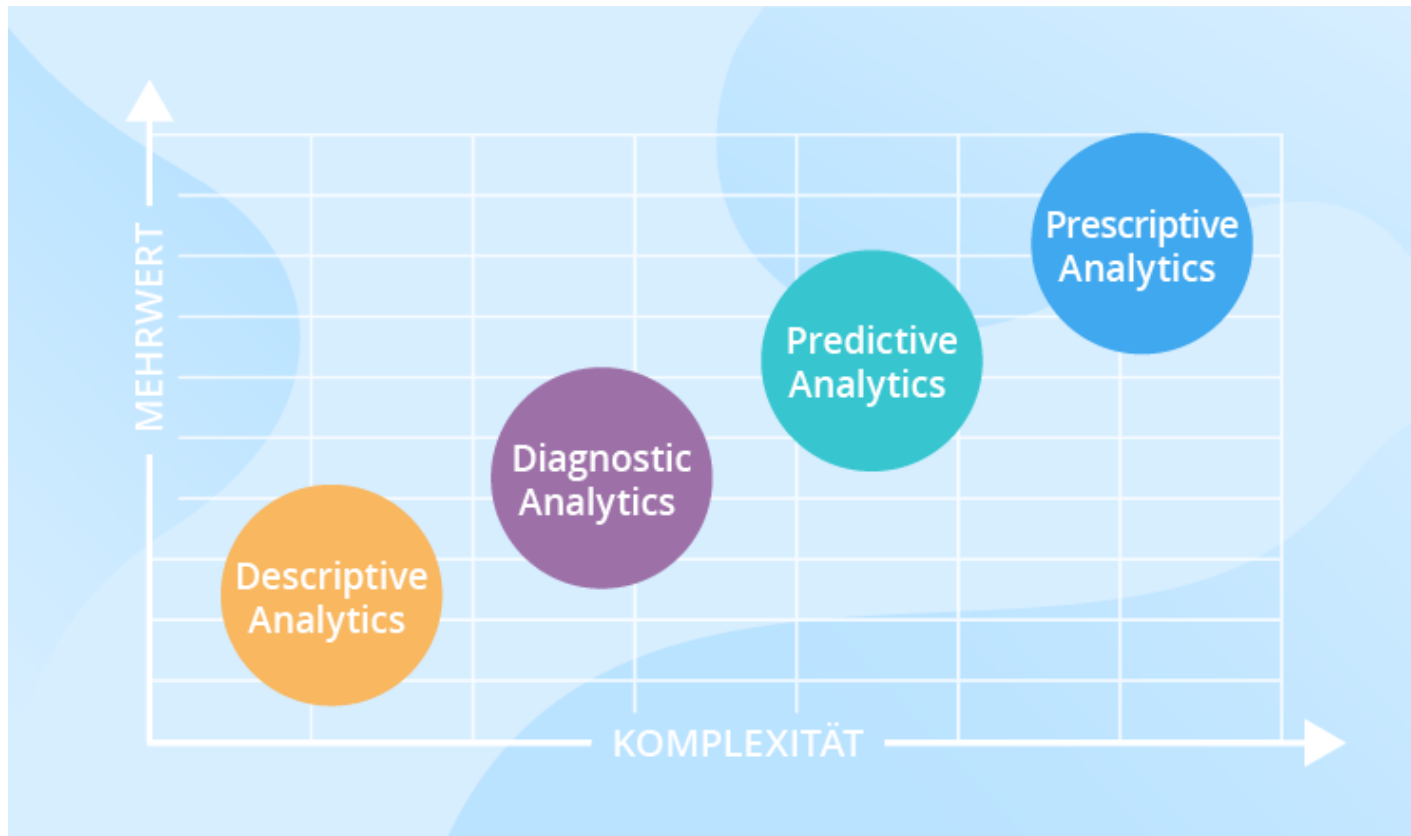
Was ist Datenanalyse?

Die Datenanalyse oder Datenauswertung beschreibt den Prozess der Gewinnung von wertvollen Informationen aus Rohdaten.

- Verwendung von statistischen Methoden für die Analyse von bereits bestehenden Daten
- Ergebnisse der Analyse werden in Form von Zahlen, Fakten, Datenvisualisierungen (z.B. Diagramme, Tabellen) veranschaulicht und zur Weiterverwendung bereitgestellt
- Basis für die Entscheidungsfindung bei Unternehmen
 - Planung von zukünftigen Unternehmensschritten
 - Bewertung von bereits getätigten Entscheidungen
 - Minimierung von Risiken

Datenanalyse

Methoden der Datenanalyse



Datenanalyse

Descriptive Analytics

Die deskriptive oder beschreibende Datenanalyse konzentriert sich auf Daten aus der Vergangenheit, um daran zu erkennen, **was passiert ist**.

Einfachste Form der Datenanalyse, weil sie Trends und Beziehungen beschreibt bzw. visualisiert, aber nicht tiefer geht.

Beispiele:

- Reporting von Finanzen
- Reporting von Logistik
- Reporting von Kundenverhalten
- Umsatz im letzten Quartal
- wie viele Artikel in einem Geschäft zurückgegeben wurden
- etc...

Datenanalyse

Diagnostic Analytics

Bei der diagnostischen Datenanalyse werden historische und andere Daten miteinander verglichen, um Muster zu erkennen, Zusammenhänge aufzudecken und Ursachen und gegenseitige Wechselwirkungen zu identifizieren. Es geht dabei darum, herauszufinden, **warum etwas passiert ist**.

Logisches Schritt nach der deskriptiven Analyse.

Beispiele:

- Untersuchung der Marktnachfrage
- Erklärung des Kundenverhaltens
- Verbesserung der Unternehmenskultur
- etc...

Datenanalyse

Predictive Analytics

Die prädiktive oder vorhersagende Analyse hilft dabei, einen Blick in die Zukunft zu werfen. Anhand dieser Analysemethode soll herausgefunden werden, **was in der Zukunft passiert bzw. passieren könnte**.

Um Vorhersagen treffen zu können, werden die Ergebnisse bereits durchgeführter deskriptiver oder diagnostischer Analysen verwendet.

Beispiele:

- Vorhersage der zukünftigen Umsätzen
- Bestimmung des Personalbedarfs
- Risikoanalyse
- etc...

Datenanalyse

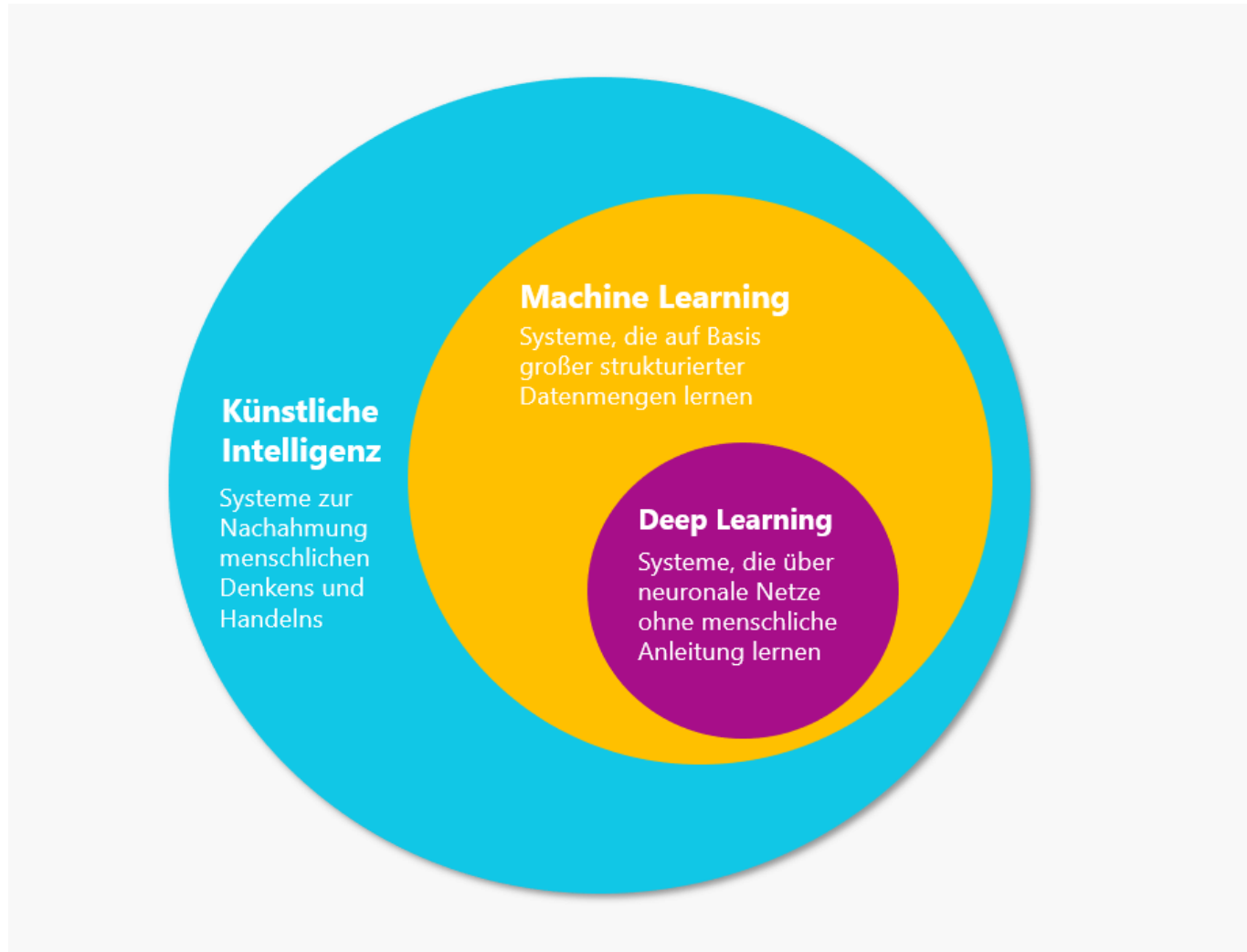
Prescriptive Analytics

Die präskriptive oder verordnende Datenanalyse ist die komplexeste und aufwendigste Analysemethode. Gleichzeitig ist sie für Unternehmen von unglaublich großem Wert, da anhand ihrer Ergebnisse entschieden werden soll, **welche Maßnahmen zu treffen sind**, um etwa ein künftiges Problem zu verhindern, zu beseitigen oder kommende Trends zu nutzen.

Beispiele:

- Marketing Maßnahmen
- Entscheidungen von Führungskräften
- etc....

KI vs Machine Learning vs Deep Learning

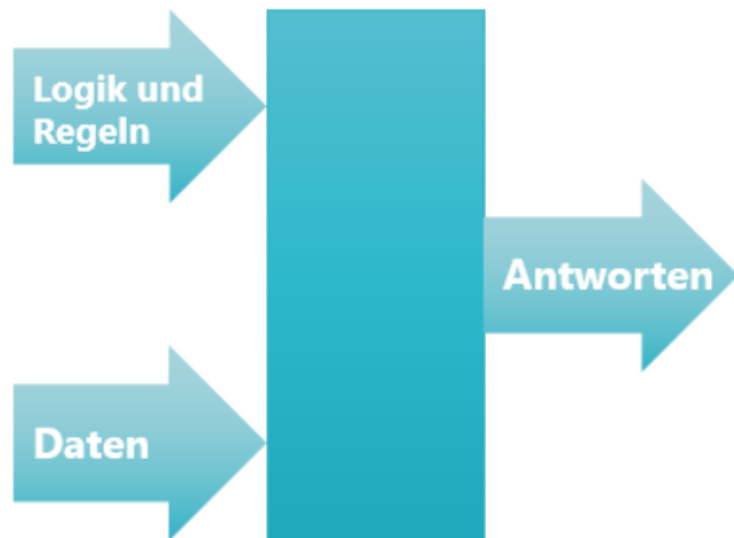


Machine Learning

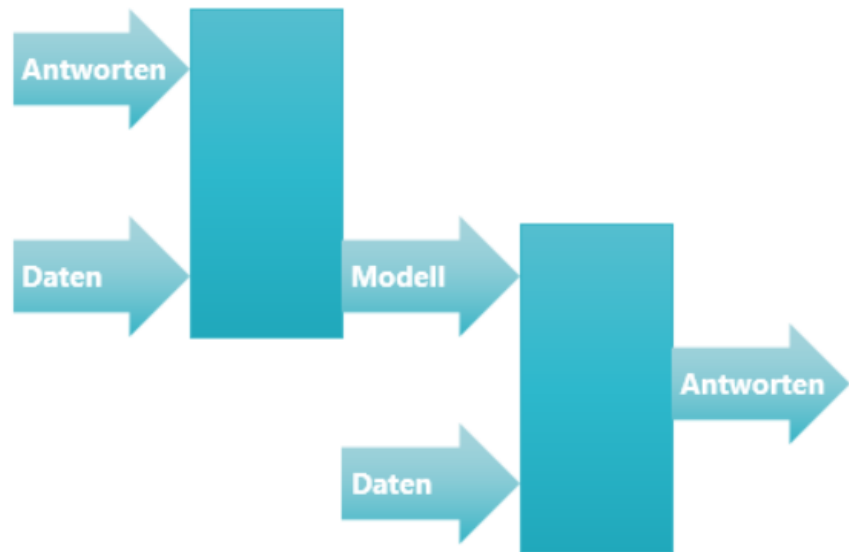
- Maschinelles Lernen ist der elementare Bereich künstlicher Intelligenz
- mit Hilfe von modellbasierten, flexiblen Algorithmen können eigenständig Lösungen für neue und unbekannte Probleme gefunden werden
- Erstellung von Systemen, die aus Daten selbstständig lernen können

Machine Learning

Klassische Programmierung

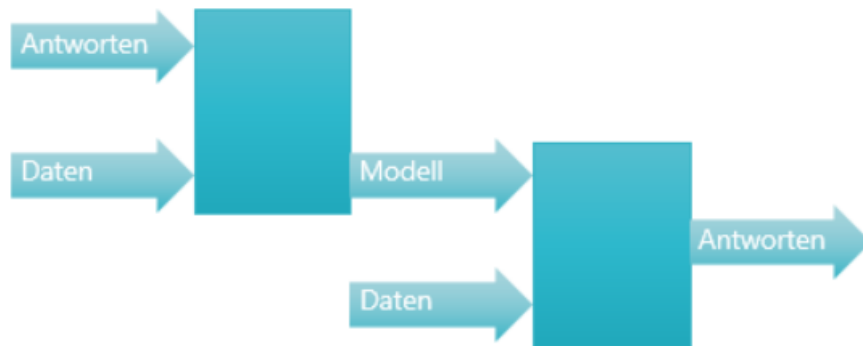


Maschinelles Lernen



Machine Learning

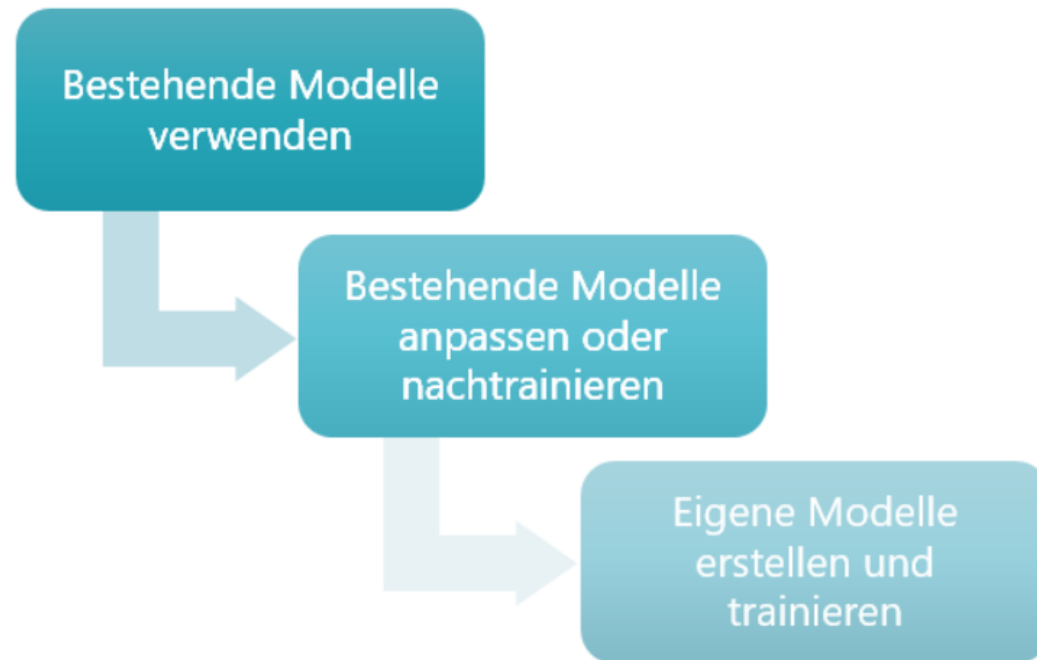
Training



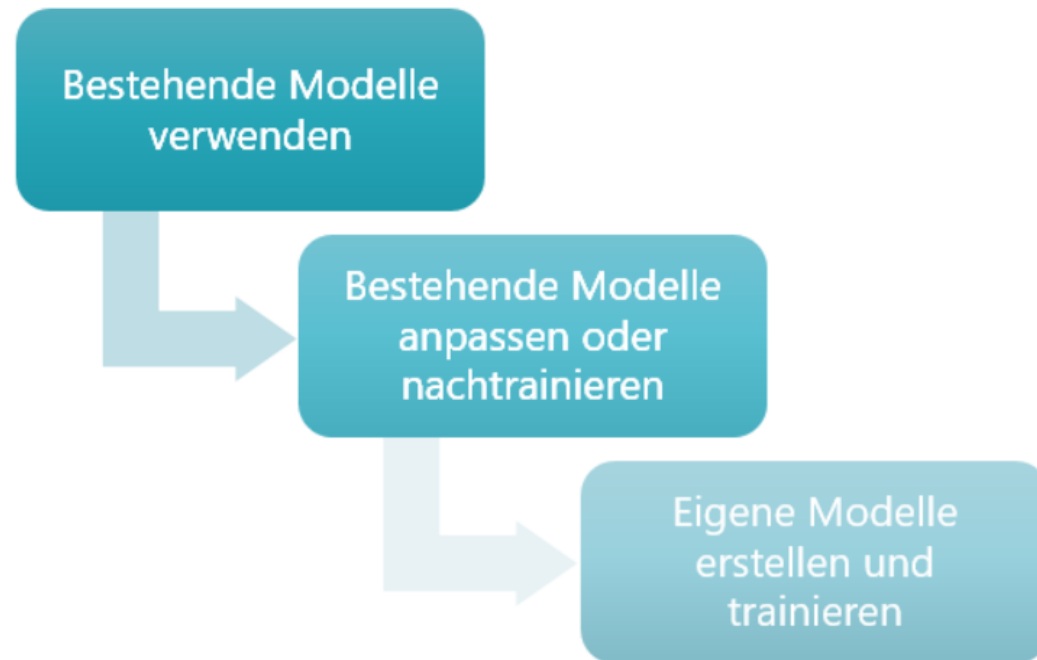
Produktivsetzung



Machine Learning



Machine Learning



Machine Learning

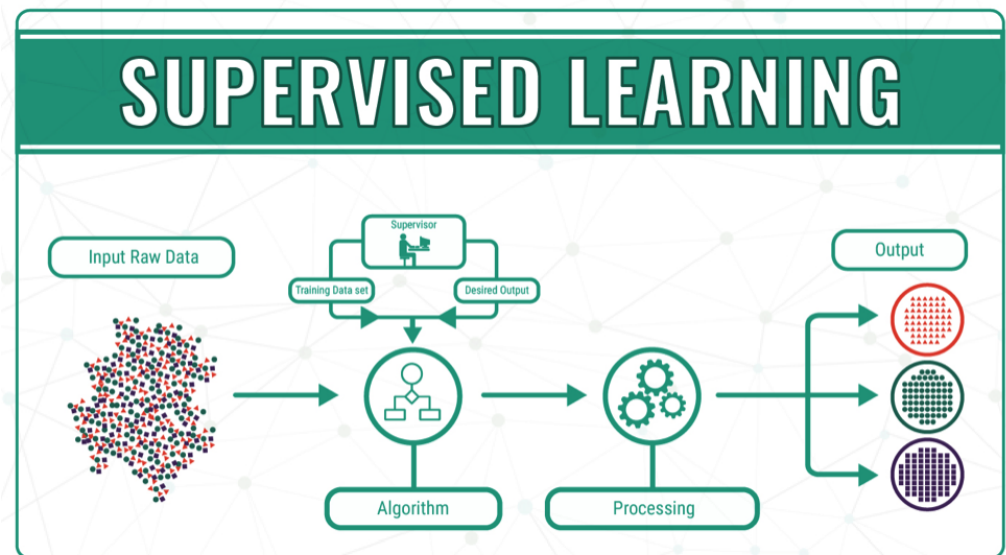
- Überwachtes Lernen
(supervised learning)
- Werden mit Hilfe von
positiven/negativen
Beispielen trainiert

Regression: „durchgehende“ Ausgabe

- Für jeden Input liefert das Modell
einen durchgehenden Wert

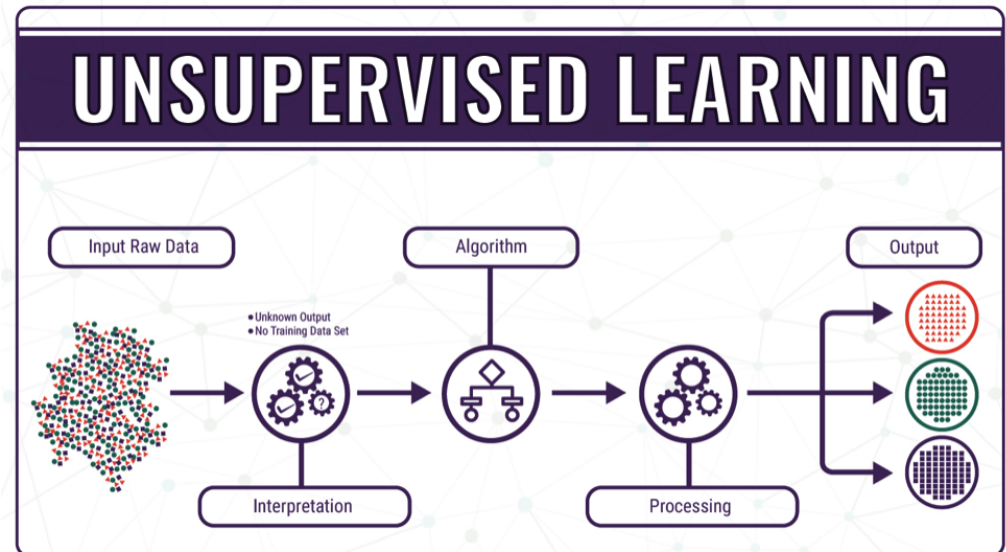
Classification: bestimmte Ausgabe

- Für jeden Input liefert das Modell
einen von speziellen Werten

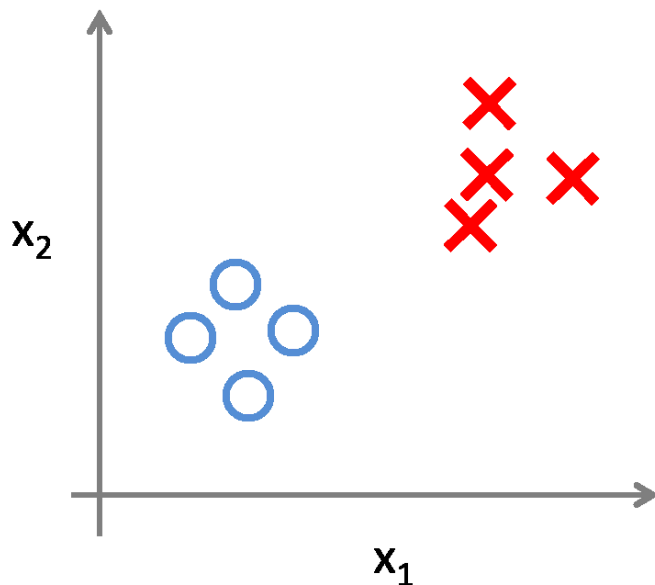


Machine Learning

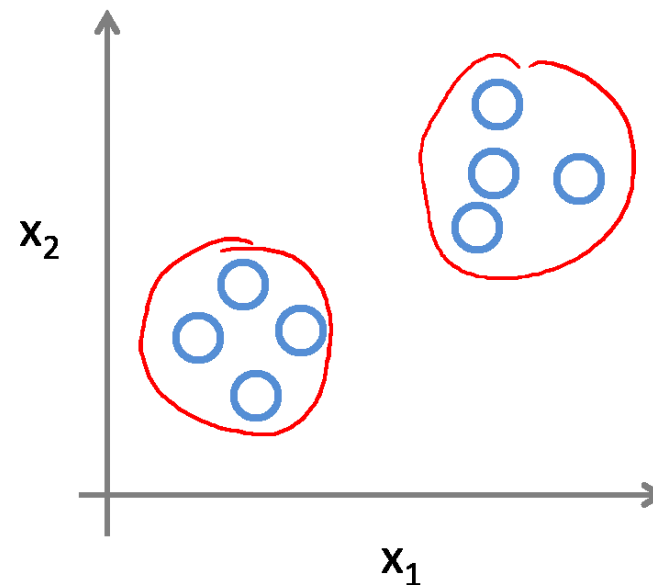
- findet (selbstständig)
Strukturen in Daten
- weiß nicht, was richtig ist
- Bewertung der Güte der
Aktion
- misst den Effekt selbst



Supervised Learning



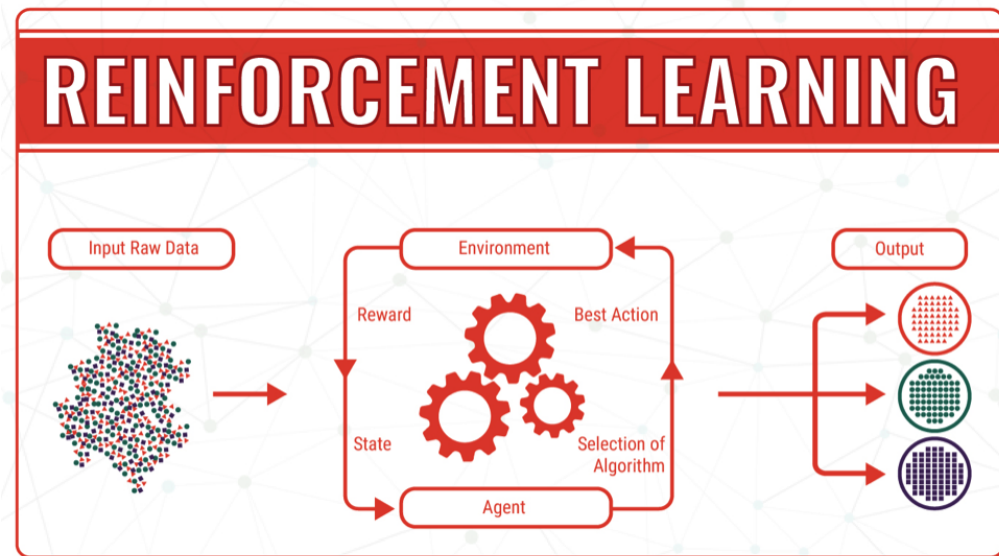
Unsupervised Learning



Während wir bei Supervised Learning schon im Voraus wissen, dass es genau zwei Kategorien gibt (blauer Kreis oder rotes Kreuz), ergeben sich bei Unsupervised Learning zwei Gruppen.

Machine Learning

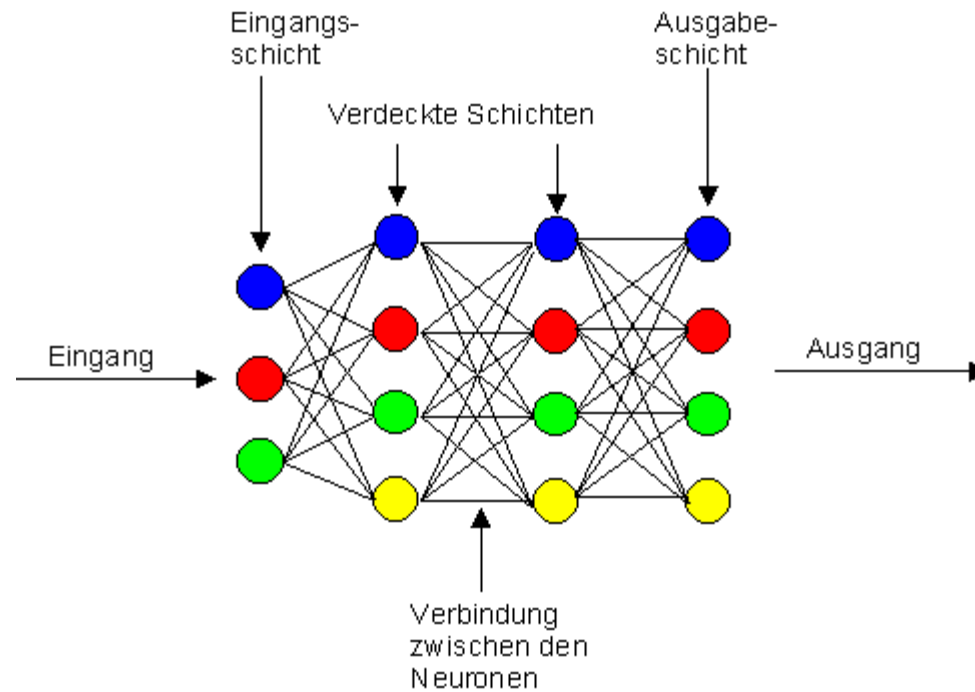
- Lernen durch Belohnung/Bestrafung
- belohnen gute Aktion, bestrafen schlechte Aktion
- Aktion ist grundsätzlich bewertbar (Gewichtung dynamisch)



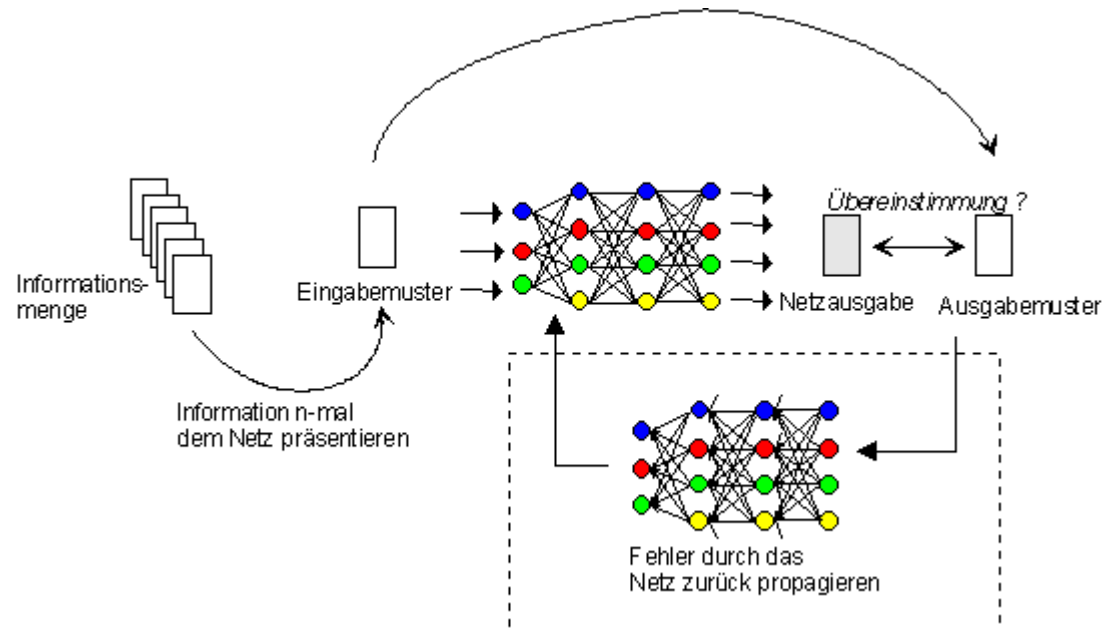
Deep Learning

Deep Learning ist ein Teilbereich des Maschinellen Lernens
Grundlage sind Neuronale Netze mit mindestens einem Hidden-Layer

Deep Learning



Deep Learning



Deep Learning

Die bekannteste **Arten** von neuronalen Netzen sind:

- Feedforward Neuronales Netz
- Rekurrentes Neuronales Netz
- Convolutional Neuronales Netz

Feedforward Neuronales Netz

Definition:

- Neuronales Netzwerk aus künstlichen Neuronen, das keine Rückkopplungen aufweist, d.h. die Signale laufen von der Eingangsschicht immer nur in Richtung der Ausgangsschicht

Anwendung:

- Klassifikation, Mustererkennung, Zeitreihenanalyse, Medizinische Diagnostik, ...

Deep Learning

Rekurrentes Neuronales Netz

Definition:

- Neuronales Netzwerk aus künstlichen Neuronen, das Rückkopplungen aufweist, d.h. Neuronen einer Schicht verbinden sich mit Neuronen derselben oder einer vorangegangenen Schicht

Anwendung:

- Spracherkennung, Prognose über die Zukunft, Simulation von menschlichen Verhaltenssequenzen,....

Convolutional Neuronales Netz

Definition:

- Üblicherweise besteht ein CNN aus mindestens 5 Schichten. Innerhalb jeder dieser Schichten wird eine Mustererkennung durchgeführt. Jede Schicht präzisiert dabei die Mustererkennung auf Basis des Outputs der vorherigen Schicht.

Anwendung:

- Bilderkennung, Audioverarbeitung.....

Deep Learning

Optimierungsmöglichkeiten

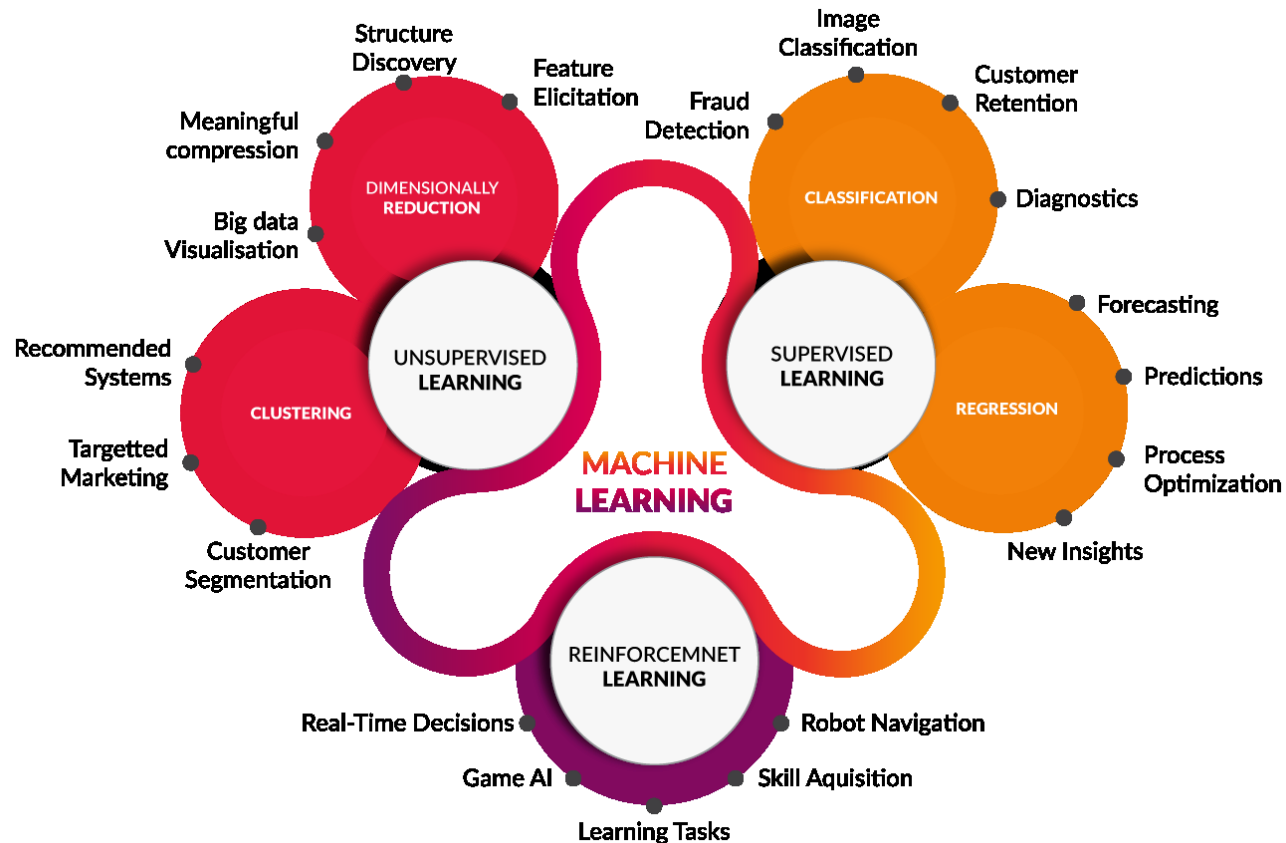
Netzarchitektur

- Anzahl von Hidden-Layers: 1 bis 2 reichen
- Anzahl von Neuronen: 2 *Daumenregel*, höchstens 75% der Neuronen des Input-Layers, oder geometrische Mittel aus der Zahl der Input- und Outputneuronen

Hyperparameter

- Batch-Size:
 - Kleine Datensätze -> alle Daten als ein Batch
 - Große Datensätze -> Mini-Batches
- Learning-Rate: 0.01 anfangen und schrittweise reduzieren bis 0.0001
- Epochen: 100, 1000, bis 10000

Anwendungsfälle KI



Anwendungsfälle KI

Natural Language Processing (NLP)

- Chatbots, Sprachverarbeitung und -verständnis

Regression

- Vorhersagen

Klassifizierung

Anomalie-Erkennung

***Vielen Dank
Für Ihre
Aufmerksamkeit!***