

Methode	Anwendungs- Beispiel	Analogie für intuitives Verständnis	Funktionsweise	Vorteile	Nachteile
<b><u>Supervised Learning</u></b>	<b>Kundenklassifikation im Marketing</b>	<b>Ein Lehrer (Daten) lehrt einen Schüler (Modell) auf Basis von Beispielaufgaben.</b>	<b>Modell wird mit gekennzeichneten Daten trainiert. Es lernt, Vorhersagen zu treffen.</b>	<b>Hohe Genauigkeit bei Vorhersagen, breite Anwendbarkeit</b>	<b>Benötigt große, gekennzeichnete Datensätze</b>
<b>Prognosen (Lineare Regression)</b>	Umsatzvorhersage	Eine Linie, die den besten Fit für die Datenpunkte darstellt.	Ermittelt die Beziehung zwischen unabhängigen und abhängigen Variablen. Nutzt eine Linie, um den besten Fit für die Datenpunkte darzustellen.	Einfach zu interpretieren, gut für lineare Beziehungen	Ungeeignet für nicht-lineare Zusammenhänge
<b>Klassifikation (Decision Trees)</b>	Kreditwürdigkeitsprüfung	Entscheidungsprozess wie ein Entscheidungsbaum im Alltag.	Entscheidungen werden basierend auf Datenattributen in einer Baumstruktur getroffen. Der Baum teilt die Daten in immer kleinere Gruppen auf.	Leicht zu verstehen und zu visualisieren	Neigung zu Überanpassung (Overfitting)
<b>Klassifikation (Random Forest)</b>	Betrugserkennung	Abstimmung einer großen Gruppe von Experten.	Kombination vieler Entscheidungsbäume, um Vorhersagen zu verbessern. Jeder Baum trägt zur finalen Entscheidung bei.	Hohe Genauigkeit, reduziert Überanpassung	Komplex, schwer zu interpretieren
<b>Klassifikation (XGBoost)</b>	Zielgruppensegmentierung	Training eines Expertenteams, wobei jeder Experte auf Fehler des anderen aufbaut.	Optimierter Boosting-Algorithmus kombiniert viele Entscheidungsbäume. Er lernt aus den Fehlern vorheriger Bäume.	Sehr hohe Leistung und Effizienz	Komplexität, erfordert gründliche Parametereinstellung
<b>Support Vector Machines (SVMs)</b>	Bildklassifikation	Wie das Spannen einer Seillinie, um zwei Gruppen voneinander zu trennen.	SVMs trennen Datenpunkte durch eine Hyper-Ebene, die die größte Trennung zwischen den Klassen maximiert.	Effektiv bei hohen Dimensionen, robust gegenüber Ausreißern	Schwer zu interpretieren, hohe Rechenkomplexität
<b>K Nearest Neighbors (KNN)</b>	Empfehlungsdienste	Wie das Klassifizieren einer Person basierend auf den Eigenschaften ihrer Freunde.	Klassifiziert Datenpunkte basierend auf den Klassen der nächsten Nachbarn. Ein Punkt wird der häufigsten Klasse seiner k nächsten Nachbarn zugeordnet.	Einfach zu implementieren, keine Trainingsphase erforderlich	Performance sinkt mit zunehmender Anzahl der Datenpunkte
<b>Naive Bayes</b>	Spam-Erkennung in E-Mails	Wie das Einschätzen der Wahrscheinlichkeit eines Ereignisses basierend auf früheren Erfahrungen.	Nutzt Bayes' Theorem, um die Wahrscheinlichkeit einer Klasse basierend auf den Attributen der Datenpunkte zu berechnen. Annahme der Unabhängigkeit zwischen den Attributen.	Schnell, effizient, gut bei großen Datensätzen	Annahme der Unabhängigkeit ist oft unrealistisch
<b>Logistische Regression</b>	Kundenzufriedenheitsanalyse	Wie das Vorhersagen, ob ein Licht an- oder ausgeht basierend auf dem Schalterzustand.	Modelliert die Wahrscheinlichkeit eines binären Ergebnisses basierend auf einer linearen Kombination von Prädiktoren.	Einfach zu interpretieren, effizient zu berechnen	Ungeeignet für nicht-lineare Zusammenhänge, eingeschränkte Flexibilität
<b>Neurale Netze</b>	Bildererkennung, Spracherkennung	Wie das Lernen und Erkennen von Mustern durch das menschliche Gehirn.	Simuliert die Funktionsweise des menschlichen Gehirns durch miteinander verbundene Knoten (Neuronen). Lerne durch Anpassung der Verbindungsgewichte.	Hohe Genauigkeit, vielseitig	Erfordert große Datenmengen und Rechenressourcen, schwer zu interpretieren

Methode	Anwendungs- Beispiel	Analogie für intuitives Verständnis	Funktionsweise	Vorteile	Nachteile
<b><u>Unsupervised Learning</u></b>	<b>Kundensegmentierung</b>	<b>Entdecker erkundet unbekanntes Land ohne Karte.</b>	<b>Modell lernt selbstständig Muster und Strukturen in den Daten. Es wird nicht mit gekennzeichneten Daten trainiert.</b>	<b>Erkennt verborgene Muster ohne Vorwissen</b>	<b>Ergebnisse können schwer interpretierbar sein</b>
<b>Clustering (Hierarchisch)</b>	Marktsegmentierung	Sortieren von Dokumenten in immer kleinere Ordner.	Daten werden in eine Baumstruktur von Clustern gruppiert. Jede Gruppe wird weiter unterteilt.	Identifiziert hierarchische Beziehungen zwischen Daten	Rechenintensiv, schwer zu skalieren
<b>Clustering (k-Means)</b>	Segmentierung von Zielgruppen	Gruppierung von Menschen auf Basis von Ähnlichkeiten.	Daten werden in k Cluster aufgeteilt. Jeder Datenpunkt wird dem nächsten Zentrum zugewiesen.	Schnell, einfach zu implementieren	Anzahl der Cluster muss vorab festgelegt werden
<b>Principal Component Analysis (PCA)</b>	Risikomodellierung	Verkleinern einer großen Datenmenge auf ihre wichtigsten Elemente.	Reduziert die Anzahl der Zufallsvariablen durch Transformation der Daten. Maximiert die Varianz in den Hauptkomponenten.	Reduziert die Komplexität der Daten	Verlust von Information möglich
<b>Isolation Forest</b>	Anomalieerkennung in Netzwerkdaten	Finden von Eindringlingen in einem Netzwerk.	Isoliert Datenpunkte durch zufällige Entscheidungsbäume. Abweichungen benötigen weniger Partitionen.	Effizient für große Datensätze	Möglicherweise schwer zu interpretieren
<b>Local Outlier Factor (LOF)</b>	Betrugserkennung	Auffinden eines verdächtigen Kundenverhaltens in einer Bankfiliale.	Bewertet die Dichte eines Datenpunkts im Vergleich zu seinen Nachbarn. Identifiziert lokale Anomalien.	Erkennt lokale Abweichungen gut	Sensitivität gegenüber Parameterwahl
<b>Apriori Algorithm</b>	Warenkorbanalyse	Finden von häufig zusammen gekauften Produkten in einem Supermarkt.	Identifiziert häufig auftretende Itemsets und erstellt Assoziationsregeln. Iteriert über alle möglichen Itemsets.	Einfach zu implementieren, interpretiert leicht	Rechenintensiv bei großen Datensätzen
<b>FP-Growth (Frequent Pattern Growth)</b>	Empfehlungsdienste	Wie das Lernen von Kaufmustern durch wiederholtes Scannen eines Katalogs.	Nutzt eine Baumstruktur, um häufige Muster effizient zu extrahieren. Reduziert die Anzahl der Durchläufe durch die Daten.	Effizienter als Apriori, besonders bei großen Datensätzen	Komplexer zu implementieren
<b>DBSCAN (Density-Based Spatial Clustering of Applications with Noise)</b>	Geodatenanalyse	Gruppieren von Städten basierend auf Bevölkerungsdichte.	Identifiziert Cluster basierend auf der Dichte von Punkten. Kann Rauschen und Ausreißer erkennen.	Erkennt Cluster beliebiger Form, identifiziert Rauschen	Sensitivität gegenüber Parameterwahl
<b>Self-Organizing Maps (SOMs)</b>	Mustererkennung in Bildern	Wie das Sortieren von Fotos in einem Album nach Ähnlichkeiten.	Neuronales Netzwerk, das Daten auf eine niedrigdimensionale Karte projiziert. Lernt durch Wettbewerbslernen.	Gute Visualisierung, identifiziert Muster	Komplexität, schwer zu interpretieren
<b>Gaussian Mixture Models (GMMs)</b>	Gesichtserkennung	Wie das Mischen mehrerer Farben, um einen neuen Farbton zu erzeugen.	Modelliert die Daten als Mischung mehrerer normalverteilter Komponenten. Berechnet die Wahrscheinlichkeit, dass ein Punkt zu einer Komponente gehört.	Flexibel, gut für Cluster beliebiger Form	Erfordert die Anzahl der Komponenten im Voraus
<b>Variational Autoencoders (VAEs)</b>	Krankheitsvorhersage	Wie das Erstellen einer komprimierten Version eines Bildes, das dann dekomprimiert wird.	Neuronales Netzwerk, das die zugrundeliegende Datenverteilung lernt. Komprimiert und rekonstruiert die Daten.	Generiert realistische Daten, robust gegenüber Rauschen	Komplex, erfordert tiefes Wissen über neuronale Netze
<b>Generative Adversarial Networks (GANs)</b>	Erstellung realistischer Bilder	Wie ein Künstler, der versucht, ein Gemälde zu fälschen, während ein Experte die Fälschung erkennt.	Zwei konkurrierende Netzwerke lernen, realistische Daten zu generieren. Ein Netzwerk generiert, das andere bewertet die Echtheit.	Hervorragend für die Erstellung realistischer Daten	Schwierig zu trainieren, kann instabil sein
<b>Large Language Models (LLMs)</b>	Textgenerierung und -analyse	Wie ein sehr belesener Mensch, der in der Lage ist, auf verschiedene Themen zu antworten.	Modelle wie GPT-4 analysieren und generieren natürliche Sprache durch Training auf großen Textkorpora. Sie lernen, Sprache zu verstehen und kontextbezogen zu antworten.	Leistungsfähig bei der Verarbeitung und Generierung natürlicher Sprache	Erfordert immense Datenmengen und Rechenressourcen, mögliche Verzerrungen in den Daten