

Interessante Muster	<p>Valide: Allgemein anwendbar und nicht zufällig.</p> <p>Neu: Bietet neue Erkenntnisse oder widerspricht Bekanntem.</p> <p>Nützlich: Hat wirtschaftlichen oder praktischen Mehrwert.</p> <p>Nachvollziehbar: Logisch und verständlich erklärbar.</p>
Formen maschinelles lernen	<p>Supervised Learning:</p> <ul style="list-style-type: none"> Überwachtes Lernen: Algorithmen lernen aus gekennzeichneten Daten (Eingabe + erwartete Ausgabe). <p>Regression:</p> <ul style="list-style-type: none"> Vorhersage von kontinuierlichen Werten (z.B. Preisvorhersage). <p>Classification:</p> <ul style="list-style-type: none"> Einordnung in Kategorien (z.B. Spam oder Nicht-Spam). <p>Unsupervised Learning:</p> <ul style="list-style-type: none"> Unüberwachtes Lernen: Algorithmen lernen aus unbeschrifteten Daten (nur Eingaben). <p>Clustering:</p> <ul style="list-style-type: none"> Gruppierung ähnlicher Datenpunkte (z.B. Kundensegmentierung). <p>Reinforcement Learning:</p> <ul style="list-style-type: none"> Verstärkendes Lernen: Ein Agent lernt durch Versuch und Irrtum, basierend auf Belohnungen und Strafen. <p>Agent:</p> <ul style="list-style-type: none"> Der lernende Akteur, der Entscheidungen trifft. <p>Action:</p> <ul style="list-style-type: none"> Die vom Agenten ausgeführte Handlung. <p>Reward:</p> <ul style="list-style-type: none"> Rückmeldung, ob die Aktion gut oder schlecht war. <p>Environment:</p> <ul style="list-style-type: none"> Die Umgebung, in der der Agent handelt und lernt.
Business Anwendungen Klassifikation	<p>Kreditvergabe: Entscheidung, ob ein Kredit genehmigt wird.</p> <p>Betrugserkennung: Erkennen von betrügerischen Aktivitäten.</p> <p>Targeted Marketing: Zielgerichtete Werbung Kundengruppen</p> <p>Kundenabwanderung (Churn): Vorhersage, ob ein Kunde abspringt.</p> <p>Inhaltsfilterung: Zum Beispiel Spam-Erkennung bei E-Mails.</p>
Business Anwendungen Regression	<p>Nachfragevorhersage:</p> <ul style="list-style-type: none"> Prognose der zukünftigen Nachfrage nach Produkten. (Wie oft wird Produkt gekauft) <p>Preissensitivität:</p> <ul style="list-style-type: none"> Analyse, wie Preisänderungen die Nachfrage beeinflussen. (Wie viel ist jemand bereit etwas zu bezahlen) <p>Kosten-/Risikovorhersage:</p> <ul style="list-style-type: none"> Einschätzung finanzieller Risiken oder Erträge. <p>Erfolgsvorhersage:</p> <ul style="list-style-type: none"> Vorhersage des Erfolgs von Marketingkampagnen.

CRISP-DM-Zyklus	<ol style="list-style-type: none"> Business Understanding: Verstehen, was das Problem aus geschäftlicher Sicht ist. Data Understanding: Analysieren, welche Daten verfügbar sind und wie sie genutzt werden können. Data Preparation: Daten bereinigen und aufbereiten. Modeling: Ein Modell erstellen, um das Problem zu lösen. Evaluation: Überprüfen, ob das Modell das Problem gut löst. Deployment: Das Modell in der Praxis anwenden.
1. CRISP-DM: Problemformalisierung Business Understanding / Data Understanding	<ul style="list-style-type: none"> Formalisierte Problembeschreibung Eine klare und strukturierte Problemformulierung erstellen, basierend auf dem Verständnis von Geschäftsprozessen und Daten. (Wie bereite ich Daten so vor damit Orange es versteht)
2. Daten vorbereiten Data Preperation	<p>Hier geht es darum fehlende Rohdaten aufzuarbeiten damit sie für ML geeignet sind.</p> <ol style="list-style-type: none"> Fehlende Werte behandeln Skalierung (zwischen 0 & 1) Diskretisierung (Altersgruppen statt Altersangabe)
3. Modell bauen Modelling	<p>Erstellung eines maschinellen Lernmodells basierend auf den vorbereiteten Daten.</p> <ol style="list-style-type: none"> Verfahren auswählen: korrekter ML Algorithmus wählen Vor & Nachteile: verstehen ML Algorithmen Hyperparameter-Tuning: Optimierung Modellparameter & Vorhersagegenauigkeit <p>Ziel: Leistungsfähiges modell entwickeln</p>
4. Modell evaluieren Evaluation	<p>Das Modell wird anhand verschiedener Metriken & Tests bewertet. Es wird geprüft ob das Modell die gewünschten Ergebnisse liefert & ob Probleme auftreten.</p> <ol style="list-style-type: none"> Ergebnisse beurteilen (Precision, Recall, Kreuzvalidierung) Problem erkennen (Overfitting, Underfitting) <p>Ziel: Sicherstellen das Modell auf neuen unbekannten Daten funktioniert & Fehler frühzeitig erkennen.</p>
Zielformat (Datenaufbereitung)	<p>Aufbau:</p> <ul style="list-style-type: none"> Header: Die oberste Zeile enthält die Spaltennamen (z.B. X1, X2... Y). Merkmale (Features): Spalten X1, X2, ... repräsentieren die Eingangsvariablen (Eigenschaften) des Modells. Zielvariable (Label): Die Spalte Y enthält die Ausgabe oder Klassifikation (z.B. A, B). Datenwerte: Jede Zeile stellt einen Datensatz dar, bestehend aus Werten für die Merkmale und dem zugehörigen Label.
Formalisierung Klassifikation	<p>Instanzen: Kunden, Patienten, Kunden ID Klassenattribut: Spam/nicht Spam, Krank/Gesund Attribute: Geschlecht, Alter, Region, Vertragsart Algorithmen: Entscheidungsbaum, Random Forest, KNN</p>

Formalisierung Regression	Instanzen: Aktienkurse, Wetterdaten Zielvariable: Preis, Temperatur, Umsatz, Anzahl Besucher Attribute: Fläche Zimmeranzahl, Standort Algorithmen: Lineare Regression, Lasso Ridge, KNN
Feature Engineering	Sinnvolle Merkmale (Attribute) für ein maschinelles Lernmodell zu finden & zu erstellen. 1. Brainstorming: <ul style="list-style-type: none"> Welche Eigenschaften hilfreich / Business Understanding 2. Alles erlaubt: <ul style="list-style-type: none"> Jede Idee ist willkommen 3. Korrelation prüfen <ul style="list-style-type: none"> Prüfen ob gewählte Merkmale tatsächlich mit Zielvariable zusammenhängen 4. Einfach ausprobieren
Was eignet sich nicht als Attribut	<ul style="list-style-type: none"> IDs: Jeder Kunde hat eigene ID, sagt nichts über Zusammenhänge Konstante Werte: Wenn Attribut denselben Wert hat (Alle Kunden aus der gleichen Stadt) Zukünftige Werte: Werte die zum Zeitpunkt Vorhersage unbekannt sind (zukünftige Umsätze)