**Zusammenfassung Theorie M259**

Inhaltsverzeichnis

[Künstliche Intelligenz 2](#_Toc159966970)

[Übersicht AI 2](#_Toc159966971)

[Supervised Learning 3](#_Toc159966972)

[Unsupervised Learning 3](#_Toc159966973)

[Reinforcement Learning 3](#_Toc159966974)

[Kultur und Limitationen 3](#_Toc159966975)

[Limitationen 3](#_Toc159966976)

[Abstraktes Denkvermögen 3](#_Toc159966977)

[Rekursivität 4](#_Toc159966978)

[Transparenz 4](#_Toc159966979)

[Emotionen 4](#_Toc159966980)

[Grenzen 4](#_Toc159966981)

[Datenschutz und Recht 4](#_Toc159966982)

[Datenschutzgesetz 4](#_Toc159966983)

[Datensicherheit 5](#_Toc159966984)

[Code of Ethics 5](#_Toc159966985)

[Datentypen 5](#_Toc159966986)

[Kategorische Datentypen 5](#_Toc159966987)

[Numerische Datentypen 6](#_Toc159966988)

[Data Science Process / CRISP-DM 6](#_Toc159966989)

[EDA 7](#_Toc159966990)

[Daten aufbereiten 7](#_Toc159966991)

[Daten modellieren 7](#_Toc159966992)

# Künstliche Intelligenz

* Künstliche Intelligenz ist die Fähigkeit einer Maschine, menschliche Fähigkeiten zu imitieren 🡪 keine Emotionen
* Umwelt wahrnehmen und umgehen, Probleme lösen um ein gewünschtes Ziel zu erreichen
* Empfängt Daten, verarbeitet und reagiertEin Bild, das Text, Screenshot, Schrift, Diagramm enthält.

  Automatisch generierte Beschreibung

# Übersicht AI

Ein Bild, das Text, Screenshot, Diagramm enthält.

Automatisch generierte Beschreibung

## Supervised Learning

* Überwachtes lernen
* Erwartete Resultate beim Training übergeben
* Maschine lernt bei welchen Inputs welche Resultate zu erwarten sind
* Beim lernen:
  + Input: Answers, Data
  + Output: Rules
* Bei Verwendung:
  + Input: Data
  + Output: Predictions
* Klassifikation: Gesichtserkennung, E-Mail-Spam Erkennung
* Regression: Immobilienpreisvorhersage, Temperaturvorhersage

## Unsupervised Learning

* Unüberwachtes lernen
* Generative Aufgabe statt beschreibende
* Bild oder Lied erstellen anstatt erkennen was auf Bild ist
* Erwartete Resultate werden nicht mitgegeben
* Maschine versucht selber Patterns zu erkennen

## Reinforcement Learning

* Bestärkenden Lernen
* Mit Belohnungen und Bestrafungen arbeiten
* Keine Aktion vorgemacht
* Aktion führt zu positiven oder negativen Bestätigung 🡪 Belohnung/Bestrafung
* Gameentwicklung oder Robotik

# Kultur und Limitationen

## Limitationen

AI hat einige Limitationen

## Abstraktes Denkvermögen

* Mit richtigen Daten kann AI «denken» aber abstraktes Denkvermögen nicht vorhanden
* AI lernt von Mustern und ist stark in Mustererkennung und Vorhersage
* Aktuell kann nicht mit kleineren Datenmengen oder Verädnerungen umgehen

Beispiel:  
Im Feld der "Driving Assistance" ist dies ein grosses Problem. Ein Stop Schild welches fest von Schnee abgedeckt ist erkennt weiterhin jeder aufmerksame Fahrer. Die AI kann aber mit der geringen Pixelmasse des Stoppschildes nicht umgehen.

## Rekursivität

* AI kann sich nicht selber verbessern
* Kann aber selbständig lernen und intelligenter werden
* Nur Menschen können Limitationen erkennen und das Modell verbessern
* Algorithmus ist abhängig von Hardware, Modellen und Optimierung

## Transparenz

* Oft nur schwer nachvollziehbar wie eine Maschine zur Entscheidung gekommen ist
* Verhinderung von Sexismus und Rassismus muss man mit Transparenz berücksichtigen

## Emotionen

* AI kann keine Emotionen fühlen nur erkennen
* Emotionen zu erkennen ist schwer für die Maschine
* Kann Denken eines Menschen nicht übernehmen

## Grenzen

* AI spezifisch bauen

# Datenschutz und Recht

* Arbeit mit Daten 🡪 diverese Gesetze und best practices im Spiel

## Datenschutzgesetz

Das Datenschutzgesetz (DSG) der Schweiz bezweckt "Dieses Gesetz bezweckt den Schutz der Persönlichkeit und der Grundrechte von Personen, über die Daten bearbeitet werden."

* Drei Arten von Daten:
  + Personendaten
  + Besonders schützenswerte Personendaten
  + Persönlichkeitsprofile

**Bestimmt vs Bestimmbar**

Bestimmt: wenn Name steht

Bestimmbar: wenn Name durch ID ersetzt wurde

**Allgemeine Regelungen**

Personendaten dürfen nur Rechtmässig bearbeitet werden:

* Treu und Glaube
* Nur für den Zweck benutzten welchen man angegeben hat
* Beschaffung und Bearbeitung müssen für die betroffene Person erkennbar sein
* Braucht Einwilligung einer Person

## Datensicherheit

Daten vor Diebstahl, Ausfall oder Verlust schützen. Dies muss bei der Entwicklung sichergestellt werden.

## Code of Ethics

Ein Roboter:

* Keine Menschen verletzten oder durch seine Untätigkeit Menschen verletzen lassen
* Befolgt Befehle
* Beschützt sich selber
* Solange die vorderen Gesetze nicht verletzt werden

# Datentypen

Ein Bild, das Diagramm, Plan, Text, technische Zeichnung enthält.

Automatisch generierte Beschreibung

## Kategorische Datentypen

Kategorische Datentypen sind Daten welche eine gewisse Charakteristik besitzen. Dies z.B. Haarfarbe, Sprache oder Lieblingsfilmgenre.

**Ordinal**

* Haben gewisse Ordnung/Reihenfolge
* Ein Beispiel dazu sind T-Shirt Grössen (XS, S, M, L, XL)

**Nominal**

* Haben keine bestimmte Reihenfolge
* Ein Beispiel hierzu ist eine Sprachauswahl (Deutsch, Englisch, Französisch, Italienisch)

## Numerische Datentypen

Numerische Datentypen sind, wie der Name schon sagt, Nummern. Diese können in zwei Ausprägungen auftauchen:

**Diskret**

* Diskrete Daten sind eindeutig und voneinander getrennt. Solche Daten können nicht gemessen sondern nur gezählt werden
* Zum Beispiel: Anzahl Lernende in einer Klasse.

**Kontinuierlich**

* Kontinuierliche Daten sind messbare Daten
* zum Beispiel die Grösse einer Person oder die Temperatur.

# Data Science Process / CRISP-DM

Das CRISP-DM (Cross Industry Standard Process for Data Mining) ist ein Standard Prozess für Data Mining. Dieser kann für jedes Datenprojekt angewendet werden und hilft dabei, vollständige und gute Datenmodelle zu bauen.

Ein Bild, das Diagramm, Kreis, Entwurf, Zeichnung enthält.

Automatisch generierte Beschreibung

Für genauere Beschreibung siehe gitbook.

# EDA

Ziel von EDA ist ein Grundverständnis über den Datenbestand zu schaffen.

EDA hat folgende Ziele:

* Einblick in das vorliegende Datenset gewinnen
* Die Struktur des Datensets erkennen
* Wichtige Variablen herausfinden
* Ausreisser und Anomalien erkennen
* Erste Hypothesen testen
* Sparsame erste Modelle entwickeln
* Ideale Einstellungen für Modelle erkennen

# Daten aufbereiten

Reale Datensätze sind oft schmutzig

Gründe:

* Falsch geschrieben
* Werte in falschen Feldern
* Irregularitäten
* Duplikate
* Fehlende Werte 🡪 NULL Values

# Daten modellieren

Data Science hauptsächlich daraus, wirtschaftliche Fragestellungen in datengetriebene Fragestellungen zu übersetzen und dann Daten zu sammeln, zu verstehen, zu säubern und zu formatieren.

Im generellen ist ein Modell eine Abbildung der Realität. Meist aber nur eine Annährung der Realität 🡪 oft simpler als im echten Leben

Oft zu wenig Datenpunkte als zu viel.

Kategorische Datentypen als numerischer Datentyp abspeichern.