Projekte. Beratung. Spezialisten.

Individuelle Softwarelösungen

Daten – Information – Wissen

Möglichkeiten und Grenzen des Machine Learning

IKS-Thementag

Autor: Martin Gossen

14.05.2019

IKS Gesellschaft für Informations-und Kommunikationssysteme mbH

T. +49 2103-5872-0 | www.iks-gmbh.com

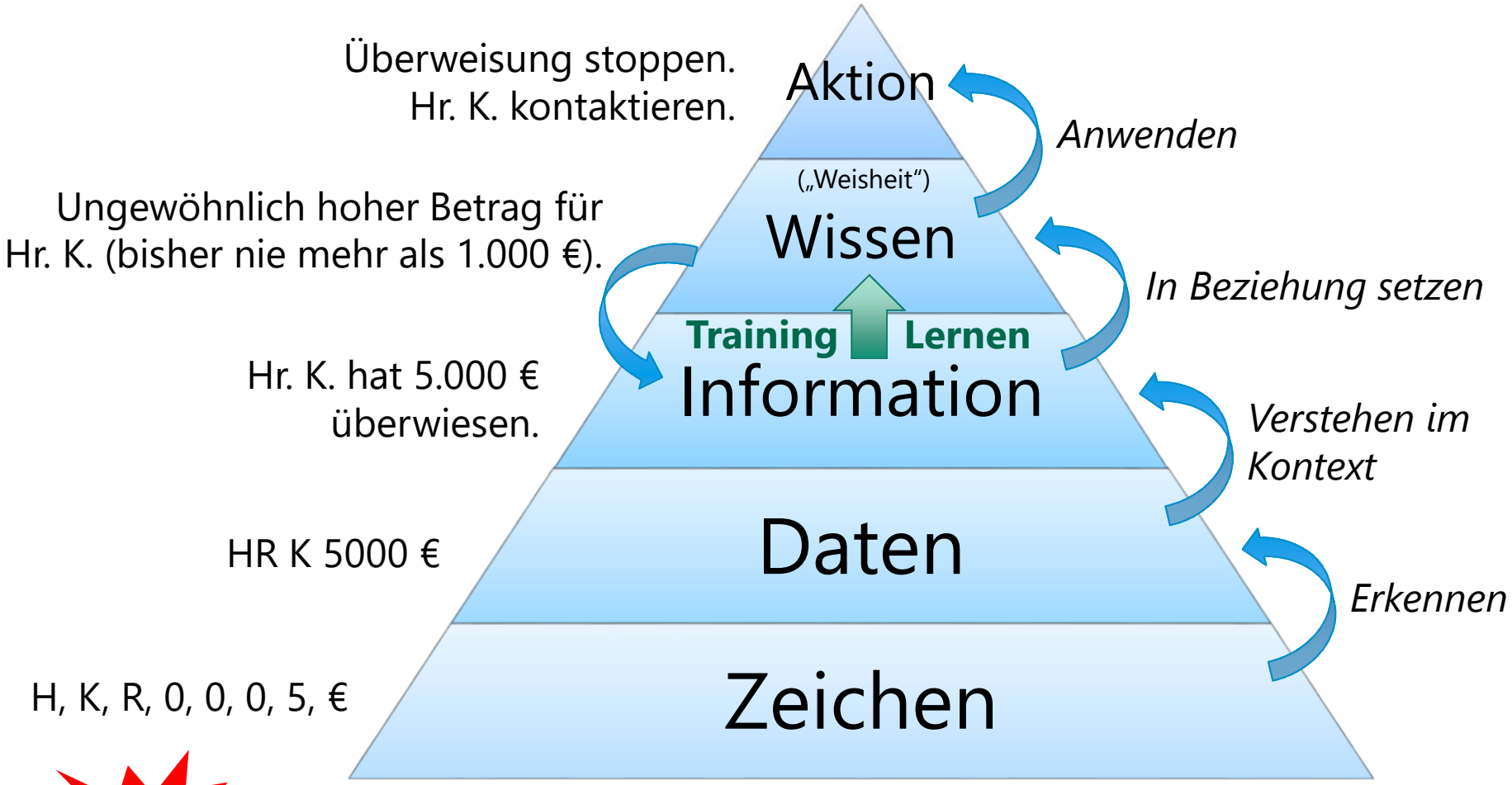
Agenda

- „Wissen“
- Technische Ansätze
- Anwendungsfälle
- Datenquellen
- Know-how
- Grenzen

Agenda

- **„Wissen“**
- Technische Ansätze
- Anwendungsfälle
- Datenquellen
- Know-how
- Grenzen

Was ist Wissen?



nach Aamodt und Nygård, 1995

Was ist Wissen?

Arbeitsdefinition

Wissen

Erkenntnisse, Muster und Schlüsse, die aus der Verknüpfung von Informationen gewonnen werden und sinnvolles Handeln ermöglichen

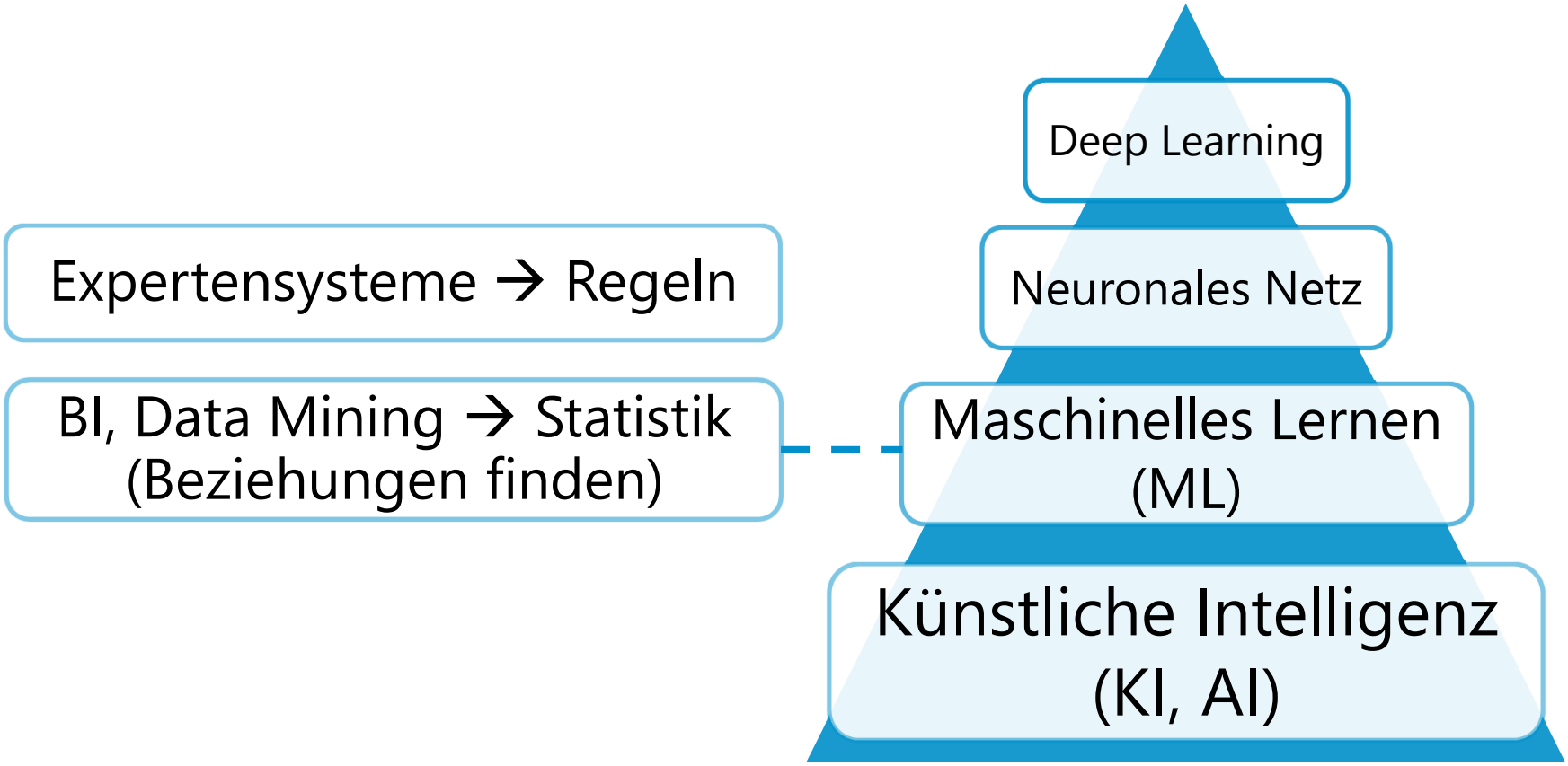
(…um ein bestimmtes Ziel zu erreichen)

…wird i.d.R. durch Training erlernt

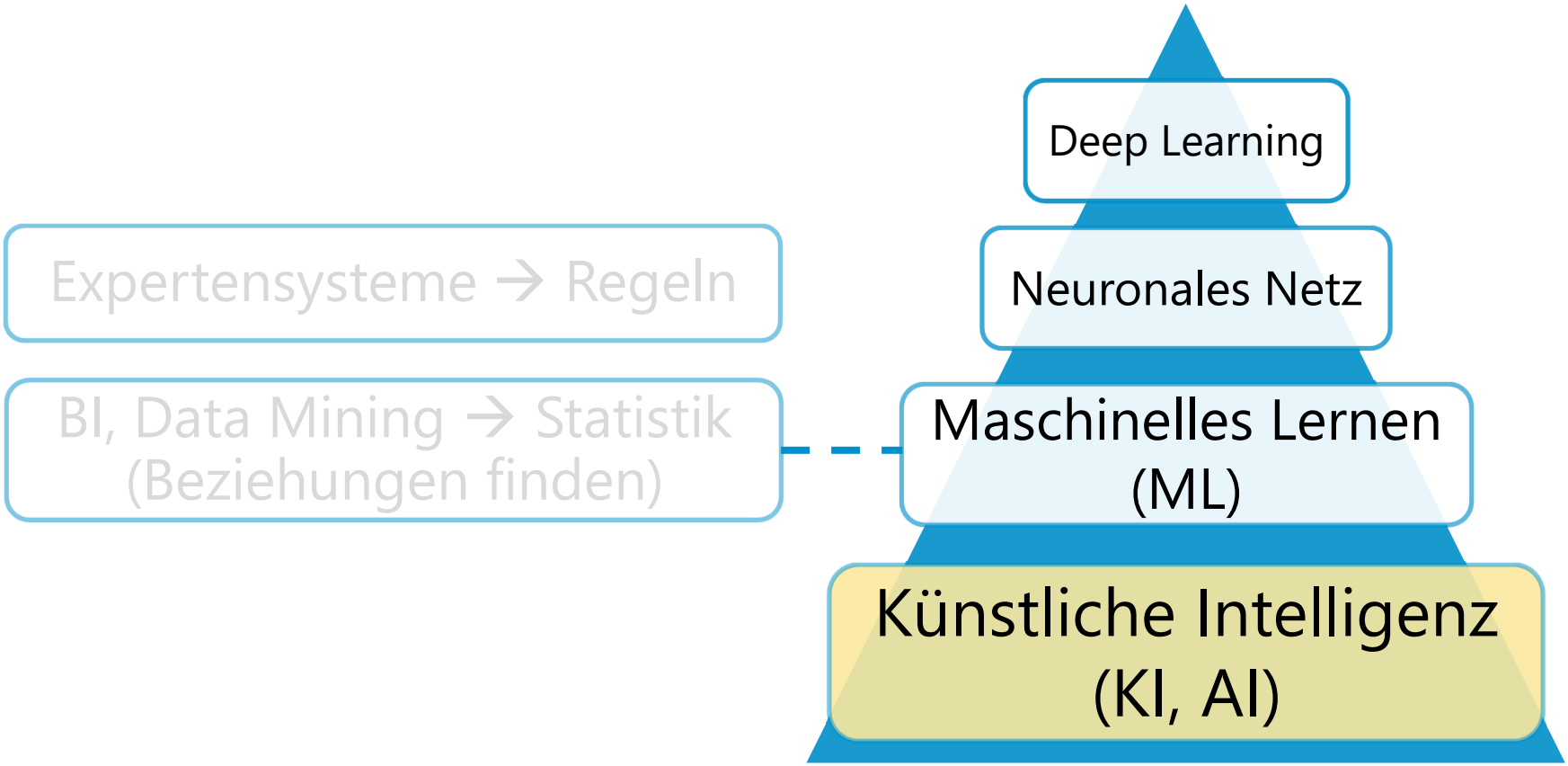
Agenda

- „Wissen“
- **Technische Ansätze**
- Anwendungsfälle
- Datenquellen
- Know-how
- Grenzen

Begriffe


Expertensysteme → Regeln
BI, Data Mining → Statistik
(Beziehungen finden)
Deep Learning
Neuronales Netz
Maschinelles Lernen
(ML)
Künstliche Intelligenz
(KI, AI)

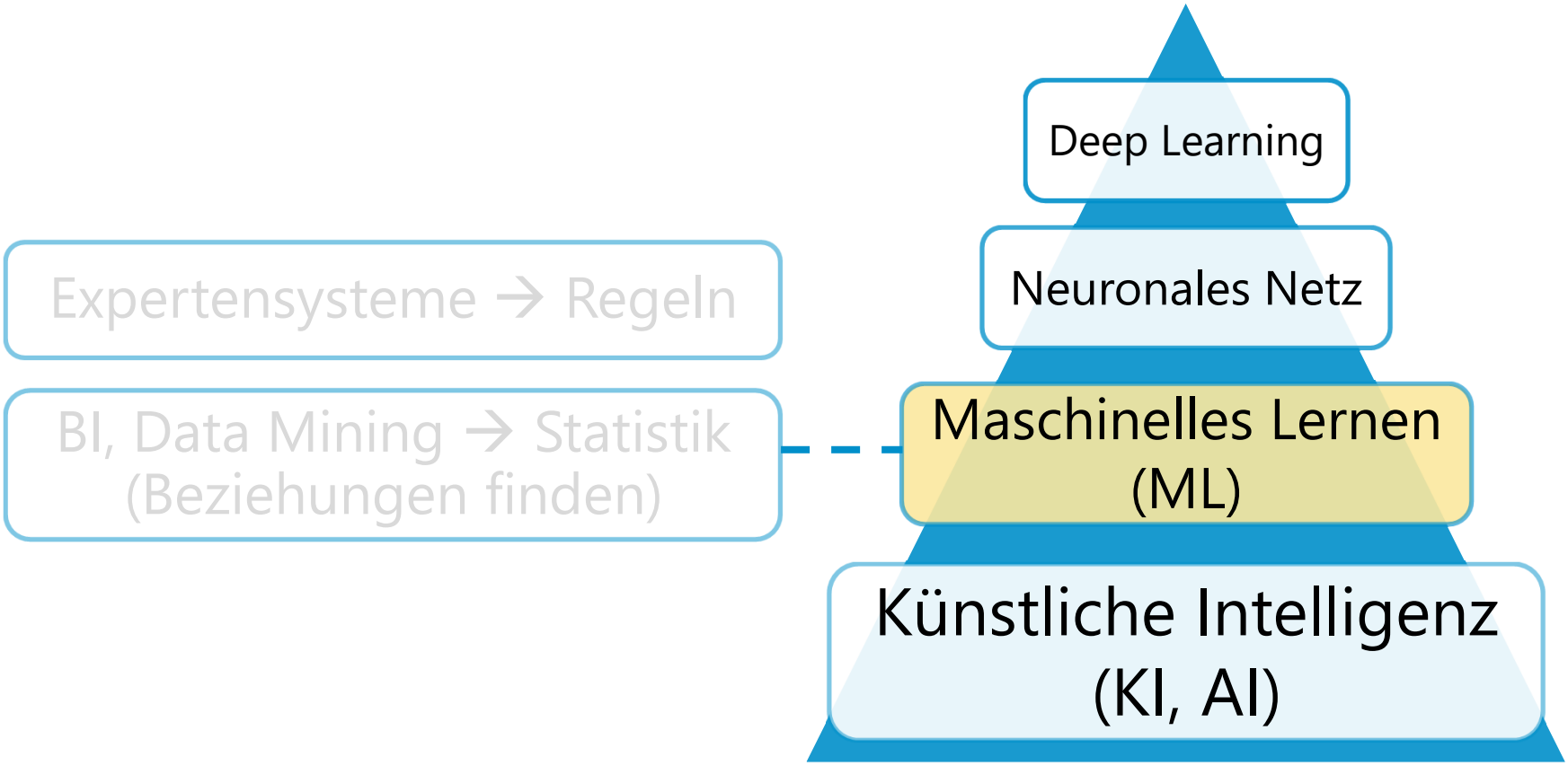
Begriffe


Expertensysteme → Regeln
BI, Data Mining → Statistik
(Beziehungen finden)
Deep Learning
Neuronales Netz
Maschinelles Lernen
(ML)
Künstliche Intelligenz
(KI, AI)

Künstliche Intelligenz

- **Kognitive** Intelligenz (Aufnahme und Erlernen von Wissen, Kombinieren, Schlussfolgern)
- **Sensomotorische** Intelligenz (Wahrnehmung und physische Interaktion)
- **Emotionale** Intelligenz („Mitgefühl“)
- **Soziale** Intelligenz („Teamgeist“)

Begriffe

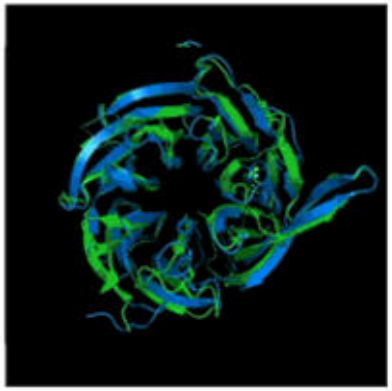

Expertensysteme → Regeln
BI, Data Mining → Statistik (Beziehungen finden)
Deep Learning
Neuronales Netz
Maschinelles Lernen (ML)
Künstliche Intelligenz (KI, AI)

Maschinelles Lernen

Deep Mind:

Alpha Go zerlegt Go-Community

Deep Mind hat die neue Version seiner Go-KI heimlich in eine der besten Go-Communitys eingeschleust. Dort hat die künstliche Intelligenz ohne Niederlage die besten Spieler der Welt besiegt.



Deepmind-KI faltet Proteine

Die KI "Alphafold" der Google-Tochter Deepmind gewinnt mit einigem Abstand den CASP-Wettbewerb zum Vorhersagen der 3D-Struktur komplexer Proteine.

Maschinelles Lernen

- Maschinen, die lernen (durch Training Wissen aneignen)
- Anwendung wissenschaftlich fundierter Methoden, Prozesse, Algorithmen und Systeme

Vorgehen

1. Das Problem verstehen
2. Daten besorgen und normalisieren
3. Algorithmus auswählen
4. Modell trainieren
5. Bewerten
6. Anwenden

Maschinelles Lernen

Algorithmen

- K-Means
- K-Nearest-Neighbors
- Regression (linear, polynomial)
- Regression (logistisch) (Klassifikation!)
- Decision Tree, Random Forest
- Naive Bayes (anteilige Wahrscheinlichkeiten)
- Hidden-Markov-Model
- Hauptkomponentenanalyse
- Support Vector Machine
- Learning Vector Quantization
- Kombinierte Modelle
- Neuronale Netze

Maschinelles Lernen

IKS
Individuelle Softwarelösungen

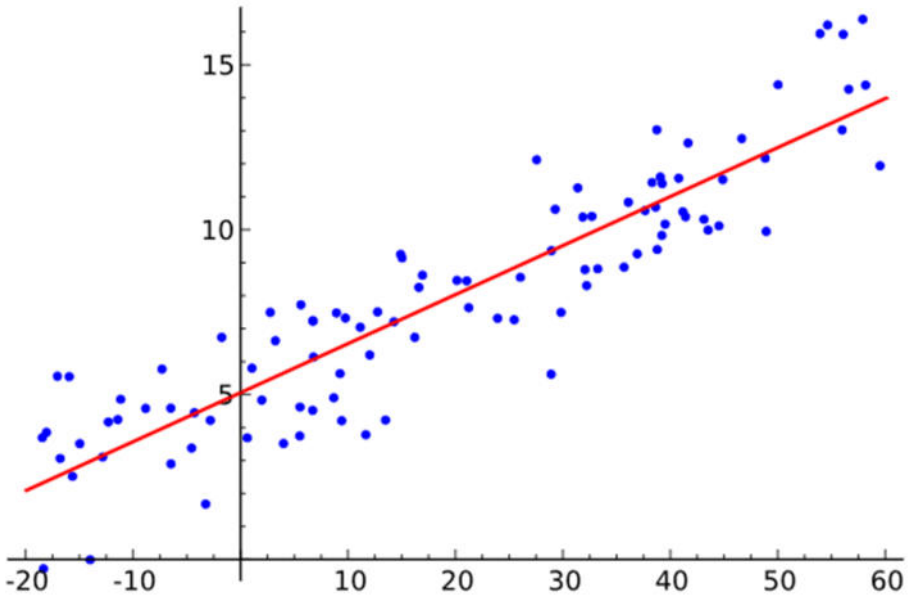
Einteilung der Algorithmen

1) Art des Ergebnisses
   a) Regression: Wie viel(e)?
   b) Klassifikation: Ist dies A oder B (oder C)?
   c) Clustering und Anomalie-Erkennung: Wie gehört das zusammen?
   d) Strategieentscheidung: Was ist zu tun?
   e) Sonderfälle: z.B. wie kann das vereinfacht werden?

2) Trainingsmethode
   a) unbeaufsichtigt: anhand vorgegebener Regeln
   b) beaufsichtigt: aus Beispielen
   c) bestärkend: über Belohnungen

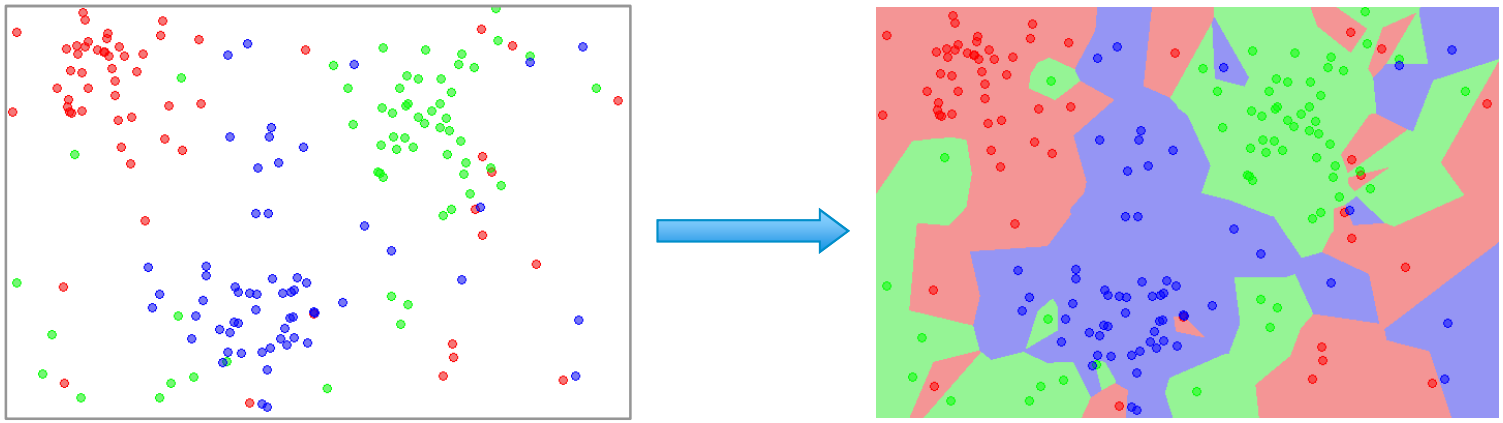
Maschinelles Lernen

Beispiel 1 : Regression (linear)

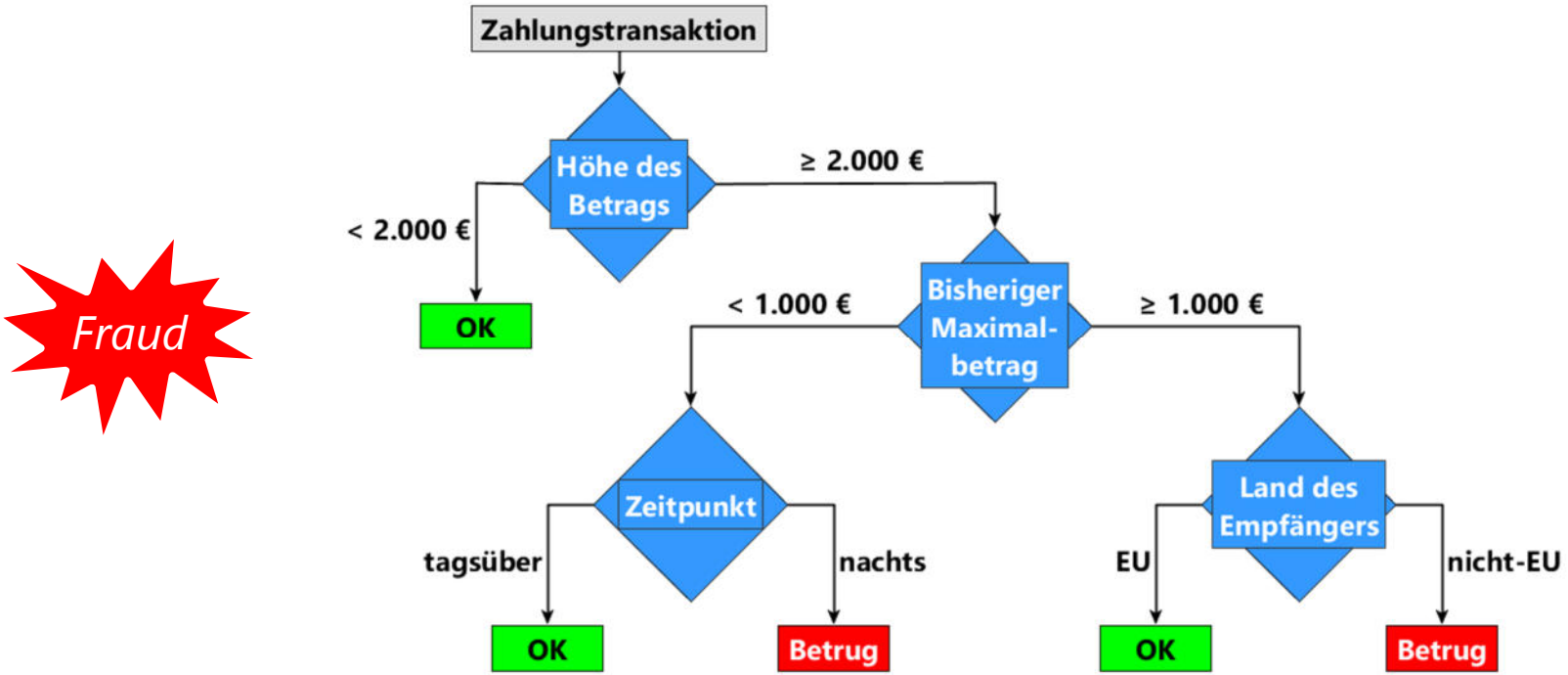


Maschinelles Lernen

Beispiel 2: K-Nearest-Neighbors



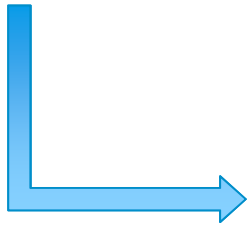
Beispiel 3: Decision Tree

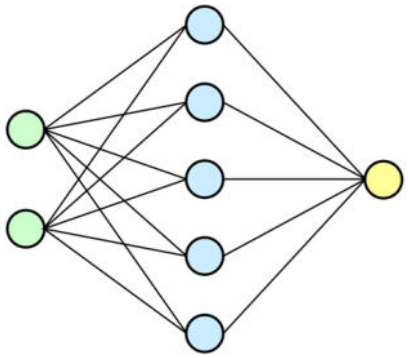


Maschinelles Lernen

Beispiel 4: Neuronales Netz (beaufsichtigt)

Betrag [€]	Bish. Max.-betrag [€]	Zeitpunkt	Empfängerland	Betrug?
45,30	800,00	Di, 12.03.2019 17:04	Deutschland	**Nein**
270,00	800,00	Di, 12.03.2019 17:07	Deutschland	**Nein**
14,55	800,00	Do, 14.03.2019 12:01	Frankreich	**Nein**
2.500,00	800,00	So, 17.03.2019 3:04	Kolumbien	**Ja**
183,22	800,00	Mo, 18.03.2019 9:56	Deutschland	**Nein**

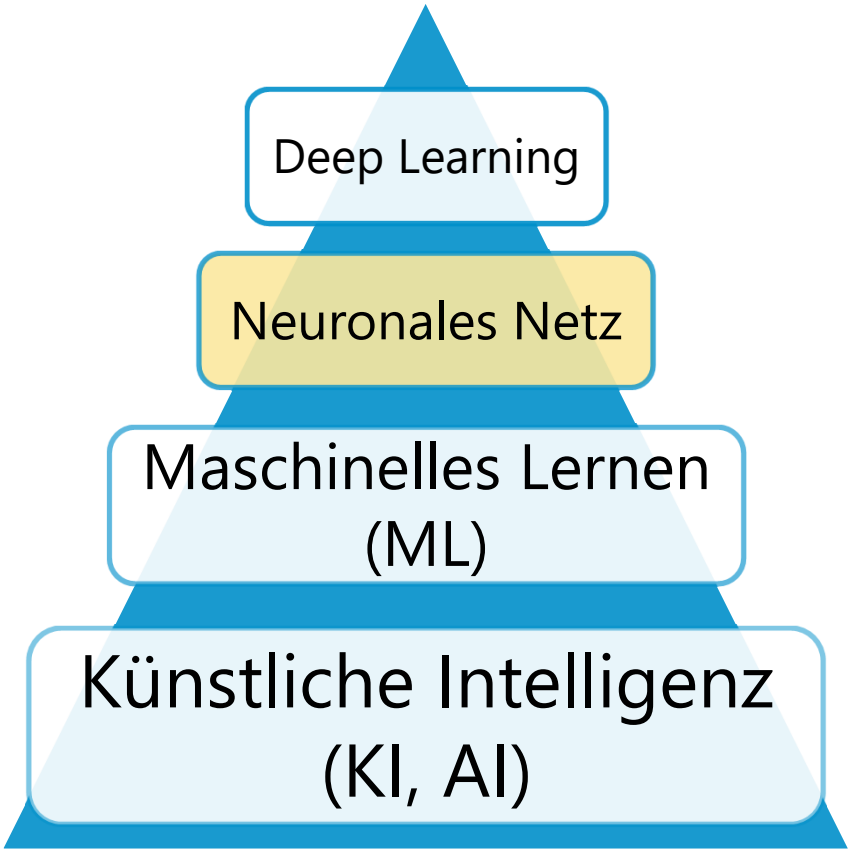




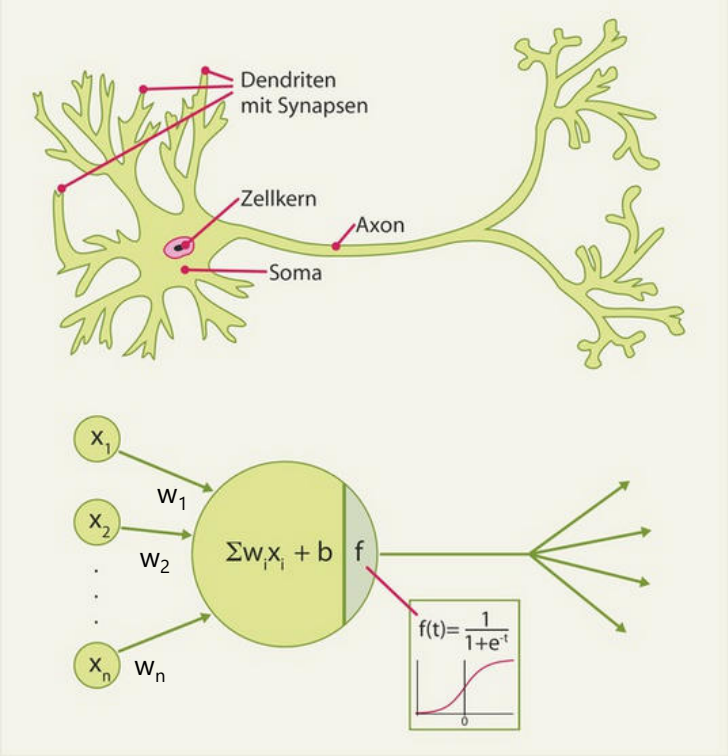
Begriffe

Expertensysteme → Regeln

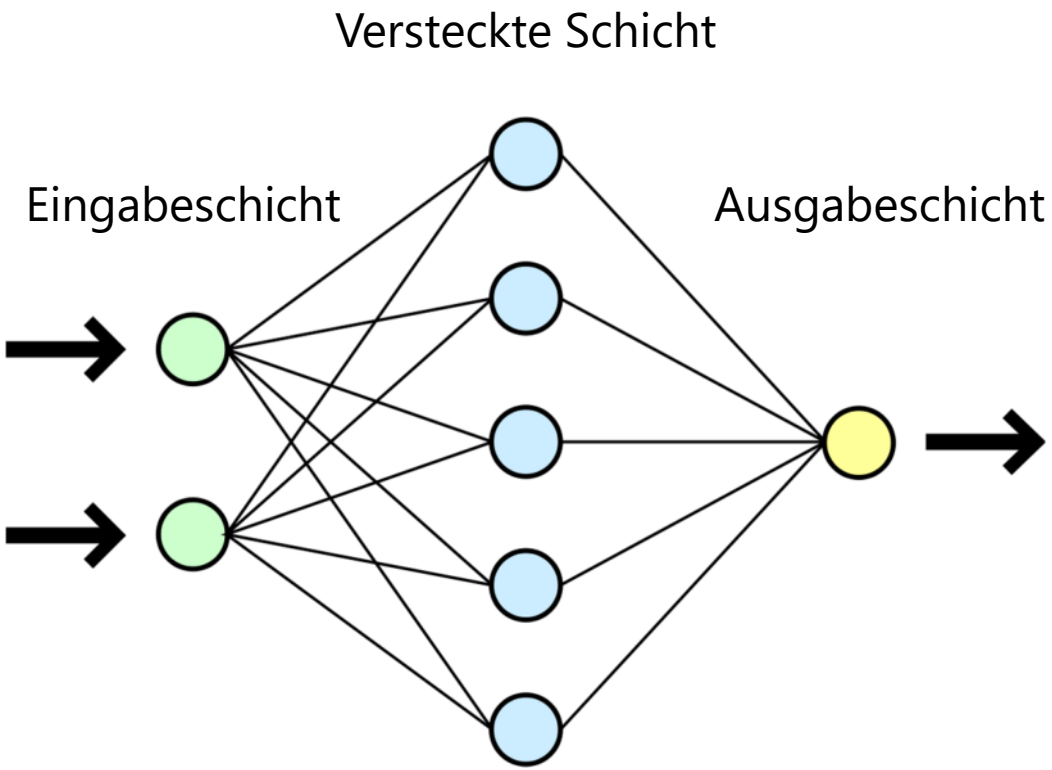
BI, Data Mining → Statistik (Beziehungen finden)



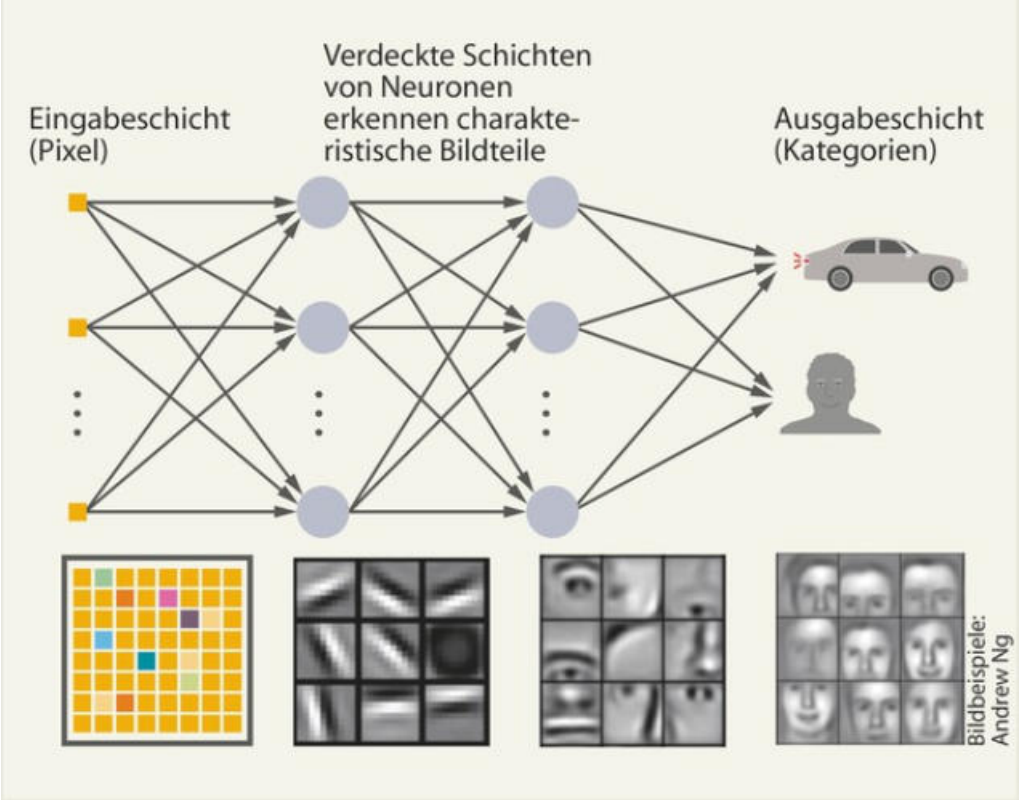
Neuron


Dendriten
mit Synapsen
Zellkern
Axon
Soma
x1
x2
xn
w1
w2
wn
Σwixi + b
f
f(t)= 1/(1+e-t)
0

Neuronales Netz



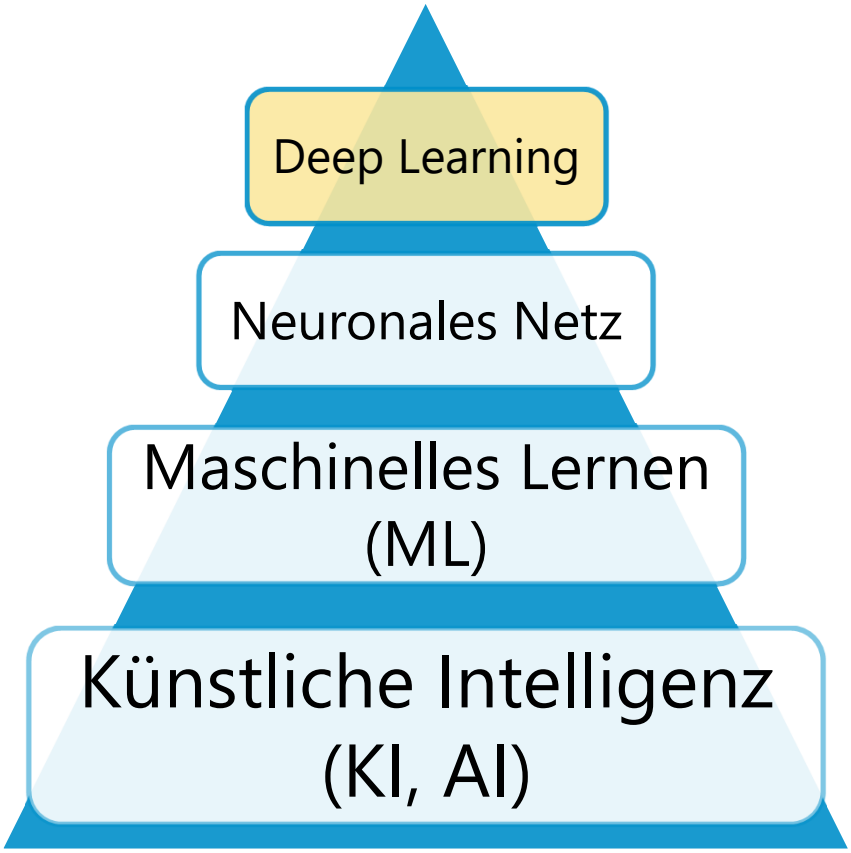
Neuronales Netz



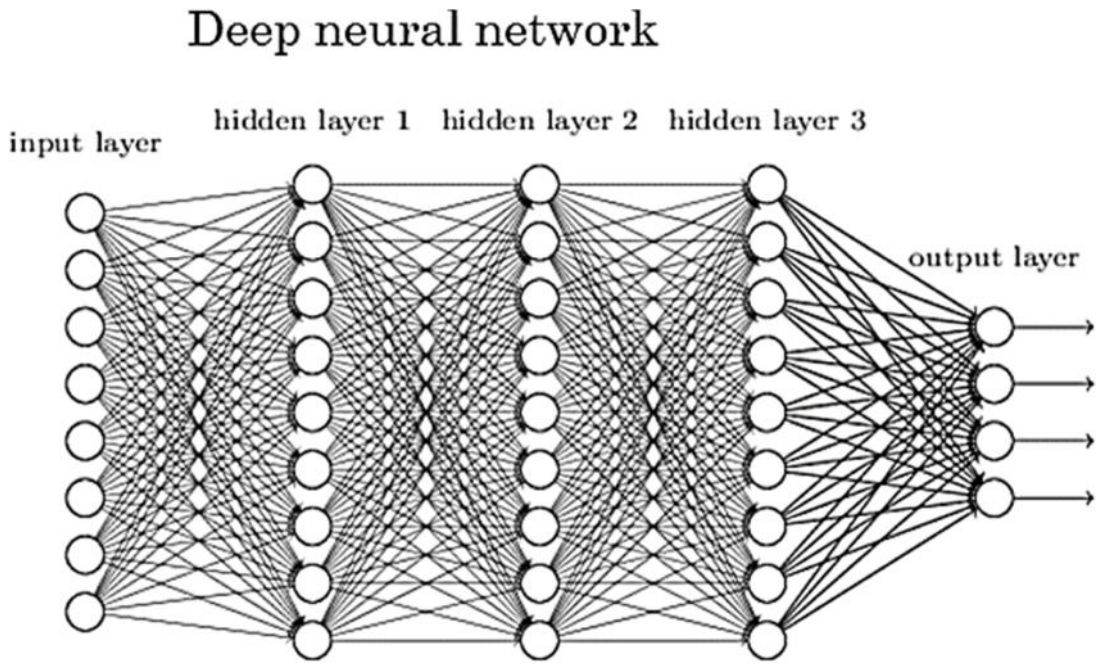
Begriffe

Expertensysteme → Regeln

BI, Data Mining → Statistik (Beziehungen finden)



Deep Learning



Eigenheiten der Algorithmen

Algorithmus	Genauigkeit	Geschw. des Trainings	Datenmenge	Speicherverbrauch	Transparenz
Regression		✓	✓		✓
Decision Tree		✓		✘	✓
Random Forest	✓	✓		✘	
Neuronales Netz	✓	✘	✓		✘

Oft müssen mehrere Algorithmen ausprobiert werden, um den optimalen Algorithmus zu finden.

Agenda

- „Wissen“
- Technische Ansätze
- **Anwendungsfälle**
- Datenquellen
- Know-how
- Grenzen

Einsatz in Unternehmen

16.02.2017

Ein Fünftel der deutschen Unternehmen nutzt maschinelles Lernen bereits produktiv

Anwendungsfälle nach Abteilung

- Unternehmensführung / Marketing / Vertrieb
  - Prognose Unternehmenskennzahlen (z. B. Absatz)
  - Operative Effizienz (Echtzeit-Verkaufszahlen)
  - Prognose Konsumentenverhalten (Kaufverhalten, Reklamationen)
  - Empfehlungsdienste
  - Preisoptimierung / personalisierte Preise

edgecase

HubSpot

- Support
  - Prognose Supportaufkommen
  - Fallkategorisierung
  - Stimmungsanalyse
  - Chatbots

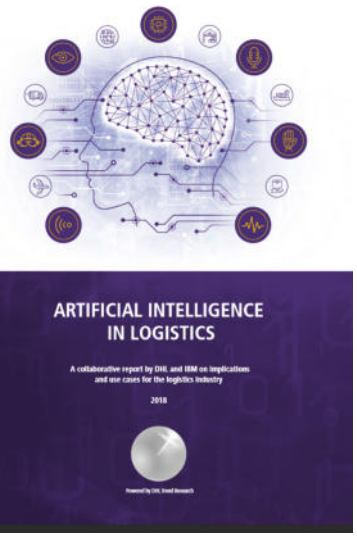
bitext

salesforce

ChatBot

Anwendungsfälle nach Abteilung

- Produktion / Logistik
  - Ausfall- / Wartungsprognose
  - Erkennung struktureller Defekte
  - Optimierung von Transport, Lagerung, Inventur



- IT
  - Archivierung (Active Archive)
  - Malware-Erkennung
  - Spam-Erkennung
  - Erkennung von Cyberangriffen
  - Datenkompression, -vereinfachung
  - Finden von Programmierfehlern

Anwendungsfälle nach Branche

- Banken / Finanzinstitute
  - Betrugsfälle
  - Einschätzung der Kreditwürdigkeit
  - Börsenentwicklung
- Versicherungen
  - Prognose eintretender Versicherungsfälle
  - Individuelle Risikovoraussage (z.B. Fahrstilbewertung)
  - Betrugsfälle

Anwendungsfälle nach Branche

- Verkehr / Umwelt
  - Regelsysteme
  - Frühwarnsysteme
  - Wettervorhersage
  - Bilden von Verhaltensstrategien (Roboter)

- Gesundheit
  - Med. Diagnosen (z.B. Tumorerkennung)
  - Personalisierte Medizin / Medikamente
  - Ausbreitung von Epidemien

Anwendungsfälle allgemein

- Identifikationssysteme
  - Objekterkennung
  - Gesichtserkennung
  - Spracherkennung
  - Handschrifterkennung
- Textverständnis
- Kreative Arbeiten

thispersondoesnotexist.com

Agenda

- „Wissen“
- Technische Ansätze
- Anwendungsfälle
- **Datenquellen**
- Know-how
- Grenzen

Daten selbst beschaffen

- Datenbanken / DWH → Direktzugriff und Reporting-Tools
- Dateisysteme
- Geschäftsanwendungen → API
  - ERP
  - CRM
- Archivsysteme → API
  - elektronische Dokumente (pdf, doc, ppt)
  - gescannte Dokumente
  - E-Mail-Korrespondenz
  - Bilder, Videos

Daten selbst beschaffen

- Nutzungsprotokolle
  - Geräte
  - Programme / Apps
  - Kredit- / Debitkarten
  - Kunden- / Bonuskarten
- Aktivitätsprotokolle (Logs)
  - Software-Anwendungen
  - Netzwerk
  - Datenbanken
- Soziale Medien → API
- Webseiten → Web-Scraping

Daten selbst beschaffen

- Sensoren (Kamera, Mikrofon, Temperatur, Beschleunigung, Luftdruck, Vibration, Stromstärke, Spannung, …)
  - Mobilgeräte
  - Wearables
  - Messgeräte
  - Überwachungssysteme
  - Smart Home
  - Fahrzeuge
  - Produktionsketten
  - GPS / Lokalisierungsdienste

Internet of Things

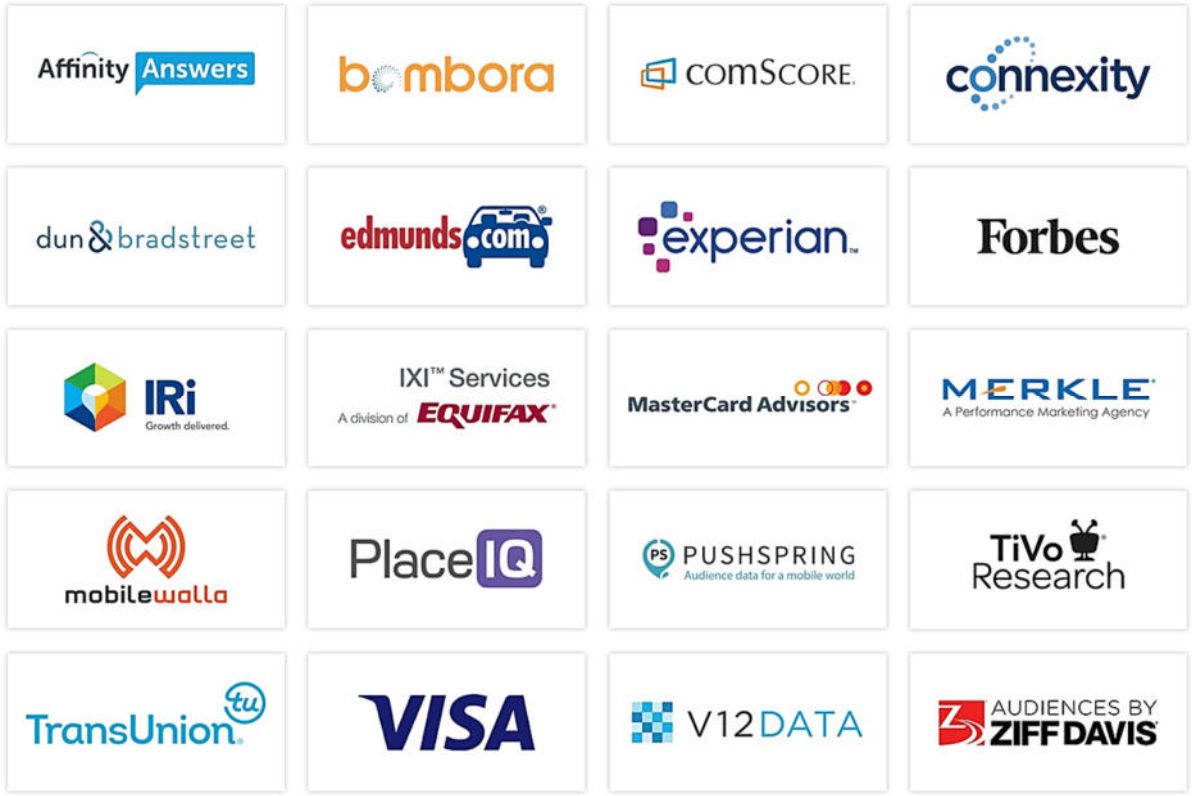
Daten von Drittanbietern

- Betreiber sozialer Medien
- Statistische Ämter des Bundes und der Länder, OECD
- Meinungsforschungsinstitute
- Börsen
- Mozilla → Sprachdatensammlung „Common Voice“
- Öffentliche Trainingsdaten

Daten von Drittanbietern

„Data as a Service“ (DaaS)

Oracle DaaS (Auswahl)



Agenda

- „Wissen“
- Technische Ansätze
- Anwendungsfälle
- Datenquellen
- **Know-how**
- Grenzen

Know-how & Werkzeuge

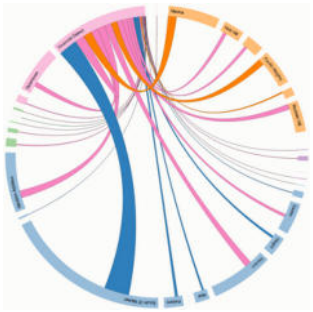
- Generelles Know-how
  - Informatik, Statistik
  - Machine Learning: Prinzipien, Algorithmen
  - Unstrukturierte Daten: Video, Audio, Texte, …
- Programmiersprachen
  - R
  - Python
  - SQL
  - (Java, C/C++, C#)
- Big-Data-Infrastruktur
- Cloud

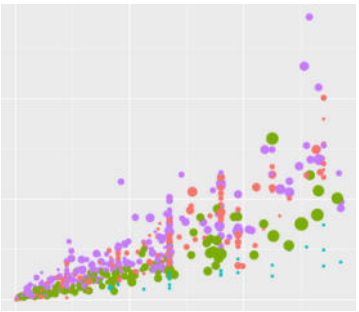
Know-how & Werkzeuge

- Entwickler-Werkzeuge
  - GNU R
  - Caffe
  - Deeplearning4j
  - Auto-Keras
  - TensorFlow
  - PyTorch
  - Microsoft Cognitive Toolkit
  - ML.NET
  - Chainer
  - Theano
  - MATLAB
  - …

en.wikipedia.org/wiki/Comparison_of_deep_learning_software

- Datenvisualisierung
  - Ggplot
  - d3.js
  - Matplottlib
  - Tableau
  - …





Agenda

- „Wissen“
- Technische Ansätze
- Anwendungsfälle
- Datenquellen
- Know-how
- **Grenzen**

Grenzen des maschinellen Lernens

IKS Individuelle Softwarelösungen

- Abhängig von Zugang und Qualität von Daten
  - Unvollständige Daten (fehlende Werte, falsches Datenformat)
  - Ungenaue Daten (reelle Zahlen)
  - Falsche Daten (z.B. Messfehler)
  - Nicht-repräsentative Daten
  - Unausgewogene Daten

- Durchführung nicht immer trivial
  - Trainingsdaten müssen manuell vorbereitet werden (z.B. Bereinigung)
  - Verschiedene Algorithmen möglich
  - Qualität muss bewertet werden
  - Dedizierte Tools

Grenzen des maschinellen Lernens

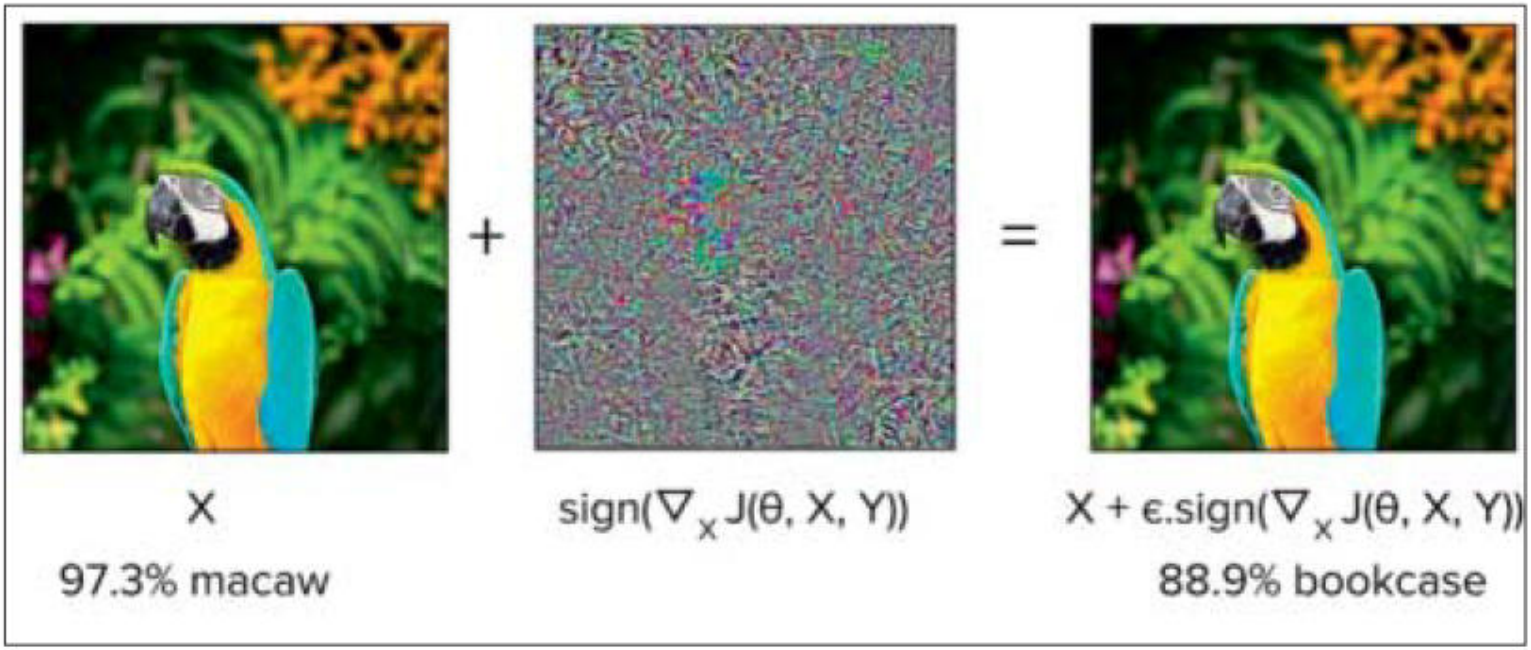
- Liefert nicht immer gute Ergebnisse
  - Interpolation gut, Extrapolation eher nicht
  - Alle relevanten Aspekte müssen in den Trainingsdaten vorliegen
  - Overfitting, underfitting
- Entscheidungsprozesse in tiefen neuronalen Netzen nicht überschaubar
  - Unklar, ob globales Optimum gefunden wurde
  - Möglichkeiten und Limitierungen nicht ersichtlich
  - Entscheidungskriterien können nicht offengelegt werden → rechtliche Probleme?
- Potentiell ressourcenintensiv

Grenzen des maschinellen Lernens

- Face-Swap

Grenzen des maschinellen Lernens

- Subtile Manipulation



Grenzen des maschinellen Lernens

IKS
Individuelle Softwarelösungen

Lösungsansätze

- Klein anfangen („ausreichend viele Daten“)
- Wenn nötig, skalierbare Ressourcen einsetzen
- Möglichst vollständige, repräsentative Daten nutzen
- Know-how im Bereich Data Science aufbauen

IKS
Individuelle Softwarelösungen

Vielen Dank!

Fragen?

Weiterführende Literatur

- Fuchs-Kittowski K: Wissens-Ko-Produktion – Organisationsinformatik
  http://www.wissenschaftsforschung.de/JB00_9-88.pdf

- Aamodt A, Nygård M (1995): Different roles and mutual dependencies of data, information, and knowledge – An AI perspective on their integration
  Data and Knowledge Engineering, Elsevier, Bd. 16, Nr. 3, S. 191ff.

- Schulz HJ, Nocke T: Maschinelle Datenanalyse im Informationszeitalter – Können oder müssen wir ihr vertrauen?
  https://www.pik-potsdam.de/members/nocke/.personal/Dispositiv05.pdf

Referenzen

- A Tour of The Top 10 Algorithms for Machine Learning Newbies
  https://towardsdatascience.com/a-tour-of-the-top-10-algorithms-for-machine-learning-newbies-dde4edffae11?gi=895089bb957e
- The 10 Algorithms Machine Learning Engineers Need to Know
  https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html
- Essentials of Machine Learning Algorithms
  https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
- Winfwiki: Überblick Ansätze des Deep Learning
  http://winfwiki.wi-fom.de/index.php/%C3%9Cberblick_Ans%C3%A4tze_des_Deep_Learning
- Netzgespinste
  http://www.heise.de/ct/ausgabe/2016-6-Die-Mathematik-neuronaler-Netze-einfache-Mechanismen-komplexe-Konstruktion-3120565.html
- Machine learning
  https://en.wikipedia.org/wiki/Machine_learning
- Trainingsdaten
  https://en.wikipedia.org/wiki/List_of_datasets_for_machine_learning_research
- Detecting Financial Fraud Using Machine Learning
  https://towardsdatascience.com/detecting-financial-fraud-using-machine-learning-three-ways-of-winning-the-war-against-imbalanced-a03f8815cce9
- Mozilla will mit Machine-Learning Coding-Fehler finden
  https://www.golem.de/news/entwicklung-mozilla-will-mit-machine-learning-coding-fehler-finden-1902-139372.html
- Artificial Intelligence in Logistics, DHL
  https://www.logistics.dhl/content/dam/dhl/global/core/documents/pdf/glo-ai-in-logistics-white-paper.pdf

IKS
Individuelle Softwarelösungen

WWW.IKS-GMBH.COM

Exkurs Maschinelles Lernen und Data Mining

Machine Learning und Data Mining verwenden oft die gleichen Methoden und überschneiden sich signifikant

Machine Learning

- Vorhersage/Prognose, basierend auf bekannten Eigenschaften, die aus den Trainingsdaten gelernt wurden
- Fähigkeit, bekanntes Wissen zu reproduzieren
- setzt auch Data Mining-Methoden als "unbeaufsichtigtes Lernen" oder als Vorverarbeitungsschritt zur Verbesserung der Lerngenauigkeit ein

Data Mining

- Explorative Suche = Entdeckung von (bisher) unbekannten Wissens in den Daten
- verwendet auch Lernmethoden des Machine Learning

Exkurs Maschinelles Lernen und Data Mining

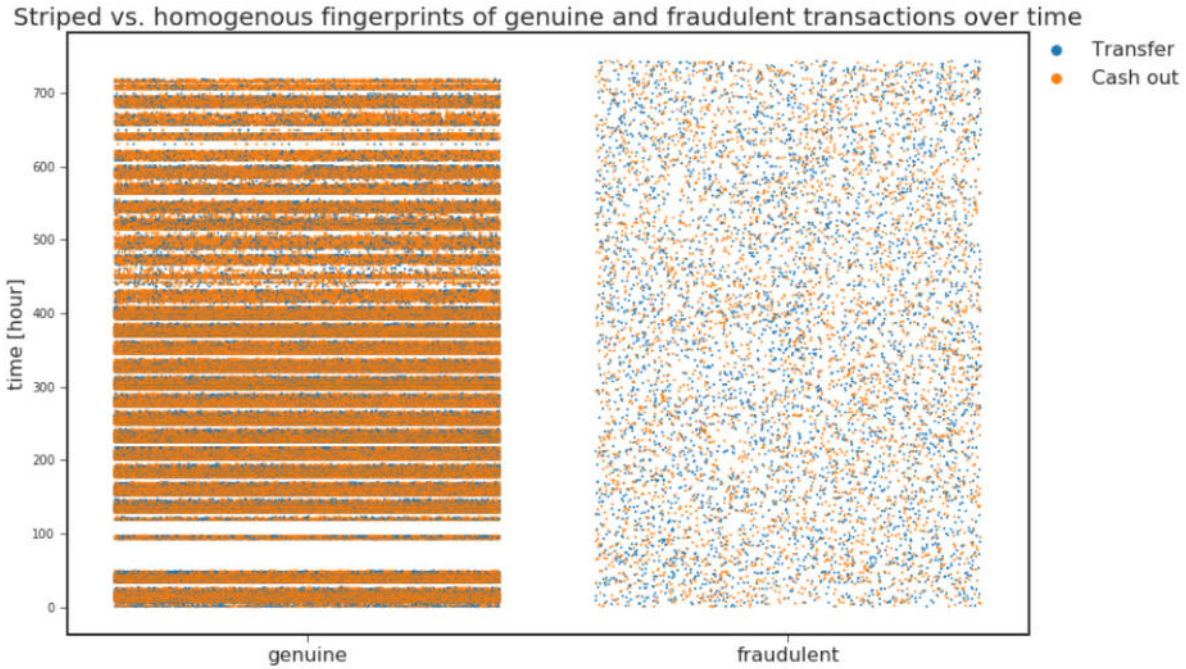
IKS Individuelle Softwarelösungen

Typische Aufgabenstellungen des Data-Mining

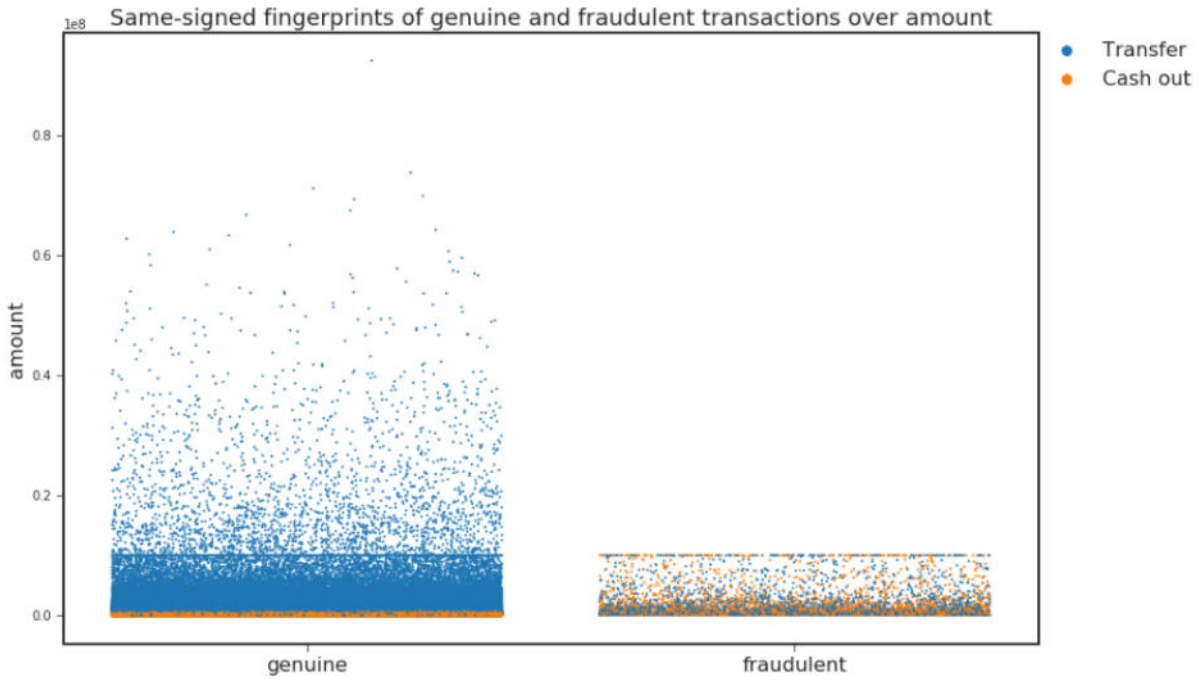
- Identifizierung von ungewöhnlichen Daten (Ausreißer, Fehler, Änderungen)
- Clusteranalyse: Gruppierung von Objekten aufgrund von Ähnlichkeiten
- Klassifikation: bisher nicht Klassen zugeordnete Elemente werden bestehenden Klassen zugeordnet
- Assoziationsanalyse: Identifizierung von Zusammenhängen und Abhängigkeiten in den Daten in Form von Regeln wie „Aus A und B folgt normalerweise C“.
- Regressionsanalyse: Identifizierung von Beziehungen zwischen (mehreren) abhängigen und unabhängigen Variablen
- Zusammenfassung: Reduktion des Datensatzes in eine kompaktere Beschreibung ohne wesentlichen Informationsverlust

Beispiel: Betrugserkennung

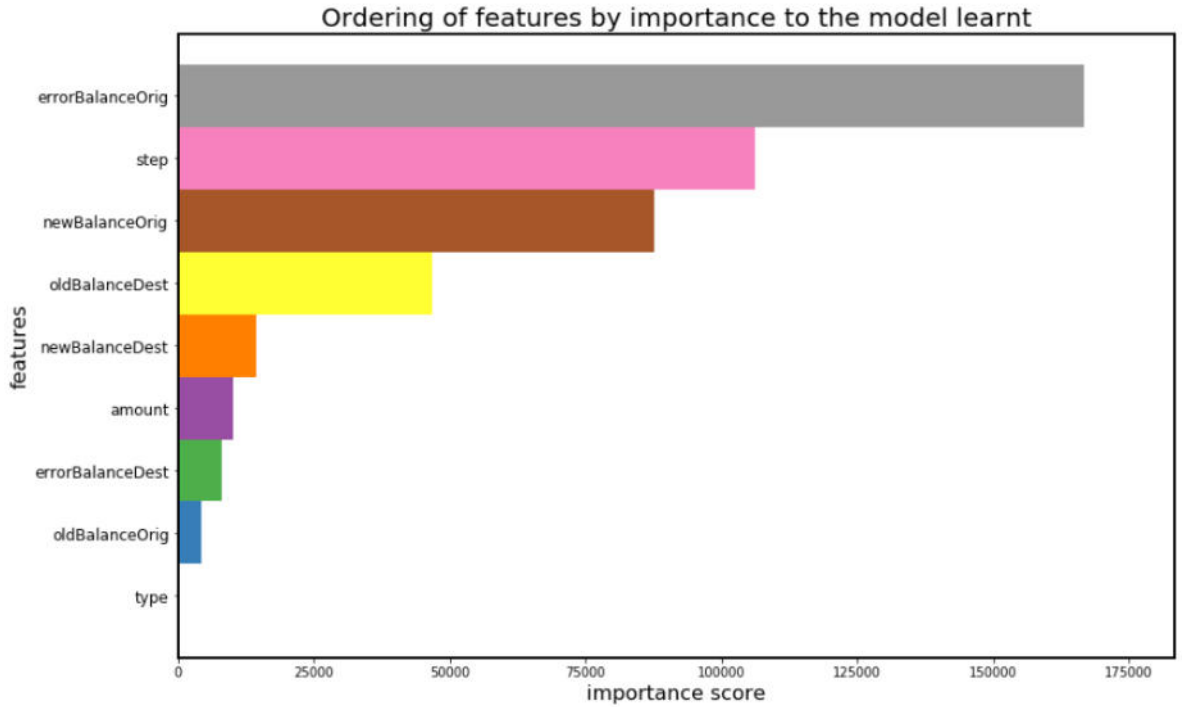
- Datananalyse



Beispiel: Betrugserkennung

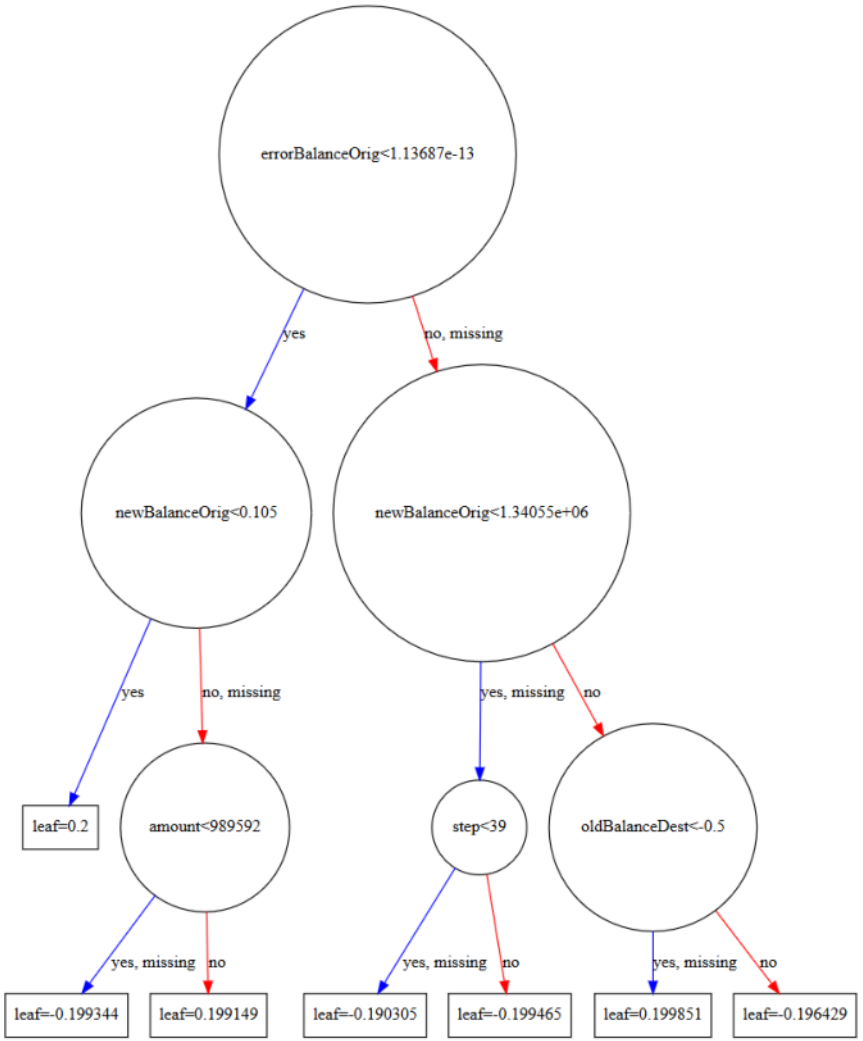


Beispiel: Betrugserkennung



Beispiel: Betrugserkennung

- Anwendung: Random Forest (geeignet für große Mengen- unterschiede der beiden Klassen)



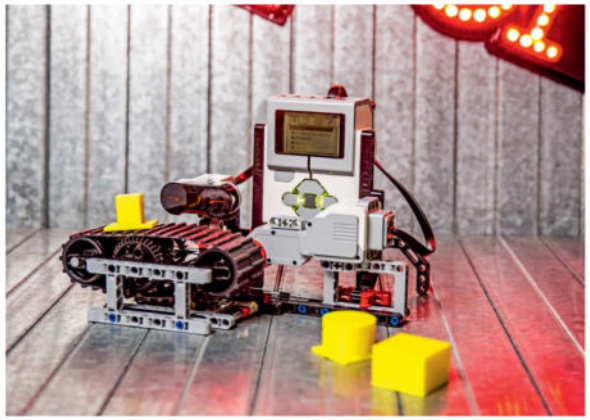
Soziale Implikationen

- 4 Todesfälle durch selbstfahrende PKW
- Risikobewertung für kriminelles Verhalten → voreingenommen gegenüber Schwarzen
- Google: Bilderkennung von Schwarzen als Gorillas
- Microsoft: Chatbot, der von Twitter gelernt hatte, nahm schnell rassistische und sexistische Sprache auf
- Befürchtung: Ablehnung von Bewerbern
- Befürchtung: Med. Diagnose steht hinter Profit zurück

Maschinelles Lernen

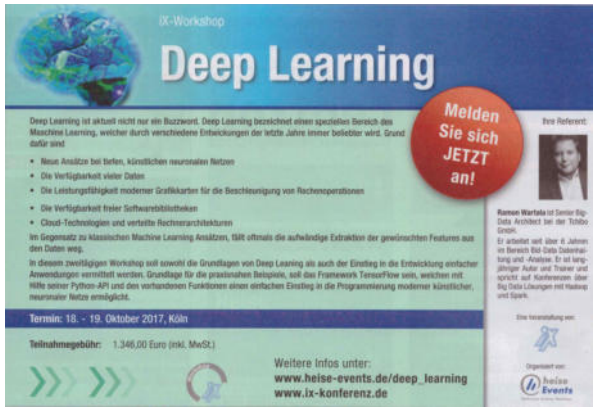
Matlab: Deep Learning mit der Grafikkarte

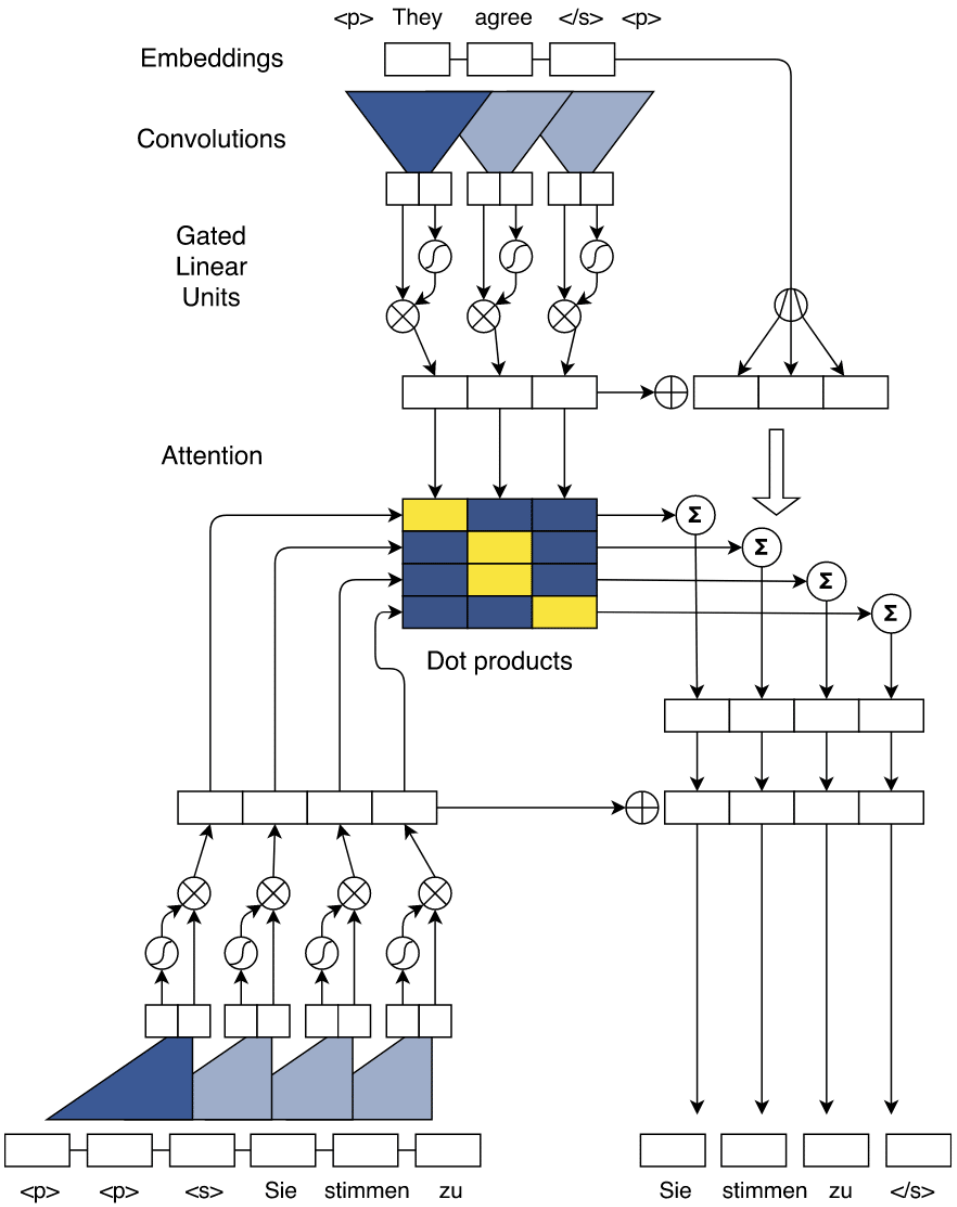
Anhand von Trainingsbildern klassifiziert Matlab ganze Bilder und erkennt, was in einzelnen Bereichen zu sehen ist.



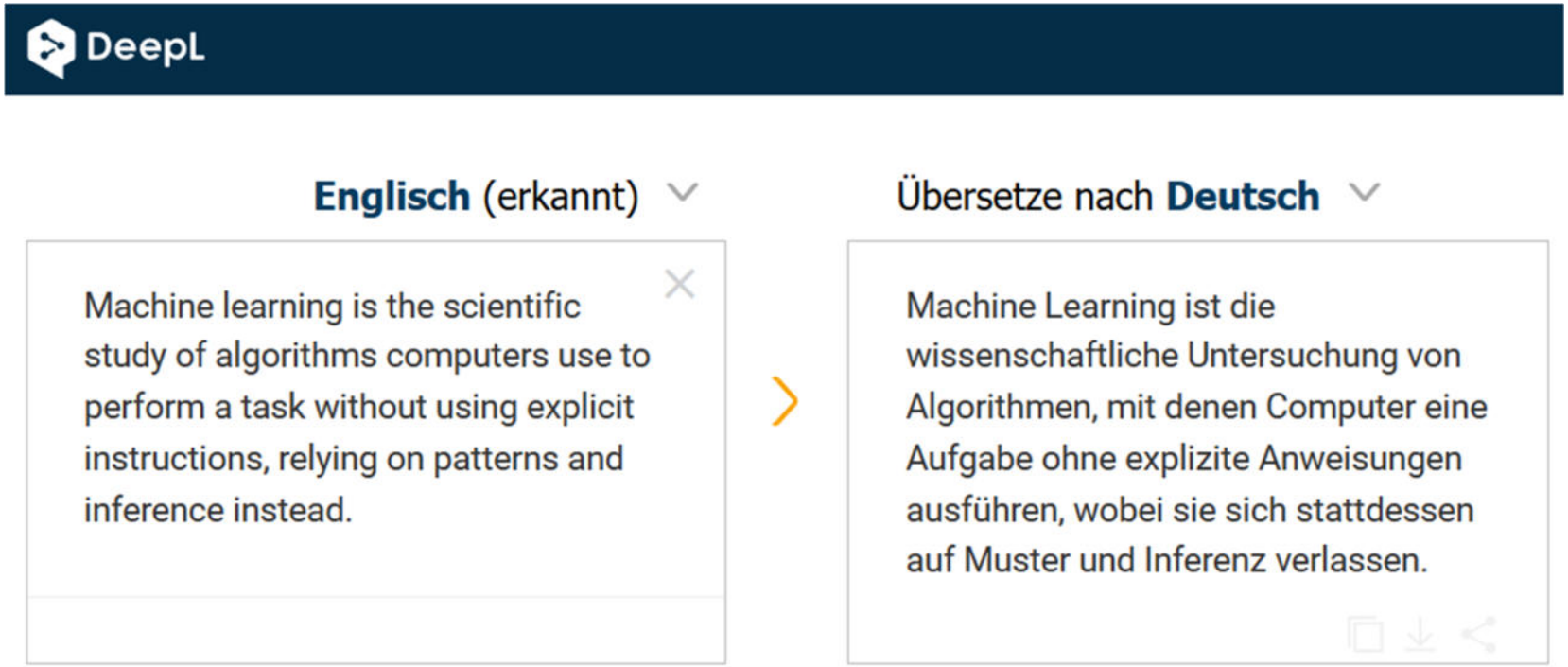
KI im Infrarotlichtbezirk

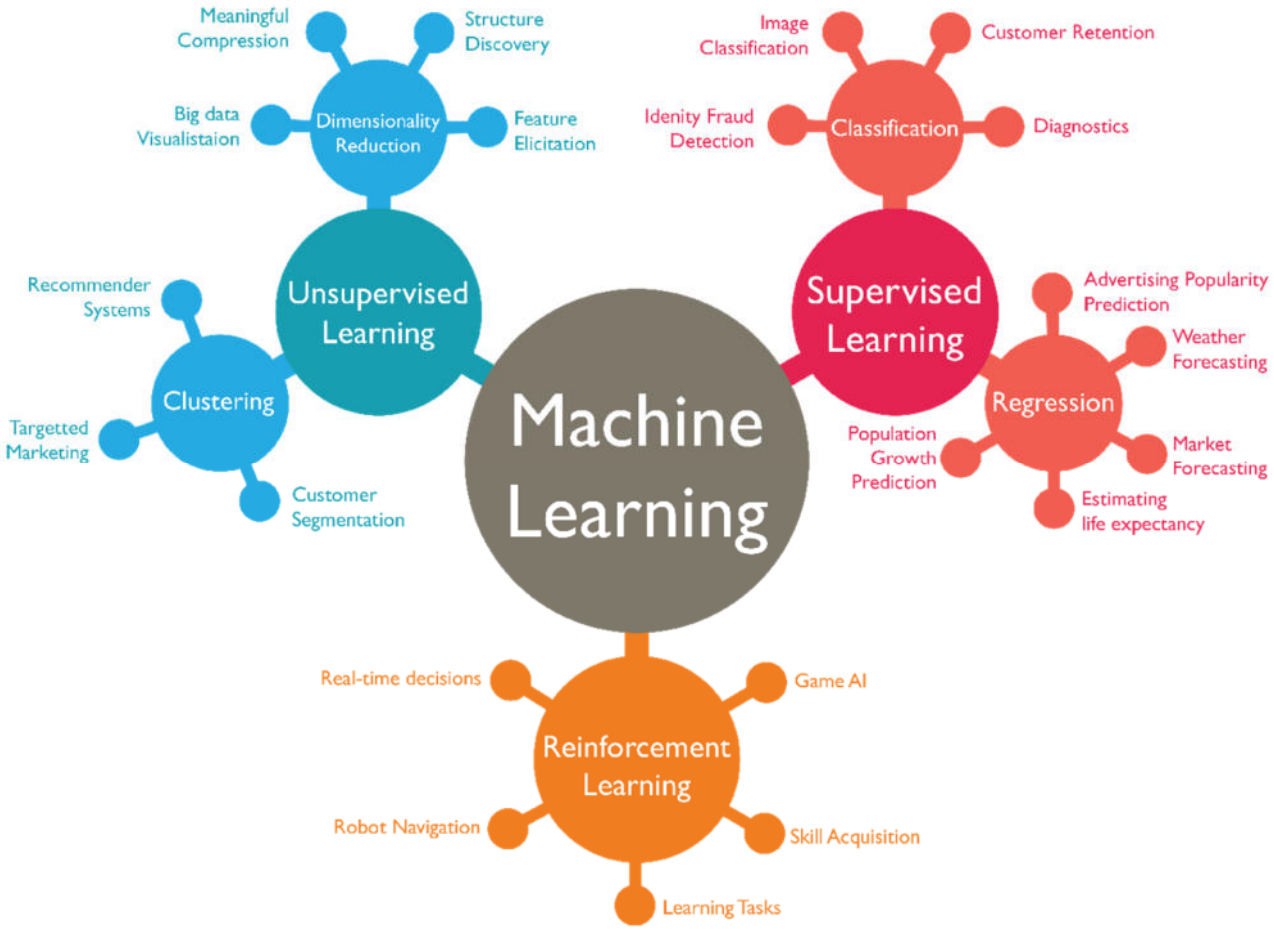
Einstieg ins maschinelle Lernen mit dem Roboterbaukasten LEGO Mindstorms EV3




<p> They agree </s> <p>
Embeddings
Convolutions
Gated
Linear
Units
Attention
Dot products
<p> <p> <s> Sie stimmen zu
Sie stimmen zu </s>

Deep Learning




Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating life expectancy
Population Growth Prediction
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks