

Machine Learning Chancen und Herausforderungen

Prof. Dr. Stephan Trahasch

<https://imla.hs-offenburg.de>



INSTITUTE FOR MACHINE
LEARNING AND ANALYTICS

 Hochschule Offenburg
offenburg.university

4400+ Studierende; 12 % International Studierende
45 Studiengänge Bachelor und Master

Fakultäten

1. Betriebswirtschaft und Wirtschaftsingenieurwesen
2. Elektrotechnik, Medizintechnik und Informatik
3. Maschinenbau und Verfahrenstechnik
4. Medien und Informationswesen



Die Hochschule Offenburg zählt zu den forschungstärksten HAWs in Baden-Württemberg.



Institut für Energiesystemtechnik



Peter Osypka Institut



Institute for Unmanned Aerial Systems



Affective and Cognitive Institute



Institut für verlässliche
Embedded Systems und
Kommunikationselektronik



Institute for Trade and Innovation



K. Dorer

Deep Learning, Autonome Systeme

R. Gasper

Grey Box Modeling

H. Grabowski

Smart Mobile Systems

T. Hagen

Business Analytics, Data Warehousing

S. Hensel

Maschinelles Sehen,
mobile autonome Systeme

J. Keuper

Large Scale Machine Learning,
Deep Learning



T. Lauer

Parallel Computing, Data Clustering

G. Meier

Ethics (external member)

J. Münchenberg

Business Analytics, Digitalisierung

J. Reiter

Data Mining, Predictive Analytics

V. Sängler

Data Mining, Web Mining, Learning Analytics,
Multimedia Datenbanken

S. Trahasch

Machine Learning, Big Data

- Predictive Maintenance (Carl-Zeiss-Stiftung)
in Zusammenarbeit mit zwei Industriepartnern
- Menschen Lernen Maschinelles Lernen (BMBF)
Weiterbildung für Unternehmensmitarbeiter
- Abverkaufsanalysen
mit Apache Hadoop, Spark und SAP HANA
- Production Intelligence (BMW I)
Datengetriebene Echtzeitanalyse automatisierter Produktion
- HyPar
Hybride CPU/GPU-Parallelisierung zur Beschleunigung von Clustering-Verfahren
- Vizeweltmeister RoboCup - Sweaty und Magma



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



Bundesministerium
für Wirtschaft
und Energie



- Data Mining / Advanced Analytics als Modul in vielen Masterstudiengängen
- Vertiefung Data Science und Analytics im Informatik-Master
- NEU
Ab WS20/21 Bachelor-Studiengang **Angewandte Künstliche Intelligenz**





Machine Learning 101

Trends

Herausforderungen



Wenn die Roboter kommen

ZEIT ONLINE - 04.09.2019

Wenn wir über den Einfluss **Künstlicher Intelligenz** auf die Ökonomie und die Arbeit sprechen, denken wir an die Zukunft, egal ob in drei, zehn ...



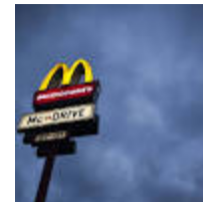
Zeit und Geld sparen mit **Künstlicher Intelligenz**

Tagesspiegel - 05.09.2019

Die Hersteller entwickeln Haushaltsgeräte wie Kühlschränke, Backöfen oder Waschmaschinen mit **Künstlicher Intelligenz** (KI) weiter um die ...

Wenn die Schwarmintelligenz den Braten im Ofen steuert

rbb|24 - 05.09.2019



McDonald's will Drive-In mit **künstlicher Intelligenz** verbessern

WEB.DE News - 13.09.2019

Nun will auch die Fast-Food-Kette McDonald's die eigene Effizienz steigern – und den Drive-In-Schalter mit einer **künstlichen Intelligenz** ...

McDonald's: Sprechen wir bald mit **künstlicher Intelligenz**?

eKitchen - 12.09.2019



Wie **Künstliche Intelligenz** den Recruiting-Prozess intelligenter ...

IT-BUSINESS - 09.09.2019

Künstliche Intelligenz dringt in immer mehr Bereiche des Berufslebens vor und kann auch beim Recruiting eine Vielzahl an Vorteilen bieten, ...



Digital Forecast und Dynamic Pricing als Schwerpunkte von ...

Haufe - News & Fachwissen - 11.09.2019

Was nützen Big Data und **Machine Learning** bei der Unternehmenssteuerung? Controller profitieren von genauerer Planung und ...



Bei Yoono designt KI die Mode

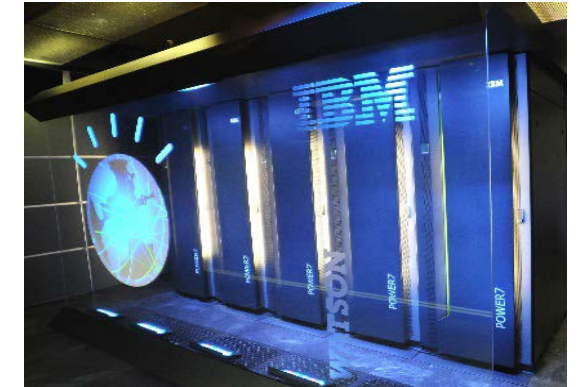
CIO - 05.09.2019

Yoono bietet individuelles On-Demand-Modedesign mithilfe von neuronalen Netzen und **Machine Learning**. Dabei sollen Prozesse verkürzt ...

KI in Filmen



KI in der Realität



Match	Black	White	Result
1	Lee Sedol	AlphaGo	W + Res
2	AlphaGo	Lee Sedol	B + Res
3			
4			
5			

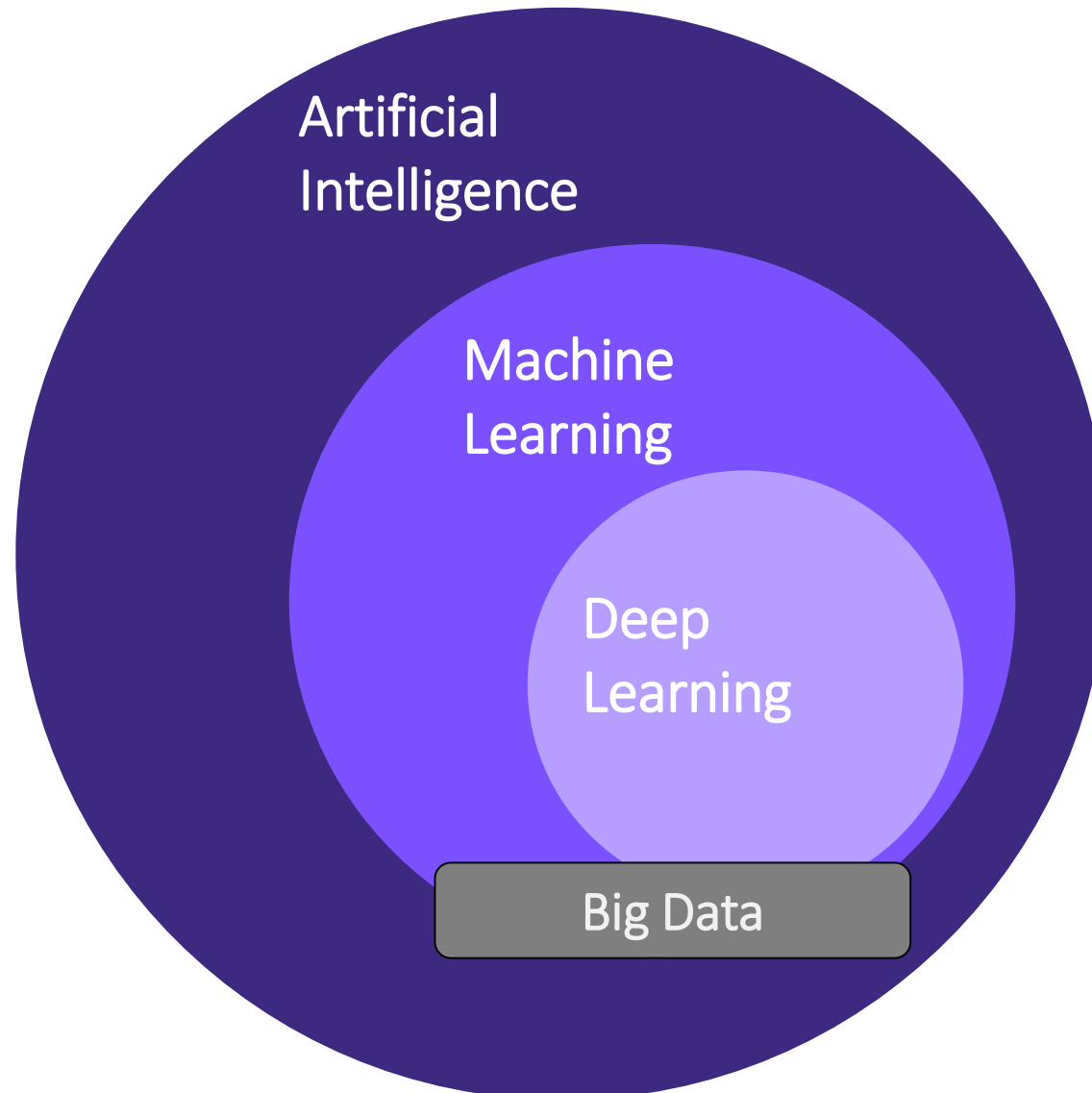
FINAL SCORES  Google DeepMind
Challenge Match
8 - 15 March 2016

LEE SEDOL
00:00:57

ALPHAGO
00:01:00



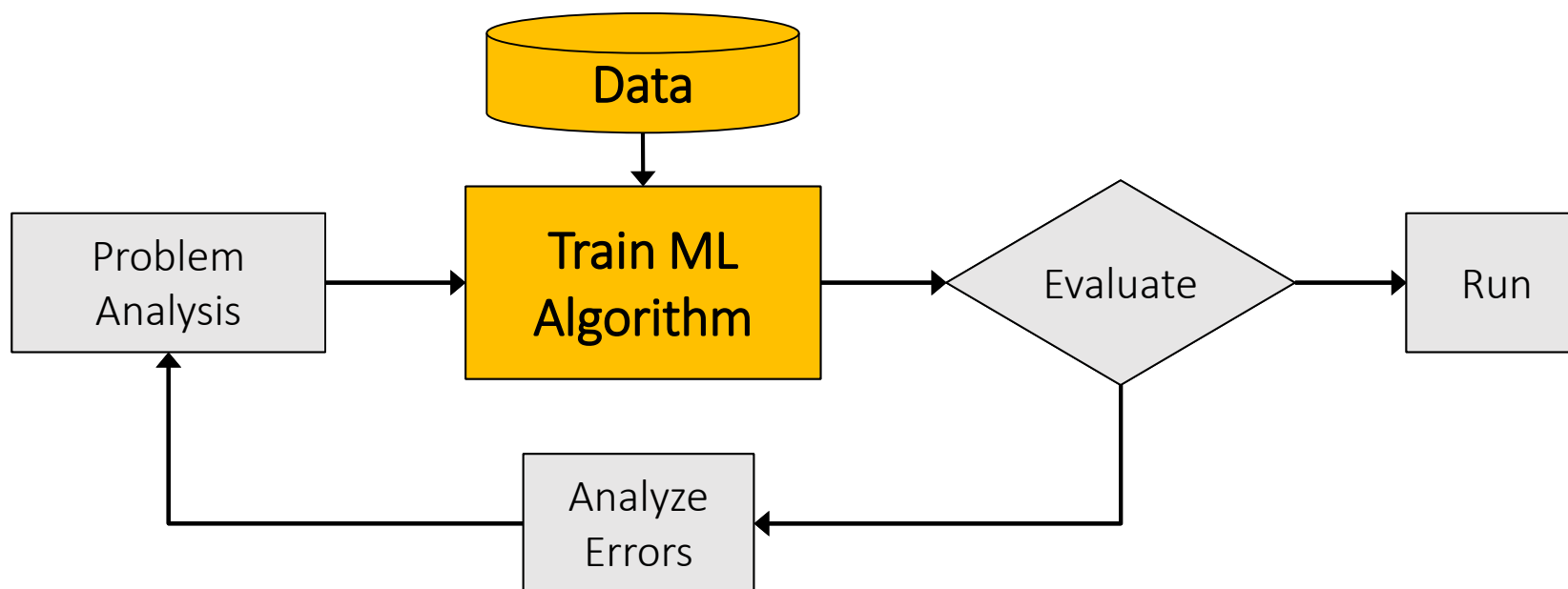
KI ist ein Oberbegriff, der
mehr als Machine Learning und Deep Learning
umfasst.





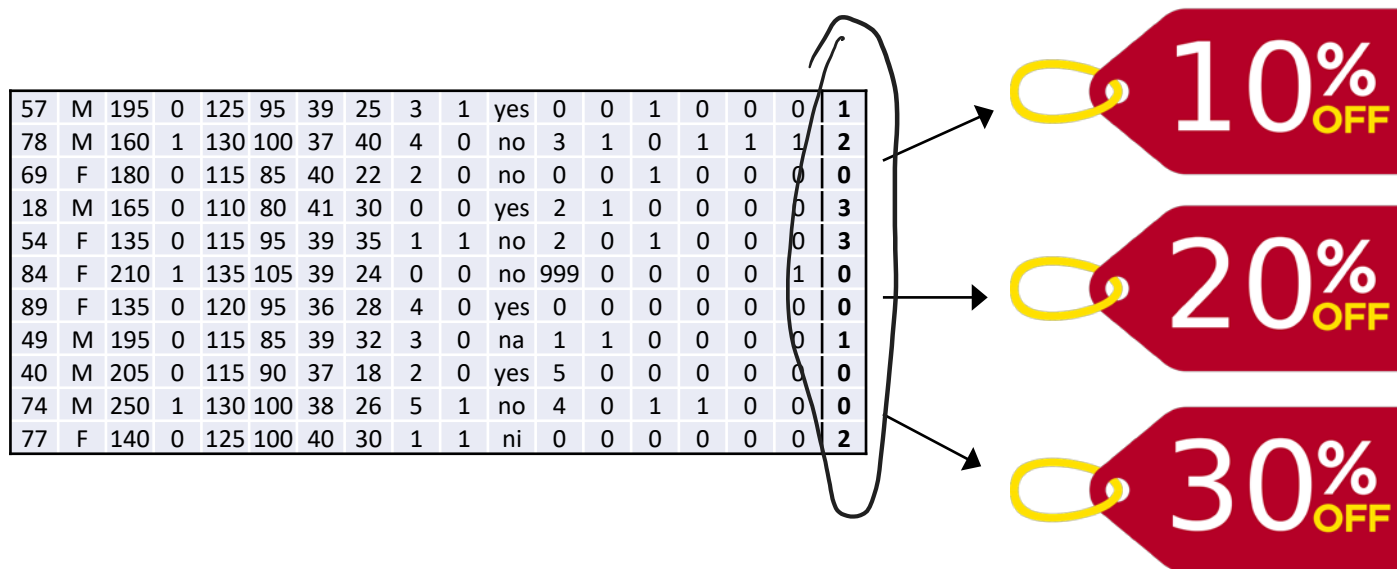
[Machine Learning is the] field of study that gives the computer the ability to learn without being explicitly programmed.

Arthur Samuel, 1959



Überwachtes Lernen (Supervised Learning)

1. Ziel: Vorhersage einer Klasse
2. Trainingsdaten mit Features und einem "Label" (Target Feature)
3. Label kann eine Klasse oder ein numerischer Wert (Regression) sein.
4. Modell soll möglichst gut die tatsächliche Klasse vorhersagen.

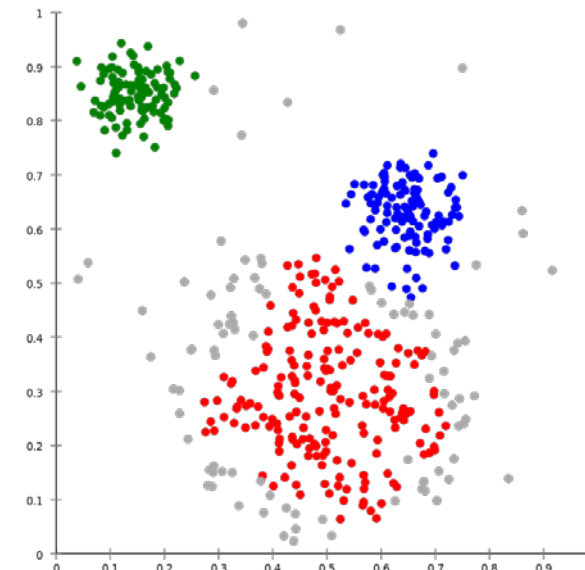


Überwachtes Lernen (Supervised Learning)

Unüberwachtes Lernen (Unsupervised Learning)

1. Ziel: Ähnlichkeiten zwischen Daten ermitteln
2. Trainingsdaten mit Features, jedoch **keinem "Label"**
3. Modell soll möglichst gut Ähnlichkeiten identifizieren.

57	M	195	0	125	95	39	25	3	1	yes	0	0	1	0	0	0
78	M	160	1	130	100	37	40	4	0	no	3	1	0	1	1	1
69	F	180	0	115	85	40	22	2	0	no	0	0	1	0	0	0
18	M	165	0	110	80	41	30	0	0	yes	2	1	0	0	0	0
54	F	135	0	115	95	39	35	1	1	no	2	0	1	0	0	0
84	F	210	1	135	105	39	24	0	0	no	999	0	0	0	0	1
89	F	135	0	120	95	36	28	4	0	yes	0	0	0	0	0	0
49	M	195	0	115	85	39	32	3	0	na	1	1	0	0	0	0
40	M	205	0	115	90	37	18	2	0	yes	5	0	0	0	0	0
74	M	250	1	130	100	38	26	5	1	no	4	0	1	1	0	0
77	F	140	0	125	100	40	30	1	1	ni	0	0	0	0	0	0

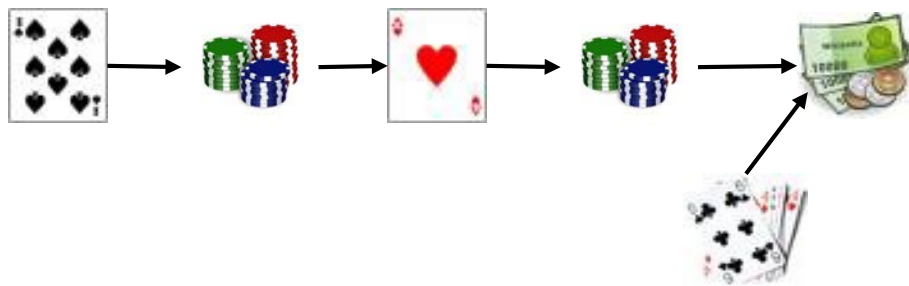


Überwachtes Lernen (Supervised Learning)

Unüberwachtes Lernen (Unsupervised Learning)

Bestärkendes Lernen (Reinforcement Learning)

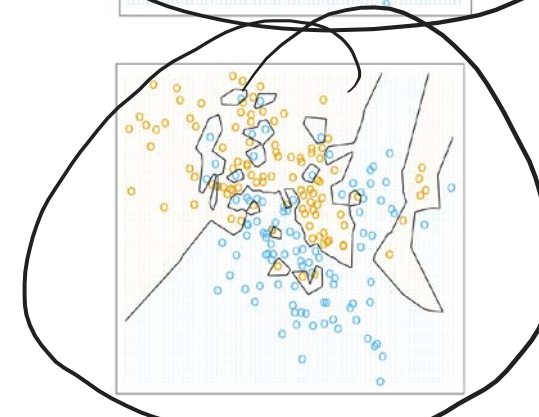
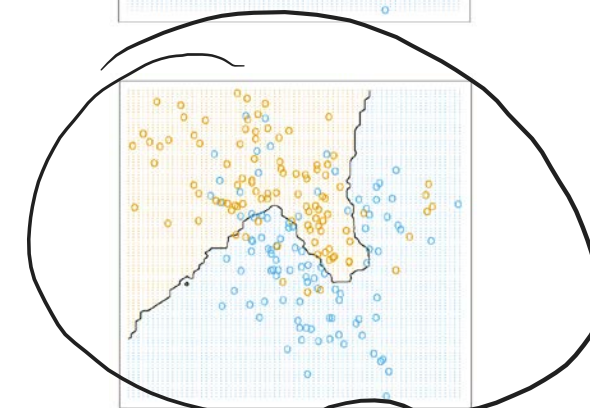
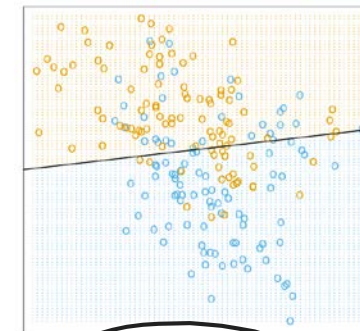
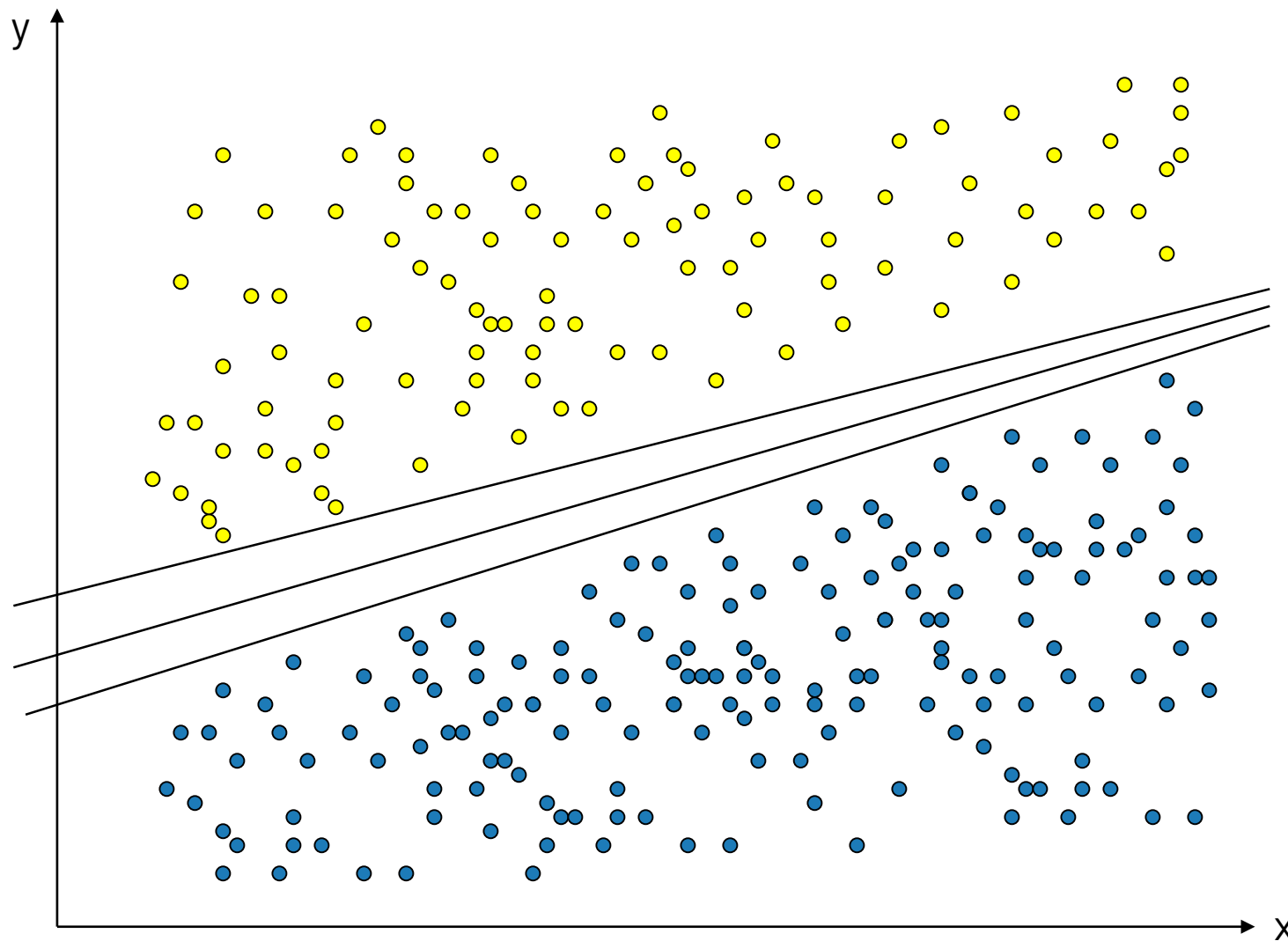
1. Ziel: Lerne Strategie, um erhaltene Belohnungen zu maximieren.
2. Indirektes Feedback für die Qualität einer Folge von Aktionen
3. Keine direkte Antwort, sondern nur besser oder schlechter
4. Feedback erfolgt zeitlich verzögert.



CADE METZ BUSINESS 02.01.17 07:00 AM

**INSIDE LIBRATUS, THE POKER
AI THAT OUT-BLUFFED THE
BEST HUMANS**

Welche Funktion trennt die Bereiche am Besten?



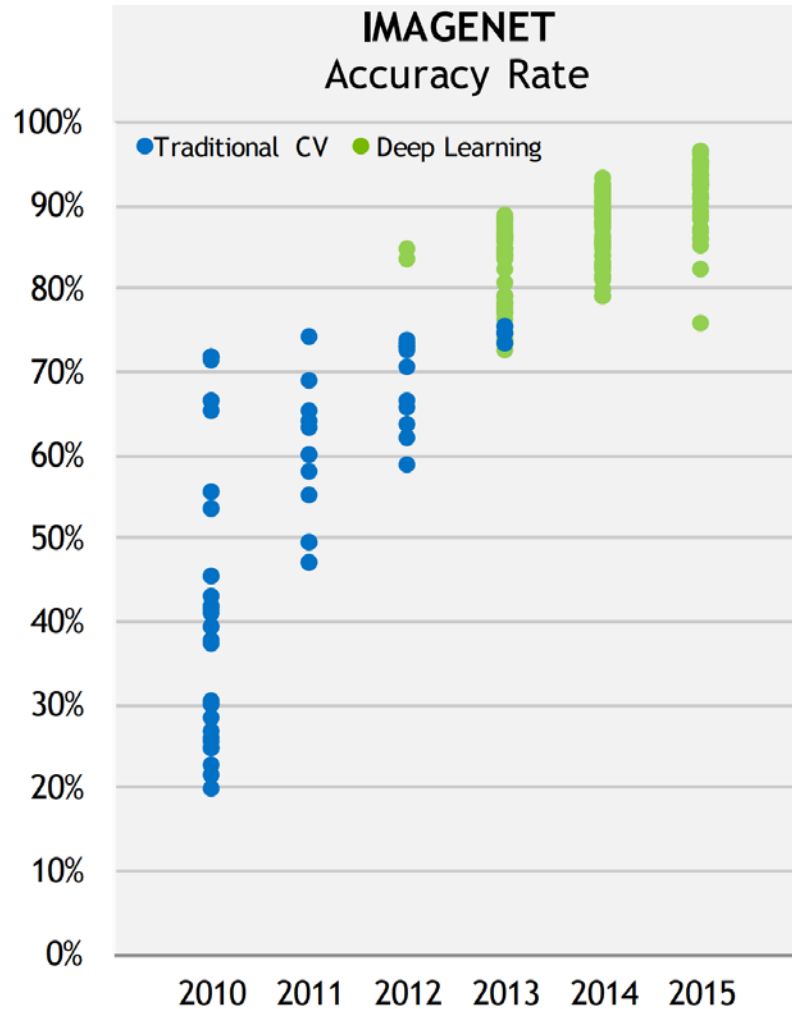


Machine Learning 101

Trends

Herausforderungen

Deep Learning Revolution



<https://devblogs.nvidia.com/wp-content/uploads/2017/05/SXM2-VoltaChipDetails.png>

Generative adversarial networks (GANs)



Zebras ↔ Horses



zebra → horse



winter Yosemite



summer Yosemite

Input winter image



AI-generated summer image



Input sunny image



AI-generated rainy image



GANs: From faces to (small) full-body synthesis



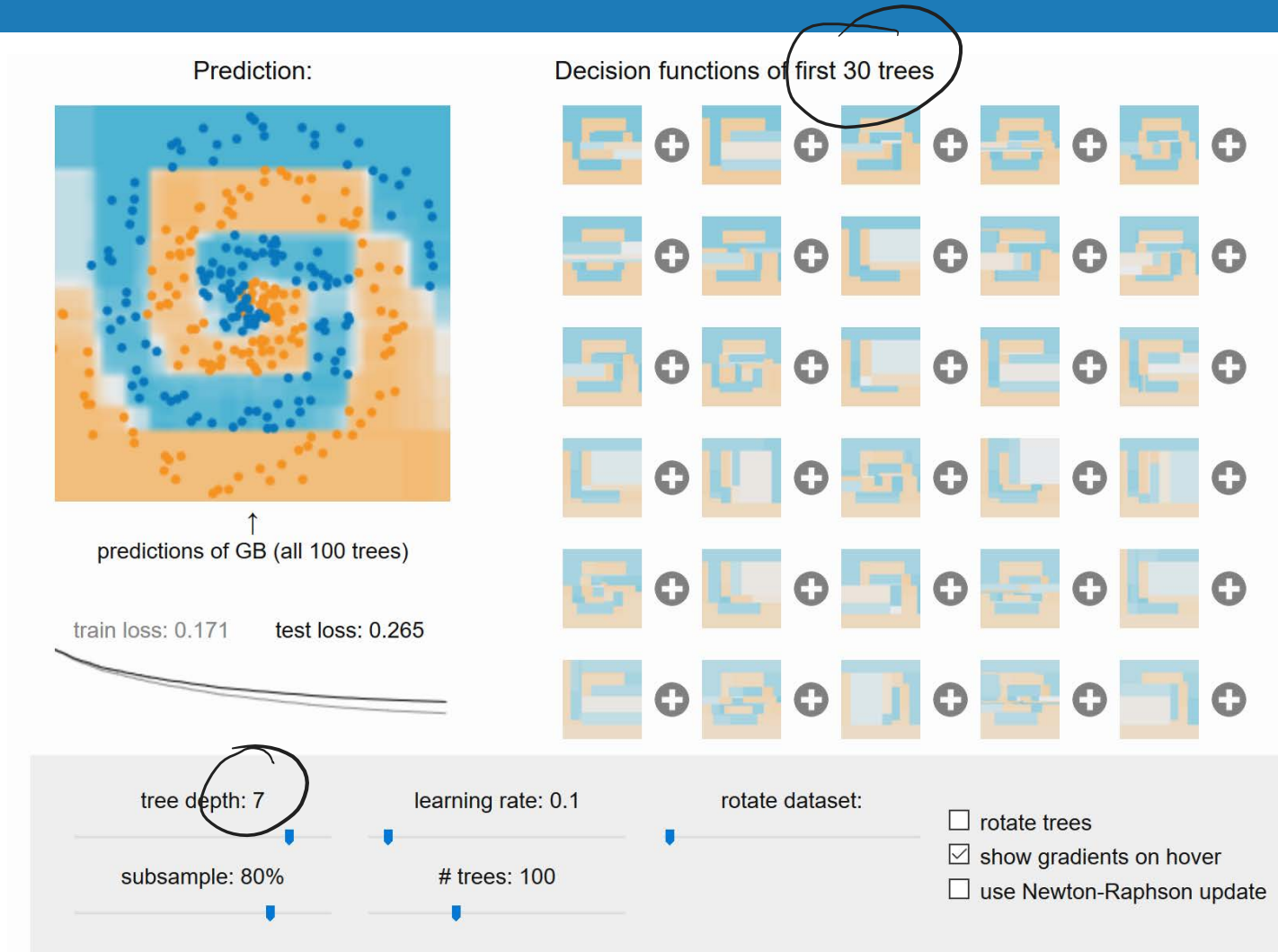
synthesia



 DataGrid

Gruppen sind klüger als Einzelne!?

Gradient Boosted Decision Trees



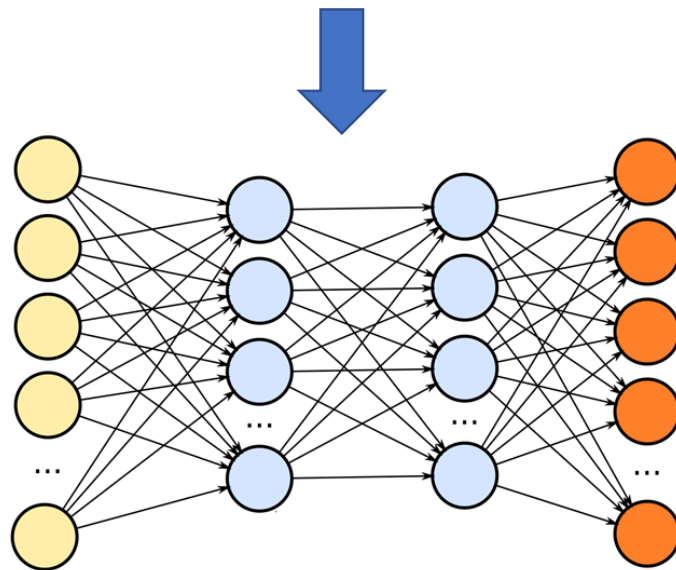
The most frequently used tool by data science competition winners: XGBoost

Ron Bekkerman:

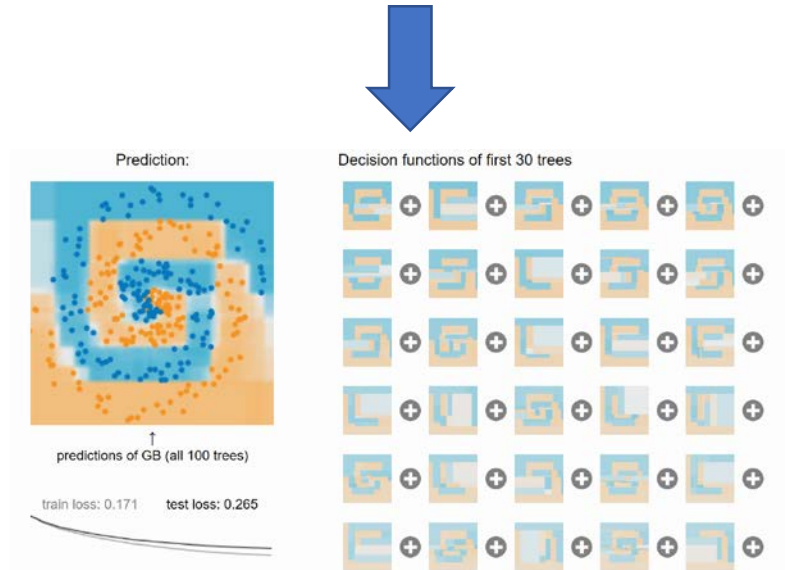
“Something dramatic happened in Machine Learning over the past couple of years. It is called XGBoost – a package implementing **Gradient Boosted Decision Trees that works wonders in data classification**. Apparently, every winning team used XGBoost, mostly in ensembles with other classifiers. Most surprisingly, the winning teams report very minor improvements that ensembles bring over a single well-configured XGBoost.”

Unstrukturierte Daten: Deep Learning

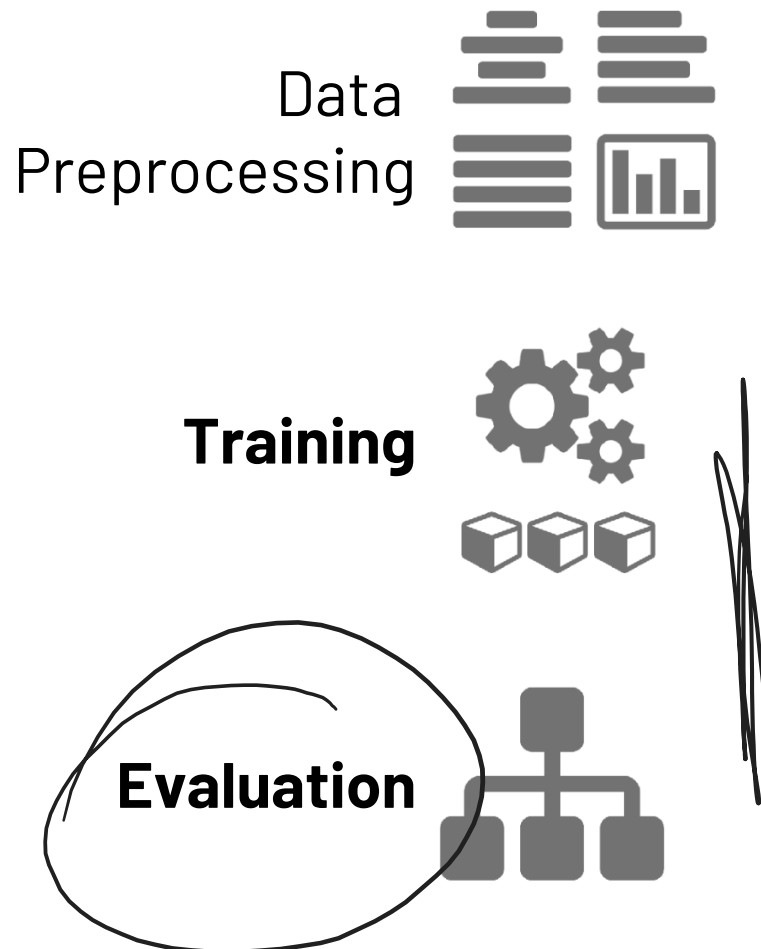
Strukturierte Daten: Gradient Boosting Trees und weitere



57	M	195	0	125	95	39	25	3	1	yes	0	0	1	0	0	0	1
78	M	160	1	130	100	37	40	4	0	no	3	1	0	1	1	1	2
69	F	180	0	115	85	40	22	2	0	no	0	0	1	0	0	0	0
18	M	165	0	110	80	41	30	0	0	yes	2	1	0	0	0	0	5
54	F	135	0	115	95	39	35	1	1	no	2	0	1	0	0	0	3
84	F	210	1	135	105	39	24	0	0	no	999	0	0	0	0	1	0
89	F	135	0	120	95	36	28	4	0	yes	0	0	0	0	0	0	0
49	M	195	0	115	85	39	32	3	0	na	1	1	0	0	0	0	1
40	M	205	0	115	90	37	18	2	0	yes	5	0	0	0	0	0	0
74	M	250	1	130	100	38	26	5	1	no	4	0	1	1	0	0	0
77	F	140	0	125	100	40	30	1	1	ni	0	0	0	0	0	0	4

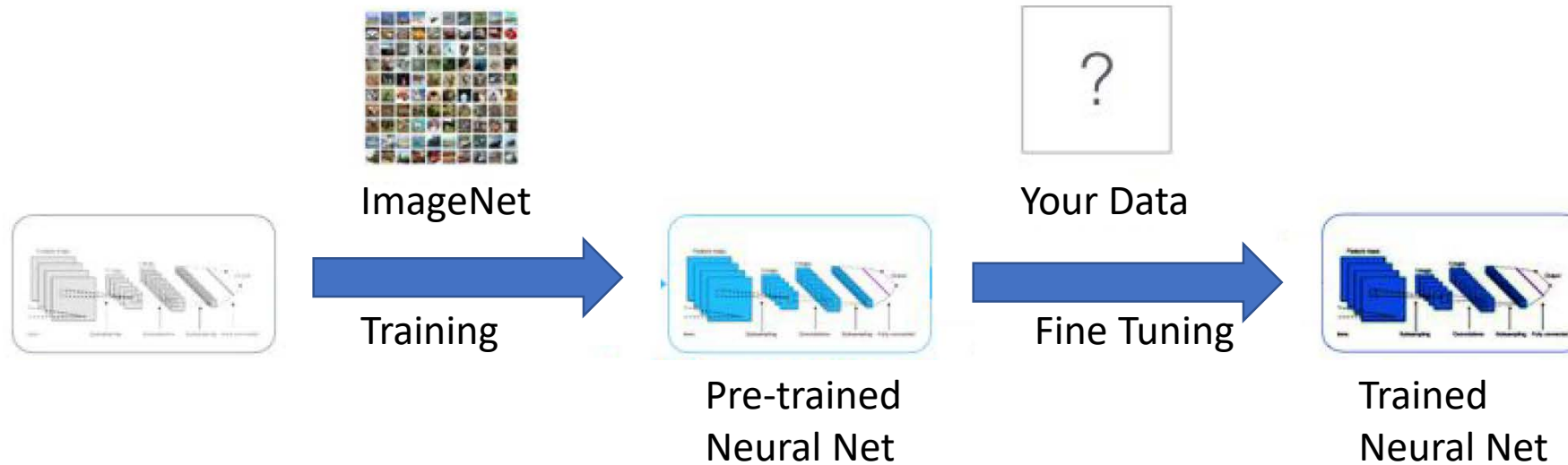


Training und Evaluation von verschiedenen ML-Verfahren ist zeitaufwändig.



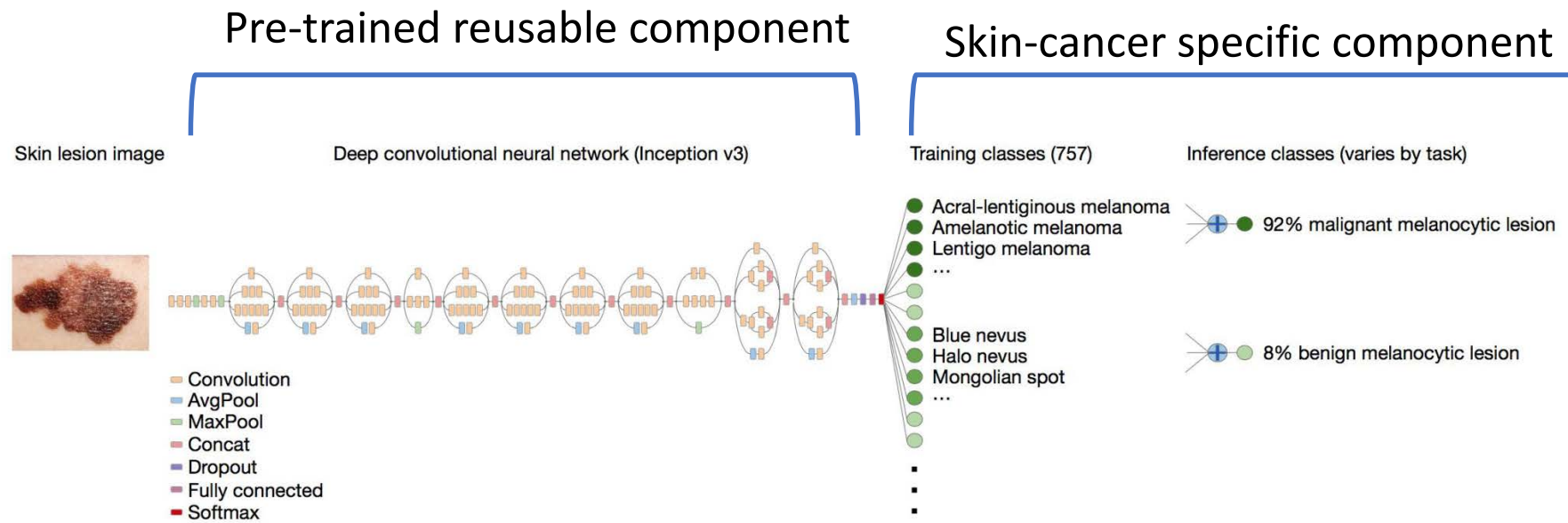
Automated Machine Learning

- automatisierte Verfahren zur
 - Modellauswahl und
 - Hyperparameteroptimierung
 - NICHT automatisiertes ML auf „Knopfdruck“
- › Time-to-Value
- › Kostenreduktion
- › Standards und Best Practices
- › Fokus auf Fragestellung



- Wiederverwendung von vorhandenen Modellen reduziert Zeit und Datenmenge, die ein Modell benötigt, um eine neue Aufgabe zu lernen.
- Der Aufwand zur Erklär- und Nachvollziehbarkeit eines Modells, das aus wiederverwendeten Komponenten besteht, wird reduziert.
- wichtiger Schritt zur Verallgemeinerung eines Modells

Repurposing Google's InceptionV3 image recognition network for skin cancer detection



- 2018/2019 **beeindruckende** Neuerungen in der Sprachverarbeitung (Natural Language Processing, NLP)
- **Textgenerierungsmodell von OpenAI**
Aus ethischen Gründen entschieden sich Entwickler, den Code nicht zu veröffentlichen.
- Dezember 2018 Veröffentlichung von **BERT** durch Google
„NLP-Image-Net-Moment“

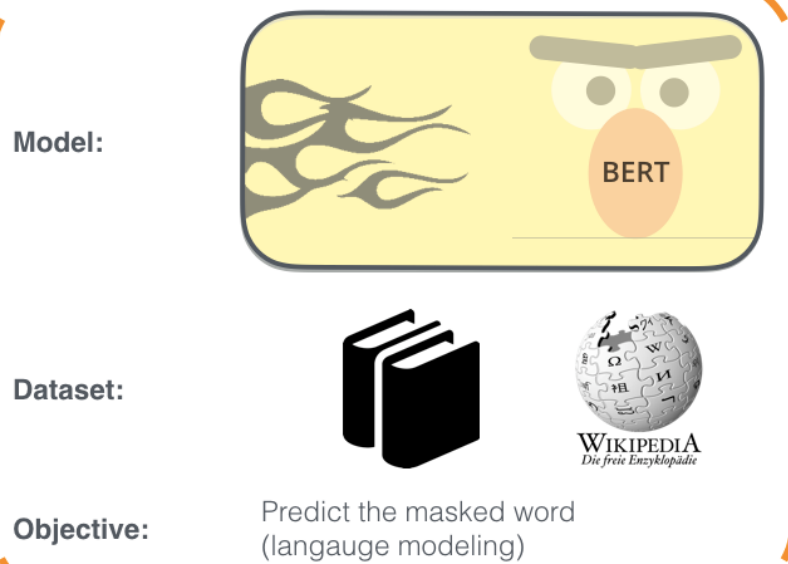
BERT

Bidirectional Encoder Representations from Transformers

1 - **Semi-supervised** training on large amounts of text (books, wikipedia..etc).

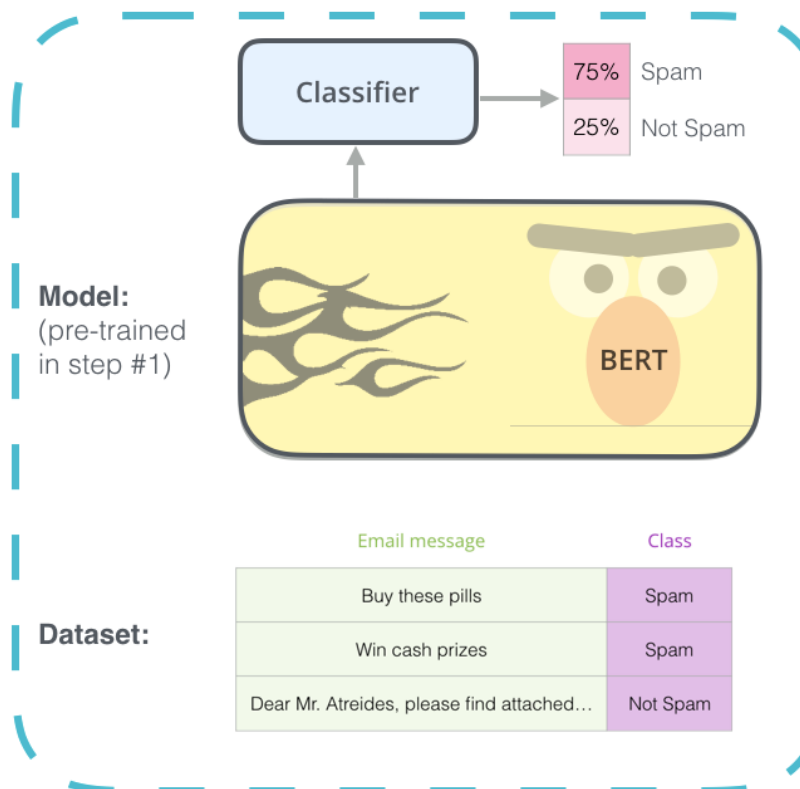
The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step



2 - **Supervised** training on a specific task with a labeled dataset.

Supervised Learning Step



NLP-Benchmarks:

- General Language Understanding Evaluation GLUE
GLUE-Score: 80,4%
7,6% Verbesserung gegenüber dem bisherigen besten Ergebnis.
- SQuAD 1.1
Genauigkeit von 93,2%



Machine Learning 101

Trends

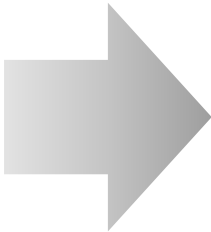
Herausforderungen



Lab to Fab

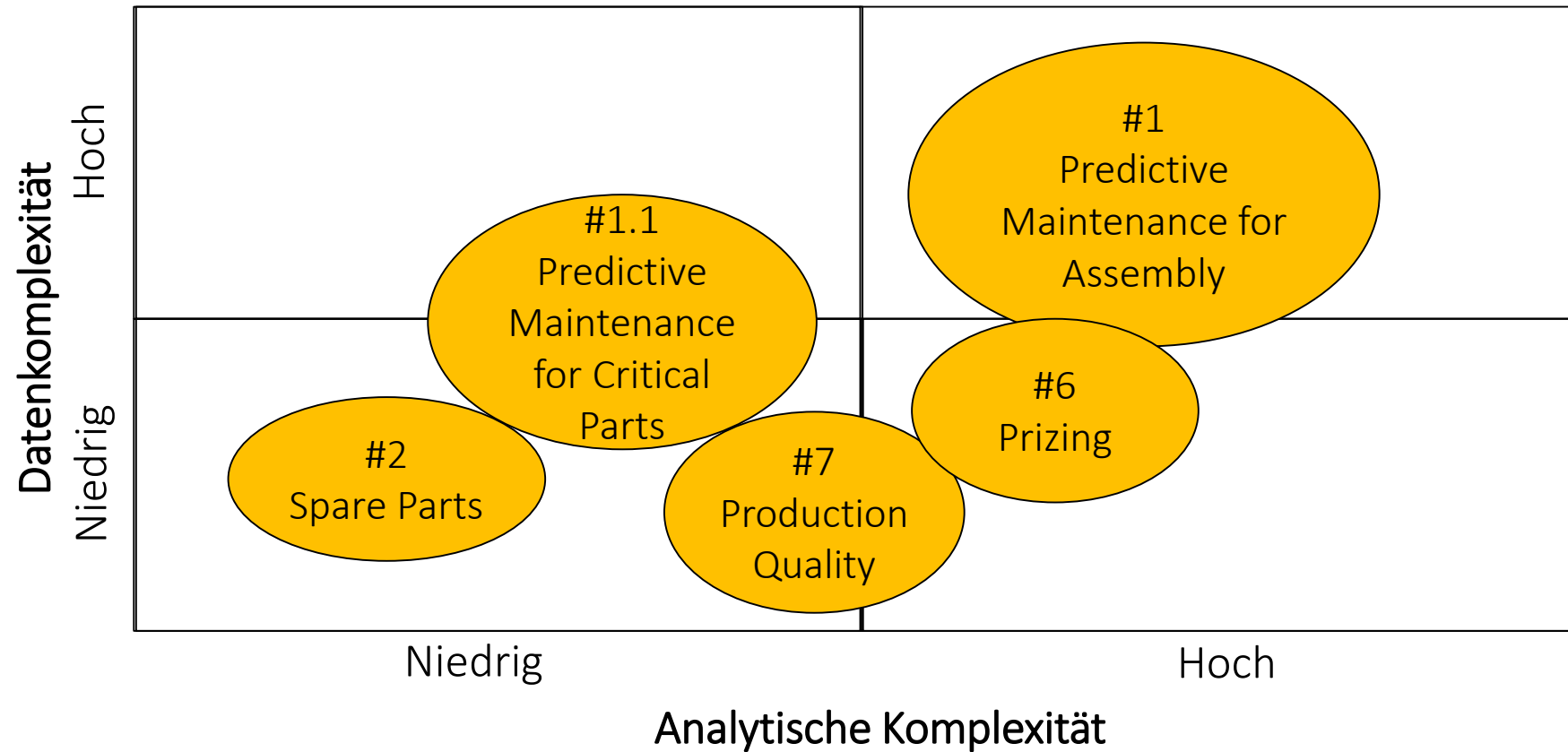


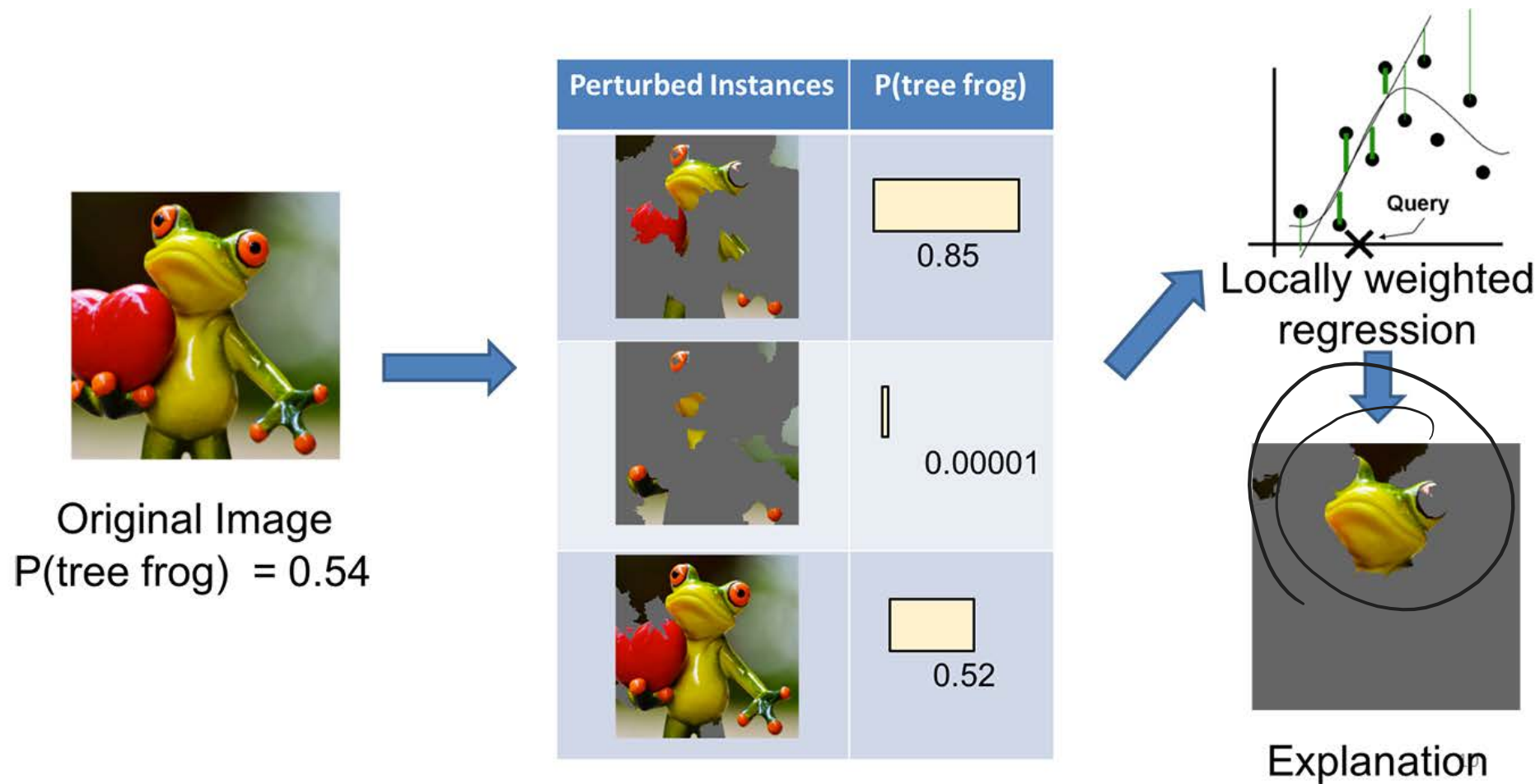
INSTITUTE FOR MACHINE
LEARNING AND ANALYTICS



Zielsetzung und Roadmap klären

No business case – no value generation






Prominent incidents that highlight the effect of algorithmic bias

- **2009** Hewlett-Packard investigates instances of so-called "racist camera software" which had trouble recognizing dark-skinned people
- **2015** A Carnegie Mellon University study determines that some personalized ads from sites such as Google and Facebook are gender-biased
- **2015** A Google algorithm mistakenly captions photos of black people as "Gorillas"
- **2016** Microsoft shuts down AI chatbot Tay after it starts using racist language
- **2016** ProPublica investigation finds that a computer program used to track future criminals in the US is racially biased
- **2016** Machine-learning algorithms used to judge an international beauty contest displays bias against dark-skinned contestants

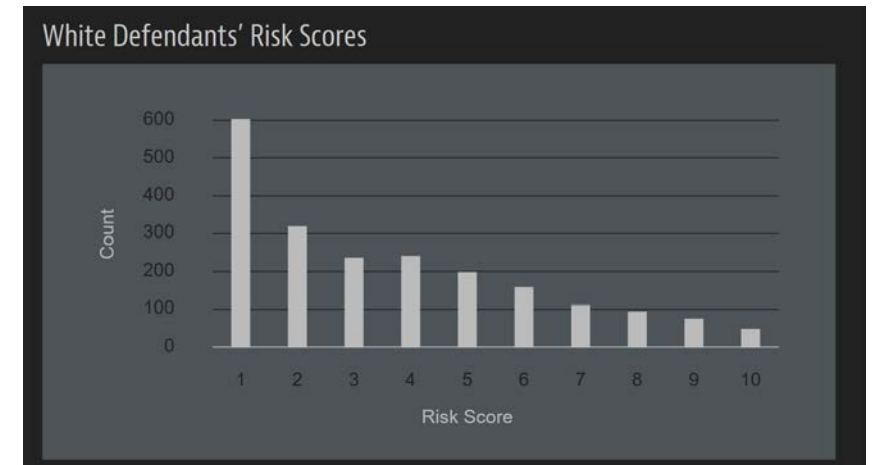
Two Petty Theft Arrests



VERNON PRATER BRISHA BORDEN

LOW RISK 3 **HIGH RISK 8**

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.



<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Fairness, keine Verzerrung, Robustheit und Reproduzierbarkeit!



Forschung

- Transfer Learning
- Automated ML
- Erklärbarkeit und Nachvollziehbarkeit

Unternehmen

- Datenverfügbarkeit und Datenqualität
- Fokus: Zielsetzung, Rollen, Kommunikation
- Lab – Factory
- Automatisierung
- Talente

Gesellschaft

- Verzerrung, Diskriminierung
- Kontrolle von Algorithmen
- Erwartungshaltung
- Datenschutz