

Advanced Machine Learning

Unbekannte Muster und Gesetzmäßigkeiten in Daten erkennen

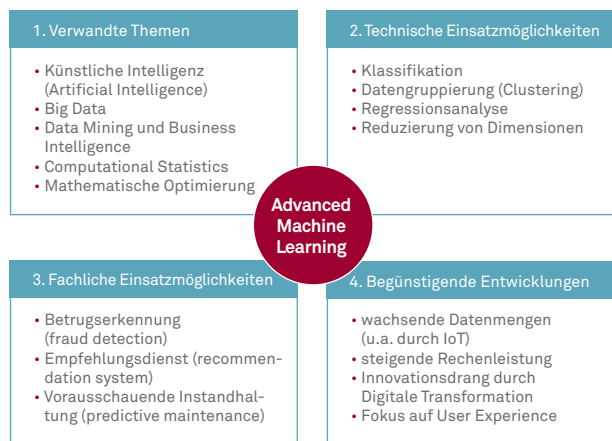
Befeuert durch den Big Data-Trend und die im Übermaß verfügbare Rechenleistung hat Machine Learning den Schritt zum produktiven Einsatz geschafft. Anwendungsfelder wie Prognosemodelle, Bilderkennung und Sprachanalyse, für die sich aufgrund der Komplexität und Unschärfe kaum Algorithmen entwickeln lassen, profitieren von Machine Learning.

Definition

Als Teildisziplin von Artificial Intelligence bezieht sich Machine Learning auf Systeme, die aus beispielhaften oder historischen Daten eigenständig Erkenntnisse gewinnen und diese Erkenntnisse auf andere Daten anwenden können. Die dabei zum Einsatz kommenden Machine Learning-Algorithmen stammen zwar zum Teil noch aus den 1960er Jahren. Allerdings existieren erst jetzt Systeme, die Daten in so großen Mengen bereitstellen und verarbeiten können, dass sich Machine Learning in der Breite praktisch einsetzen lässt.

Der Machine Learning-Einsatz bietet sich an, wenn konkrete fachliche Algorithmen und dedizierte Lösungen nicht mehr funktionieren, wenn die Entwicklung entsprechender Algorithmen nicht mehr mit vertretbarem Aufwand möglich ist, wenn Komplexität und Unschärfe zu groß werden, wenn die Zusammenhänge zwischen Eingabe und Ausgabe nicht verstanden sind oder wenn es zu starken individuellen Unterschieden kommt.

Vor einem sinnvollen Machine Learning-Einsatz steht eine penible Datenaufbereitung. Ein Teil der vorbereiteten Daten dient etwa beim Supervised Learning dazu, das Machine Learning-Modell zu trainieren, der andere Teil dieser Daten hingegen der Validierung des Modells. Erst ein entsprechend verifiziertes Modell lässt sich produktiv einsetzen. Gute Machine Learning-Systeme sind zudem in der Lage, die Modelle kontinuierlich selbst zu verbessern und Effizienz sowie Wirksamkeit der Modelle zu steigern.



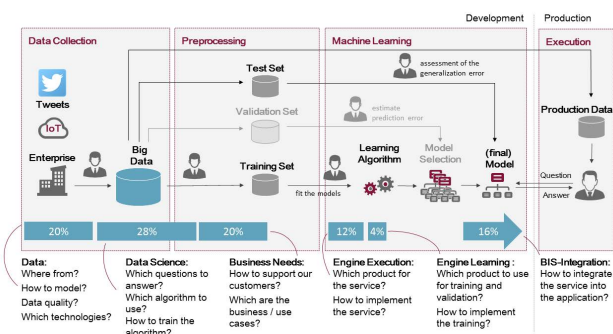
Der Datenvorauswahl und der Zusammenstellung kommt eine besondere Bedeutung zu, weil fehlende Informationen und Unklarheiten in den Trainings- und Validierungsdaten zu stark verfälschten Ergebnissen im Produktiveinsatz führen würden. Ergebnisfehler müssen beim Einsatz von Machine Learning aber dennoch akzeptabel sein und lassen sich höchstens minimieren, nicht vermeiden. Obwohl die Datenvorverarbeitung entscheidend ist, gehört Data Mining nicht zur Disziplin des Machine Learnings.

Referenzszenario

Um Ausfälle der Produktionsanlagen zu vermeiden, möchte ein Unternehmen einen Algorithmus entwickeln, der nach bestimmten Auffälligkeiten sucht und frühzeitig Alarm schlägt. Ein Techniker soll dann die Anlage warten, um einem teuren Ausfall zuvorzukommen.

Die Entwicklung des Algorithmus stellt sich allerdings als zu komplex heraus. Es müssen nicht nur spezifische Merkmale und Abhängigkeiten in den Messwerten erkannt werden, je nach Produktionsanlage variieren die Merkmale auch noch sehr stark.

Das Unternehmen nutzt stattdessen Machine Learning, um Predictive Maintenance in die Praxis umzusetzen. Es kann



dabei auf historische Messwerte der Produktionsanlage zurückgreifen, die erfasst, archiviert und qualifiziert wurden. Das auf Basis dieser Daten trainierte Machine Learning-Modell überprüft künftig in Echtzeit die Messwerte der Anlagen und informiert rechtzeitig die Wartungstechniker.

Dabei nimmt das Unternehmen in Kauf, einen Techniker aufgrund der unvermeidbaren Fehlerwahrscheinlichkeit des Machine Learning-Modells unnötig loszuschicken.

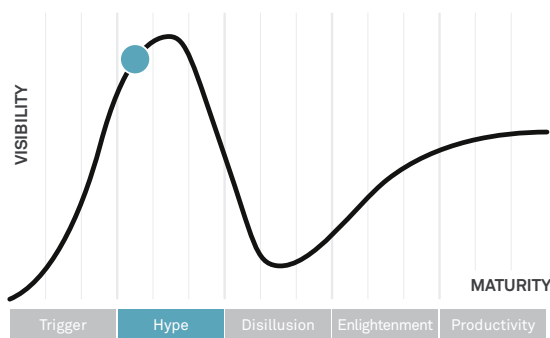
Business Impact

Mit Machine Learning lassen sich ganz neue Ideen in Geschäftsanwendungen umsetzen oder bestehende Szenarien hinsichtlich der User Experience merklich verbessern. Gleichzeitig lassen sich mit Machine Learning Prozesse optimieren sowie Prognosen und Entscheidungen automatisieren.

In den bei Unternehmen vorhandenen großen Datenmengen lassen sich mit Hilfe von Machine Learning auch komplexe Zusammenhänge entdecken und für Anwendungen nutzbar machen. Hieraus entstehen häufig auch innovative Ideen im Kontext der Digitalen Transformation.

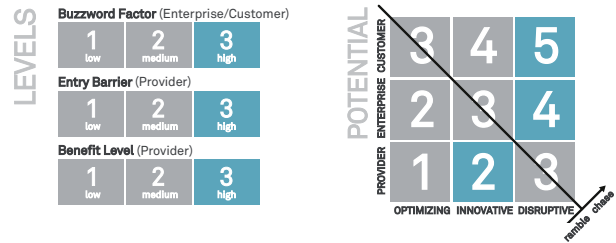
Reifegrad

Verschiedene als Open Source verfügbare Bibliotheken ermöglichen es, Machine Learning-Algorithmen theoretisch in jede beliebige Anwendung zu integrieren. In der Praxis stellt sich allerdings nicht nur die Integration selbst als Hürde dar. Auch der Aufwand für die fachliche und technische Analyse sowie für die Aufbereitung der Datenbasis wird mitunter erheblich unterschätzt. Selbst vermeintlich fertige Machine Learning-Services großer Anbieter müssen die Unternehmen für jeden Anwendungsfall einzeln mit passenden Trainingsdaten anlernen und validieren. Das hierfür notwendige theoretische Wissen und vor allem die praktische Erfahrung im Geschäftsumfeld ist bei vielen Unternehmen kaum vorhanden.



Marktübersicht

Der Markt um Machine Learning wächst stetig. Im Bereich der Bibliotheken gehören TensorFlow, Apache Spark MLlib, Apache Singa, Accord.NET, Torch, OpenCV, Caffe, Shogun, SpaCy und Weka zu den bekanntesten Vertretern. Zu den kostenpflichtigen Lösungen in Form von Machine-Learning-as-a-Service zählen Amazon Web Services, Google Cloud Platform, IBM Bluemix und Microsoft Azure.



Alternativen

Spüren Zeit und Aufwand keine Rolle, dann lassen sich die fachlichen Anforderungen auch in Form eines individuellen Algorithmus ausprogrammieren. Schlussendlich sind auch mitunter fehleranfällige Heuristiken eine Alternative zu Algorithmen und Machine Learning-Modellen.

Pro	Contra
Neue fachliche Lösungen möglich	Moralische, rechtliche und Datenschutzaspekte
Bibliotheken, Verfahren und Dienste sind ausgereift	Auswahl der Verfahren und die Datenmodellierung hochgradig individuell
Praktikabel auf aktuellen Geräten	Technologiezoo
	Bibliotheken zum Teil Bastellösungen
	eingeschränkt nutzbare Dienste