

Einblicke in die Arbeitswelt eines Data Scientist

Oliver Baumfeld, Data Scientist¹

„In the early days of artificial intelligence, the field rapidly tackled and solved problems that are intellectually difficult for human beings but relatively straightforward for computers—problems that can be described by a list of formal, mathematical rules. The true challenge to artificial intelligence proved to be solving the tasks that are easy for people to perform but hard for people to describe formally—problems that we solve intuitively, that feel automatic, like recognizing spoken words or faces in images.“

- from „Deep Learning“ by Goodfellow et al.
Siehe <https://www.deeplearningbook.org/>

+

Die Begriffe Data Science und Artificial Intelligence haben in den letzten Jahren sehr viel Aufwind erlebt, befeuert durch bemerkenswerte Erfolge im wissenschaftlichen Bereich. So sind Algorithmen mittlerweile in der Lage manche einfachen kognitive Aufgaben ähnlich gut wie Menschen durchzuführen. Die technischen Errungenschaften sind aber noch nicht im Business Bereich flächendeckend angekommen.

Es ist unsere Aufgabe die technischen Errungenschaften der letzten Jahre in einen Businesskontext zu übersetzen und für Unternehmen nutzbar zu machen.

Nicht alles was in den Kundenprojekten methodisch eingesetzt wird ist neu. Neu ist ...

- ... das Bewusstsein, man könnte mit den Daten doch etwas machen.
- ... die Menge an Daten, die erzeugt werden, verknüpft und verfügbar sind.
- ... die große Anzahl an Werkzeugen die für die Datenanalyse zu Verfügung stehen, sowohl von den großen Plattformbetreibern (Google, Microsoft, IBM, Amazon, ...) als auch im Open Source Bereich (R, Python, ...).

Was ist Data Science?

Data Science ist ein Bereich, der sich aus Statistik und Computer Science heraus entwickelt hat, es gibt keine allgemeingültige Definition. In Zukunft wird sich der Beruf in eine feinere Rollenaufteilung ausdifferenzieren (Data Engineer, Machine Learning Expert, Translator, ...).

„Data Science is an **interdisciplinary** field [...] to **extract knowledge** or insights from data.“

- Wikipedia

„Data Science is the study of the **generalizable** extraction of knowledge from data.“

- Dhar, V. (2013)

„Bei Data Science Projekten geht es um die Beantworten von Fragen, oder Lösen von Problemen, mit einem **messbaren Ergebnis**, durch die Verarbeitung von Daten.“

- Christof Stocker (mein Kollege)

Interdisziplinär: Unterschiedliche Experten müssen zusammenkommen (Mathematik, Computer Science, Fachbereich, Data Scientist, IT, Business, ...).

Generalisierbar: Die gewonnenen Erkenntnisse müssen auf zukünftige Situationen anwendbar sein.

Messbar: Die Erkenntnisse müssen verifizierbar sein damit es bei der Umsetzung in die Praxis keine Überraschungen gibt.

¹ oliver.baumfeld@kapsch.net, <https://www.linkedin.com/in/oliver-baumfeld-95829099>

Anwendungsbeispiele

Kundenkontakt

- Automatisches Bearbeiten von eingehenden Kundenanfragen z.B.: für Versicherungen
 - o Emails, Abrechnungen, Arztbriefe, ...
 - o Klassifikation der Dokumente
 - o Extraktion von Informationen
- Aufbereitung von Informationen für Sachbearbeiter z.B. beim Service Desk
 - o Was ist bisher geschehen?
 - o Antwortvorschläge

Industrie

- Prozessoptimierung, optimale Wartung von Maschinen
- Erkennen fehlerhafter Werkstücke um Ausschüsse zu reduzieren und die Produktqualität sicher zu stellen.
- Anbindung aller Maschinen für eine Prozessübersicht. Viele Unternehmen haben Datensilos, daher ist bei Ihnen keine Datenanalyse über den Gesamtprozess möglich.

Herausforderungen bei der Umsetzung von Projekten

- Die **Unwissenheit** ist ein inhärentes Problem: Was ist in den Daten? Diese Frage kann erst beantwortet werden, wenn mit den Daten gearbeitet wird. Die Datenqualität ist ein kritischer Faktor, die explorative Datenanalyse ein wichtiger erster Schritt.
- **Unstimmigkeiten** zwischen Fachbereichen und IT. Interne Machtkämpfe und Unternehmenspolitik können Data Science Projekte torpedieren.
- **Fehlende Priorisierung** durch das Management, dadurch keine Konfliktlösung und fehlende Ressourcen beim Kunden für das Projekt.
- Management der **Erwartungshaltung**: Medienhype/Marketing vs. Realität. Es braucht Transparenz in der Umsetzung.
- **Auswirkungen** auf die Organisation und Prozesse eines Unternehmens

Produkt vs. Dienstleistung

Es gibt vereinfacht gesagt 2 Arten von AI-Unternehmen: Produktentwickler und Dienstleister. Wir sind Dienstleister und beschäftigen uns ständig mit neuen Themen, von denen wir selbst wenig Ahnung haben. Wir brauchen daher Mitarbeiter, die die Kreativität mitbringen um allgemeine Problemlösung zu betreiben (Generalisten) und Methoden, die es uns erlauben mit geringem Risiko durch die unsicheren Projektphasen zu navigieren.

Iterative, agile Vorgehensweise



Wir sind selbst ständig auf der Suche unsere Vorgehensweisen zu optimieren. Ein Problem ist, dass der Kunde oft die Vorgehensweise vorgibt (er ist das auch so gewohnt). Es stellt sich für uns daher die Frage, wie überzeuge ich den Kunden von unserer Vorgehensweise? Unsere Lösung ist mehr auf eine **partnerschaftliche** Zusammenarbeit zu setzen statt auf klassische Auftraggeber vs. Auftragnehmer Beziehungen. Das braucht viel Vertrauen, das erst aufgebaut werden muss. Daher setzen wir auf eine enge Kommunikation mit dem Kunden zum **Managen der Unsicherheiten und Erwartungshaltungen**.